

# Dynamic-Model-Based Seasonal Prediction of Meteorological Drought over the Contiguous United States

JIN-HO YOON

*Pacific Northwest National Laboratory, Richland, Washington*

KINGTSE MO

*NCEP/NWS/NOAA/Climate Prediction Center, Camp Springs, Maryland*

ERIC F. WOOD

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

(Manuscript received 18 April 2011, in final form 13 September 2011)

## ABSTRACT

A simple method was developed to forecast 3- and 6-month standardized precipitation indices (SPIs) for the prediction of meteorological drought over the contiguous United States based on precipitation seasonal forecasts from the NCEP Climate Forecast System (CFS). Before predicting SPI, the precipitation ( $P$ ) forecasts from the coarse-resolution CFS global model were bias corrected and downscaled to a regional grid of 50 km. The downscaled CFS  $P$  forecasts, out to 9 months, were appended to the  $P$  analyses to form an extended  $P$  dataset. The SPIs were calculated from this new time series. Five downscaling methods were tested: 1) bilinear interpolation; 2) a bias correction and spatial downscaling (BCSD) method based on the probability distribution functions; 3) a conditional probability estimation approach using the mean  $P$  ensemble forecasts developed by J. Schaake, 4) a Bayesian approach that bias corrects and downscales  $P$  using all ensemble forecast members, as developed by the Princeton University group; and 5) multimethod ensemble as the equally weighted mean of the BCSD, Schaake, and Bayesian forecasts. For initial conditions from April to May, statistical downscaling methods were compared with dynamic downscaling based on the NCEP regional spectral model and forecasts from a high-resolution CFS T382 model. The skill is regionally and seasonally dependent. Overall, the 6-month SPI is skillful out to 3–4 months. For the first 3-month lead times, forecast skill comes from the  $P$  analyses prior to the forecast time. After 3 months, the multimethod ensemble has small advantages, but forecast skill may be too low to be useful in practice.

## 1. Introduction

Drought and persistent wet spells are costly disasters in the United States. Losses are easily in the tens of billions (U.S. dollars) (e.g., Lott and Ross 2006). If drought can be predicted a few months in advance, then the losses may be reduced. A drought early warning system can help local and federal governments to relocate their resources for mitigating the upcoming drought (Hayes et al. 2004). There are many aspects of drought (Keyantash and Dracup 2002). Meteorological drought is defined by precipitation ( $P$ ) deficits. Agricultural drought is caused

by soil moisture (SM) deficits and hydrological drought is reflective of runoff or streamflow deficits. In the recent decade, advances have been made in hydrologic prediction of streamflow and soil moisture by using climate forecasts at seasonal lead times (Wood and Lettenmaier 2006; Luo et al. 2007). Based on the National Centers for Environmental Prediction (NCEP) Climate Forecast System (CFS) seasonal forecasts (Saha et al. 2006), the Environmental Modeling Center (EMC) produces seasonal forecasts of soil moisture and streamflow over the United States based on the hydrologic prediction system developed by the Princeton group (Luo et al. 2007; Luo and Wood 2008). The streamflow and soil moisture forecasts are produced by land surface models or hydrologic models driven by downscaled  $P$  and temperatures from global model forecasts. The prediction of

---

*Corresponding author address:* Dr. Jinho Yoon, Pacific Northwest National Laboratory, P.O. Box 999, Richland, WA 99352.  
E-mail: jin-ho.yoon@pnnl.gov

meteorological drought measured by the  $P$  deficit is easier because it does not require any land or hydrologic model. In this paper, we will extend the hydrologic prediction to cover meteorological drought based on the CFS forecasts.

The meteorological drought is often measured by the standardized precipitation index (SPI; Hayes et al. 1999). The advantage is that it is a function of  $P$  alone and can be applied to station data. It covers a spectrum of time scales from 1 to 24 months or beyond. The SPI has been recently identified as a universal meteorological drought index for effective drought monitoring and climate risk management by many international experts and representatives from national meteorological and hydrological services around the world [the World Meteorological Organization (WMO); [http://www.wmo.int/pages/mediacentre/press\\_releases/pr\\_872\\_en.html](http://www.wmo.int/pages/mediacentre/press_releases/pr_872_en.html)]. There are circumstances that SPI may not be the best index to classify drought because SPI is based only on precipitation. Drought conditions over the western interior states in spring are influenced by snow packs in the previous winter and snowmelt in spring. SPI does not take that into consideration. In the deserts or places where evaporation is an important moisture source, the SPI is not a good index to measure drought.

