

CFSv2-Based Seasonal Hydroclimatic Forecasts over the Conterminous United States

XING YUAN, ERIC F. WOOD, JOSHUA K. ROUNDY, AND MING PAN

Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey

(Manuscript received 20 September 2012, in final form 2 January 2013)

ABSTRACT

There is a long history of debate on the usefulness of climate model–based seasonal hydroclimatic forecasts as compared to ensemble streamflow prediction (ESP). In this study, the authors use NCEP’s operational forecast system, the Climate Forecast System version 2 (CFSv2), and its previous version, CFSv1, to investigate the value of climate models by conducting a set of 27-yr seasonal hydroclimatic hindcasts over the conterminous United States (CONUS). Through Bayesian downscaling, climate models have higher squared correlation R^2 and smaller error than ESP for monthly precipitation, and the forecasts conditional on ENSO have further improvements over southern basins out to 4 months. Verification of streamflow forecasts over 1734 U.S. Geological Survey (USGS) gauges shows that CFSv2 has moderately smaller error than ESP, but all three approaches have limited added skill against climatology beyond 1 month because of overforecasting or underdispersion errors. Using a postprocessor, 60%–70% of probabilistic streamflow forecasts are more skillful than climatology. All three approaches have plausible predictions of soil moisture drought frequency over the central United States out to 6 months, and climate models provide better results over the central and eastern United States. The R^2 of drought extent is higher for arid basins and for the forecasts initiated during dry seasons, but significant improvements from CFSv2 occur in different seasons for different basins. The R^2 of drought severity accumulated over CONUS is higher during winter, and climate models present added value, especially at long leads. This study indicates that climate models can provide better seasonal hydroclimatic forecasts than ESP through appropriate downscaling procedures, but significant improvements are dependent on the variables, seasons, and regions.

1. Introduction

Although there is a theoretical threshold of about 2 weeks in weather prediction due to the chaotic nature of the climate system, forecasting seasonal climate is potentially possible because of additional memory imparted from the slowly evolved components such as ocean and land through land–atmosphere–ocean interaction. In fact, seasonal forecasting became popular in the 1970s, when people started to notice the relationship between the Southern Oscillation (SO) and El Niño (Troccoli 2010) and their corresponding changes in the strength of trade winds, position of convection, and teleconnection patterns that cause rainfall variations in many parts of the world. With gradual improvements in observations and assimilation systems, physical parameterizations, spatial resolutions, and the understandings

of ENSO-related ocean–atmosphere interaction, the coupled atmosphere–ocean–land general circulation models (CGCMs) now have higher ENSO prediction skill than statistical models (Barnston et al. 2012). The recent investments in decadal prediction will further benefit seasonal forecast thanks to similar development strategies (Goddard et al. 2012).

The progress in dynamical seasonal forecasts provides a potential opportunity to predict hydroclimatic variables (e.g., precipitation, streamflow, soil moisture) at long lead times, which is important for agriculture and water resources management, drought detection, and mitigation. In this context, climate model–based seasonal hydroclimatic forecasting plays an important role in transitioning the scientific advances from the hydroclimatic research community to the end users of society. On the other hand, the increasing service needs from the society also bring new challenges to the scientific community and will stimulate and accelerate the development of climate prediction as well as seasonal hydroclimatic forecast.

One of the first attempts to produce seasonal hydroclimatic forecast by using CGCM output was made

Corresponding author address: Xing Yuan, Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ 08544.
E-mail: xingy@princeton.edu

by Wood et al. (2002). They implemented an equal-quantile mapping method to bias correct and downscale seasonal forecasts of monthly precipitation and 2-m air temperature from National Centers for Environmental Prediction (NCEP) Global Spectral Model (GSM) and used the downscaled monthly forcings to scale randomly selected daily series. Then the scaled series were used as input to drive a hydrologic model to provide forecasts over the eastern United States in the summer of 2000 and winter of 1997/98. Later, Wood et al. (2005) assessed the hydroclimatic forecast skill over western United States during 1979–99. They found that CGCM-based seasonal hydroclimatic forecast could propagate climate forecast signals into hydrologic variables in a qualitative perspective but had negligible skill improvement over the traditional ensemble streamflow prediction (ESP; Twedt et al. 1977) method. ESP is based on resampling of the historical climate forcings that are used as inputs to a hydrologic model with nowcast initial conditions, which indicates that seasonal streamflow forecast is essentially reduced to an initial value problem in their western U.S. cases. Luo and Wood (2008) utilized a Bayesian method (Luo et al. 2007) to merge climate forecasts of multiple CGCMs from NCEP and European Union Development of a European Multimodel Ensemble System for Seasonal-to-Interannual Prediction (DEMETER) project and conducted a 19-yr seasonal hydroclimatic forecast over the Ohio River basin in the eastern United States. The multimodel forecast was more skillful than the ESP forecast during the first two months. However, the advantage was marginal to moderate if only using one CGCM, such as NCEP's Climate Forecast System version 1 (CFSv1; the successor to GSM). Therefore, even though some pilot studies have been carried out for CGCM-based seasonal hydroclimatic prediction, the relatively low forecast skill is still an unresolved question.

Similar to the short-term river forecast (Schaake et al. 2007), skillful CGCM-based seasonal hydroclimatic forecast relies on accurately downscaled CGCM precipitation and temperature predictions, refined initial hydrologic conditions that contain additional memory (e.g., soil moisture, snow), and advanced land surface hydrologic models that reasonably represent the terrestrial water and energy budgets. After decades of development, land surface hydrologic models have been successfully implemented at local, regional, and global scales. Moreover, the uncertainty from hydrologic models can be reduced substantially through parameter calibration and regionalization procedures (Duan et al. 2006). Besides the development of hydrologic models, the role of initial conditions on the seasonal hydrologic forecast has been receiving more attention recently (Li

et al. 2009; Koster et al. 2010; Shukla and Lettenmaier 2011). One of the routine ways to generate initial conditions for operational hydrologic nowcast and forecast is to run the hydrologic model with antecedent climatic observations (Wood et al. 2002; Luo and Wood 2008; Xia et al. 2012), though more complicated methods such as satellite data assimilation might be possible (Crow and Wood 2003; Pan et al. 2008).

Unlike the uncertainties from initial conditions and hydrologic models that could be constrained in a relatively acceptable extent, the uncertainties from climate forecast models seem to be the major sources for the low skill in seasonal hydroclimatic forecasting. The large uncertainty from CGCM precipitation prediction raises the question whether seasonal hydrologic forecast should be treated as an initial value problem in the hydrologic community. In other words, is there any added value from a CGCM-based seasonal hydrologic forecast against the traditional ESP forecast? The answer may differ by region and season. Therefore, with the gradual improvement of CGCMs, it is necessary to conduct comprehensive seasonal hydroclimatic reforecasts and to investigate whether or how much of the improvement from climate forecast models result in improved skill in the seasonal hydroclimatic forecasts and whether the state-of-the-art CGCM-based streamflow and drought forecasts outperform those from ESP forecasts.

Recently, NCEP has upgraded their operational seasonal forecast system with a new CGCM, the second version of CFS (CFSv2), where a number of new physical packages and a new atmosphere–ocean–land data assimilation system have been incorporated (Saha et al. 2010). Yuan et al. (2011) assessed the surface air temperature and precipitation predictions, and found that CFSv2 achieved significant improvement against CFSv1. Here, we use a Bayesian merging method (Luo et al. 2007) to bias correct and downscale the temperature and precipitation reforecasts from CFSv1 and CFSv2, which are then used as inputs to the Variable Infiltration Capacity (VIC; Liang et al. 1996) land surface model to generate hydrologic forecasts over the conterminous United States (CONUS) during 1982–2008. For comparison, a parallel run using the ESP forecast method is also carried out. Systematic evaluations of hydroclimatic variables such as precipitation, streamflow, and soil moisture are conducted to investigate the role of climate forecast models and initial conditions in seasonal flood and drought forecasting.

This paper is arranged as follows: Section 2 introduces observational data, climate and hydrologic models, the downscaling method, and the experimental design. Section 3 assesses the skill of downscaled precipitation forecasts. Sections 4 and 5 verify the seasonal streamflow

and soil moisture drought predictions, and the discussion and summary are given in sections 6 and 7, respectively.

2. Data, models, and experimental design

a. Climate forecast models and reforecast data

CFSv1 is a fully coupled ocean–land–atmosphere dynamical seasonal prediction system, which became operational at NCEP in 2004 (Saha et al. 2006). It consists of the NCEP Global Forecast System (GFS) at T62L64 ($\sim 1.875^\circ$) resolution, the Geophysical Fluid Dynamics Laboratory (GFDL) Modular Ocean Model version 3.0 (MOM3) at $\frac{1}{3}$ – 1° grid spacing, and the two-layer Oregon State University (OSU) land surface model. The monthly CFSv1 reforecasts during 1981–2008 are used in this study, with 15 ensemble members for each target month. The first and second five-member groupings have initial dates within the 9th–13th and the 19th–23rd of the month before target month, and the last five members have initial dates between the second-to-last day of the month before the target month and the 3rd of the target month (Saha et al. 2006).

CFSv2 used in the reforecast consists of the NCEP GFS at T126 ($\sim 0.938^\circ$) resolution, the GFDL MOM4 at 0.25° – 0.5° grid spacing coupled with a two-layer sea ice model, and the four-layer Noah land surface model (Saha et al. 2010). The 28-yr (1982–2009) monthly ensemble reforecasts with 24 ensemble members are now available at NCEP and National Climatic Data Center (NCDC). Beginning on 1 January, 9-month reforecasts are initiated every 5 days with 4 cycles on those days. NCEP compiled the monthly estimates as follows: for each calendar month, the reforecasts with initial dates after the 7th of that month are used as the ensemble members of the next month. For instance, the starting dates for the February ensemble members are 11, 16, 21, 26, and 31 January and 5 February (Yuan et al. 2011). NCEP has been producing real-time seasonal forecast using CFSv2 since 30 March 2011.

b. Hydrologic models and initial conditions

The VIC model (Liang et al. 1996) version 4.0.5 is used to predict soil moisture and runoff in this study. It is a semidistributed, grid-based hydrologic model with a mosaic representation of land cover and soil water storage capacity. Although there are several newer versions of VIC model, the version 4.0.5 is used as the core land surface model for our hydroclimatic forecasting system and in this study, because it has been recoded to image mode (i.e., to use flat spatial forcing files at each time step), which is the most suitable form for forecasting. Nonetheless, there is hope to update a

newer VIC version to image mode in the future. By using the efficient calibration method of Troy et al. (2008), we recalibrated the soil parameters for the water-budget version of VIC model grid by grid with runoff ratio data derived from over 1700 U.S. Geological Survey (USGS) gauges (J. K. Roundy et al. 2012, unpublished manuscript). In the water budget mode, only the daily precipitation, maximum and minimum air temperature, and surface wind are needed to drive the model. Runoff generated within a grid cell is routed to the stream gauge location using a linear routing model developed by Lohmann et al. (2004).

Although a hydrologic model can reach equilibrium after a few years, in this study, the VIC and routing models are spun up from 1949 to 1979 using the University of Washington (UW) observed forcing data of Maurer et al. (2002) and then forced by National Land Data Assimilation System (NLDAS-2) data (Xia et al. 2012) over 1980–2008 to generate initial conditions for the seasonal hydrologic forecasts. The UW and NLDAS-2 data cover the periods 1949–2000 and 1979–present, respectively. The datasets were not blended during their overlapping period (1979–2000) because of differences between the two datasets. Based on the multiyear (1979–2000) mean statistics, we found that the two datasets disagree in precipitation for some grid cells over Rocky and Appalachian Mountain regions; NLDAS-2 provides systematically higher 2-m maximum air temperature than UW data over central United States and higher 2-m minimum air temperature than the latter throughout most of CONUS. Given that our hindcast period is 1982–2008, all the initial hydrologic conditions are consistently generated from the NLDAS-2-driven simulation, and the soil moisture drought evaluation in section 5 is also based on the same simulation. We believe there is limited impact from using UW data during the spinup (1949–79) on the initial conditions and hydrologic simulation because of the 2-yr transition period (1980–81) with NLDAS-2 and an improvement (especially for deeper soil moisture stores) to the spinup only using 3-yr (1979–81) NLDAS-2 data. Although using both datasets to construct the climatology in the Bayesian downscaling and ESP as described in section 2c is suboptimal, both the climate model-based approach and ESP are based on the same merged data, making their comparisons consistent.

c. Bayesian downscaling and experimental design

Figure 1 is the flowchart of the seasonal hydroclimatic forecast system. Monthly precipitation and temperature reforecasts during 1982–2008 from CFSv1 and CFSv2 are bias corrected and downscaled to $\frac{1}{8}^\circ$ over CONUS using the Bayesian merging method of Luo et al. (2007).

