

The NCEP Climate Forecast System Version 2

SURANJANA SAHA,^{*} SHRINIVAS MOORTHI,^{*} XINGREN WU,⁺ JIANDE WANG,[#] SUDHIR NADIGA,⁺
 PATRICK TRIPP,⁺ DAVID BEHRINGER,^{*} YU-TAI HOU,^{*} HUI-YA CHUANG,^{*} MARK IREDELL,^{*}
 MICHAEL EK,^{*} JESSE MENG,⁺ RONGQIAN YANG,⁺ MALAQUÍAS PEÑA MENDEZ,⁺
 HUUG VAN DEN DOOL,[@] QIN ZHANG,[@] WANQIU WANG,[@] MINGYUE CHEN,[@] AND EMILY BECKER[&]

^{*} Environmental Modeling Center, NOAA/NWS/NCEP, College Park, Maryland

⁺ I. M. Systems Group, Inc., Rockville, Maryland

[#] Science Systems and Applications, Inc., Largo, Maryland

[@] Climate Prediction Center, NOAA/NWS/NCEP, College Park, Maryland

[&] Wyle Lab, Inc., Arlington, Virginia

(Manuscript received 23 November 2012, in final form 28 May 2013)

ABSTRACT

The second version of the NCEP Climate Forecast System (CFSv2) was made operational at NCEP in March 2011. This version has upgrades to nearly all aspects of the data assimilation and forecast model components of the system. A coupled reanalysis was made over a 32-yr period (1979–2010), which provided the initial conditions to carry out a comprehensive reforecast over 29 years (1982–2010). This was done to obtain consistent and stable calibrations, as well as skill estimates for the operational subseasonal and seasonal predictions at NCEP with CFSv2. The operational implementation of the full system ensures a continuity of the climate record and provides a valuable up-to-date dataset to study many aspects of predictability on the seasonal and subseasonal scales. Evaluation of the reforecasts show that the CFSv2 increases the length of skillful MJO forecasts from 6 to 17 days (dramatically improving subseasonal forecasts), nearly doubles the skill of seasonal forecasts of 2-m temperatures over the United States, and significantly improves global SST forecasts over its predecessor. The CFSv2 not only provides greatly improved guidance at these time scales but also creates many more products for subseasonal and seasonal forecasting with an extensive set of retrospective forecasts for users to calibrate their forecast products. These retrospective and real-time operational forecasts will be used by a wide community of users in their decision making processes in areas such as water management for rivers and agriculture, transportation, energy use by utilities, wind and other sustainable energy, and seasonal prediction of the hurricane season.

1. Introduction

In this paper, we describe the development of the National Centers for Environmental Prediction's (NCEP's) Climate Forecast System, version 2 (CFSv2; <http://cfs.ncep.noaa.gov>). We intend to be fairly complete about this development and the generation of its retrospective data. We also present some limited analysis of the performance of CFSv2.

The first Climate Forecast System (CFS), retroactively called CFSv1, was implemented into operations at NCEP in August 2004 and was the first quasi-global,

fully coupled atmosphere–ocean–land model used at NCEP for seasonal prediction (Saha et al. 2006, hereafter S06). Earlier coupled models at NCEP had full ocean coupling restricted to only the tropical Pacific Ocean. CFSv1 was developed from four independently designed pieces of technology, namely the NCEP–Department of Energy (DOE) Global Reanalysis 2 (R2; Kanamitsu et al. 2002) that provided the atmospheric and land surface initial conditions, a global ocean data assimilation system (GODAS) operational at NCEP in 2003 (Behringer 2007) that provided the ocean initial states, NCEP's Global Forecast System (GFS) operational in 2003 that was the atmospheric model run at a lower resolution of T62L64, and the Modular Ocean Model, version 3 (MOM3), from the Geophysical Fluid Dynamics Laboratory (GFDL). The CFSv1 system worked well enough that it became difficult to terminate

Corresponding author address: Dr. Suranjana Saha, NOAA Center for Weather and Climate Prediction, 5830 University Research Court, College Park, MD 20740.
 E-mail: suranjana.saha@noaa.gov

it, as it was used by many in the community, even after the CFSv2 was implemented into operations in March 2011. It was finally decommissioned in late September 2012.

Obviously CFSv2 has improvements in all four components mentioned above, namely the two forecast models and the two data assimilation systems. CFSv2 also has a few novelties: an upgraded four-level soil model, an interactive three-layer sea ice model, and historically prescribed (i.e., rising) CO₂ concentrations. But above all, CFSv2 was designed to improve consistency between the model states and the initial states produced by the data assimilation system. It took nearly seven years to complete the following aspects:

- 1) carry out extensive testing of a new atmosphere–ocean–sea ice–land model configuration including decisions on resolution, etc.;
- 2) make a coupled atmosphere–ocean–sea ice–land reanalysis for 1979–2010¹ with the new system [resulting in the Climate Forecast System Reanalysis (CFSR)] for the purpose of creating initial conditions for CFSv2 retrospective forecasts;
- 3) make retrospective forecasts with the new system using initial states from CFSR from 1982 to 2010 and onward to calibrate operational subsequent real time subseasonal and seasonal predictions; and
- 4) operationally implement CFSv2.

Items 1 and 2 have already been described in Saha et al. (2010), and item 4 does not need to be treated in any great detail in a scientific paper, other than to mention that CFSv2 is run in near-real time with a very short data cutoff time, thereby increasing its applicability to the shorter time scales relative to CFSv1, which was late by about 36 h after real time. So, in this paper we mainly describe the CFSv2 model, the design of the retrospective forecasts, and some results from these forecasts.

The performance of the CFSv2 retrospective forecasts can be split into four time scales. The shortest time scale of interest is the subseasonal, mainly geared toward the prediction of the Madden–Julian oscillation (MJO) and more generally forecasts for the week 2–6 period over the United States (or any other part of the globe). The next time scale is the “long lead” seasonal prediction, out to 9 months, for which these systems are ostensibly designed. For both the subseasonal and seasonal, we have a very precise comparison between skill of prediction by the CFSv1 and CFSv2 systems evaluated over exactly the same hindcast years. The final two time

scales are decadal and centennial. Here the emphasis is less on forecast skill, and more on the general behavior of the model in extended integrations for climate studies.

Structurally, this paper makes a number of simple comparisons between aspects of CFSv1 and CFSv2 performance and discusses changes relative to CFSv1. For the background details of most of these changes, we refer to the CFSR paper (Saha et al. 2010) where all model development over the period 2003–09 has been laid out. In addition, some new changes were made relative to the models used in CFSR. These changes to the atmospheric and land model in the CFSR were deemed necessary when they were used for making the CFSv2 hindcasts. For instance, changes had to be made to combat a growing warm bias in the surface air temperature over land, and a decrease in the tropical Pacific sea surface temperature in long integrations.

The layout of the paper is as follows: Section 2 deals with changes in model components relative to CFSR. In section 3 the design of the hindcasts are described. Model performance in terms of forecast skill for intra-seasonal to long lead seasonal prediction is given in section 4. Section 5 describes other aspects of performance, including the evolution of the systematic error, diagnostics of the land surface, and behavior of sea ice. Model behavior in very long integrations, both decadal and centennial, is described in section 6. Conclusions and some discussion are presented in section 7. We also include four appendices that include the retrospective forecast calendar, reforecast and operational configuration of the CFSv2, and most importantly a summary of the availability of the CFSv2 data.

2. Overview of the Climate Forecast System Model

The coupled forecast model used for the seasonal retrospective and operational forecasts is different from the model used for obtaining the first-guess forecast for CFSR and operational Climate Data Assimilation System (CDAS) analyses (CDAS is the real time continuation of CFSR). The ocean and sea ice models are identical to those used in CFSR (Saha et al. 2010). The atmospheric and the land surface components, however, are somewhat different and these differences are briefly described below.

The atmospheric model has a spectral triangular truncation of 126 waves (T126) in the horizontal (equivalent to nearly a 100-km grid resolution) and a finite differencing in the vertical with 64 sigma-pressure hybrid layers. The vertical coordinate is the same as that in the operational CDAS. Differences between the model used here and in CFSR are mainly in the physical parameterizations of the atmospheric model and some tuning parameters in the land surface model and are as follows:

¹This paper describes the CFS reanalysis data from 1979–2010 and the CFSv2 retrospective data from 1982–2010. However, both datasets are being updated in real-time operations at NCEP.

- We use virtual temperature as the prognostic variable, in place of enthalpy that was used in major portions of CFSR. This decision was made with an eye toward unifying the GFS (which uses virtual temperature) and CFS, as well as the fact that the operational CDAS with CFSv2 currently uses virtual temperature.
- We also disabled two simple modifications made in CFSR to improve the prediction of marine stratus (Moorthi et al. 2010; Saha et al. 2010; Sun et al. 2010). This was done because including these changes resulted in excessive low marine clouds, which led to increased cold sea surface temperatures over the equatorial oceans in long integrations of the coupled model.
- We added a new parameterization of gravity wave drag induced by cumulus convection based on the approach of Chun and Baik (1998) (Å. Johansson 2009, personal communication). The occurrence of deep cumulus convection is associated with the generation of vertically propagating gravity waves. While the generated gravity waves usually have eastward or westward propagating components, in our implementation only the component with zero horizontal phase speed is considered. This scheme approximates the impact of stationary gravity waves generated by deep convection. The base stress generated by convection is parameterized as a function of total column convective heating and applied at the cloud top. Above the cloud top the vertically propagating gravity waves are dissipated following the same dissipation algorithm used in the orographic gravity wave formulation.
- As in CFSR, we use the Rapid Radiative Transfer Model (RRTM) adapted from AER Inc. (e.g., Mlawer et al. 1997; Iacono et al. 2000; Clough et al. 2005). The radiation package used in the retrospective forecasts is similar to the one used in the CFSR but with important differences in the cloud–radiation calculation. In CFSR, a standard cloud treatment is employed in both the RRTM longwave and shortwave parameterizations, namely that layers of homogeneous clouds are assumed in fractionally covered model grids. In the new CFS model, an advanced cloud–radiation interaction scheme is applied to the RRTM to address the unresolved variability of layered cloud. One accurate method would be to divide the clouds in a model grid into independent subcolumns. The domain averaged result from those individually computed subcolumn radiative profiles can then represent the domain approximation. Because of the exorbitant computational cost of a fully independent column approximation (ICA) method, an alternate approach, which is a Monte Carlo independent column approximation (McICA) (Barker et al. 2002; Pincus et al. 2003), is used in the new CFS model. In McICA, a random column cloud generator samples the model layered cloud into subcolumns and pairs each column with a pseudomonochromatic calculation in the radiative transfer model. Thus the radiative computational expense does not increase, except for a small amount of overhead cost attributed to the random number generator.
- In calculating cloud optical thickness, all the cloud condensate in a grid box is assumed to be in the cloudy region. So the in-cloud condensate mixing ratio is computed by the ratio of grid mean condensate mixing ratio and cloud fraction when the latter is greater than zero.
- The CO₂ mixing ratio used in these retrospective forecasts includes a climatological seasonal cycle superimposed on the observed estimate at the initial time.
- The Noah land surface model (Ek et al. 2003) used in CFSv2 was first implemented in the GFS for operational medium-range weather forecasts (Mitchell et al. 2005) and then in the CFSR (Saha et al. 2010). Within CFSv2, Noah is employed in both the coupled land–atmosphere–ocean model to provide land surface prediction of surface fluxes (surface boundary conditions), and in the Global Land Data Assimilation System (GLDAS) to provide the land surface analysis and evolving land states. While assessing the predicted low-level temperature and the land surface energy and water budgets in the CFSR reforecast experiments, two changes to CFSv2/Noah were made. First, to address low-level warm bias (notable in midlatitudes), the CFSv2/Noah vegetation parameters and rooting depths were refined to increase evapotranspiration, which, along with a change to the radiation scheme (RRTM in GFS and CFSR, and now McICA in CFSv2), helped to improve the predicted 2-m air temperature over land. Second, to accommodate a change in soil moisture climatology from GFS to CFSv2, Noah land surface runoff parameters were nominally adjusted to favorably increase the predicted runoff (see section 5 for more comments).

3. The design of the retrospective and real-time forecasts: Considerations for operational implementation

a. Nine-month retrospective predictions

The official release of the Climate Prediction Center (CPC) operational seasonal prediction is on the third Thursday of each month. In this case, given operational protocol (several teleconference meetings with partners must be held prior to the release) products must be ready almost one week earlier (i.e., by the second Friday of the month). For these products to be ready, the latest CFSv2 run that can be used is from the 7th of each month. These considerations are adhered to in the hindcasts

(even when the release date is after the 15th, since the very latest date of release can be the 21st of a month).

The retrospective 9-month forecasts have initial conditions of the 0000, 0600, 1200, and 1800 UTC cycles for every 5th day, starting from 0000 UTC 1 January of every year, over the 29-yr period 1982–2010. There are 292 forecasts for every year for a total of 8468 forecasts (see appendix A). Selected data from these forecasts may be downloaded from the National Climatic Data Center (NCDC) web servers (see appendix D).

The retrospective forecast calendar (appendix B) outlines the forecasts that are used each calendar month, to estimate proper calibration and skill estimates, in such a way as to mimic CPC operations.

This results in an ensemble size of 24 forecasts for each month except November, which has 28 forecasts.

Smoothed calibration climatologies have been prepared from the forecast monthly means and time series of selected variables and is available for download (see appendix D).

Having a robust interpolated calibration for each cycle, each day, and each calendar month allows CPC to use real-time ensemble members (described in section 3c) as close as possible to release time.

b. First season and 45-day retrospective forecasts

These retrospective forecasts have initial conditions from every cycle (0000, 0600, 1200, and 1800 UTC) of every day over the 12-yr period from January 1999 to December 2010. Thus, there are approximately 365×4 forecasts per year, for a total of 17520 forecasts. The forecast from the 0000 UTC cycle was run out to a full season, while the forecasts from the other three cycles (0600, 1200, and 1800 UTC) were run out to exactly 45 days (see appendix A for the reforecast configuration). Selected data from these forecasts may be downloaded from the NCDC (see appendix D).

Smoothed calibration climatologies have been prepared from the forecast time series of selected variables (<http://cfs.ncep.noaa.gov/cfsv2.info/CFSV2.Calibration.Data.doc>) and is available for download (see appendix D). It is essential that some smoothing is done when preparing the climatologies of the daily time series, which are quite noisy.