The impacts of drought are regional. Information is often needed for small river basins. Therefore, drought assessment and prediction need to be delivered at the small regional level. Because the CFS model has the horizontal resolution of T62, which is approximately 250 km, it is too coarse to resolve detailed terrain-related features. To provide regional drought information, spatial downscaling is needed. In general, there are two categories of downscaling techniques (Wilby and Wigley 1997). One is statistical downscaling based on the historical relationship between global model hindcasts and high-resolution observation, and the other one is dynamic downscaling method using limited-area, high-resolution models. To perform statistical downscaling, model errors are estimated by comparing the hindcasts from the same global model at the same lead times as forecasts with high-resolution corresponding observations. The implicit assumption is that the relationships between hindcasts and observations are still valid for the forecasts. For dynamic downscaling, a high-resolution regional climate model (RCM) is nested inside the global climate forecast model to produce finer-resolution forecast products. The coarse-resolution climate forecasts provide both the boundary and initial conditions for the RCM. Therefore, the errors in the global model can influence the regional model prediction. Of course, the regional model can also have its own errors. Therefore, the calibration with observed conditions is necessary.

In this study, five statistical downscaling methods were tested. The simplest method is to utilize the bilinear interpolation (BI) to downscale the climate forecasts from a coarse grid to a finer grid and correct only the mean biases defined as the differences between the hindcasts' climatology and the observed climatology. The second method is a bias correction and spatial downscaling (BCSD) method. This has been used extensively in the hydrologic prediction community (Wood et al. 2002, 2005). The third method is a parametric approach. It updates the probability distribution on  $P$  by deriving the conditional distribution based on the correlation between the hindcast ensemble means and the corresponding observations (Schaake et al. 2007; Wood and Schaake 2008). It is referred to here as the Schaake method. The fourth method is the Bayesian merging technique using all ensemble members (Luo et al. 2007; Luo and Wood 2008; Coelho et al. 2004). Because different downscaling methods produce skillful forecasts at different spatial locations, the multimethod ensemble forecasts defined as the equally weighted mean of downscaled forecasts from the above methods may have higher skill than a single method. There is another downscaling method based on the constructed analogs (Hidalgo et al. 2007; Maurer and Hidalgo 2008). The hindcasts' datasets here are not long enough to test this method.

The statistical downscaling methods will be compared with dynamic downscaling based on the NCEP regional spectral model (RSM; Juang and Kanamitsu 1994; Juang et al. 1997) for April–May initial conditions. They are also compared directly with the global high-resolution forecasts by the CFS T382 model. The T382 experiments were originally developed for dynamic hurricane seasonal forecasts. Therefore, only hindcasts initialized in April–May are available. To forecast SPI, the downscaled monthly mean forecasts are appended to the  $P$  analyses to form a longer time series. The SPI indices are computed from this extended time series. Although some previous effort was made to make SPI prediction using log-linear models (Moreira et al. 2008) over Europe, or using regression method based on sea surface temperature (SST) modes and outgoing longwave radiation (Barros and Bowden 2008), no attempts have been made with dynamic seasonal forecast outputs at regional scales.

The objectives of this paper are 1) to compare the forecast skill from five different methods of statistical downscaling of monthly mean  $P$  forecasts from the CFS, 2) to forecast SPI from one season to 9 months based on downscaled  $P$  forecasts, and 3) to compare the SPI forecasts based on statistical downscaling with these from the RSM and directly from the T382 model. The area of this study is the contiguous United States. The datasets used and the five statistical downscaling

methods are briefly reviewed in section 2. The skill of downscaled seasonal  $P$  forecasts from different methods is presented and cross validated in section 3. The SPI prediction and its skill are discussed in section 4 and they are compared directly from forecasts from the NCEP RSM and high-resolution T382 CFS in section 5. Conclusions are given in section 6.