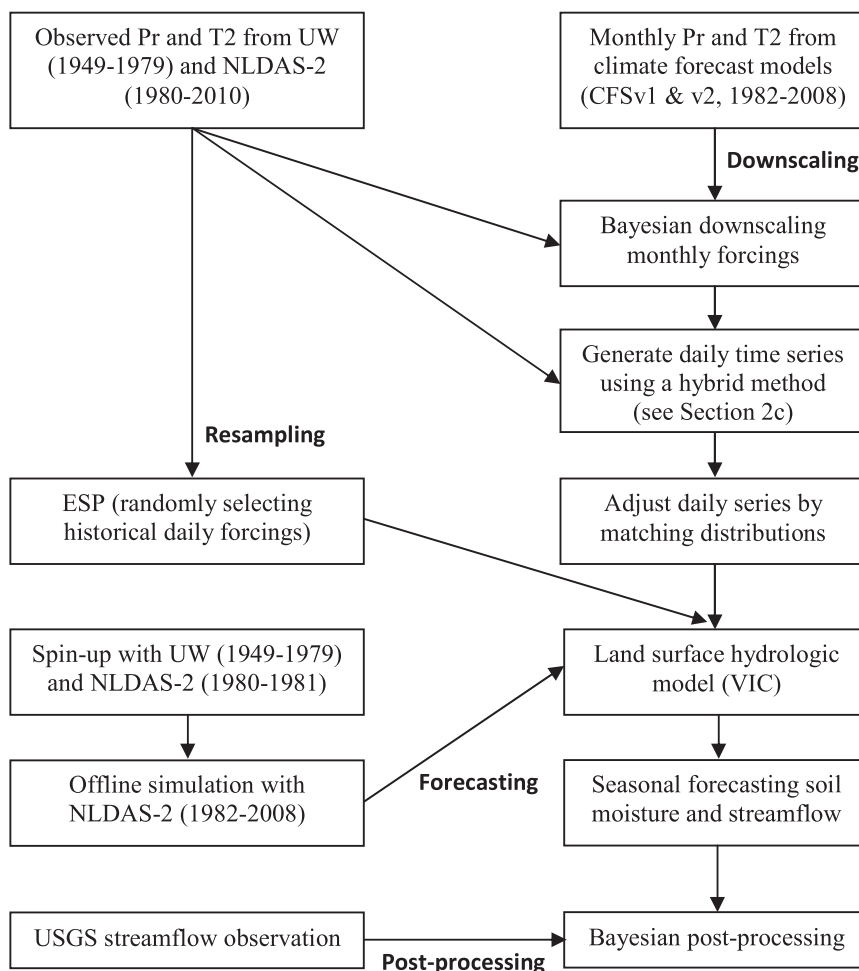


FIG. 1. Flowchart of the seasonal hydroclimatic forecast system including data and methodology.

For each calendar month, all ensemble members of climate forecasts (15 from CFSv1 and 24 from CFSv2) are used to construct a cumulative distribution function (CDF) of the forecasts. The 31-yr (1949–79) UW forcing, merged with the 31-yr (1980–2010) NLDAS-2 observations in that calendar month, are used to construct a CDF of observations. Both CDFs are then transformed to normal space through quantile-mapping. A linear regression function between the transformed observation and ensemble mean of the forecast is fitted to create the likelihood function in the Bayesian method. The regression procedure is repeated 100 times and each time the data in the target year and two other random years are excluded for cross validation, and then the average coefficients are computed for the likelihood function. Consequently, the posterior distribution in the normal space is obtained using the likelihood function, the target forecast, and a prior forecast based on 62-yr observation climatology (UW merged with NLDAS-2). The resulting posterior distribution is finally transformed

back to its original space through the same quantile mapping. In Luo et al. (2007), the posterior variance consists of both linear regression error and the ensemble spread, while in this study, to improve the interannual variability of the downscaled climate forcings, we drop the ensemble spread part in posterior variance according to the work of Yoon et al. (2012).

With the mean and variance of the posterior distribution for the spatially downscaled monthly variables, a hybrid method including both the historical-analog criterion and random selection is used to generate 20 daily time series (Luo and Wood 2008). In the historical-analog criterion, 10 historical daily forcings that have similar spatiotemporal patterns to the mean of posterior distribution are selected. To avoid overconfident forecasts, another 10 daily series are randomly selected from history. In Luo and Wood (2008), they adjusted the daily series by matching the observation distribution with forecast distribution. In that manner, the rank structure of historical observations is retained but not their actual

values. However, we find that such a matching method causes obvious bias when the forecast distributions significantly deviate from observations. Therefore, we use a simple scaling method to adjust the resulted series according to the monthly ensemble mean of the posterior distribution. For the ESP experiment, the 20-member ensemble forcings are randomly chosen from the 62-yr merged data, excluding the target year.

Based on the downscaled climate forcings from CFSv1 and CFSv2, randomly selected forcings through the ESP procedure, and the initial conditions from the offline simulation, we use VIC model to produce 6-month, 20-member ensemble hydrologic forecasts over CONUS starting on the 1st of each calendar month during 1982–2008. This is equivalent to 9720-yr VIC simulation at $1/8^\circ$ over CONUS. As introduced in section 2a, some ensemble members of CFSv1 and v2 are about 20 days old at the beginning of the target month. Therefore, hereafter, we call the month-1 forecast as forecast at 0.5-month lead, the month-2 forecast as 1.5-month lead, and so on.

3. Predictive skill of downscaled precipitation

To have an overview of the performance in the deterministic forecast, we calculate the squared correlation R^2 (coefficient of determination), for basin-averaged ensemble mean monthly precipitation. For each season, ensemble mean values of monthly precipitation from downscaled climate forecasts or ESP during 1982–2008 are compared to the corresponding observations by plotting a scatterplot and then calculating R^2 . The R^2 values represent the fraction of variance explained by the forecasts.

The differences in R^2 for the ensemble mean monthly precipitation at 0.5-month lead (month-1 forecast) between the downscaled CFSv1/CFSv2 forecasts and ESP are shown in Fig. 2 for 14 large basins (numbered in Fig. 2b). These basins will be used throughout the remainder of this paper to compare the spatial differences in the forecast predictions. A summary of the basin characteristics is provided in Table 1, including a list of basin area and wet–dry seasons. CFSv1 shows generally higher predictive skill than ESP over the 14 large basins, except for lower Mississippi during winter and summer (Figs. 2a,e), NE during spring and fall (Figs. 2c,g), and the Ohio basin during winter (Fig. 2a). A common feature for the degradation is that they do not happen in wet or dry seasons (Table 1), indicating some challenges in predicting precipitation during transition seasons over those regions. Most of the significant improvements beyond ESP occur in wet seasons for the Great Lakes region (Fig. 2e), Ohio (Fig. 2c), upper Mississippi (Fig. 2e),

Missouri (Fig. 2e), and south central (Fig. 2g) in the eastern and central United States. For the western United States, only the California region shows obvious improvement during wet seasons (Fig. 2a). As compared with CFSv1, CFSv2 improves predictive skill over many regions, especially during winter and spring (Figs. 2b,d). Unlike CFSv1, which shows lower skill than ESP in some regimes, CFSv2 shows consistent improvement against ESP throughout CONUS for all seasons. In particular, CFSv2 has significantly higher skill than ESP (R^2 difference larger than 0.05) during wet seasons for every large basin, except for the SE (Fig. 2f). Note that, in this study, we downscale the climate forecasts to $1/8^\circ$ for the inputs to the hydrologic model and then aggregate them to large basins to have a general assessment before evaluating their impact on hydrologic forecasts. Thus, we can expect higher skill if we downscale the basin-averaged forecasts directly because of higher skill for climate model forecasts over larger domains.

Besides the R^2 , the evolution of root-mean-square error skill score (SS_{RMSE}) for the ensemble mean monthly precipitation at different lead times is also analyzed (Fig. 3). The SS_{RMSE} is defined as $1 - (\text{RMSE}/\text{RMSE}_{\text{ESP}})$, where RMSE and RMSE_{ESP} are the root-mean-square error for climate models (CFSv1/CFSv2) and ESP, respectively. Here, $SS_{\text{RMSE}} = 1$ indicates perfect forecast, while SS_{RMSE} less than zero means the CFSv1 or CFSv2 forecast is worse than ESP. The solid thick lines in Fig. 3 demonstrate that skill scores are generally above zero for the 14 basins over all leads, except for the long lead forecasts in Ohio (Fig. 3c) and California (Fig. 3n). Thus, the climate model forecasts have less error than ESP in most cases. Most basins show obvious skill decline over lead times in the first 2–5 months, while there are no obvious declines in the NE (Fig. 3a) and Arkansas Red (Fig. 3h). The abnormal increase in skill at long leads is due to the seasonal variation of precipitation that significantly affects the magnitude of errors for the models and ESP, and should be interpreted as noise. For the first month forecasts, CFSv2 has significantly higher skill than CFSv1 over 9 (Figs. 3a,c,d,f,g,i–l) of the 14 large basins; lower skill over Great Lakes region (Fig. 3b) and upper Mississippi (Fig. 3e); and comparable skill over Arkansas Red, Great Basin (Fig. 3m), and California.

Wood et al. (2005) investigated the influence of ENSO on the skill of basin-averaged precipitation over western United States. Here we also compare the unconditional SS_{RMSE} calculated using all forecasts with SS_{RMSE} conditional on the forecasts having strong Niño-3.4 SST anomalies (>1.0 K) 1 month before the target initiation month. For instance, for all 6-month forecasts starting from December, we selected the forecasts if there is a strong SST anomaly in November. Figure 3 shows that