Having a robust calibration for each cycle, each day, and each calendar month allows CPC to use ensemble members very close to the release time of their 6–10-day and week 2 forecasts. They are also exploring the possibility of using the CFSv2 predictions in the week 3–6 range.

c. Operational configuration

The initial conditions for the CFSv2 retrospective forecasts are obtained from the CFSR, while the real-time operational forecasts obtain their initial conditions from the real-time operational CDAS, version 2 (CDASv2).

Great care was made to unify the CFSR and CDASv2 in terms of the same cutoff times for data input to the atmosphere, ocean, and land surface components in the data assimilation system. Therefore, there is greater utility of the new system, as compared to CFSv1 (which had a lag of a few days), since the CFSv2 initial conditions are made completely in real time. This makes it possible to use them for the subseasonal (week 1–6) forecasts. There are 16 CFSv2 runs per day in operations; four out to 9 months, three out to 1 season, and nine out to 45 days (see appendix C). Operational real-time data may be downloaded from the official site (see appendix D).

4. Results in terms of skill

In this section, we present a limited analysis of skill of CFSv2 prediction for the subseasonal range, “deterministic” seasonal prediction, placing CFSv2 in context with other models, and probabilistic long lead prediction. More detailed analyses will be published in subsequent papers.

a. Subseasonal prediction

Figure 1 shows the skill, as per the bivariate anomaly correlation (BAC) [Lin et al. 2008, their Eq. (1)] of CFSv2 forecasts in predicting the MJO, as expressed by the Wheeler and Hendon (2004) index, using two EOFs of combined zonal wind and outgoing longwave radiation (OLR) at the top of the atmosphere. The period is 1999–2009. On the left is CFSv2; on the right is CFSv1. Both are subjected to systematic error correction (SEC) as described in detail in Zhang and Van den Dool (2012, hereafter ZV). The BAC stays above the 0.5 level (the black line) for two to three weeks in the new system, whereas it was at only one week in the old system. Both models show a similar seasonal cycle in forecast skill with maxima in May–June and November–December respectively, and minima in between. Correlations were calculated as a function of lead for each starting day; that is, for any given lead, there were only 11 cases, one case for each year. Figure 1 (both panels) was then plotted with day of the year along the vertical axis (months are labeled for reference) and forecast lead along the horizontal axis, with the correlation multiplied by 100 being contoured. To suppress noise, a light smoothing was applied in the vertical (i.e., over adjacent starting days). The right panel in Fig. 1 for CFSv1 would have holes because no CFSv1 forecasts originated from the 4th through 8th, 14th through 18th, and 24th through 28th of each month. In the CFSv1 graph, the smoothing also serves to mask these holes.

Note that, consistent with CPC operations, which still uses the older R2 reanalysis (Kanamitsu et al. 2002) for the MJO, we verify both CFSv1 and CFSv2 against R2 based observations of the real-time multivariate MJO (RMM) 1 and 2 indices, using an observed climatology

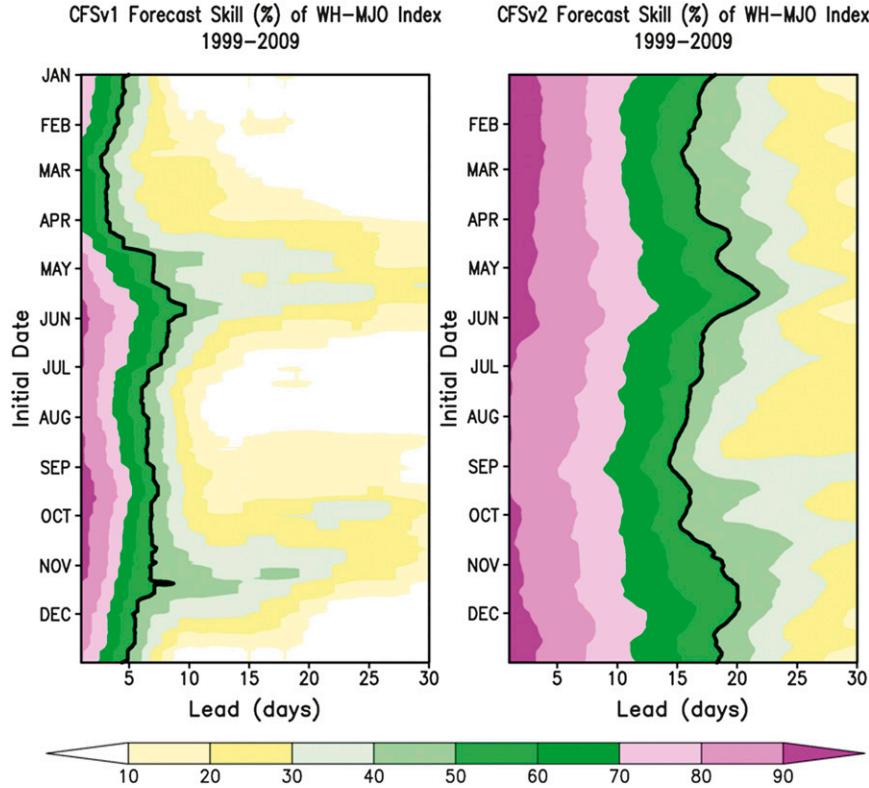


FIG. 1. The bivariate anomaly correlation (BAC) $\times 100$ of CFS in predicting the MJO for period 1999–2009, as expressed by the Wheeler and Hendon (2004) index (two EOFs of combined zonal wind and OLR), for (left) CFSv1 and (right) CFSv2. Both are subjected to systematic error correction. The black lines indicate the 0.5 level of BAC.

(1981–2004) based on R2 winds and satellite OLR. Note further that although the hindcasts are for 1999–2009, one can express anomalies relative to some other period (here for 1981–2004; see ZV for details on how that was done).

It is quite clear that CFSv2 has much higher skill than CFSv1 throughout the year. In fact, this is the improvement made by half a generation (~ 15 yr) of work by many in both data assimilation and modeling fields (taking into account that CFSv1 has rather old R2 atmospheric initial conditions as its weakest component). One rarely sees such a demonstration of improvement. This is because operational atmospheric NWP models are normally abandoned when a new model comes in. But in the application to seasonal climate forecasting, systems tend to have a longer lifetime. This gave us a rare opportunity to compare two frozen models that are about 15 years apart in vintage.

The causes for the enormous improvement seen in Fig. 1 are probably very many, but especially the improved initial states in the tropical atmosphere and the consistency of the initial state and the model used to make the forecasts play a role. Further research should bring out the importance of coupling to the ocean (Vitart et al. 2007) and its quantitative contribution to skill.

Further results and discussion on the MJO in CFSv1/v2 can be found in ZV.

We studied the MJO results with and without the benefit of SEC for both CFSv1 and CFSv2. We found that SEC results in improvements for either CFS over raw forecasts, more often than not, and overall the improvement in CFSv2 is between 5 and 10 points (see Fig. 2 in ZV), which could be the equivalent of several new model implementations. This is a strong justification for making hindcasts.

As is the case with CFSv2, version 1 did benefit noticeably from the availability of its hindcasts. While the distribution of the improvement with lead and season is different for CFSv1, the overall annual mean improvement is quite comparable (see Fig. 3 in ZV). Both CFSv1 and CFSv2 appear to gain about 2–3 days of prediction skill by applying an SEC. Obviously, the model and data assimilation improvements between 1995 and 2010 count for much more than the availability of the hindcasts, but the latter do correspond to a few years of model improvement.

b. Seasonal prediction out to 9 months

The anomaly correlation of 3-month mean sea surface temperature (SST) forecasts is shown in Fig. 2 for 3- and

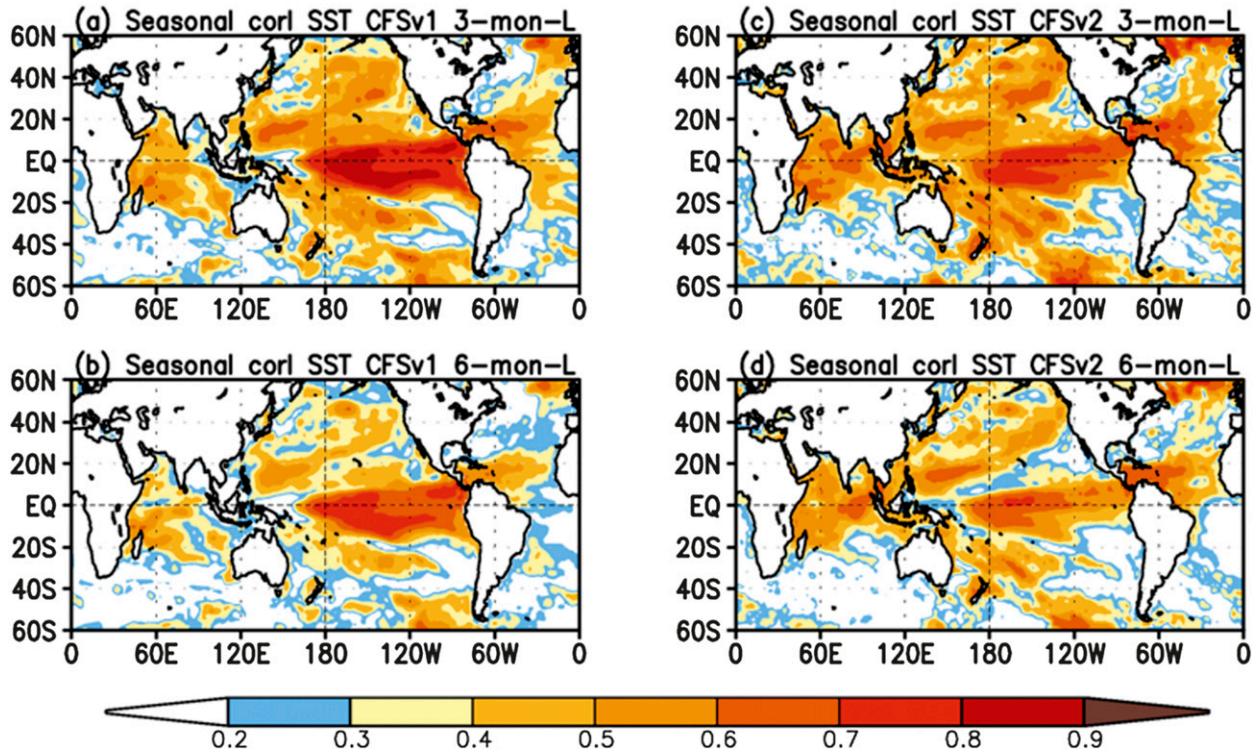


FIG. 2. Anomaly correlation of 3-month-mean SST between model forecasts and observation: (a) 3-month lead CFSv1, (b) 6-month lead CFSv1, (c) 3-month lead CFSv2, and (d) 6-month lead CFSv2. Contours are plotted at an interval of 0.1.

6-month lead times. The forecasts are verified against optimum interpolation SST, version 2 (OISST2; Reynolds et al. 2002). A lagged ensemble mean of 20 members from each starting month is used to compute the correlation. Similar spatial distributions of the correlation are seen in both CFS versions, with relatively higher skill in the tropical Pacific than the rest of the globe. Overall, the skill for CFSv2 is improved in the extratropics with an average anomaly correlation poleward of 20°S and 20°N of 0.34 (0.27) for 3-month lead (6-month lead) compared to the corresponding CFSv1 anomaly correlation of 0.31 (0.24). In the tropical Pacific, the CFSv2 skill is slightly lower than that of CFSv1 for NH winter target periods [e.g., December–February (DJF)], but has less of a spring and summer minimum. This lower CFSv2 skill in DJF is related to the climatology shift with significantly warmer mean predicted SST in the tropical Pacific after 1999, compared to that before 1999, which is likely due to the start of assimilating the Advanced Microwave Sounding Unit (AMSU) satellite observations in the CFSR initial conditions in 1999 [see section 5a and Kumar et al. (2012) for a lengthier discussion].

Figure 3 compares the amplitude of interannual variability between the SST observation and forecasts at 3- and 6-month lead times. The largest variability over the globe is related to the ENSO variability in the tropical

Pacific. The variability of the forecast is computed as the standard deviation based on anomalies of individual members (rather than the ensemble mean). Both CFSv1 and CFSv2 are found to generate stronger variability than observed over most of the globe. In particular, the forecast amplitude is larger than the observed in the tropical Indian Ocean, eastern Pacific, and northern Atlantic. Compared to CFSv1, CFSv2 produced more reasonable amplitude. For example, the strong variability in CFSv1 in the tropical Pacific is substantially reduced, and the variability in CFSv2 in the northern Pacific is comparable to the observation (Figs. 3b,c), while the CFSv1 variability in this region is too strong (Figs. 3d,e).

