

## Prediction Skill and Bias of Tropical Pacific Sea Surface Temperatures in the NCEP Climate Forecast System Version 2

YAN XUE, MINGYUE CHEN, ARUN KUMAR, ZENG-ZHEN HU, AND WANQIU WANG

*Climate Prediction Center, NOAA, College Park, Maryland*

(Manuscript received 9 August 2012, in final form 22 January 2013)

### ABSTRACT

The prediction skill and bias of tropical Pacific sea surface temperature (SST) in the retrospective forecasts of the Climate Forecast System, version 2 (CFSv2), of the National Centers for Environmental Prediction were examined. The CFSv2 was initialized from the Climate Forecast System Reanalysis (CFSR) over 1982–2010. There was a systematic cold bias in the central–eastern equatorial Pacific during summer/fall. The cold bias in the Niño-3.4 index was about  $-2.5^{\circ}\text{C}$  in summer/fall before 1999 but suddenly changed to  $-1^{\circ}\text{C}$  around 1999, related to a sudden shift in the trade winds and equatorial subsurface temperature in the CFSR.

The SST anomaly (SSTA) was computed by removing model climatology for the periods 1982–98 and 1999–2010 separately. The standard deviation (STD) of forecast SSTA agreed well with that of observations in 1982–98, but in 1999–2010 it was about 200% too strong in the eastern Pacific and 50% too weak near the date line during winter/spring. The shift in STD bias was partially related to change of ENSO characteristics: central Pacific (CP) El Niños were more frequent than eastern Pacific (EP) El Niños after 2000. The composites analysis shows that the CFSv2 had a tendency to delay the onset phase of the EP El Niños in the 1980s and 1990s but predicted their decay phases well. In contrast, the CFSv2 predicted the onset phase of the CP El Niños well but prolonged their decay phase. The hit rate for both El Niño and La Niña was lower in the later period than in the early period, and the false alarm for La Niña increased appreciably from the early to the later period.

### 1. Introduction

Seasonal climate predictions are now routinely made at operational centers using coupled dynamical models (e.g., Ji et al. 1998; Saha et al. 2006; Stockdale et al. 2011; Barnston et al. 2012). In the development of seasonal prediction systems, prediction skill of the tropical Pacific sea surface temperature (SST) anomaly associated with El Niño–Southern Oscillation (ENSO) is commonly used as a benchmark for evaluating progress (Saha et al. 2006; Jin et al. 2008; Stockdale et al. 2011). In this study, we document the prediction skill of ENSO and the biases in the new coupled dynamical model, which is referred to as Climate Forecast System, version 2 (CFSv2; Saha et al. 2012, manuscript submitted to *J. Climate*), implemented at the National Centers for Environmental Prediction (NCEP) in early 2011. Prediction skill and systematic biases of the CFSv2 were

examined based on a set of retrospective forecasts (often referred to as hindcasts) over the 1982–2010 period.

A common practice in the assessment of seasonal prediction skill is the removal of systematic biases estimated based on hindcasts. Systematic biases depend on the growth of model errors and can be initial condition dependent. For example, for SST evolution in the tropical Pacific, the growth rate of initial condition (IC) perturbations typically varies with the seasonal cycle because of the seasonality in the coupled instability (Battisti 1988; Tziperman et al. 1998). This also has an influence on the growth of model errors resulting from mismatches between the physics in the model and in the real world. Because of seasonality in the systematic biases (also referred to as the forecast biases), they are generally calculated as a function of IC month and lead time. A basic assumption in the estimation and removal of systematic biases, however, is that they are stationary with time (over the hindcast and as well as over the real-time forecast period).

Since the memory of the coupled system resides in the ocean, operational seasonal forecast models are generally initialized with operational ocean reanalysis products

---

Corresponding author address: Yan Xue, NOAA/Center for Weather and Climate Prediction, 5830 University Research Court, Room 3005, College Park, MD 20740.  
E-mail: yan.xue@noaa.gov

that combine information from ocean models, atmospheric fluxes, and ocean observations using data assimilation methods (Balmaseda et al. 2009; Xue et al. 2012). The retrospective forecasts of CFSv2 were initialized from a new reanalysis of the atmosphere, ocean, sea ice, and land for 1979–present, referred to as Climate Forecast System Reanalysis (CFSR; Saha et al. 2010). Recent studies documented the characteristics of various CFSR analyzed fields (e.g., equatorial tropical winds and subsurface ocean temperatures) and found that there was an abrupt shift in many of those fields around 1999 (Xue et al. 2011; Zhang et al. 2012). Because of the influence of IC error growth during the forecasts, it is conceivable that the systematic biases in the CFSv2 hindcasts were not stationary, and changes in the characteristics of initial conditions have a fingerprint on forecast biases. This is indeed the case for the CFSv2, which was initialized from the CFSR. Kumar et al. (2012) documented that the characteristics of the forecast bias for SST in the equatorial Pacific had a dramatic change around 1999. They attributed the change in the SST forecast bias to the change in the CFSR from which the ocean initial conditions for hindcasts were taken. Kumar et al. (2012) show that the CFSv2 SST averaged in the Niño-3.4 region ( $5^{\circ}\text{S}$ – $5^{\circ}\text{N}$ ,  $170^{\circ}$ – $120^{\circ}\text{W}$ ) was dominated by negative errors before 1999 that reached  $-2^{\circ}\text{C}$  in summer/fall, but after 1999 the negative errors reduced to about  $-1^{\circ}\text{C}$  in summer/fall, while positive errors of  $+0.5^{\circ}\text{C}$  appeared in late winter and spring (Fig. 1 in Kumar et al. 2012). A consequence of period dependent forecast bias is that if it is not taken into account in the computation of forecast anomalies, it might lead to a reduction in the estimate of skill.

Toward documenting the characteristics of CFSv2 SST forecasts in the tropical Pacific, the first objective of the paper is to show the differences in systematic biases in hindcast SST over two periods, before and after 1999, and how the differences in systematic biases relate to the differences in ICs. The second objective of the paper is to evaluate the variability, prediction skill, and predictability of ENSO in CFSv2 over two periods, 1982–98 and 1999–2010, separately. A two-period approach allows us to avoid the adverse influence of the differences in systematic biases before and after 1999 on the assessment of skill if a single climatology based on the 1982–2010 period is used (Kim et al. 2012).

We compared the ENSO prediction skill between the CFSv1 and CFSv2 and found that the prediction skill of ENSO, quantified by anomaly correlation coefficient (ACC) and root-mean-square error (RMSE) of the Niño-3.4 SST index, is generally lower in CFSv2 than in CFSv1 when anomalies are based on climatology for the common period in hindcast runs (1982–2004). For

example, for Niño-3.4, the ACC in CFSv1 is 0.9 at 5-month lead from May to July ICs, but it is 0.7 in CFSv2. However, when anomalies are based on the climatology for 1982–98 and 1999–2010 separately, the ACC in CFSv2 is slightly higher than that in CFSv1. Since Jin and Kinter (2009) have done a thorough evaluation of the ENSO prediction skill in CFSv1, the current paper focuses on providing a comprehensive evaluation of the prediction skill of ENSO in CFSv2.

In addition to minimizing the influence of period dependent systematic biases on forecast skill, the two-period approach also allows us to examine the change of ENSO variability and predictability that occurred since 2000 (McPhaden 2012; Horii et al. 2012; Xiang et al. 2013; Hu et al. 2013). El Niño events in the 1980s were mostly characterized by maximum warm anomaly in the eastern equatorial Pacific, often referred to as eastern Pacific El Niño (EP event). In contrast, the El Niños in the 1990s and 2000s tended to have largest anomalous warming in the central equatorial Pacific (Larkin and Harrison 2005; Ashok et al. 2007; Kug et al. 2009; Kao and Yu 2009), often referred to as central Pacific El Niño (CP event). There is a suggestion that the CP events are becoming more frequent and more intense in recent decades (Yeh et al. 2009; Lee and McPhaden 2010). It has also been suggested that there is a breakdown of traditional ENSO predictors since 2000 related to a change in the recharge–discharge–SST phase relationship and change in the characteristics of atmospheric intraseasonal forcing (Horii et al. 2012). Associated with the decadal change of ENSO characteristics, the real-time predictions of ENSO during the 2002–11 period have been somewhat less skillful than during the 1980s and 1990s (Wang et al. 2010; Barnston et al. 2012), and it is suggested that model errors limit their capability in distinguishing between the CP and EP events (Hendon et al. 2009; Hu et al. 2012). The third objective of the paper is to address the decadal change of the prediction skill and predictability of ENSO since 2000.

Section 2 introduces the model hindcasts and validation data and methodologies used in our analysis. Section 3 describes the systematic biases in SST hindcasts over two periods and how they relate to an abrupt shift in the CFSR ICs. In section 4, we discuss the decadal change of variability of SST anomaly from 1982–98 to 1999–2010 and how well the model simulates the decadal change of SST variability. In section 5, we provide a comprehensive evaluation of the model prediction skill of ENSO. The questions that we will address are as follows: What are the model deterministic skills of ENSO predictions before and after 1999? What are the gaps between the model forecast skill and persistence forecast skill? What are the probabilistic forecasts of

ENSO? What are the hit rate and false alarm rate for El Niño, ENSO neutral, and La Niña? What are the prospects for predicting the two flavors of El Niño? The summary and discussion are given in section 6.

## 2. Data and methodologies

The retrospective forecasts of CFSv2 were initialized from the CFSR (Saha et al. 2010). In the CFSR, the atmospheric component is the NCEP Global Forecast System (GFS) with a horizontal resolution of T382 (~38 km) and 64 vertical levels extending from the surface to 0.26 hPa. The oceanic component is the Geophysical Fluid Dynamics Laboratory Modular Ocean Model version 4 with 40 levels in the vertical, a zonal resolution of  $0.5^\circ$  and a meridional resolution of  $0.25^\circ$  between  $10^\circ\text{S}$  and  $10^\circ\text{N}$ , gradually increasing to  $0.5^\circ$  poleward of  $30^\circ\text{S}$  and  $30^\circ\text{N}$ . Comprehensive assessments of the salient features in the CFSR, including the ocean variability (Xue et al. 2011), the surface climate (Wang et al. 2011), and troposphere circulation (Chelliah et al. 2011), have been made. An important finding from those studies is that the CFSR has a sudden shift in both atmospheric and oceanic fields near the end of 1998 associated with the use of the advent of Advanced Television Infrared Observation Satellite (TIROS) Operational Vertical Sounder (ATOVS) radiance data (Xue et al. 2011; Zhang et al. 2012). The shift in the CFSR IC also imprinted as shift in systematic bias in retrospective forecasts (Kumar et al. 2012).

The CFSv2 coupled model is similar to that used in the CFSR but with a reduced horizontal resolution at T126 (~100 km) for the atmospheric component (Saha et al. 2012, manuscript submitted to *J. Climate*). For CFSv2 hindcasts, four runs, each for 9-month integration, were made every 5 days starting from December 1981 to December 2010. For each IC month, an ensemble mean of 20 runs initialized prior to the first day of first target month is used in the analysis. For example, forecasts from 11, 16, 21, 26, and 31 January, with four runs from each day, were used to compute ensemble mean forecast for nine subsequent months after January. To improve signal to noise ratio, monthly mean forecasts were further averaged into seasonal mean forecasts. Following this example, forecasts from January ICs for target seasons of February–April, March–May, etc., are referred to as forecasts at the lead time of 0 month, 1 month, and so on. The seasonal mean SST forecast from 0-month lead to 6-month lead is analyzed. Individual forecast members were also used in the assessment of the probabilistic forecast skill of ENSO.

The hindcast SSTs were verified against the weekly optimum interpolation SST analysis version 2 (OIv2) of

Reynolds et al. (2002). For model verification, the weekly OIv2 SST was interpolated onto the model grid and averaged into seasonal mean. Systematic bias in seasonal mean SST forecast is calculated as a function of IC months and lead months for the periods 1982–98 and 1999–2010 separately. After the systematic bias is removed, the model SST anomaly (SSTA) is compared with the OIv2 SSTA, which was derived by removing the climatology for 1982–98 and 1999–2010 separately in a similar fashion as it was done for the forecast SST.

It is worth pointing out again that all the analyses discussed below are done for the periods 1982–98 and 1999–2010 separately. The standard deviation of model SSTA, calculated using individual ensemble members, is compared with that of observed SSTA. The ACC and RMSE between ensemble mean forecast and observed SSTA are calculated as functions of IC months and lead months. The ACC and RMSE of CFSv2 are compared with those of persistence forecast in which the seasonal mean SSTA prior to first target month is persisted for 6 lead months. To quantify potential predictability of SSTA, 1 of 20 ensemble members is treated as the proxy for “observation” and the ensemble mean of other 19 members is validated against the observation. The perfect-model skill refers to the average of 20 ACCs, where each of 20 ensemble members is alternatively treated as observation, and the model ensemble spread refers to the average of 20 RMSEs, where each of 20 ensemble members is alternatively treated as observation. The difference between the model skill and perfect-model skill measures the difference in predictable component in the model and observation world, and to what extent the potential predictability is realized in hindcasts. On the other hand, the differences between the model RMSE and model ensemble spread (ES) measure to what extent ensemble forecast errors can be attributed to internal variability in the model (Stockdale et al. 2011). For an unbiased model, the model ES is approximately the same as the model RMSE. However, for an overconfident (underconfident) model prediction, the model ES is smaller (larger) than the model RMSE.