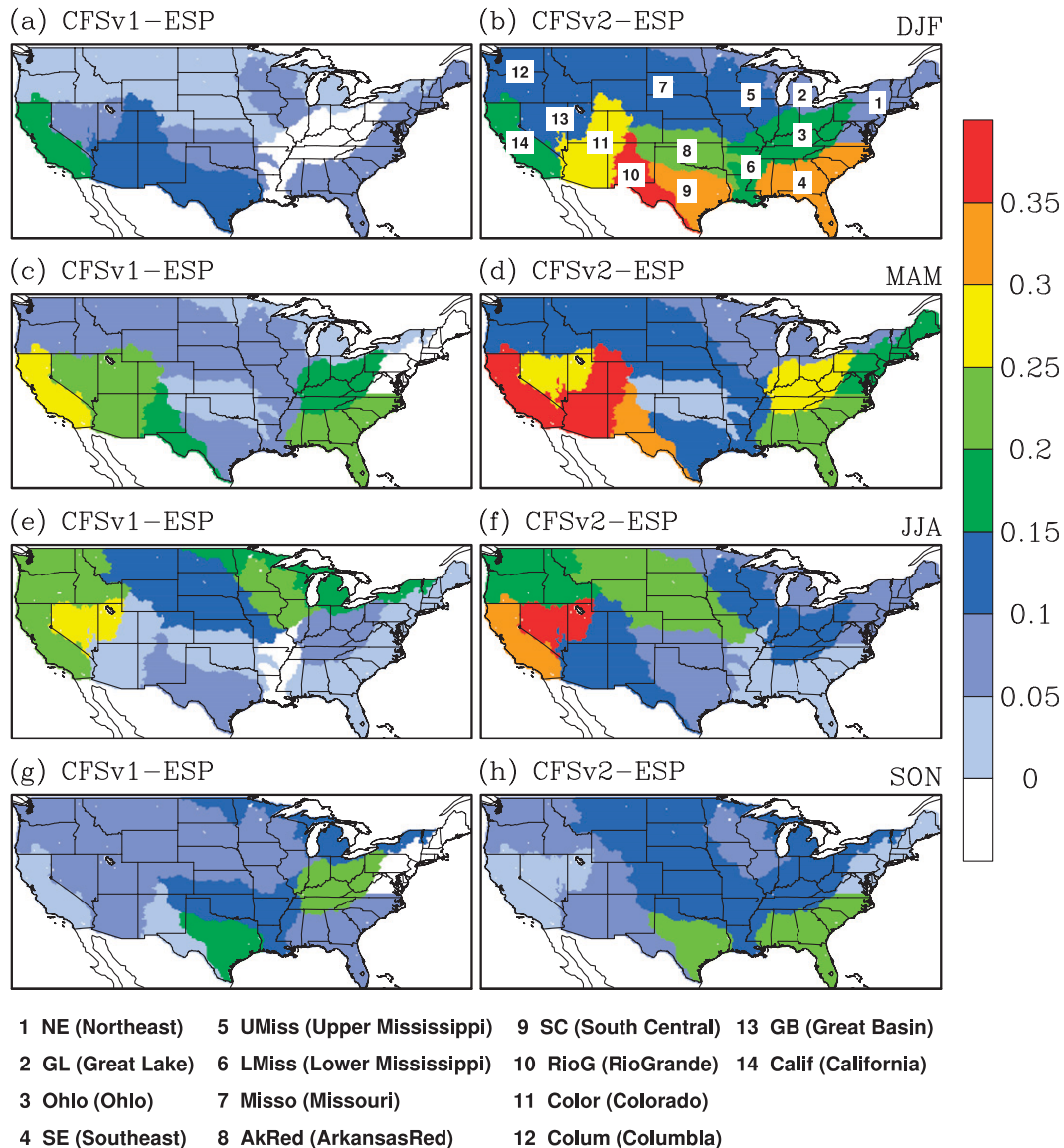


FIG. 2. Differences in squared correlation R^2 of basin-averaged ensemble mean monthly precipitation at 0.5-month lead between downscaled climate forecasts and ESP: (top to bottom) DJF–September to November (SON) and (left) CFSv1 and (right) CFSv2.

not all basins over CONUS are positively affected by ENSO, such as Missouri and Arkansas Red (Figs. 3g,h). However, the forecasts conditional on ENSO do have significantly higher skill than the unconditional forecasts over the SE (Fig. 3d), Colorado (Fig. 3k), and California (Fig. 3n) in the first 3–4 months, indicating the positive effects of tropical Pacific SST anomalies on precipitation forecasts through teleconnections.

The above results are based on deterministic and basin-scale analysis. To evaluate probabilistic and grid-scale precipitation forecast skill, we use the ranked probability skill score (RPSS; Wilks 2011) defined as

$1 - (RPS/RPS_{ESP})$, where RPS and RPS_{ESP} are the ranked probability scores for CFSv1/CFSv2 forecasts and ESP, respectively. RPSS is quite similar to SS_{RMSE} , except that the former is based on multicategory probabilistic forecasts using all ensemble members. In this study, we classify monthly precipitation into three categories, small (<33th percentile), median, and heavy (>67th percentile), based on the climatological distributions for each calendar month using observations. We calculate RPSS grid by grid and then compute the percentages of forecasts with positive values: that is, percentages of forecasts that are more skillful than ESP.

TABLE 1. Information for the 14 large basins. Wet and dry seasons (DJF, MAM, JJA, and SON) are determined by 27-yr (1982–2008) areal-averaged precipitation observations.

Basin	Full name	Area (km ²)	Wet season (mm day ⁻¹)	Dry season (mm day ⁻¹)	No. of gauges
1. NE	Northeast	392 420	JJA (3.20)	DJF (2.64)	279
2. GL	Great Lakes	482 020	JJA (2.86)	DJF (1.27)	147
3. Ohio	Ohio	494 060	MAM (3.46)	SON (2.75)	263
4. SE	Southeast	601 440	JJA (4.34)	SON (3.05)	283
5. UMiss	Upper Mississippi	451 080	JJA (3.38)	DJF (0.99)	192
6. LMiss	Lower Mississippi	290 640	MAM (4.00)	SON (3.25)	73
7. Misso	Missouri	1 311 520	JJA (2.17)	DJF (0.62)	101
8. AkRed	Arkansas Red	534 380	JJA (2.59)	DJF (1.23)	53
9. SC	South Central	393 680	SON (2.36)	DJF (1.61)	57
10. RioG	Rio Grande	311 080	JJA (1.57)	DJF (0.58)	7
11. Color	Colorado	590 800	JJA (1.04)	MAM (0.81)	52
12. Colum	Columbia	724 920	DJF (3.51)	JJA (1.04)	130
13. GB	Great Basin	356 020	DJF (1.05)	JJA (0.58)	19
14. Calif	California	393 680	DJF (3.22)	JJA (0.25)	78

The results for different seasons and all months (ANN) are shown in Fig. 4. It is found that generally more than 50%–60% of climate forecasts are more skillful than ESP over 14 basins in different seasons for different leads. Unlike the RMSE analysis for basin-averaged ensemble mean precipitation, there is no clear decline of RPSS over lead times except in the first 2 months. This might be due to the low skill of climate models at local scales after the first month forecasts. Thus, the downscaled forecasts tend to approximate the prior distribution: that is, the climatological distribution, which is similar to ESP. Therefore, though there are some skill enhancements beyond ESP at long leads [e.g., June–August (JJA) forecasts over NE], they are random score excursions. Focusing on the first month forecasts, CFSv2 has moderate improvements against CFSv1 for the annual average results (Fig. 4, right column). The improvements are obvious in the wet seasons for some humid basins, such as NE, Ohio, and Columbia, which is encouraging since heavy rainfall is more challenging for categorical forecasts.

4. Verification of streamflow forecast

Before using the ESP generated or downscaled forcings to drive the hydrologic model, we evaluated the capability of the recalibrated VIC model in capturing interannual variations of streamflow given observed forcings. The Nash–Sutcliffe efficiency (NSE) coefficients for monthly streamflow are calculated at 1734 USGS gauges for the offline observation-driven simulation during 1982–2009, and the results are shown in Fig. 5. The gauges used in this study cover most of the eastern United States and West Coast area. The drainage areas vary from 140 (the area of a $\frac{1}{8}^\circ$ grid cell) to 525 768 km². All gauges have at least 5 yr of streamflow records, and the average fraction of

missing data is about 9%. To fill the missing values, we use offline simulated streamflow and all available gauge records to fit a linear regression function and then use it to estimate the missing observation based on offline simulation. This procedure is similar to Dai et al. (2009), but here the regression is being used regardless of the significance of the regression (Pan et al. 2012). The filled streamflow observations are used to calculate NSE and other forecast verifications. Figure 5 shows that most gauges have NSE > 0.7 in the eastern and western parts of the country and some gauges have NSE < 0.5 over the central United States. Note that all 1734 gauges used in this study have NSE > 0.3 based on the records with unfilled observations.

With downscaled forcings and offline simulated initial conditions, 6-month hydrologic forecasts starting from the 1st of each calendar month in each year are carried out over CONUS during 1982–2008. Given the limited skill of precipitation after one season, streamflow analysis is only carried out for the first 3 months in this study. Figure 6 shows the SS_{RMSE} for ensemble mean monthly streamflow from climate model-based forecasts. Similar to the precipitation analysis, the streamflow from ESP is used as the reference forecast. During the first month, CFSv1 and CFSv2 have moderate improvements against ESP over most watersheds, with high skill score (>0.1) occurring over California (Figs. 6a,b) and some over the Ohio basin for CFSv2. Based on the skill scores averaged over the 14 large basins, CFSv2 reduces RMSE for streamflow forecasts from ESP by about 4%–7% in the eastern United States and less than 4% over the west, and the errors are reduced by 10% over California. In the second and third months, climate model-based forecasts are comparable to the ESP at more gauges (Figs. 6c–f: green dots). The large basin-averaged skill scores indicate that the error reductions are below 3%

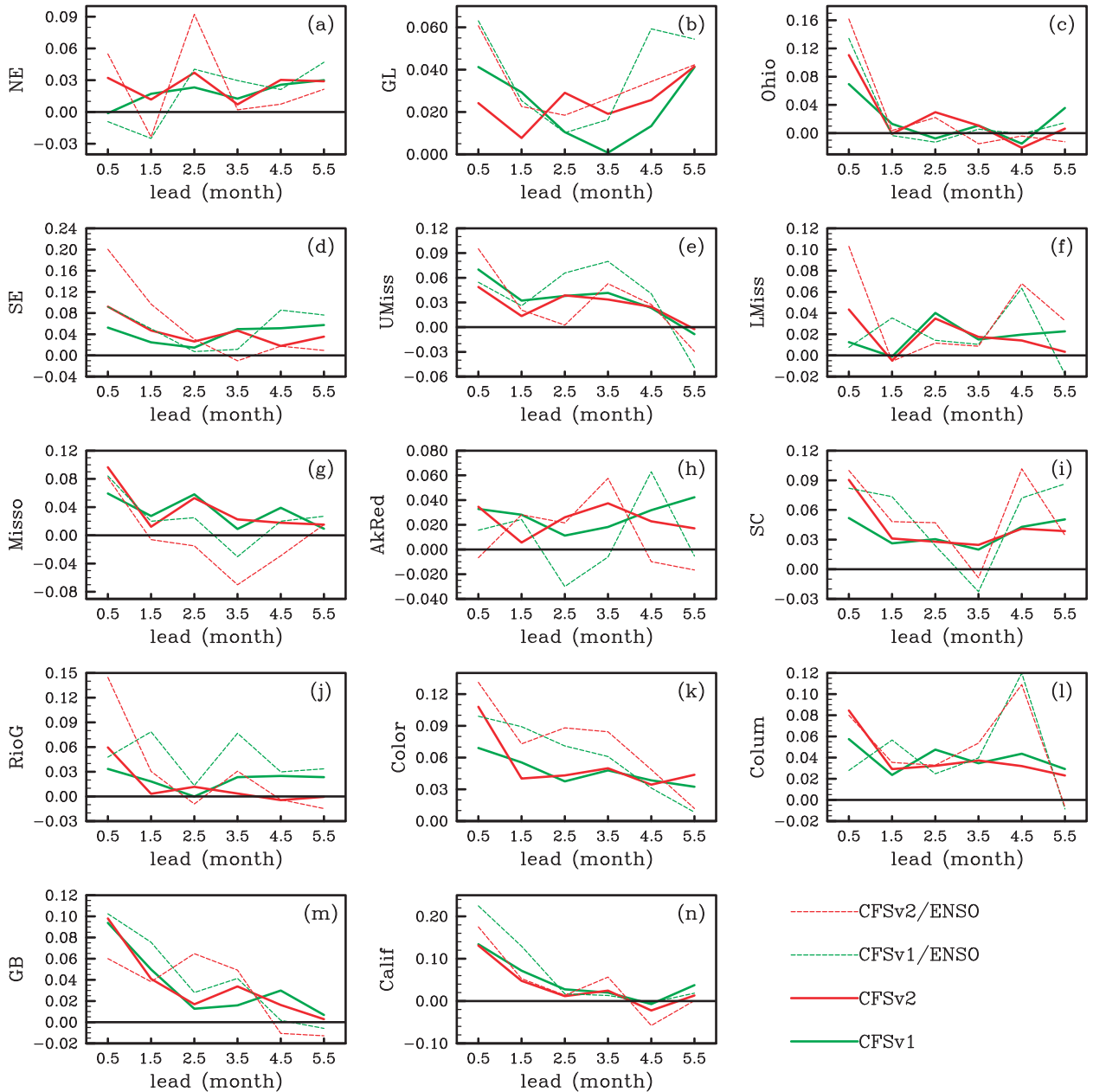


FIG. 3. The SS_{RMSE} for basin-averaged ensemble mean monthly precipitation from downscaled CFSv1 and CFSv2 forecasts at different leads based on unconditioned (all forecasts; solid thick lines) and ENSO conditioned (forecasts having Niño-3.4 SST anomaly > 1.0 K at 1 month before the target initiation month; dashed thin lines) for (a)–(n) the 14 basins. The reference is ESP.