Figure 4 provides a grand summary of the skill of monthly prediction as a function of target month (horizontal axis) and lead (vertical axis). For precipitation and 2-m temperature the area is all of NH extratropical land, and the measure is the anomaly correlation evaluated over all years (1982–2010). We compare CFSv2 directly to CFSv1, over the same years. One may also compare this to Figs. 1 and 7 in S06 for CFSv1 alone (and six fewer years). The top panels of Fig. 4 show that prediction of temperature has substantially improved for all leads and all target months from CFSv1 to CFSv2. The statistical significance is evident. We believe this is caused primarily by increasing CO₂ in the initial

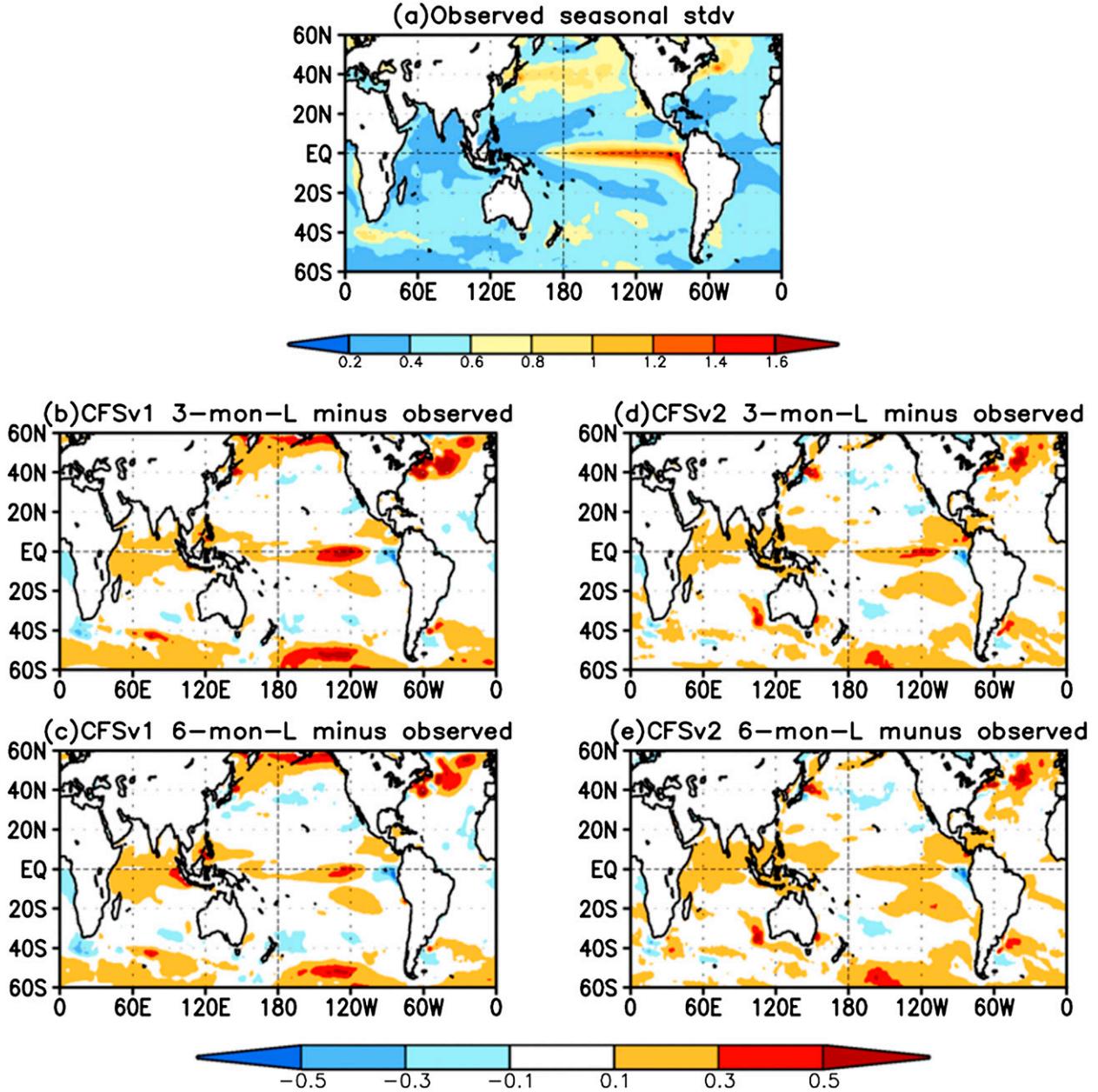


FIG. 3. Standard deviation of 3-month-mean SST forecasts (K): (a) observation, (b) 3-month lead CFSv1 minus observation, (c) 6-month lead CFSv1 minus observation, (d) 3-month lead CFSv2 minus observation, and (e) 6-month lead CFSv2 minus observation. Contours are plotted at an interval of 0.2 from 0.2 to 1.6 in (a) and from -0.5 to 0.5 in (b)–(e).

conditions and hindcasts,² and possibly by eliminating some soil moisture errors (and too cold temperatures) that have plagued CFSv1 in real time in recent years. The

positive impact of increasing CO₂ was to be expected as analyzed by Cai et al. (2009) for CFSv1, especially at long leads. Still, the skill is only modest, a mere correlation of 0.20.

While the skill for 2-m temperature is modest, the skill for precipitation forecasts (middle panels of Fig. 4) for monthly mean conditions over NH land remains less than modest. Except for the first month (lead 0), which is essentially weather prediction in the first 2 weeks, there

²CO₂ is not increased during a particular hindcast, but through the initial conditions; for example, hindcasts for 2010 are run at much higher CO₂ (which is maintained throughout the forecast) than for hindcasts in 1982. In CFSv1, a single CO₂ value valid in 1988 was used for all years.

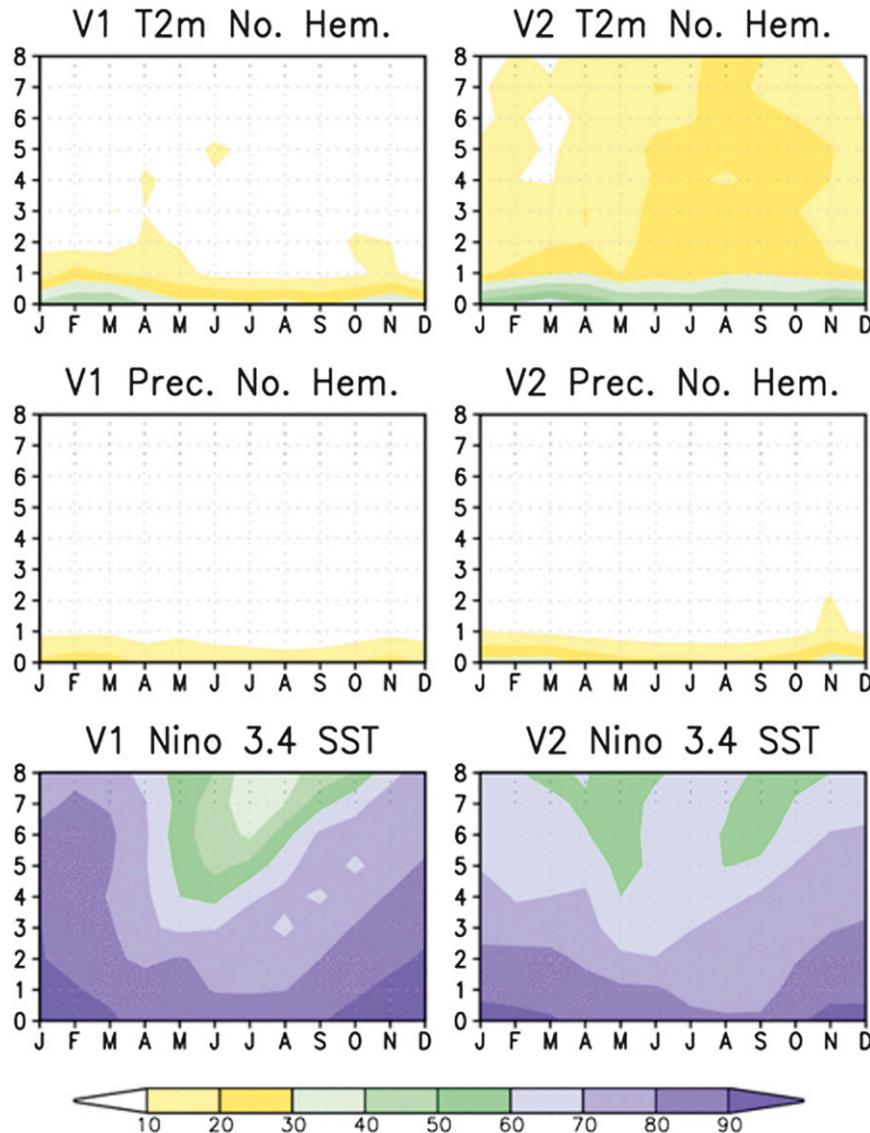


FIG. 4. Evaluation of anomaly correlation as a function of target month (horizontal axis) and forecast lead (vertical axis) for (left) CFSv1 and (right) CFSv2. (top) Monthly 2-m temperature over NH land, (middle) monthly precipitation over NH land, and (bottom) SST in the Niño-3.4 area. The scale is the same for all six panels. Except for the years added, the CFSv1 entries in this figure should correspond to the figures in S06.

is no skill at all (correlation over 0.1), which is a sobering conclusion. CFSv2 is not better than CFSv1. Although these systems have skill in precipitation prediction over the ocean (in conjunction with ENSO), the benefit of ENSO skill in precipitation over land appears small or washed away by other factors.

The bottom panels of Fig. 4 show that both systems have decent skill in predicting the SST at grid points inside the Niño-3.4 area (5°S – 5°N , 170° – 120°W). The skill for the Niño-3.4 area overall has not improved for CFSv2 versus CFSv1, but the seasonality has changed. Skill has

become lower at long lead for winter target months and higher for summer target months, thereby decreasing the spring barrier. In general, CFSv2 is better in the tropics than CFSv1 for SST prediction (see Fig. 2), but Niño-3.4 is the only area where this is not so.

c. CFSv2 seasonal prediction in context of other model predictions

The development of CFSv2 can be placed in context by making a comparison to other models (with similar applications to seasonal prediction) such as the ones

TABLE 1. The anomaly correlation (AC) of 0.5-month lead, 3-month mean prediction for 1982–2010 by seven models used in the U.S. National Multimodel Ensemble (NMME) and the equal-weights average at the bottom. Apart from NCEP’s CFSv1 and v2 the other models are only identified by letters A–E. The AC was calculated for NH T2m and prate over land, NH SST (ocean), and Niño-3.4 index, for all start months combined.

	NH SST (K)	Niño-3.4 SST (K)	NH T2m (K)	NH prate (mm day ⁻¹)
CFSv1	0.29	0.82	0.11	0.10
CFSv2	0.41	0.82	0.29	0.12
A	0.27	0.81	0.12	0.10
B	0.27	0.82	0.12	0.11
C	0.42	0.80	0.25	0.12
D	0.34	0.78	0.23	0.08
E	0.14	0.80	0	0.04
NMME	0.45	0.87	0.27	0.17

used in the U.S. National Multimodel Ensemble (NMME). NCEP plays a central role in this activity that was started in real time in August 2011. The seven participating models are all global coupled atmosphere–ocean models developed in the United States [see Kirtman et al. (2014) for an overview]. Predictions made by all these models—CFSv1, CFSv2, models from the National Aeronautics and Space Administration (NASA), GFDL, and the National Center for Atmospheric Research (NCAR), and two models from the International Research Institute for Climate and Society (IRI)—were verified over exactly the same years.

Table 1 shows the anomaly correlation for 0.5-month lead seasonal prediction for SST, 2-m air temperature (T2m), and precipitation rate (prate). These are aggregate numbers for all start months and large areas combined. For SST (whether it is NH or Niño-3.4 SST) CFSv2 performs well, but so do several or all of the other models, and the equal-weight NMME (shown on the bottom row of Table 1) is the best of all. The same applies for prate, but we note that the skill for prate over NH land is extremely low for all the models. However, for NH T2m over land, CFSv2 is the best model to such a degree that the NMME average of all models drags down the score of CFSv2.

Table 2 shows the interannual standard deviation of individual members around the model climatology, all start months combined. This distributional property, in a grandly aggregated sense, is at least as large as that observed for any model (bottom row), and CFSv2 is no exception. Not long ago, models were deemed to be underdispersive, and that was the main reason why the multimodel approach would improve scores, especially probabilistic scores. But for the 3-month-mean variables shown here, this is no longer true.

TABLE 2. As in Table 1, but for the models’ standard deviation around its own climatology.

	NH SST (K)	Niño-3.4 SST (K)	NH T2m (K)	NH prate (mm day ⁻¹)
CFSv1	0.76	1.20	1.53	0.68
CFSv2	0.85	1.13	1.42	0.61
A	0.73	0.97	1.44	0.49
B	0.73	1.19	1.40	0.48
C	0.69	1.20	1.62	0.57
D	0.81	1.19	1.42	0.64
E	0.82	1.17	1.51	0.61
Obs	0.65	0.89	1.47	0.46

The distributional parameters being roughly correct in a grand sense does not preclude standard deviations being too small, or too large, in specific areas and specific seasons, as we saw already in section 4b. Additional insights can be gained from verification of probabilistic verification in the next section.

d. Probabilistic seasonal prediction verification

This section follows the CFSv1 paper (S06, section 4b, 3495–3501) quite precisely, both in terms of the definition of “reliability” and the Brier skill score (BSS) and the corresponding figures (Figs. 17 and 18 in S06) that will be shown. The difference is an additional six years for CFSv1, and an exact comparison between CFSv1 and CFSv2 over the period 1982–2009, all start months, for a probabilistic prediction of the terciles of monthly Niño-3.4 SST.

Figure 5 shows the reliability comparison, which is often considered a make or break selling point for probabilistic prediction. Plotted are observed frequency against predicted probability in four bins, for each of the three terciles. Compared to perfection (the black line at 45°), we see a clear model improvement from CFSv1 to CFSv2. Keep in mind that CFSv2 was reduced to 15 members only (more are available) to be on an equal footing with CFSv1 in this display, as far as the number of ensemble members is concerned. With 15 members each, CFSv2 has better reliability than CFSv1. One can see this especially at lead 8 months, and for the notoriously difficult “near normal” tercile. Using more ensemble members (not shown) further improves reliability, so CFSv2 is a large improvement over CFSv1 in reliability, even though some problems were noted in section 4b.