In addition to the deterministic skills (ACC, RMSE), the probabilistic skill is assessed using 20 ensemble members. The probabilistic forecast for El Niño, ENSO neutral, and La Niña is assessed using the threshold of  $\pm 0.5^\circ\text{C}$  for 3-month running mean Niño-3.4 similar to the National Oceanic and Atmospheric Administration (NOAA) official ENSO definition. Following this definition, the probabilistic forecast for CFSv2 is defined as the percentage of ensemble members out of 20 that meets the threshold of  $\pm 0.5^\circ\text{C}$ . The probabilistic forecast for the 10 El Niños and 10 La Niñas during 1982–2010 based on the NOAA definition

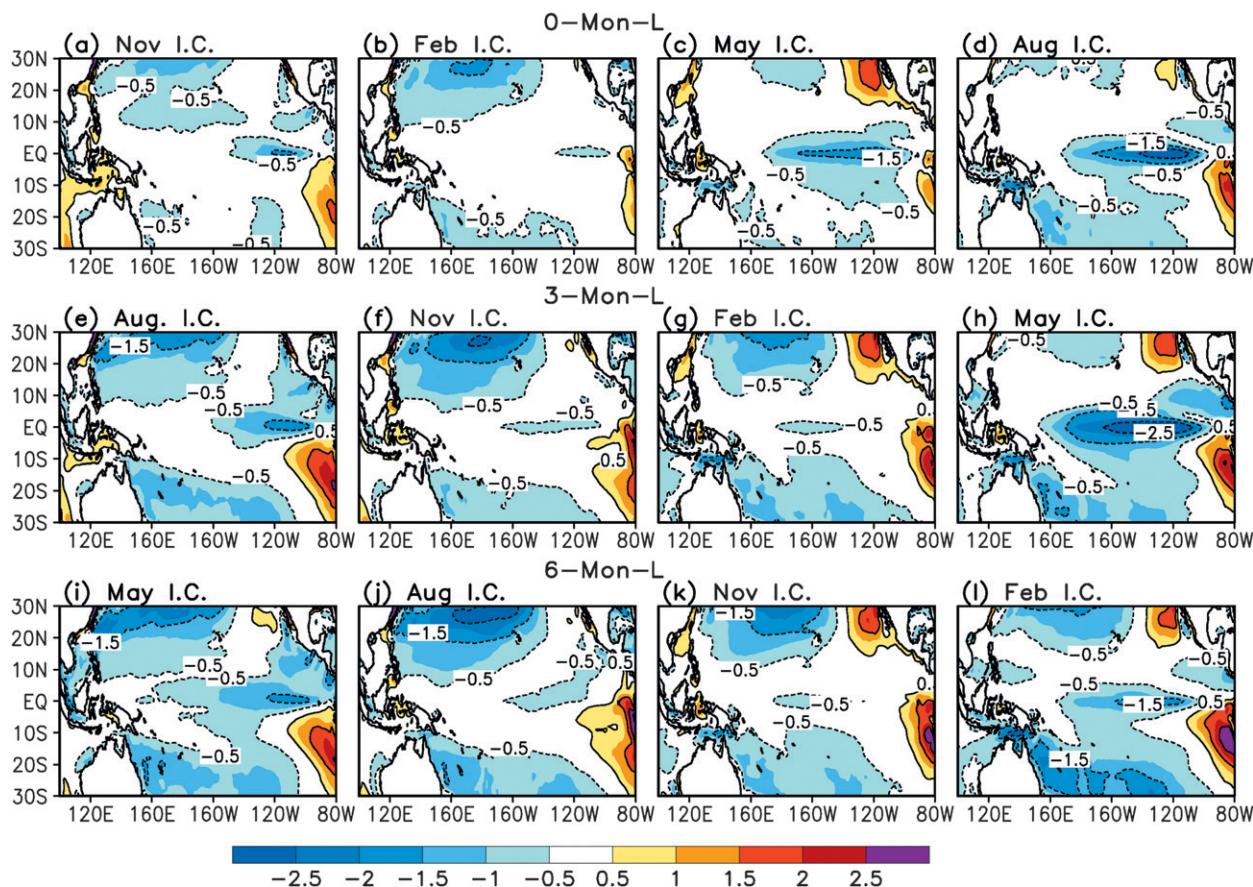


FIG. 1. Spatial patterns of seasonal mean SST forecast departure ( $^{\circ}\text{C}$ ) from OIv2 SST averaged in 1982–98 at (top) 0-, (middle) 3-, and (bottom) 6-month lead for (left to right) 4 target seasons and 4 initial months, labeled on the top of each panel. The contour interval is  $1^{\circ}\text{C}$ .

([http://www.cpc.ncep.noaa.gov/products/analysis\\_monitoring/ensostuff/ensoyears.shtml](http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml)) is assessed.

Another measure of prediction skill is to see how often the model forecasts follow into the correct categories of El Niño, ENSO neutral, and La Niña and how often they do not. The former is the hit rate, and the latter is the false alarm rate. The hit and false alarm rates have been used to assess the status of ENSO forecast skill (Kirtman et al. 2002). To assess the decadal change of ENSO predictability, the hit and false alarm rates are calculated for the periods 1982–98 and 1999–2010 separately.

To assess how well the model forecasts capture the differences in SST anomaly patterns for the EP and CP events, the El Niño composites are calculated. There are different ways to define EP and CP events (see a summary in Singh et al. 2011). We adopted the definition of McPhaden et al. (2011), who identified five EP (1982/83, 1986/87, 1991/92, 1997/98, and 2006/07) and five CP (1987/88, 1994/95, 2002/03, 2004/05, and 2009/10) events in the period 1980–2010 (Fig. 1 in McPhaden et al.

2011). The composites for the EP and CP events were constructed with four EP and four CP events with the 1986/87 and 1987/88 events excluded because of their unique characteristics of seasonal evolution.

### 3. Systematic biases in hindcast SST

The systematic SST biases calculated for ICs from December 1981 to December 1998 are shown at 0-, 3-, and 6-month leads for target seasons December–February (DJF), March–May (MAM), June–August (JJA), and September–November (SON) (Fig. 1). A substantial cold bias ( $-1.5^{\circ}\text{C}$ ) was found in the equatorial central and eastern Pacific at 0-month lead from May and August IC (Figs. 1c,d). The cold bias from May IC was initially around  $-1.5^{\circ}\text{C}$  in the first target season (JJA; Fig. 1c), enlarged to  $-2.5^{\circ}\text{C}$  in the second target season (SON; Fig. 1h), and then reduced in DJF (Fig. 1i). In contrast, the cold bias from August IC peaked in the first target season (SON; Fig. 1d), reduced in DJF (Fig. 1e), and then dissipated in MAM (Fig. 1j). For February IC,

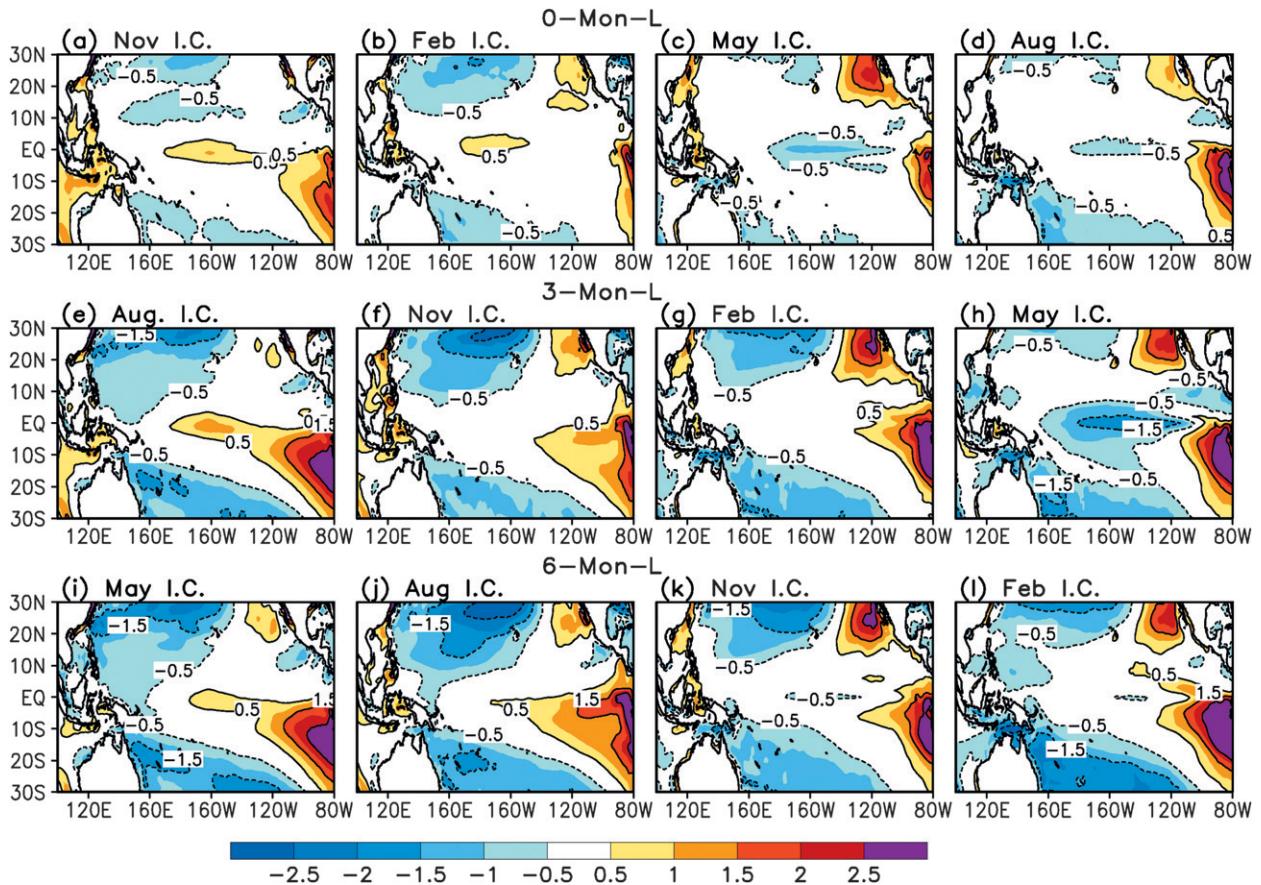


FIG. 2. As in Fig. 1, but for 1999–2010.

the systematic bias was small in the first two target seasons (Figs. 1b,g), and then cold bias of  $-1.5^{\circ}\text{C}$  developed in the third target season in SON (Fig. 1i), indicating a strong growth of cold bias during fall, consistent with that from May IC. For November IC, the systematic bias was generally small at all three target seasons (DJF, MAM, and JJA; Figs. 1a,f,k). The results suggest that there is a systematic growth of cold bias during fall. We found that the cold bias in SST near the equator is closely related to a cold bias in the subsurface temperature near the equator (not shown; see Kumar et al. 2012). Further analysis showed that the cold subsurface temperature bias is closely related to easterly wind bias near the date line that is largest during summer/fall (not shown). Future analysis is needed to fully understand the causes for the systematic biases and their impacts on the prediction skill of ENSO.

Systematic cold biases were also found in the western and central subtropical Pacific in both hemispheres. The cold bias in Northern Hemisphere was slightly larger than that in Southern Hemisphere and was about  $-1.5^{\circ}\text{C}$  near  $20^{\circ}\text{N}$  and  $-1^{\circ}\text{C}$  near  $20^{\circ}\text{S}$ . A systematic warm bias

was found in the northeast and southeast Pacific. The warm bias was about  $+2^{\circ}\text{C}$  near  $15^{\circ}\text{S}$  and  $85^{\circ}\text{W}$  and about  $+1.5^{\circ}\text{C}$  near  $25^{\circ}\text{N}$  and  $120^{\circ}\text{W}$ .

With the abrupt warming in the equatorial subsurface temperature in the CFSR around 1999 (Xue et al. 2011), the systematic cold bias in SST near the equator decreased substantially during 1999–2010 (Fig. 2). At 0-month lead, the bias was about  $+0.5^{\circ}\text{C}$  from November IC (Fig. 2a) and February IC (Fig. 2b) and  $-0.5^{\circ}\text{C}$  from May IC (Fig. 2c) and August IC (Fig. 2d). The small cold bias from August IC in SON transitioned into a warm bias of  $+1^{\circ}\text{C}$  in DJF (Fig. 2e), indicating a strong warming tendency during winter. The warming tendency from fall to winter was also observed in the period 1982–98 (Figs. 1d,e). The cold bias from May IC strengthened during fall (Fig. 2h) and then rapidly reduced and switched to a warm bias during winter (Fig. 2i), consistent with the strong warming tendency from fall to winter observed in August IC. From November IC, the systematic bias near the equator was small for all three target seasons (DJF, MAM, and JJA; Figs. 2a,f,k), except near the coast of South America, similar to that in