(except the 1.5-month forecasts over California), and CFSv1 is slightly worse than ESP over the Ohio basin. The blue dots indicate that, although there are some promising forecast results for large basin-averaged precipitation, streamflow forecasts over some small catchments are still quite challenging.

Similar to the precipitation analysis, we also calculate probabilistic skill score for the monthly streamflow forecasts. Table 2 lists the percentages of positive RPSS

for the streamflow forecasts averaged over the 14 large basins. For the 0.5-month lead, 57%–76% of the CFSv1 forecasts and 60%–77% of the CFSv2 forecasts over different large basins are more skillful than ESP, respectively. Figure 6c shows that there are negative SS_{RMSE} at some gauges over Ohio basin for CFSv1 at 1.5-month lead, and Table 2 also illustrates that CFSv1 is comparable or slightly worse than ESP over Ohio beyond 1 month. For CFSv2, the skill also decreases over lead

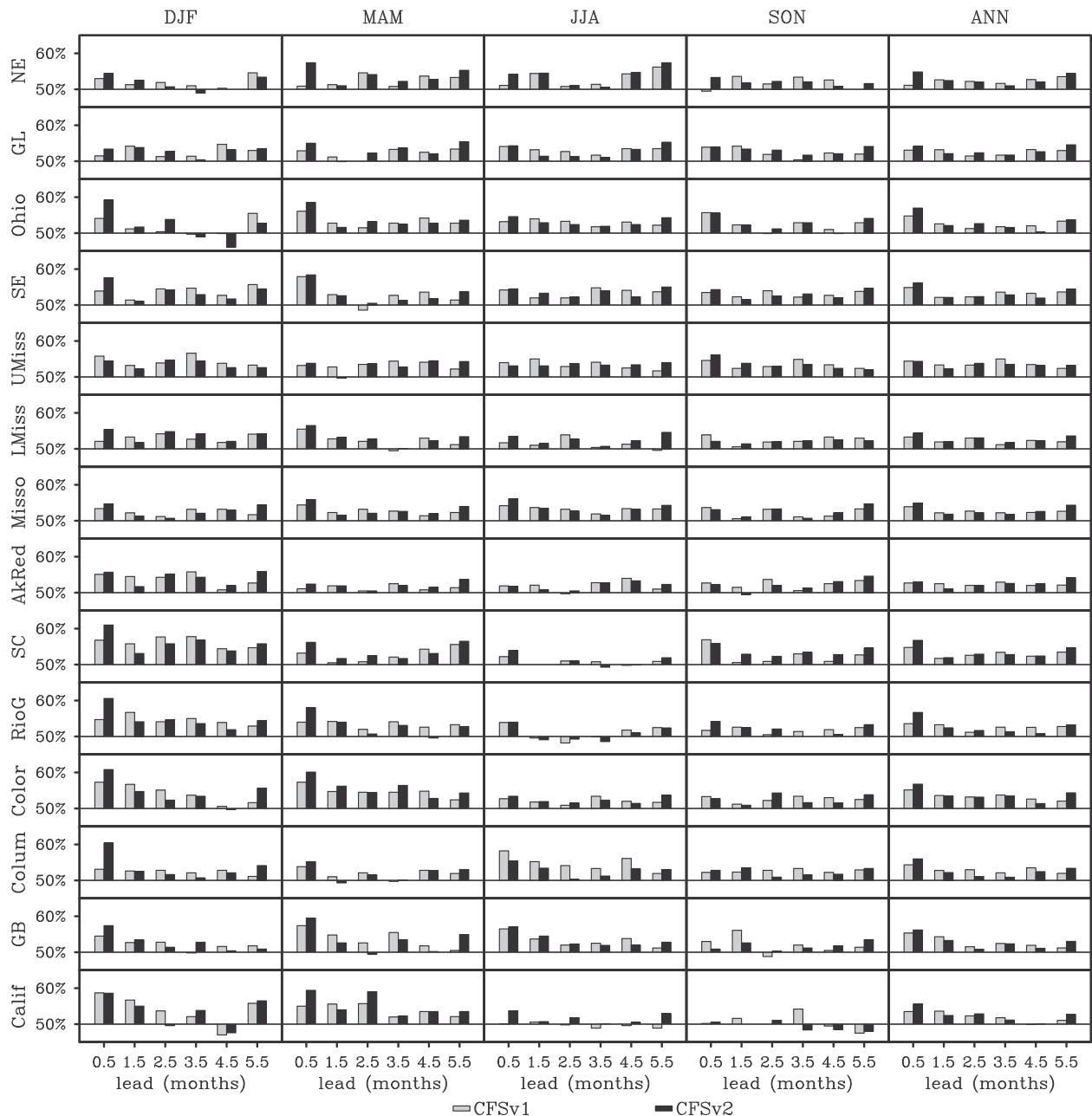


FIG. 4. Percentages of skillful probabilistic forecasts ($RPSS > 0$) averaged over (top to bottom) the 14 basins for downscaled CFSv1 and CFSv2 monthly grid-scale precipitation at different leads for (left to right) DJF–ANN. The reference is ESP.

times, but at least more than 51% of the forecasts are more skillful than ESP for each large basin.

As shown in Table 2, climate model-based streamflow forecasts have more advantages against ESP over western basins than those over the east. For instance, climate models have more than 60% of the skillful forecasts over south central, Rio Grande, Colorado, Great Basin, and California out to 3 months. Does that mean the climate model-based streamflow probabilistic forecasts have higher

accuracy over arid/semiarid regions (western United States)? To answer the question, we calculate the percentages of positive RPSS for ESP, CFSv1, and CFSv2 against the climatological forecasts. Here the climatological forecasts for each calendar month are constructed by using all historical USGS observations in the same month except the target year as ensemble members. Figures 7a1–n1 show that all three approaches have more skill against climatological forecasts over the eastern humid basins than

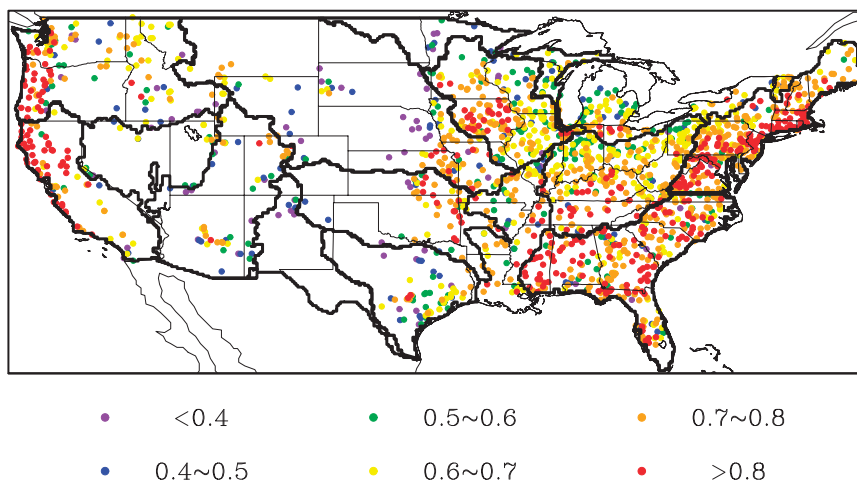


FIG. 5. The NSE coefficients for monthly streamflow simulated by the recalibrated water-budget VIC model at 1734 USGS gauges.

those over western arid basins. Therefore, although CFSv1 and v2 have more advantage against ESP over arid regions, their forecasting skill is still quite low.

Figures 7a1–n1 also demonstrate that the three forecasting approaches are comparable to or even poorer than the climatological forecasts over most basins at 1.5-month lead, which might come from systematic errors of hydrologic models or underestimated uncertainties from initial conditions. For example, Yuan and Wood (2012a) found underforecasting or underdispersion errors for climate model-based ensemble streamflow forecasts over 50 gauges in the Ohio basin based on the analysis of verification rank histograms (Wilks 2011). So, following that work, we also use them to verify the ensemble characteristics in this study. The rank histograms for each basin are plotted in Fig. 8. Given that there are 20 ensemble members for each forecast, rank 1 represents the proportion of the forecasts with all 20 ensemble members larger than the observation, rank 2 represents the proportion of the forecasts with 19 members larger than the observation and one member smaller than the observation, and rank 21 represents the proportion of the forecasts with all 20 members smaller than the observation. Ideally, the rank histogram should be uniform. For simplicity, we only show the histograms for the month-2 CFSv2 streamflow forecasts, although the histograms slightly differ at different leads and among different forecasting approaches. Similar to Yuan and Wood (2012a), Fig. 8 shows that CFSv2 has underdispersion errors in many basins, especially over the western United States. (Fig. 8k–n), which indicates that the ensemble members are too similar to each other, different from verification, and the ensemble forecast is too sharp. CFSv2 also has overforecasting biases over arid-semiarid basins, where more than 25% of the forecasts

having all ensemble members systematically larger than observations (Figs. 8i–k,m,n). These systematic biases make CFSv2 worse than the climatological forecasts over south central, Columbia, and California at 1.5-month lead (Figs. 7i1,l1,n1). For the humid basins, CFSv2 is worse than climatological forecasts over the Great Lakes (Fig. 7b1) and Ohio (Fig. 7c1), due to the underdispersion (Fig. 8b) and overforecasting (Fig. 8c) errors, respectively.

To correct the above errors, Yuan and Wood (2012a) proposed a postprocessing procedure based on Bayesian procedures. The Bayesian method is similar to Luo et al. (2007) and is discussed in section 2c. In this section, it is being applied to the monthly streamflow forecasts instead of precipitation. The postprocessing procedure is applied to the ESP, CFSv1, and CFSv2 forecasts by matching CDFs of USGS streamflow observations through cross validation, resulting in significant improvements. Figures 7a2–n2 show that, with the postprocessing, all three approaches have more than 60% skillful forecasts against climatological forecasts out to 2 months. The resulting rank histograms are also corrected to approximately uniform distributions (not shown). Therefore, postprocessing is essential for accurate and reliable probabilistic streamflow forecasts. After postprocessing, CFSv1/CFSv2 forecasts are quite similar to ESP over the upper Mississippi, Missouri, Rio Grande, Great Basin, and California (Figs. 7e2,g2,j2,m2,n2) but are better than ESP over the other nine basins.

5. Assessment of drought prediction

Besides evaluating the streamflow forecast, we also assess the seasonal predictions of soil moisture drought. Many previous studies focused on identifying and