Figure 6 shows a comparison of the BSS, with CFSv1 (CFSv2) on the left (right). The BSS (solid line) has been decomposed in the usual contributions to BSS by reliability (dashed dotted) and resolution (dotted). We do not show the third component called uncertainty since, by definition, this is the same for both systems. Keep in mind that reliability (shown in another way in Fig. 5) has

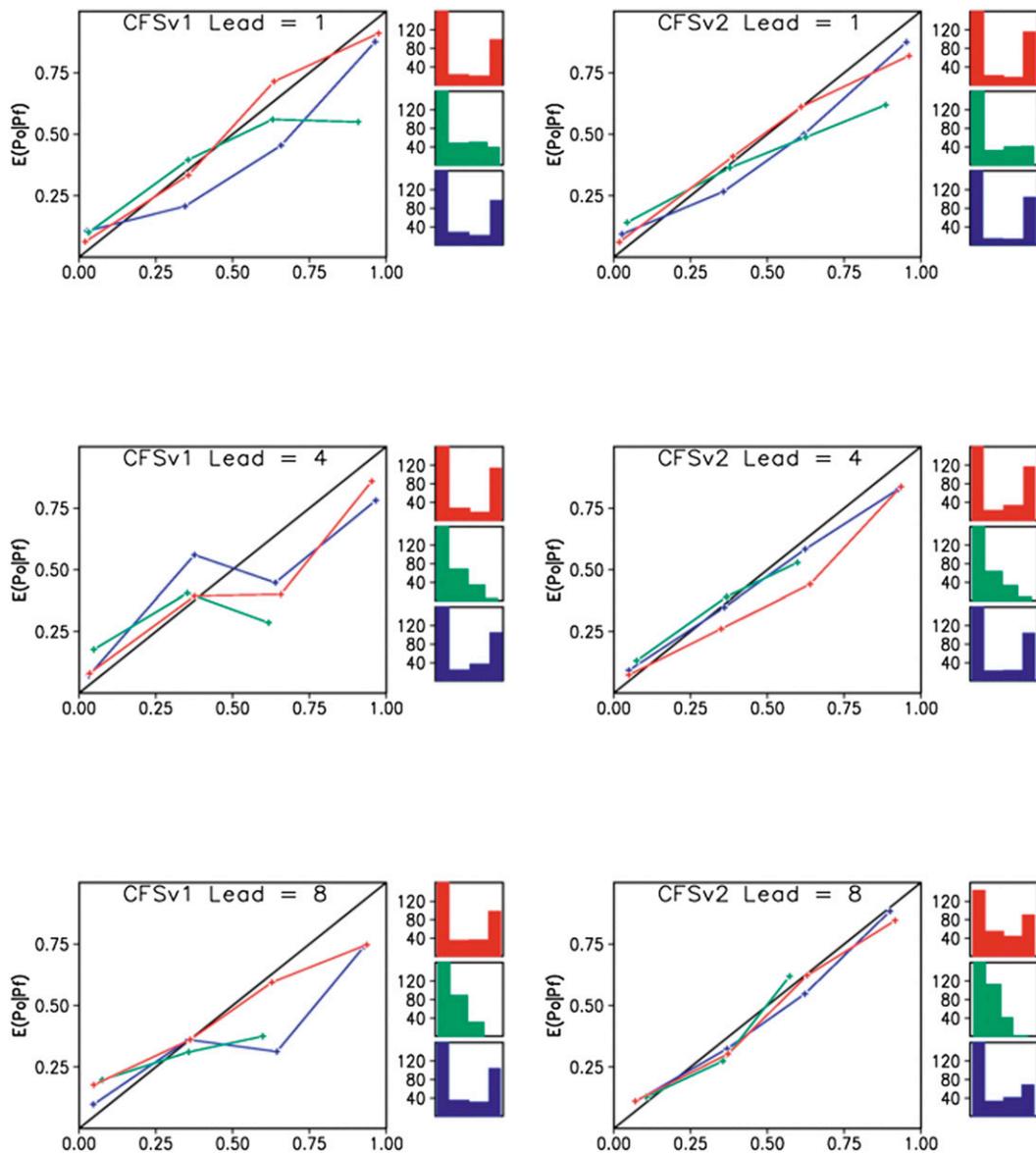


FIG. 5. Reliability diagrams of CFS probability predictions that Niño-3.4 SST prediction will fall in the upper (red), middle (green), or lower (blue) tercile of the observed climatological distribution for (top) lead 1, (middle) lead 4, and (bottom) lead 8 months for (left) CFSv1 and (right) CFSv2, both for the period 1982–2009. The color-coded small histograms indicate the frequency of forecasts in the bins 0–0.25, 0.25–0.50, 0.50–0.75 and 0.75–1.00, respectively. The black line at 45° is for perfect reliability. Data period is 1982–2009; cross-validation (2 yr withheld) was applied.

to be numerically small and resolution numerically high for a well-calibrated system (i.e., to contribute to a high BSS). Comparison of the left and right diagrams in Fig. 6 indicates that CFSv2 is an improvement over CFSv1, especially for longer leads and the near-normal tercile. In terms of their contribution to the total BSS, both resolution and reliability have helped to make CFSv2 better.

We did calculate the BSS for T2m over the United States (presented as maps in Fig. 7), but neither CFSv1

nor CFSv2 has positive BSS overall for this domain, unless a very laborious calibration is carried out. When only the mean and the standard deviation are corrected and both systems are allowed 15 members (the maximum for CFSv1), the BSS scores for CFSv1 are slightly negative while those for CFSv2 are also negative, but closer to zero. It is only when all 24 members are used that CFSv2 has positive BSS scores overall (see bottom row). The skill is very modest nevertheless, with values such as +0.02 compared to 0.4–0.5 for Niño-3.4 SST in

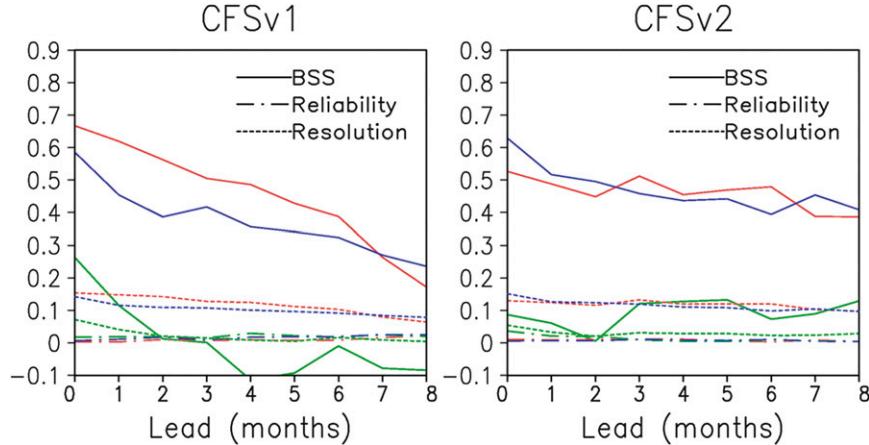


FIG. 6. The BSS (solid), reliability (dashed dotted), and resolution (dashed) as function of the lead time, for Niño-3.4 SST prediction, for (left) CFSv1 and (right) CFSv2, both for the period 1982–2009. The three terciles are upper (red), middle (green), and lower (blue). A cross validation was applied (2 yr withheld).

Fig. 6. More aggressive suppression of noise and more calibration may improve the outcome further, but this is outside the scope of this paper. In spite of many (modest) improvements in these global models, we continue with the same basic discrepancy of having high skill for SST in the tropics, but small and often negligible skill for T2m and especially prate over land.

5. Diagnostics

While section 4 contains results of CFSv2 (versus CFSv1) in terms of forecast skill, we also need to report on some diagnostics that describe model behavior. Even without strict verification, one may judge models as being “reasonable” or not. In section 5a we compare the systematic errors globally in SST, T2m, and prate between CFSv2 and CFSv1. Next the surface water budget, which was mentioned in section 2 as being the subject of tuning, is discussed in section 5b. We also present some results on sea ice prediction (without a strict verification) since this is an important emerging aspect of global coupled models. CFSv1 had an interactive ocean only between 65°N and 75°S latitudes, with climatological sea ice in the polar areas. The aspect of a global ocean and interactive sea ice model in the CFSv2 is new in the seasonal modeling context at NCEP.

a. Evolution of systematic error

The systematic error is approximated as the difference in the predicted and observed climatology over a common period (1982–2009). We describe the systematic error here under the header “model diagnostics” because it describes one of the net effects of modeling

errors. While the systematic error has a bearing on the forecast verification in section 4, its impact on the verification was largely removed since we made hindcasts to apply the correction. Figure 8 shows global maps of the annual mean systematic error for the variables, from top to bottom, T2m, prate, and SST. On the left is CFSv1 and on the right CFSv2, so this is the evolution of the systematic error in an NCEP model from about 2003 to about 2010. The headers display numbers for the mean and the root-mean-square (rms) difference averaged over the map. For all three parameters CFSv2 has lower rms values, which is a definite sign of a better model. Lower rms values globally do not preclude some areas having a larger systematic error; for instance, the cold bias over the eastern United States is stronger in CFSv2. Figure 8 is for a lead of 3 months, but these maps look very similar for all leads from 1 to 8 months. Apparently these models settle quickly in their respective climatological distributions. The systematic error has a sign, so the map mean shows a cold bias (-0.3 K) and a wet bias (from $+0.6$ to -0.7 mm day^{-1}) globally averaged in both models. Of these three maps, the one for T2m has changed the least between the CFSv1 and CFSv2 versions, and the maps for prate have changed more, especially in the tropics, but note that the SST systematic error has changed beyond recognition from CFSv1 to CFSv2.

Another “evolution” of the systematic error is displayed in Fig. 9 where we compare, just for CFSv2, the systematic error as calculated for 1982–98 (left) and 1999–2009 (right). In a constant frozen system the maps on the left and right should be the same, except for sampling error. From a global standpoint these maps are quite similar, but if one focuses on the tropical Pacific we

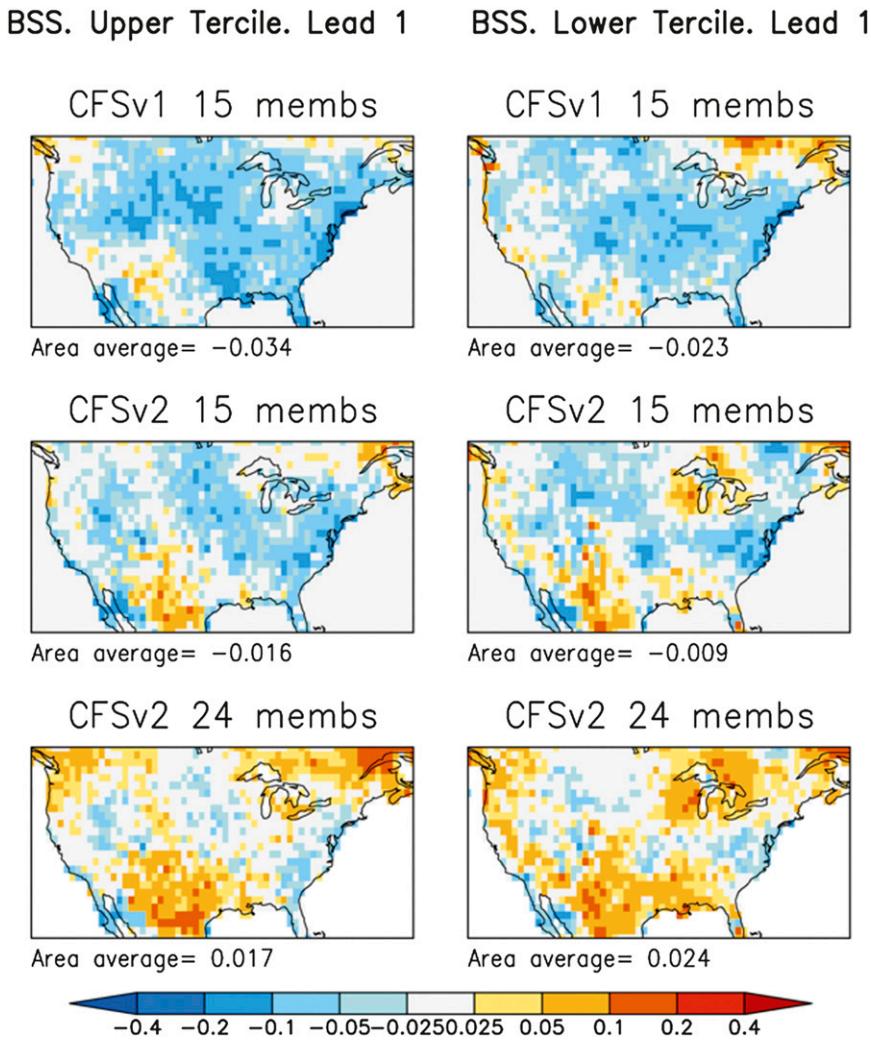


FIG. 7. The BSS of prediction of the probability of terciles of monthly T2m at lead 1 month, for the (left) upper and (right) lower tercile, for (top) CFSv1 with its 15 members, (middle) CFSv2 with only 15 members, and (bottom) CFSv2 with all 24 members. All start months are combined. Period is 1982–2009. Below each map is the map integrated BSS value.

should point out a difference in the SST maps in the Niño-3.4 area. The later years (after 1998) have a negligible systematic error, whereas the earlier years have a modest cold bias. Perhaps this makes perfect sense because in later years the models are initialized with much more data. On the other hand, it is a problem in systematic error correction if the systematic error is nonstationary (Kumar et al. 2012).

The SST in the Niño-3.4 area is important as this area is often chosen as the most sensitive single indicator of ENSO. One may surmise that changes in the systematic error in prate are caused by the model predicted SST being warmer in later years. Indeed, one can see large changes in the Pacific basin in the ITCZ in the NH, the South Pacific convergence zone (SPCZ) in the SH, and

the rainfall in the western Pacific (see middle row in Fig. 9). The rest of the globe is not impacted so obviously in terms of either SST or prate, not even the tropical Atlantic and Indian Oceans. The systematic error in T2m over land appears oblivious to changes in SST in the Pacific.