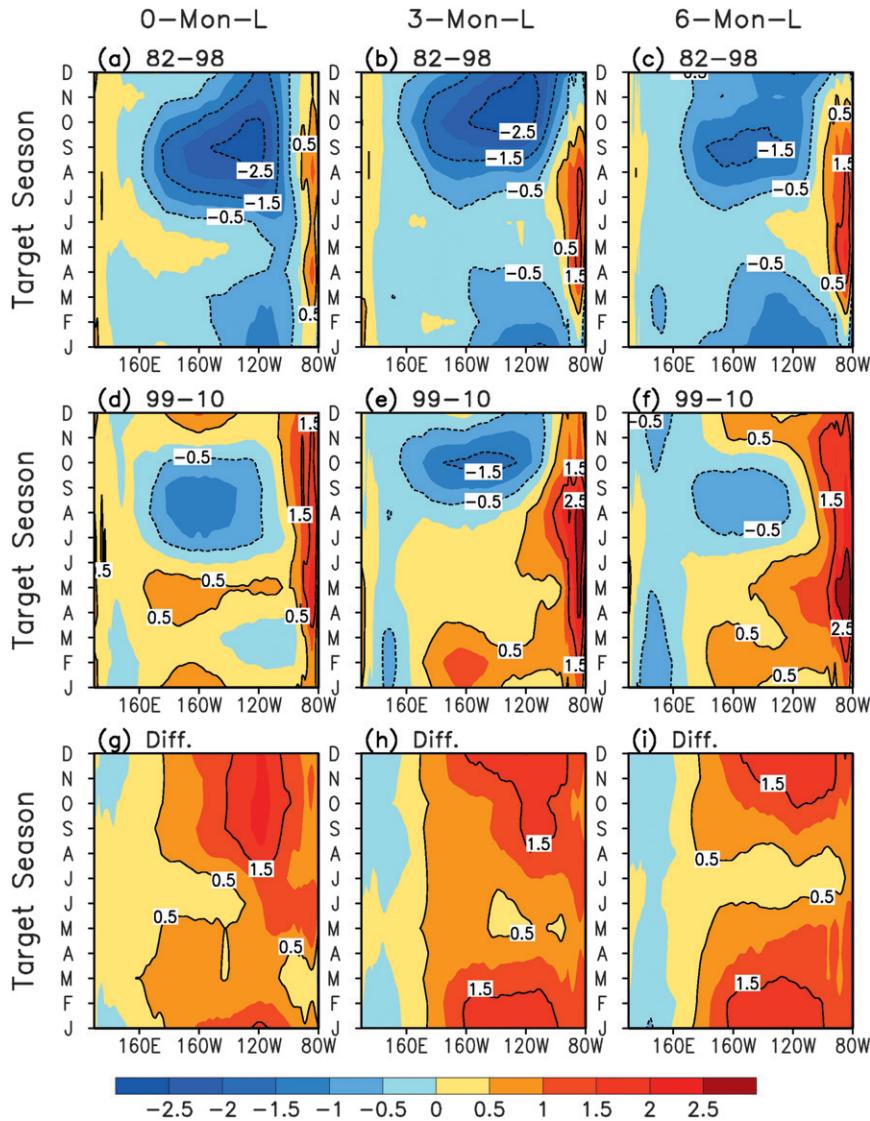


FIG. 3. Time-longitude plots of seasonal mean SST ( $^{\circ}\text{C}$ ) forecast departure from OIv2 SST averaged in  $2^{\circ}\text{S}$ – $2^{\circ}\text{N}$  in (top) 1982–98 and (middle) 1999–2010 and (bottom) difference of the departures between 1999–2010 and 1982–98 at (left) 0-, (middle) 3-, and (right) 6-month lead. The contour interval is  $1^{\circ}\text{C}$ .

1982–98 (Figs. 1a,f,k). From February IC, the systematic bias was also small for all three target seasons (Figs. 2b,g,l). These results suggest that the model has smaller systematic cold bias near the equator for forecasts starting from May to August IC in 1999–2010 than in 1982–98.

The warm bias in the southeast Pacific strengthened from 1982–98 to 1999–2010 and covered a larger area and had a larger amplitude (Figs. 1, 2). The systematic cold bias in the subtropics was slightly enhanced in Southern Hemisphere and extended to the far western equatorial Pacific at the long leads. The coexistence of

the cold bias west of the date line and the warm bias east of the date line reduced the mean zonal SST gradient near the eastern edge of the warm pool, which is likely to lead to a reduction in the zonal advective feedbacks for ENSO (Picaut et al. 1997; Jin and An 1999). The reduction of the zonal SST gradient was most prominent in the DJF target season (Figs. 2a,e,i).

The systematic bias near the equator as a function of target seasons at 0-, 3-, and 6-month leads is shown in Fig. 3. The systematic bias in 1982–98 was dominated by a pronounced negative bias, with largest amplitude from July to October. The season for the center of

negative bias varies slightly with lead times, changing from September at 0-month lead to October at 3-month lead and then back to September at 6-month lead (Figs. 3a–c). The amplitude of the negative bias changed little from 0- to 3-month leads but weakened substantially from 3- to 6-month leads. This suggests that a cold bias developed quickly at 0-month lead, probably because of initial error growth and initialization shock. The cold bias was then governed by the slow mode of coupled instability that had the fastest growth during fall. The fact that the cold bias at 6-month lead is smaller than that at 3-month lead may imply that the impacts of ICs began to fade and the model started to drift toward its climatology that has smaller cold bias than that at 3-month lead.

The cold bias during summer/fall decreased substantially from 1982–98 to 1999–2010 (Figs. 3d–f). A warm bias of about  $+0.5^{\circ}\text{C}$  developed during winter and spring. In addition, the warm bias near the coast of South America became larger and also extended westward. The differences in the systematic bias from 1999–2010 to 1982–98 are all positive except in the far western Pacific (Figs. 3g–i). The differences are largest during fall and winter while the signal of ENSO is the largest. In addition, the differences have a phase propagation with lead months with little change in amplitude. Further analysis shows that the differences developed at 0-month lead and then largely persisted at different lead times (not shown).

The differences in the systematic SST forecast bias for 1982–98 and 1999–2010 can be traced back to the differences in the IC bias in the two periods. For example, compared to the NCEP–U.S. Department of Energy (DOE) atmospheric reanalysis (Kanamitsu et al. 2002), the CFSR (from which the ICs for the CFSv2 were taken) had easterly wind bias in the central equatorial Pacific that was as large as  $-1.5\text{ m s}^{-1}$  during June–August in 1982–98 (Fig. 8 in Kumar et al. 2012). Consistently, the equatorial subsurface temperature in the CFSR had negative bias during July–September (Fig. 7 in Kumar et al. 2012). The negative subsurface temperature bias lagged the easterly wind bias by 1–2 months, implying that the easterly wind bias (during the analysis) may be the cause for the negative subsurface temperature bias. The easterly wind bias was largely corrected when the ATOVS satellite observations were assimilated starting around October 1998 (Xue et al. 2011; Zhang et al. 2012). Interestingly, positive subsurface temperature bias developed in 1999–2010 when the easterly wind bias was significantly reduced (Fig. 7 in Kumar et al. 2012). The appearance of warm bias around 1999 was believed to be related to the change of background error covariance, related to the change of

wind bias, which was not captured in the ocean data assimilation system (Xue et al. 2011). The positive subsurface temperature bias would favor positive SST bias, which partially canceled out the cold SST bias during fall but led to warm SST bias in winter and spring (Fig. 3d).

#### 4. Standard deviation of SST anomaly

The standard deviation (STD) of SSTA is used to measure interannual variability of SSTA. The STD of hindcast SSTA is compared with that of observations in 1982–98 and 1999–2010 separately (Figs. 4, 5). During 1982–98, the difference between the STD of model and observation (shading) is much smaller than the STD of observation (contours), suggesting the model predicted well the overall observed STD pattern. However, the model underestimated the STD in the far eastern tropical Pacific, and overestimated the STD between  $140^{\circ}$  and  $100^{\circ}\text{W}$ , indicating a westward shift of the maximum center in interannual variability of SSTA, a typical bias in climate models (Jin et al. 2008). Another feature is that the model overestimated the STD around the Maritime Continent and the western and northwestern tropical Pacific by as much as 50% (not shown). In SON, the model generally overestimated the STD in the central Pacific for all leads. The overestimation of the STD near the date line during fall was related to the cold bias in the central–eastern Pacific (Figs. 1d,h,l), which enhanced the mean zonal SST gradient and therefore the zonal advective feedbacks for ENSO.

There was a distinct change of STD from 1982–98 to 1999–2010 (Fig. 5). With more frequent occurrence of the CP events (Yeh et al. 2009; Lee and McPhaden 2010), the center of maximum SST variability moved to near the date line during winter and spring (Figs. 5a,b). The STD in the eastern Pacific decreased substantially over all the seasons (Figs. 4, 5). The model underestimated (overestimated) the STD near the date line (in the eastern Pacific) particularly during winter and spring. The too weak variability near the date line was related to the systematic biases in SST, which reduced the mean zonal SST gradient (Figs. 2a,e,i) and weakened the zonal advective feedbacks for ENSO. The STD in the eastern Pacific was overestimated by as much as 200% in spring when the observed STD decreased substantially from 1982–98 to 1999–2010. Similar to the period 1982–98, the model overestimated the STD in the far western tropical Pacific in 1999–2010. The STD was also overestimated by 200% in the northwest Pacific during fall (not shown). The above results suggest that the model has large systematic errors near the date line during winter, in the eastern Pacific during winter/spring, and in the northwest

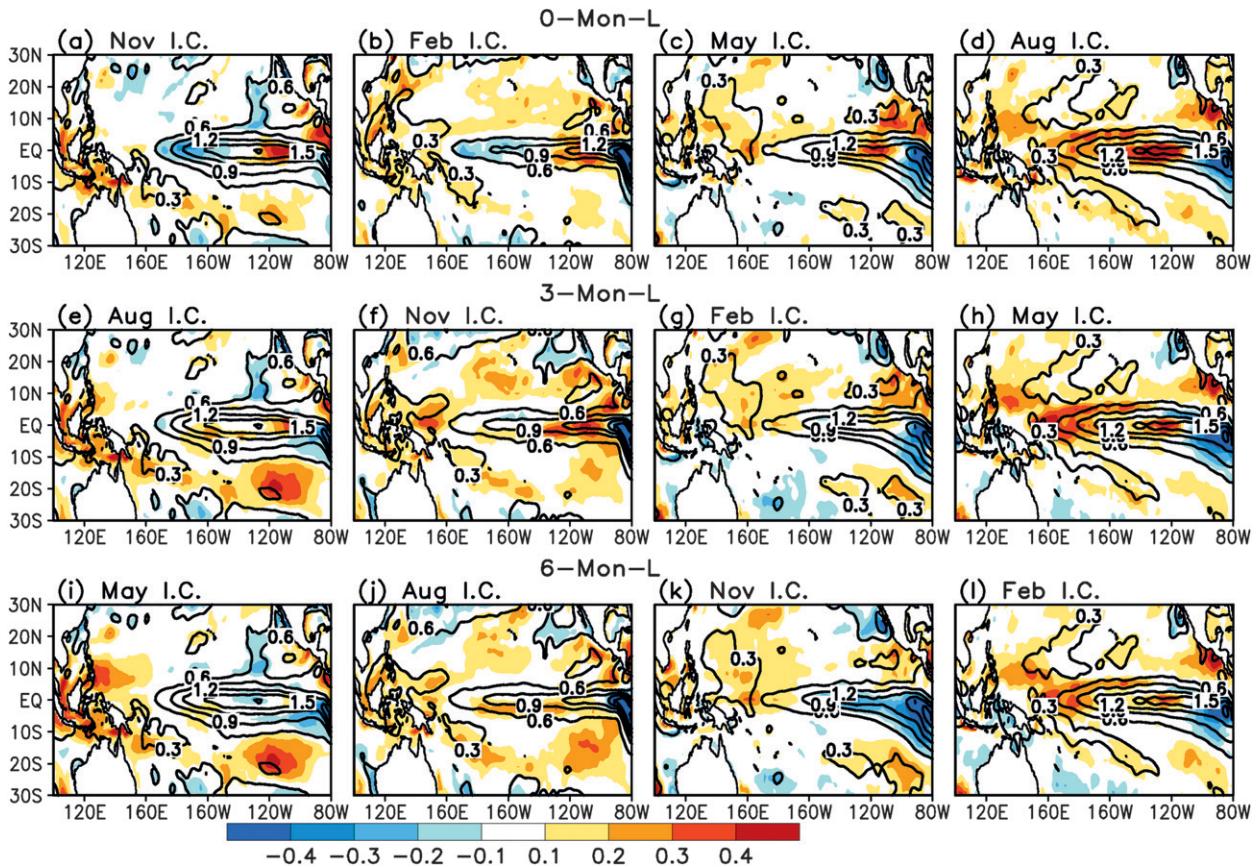


FIG. 4. As in Fig. 1, but for the STD of SST ( $^{\circ}\text{C}$ ) anomaly. Shown are the STD of observed SST anomaly (contours) and the difference between the STD of the model and observed SST anomaly (shading). The contour interval is  $0.3^{\circ}\text{C}$ .

Pacific during fall. The sources of the model errors are beyond the scope of the current study.

The decadal change of ENSO variability is further analyzed using the STDs of the Niño-4 ( $5^{\circ}\text{S}$ – $5^{\circ}\text{N}$ ,  $160^{\circ}\text{E}$ – $150^{\circ}\text{W}$ ), Niño-3 ( $5^{\circ}\text{S}$ – $5^{\circ}\text{N}$ ,  $150^{\circ}$ – $90^{\circ}\text{W}$ ), and Niño-3.4 SSTA in the period prior to and after 1999 (Fig. 6). During 1982–98, the model overestimated the STD of Niño-4 during fall (dashed versus solid black line), which is consistent with the too strong variability near the date line in SON (Figs. 4d,h,l). McPhaden (2012) suggested that the STD of Niño-4 increased since 2000 during winter and early spring, which contributes to enhanced spring predictability barrier in Niño-4. Figure 6 shows that the model did not capture the increase of Niño-4 STD. Instead, the STD of Niño-4 in the model forecast decreased during winter and spring at 3- and 6-month leads (Figs. 6d,g).

The model predicted the STD of Niño-3 in 1982–98 reasonably well. The STD of Niño-3 in observation decreased substantially from the early to late period, with the largest decrease (from  $0.9^{\circ}$  to  $0.3^{\circ}\text{C}$ ) in spring (Fig. 6b). At 0-month lead, the model captured the

decreasing Niño-3 STD from the early to late period (Fig. 6b). However, at 3- and 6-month leads, the model failed to capture the decrease of STD in observation (Figs. 6e,h). This suggests that the model drifted away from ICs quickly and could not maintain the decadal shift of STD in the observation. Consistent with the Niño-3, the STD of Niño-3.4 in observation also decreased from the early to late period, and the model also failed to capture the weakening of STD (Figs. 6f,i). Overall, the model did a poor job in capturing the observed weakening tendency of SST variability in the eastern Pacific from 1982–98 to 1999–2010 (Wang et al. 2010; Hu et al. 2013).