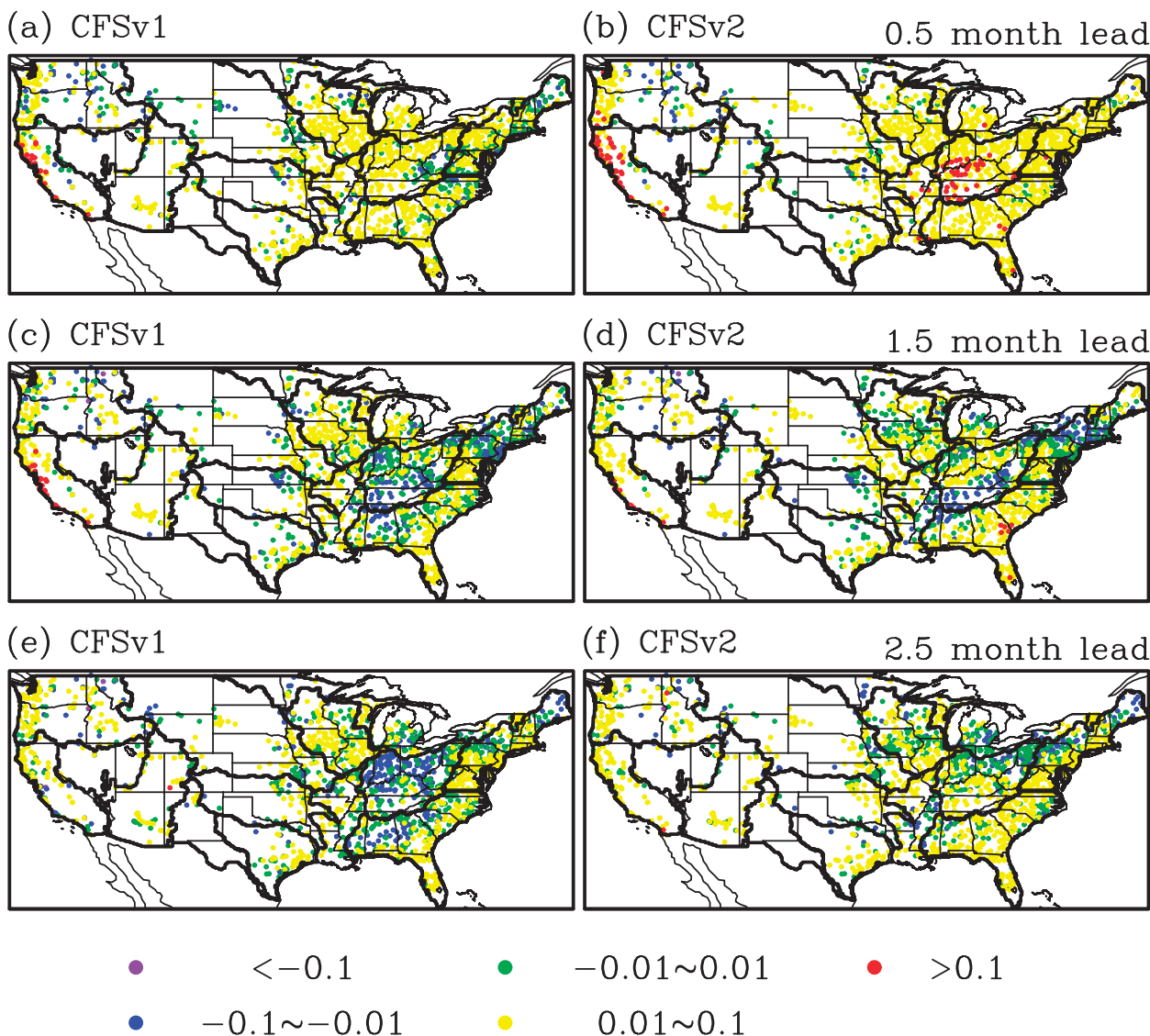


FIG. 6. The SS_{RMSE} for ensemble mean monthly streamflow forecasts from (left) CFSv1 and (right) CFSv2 calculated at 1734 gauges for (top to bottom) lead times of 0.5–2.5 months. The reference is ESP.

characterizing historical soil moisture drought (Sheffield et al. 2004; Andreadis et al. 2005; Wang et al. 2009) and projecting future changes in drought occurrence (Sheffield and Wood 2008; Dai 2011), but only a few were dedicated to seasonal drought forecasts (Luo and Wood 2007; Mo 2011; Yuan et al. 2011). In this study, we assess the models' capability in forecasting short-term drought in terms of frequency, area, and severity. Monthly percentiles for the forecasted and offline simulated soil moisture are calculated for each grid cell based on the climatology of a 31-yr (1980–2010) VIC offline simulation. According to CPC's drought classification, we define drought occurrence when the soil moisture percentile is below 20%.

Figure 9 shows the ratio of forecasted over offline simulated drought frequency averaged among all forecasts during 1982–2008. For the droughts that last for at least 1 month (Fig. 9, first row), more than 80% of them are captured by ESP forecasts over central United States; 60%–70% are captured over SE; and less than 50% can be predicted over NE, Ohio, and western coast areas. CFSv1 has improved forecasts over the lower Mississippi and California, while CFSv2 has even further enhancements, especially over the eastern United States. For the 2-month-duration droughts (Fig. 9, second row), offline simulation indicates that the occurrence frequency of these events is less than 10% over eastern United States and Pacific Northwest. Thus, all three approaches

TABLE 2. Percentages (%) of positive RPSS for monthly streamflow forecasts at 1734 gauges averaged over the 14 basins. The reference is ESP.

Basin	0.5-month lead		1.5-month lead		2.5-month lead	
	CFSv1	CFSv2	CFSv1	CFSv2	CFSv1	CFSv2
1. NE	57	61	52	52	52	53
2. GL	58	60	54	54	52	53
3. Ohio	59	63	50	52	49	51
4. SE	59	62	54	56	53	54
5. UMiss	60	60	56	56	55	55
6. LMiss	59	62	54	55	54	55
7. Misso	64	66	59	60	56	57
8. AkRed	62	63	56	57	56	56
9. SC	70	71	62	63	61	62
10. RioG	76	76	68	69	61	62
11. Color	73	74	67	68	63	64
12. Colum	64	67	59	61	58	58
13. GB	76	77	70	70	66	66
14. Calif	74	75	70	70	67	68

have lower skill in forecasting them over these areas. However, CFSv2 forecasts have a consistent improvement over ESP and CFSv1 across the country. As the drought duration increases from 3 to 6 months, the chance of successfully forecasting their frequency decreases, especially in the eastern United States and West Coast areas (Fig. 9, third and fourth rows). In contrast, all three approaches have plausible performance in predicting drought frequency over central United States out to 6 months.

Another important characteristic for drought is its spatial extent. To assess the models' capability in forecasting interannual variations of drought area, R^2 is calculated between forecasted and offline simulated extents for each large basin. Figure 10 shows the R^2 of ESP forecasted drought area during 1982–2008, where forecast lead in months is along the y axis and the target or verification month is along the x axis, so the value of July with 1.5-month lead is the predictive skill for July drought area from the forecast that is initiated in June. Similar to the frequency prediction above, ESP has higher predictive skill for drought area over the basins in central United States, such as the Missouri (Fig. 10g), Rio Grande (Fig. 10j), and Colorado (Fig. 10k), and the skill is generally lower over humid basins than the arid basins. Given that there is no forecast information in the climate forcings for the ESP method, its skill mostly reflects the impact of initial soil moisture conditions on drought prediction. In the Missouri and Colorado basins, ESP can predict more than 50% of the interannual variations out to 3 months for the forecasts initiated during December–February (DJF) and March–May (MAM). These seasons are indeed their dry seasons (Table 1), which indicates that soil moisture initial

condition may have more impact on the predictive skill for the forecasts initiated during dry seasons over the two basins. Other arid–semiarid basins (upper Mississippi, Arkansas Red, south central, Rio Grande, Great Basin, and California) also show higher predictive skill for forecasts initiated from dry seasons. However, it becomes more complicated for the humid basins. The predictive skill is higher for the forecasts initiated both during wet and dry seasons over the northern humid basins, such as the NE (Fig. 10a), Great Lakes (Fig. 10b), and Columbia (Fig. 10l), while for the southern humid basins high predictive skill usually occurs in the forecasts initiated during transition seasons, such as Ohio in JJA (Fig. 10c) and SE and lower Mississippi in DJF (Figs. 10d,f).

To compare the climate model–based forecasts with ESP in terms of predictive skill for drought area, the differences in R^2 are provided in Fig. 11. For the humid basins, CFSv2 shows significant improvements for the forecasts initiated during transition seasons over the NE (Fig. 11a2) and the dry season over Ohio, SE, and lower Mississippi (Figs. 11c2,d2,f2). For the arid–semiarid basins, there are obvious improvements for wet season over the Missouri (Fig. 11g2) and California (Fig. 11n2), and for dry seasons over other basins (Figs. 11h2–k2). CFSv1 is generally worse than CFSv2, except for JJA over the Great Lakes and SE (Figs. 11b1,d1) and for MAM over the Arkansas Red (Fig. 11h1). Figure 11 also shows that both CFSv1 and CFSv2 have lower predictive skill than ESP for the forecasts over the lower Mississippi in September (Figs. 11f1,f2), which might be due to noise from the ESP. In fact, ESP has unexpectedly higher skill at long leads than those at short leads for the forecasts in September (Fig. 10f). The attribution of this anomaly is unresolved at this time.

Besides the drought frequency and area, the predictive skill for drought severity is also evaluated. To have a general comparison among three forecasting approaches over CONUS, we define the regional accumulated severity S as $S = \sum_{i=1}^n \sum_{j=1}^t (1 - P_{i,j})$, where n and t represent that the number n of drought grid cells with drought durations of t months and $P_{i,j}$ is the monthly percentile of soil moisture for a specific grid cell in a specific drought month. Note that n and $P_{i,j}$ may be different between offline simulation and the forecasts; thus, the defined S is used to quantify regional accumulated soil water deficit during the drought period. Figure 12 shows the predictive skill of S for three types of forecasts over the CONUS region for different durations. To be consistent with the 3-month-duration droughts, which have four different leads based on 6-month forecasts, only the results with lead times up to 3.5 months are shown. The predictive skill of severity tends to be higher during winter because of the strong

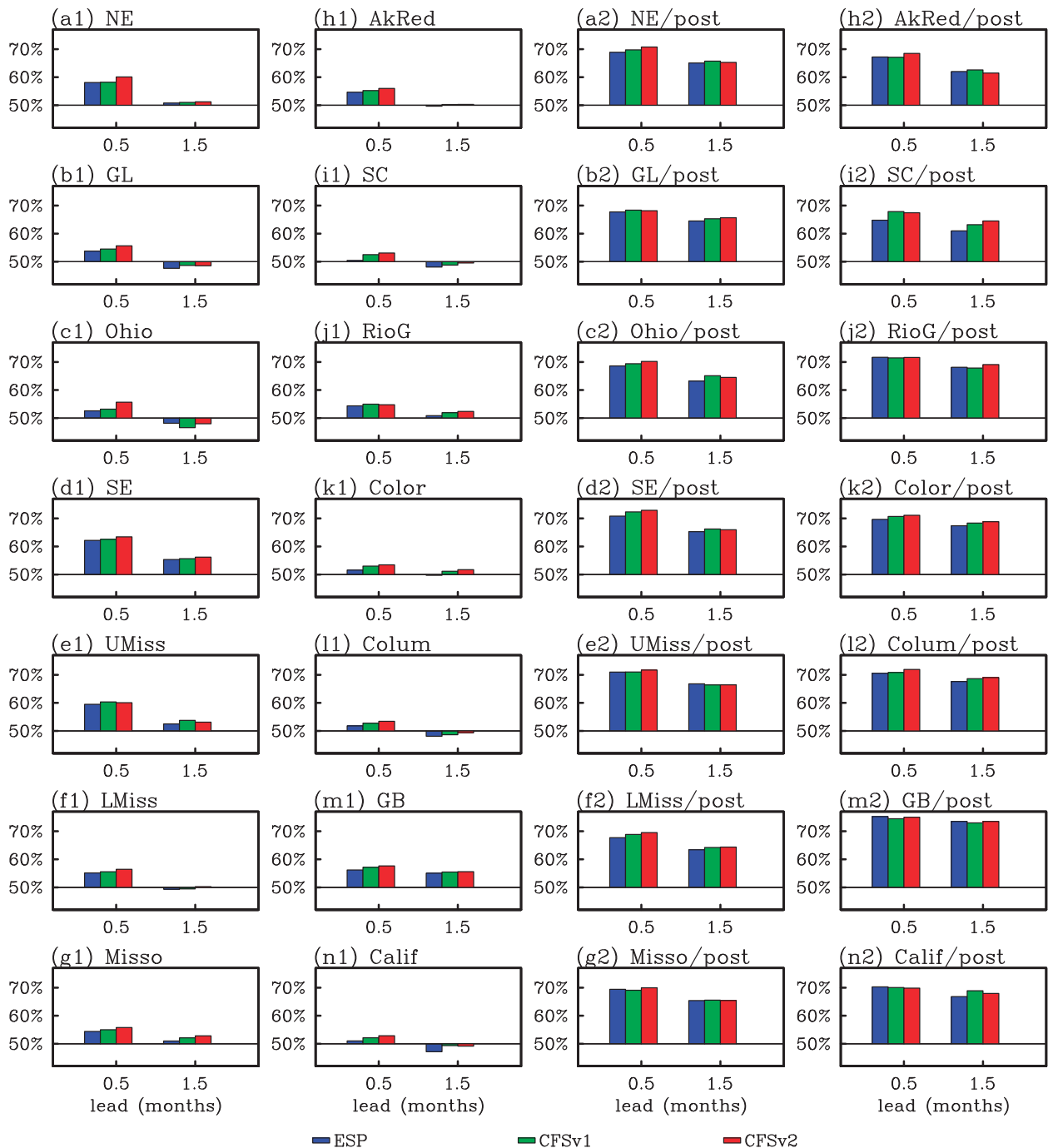


FIG. 7. Percentages of positive RPSS for monthly streamflow forecasts (a1)–(n1) without and (a2)–(n2) with postprocessing at 1734 gauges averaged over the 14 large basins. The reference is climatological forecast.