The causes of this discontinuity are most probably related to ingest of new data systems, most notably AMSU in late 1998 (Saha et al. 2010, 3495–3501), which caused an enormous increase in satellite data to be assimilated. Such issues need to be addressed in version 3 of the CFS (CFSv3), and specifically in any reanalyses that are made in the future to create initial conditions (land, ocean, and atmosphere) for CFSv3 or systems elsewhere. But, for the time being, we need to address

Annual Mean Systematic Error (1982–2009) for Lead 3

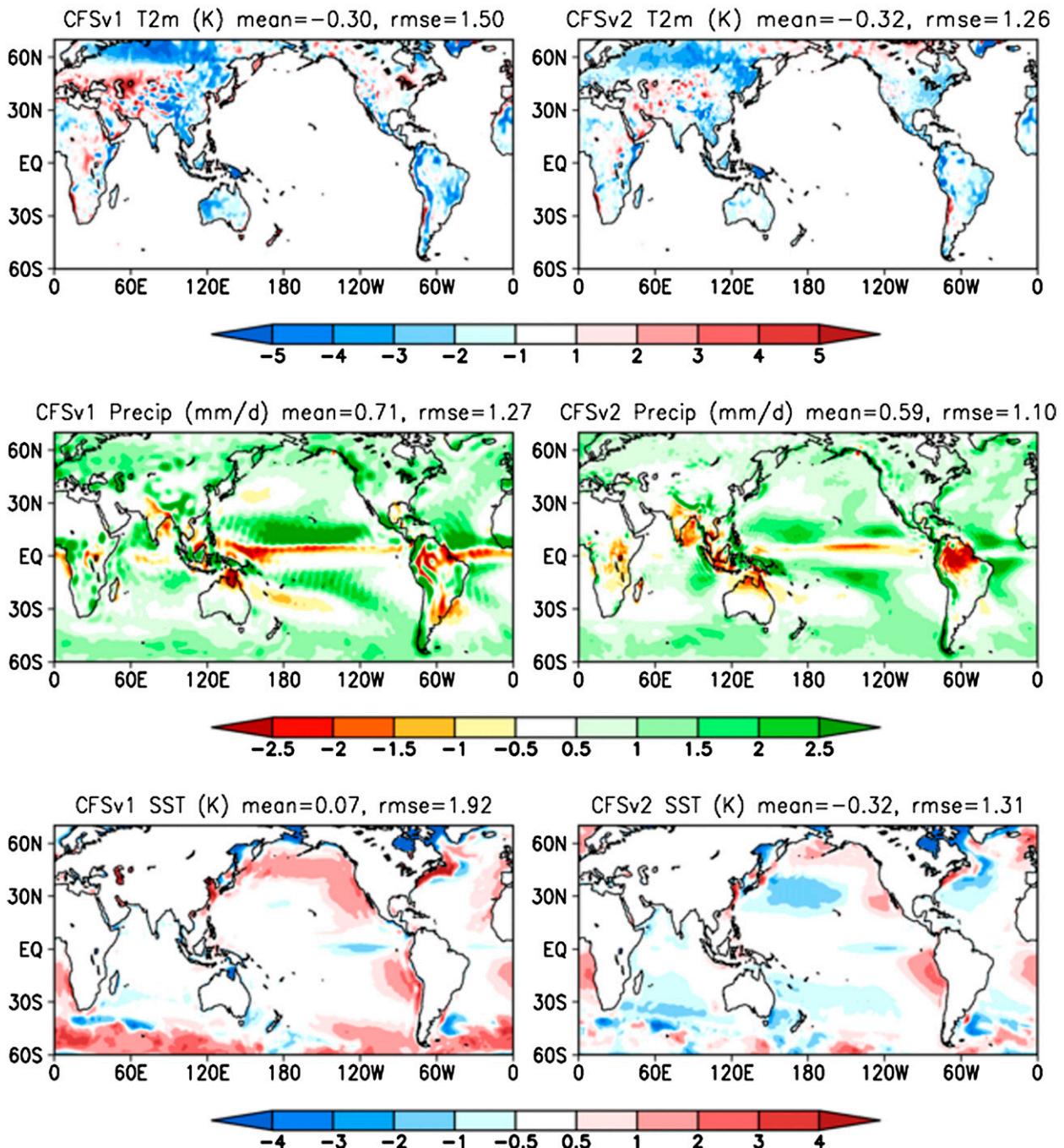


FIG. 8. The annual mean systematic error in three parameters [SST (K), T2m (K), and prate (mm day^{-1})] at lead 3 months evaluated as the difference between the predicted and observed climatology for the full period 1982–2009, for (left) CFSv1 and (right) CFSv2. The header in each panel contains the root-mean-square difference, as well as the spatial mean difference. Contours and colors are as indicated by the bar underneath.

CFSv2 Annual Mean Systematic Error for Lead 3

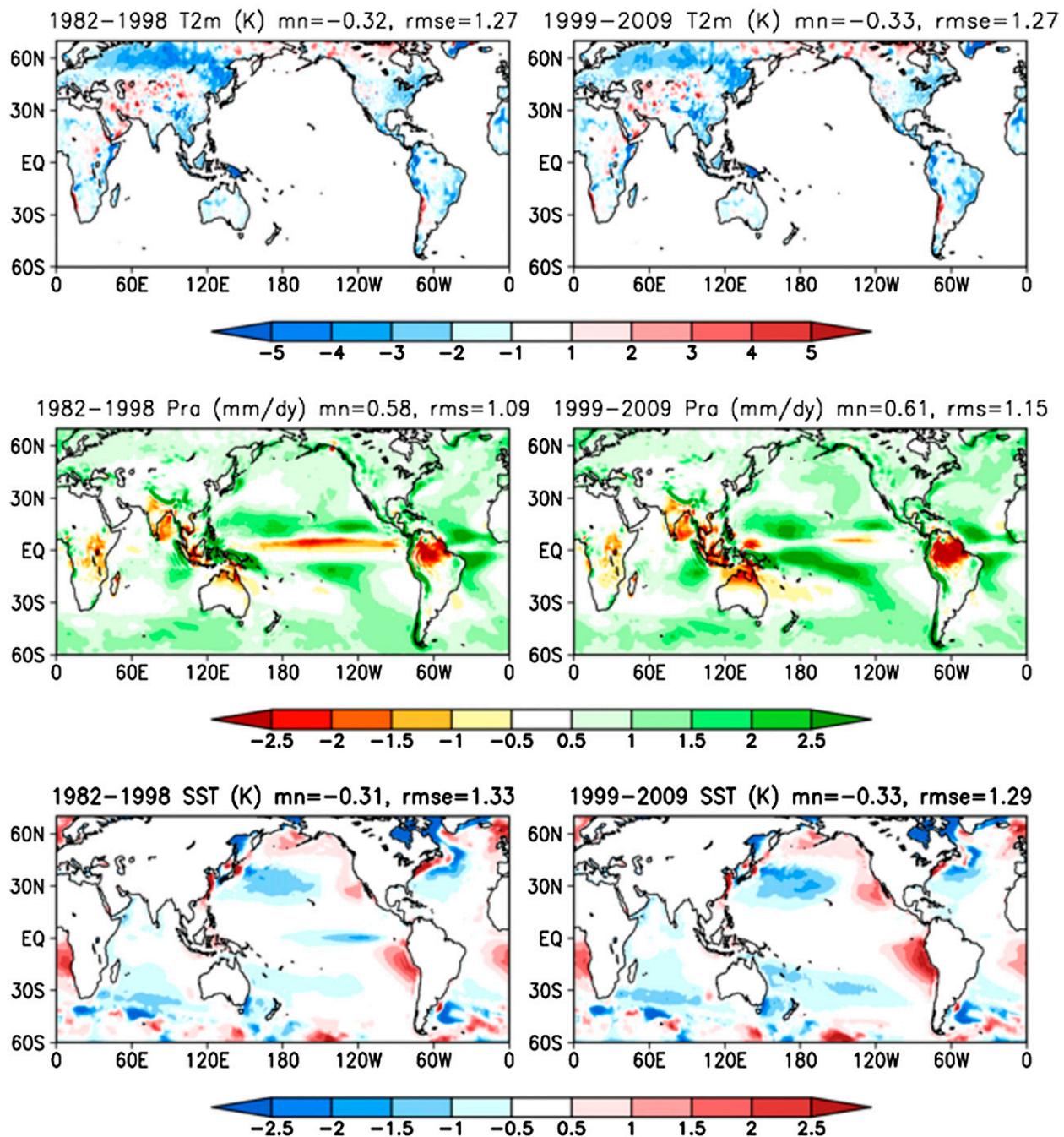


FIG. 9. The annual mean systematic error in three parameters [SST (K), T_{2m} (K), and prate (mm day^{-1})] at lead 3 months evaluated as the difference between CFSv2's predicted and observed climatology, for (left) 1982–98 and (right) 1999–2009. The header in each panel contains the root-mean-square difference, as well as the spatial mean difference. Contours and colors are as indicated by the bar underneath.

TABLE 3. Surface water budget comparison of CFSv1, CFSR, and CFSv2 for summer (JJA) and winter (DJF). Values are averages for NH land.

	CFSv1 (JJA/DJF)	CFSR (JJA/DJF)	CFSv2 (JJA/DJF)
Precipitation (mm day ⁻¹)	3.3/1.6	3.2/1.4	2.7/1.3
Evaporation (mm day ⁻¹)	2.5/1.1	2.2/0.89	2.1/0.71
Run off (mm day ⁻¹)	0.56/0.16	0.16/0.04	0.22/0.06
Soil moisture (mm)	441/476	510/514	502.43/501.37
Snow water (mm)	0.09/4.1	0.02/4.2	0.01/6.5

how we apply the systematic error correction in the CFSv2 hindcasts, and in real-time (subsequent) CFSv2 forecasts. Our recommendation is that the full 30-yr period (1982–2012 is now available for CFSv2) be used for all fields globally with the exception of SST and prate in the Pacific Ocean basin where it seems better to use a split climatology. Therefore, for real-time forecasts, the systematic error correction for prate and SST in the Pacific should be based on 1999–present. This does not mean that anomalies should be presented as departures from the 1999–present climatology (see ZV for that distinction).

b. Land surface

Table 3 shows a comparison of surface water budget terms averaged over the Northern Hemisphere land between CFSv1 and CFSv2 and with CFSR. The quantities in CFSv1 and CFSv2 are computed from seasonal ensemble means covering a 29-yr period (1982–2010), where the CFSv1 is based on seasonal predictions from 15 ensemble members whose initial conditions are from mid-April to early May (9–13 April, 19–23 April, and 29 April–3 May at 0000 UTC) for the summer season [June–August (JJA)], and from mid-October to early November (9–13 October, 19–23 October, and 29 October–3 November) for the winter season (DJF), while the CFSv2 is based on 24 ensemble members (initial conditions from 4 cycles of the 6 days between 11 April and 6 May with 5 days apart) for summer and 28 ensemble members (initial conditions from 4 cycles of 7 days between 8 October and 7 November with 5 days apart) for winter season, respectively.

Compared to the CFSR, precipitation (snow in winter) in the CFSv1 is higher in both seasons, which yields higher values for both evaporation and runoff. The higher evaporation in the summer season in the CFSv1 yields a much larger seasonal variation in soil moisture (but lower absolute values) than in both CFSR and CFSv2. In contrast, precipitation in the CFSv2 is considerably lower than in both CFSv1 and CFSR, consistent with lower evaporation in the CFSv2. While less than in the CFSv1, runoff in the CFSv2 is more than in CFSR, indicating that soil moisture is a more important source

for surface evaporation in the CFSv2; this higher runoff in winter season leads to a damped seasonal variation in soil moisture since soil moisture is recharged in winter when evaporation is at its minimum. The increases in both surface evaporation from root-zone soil water and runoff production are consistent with the changes made to vegetation parameters and rooting depths in CFSv2 (see comments in section 2) to address high biases in predicted T2m, and the accommodated changes in soil moisture climatology and surface runoff parameters. The good agreement in soil moisture between CFSR and CFSv2 is expected because they use the same Noah land model.

c. Sea ice

Sea ice prediction is challenging and relatively new in the context of seasonal climate prediction models. Sea ice can form or melt and can move with wind and/or ocean current. Sea ice interacts with both the air above and the ocean beneath and it is influenced by, and has an impact on, the air and ocean conditions. The CFSv2 sea ice component includes a dynamic/thermodynamic sea ice model and a simple assimilation scheme, which are described in detail in Saha et al. (2010). One of the most important developments in CFSv2, compared to CFSv1, is the extension of the CFS ocean domain to the global high latitudes and the incorporation of a sea ice component.

The initial condition (IC) for ice in the CFSv2 hindcasts is from CFSR as described in Saha et al. (2010). For sea ice thickness, there are no data available for assimilation, and we suspect there is a significant bias of sea ice thickness in the CFSv2 model that causes the sea ice to be too thick in the IC. For the sea ice prediction, sea ice appears too thick and certainly too extensive in the spring and summer. Figure 10 shows the mean September sea ice concentration from 1982 to 2010, and the bias in the predicted mean condition at lead times of 1 month (15 August IC), 3 months (15 June IC), and 6 months (15 March IC). The model shows a consistent high bias in its forecasts of September ice extent. The corresponding predicted model variability at the three different lead times is shown in Fig. 11. The variability from the model prediction is underestimated near the mean September ice pack and overestimated outside the observed mean

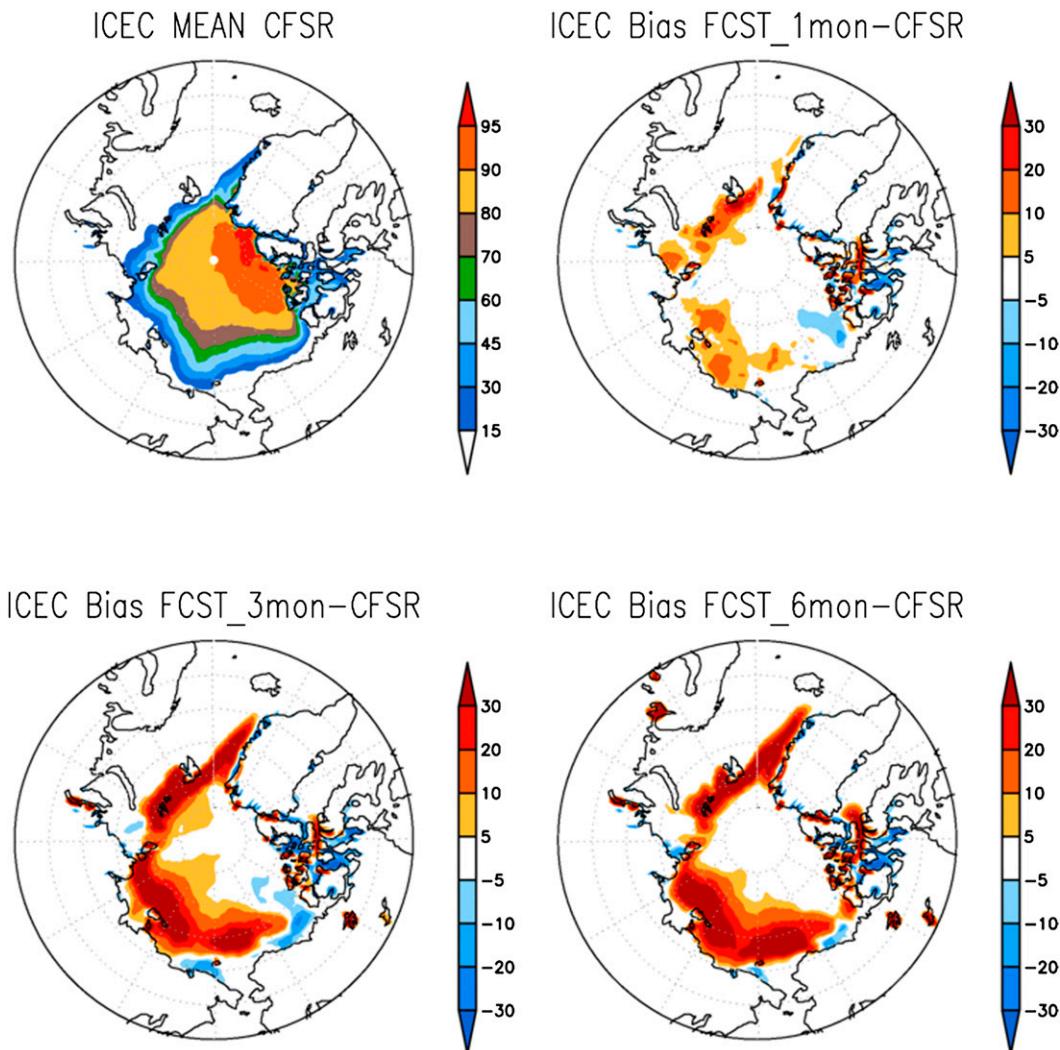


FIG. 10. (top left) The mean September sea ice concentration from 1982 to 2010 from CFSR, and the bias from the predicted mean condition for the September sea ice concentration with a lead time of (top right) 1 month (15 Aug IC), (bottom left) 3 months (15 Jun IC), and (bottom right) 6 months (15 Mar IC).