## 5. Prediction skill of SST anomaly

### a. ACC and RMSE

The deterministic prediction skill of SSTA is quantified using ACC and RMSE. The model ACCs at 3-month lead over 1982–98 and 1999–2010 are shown in Fig. 7 along with the perfect-model ACCs. Note

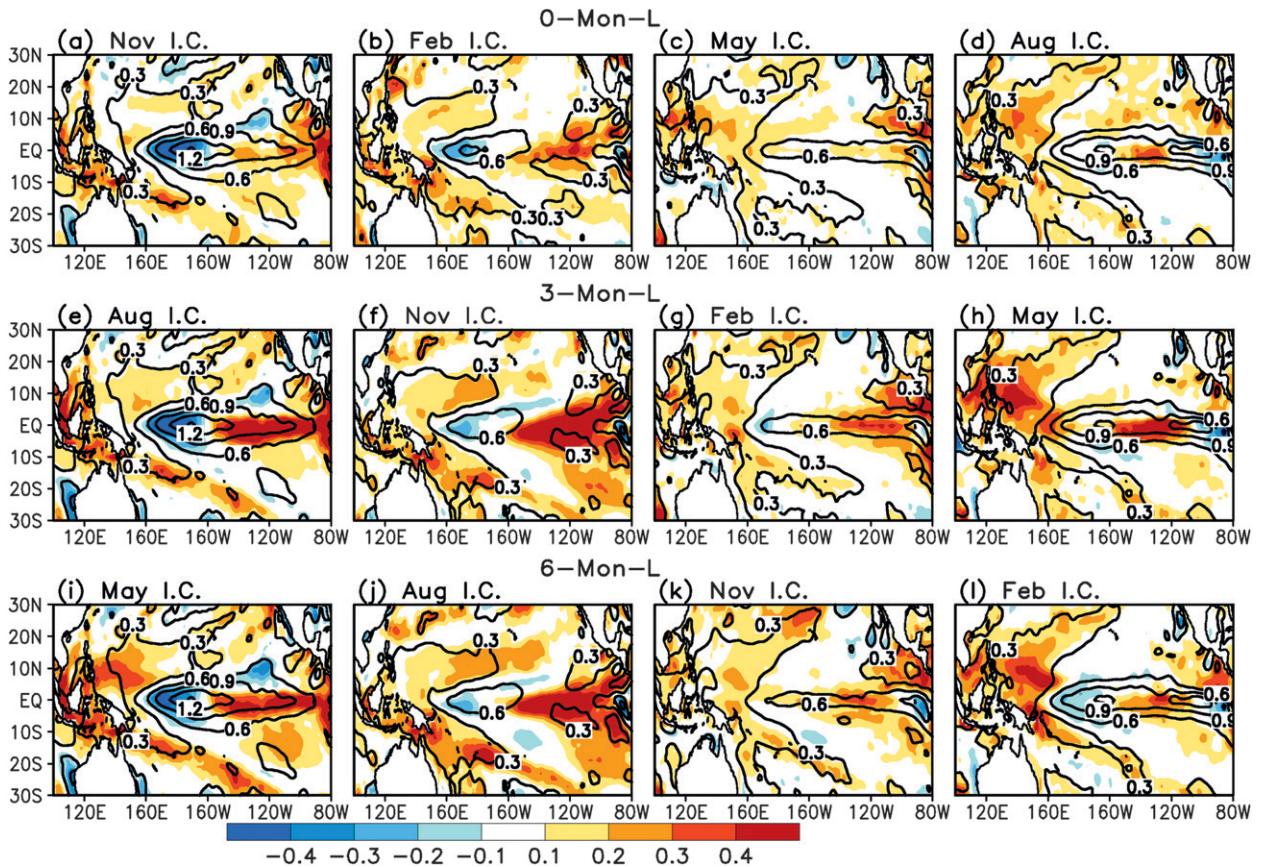


FIG. 5. As in Fig. 4, but for 1999–2010.

that only ACC values higher than the 95% significance value are shown. The significance value is based on Student's *t* test. The model ACC in 1982–98 reaches the highest value in the central and eastern Pacific during winter when the STD of SSTA is the largest. The model ACC in the central–eastern Pacific is about 0.9 in winter; reduces to about 0.7 in spring; and further reduces to about 0.5 in summer because of spring predictability barrier, which is common for numerical and statistical models (Xue et al. 1994; Jin et al. 2008; Stockdale et al. 2011). It is interesting that the ACC in the equatorial western and northwestern Pacific is above 0.6 during winter and spring and becomes less than 0.6 in the far western Pacific during summer and fall (Figs. 7c,d). The reduction of skill in the far western Pacific in summer and fall might be related to too strong spurious variability over those regions (Figs. 4c,d,h,j–l). Compared to the model ACC, the perfect-model ACC (Figs. 7e–h) is somewhat larger in the eastern Pacific during summer and in the far western Pacific during fall. Overall, the perfect-model ACC is a good indication of the actual model skill, implying that the CFSv2 represents well the coupled variability in the real world during 1982–98.

Compared to the early period, the model ACC in the late period is generally lower. The model ACC is particularly lower in the northwest subtropical Pacific during spring and summer (Figs. 7j,k). However, there is little change in the perfect model ACC from the early to late period, indicating inconsistency in the change of prediction skill and potential predictability. During 1999–2010, the perfect-model ACC is substantially higher than the model ACC in the southeastern Pacific during winter/spring, and in the far western and northwestern Pacific throughout the annual cycle. Overall, the differences between the CFSv2 hindcast ACC and the perfect-model ACC are much larger in 1999–2010 than in 1982–98, indicating that there may be larger potential for improvement in the late period than in the early period.

The ACCs of Niño-3, Niño-4, and Niño-3.4 SSTA over the two periods are compared with those of persistence forecast (Fig. 8). The ACC, as a function of IC months and lead time, shows a fast decline across the March–May season in both the model and persistence forecast, which is often referred to as the spring predictability barrier. It is noted that the spring predictability

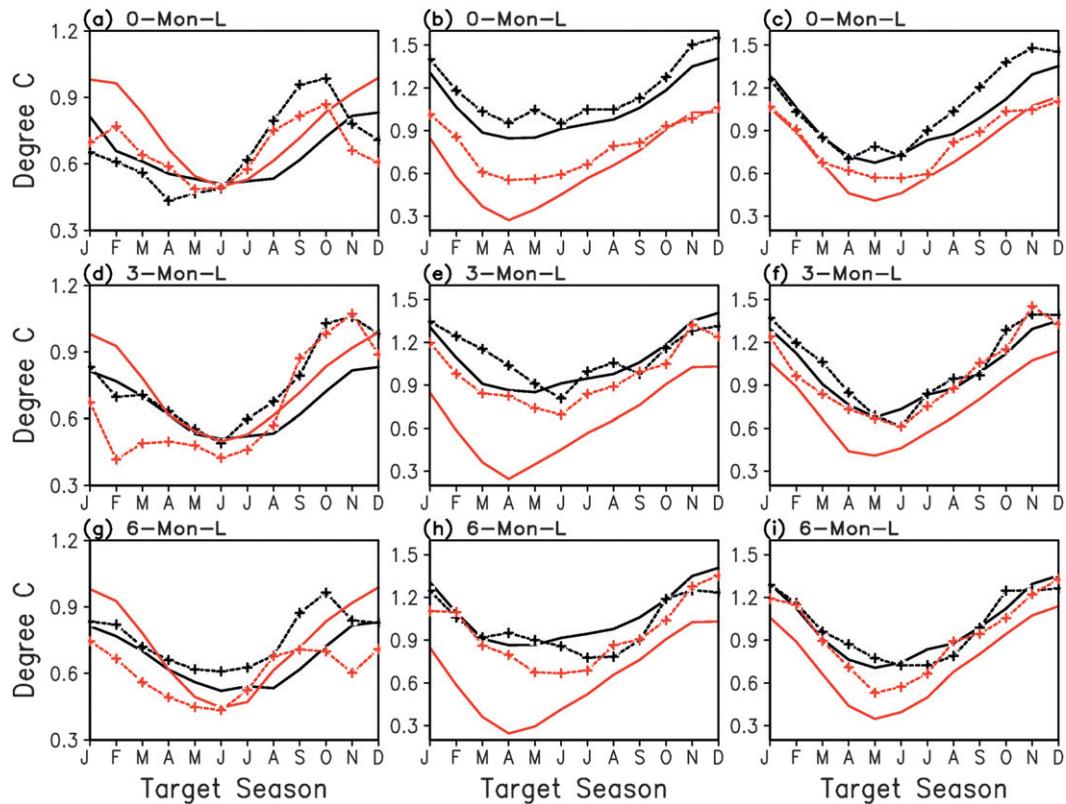


FIG. 6. Annual cycle of STD of SST ( $^{\circ}\text{C}$ ) anomaly for (left to right) the Niño-4 ( $5^{\circ}\text{S}$ – $5^{\circ}\text{N}$ ,  $160^{\circ}\text{E}$ – $150^{\circ}\text{W}$ ), Niño-3 ( $5^{\circ}\text{S}$ – $5^{\circ}\text{N}$ ,  $150^{\circ}$ – $90^{\circ}\text{W}$ ), and Niño-3.4 ( $5^{\circ}\text{S}$ – $5^{\circ}\text{N}$ ,  $170^{\circ}$ – $120^{\circ}\text{W}$ ) region at (top) 0-, (middle) 3-, and (bottom) 6-month lead. Shown are the STD of observed (solid line) and forecast (dashed line) SST anomaly in 1982–98 (black line) and 1999–2010 (red line).

barrier in Niño-4 is weaker than that in Niño-3 in both the model and persistence forecast. In 1982–98, the Niño-4 in the model has a larger predictability than the Niño-3 from July to December IC, while from January to June IC the ACC drops quickly at short lead times and recovers at long lead times (Fig. 8a). In 1999–2010, the STD of Niño-4 increased during winter and early spring and the spring predictability barrier in Niño-4 is enhanced (McPhaden 2012). Consistently, the ACC of Niño-4 declined rapidly across spring for forecasts starting from winter and early spring IC (Fig. 8d). However, the ACC in the model forecast starting from summer IC declines more rapidly in 1999–2010 than in 1982–98, while the ACC in the persistence forecast declines more slowly in 1999–2010 than in 1982–98, which may indicate model forecast errors from summer ICs.

The spring predictability barrier in Niño-3 strengthened substantially from the early to late period, consistent with the decrease of STD in the eastern Pacific in spring (Figs. 4b, 5b). The spring predictability barrier in the CFSv2 was consistent with that of persistence forecast. Similarly, the spring predictability barrier in

Niño-3.4 was also enhanced from the early to late period. Therefore, for 1999–2010, it was more difficult to forecast summer conditions starting from winter and early spring IC because of the enhanced spring predictability barrier.

The ratios of RMSE between the model and persistence forecast are shown in Fig. 9 along with the differences in ACC. For RMSE ratio less than 1 (yellow–red shading), the model forecast is more skillful than the persistence forecast. In 1982–98, the model beats the persistence in Niño-3 and Niño-3.4 for almost all initial months and lead months. In 1999–2010, the model forecast is more skillful than the persistence forecast, except starting from summer/fall ICs. The lower skill, indicated by RMSE ratio larger than 1 (blue shading), has target seasons varying from November to February. We will show later that the larger RMSE from summer/fall IC was related to the model bias in STD in winter/spring target seasons (Fig. 5).

The model RMSE can also be compared with the model ES. If the CFSv2 is unbiased, the model RMSE is expected to be close to the model ES. Figure 10 shows

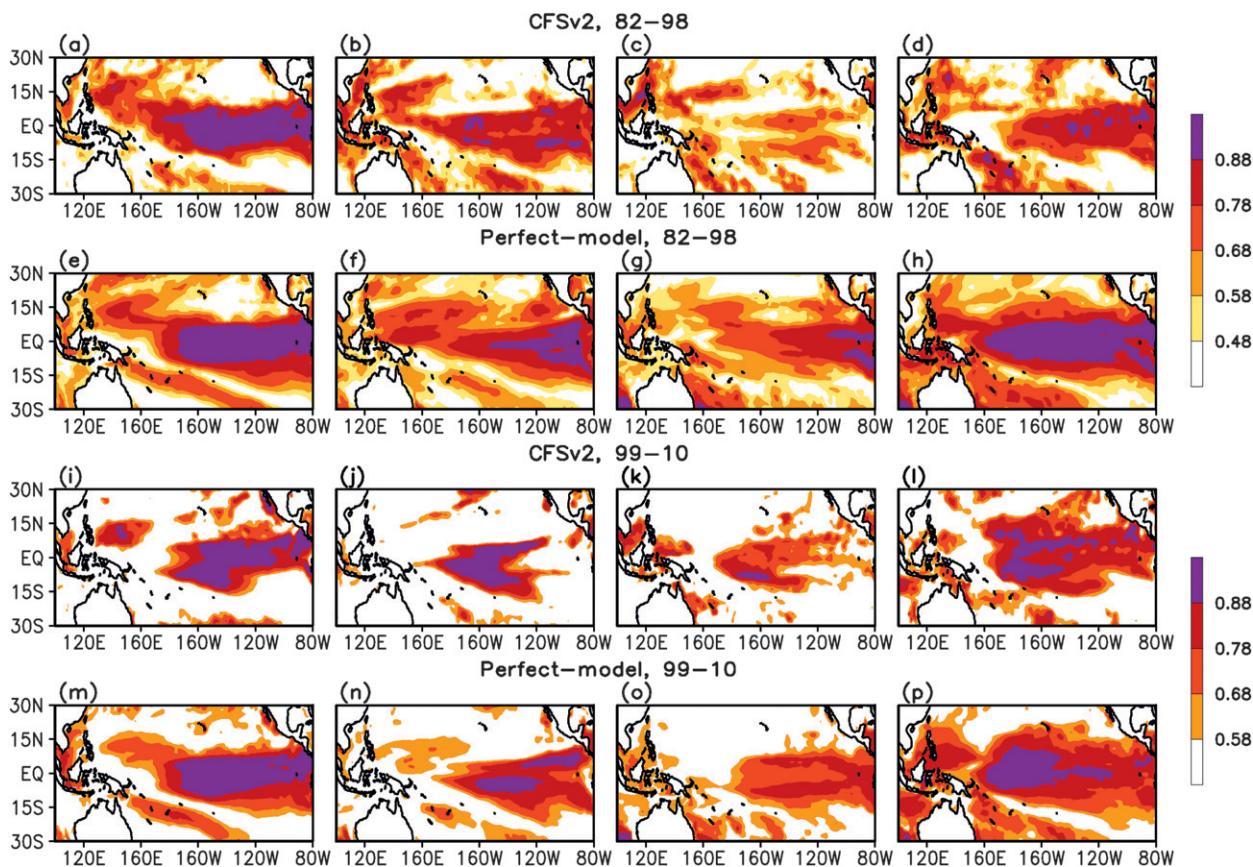


FIG. 7. The ACC between the CFSv2 and observed SST anomaly for (a)–(d) 1982–98 and (i)–(l) 1999–2010 for 4 target seasons and 4 initial months, shown on the top of each panel. Also shown are the ACC between the perfect model and observed SST anomaly for (e)–(h) 1982–98 and (m)–(p) 1999–2010. Only values  $>95\%$  significance are shown.