initial soil moisture control and/or better precipitation prediction skill from the climate forecast models. However, it is not as low as we expected during the summer, while the lowest skill occurs during the spring and fall. Therefore, the predictive skill for short-term drought has different seasonal characteristics as compared

with precipitation, indicating the important role of initial soil moisture condition in the forecasting. As compared with ESP, climate models do offer added value, and their advantages become clearer at long leads where there is less impact from initial conditions. Although CFSv2 is better than CFSv1 during the winter, summer, and fall

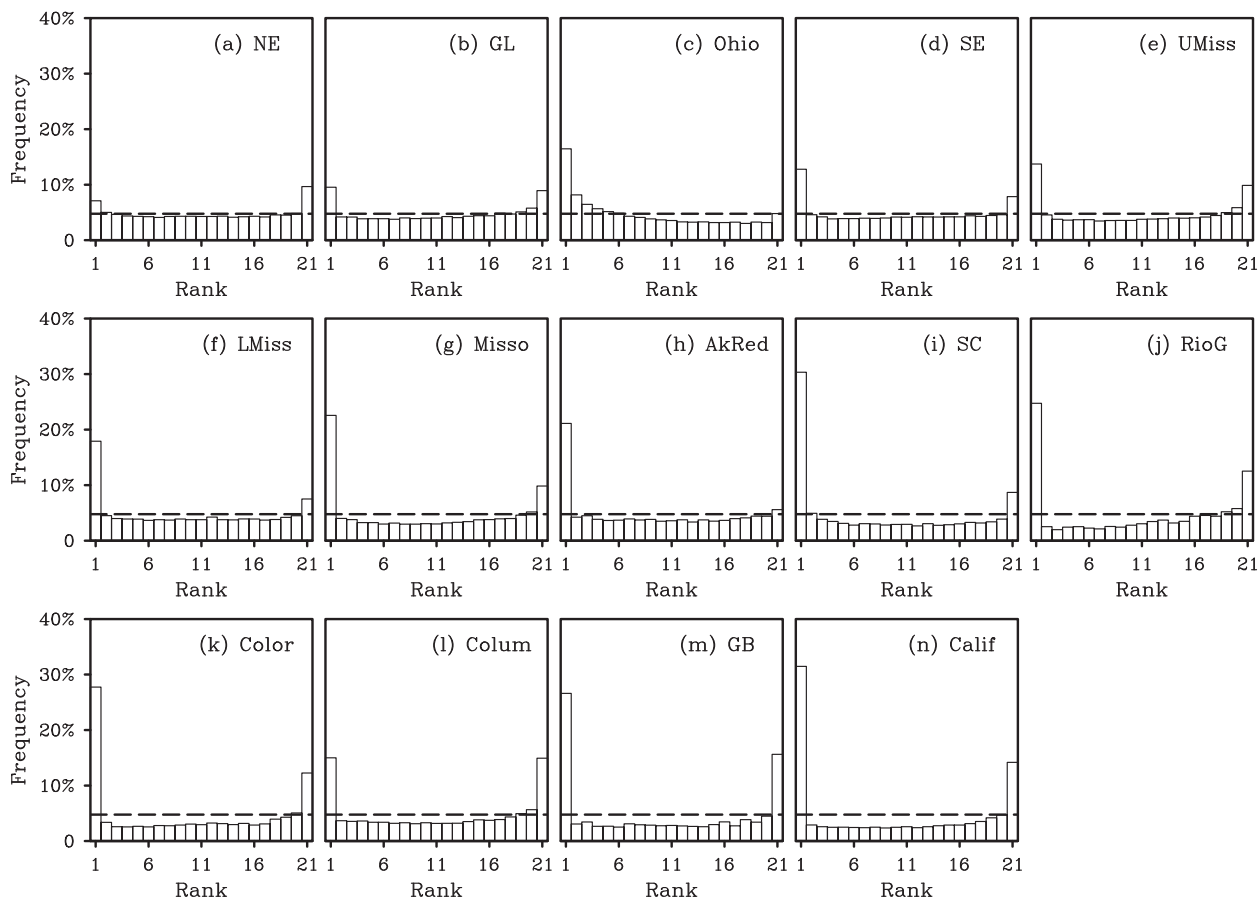


FIG. 8. Verification rank histograms for month-2 CFSv2 streamflow forecasts without postprocessing at 1734 gauges averaged over the 14 basins. Perfect rank uniformity is indicated by the horizontal dashed lines.

seasons, it is slightly worse than the latter during the spring.

6. Discussion

The climate model-based seasonal hydroclimatic forecasts show visible skill enhancements against ESP, but they also have certain deficiencies. For the precipitation forecasts, although we know that models' skill decreases greatly after 1 month (Yuan et al. 2011), Figs. 2–4 show that they have negligible improvement against ESP over specific regions and seasons even for the month-1 forecasts, such as the R^2 over southeastern basins during summer (Figs. 2e,f). The current downscaling system used in the study is based on the monthly mean precipitation, where the ensemble members have different leads. For example, CFSv2 provides forecasts every 5 days with four forecasts on those days, so it is expected that the ensemble members with shorter leads have higher skill. Thus, assigning optimal weights to the 24 members seems appropriate. However, when we try to

use an optimization procedure to weigh them, it is quite difficult to obtain stable weights. The computed weights differ year by year through the cross-validation procedure, although the weighed ensemble can have significantly higher skill without cross validation. This indicates that there may well be an optimal combination among different ensemble members, but the 27-yr hindcasts are not long enough to get stable weights.

Another approach is to make use of the forecast information in the first 2 weeks. Within the Bayesian framework and perhaps common in many other statistical downscaling methods, the posterior distribution will be very similar to the climatology if the forecast skill is very low. Therefore, such downscaling methods can have added value only if a forecast has skill. During the first month, climate models usually have higher skill in the first 2 weeks than in weeks 3 and 4. Thus, developing independent likelihood functions for the first and second half of the first month might provide additional information. Furthermore, integrating the first 2-week forecasts from weather prediction models that have high

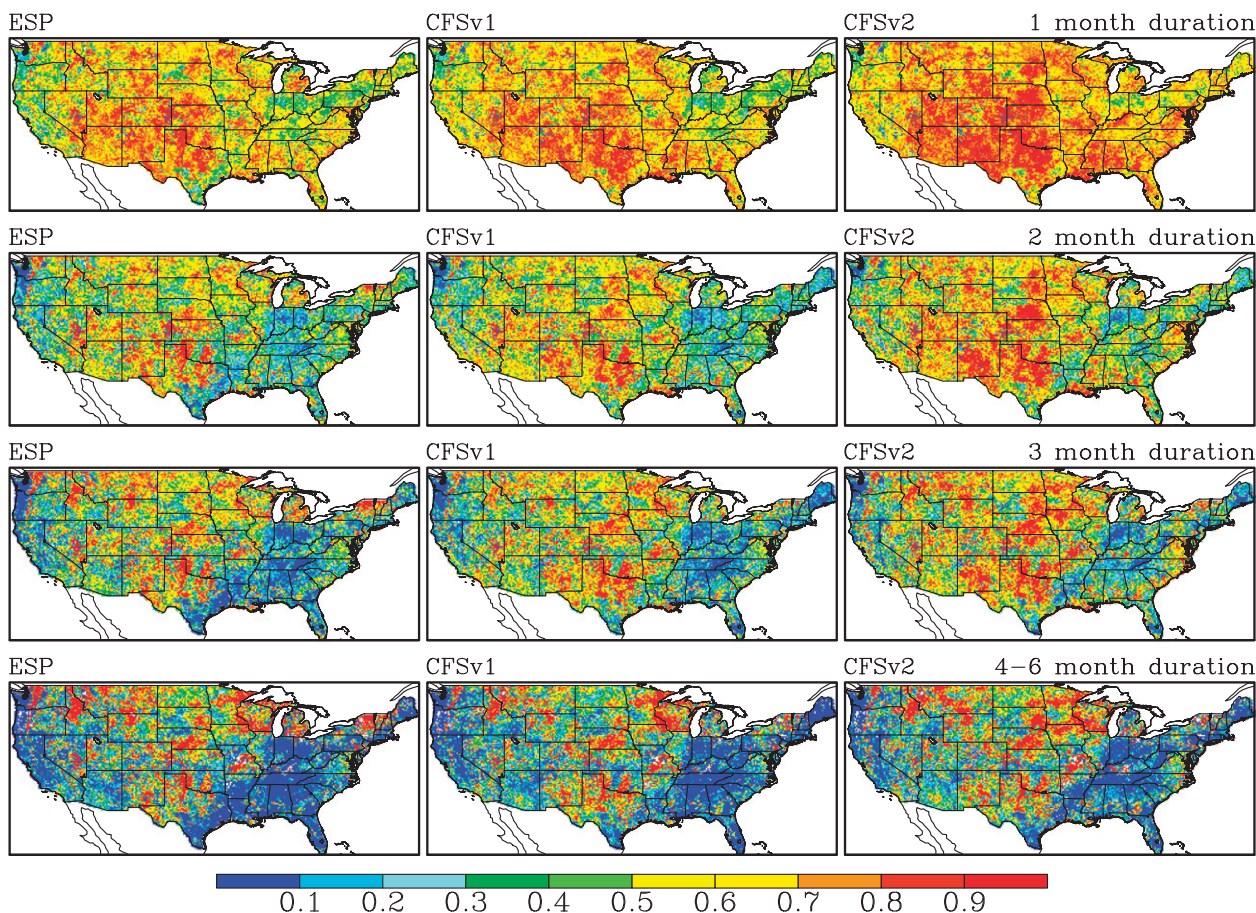


FIG. 9. Ratio of the ensemble mean forecasts of soil moisture drought frequency averaged among all forecasts in each calendar month at 0.5-month lead divided by those from offline simulation: (left) to (right) ESP, CFSv1, and CFSv2 and (top to bottom) 1 to 4–6 months duration. As an example, the frequency of 3-month-duration drought is counted by using all forecasts that have continuously dry conditions (<20%) in the first 3 months.

spatiotemporal resolutions and advanced data assimilation procedures with the current seasonal forecasting system will have a fundamental influence on seasonal hydroclimatic forecast. Ongoing research in this area, using the hindcasts from the National Oceanic and Atmospheric Administration's (NOAA's) global medium range ensemble forecast (GEFS; T. M. Hamill 2012, personal communication) by the authors will test the above hypothesis with the goal of integrating weather and seasonal climate forecasts to develop a seamless prediction system that can handle hydroclimatic forecasts at daily, monthly, and seasonal scales.

Besides developing long-term hindcast data and utilizing information from medium-range weather forecasts, creating ensemble forecasts based on multiple seasonal climate models is also an effective way to increase the skill for hydroclimatic predictions. In fact, many multimodel ensemble forecast systems already exist, such as European Operational Seasonal to

Interannual Prediction (EUROSIP), the World Meteorological Organization (WMO) Lead Centre for Long-Range Forecast Multimodel Ensemble, and the newly funded experimental National Multimodel Ensemble (NMME) system. However, many models are sharing similar atmospheric or oceanic components, and consequently they may produce similar seasonal forecast results. For example, by analyzing hindcasts from 12 seasonal forecast models, Yuan and Wood (2012b) found that a multimodel ensemble does not necessarily have higher predictive skill than a single model, while a proposed cluster ensemble has consistent improvement against individual models. Therefore, identifying the covariance structure of the multimodel system is quite necessary before applying them to the hydroclimatic forecast fields.