September ice pack. Although the CFSv2 captured the observed seasonal cycle, long-term trend, and interannual variability to some extent, large errors exist in its representation of the observed mean state and anomalies, as shown in Figs. 10 and 11. Therefore in the CFSv2, when the sea ice predictions are used for practical applications, bias correction is necessary. The bias can be obtained from the hindcast data for the period 1982–2010, which are available from NCDC.

In spite of the above reported shortcomings, when the model was used for the prediction of the September minimum sea ice extent organized by the Study of Environmental Arctic Change (SEARCH) during 2009 and 2011, CFSv2 (with bias correction applied) was among the best prediction models. In the future we plan

to assimilate the sea ice thickness data into the CFS assuming that would reduce the bias and improve the sea ice prediction.

6. Model behavior in very long integrations

a. Decadal prediction

The protocol for the 2013/14 Intergovernmental Panel on Climate Change (IPCC) model runs [i.e., the Fifth Assessment Report (AR5)] recommended creating decadal predictions to assist in the study of climate change (see <http://www.ipcc.ch/activities/activities.shtml#.UGyOHpH4Jw0>).

These decadal runs may bring in elements of the initial states in terms of land, ocean, sea ice, and atmosphere

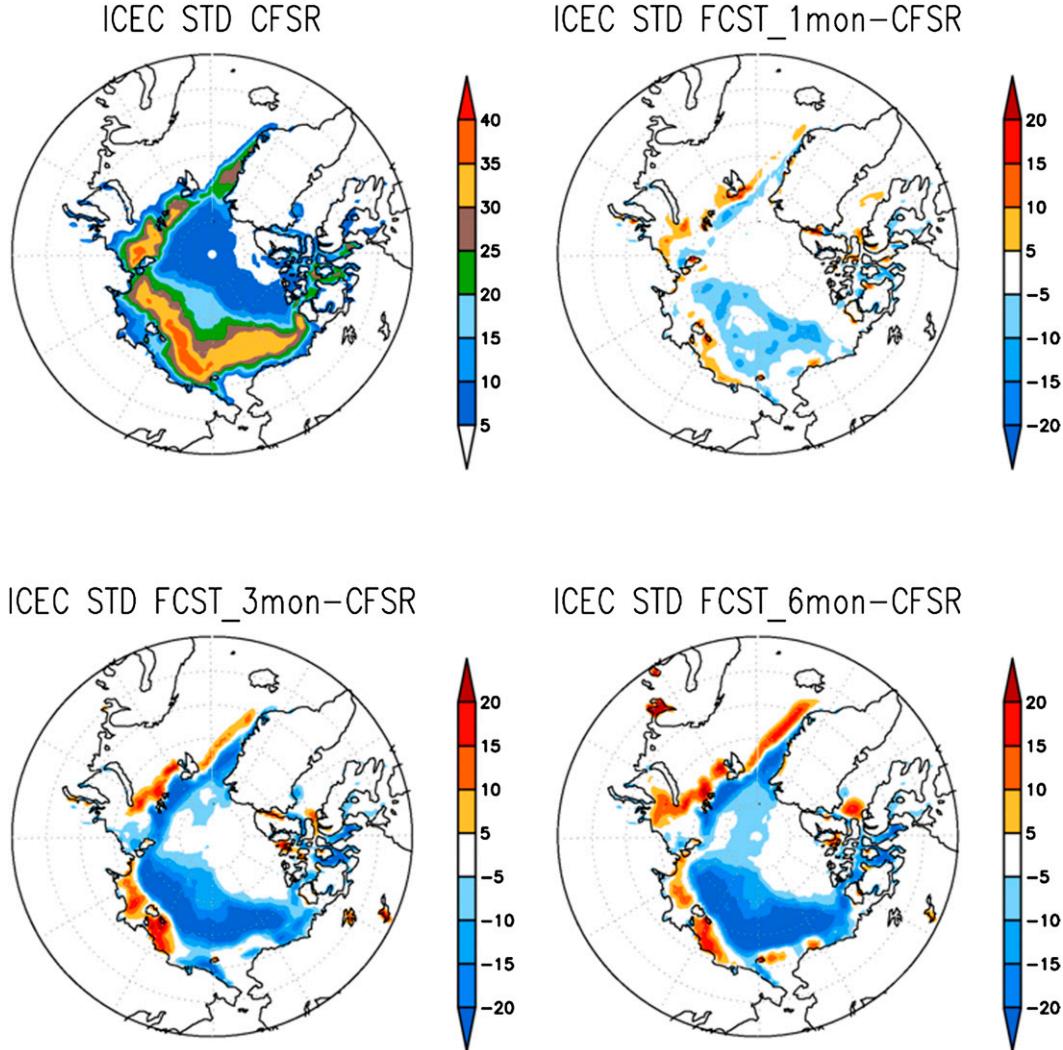


FIG. 11. (top left) The standard deviation of the September sea ice concentration from 1982 to 2010 from CFSR, and the difference of the standard deviation between the model prediction and that from the CFSR for the September sea ice concentration with a lead time of (top right) 1 month (15 Aug IC), (bottom left) 3 months (15 Jun IC), and (bottom right) 6 months (15 Mar IC).

and thus perhaps add information in the first 10 years, in addition to the general warming that most models may predict when greenhouse gases (GHG) increase. Following this recommendation, sixty 10-yr runs were completed from initial conditions on 1 November for 0000, 0600, 1200, and 1800 UTC cycles (i.e., 4 members), for the following years: 1980, 1981, 1983, 1985, 1990, 1993, 1995, 1996, 1998, 2000, 2003, 2005, 2006, 2009, and 2010 (every fifth year from 1980 to 2010 with some climatologically interesting intermediate years). Each run was 122 months long (the first 2 months were not used to avoid spinup). The forcing for these decadal runs included both shortwave and longwave tropospheric aerosol effects and is from a monthly climatology that

repeats its values year after year (described in Hou et al. 2002). Also, included in the runs are historical stratospheric volcanic aerosol effects on both shortwave and longwave radiation, which end in 1999, after which a minimum value of optical depth (10^{-4}) was used (Sato et al. 1993). The runs also used the latest observed CO₂ data when available [World Meteorological Organization (WMO) Global Atmospheric Watch; see also <http://ds.data.jma.go.jp/gmd/wdcgg/>] and an extrapolation was done into the future with a fixed growth rate of 2 ppmv. Solar constant variations were applied annually as described in Van den Dool (2011).

Results using only monthly mean data from the 60 decadal runs are presented in this paper. The variable X

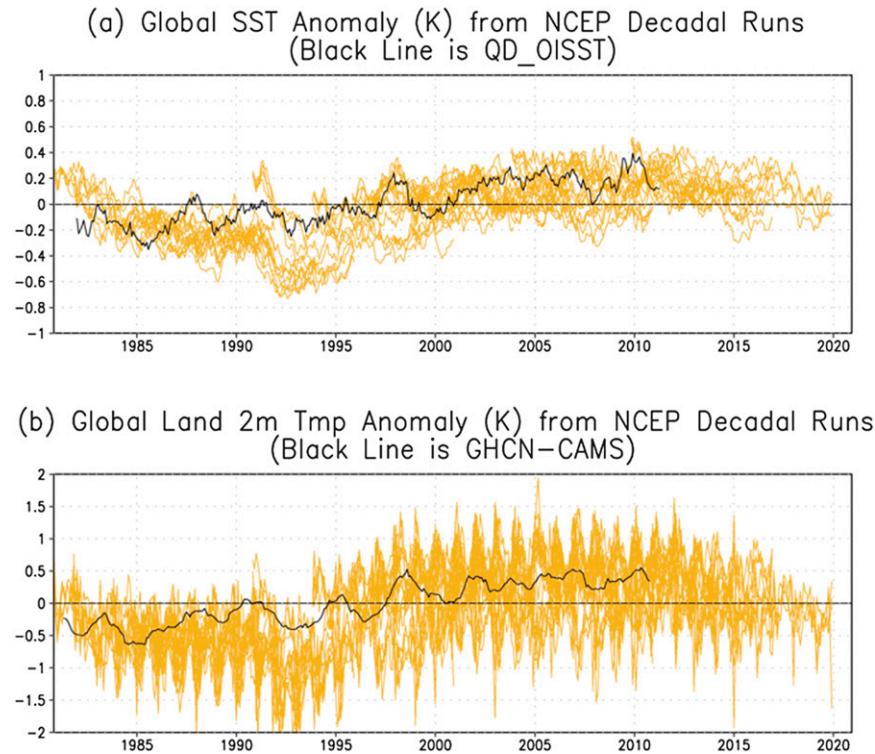


FIG. 12. (a) Globally averaged SST anomaly (K) in NCEP decadal integrations. Sixty-two 10-yr integrations were made and they are plotted as yellow traces. The observed single trace of 30+ yr is given in black. The definition of anomaly is given in the text. (b) As in (a), but for the globally averaged 2-m temperature anomaly (K) over land.

in an individual run can be denoted as $X_{j,m}$, where j and m are the target year and month. How “anomalies” are obtained is not obvious in these type of decadal runs. We proceeded as follows: first a 60-run mean was formed; that is, $\langle X_{j,m} \rangle$, where $j = 1, \dots, 10$ and $m = 1, \dots, 120$. Averaging across all years, we get $\langle \langle X_m \rangle \rangle$. The anomaly is then computed as $X_{j,m} - \langle \langle X_m \rangle \rangle$. Figure 12a shows the global mean SST anomalies (here X is SST). There are 60 yellow traces, each of 10-yr length. The observations (Reynolds et al. 2007) are shown as the solid black line, and the monthly anomaly is formed as the departure from the 1982–2010 climatology. One can conclude that the observations are in the cloud of model traces produced by CFSv2, especially after 1995 and before 1987 when the observations are near the middle of the cloud. The model appears somewhat cold in the late eighties and early nineties. Figure 12b shows the same thing, but for global mean land temperature. The black line, from GHCN-CAMS (Fan and Van den Dool 2008), which is a combination of the Global Historical Climate Network (GHCN) with the observation in the CPC’s Climate Anomaly Monitoring System (CAMS), is comfortably inside the cloud of model traces, except around 1993 when perhaps the model

overdid the aerosol impact of the Pinatubo volcanic eruption. The spread produced by the model is much higher in Fig. 12b than in Fig. 12a, not only because the land area is smaller than the oceanic area but also because the air temperature is much more variable to start with. This model, never before exposed to such long integrations, passed the zero-order test, in that it produced some warming over the period from 1980 to the present and has enough spread to cover what was observed (essentially a single model trace). In this paper there is no attempt to address any model prediction skill over and beyond a capability to show general warming and uncertainty.

Some monthly mean and 3-hourly time series data from the NCEP decadal runs are available for download (see appendix D).

b. Long “free” runs

For very long time scales, a few single runs were made lasting from 43 to 100 yr, which were designated as Coupled Model Intercomparison Project (CMIP) runs. There is nothing that reminds these runs of the calendar years they are in except for GHG levels, which are prescribed when available (see section 4c), and CO₂ is

projected to increase by 2 ppm in future years. Here, we are interested in behavioral aspects, including a test as to whether the system is stable or drifting due to assorted technical issues. The initial conditions were chosen for January of three years, 1987, 1995, and 2001 (similar runs were made with CFSv1). Allowing for a spinup of 1 year, data were saved for 1988–2030 (43 yr), 1996–2047 (52 yr), and 2002–2101 (100 yr) from these three runs, one of which is truly centennial. None of these runs became unstable or produced completely unreasonable results. A common undesirable feature (i.e., not a real forecast!) was a slow cooling of the upper ocean for the first 15–20 yr. Only after this temperature decline stabilized, a global warming of the sea surface temperature was seen, starting 25–35 yr after initial time. In contrast, the water at the bottom of the ocean showed a small warming from the beginning to end, which is unlikely to be correct.

An important issue was to examine the onset and decay of warm and cold events (e.g., El Niño and La Niña) and ascertain how regular they were. The CFSv1 was found to be too regular and very close to being periodic in its CMIP runs (Penland and Saha 2006) when diagnosed via a spectral analysis of Niño-3.4 monthly values. Figure 13 shows the spectra of Niño-3.4 for the observations from 1950 to 2011 (upper left) and the three CFSv2 CMIP runs. A harmonic analysis was conducted on monthly mean data with a monthly climatology removed. Raw power was estimated as half of the amplitude (of the harmonic) squared. The curves shown were smoothed by a 1–2–1 filter. The variance of all the CMIP runs is higher than observed by at least 25%, therefore the integral under the modeled and observed curves differs. The model variance being too large was already noted in Fig. 3 for leads of 3 and 6 months, and in Tables 1 and 2 for many other fields and areas. The observations have a broad spectral maximum from 0.15 to 0.45 cpy. The shortest of the CMIP runs (upper right) resembles the broad spectral maximum quite well; the longer runs are somewhat more sharply peaked but are not nearly as periodic as in CMIP runs made by CFSv1, especially when T62 resolution was used (Penland and Saha 2006). On the whole, the behavioral aspects of ENSO (well beyond prediction) appear acceptable. One may also consider the possibility that certain segments of 43 yr from the 100-yr run may look like the upper right entry. Or by the same token, that the behavior of observations for 1951–2011 is not necessarily reproduced exactly when a longer period could be considered, or a period without mega-events like the 1982/83 and 1997/98 ENSO events. Some data from these CMIP runs are available for download from the CFS website (see appendix D).