that the model RMSE indeed agrees well with the model ES during 1982–98, particularly at lead times longer than 3 months. At lead times shorter than 3 months, the model RMSE is larger than the model ES, indicating that the model is overconfident. In comparison, during 1999–2010, the agreement between the model RMSE and model ES is generally lower than during 1982–98, particularly in Niño-4. It is seen in Fig. 10d that the model RMSE is twice as large as the model ES during target seasons varying from December to March, when the model RMSE is larger than the RMSE of persistence forecast (Fig. 9d). This is related to a model bias that underestimates SST variability near the date line during winter and spring (Fig. 5). For Niño-3, the model RMSE is larger than the model ES during target seasons varying from February to April. This is related to the fact that the model overestimates SST variability in the eastern Pacific during winter/spring (Fig. 5). For Niño-3.4, the model RMSE is actually smaller than or equal to the model ES from September to May ICs, which may suggest the model is unbiased. However, from June to

August ICs, the model RMSE is larger than the model ES, indicating the model is overconfident. This is consistent with the fact that the model forecast has larger RMSE than the persistence forecast from summer IC (Fig. 9f). Therefore, in 1999–2010, the model Niño-4 and Niño-3 forecasts at winter/spring target seasons are biased and the model Niño-3.4 forecasts starting from summer IC are biased.

#### b. Probability forecasts for ENSO events

The ensemble forecasts from the CFSv2 provide an opportunity to issue a probabilistic forecast. We used three categories—El Niño, ENSO neutral, and La Niña—to develop probabilistic forecast. Figure 11 shows the probability forecast (bar) at 3-month lead for the 10 El Niño (1982/83, 1986/87, 1987/88, 1991/92, 1994/95, 1997/98, 2002/03, 2004/05, 2006/07, and 2009/10) and 10 La Niña (1983/84, 1984/85, 1988/89, 1995/96, 1998/99, 1999/2000, 2000/01, 2005/06, 2007/08, and 2010/11) events. The probability forecast is defined as the percentage of individual ensemble members that meet the threshold

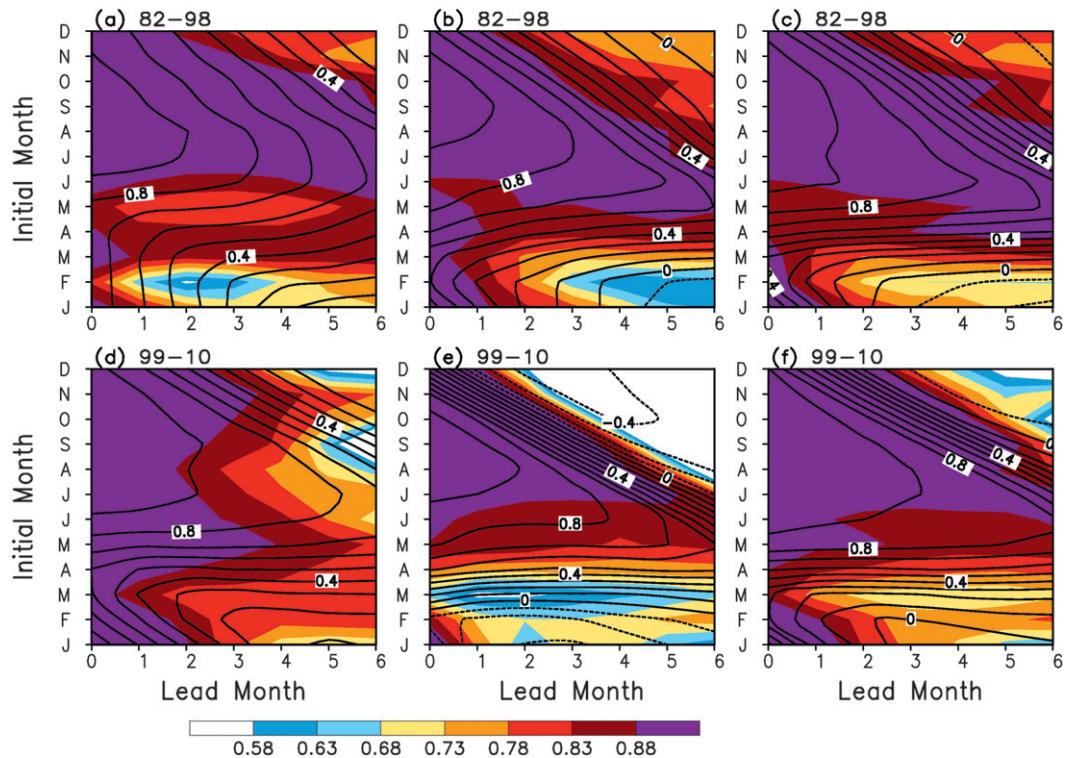


FIG. 8. The ACC of the CFSv2 (shading) and persistence-forecast SST anomaly (contours) in the (left) Niño-4, (middle) Niño-3, and (right) Niño-3.4 regions for the periods (top) 1982–98 and (bottom) 1999–2010 as functions of lead months ( $x$  axis) and initial months ( $y$  axis). The contour interval is 0.1. Only values with  $>95\%$  significance were shown.

of  $\pm 0.5^{\circ}\text{C}$ . Also shown in Fig. 11 are the ensemble mean forecast and observed Niño-3.4 for each event. The probability forecast is generally consistent with the ensemble mean forecast. Whenever the ensemble mean forecast meets the threshold  $\pm 0.5^{\circ}\text{C}$ , the probability forecast often exceeds 50%. The consistency between the probability forecasts and ensemble mean forecasts suggests that the probability forecast from CFSv2 is a reliable indicator for ENSO development, and it is useful, in addition to the ensemble mean forecast, for operational ENSO prediction.

### c. Hit rate and false alarm rate

In addition to the probability forecasts, we also investigated the hit and false alarm rate for El Niño, ENSO neutral, and La Niña (Kirtman et al. 2002). Figure 12 shows that the hit and false alarm rates for 1982–98 and 1999–2010 separately. During 1982–98, the hit and false alarm rates are well separated for El Niño (red dots), ENSO neutral (yellow dots), and La Niña (blue dots). Although the hit rate is similar, the false alarm rate increases from La Niña to El Niño to ENSO neutral. This implies that La Niña and El Niño forecast are more reliable than ENSO-neutral forecast in

1982–98, which is consistent with the assessment of multiple model ENSO forecast skill by Kirtman et al. (2002).

During 1999–2010, the hit and false alarm rates are mixed among El Niño, ENSO neutral, and La Niña. For El Niño and La Niña, the hit rate decreased appreciably from 3- to 6-month lead, suggesting the predictability for El Niño–La Niña is lower in 1999–2010 than in 1982–2010, consistent with the anomaly correlation score. It is interesting to note that the false alarm rate for La Niña increased appreciably from 1982–2010 to 1999–2010. This suggests that the CFSv2 became less reliable for forecasting La Niña after 1999 than before 1999. Further study is needed to investigate why the false alarm for La Niña increased after 1999.

### d. El Niño composites

To further examine the prediction skill of ENSO, the hindcast time series of Niño-3 and Niño-4 at 0-, 3-, and 6-month leads are compared with observations (Fig. 13). For Niño-3, the model tended to delay the onset phase of the strong El Niño events (1982/83, 1986/87, and 1997/98) in the 1980s and 1990s, while it captured the onset phase of the weak to moderate

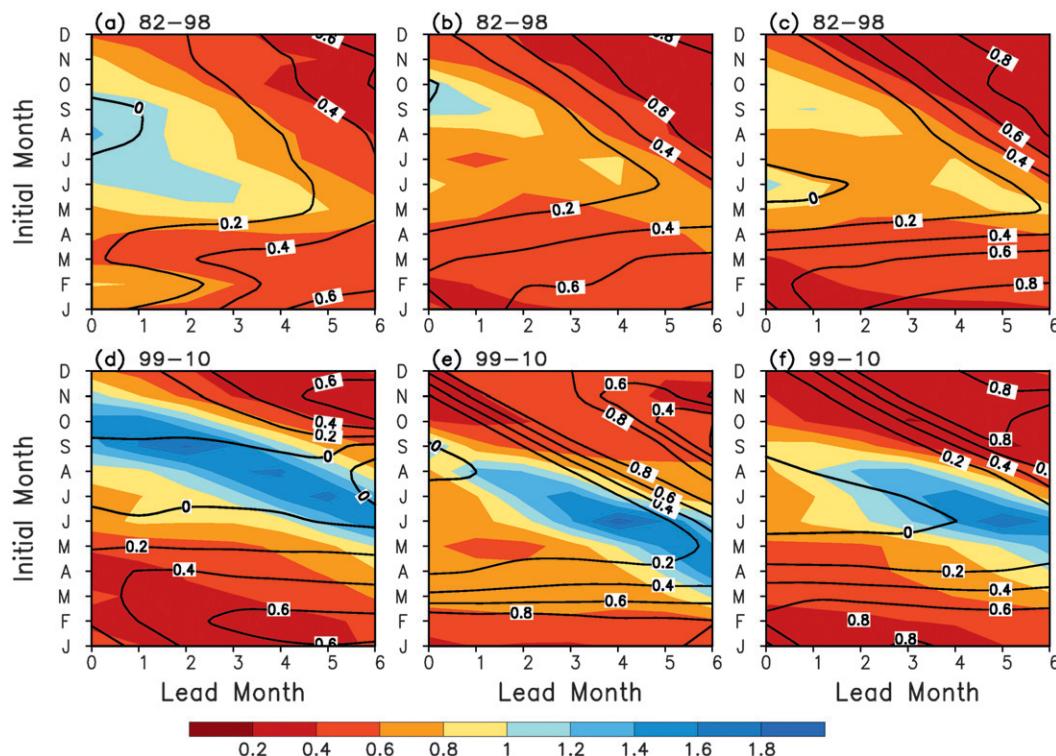


FIG. 9. The RMSE ratio (shading) between the CFSv2 and the persistence-forecast SST anomaly in the (left) Niño-4, (middle) Niño-3, and (right) Niño-3.4 regions for the periods (top) 1982–98 and (bottom) 1999–2010 as functions of lead months ( $x$  axis) and initial months ( $y$  axis). Also shown are the differences between ACC of the CFSv2 and persistence SST forecast (contours). The contour interval is 0.2.

El Niño events (1991/92, 1994/95, 2002/03, 2004/05, and 2006/07) generally well. It is also noted that the model captured the decay phase of the strong El Niños in the 1980s and 1990s well but prolonged the decay phase of the weak to moderate El Niños (i.e., 1994/95, 2002/03, and 2006/07). For Niño-4, the model overestimated the amplitude of the 1982/83 and 1997/98 El Niños but generally underestimated the amplitude of the weak to moderate El Niños.

It is worthwhile to point out that the model forecasted a false El Niño in 1990/91 and 2001/02 and a false La Niña in 2003/04, which is consistent with the fact that the model overestimated the STD of SSTA in the eastern Pacific (Fig. 5). The model forecasted all the La Niña events, except it severely underestimated the amplitude of the 2007/08 La Niña in Niño-4 SSTA.

The El Niño composites of Niño-4, Niño-3, and Niño-3.4 were calculated in 1982–98 and 1999–2010 separately (not shown). Based on the NOAA El Niño definition, six (four) El Niño winters were observed before (after) 1999. The composites were calculated with four El Niños in each period with the 1986/87 and 1987/88 El Niños excluded because of their different seasonal evolution.

During the period before 1999, El Niños typically developed in spring reached peak phase in winter and decayed in next spring, transitioning to neutral conditions in the next summer. The model tended to delay the onset phase of El Niños in Niño-3 and Niño-3.4 particularly at long leads but captured the decay phase and transition to neutral conditions quite well. In contrast, in the period after 1999, the model captured the onset phase of El Niños in Niño-3 and Niño-3.4 reasonably well but prolonged the decay phase in Niño-3 by 2–3 months. For Niño-4 SSTA, the model underestimated the amplitude from November to February even at 0-month lead. The underestimation of Niño-4 SSTA had significant impacts on the SSTA pattern of the CP events predicted by the CFSv2.