Improving the prediction of climate forcings such as precipitation is one important aspect for advancing the seasonal hydrologic forecast. Meanwhile, this paper

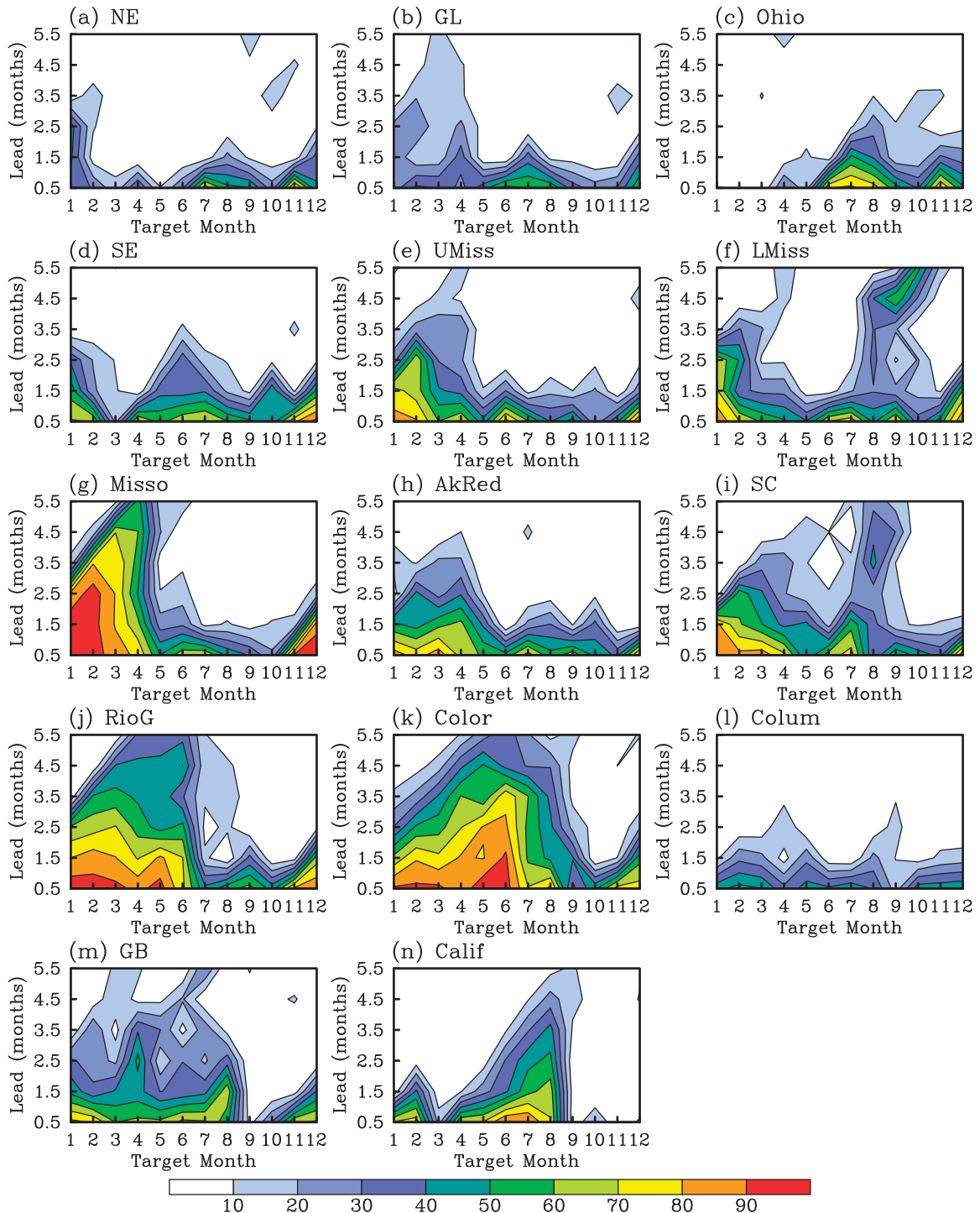


FIG. 10. The R^2 (%) of ESP ensemble mean forecasts of drought area calculated over the 14 basins as functions of target months and leads.

shows that developing an appropriate postprocessing procedure can also have significant contributions to seasonal forecast skill. The rationale of postprocessing is that the hydrologic models, though calibrated, may have

some uncorrected biases because of inadequate process representations of the nonlinear rainfall–runoff relationship. Such biases could be amplified by the river routing model either because it does not properly

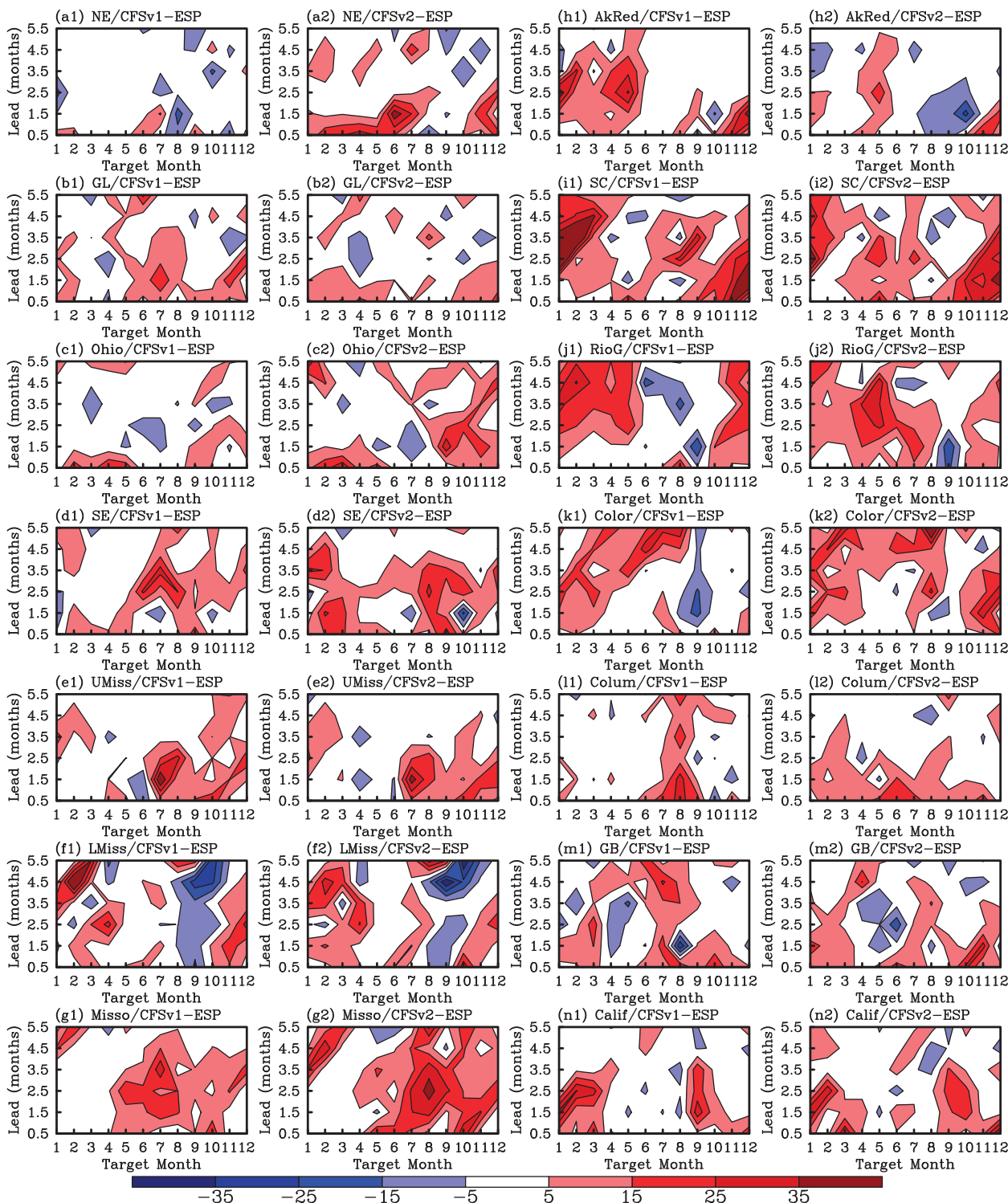


FIG. 11. Differences in R^2 (%) of drought area between climate model-based forecasts and ESP calculated over the 14 basins as functions of target months and leads: (a1)–(n1) CFSv1 – ESP and (a2)–(n2) CFSv2 – ESP.

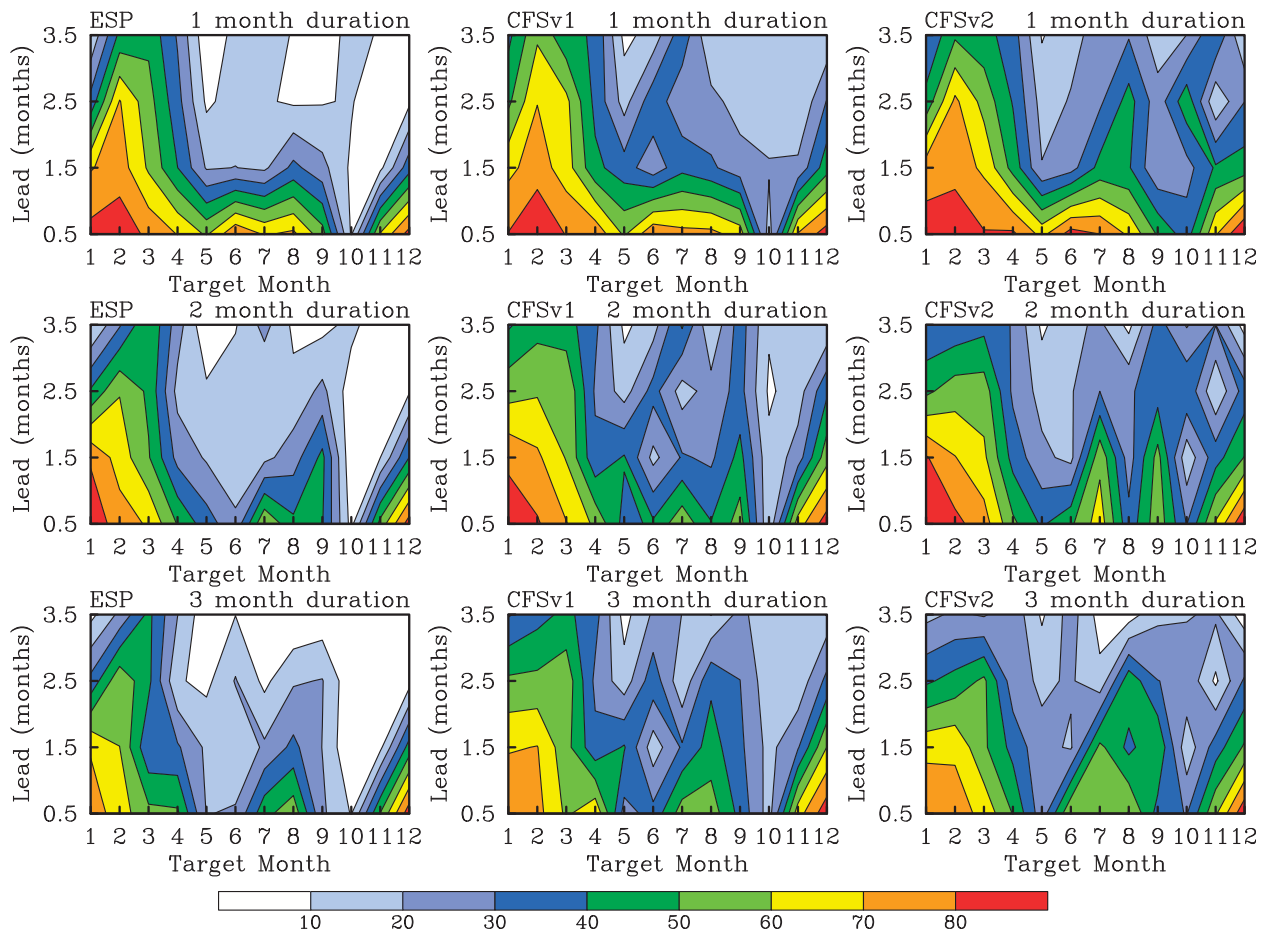


FIG. 12. The R^2 (%) of drought severity accumulated over CONUS for different durations as functions of target months and leads: (left) ESP, (middle) CFSv1, and (right) CFSv2 and (top to bottom) 1–3 months duration.

include river features like small dams or extractions or because it is uncalibrated. We calibrated the VIC land surface model, grid cell by grid cell, using observed runoff ratio data (Troy et al. 2008) but did not calibrate the routing model in this study. As seen in Fig. 7, the postprocessing procedure can correct those biases by matching the streamflow forecast distribution with the observation distribution based on statistical bias correction methods, which is also discussed in detail in Yuan and Wood (2012a). Such postprocessing procedures are also needed for the basins affected by human interventions such as reservoir management or irrigation that have not been considered in most of the land surface hydrologic and routing models. Additionally, Fig. 8 shows that the forecasted streamflow has underdispersion errors in ensemble prediction, although the climate forcings both from ESP and Bayesian downscaling are quite reliable. This indicates that there are underrepresented uncertainties from initial conditions and/or hydrologic models. Numerical experiments

by perturbing the initial conditions or model parameters are needed for further diagnosis.