7. Concluding remarks

This paper describes the transition from the CFSv1 to the CFSv2 operational systems. The Climate Forecast System (CFS), retroactively named version 1, was operationally implemented at NCEP in August 2004. The CFSv1 was described in S06. Its successor, named CFSv2, was implemented in March 2011 even though version 1 was only decommissioned in October 2012. The overlap (1.5 yr) was needed, among other things, to give users time to make their transition between the two systems. In contrast to most implementations at NCEP, the CFS is accompanied by a set of retrospective forecasts that can be applied by the user community to calibrate subsequent real-time operational forecasts made by the same system. Therefore, a new CFS takes time to develop and implement both on the part of NCEP and on the side of the user. One element that took a lot of time at NCEP to complete was a new reanalysis (CFSR), which was needed to create the initial conditions for the coupled land-atmosphere-ocean-sea ice CFSv2 retrospective forecasts. Every effort was made to create these initial conditions (for the period 1979–present) with a forecast system that was as consistent as possible with the model used to make the long-range forecasts, whether it be for the retrospective forecasts or the operational forecasts going forward in real time.

For convenience, the evolution of the model components between CFSv1 and CFSv2 has been split into two portions, namely the very large model developments between CFSv1 and CFSR and the far smaller model developments between CFSR and CFSv2. The development of model components between the time of CFSv1 (of 1996–2003 vintage) and CFSR (of 2008–10 vintage) to generate the background guess in the data assimilation has already been documented in Saha et al. (2010). Therefore, in the present paper, we only describe some further adjustments/tunings of the land surface parameters and clouds in the equatorial SST (in section 2).

The paper describes the design of both the long-lead seasonal (out to 9 months) and shorter-lead intraseasonal predictions (out to 45 days) for the retrospective forecasts and the real-time operational predictions going forward. This information is essential for any user who may want to use these forecasts. The retrospective forecasts are important for both calibration and skill estimates of subsequent real-time prediction. The size of the hindcast dataset is very large, since it spans forecasts from 1982 to the present for long-lead seasonal range (4 runs out to 9 months, every 5th day), and forecasts from 1999 to the present for intraseasonal range (3 runs each day out to 45 days, plus one run each day out to 90 days), with all model forecast output data archived at 6-h intervals for each run.

POWER SPECTRUM FOR SST OVER NINO 3.4

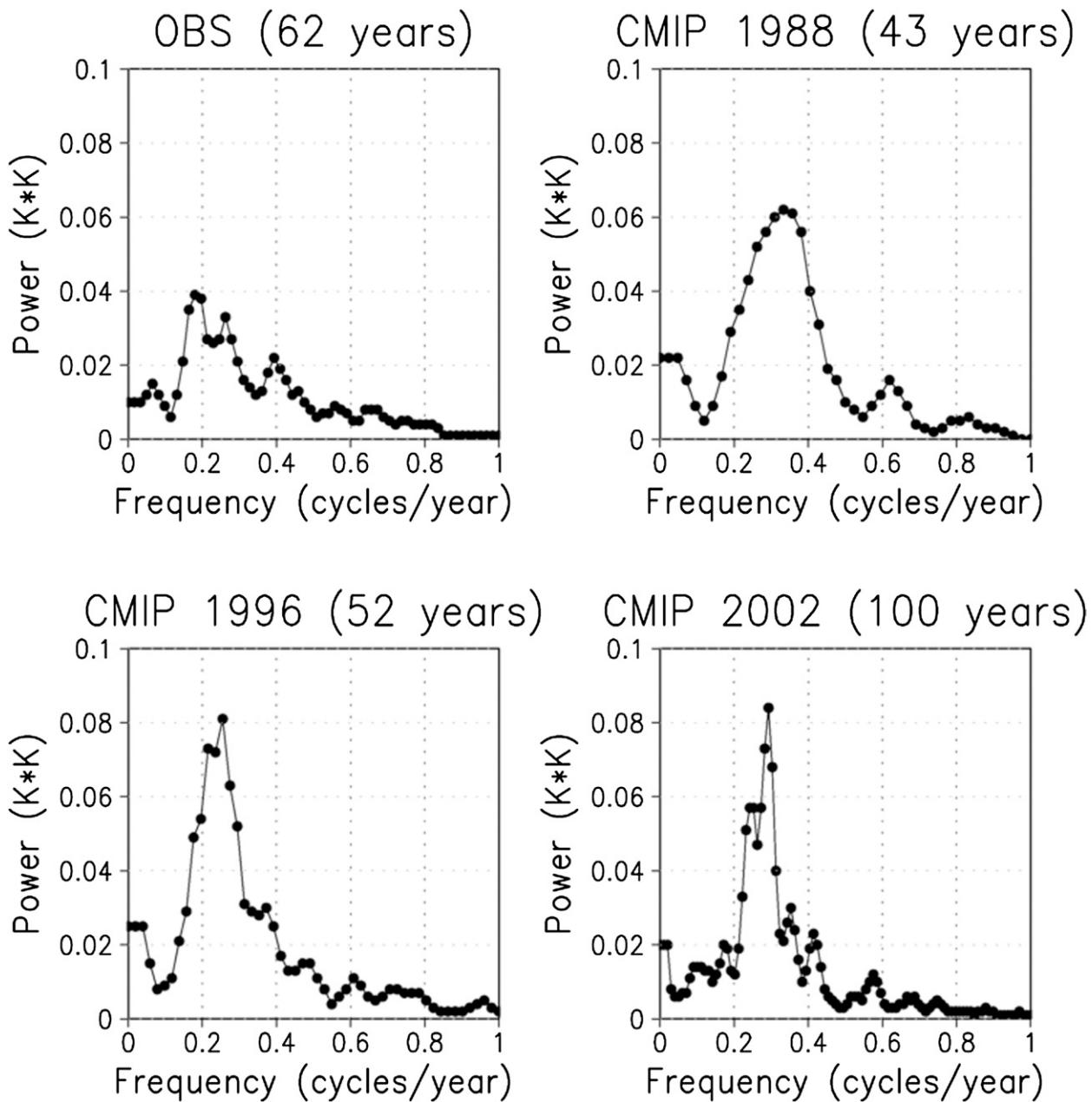


FIG. 13. Power spectra of time series of monthly anomalies of the Niño-3.4 index (average SST for 5°S – 5°N , 170° – 120°W) for (top left) the observation and CMIP runs of (top right) 43, (bottom left) 52, and (bottom right) 100 yr.

The paper also describes some of the results, in terms of the forecast skill, determined from the retrospective forecasts, for the prediction of the intraseasonal component (MJO in particular) and the seasonal prediction component (in section 4). This is done by comparing, very precisely, the CFSv2 predictions to exactly matching

CFSv1 predictions. There is no doubt that CFSv2 is superior to CFSv1 on the intraseasonal time scale; in fact, the improvement is impressive from 1 week to more than 2 weeks (at the 0.5 level of anomaly correlation) for MJO prediction. For seasonal prediction, we note a substantial improvement in 2-m temperature prediction

over global land. This is mainly a result of successfully simulating temperature trends (which are large over the 1980–2010 period and thus an integral part of any verification) by increasing the amount of prescribed greenhouse gases in the model (a feature that was missing in CFSv1). For precipitation over land, the CFSv2, unfortunately, is hardly an improvement over CFSv1. This is perhaps due to the predictability ceiling being too low to expect big leaps forward in prediction. The SST prediction has been improved modestly over most of the global oceans and extended in CFSv2 to areas where CFSv1 had prescribed SST and/or sea ice, as well as over the extratropical oceans. In the tropics, SST prediction has also improved, but least so in the much-focused-on Niño-3.4 area, where the subsurface initial states of CFSR show warming after 1998, resulting from the introduction of the AMSU satellite data. Before that time, the SST forecasts were too cold in that area, thus making the systematic error correction a challenge.

Being a community model to some extent, the CFSv2 has been (and will be) applied to decadal and centennial runs. These have not been typical NCEP endeavors in the past, so we have tested the behavior of this new model in integrations beyond the operational 9-month runs. Some results are described in section 6. The decadal runs appear reasonable in that, in the global mean, reality is within the cloud of the 60 decadal runs, both for 2-m temperature over land and for SST in the ocean. The three centennial runs did not derail (a minimal test passed) and show both reasonable and unreasonable behavior. Unreasonable, we believe, is a small but steady cooling of the global ocean surface that lasts about 15 years before GHG forced warming sets in. Equally unreasonable may be a small warming of the bottom layers of global oceans from start to finish. The better news is that the ENSO spectrum in these free runs is far more acceptable in CFSv2, in contrast to CFSv1. When run in its standard resolution of T62L64, the CFSv1 produced too regular and almost periodic ENSO in its free runs, lasting up to a century.

A few diagnostics (presented in section 5) were made in support of the need for tuning some of the land surface parameters when going from CFSR to CFSv2. The main concern was the fact that the NH mean precipitation in summer over land reduced from 3.2 mm day^{-1} in CFSR to 2.7 mm day^{-1} in CFSv2, which posed a real problem for improved prediction of evaporation, runoff, and surface air temperature. Some diagnostics are also presented for the emerging area of coupled sea ice modeling, imbedded in a global ocean. Although this topic is important for monthly and seasonal prediction, it has taken on new urgency because of concerns over

shrinking sea ice coverage (and thickness) in the Arctic. It is easy to identify some large errors in sea ice coverage and variability and it is obvious that a lot more work needs to be done in this area of sea ice modeling.

This paper is mainly to describe CFSv2 as a whole, from inception to implementation. There are many subsequent papers in preparation (or submitted/published) about detailed studies of CFSv2 prediction skill and/or diagnostics of some of the parts of CFSv2, whether it be the stratosphere, troposphere, deep oceans, land surface, or other aspects.

While there are many users for the CFS output (sometimes one finds out how many only by trying to discontinue a model), the first-line user is the Climate Prediction Center at NCEP. The CFSv2 plays a substantial role in the seasonal prediction efforts at the CPC, both directly and through joint efforts such as national and international multimodel ensembles.³ CFSv2 is also used in the subseasonal MJO prediction, and in a product called international hazards assessment. Because CFSv2 runs practically in real time (compared to CFSv1, which was about 36 h later than real time), it plays a role in the operational 6–10-day and week 2 forecasts and conceivably in the future prediction of the week 3–6 forecasts for the United States, which is on the drawing board at CPC. The appropriate forcing fields extracted from CFSv2 predictions, such as daily radiation, precipitation, wind, relative humidity, etc., are used to carry the Global Land Data Assimilation Systems (GLDAS) forward, yielding an ensemble of drought-related indices over the United States and soon globally.

Acknowledgments. The authors would like to recognize all the scientists and technical staff of the Global Climate and Weather Modeling Branch of EMC for their hard work and dedication to the development of the GFS. We would also like to extend our thanks to the scientists at GFDL for their work in developing the MOM4. George Vandenberge, Carolyn Pasti, and Julia Zhu are recognized for their critical support in the smooth running of the CFSv2 retrospective forecasts and the operational implementation of the CFSv2. We also thank Ben Kyger, Dan Starosta, Christine Magee, and Becky Cosgrove from the NCEP Central Operations (NCO) for the timely operational implementation of the CFSv2 in March 2011.

³ We should point out that what we call the International Multi-model Ensemble (IMME) has its counterpart called Eurosip in Europe. CFSv2 has been included as a member in the Eurosip ensemble, which consists of the European Centre for Medium-Range Weather Forecasts (ECMWF), Met Office (UKMO), and Météo France.

APPENDIX A

Reforecast Configuration of the CFSv2

- The 9-month hindcasts were initiated from every 5th day and run from all four cycles of that day, beginning from 1 January of each year, over the full 29-yr period (1982–2010). This is required to calibrate the operational CPC longer-term seasonal predictions (ENSO, etc.) (solid lines in Fig. A1).
- There was also a single 1-season (123 day) hindcast run, initiated from every 0000 UTC cycle between these five days, but only over the 12-yr period from 1999 to 2010. This is required to calibrate the operational CPC first season predictions for hydrological forecasts (precipitation, evaporation, runoff, streamflow, etc.) (dashed lines in Fig. A1).
- In addition, there were three 45-day hindcast runs from every 0600, 1200, and 1800 UTC cycle over the 12-yr period 1999–2010. This is required for the operational CPC week 3–6 predictions of tropical circulations [MJO, Pacific–North American (PNA) pattern, etc.] (dotted lines in Fig. A1).
- Total number of years of integration = 9447 years.

APPENDIX B

Retrospective Forecast Calendar (292 Runs per Year)

Organized by date of release of the official CPC seasonal prediction every month

As outlined in appendix A, four 9-month retrospective forecasts are made every 5th day over the period 1982–2010. The calendar always starts on 1 January and proceeds forward in the same manner each year. Forecasts are always made from the same initial dates every year. This means that in leap years 25 February and 2 March are separated by 6 days (instead of 5). Table B1 describes the grouping of the retrospective forecasts in relation to the CPC's operational schedule (all forecast products must be available a week before the earliest official release on the third Thursday of each month). For instance, for the release of the official forecast in the month of February, all retrospective forecasts made from initial conditions over the period from 11 January through 5 February for all previous years can be used for calibration and skill estimates, constituting a lagged ensemble of 24 members. Obviously one can use more

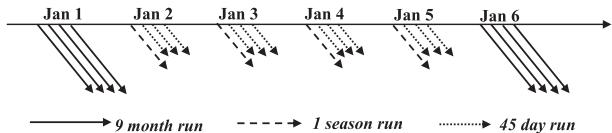


FIG. A1. Reforecast configuration of the CFSv2.

(going back farther to give a larger ensemble) or less (since older forecasts may have less skill).

All real-time forecasts that are available closest to the date of release are used (see appendix C).