To see how the model predicted the spatial patterns of SST anomaly for two flavors of El Niño, the model composites for the EP and CP events at 0-, 3-, and 6-month leads are compared with observations (Figs. 14–16). During the development phase (August–October), the maximum SSTA of the EP events extends from 110°W to the coast of South America, while the maximum SSTA of the CP events is located between

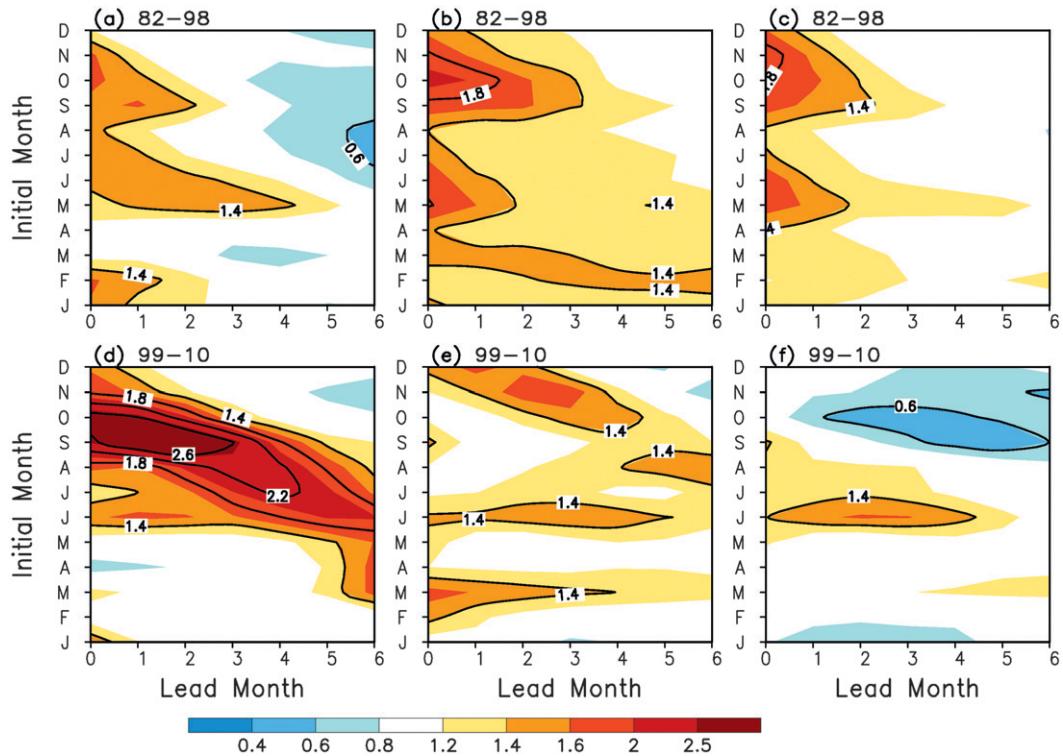


FIG. 10. The RMSE ratio (shading) between the CFSv2 and perfect-model forecast SST anomaly in the (left) Niño-4, (middle) Niño-3, and (right) Niño-3.4 regions for the periods (top) 1982–98 and (bottom) 1999–2010 as functions of lead months ( $x$  axis) and initial months ( $y$  axis).

180° and 140°W (Figs. 14a,b). The CFSv2 predicted the SSTA pattern of the EP events reasonably well, except that the maximum center was shifted westward by about 20° (Fig. 14c). For the CP events, the CFSv2 predicted both the pattern and amplitude up to 3-month lead (Fig. 14d). Both the amplitude of the EP and CP events were underestimated at 6-month lead, starting from January IC (Figs. 14g,h). The above results suggest that the CFSv2 is skillful in predicting the SSTA pattern of both the EP and CP events for the Atlantic hurricane season starting from April IC, which is particularly useful for the May outlook of the Atlantic hurricane season activity (Kim et al. 2009).

During the peak phase (DJF), the maximum SSTA of the EP events is centered around 120°W with an amplitude of 2.5°C, while the maximum SSTA of the CP events is centered around 170°W with an amplitude of 1.5°C. The CFSv2 predicted very well both the pattern and amplitude of the EP events up to 3-month lead but underestimated the amplitude at 6-month lead. On the other hand, the CFSv2 predicted the amplitude of the CP events well up to 6-month lead but shifted the maximum SSTA of the CP events eastward by about 20°. This is related to the fact that the CFSv2 underestimated

(overestimated) SSTA variability near the date line (in the eastern Pacific) in 1999–2010 when the CP events are dominant.

During the decay phase (MAM), the maximum SSTA of the EP events retreated to near the coast (Fig. 16a). The model agreed with the observation very well except that the warm anomaly was broader meridionally and the cold anomaly in the subtropics was stronger. For the CP events, the warm anomaly near the date line weakened and a cold anomaly developed in the eastern Pacific (Fig. 16b). The east–west dipole of SSTA was forecast well by CFSv2 at 0-month lead but poorly at 3- and 6-month leads. Xiang et al. (2013) speculate that the rapid decay of the SST warming in the eastern Pacific (EPW) for the CP events is related to enhanced latent heat flux, anomalous zonal advection, and vertical entrainment, while the slow decay of the EPW for the EP events is mainly due to less solar radiation. Conceivably, the fast decay of the EPW in the CP events is more challenging for models to predict than the slow decay of the EPW in the EP events. To understand the bias in the decay phase of the CP events in CFSv2, a study of the mixed layer heat budget will be needed.

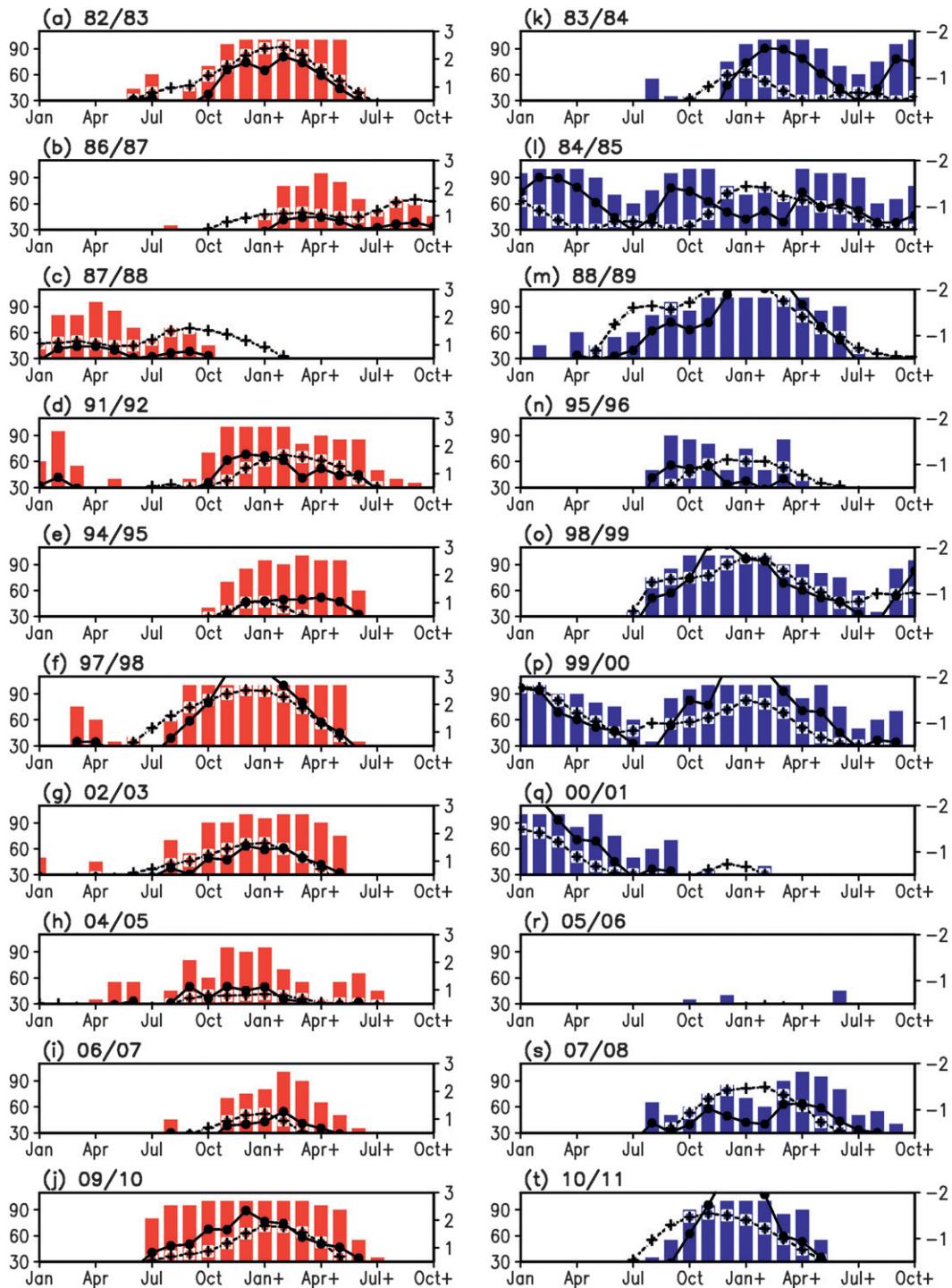


FIG. 11. Niño-3.4 SST anomaly from observation (+'s) and ensemble mean forecast (•'s) for (a) 1982/83, (b) 1986/87, (c) 1987/88, (d) 1991/92, (e) 1994/95, (f) 1997/98, (g) 2002/03, (h) 2004/05, (i) 2006/07, and (j) 2009/10 El Niño and (k) 1983/84, (l) 1984/85, (m) 1988/89, (n) 1995/96, (o) 1998/99, (p) 1999/2000, (q) 2000/01, (r) 2005/06, (s) 2007/08, and (t) 2010/11 La Niña. Also shown are the percentages (bars) of individual ensemble member forecast that are  $\geq 0.5^{\circ}\text{C}$  ( $\leq -0.5^{\circ}\text{C}$ ). The x axis is the time from January in the El Niño–La Niña year to October in the following year, and the y axis on the left is the percentage and on the right is the Niño-3.4 SST anomaly in degrees Celsius.

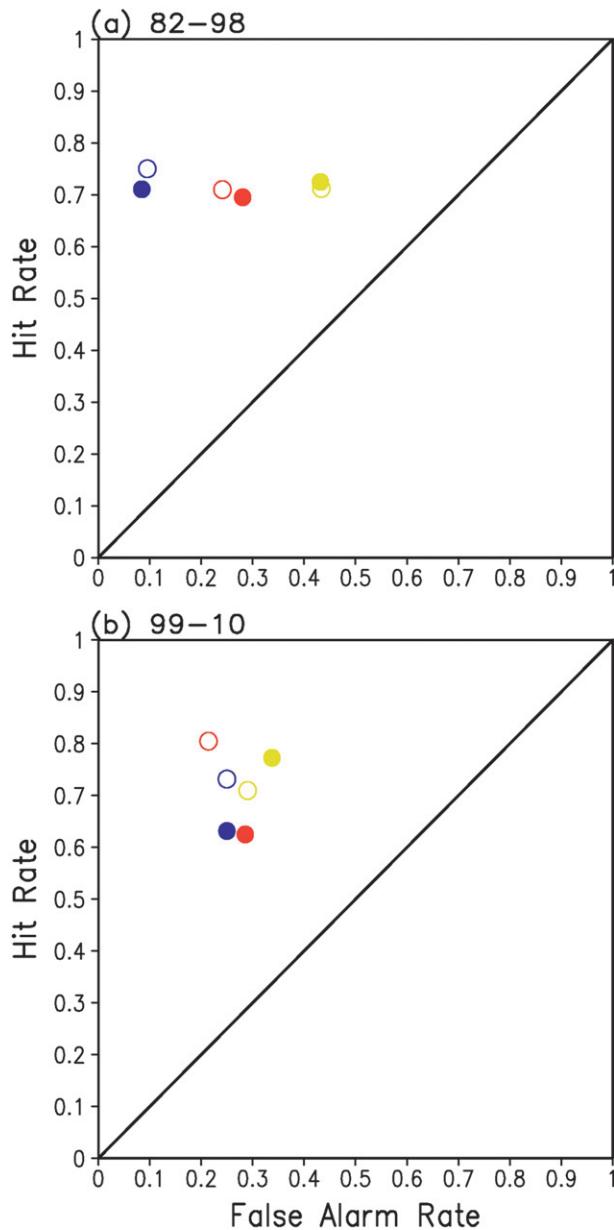


FIG. 12. Hit rate vs false alarm rate at 3-month lead ( $\circ$ 's) and 6-month lead ( $\bullet$ 's) for El Niño (red), ENSO neutral (yellow), and La Niña (blue) in the periods of (a) 1982–98 and (b) 1999–2010.

## 6. Summary and discussion

A partially coupled ocean, atmosphere, land, and sea ice climate reanalysis from 1979 to present, referred to as CFSR (Saha et al. 2010), was implemented at NCEP in 2010. Initialized with the CFSR, retrospective forecasts for the period 1982–2010 were made using a new coupled dynamical model, referred to as CFSv2 (Saha et al. 2012, manuscript submitted to *J. Climate*). The CFSv2 was implemented as an operational model in

March 2011. The purpose of the analysis is (i) to inform the users of the CFSv2 about its capability in predicting ENSO and the shift of variability and predictability around late 1990s and (ii) to document the biases in the CFSv2 prediction of tropical Pacific SSTs that need to be targeted as part of the model development and evaluation process.

For the CFSv2 SST forecasts in the tropical Pacific, there was a systematic cold bias in the central–eastern equatorial Pacific during 1982–98 that reached  $-2.5^{\circ}\text{C}$  during summer/fall. At the end of 1998, the cold bias suddenly reduced to about  $-1^{\circ}\text{C}$  during summer/fall and a warm bias of  $+0.5^{\circ}\text{C}$  developed during winter/spring. This shift of the systematic biases in hindcast SST around 1999 contributed to a spurious warming trend in forecast SSTA based on the 1982–2010 climatology (Kim et al. 2012). Given a distinct difference in mean forecast bias for the period before and after 1999, it was necessary to derive model SSTA by removing the model climatology in 1982–98 and 1999–2010 separately.