7. Summary

In this study, 6-month seasonal hindcasts of monthly precipitation from the NCEP Climate Forecast System [both version 1 (CFSv1) and version 2 (CFSv2)] for each calendar month during 1982–2008 are downscaled to $1/8^\circ$ over CONUS using a Bayesian merging method. For the predictive skill R^2 of large basin-averaged ensemble mean precipitation at 0.5-month lead, the results from CFSv1 are generally higher than ESP, which is based on randomly selected historical observations. The exception is for some basins during the transition seasons, as discussed in the text (see section 3). CFSv2 has improved precipitation predictive skill over CFSv1 for many regions, especially during the winter and spring, and shows significantly higher skill than ESP during wet seasons over all studied large basins except the SE. In terms of

root-mean-square error skill score (SS_{RMSE}), both CFSv1 and CFSv2 are better than ESP out to 6 months, except for the 4.5-month lead forecasts over Ohio and California. The SS_{RMSE} conditional on Niño-3.4 SST anomalies indicates that not all basins over CONUS are positively affected by ENSO, but the forecasts conditional on ENSO do have significantly higher skill than the unconditional ones over the SE, Colorado basin, and California in the first 3–4 months. The ranked probability skill score (RPSS) analysis demonstrated that more than 50%–60% of the probabilistic forecasts from CFSv1 and v2 are more skillful than ESP over each large basin, and CFSv2 has moderate but consistent improvement against CFSv1 for the month-1 precipitation probabilistic forecasts.

To evaluate the performance for seasonal hydrologic prediction, the downscaled climate forcings, as well as those from ESP, are used to drive VIC model to produce 6-month, 20-ensemble hydrologic forecasts during 1982–2008, with initial conditions from a 62-yr (1949–2010) continuous offline simulation. Verification over 1734 USGS gauges indicates that CFSv2 reduces errors from ESP by 4%–7% for month-1 streamflow forecasts averaged over eastern basins, 10% over California, and less than 4% over the west. The error reductions are below 3% for the forecasts beyond 1 month, and CFSv1 is slightly worse than ESP over Ohio basin. For the month-1 probabilistic streamflow forecasts, 57%–76% of the CFSv1 forecasts and 60%–77% of the CFSv2 forecasts are more skillful than ESP. The climate model-based streamflow probabilistic forecasts have more advantages against ESP over western basins than those over the east, but actually all three approaches have very low skill over western basins as compared with climatological forecasts that are based on climatological streamflow observation distributions. The verification rank histograms show that all three approaches have underdispersion errors in the ensemble and overforecasting biases for some arid basins. Those errors and biases make ESP and climate model-based forecasts worse than climatological forecasts beyond 1 month over many basins. Therefore, we use the Bayesian postprocessing procedure proposed by Yuan and Wood (2012a) to correct the forecast. After postprocessing, all three approaches produce more skillful forecasts (>60%) out to 2 months, when compared against climatological forecasts.

The assessment for drought prediction shows that climate model-based forecasts produce more reasonable drought frequency estimates than ESP over central and eastern United States. Although the droughts with durations longer than three months are very rare over the eastern United States, where less than 30% of them can be predicted, all three approaches have plausible

performances in predicting drought frequency over the central United States out to 6 months. The predictive skill of the spatial extent in drought is evaluated for each large basin. ESP has higher drought area predictive skill for the central U.S. basins than for other regions and lower skill over humid basins than arid/semiarid basins. For the arid basins and for the forecasts initiated during dry seasons, initial soil moisture conditions have more impact on the drought area predictive skill. The spatio-temporal characteristics of R^2 for drought extent from CFSv1 and CFSv2 are similar to ESP, and CFSv2 is generally better than CFSv1 and ESP. Significant improvement of CFSv2 occur for the forecasts initiated during transition or dry seasons for humid basins and during wet or dry seasons for arid basins. The predictive skill of regional accumulated drought severity is analyzed for the CONUS region, and it is higher during winter because of strong initial soil moisture control and/or better precipitation predictions from the climate models. Unlike precipitation forecasts that have low skill in the summer, the predictive skill of drought severity is higher in the summer than in spring or fall. Climate models offer added value against ESP in forecasting drought severity, especially for long lead forecasts. CFSv2 is slightly worse than CFSv1 during spring but better than the latter during other seasons.

The CFSv2 shows some clear potential in forecasting precipitation, streamflow and, soil moisture drought based on our evaluation over CONUS. To further enhance seasonal hydroclimatic forecast skill, creating long-term hindcast data, utilizing information from medium-range weather forecasts, developing multi-model ensemble system with consideration of their covariance structure, and adopting appropriate post-processing procedures are expected.

Acknowledgments. The research was supported by the NOAA Climate Program Office through Grants NA17RJ2612 and NA10OAR4310246. We thank the anonymous reviewers for their comments. We acknowledge PICSciE/OIT at Princeton University for the supercomputing support.

REFERENCES

- Andreadis, K. M., E. A. Clark, A. W. Wood, A. F. Hamlet, and D. P. Lettenmaier, 2005: Twentieth-century drought in the conterminous United States. *J. Hydrometeor.*, **6**, 985–1001.
- 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.
- Crow, W. T., and E. F. Wood, 2003: The assimilation of remotely sensed soil brightness temperature imagery into a land surface

- model using ensemble Kalman filtering: A case study based on ESTAR measurements during SGP97. *Adv. Water Resour.*, **26**, 137–149.
- Dai, A., 2011: Drought under global warming: A review. *Wiley Interdiscip. Rev. Climate Change*, **2**, 45–65, doi:10.1002/wcc.81.
- , T. Qian, K. E. Trenberth, and J. D. Milliman, 2009: Changes in continental freshwater discharge from 1948 to 2004. *J. Climate*, **22**, 2773–2792.
- Duan, Q., and Coauthors, 2006: Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. *J. Hydrol.*, **320**, 3–17.
- Goddard, L., J. W. Hurrell, B. P. Kirtman, J. Murphy, T. Stockdale, and C. Vera, 2012: Two time scales for the price of one (almost). *Bull. Amer. Meteor. Soc.*, **93**, 621–629.
- Koster, R. D., S. P. P. Mahanama, B. Livneh, D. P. Lettenmaier, and R. H. Reichle, 2010: Skill in streamflow forecasts derived from large-scale estimates of soil moisture and snow. *Nat. Geosci.*, **3**, 613–616.
- Li, H., L. Luo, E. F. Wood, and J. Schaake, 2009: The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting. *J. Geophys. Res.*, **114**, D04114, doi:10.1029/2008JD010969.
- Liang, X., E. F. Wood, and D. P. Lettenmaier, 1996: Surface soil moisture parameterization of the VIC-2L model: Evaluation and modifications. *Global Planet. Change*, **13**, 195–206.
- Lohmann, D., and Coauthors, 2004: Streamflow and water balance intercomparisons of four land surface models in the North American Land Data Assimilation System project. *J. Geophys. Res.*, **109**, D07S91, doi:10.1029/2003JD003517.
- Luo, L., and E. F. Wood, 2007: Monitoring and predicting the 2007 U.S. drought. *Geophys. Res. Lett.*, **34**, L22702, doi:10.1029/2007GL031673.
- , and —, 2008: Use of Bayesian merging techniques in a multimodel seasonal hydrologic ensemble prediction system for the eastern United States. *J. Hydrometeorol.*, **9**, 866–884.
- , —, and M. Pan, 2007: Bayesian merging of multiple climate model forecasts for seasonal hydrological predictions. *J. Geophys. Res.*, **112**, D10102, doi:10.1029/2006JD007655.
- Maurer, E. P., A. W. Wood, J. C. Adam, D. P. Lettenmaier, and B. Nijssen, 2002: A long-term hydrologically based data set of land surface fluxes and states for the conterminous United States. *J. Climate*, **15**, 3237–3251.
- Mo, K. C., 2011: Drought onset and recovery over the United States. *J. Geophys. Res.*, **116**, D20106, doi:10.1029/2011JD016168.
- Pan, M., E. F. Wood, R. Wojcik, and M. McCabe, 2008: Estimation of regional terrestrial water cycle using multi-sensor remote sensing observations and data assimilation. *Remote Sens. Environ.*, **112**, 1282–1294.
- , A. K. Sahoo, T. J. Troy, R. K. Vinukollu, J. Sheffield, and E. F. Wood, 2012: Multisource estimation of long-term terrestrial water budget for major global river basins. *J. Climate*, **25**, 3191–3206.
- 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.
- Schaake, J., and Coauthors, 2007: Precipitation and temperature ensemble forecasts from single-value forecasts. *Hydrol. Earth Syst. Sci. Discuss.*, **4**, 655–717.
- Sheffield, J., and E. F. Wood, 2008: Projected changes in drought occurrence under future global warming from multi-model, multi-scenario, IPCC AR4 simulations. *Climate Dyn.*, **31**, 79–105.
- , G. Goteti, F. Wen, and E. F. Wood, 2004: A simulated soil moisture based drought analysis for the United States. *J. Geophys. Res.*, **109**, D24108, doi:10.1029/2004JD005182.
- Shukla, S., and D. P. Lettenmaier, 2011: Seasonal hydrologic prediction in the United States: Understanding the role of initial hydrologic conditions and seasonal climate forecast skill. *Hydrol. Earth Syst. Sci.*, **15**, 3529–3538.
- Troccoli, A., 2010: Seasonal climate forecasting. *Meteor. Appl.*, **17**, 251–268.
- Troy, T. J., E. F. Wood, and J. Sheffield, 2008: An efficient calibration method for continental-scale land surface modeling. *Water Resour. Res.*, **44**, W09411, doi:10.1029/2007WR006513.
- Twedt, T. M., J. Schaake, and E. L. Peck, 1977: National Weather Service extended streamflow prediction. *Proc. 45th Western Snow Conf.*, Albuquerque, NM, Colorado State University, 52–57.
- Wang, A., T. J. Bohn, S. P. Mahanama, R. D. Koster, and D. P. Lettenmaier, 2009: Multimodel ensemble reconstruction of drought over the continental United States. *J. Climate*, **22**, 2694–2712.
- Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. International Geophysics Series, Vol. 100, Academic Press, 676 pp.
- Wood, A. W., E. P. Mauer, A. Kumar, and D. P. Lettenmaier, 2002: Long-range experimental hydrologic forecasting for the eastern United States. *J. Geophys. Res.*, **107**, 4429, doi:10.1029/2001JD000659.
- , A. Kumar, and D. P. Lettenmaier, 2005: A retrospective assessment of National Centers for Environmental Prediction climate model-based ensemble hydrologic forecasting in the western United States. *J. Geophys. Res.*, **110**, D04105, doi:10.1029/2004JD004508.
- Xia, Y., and Coauthors, 2012: Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *J. Geophys. Res.*, **117**, D03109, doi:10.1029/2011JD016048.
- Yoon, J. H., K. Mo, and E. F. Wood, 2012: Dynamic-model-based seasonal prediction of meteorological drought over the contiguous United States. *J. Hydrometeorol.*, **13**, 463–481.
- Yuan, X., and E. F. Wood, 2012a: Downscaling precipitation or bias-correcting streamflow? Some implications for coupled general circulation model (CGCM)-based ensemble seasonal hydrologic forecast. *Water Resour. Res.*, **48**, W12519, doi:10.1029/2012WR012256.
- , and —, 2012b: On the clustering of climate models in ensemble seasonal forecasting. *Geophys. Res. Lett.*, **39**, L18701, doi:10.1029/2012GL052735.
- , —, L. Luo, and M. Pan, 2011: A first look at Climate Forecast System version 2 (CFSv2) for hydrological seasonal prediction. *Geophys. Res. Lett.*, **38**, L13402, doi:10.1029/2011GL047792.