APPENDIX C

Operational Configuration of the CFSv2 for a 24-h Period

- There are four control runs per day from the 0000, 0600, 1200, and 1800 UTC cycles of the CFSv2 real-time data assimilation system, out to 9 months (full lines in Fig. C1).
- In addition to the control run of 9 months, there are three additional perturbed runs at 0000 UTC out to one season (dashed lines in Fig. C1).
- In addition to the control run of 9 months at the 0600, 1200, and 1800 UTC cycles, there are three additional perturbed runs, out to 45 days (dotted lines in Fig. C1).
- There are a total of 16 CFS runs every day, of which four runs go out to 9 months, three runs go out to 1 season, and nine runs go out to 45 days.

APPENDIX D

Availability of CFSv2 Data

- The official website for the CFSv2 is <http://cfs.ncep.noaa.gov>. Useful documentation and some model data can be downloaded from this site.
- Real-time operational data: Users must maintain their own continuing archive by downloading the real-time operational data from the 7-day rotating archive located online at <http://nomads.ncep.noaa.gov/pub/data/nccf/com/cfs/prod/>. This website includes both the initial conditions and forecasts made at each cycle of each day. Monthly means of the initial conditions are posted once a month and can be downloaded from a 6-month rotating archive at the same location given above.

TABLE B1. CFSv2 retrospective calendar (organized by date of earliest possible release of the official CPC seasonal prediction every month using 24 members unless specified).

Mid-January release	12 Dec at 0000, 0600, 1200, and 1800 UTC 17 Dec at 0000, 0600, 1200, and 1800 UTC 22 Dec at 0000, 0600, 1200, and 1800 UTC 27 Dec at 0000, 0600, 1200, and 1800 UTC 1 Jan at 0000, 0600, 1200, and 1800 UTC 6 Jan at 0000, 0600, 1200, and 1800 UTC	Mid-July release	10 Jun at 0000, 0600, 1200, and 1800 UTC 15 Jun at 0000, 0600, 1200, and 1800 UTC 20 Jun at 0000, 0600, 1200, and 1800 UTC 25 Jun at 0000, 0600, 1200, and 1800 UTC 30 Jun at 0000, 0600, 1200, and 1800 UTC 5 Jul at 0000, 0600, 1200, and 1800 UTC
Mid-February release	11 Jan at 0000, 0600, 1200, and 1800 UTC 16 Jan at 0000, 0600, 1200, and 1800 UTC 21 Jan at 0000, 0600, 1200, and 1800 UTC 26 Jan at 0000, 0600, 1200, and 1800 UTC 31 Jan at 0000, 0600, 1200, and 1800 UTC 5 Feb at 0000, 0600, 1200, and 1800 UTC	Mid-August release	10 Jul at 0000, 0600, 1200, and 1800 UTC 15 Jul at 0000, 0600, 1200, and 1800 UTC 20 Jul at 0000, 0600, 1200, and 1800 UTC 25 Jul at 0000, 0600, 1200, and 1800 UTC 30 Jul at 0000, 0600, 1200, and 1800 UTC 4 Aug at 0000, 0600, 1200, and 1800 UTC
Mid-March release	10 Feb at 0000, 0600, 1200, and 1800 UTC 15 Feb at 0000, 0600, 1200, and 1800 UTC 20 Feb at 0000, 0600, 1200, and 1800 UTC 25 Feb at 0000, 0600, 1200, and 1800 UTC 2 Mar at 0000, 0600, 1200, and 1800 UTC 7 Mar at 0000, 0600, 1200, and 1800 UTC	Mid-September release	9 Aug at 0000, 0600, 1200, and 1800 UTC 14 Aug at 0000, 0600, 1200, and 1800 UTC 19 Aug at 0000, 0600, 1200, and 1800 UTC 24 Aug at 0000, 0600, 1200, and 1800 UTC 29 Aug at 0000, 0600, 1200, and 1800 UTC 3 Sep at 0000, 0600, 1200, and 1800 UTC
Mid-April release	12 Mar at 0000, 0600, 1200, and 1800 UTC 17 Mar at 0000, 0600, 1200, and 1800 UTC 22 Mar at 0000, 0600, 1200, and 1800 UTC 27 Mar at 0000, 0600, 1200, and 1800 UTC 1 Apr at 0000, 0600, 1200, and 1800 UTC 6 Apr at 0000, 0600, 1200, and 1800 UTC	Mid-October release	8 Sep at 0000, 0600, 1200, and 1800 UTC 13 Sep at 0000, 0600, 1200, and 1800 UTC 18 Sep at 0000, 0600, 1200, and 1800 UTC 23 Sep at 0000, 0600, 1200, and 1800 UTC 28 Sep at 0000, 0600, 1200, and 1800 UTC 3 Oct at 0000, 0600, 1200, and 1800 UTC
Mid-May release	11 Apr at 0000, 0600, 1200, and 1800 UTC 16 Apr at 0000, 0600, 1200, and 1800 UTC 21 Apr at 0000, 0600, 1200, and 1800 UTC 26 Apr at 0000, 0600, 1200, and 1800 UTC 1 May at 0000, 0600, 1200, and 1800 UTC 6 May at 0000, 0600, 1200, and 1800 UTC	Mid-November release (28 members)	8 Oct at 0000, 0600, 1200, and 1800 UTC 13 Oct at 0000, 0600, 1200, and 1800 UTC 18 Oct at 0000, 0600, 1200, and 1800 UTC 23 Oct at 0000, 0600, 1200, and 1800 UTC 28 Oct at 0000, 0600, 1200, and 1800 UTC 2 Nov at 0000, 0600, 1200, and 1800 UTC 7 Nov at 0000, 0600, 1200, and 1800 UTC
Mid-June release	11 May at 0000, 0600, 1200, and 1800 UTC 16 May at 0000, 0600, 1200, and 1800 UTC 21 May at 0000, 0600, 1200, and 1800 UTC 26 May at 0000, 0600, 1200, and 1800 UTC 31 May at 0000, 0600, 1200, and 1800 UTC 5 Jun at 0000, 0600, 1200, and 1800 UTC	Mid-December release	12 Nov at 0000, 0600, 1200, and 1800 UTC 17 Nov at 0000, 0600, 1200, and 1800 UTC 22 Nov at 0000, 0600, 1200, and 1800 UTC 27 Nov at 0000, 0600, 1200, and 1800 UTC 2 Dec at 0000, 0600, 1200, and 1800 UTC 7 Dec at 0000, 0600, 1200, and 1800 UTC

- Selected data from the CFSv2 retrospective forecasts (both seasonal and subseasonal) for the forecast period 1982–2010 may be downloaded from the NCDC web servers online at <http://nomads.ncdc.noaa.gov/data.php?name=access#cfs>.
- Smoothed calibration climatologies have been prepared from the forecast monthly means and time series of selected variables and are available for download from the CFS website (<http://cfs.ncep.noaa.gov/cfsv2.info/CFSv2.Calibration.Data.doc>). Please note that two sets of climatologies have been prepared for calibration, for the full period (1982–2010) and the later period (1999–2010). We highly recommend that the climatology

prepared from the later period be used when calibrating real-time operational predictions for variables in the tropics, such as *SST* and *precipitation over oceans*. For skill estimates, we recommend that split climatologies be used for the two periods when removing the forecast bias.

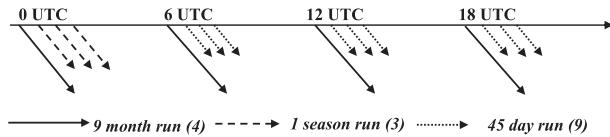


FIG. C1. Operational configuration of the CFSv2.

- A small amount of CFSv2 forecast data from 2011 to the present may be found at the CFS website at <http://cfs.ncep.noaa.gov/cfsv2/downloads.html>.
- Decadal runs: Some monthly mean and 3-hourly time series data from the NCEP decadal runs may be obtained from the Earth System Grid Federation (ESGF)/Program for Climate Model Diagnosis and Intercomparison (PCMDI) website at <http://esgf.nccs.nasa.gov/esgf-web-fe/>.
- CMIP runs: Monthly mean data from the three CMIP runs is available for download from the CFS website at <http://cfs.ncep.noaa.gov/pub/raid0/cfsv2/cmipruns>.

REFERENCES

- Barker, H. W., R. Pincus, and J.-J. Morcrette, 2002: The Monte Carlo independent column approximation: Application within large-scale models. *Extended Abstracts, GCSS-ARM Workshop on the Representation of Cloud Systems in Large-Scale Models*, Kananaskis, AB, Canada, GEWEX, 1–10.
- Behringer, D. W., 2007: The Global Ocean Data Assimilation System at NCEP. *Preprints, 11th Symp. on Integrated Observing and Assimilation Systems for Atmosphere, Oceans and Land Surface*, San Antonio, TX, Amer. Meteor. Soc., 14–18.
- Cai, M., C.-S. Shin, H. M. van den Dool, W. Wang, S. Saha, and A. Kumar, 2009: The role of long-term trends in seasonal predictions: Implication of global warming in the NCEP CFS. *Wea. Forecasting*, **24**, 965–973.
- Chun, H.-Y., and J.-J. Baik, 1998: Momentum flux by thermally induced internal gravity wave and its approximation for large-scale models. *J. Atmos. Sci.*, **55**, 3299–3310.
- Clough, S. A., M. W. Shephard, E. J. Mlawer, J. S. Delamere, M. J. Iacono, K. Cady-Pereira, S. Boukabara, and P. D. Brown, 2005: Atmospheric radiative transfer modeling: A summary of the AER codes. *J. Quant. Spectrosc. Radiat. Transfer*, **91**, 233–244.
- Ek, M., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley, 2003: Implementation of Noah land-surface model advances in the NCEP operational mesoscale Eta model. *J. Geophys. Res.*, **108**, 8851, doi:10.1029/2002JD003296.
- Fan, Y., and H. van den Dool, 2008: A global monthly land surface air temperature analysis for 1948–present. *J. Geophys. Res.*, **113**, D01103, doi:10.1029/2007JD008470.
- Hou, Y., S. Moorthi, and K. Campana, 2002: Parameterization of Solar Radiation Transfer in the NCEP Models. NCEP Office Note 441, 46 pp. [Available online at <http://www.emc.ncep.noaa.gov/officenotes/newernotes/on441.pdf>.]
- Iacono, M. J., E. J. Mlawer, S. A. Clough, and J.-J. Morcrette, 2000: Impact of an improved longwave radiation model, RRTM, on the energy budget and thermodynamic properties of the NCAR Community Climate Model, CCM3. *J. Geophys. Res.*, **105**, 14 873–14 890.
- Kanamitsu, M., W. Ebisuzaki, J. Woollen, 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.
- Kirtman, B. P., and Coauthors, 2014: The North American Multi-Model Ensemble (NMME): Phase-1, seasonal-to-interannual prediction; phase-2, toward developing intraseasonal prediction. *Bull. Amer. Meteor. Soc.*, in press.
- Kumar, A., M. Chen, L. Zhang, W. Wang, Y. Xue, C. Wen, L. Marx, and B. Huang, 2012: An analysis of the non-stationarity in the bias of sea surface temperature forecasts for the NCEP Climate Forecast System (CFS) version 2. *Mon. Wea. Rev.*, **140**, 3003–3016.
- Lin, H., G. Brunet, and J. Derome, 2008: Forecast skill of the Madden–Julian oscillation in two Canadian atmospheric models. *Mon. Wea. Rev.*, **136**, 4130–4149.
- Mitchell, K. E., H. Wei, S. Lu, G. Gayno and J. Meng, 2005: NCEP implements major upgrade to its medium-range global forecast system, including land-surface component. *GEWEX Newsletter*, No. 15 (4), International GEWEX Project Office, Silver Spring, MD, 8–9.
- Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmosphere: RRTM, a validated correlated- k model for the long-wave. *J. Geophys. Res.*, **102** (D14), 16 663–16 683.
- Moorthi, S., R. Sun, H. Xia, and C. R. Mechoso, 2010: Southeast Pacific low-cloud simulation in the NCEP GFS: Role of vertical mixing and shallow convection. NCEP Office Note 463, 28 pp. [Available online at <http://www.emc.ncep.noaa.gov/officenotes/FullTOC.html#2000>.]
- Penland, C., and S. Saha, 2006: El Niño in the Climate Forecast System: T62 vs T126. *Proc. 30th Climate Diagnostics and Prediction Workshop*, State College, PA, NOAA, P1.3. [Available online at http://www.cpc.ncep.noaa.gov/products/outreach/proceedings/cdw30_proceedings/P1.3.pdf.]
- Pincus, R., H. W. Barker, and J.-J. Morcrette, 2003: A fast, flexible, approximate technique for computing radiative transfer in inhomogeneous cloud fields. *J. Geophys. Res.*, **108**, 4376, doi:10.1029/2002JD003322.
- Reynolds, R. W., N. A. Raynor, 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.
- , T. M. Smith, C. Liu, D. B. Chelton, K. S. Casey, and M. G. Schlax, 2007: Daily high-resolution blended analyses for sea surface temperature. *J. Climate*, **20**, 5473–5496.
- 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.
- Sato, M., J. E. Hansen, M. P. McCormick, and J. B. Pollack, 1993: Stratospheric aerosol optical depths, 1850–1990. *J. Geophys. Res.*, **98**, 22 987–22 994.
- Sun, R., S. Moorthi, and C. R. Mechoso, 2010: Simulation of low clouds in the southeast Pacific by the NCEP GFS: Sensitivity to vertical mixing. *Atmos. Chem. Phys.*, **10**, 12 261–12 272.
- Van den Dool, H. M., cited 2011: Reconstruction of the solar constant back to 1750. [Available online at http://www.cpc.ncep.noaa.gov/products/people/wd51hd/vddoolpubs/solar_reconstruction.doc.]
- Vitart, F., S. Woolnough, M. A. Balmaseda, and A. M. Tompkins, 2007: Monthly forecast of the Madden–Julian oscillation using a coupled GCM. *Mon. Wea. Rev.*, **135**, 2700–2715.
- Wheeler, M., and H. H. Hendon, 2004: An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Mon. Wea. Rev.*, **132**, 1917–1932.
- Zhang, Q., and H. van den Dool, 2012: Relative merit of model improvement versus availability of retrospective forecasts: The case of Climate Forecast System MJO prediction. *Wea. Forecasting*, **27**, 1045–1051.