The differences in the systematic bias in 1982–98 and 1999–2010 can be linked to the differences in the bias in the CFSR in the two periods (Xue et al. 2011; Kumar et al. 2012). For example, the CFSR has easterly wind bias and negative ocean heat content bias in the central and eastern equatorial Pacific in 1982–98 (Xue et al. 2011). The easterly wind bias was largely corrected when the ATOVS satellite observations were assimilated around October 1998 (Zhang et al. 2012). Unfortunately, positive heat content bias developed in 1999–2010 when the easterly wind bias was significantly reduced. Positive ocean heat content bias favors positive SST bias, which partially canceled out the cold SST bias during summer/fall but led to a warm SST bias in winter/spring in 1999–2010.

There was a decadal change of standard deviation (STD) of SSTA from 1982–98 to 1999–2010. With the more frequent occurrence of the CP events since 2000 (Yeh et al. 2009; Lee and McPhaden 2010), the center of maximum SST variability moved to near the date line during winter and spring. The STD of Niño-4 in observation increased substantially during winter and early spring from 1982–98 to 1999–2010. However, the STD of Niño-4 in CFSv2 decreased during winter and spring at 3- and 6-month leads. The underestimation of STD near the date line in 1999–2010 might be related to the systematic warm (cold) bias in the central Pacific (western Pacific) that reduces the mean zonal SST gradient, therefore interannual variability through weakening zonal advective feedbacks.

There was also a large decadal change of STD in Niño-3 since 2000. For example, the STD of Niño-3 in observation decreased from  $0.9^{\circ}$  to  $0.3^{\circ}\text{C}$  during spring

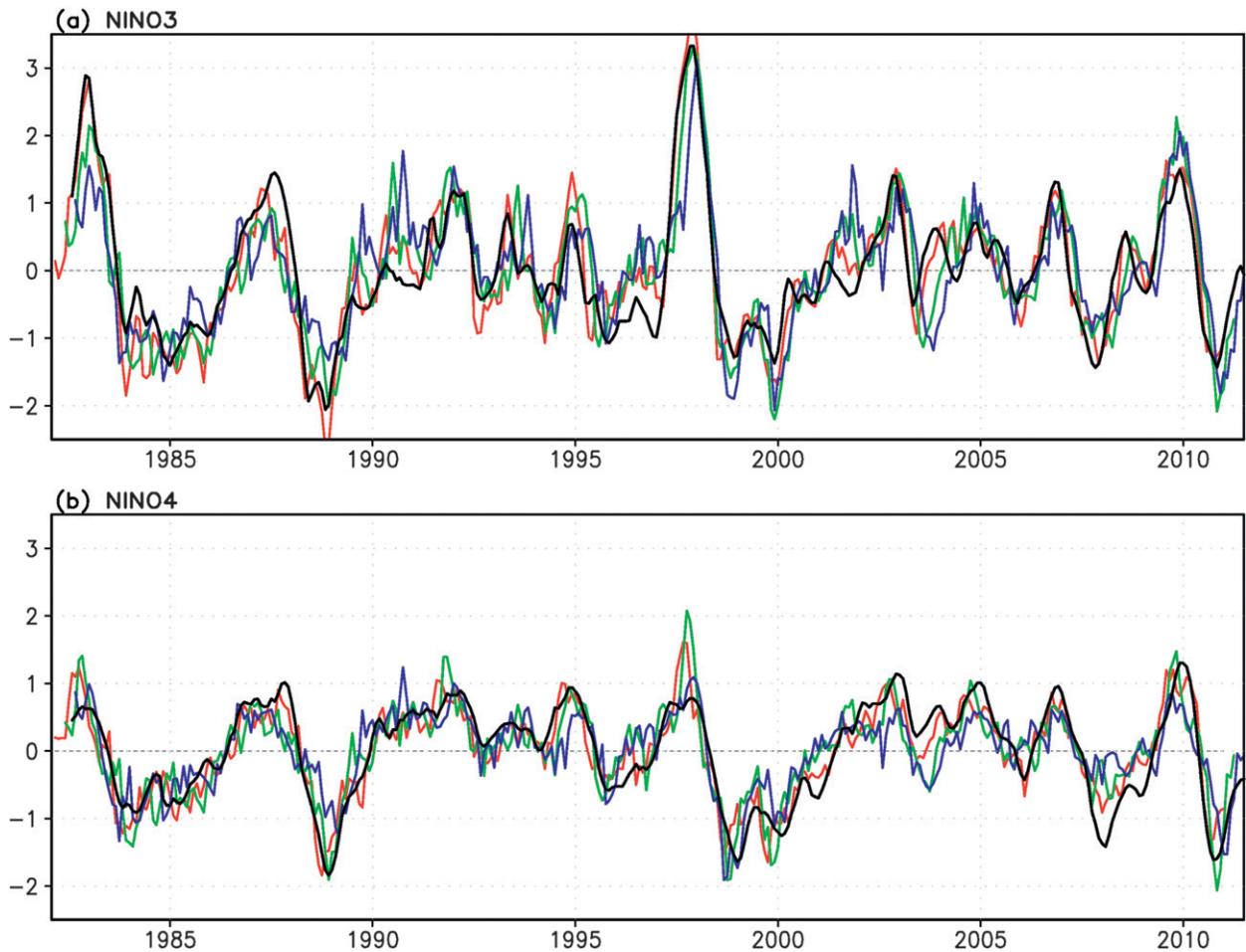


FIG. 13. Time series of the (top) Niño-3 and (bottom) Niño-4 SST anomaly ( $^{\circ}\text{C}$ ) from observation (thick black line), CFSv2 forecast at 0- (thin red line), 3- (thin green line), and 6-month lead (thin blue line) from 1982 to 2011.

from 1982–98 to 1999–2010. The CFSv2 captured the decrease of STD in Niño-3 at 0-month lead. However, at 3- and 6-month leads, the model failed to capture the decrease of STD. The reason is that the model overestimated the STD in Niño-3 by about 200% in spring during 1999–2010. The overestimated variability in Niño-3 is partially related to model deficiencies in forecasting the SSTA pattern of the CP events. The model tends to damp out SSTA west of the date line and shift the maximum SSTA eastward by  $20^{\circ}$  in the composite of the CP events.

The composite SSTAs for the EP and CP events in CFSv2 are compared with those in the observation. The CFSv2 had a tendency to delay the onset phases of the strong EP events in 1980s and 1990s (i.e., 1982/83 and 1997/98) but predicted their decay phases rather well. On the other hand, the model predicted the onset phases of the weak to moderate CP events well but prolonged the decay phase of the warming in the eastern Pacific for the CP events.

The prediction skill of SSTA is quantified using ACC and RMSE. The actual model skill agreed quite well with the perfect-model skill in 1982–98, indicating that the model replicates the coupled variability in the real world well. The differences between the actual model and perfect-model skill are much larger in 1999–2010 than in 1982–98, suggesting that there may be larger potential for improvement in the late than in the early period. During 1999–2010, the CFSv2 forecasts did not have enough dispersion (too confident) in the southeastern Pacific during winter/spring and in the far western and northwestern Pacific throughout the seasonal cycle.

On average, the model skill is higher than the persistence skill for almost all initial months and lead months in 1982–98, but in 1999–2010 the model RMSE is larger than the RMSE of persistence forecast starting from summer/fall ICs for winter target season. Two factors contributed to the lower skill than the persistence: 1) the

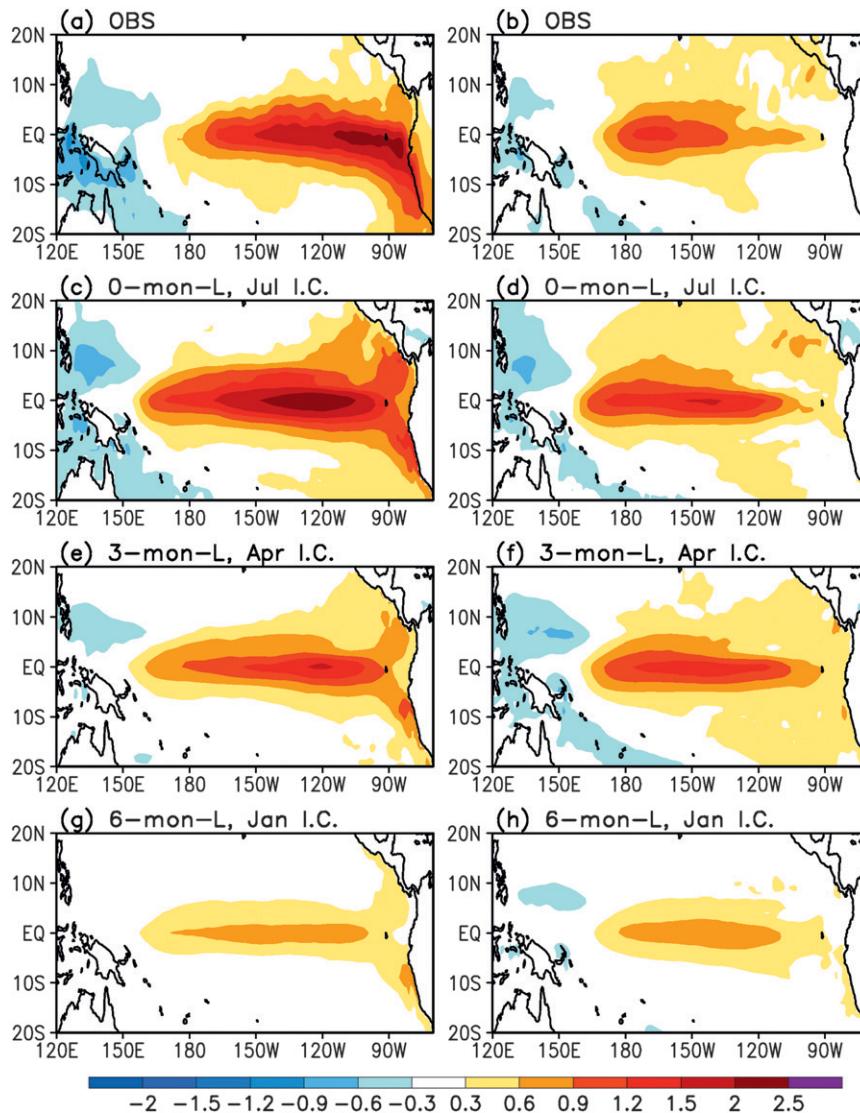


FIG. 14. The SST anomaly composite for August–October from (a),(b) observation; model forecast at (c),(d) 0-, (e),(f) 3-, and (g),(h) 6-month lead for (left) the eastern Pacific El Niños (1982/83, 1991/92, 1997/98, and 2006/07) and (right) central Pacific El Niños (1994/95, 2002/03, 2004/05, and 2009/10).

persistence skill is very high starting from summer/fall ICs and 2) the CFSv2 underestimated (overestimated) STD near the date line (in the eastern Pacific) for winter target season. This suggests there is a room for improvement in the CFSv2 forecast starting from summer/fall ICs in 1999–2010.

The ensemble forecasts from the CFSv2 provide an opportunity to develop probabilistic forecast products. We used three categories—El Niño, ENSO neutral, and La Niña—to calculate a probabilistic forecast. The probability forecast is generally consistent with the ensemble mean forecast. Whenever the ensemble mean forecast

meets the threshold  $\pm 0.5^{\circ}\text{C}$ , the probability forecast often exceeds 50%. This suggests that the probability forecast from CFSv2 is a reliable indicator for ENSO development, and it is useful, in addition to the ensemble mean forecast, for operational ENSO prediction.

We also investigated the hit and false alarm rate for El Niño, ENSO neutral, and La Niña (Kirtman et al. 2002). During 1982–98, the hit rate is about 0.7 at 6-month lead for all three categories, but the false alarm rate increases from La Niña (0.1) to El Niño (0.2) to ENSO neutral (0.4). This implies that La Niña and El Niño forecasts are more reliable than the ENSO-neutral forecast in

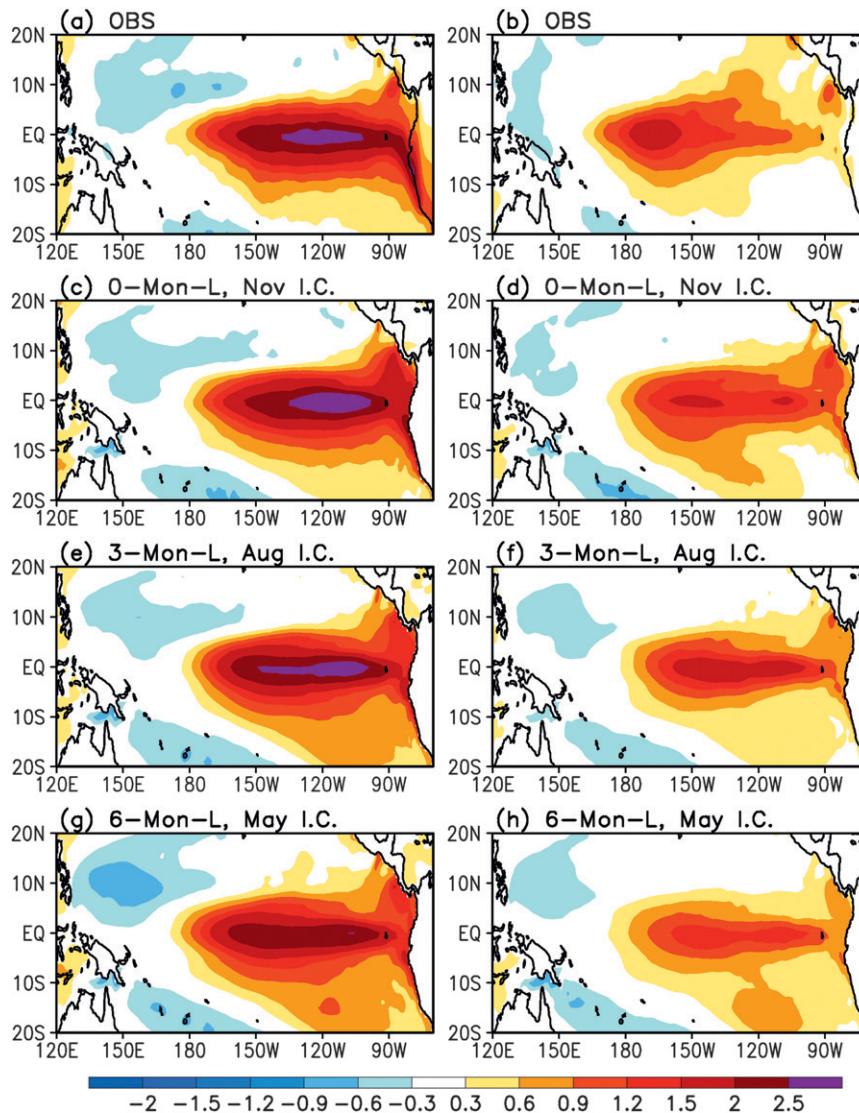


FIG. 15. As in Fig. 14, but for December–February.

1982–98. During 1999–2010, the hit rate decreased appreciably from 3-month lead (0.7) to 6-month lead (0.6), suggesting the predictability for El Niño/La Niña is lower in 1999–2010 than in 1982–2010, consistent with the anomaly correlation score. It is interesting to note that the false alarm rate for La Niña increased from 1982–2010 (0.1) to 1999–2010 (0.25). This suggests that the CFSv2 became less reliable for forecasting La Niña after 1999 than before 1999. Further study is needed to investigate why the false alarm for La Niña increased after 1999.

Many studies suggest there is a close linkage between mean and variability bias in climate models. Therefore, it is critical to reduce the systematic bias in the CFSv2 forecast. The positive ocean heat content bias in the

CFSR after 1999 contributed to a warm SST bias in the central Pacific, which reduced the mean zonal SST gradient and therefore interannual variability through weakening of zonal advective feedback for ENSO. Conceivably, removal of the positive heat content bias in ICs helps reduce the warm bias and therefore improve simulation of variability in the central Pacific. However, removal of the positive heat content bias in ICs will likely worsen the cold bias in the central–eastern Pacific during summer/fall.

Recently, Zhu et al. (2012) suggest that the hindcast skill of the CFSv2 can be improved significantly by replacing the CFSR with another ocean reanalysis for ocean initialization. However, Zhu et al. (2012) only compared the skill based on the 1979–2007 climatology

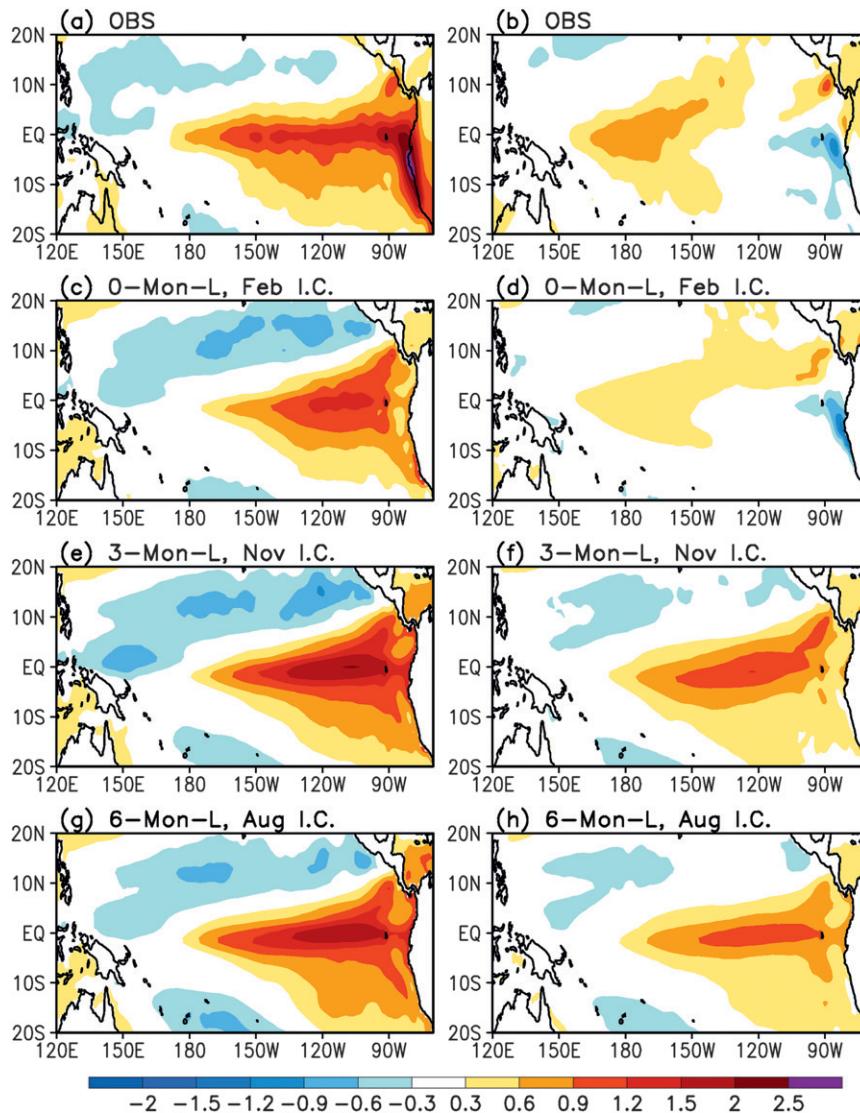


FIG. 16. As in Fig. 14, but for March–May.

when the lower skill with CFSR ICs was largely related to the shift of systematic bias around 1999. It will be interesting to further compare the skill for the periods 1982–98 and 1999–2010 separately and to verify if the prediction skill of CFSv2 is improved in 1999–2010 when another ocean reanalysis is used in ocean initialization.

**Acknowledgments.** We thank three reviewers as well as Dr. Hui Wang and Dr. Caihong Wen for their constructive comments and suggestions on this paper.

#### REFERENCES

- 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.
- Balmaseda, M. A., and Coauthors, 2009: Ocean initialization for seasonal forecasts. *Oceanography*, **22**, 154–159.
- Barnston, A. G., M. K. Tippett, 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.
- Battisti, D. S., 1988: Dynamics and thermodynamics of a warming event in a coupled tropical atmosphere–ocean model. *J. Atmos. Sci.*, **45**, 2889–2919.
- Chelliah, M., W. Ebisuzaki, S. Weaver, and A. Kumar, 2011: Evaluating the tropospheric analyses from NCEP’s Climate Forecast System Reanalysis. *J. Geophys. Res.*, **116**, D17107, doi:10.1029/2011JD015707.
- Hendon, H. H., E. Lim, G. Wang, O. Alves, and D. Hudson, 2009: Prospects for predicting two flavors of El Niño. *Geophys. Res. Lett.*, **36**, L19713, doi:10.1029/2009GL040100.
- Horii, T., I. Ueki, and K. Hanawa, 2012: Breakdown of ENSO predictors in the 2000s: Decadal changes of recharge/discharge-SST

- phase relation and atmospheric intraseasonal forcing. *Geophys. Res. Lett.*, **39**, L10707, doi:10.1029/2012GL051740.
- Hu, Z.-Z., A. Kumar, B. Jha, W. Wang, B. Huang, and B. Huang, 2012: An analysis of warm pool and cold tongue El Niños: Air-sea coupling processes, global influences, and recent trends. *Climate Dyn.*, **38**, 2017–2035, doi:10.1007/s00382-011-1224-9.
- , —, H.-L. Ren, H. Wang, M. L'Heureux, and F.-F. Jin, 2013: Weakened interannual variability in the tropical Pacific Ocean since 2000. *J. Climate*, **26**, 2601–2613.
- Ji, M., D. W. Behringer, and A. Leetmaa, 1998: An improved coupled model for ENSO prediction and implications for ocean initialization. Part II: The coupled model. *Mon. Wea. Rev.*, **126**, 1022–1034.
- Jin, E. K., and J. L. Kinter III, 2009: Characteristics of tropical Pacific SST predictability in coupled GCM forecasts using the NCEP CFS. *Climate Dyn.*, **32**, 675–691.
- , and Coauthors, 2008: Current status of ENSO prediction skill in coupled ocean-atmosphere models. *Climate Dyn.*, **31**, 647–664.
- Jin, F.-F., and S.-I. An, 1999: Thermocline and zonal advection feedbacks within the equatorial ocean recharge oscillator model for ENSO. *Geophys. Res. Lett.*, **26**, 2989–2992.
- Kanamitsu, M., W. Ebisuzaki, J. Woolen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter, 2002: NCEP-DOE AMIP-II Reanalysis (R-2). *Bull. Amer. Meteor. Soc.*, **83**, 1631–1643.
- Kao, H. Y., and J. Y. Yu, 2009: Contrasting eastern Pacific and central Pacific types of ENSO. *J. Climate*, **22**, 615–632.
- Kim, H. M., P. J. Webster, and J. A. Curry, 2009: Impact of shifting patterns of Pacific Ocean warming on North Atlantic tropical cyclones. *Science*, **325**, 77–80.
- , —, and —, 2012: Seasonal prediction skill of ECMWF system 4 and NCEP CFSv2 retrospective forecast for the Northern Hemisphere winter. *Climate Dyn.*, **39**, 2957–2973, doi:10.1007/s00382-012-1364-6.
- Kirtman, B. P., J. Shukla, M. Balmaseda, N. Graham, C. Penland, Y. Xue, and S. Zebiak, 2002: Current status of ENSO forecast skill: A report to the CLIVAR Working Group on Seasonal to Interannual Prediction. International CLIVAR Project Office Rep. 56, 24 pp.
- 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.
- Kumar, A., M. Chen, L. Zhang, W. Wang, Y. Xue, C. Wen, L. Marx, and B. Huang, 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.
- Larkin, N. K., and D. E. Harrison, 2005: On the definition of El Niño and associated seasonal average U.S. weather anomalies. *Geophys. Res. Lett.*, **32**, L13705, doi:10.1029/2005GL022738.
- 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.
- McPhaden, M. J., 2012: A 21st century shift in the relationship between ENSO SST and warm water volume anomalies. *Geophys. Res. Lett.*, **39**, L09706, doi:10.1029/2012GL051826.
- , T. Lee, and D. McClurg, 2011: El Niño and its relationship to changing background conditions in the tropical Pacific Ocean. *Geophys. Res. Lett.*, **38**, L15709, doi:10.1029/2011GL048275.
- Picaut, J., F. Masia, and Y. du Penhoat, 1997: An advective reflective conceptual model for the oscillatory nature of the ENSO. *Science*, **277**, 663–666.
- Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang, 2002: An improved in situ and satellite SST analysis for climate. *J. Climate*, **15**, 1609–1625.
- Saha, S., and Coauthors, 2006: The NCEP Climate Forecast System. *J. Climate*, **19**, 3483–3517.
- , and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1057.
- Singh, A., T. Delcroix, and S. Cravatte, 2011: Contrasting the flavors of El Niño-Southern Oscillation using sea surface salinity observations. *J. Geophys. Res.*, **116**, C06016, doi:10.1029/2010JC006862.
- Stockdale, T. N., D. Anderson, M. Balmaseda, F. Doblas-Reyes, L. Ferranti, K. Mogensen, F. Molteni, and F. Vitart, 2011: ECMWF Seasonal Forecast System 3 and its prediction of sea surface temperature. *Climate Dyn.*, **37**, 455–471, doi:10.1007/s00382-010-0947-3.
- Tziperman, E., 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.
- Wang, W., M. Chen, and A. Kumar, 2010: An assessment of the CFS real-time seasonal forecasts. *Wea. Forecasting*, **25**, 950–969.
- , P. Xie, S. H. Yo, Y. Xue, A. Kumar, and X. Wu, 2011: An assessment of the surface climate in the NCEP Climate Forecast System Reanalysis. *Climate Dyn.*, **37**, 1601–1620.
- Xiang, B., B. Wang, and T. Li, 2013: A new paradigm for the predominance of standing central Pacific warming after the late 1990s. *Climate Dyn.*, **41**, 327–340, doi:10.1007/s00382-012-1427-8.
- Xue, Y., M. A. Cane, S. E. Zebiak, and B. Blumenthal, 1994: On the prediction of ENSO: a study with a low-order Markov model. *Tellus*, **46A**, 512–528.
- , B. Huang, Z.-Z. Hu, A. Kumar, C. Wen, D. Behringer, and S. Nadiga, 2011: An assessment of oceanic variability in the NCEP Climate Forecast System Reanalysis. *Climate Dyn.*, **37**, 2511–2539, doi:10.1007/s00382-010-0954-4.
- , and Coauthors, 2012: A comparative analysis of upper ocean heat content variability from an ensemble of operational ocean reanalyses. *J. Climate*, **25**, 6905–6929.
- Yeh, S. W., J. S. Kug, B. Dewitte, M. H. Kwon, B. P. Kirtman, and F. F. Jin, 2009: El Niño in a changing climate. *Nature*, **461**, 511–514.
- Zhang, L., A. Kumar, and W. Wang, 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.
- Zhu, J., B. Huang, L. Marx, J. L. Kinter III, M. A. Balmaseda, R.-H. Zhang, and Z.-Z. Hu, 2012: Ensemble ENSO hindcasts initialized from multiple ocean analyses. *Geophys. Res. Lett.*, **39**, L09602, doi:10.1029/2012GL051503.