## 2. Data and procedures

### a. Datasets

#### 1) NCEP CFS FORECASTS

The CFS forecast system is a fully coupled ocean–atmosphere dynamical seasonal prediction system that became in operation in August 2004. The atmospheric model is the 2003 version of the Global Forecast System (Moorthi et al. 2001). The resolution is T62 with 64 sigma levels in the vertical. The ocean component is the Geophysical Fluid Dynamics Laboratory (GFDL) Modular Ocean Model version 3 (MOM3; Pacanowski and Griffies 1998). The land model is a two-layer land surface model (Mahrt and Pan 1984). The initial conditions are taken from the National Centers for Environmental Prediction (NCEP)/Department of Energy Atmospheric Model Intercomparison Project (AMIP) R2 (Kanamitsu et al. 2002). The oceanic initial conditions are taken from the NCEP Global Ocean Data Assimilation (GODAS; Behringer et al. 1998). Detailed documentation and forecast evaluation can be found in Saha et al. (2006).

Monthly mean time series of  $P$  were obtained from the CFS hindcast archive. Hindcast monthly means have 15 members initialized in each calendar month of the year for the period 1981–2008. The 15 members are clustered in three 5-member groups with initial conditions centered at the 1st, 11th, and 21st of each month. For example, members of the October–November hindcasts for 1988 consist of hindcasts initialized from 9–13 October, 19–23 October, 29–30 October, and 1–3 November 1988 conditions. This set of hindcasts is labeled as the November 1988 hindcasts. For 3 November forecast, the  $P$  analyses for 1 and 2 November are ignored. The 1-month lead hindcasts are verified against the  $P$  analyses for November 1988. The 9-month lead hindcasts are verified against  $P$  for July 1989. The ensemble mean is the equally weighted mean of all 15 members.

#### 2) NCEP RSM FORECASTS

The NCEP RSM (Juang et al. 1997) is a regional spectral model with a similar dynamic core and the same physics package as the NCEP Global Forecast System (GFS). The lateral boundary conditions are provided by the CFS forecasts every 6 h. The spatial resolution of the

NCEP RSM is 50 km. There are five members in the ensemble. They were initialized from 29 April to 3 May each year. Each member of hindcasts lasts for 7 months.

#### 3) HIGH-RESOLUTION CFS T382 MAY FORECASTS

The coupled global climate model (CGCM) was the operational version of the 2007 GFS model with T382L64 resolution coupled with the GFDL MOM3 ocean model. High-resolution hindcasts were initialized from 19–23 April each year from 1981 to 2008. There are five members in the ensemble and hindcasts last for 7 months. The ensemble mean is the equally weighted mean of five members. Although this set of forecasts has been proven to produce reasonable skill in seasonal climate hurricane forecast, rainfall anomalies over the North American monsoon region have not been well captured (Long et al. 2008). In our analysis,  $P$  and SPI forecast from the NCEP RSM and T382 are reduced to  $0.5^\circ \times 0.5^\circ$  in latitude and longitude for comparison.

#### 4) $P$ ANALYSES

The  $P$  dataset used is the CPC unified  $P$  analyses ( $P_{\text{analy}}$ ; Xie et al. 2010). The optimal interpolation (OI) method is used and it is terrain adjusted. The spatial resolution is 50 km. It covers the period from 1950 to present. For verification, the monthly mean  $P$  anomaly is defined as the departure from the monthly mean climatology from 1981 to 2008 for that month. The same dataset is also used for verification. The  $P$  analyses can only be viewed as a proxy for observations. The quality of the  $P$  analyses depends on the input station data and the analysis method. At a given location, the difference between the  $P$  analysis and rain gauge report can be large.

### b. Cross validation

The downscaled  $P$  and SPI were cross validated. The period is from 1981 to 2008. Hindcasts with initial conditions from October to November, January to February, April to May, and July to August were evaluated. The cross validation was done for each lead time and each set of initial conditions separately. The procedures are outlined below.

For the given lead and given initial conditions, we removed one year from the hindcasts. That year is denoted as the target year. The rest of the hindcast years are denoted as training years. To remove one year is sufficient because the maximum of 1-month lag correlation of  $P_{\text{analy}}$  over the contiguous United States is about 0.2, which is not statistically significant. The downscaling parameters and cumulative distribution functions (CDFs) were determined from data in the training period. Then, we applied downscaling with the same parameters and CDF to

the forecasts of the target year. The downscaled field should be verified against the corresponding  $P_{\text{analy}}$  for the target year. Two measures of forecast skill are used in this study. One is the anomaly correlation (AC) between hindcasts and  $P_{\text{analy}}$ , and the other is the root-mean-square error (RMSE). Both measure the difference between hindcasts and  $P_{\text{analy}}$ . The AC depends on the climatology used and is not sensitive to the magnitudes of hindcast anomalies. Therefore, the RMSE is also calculated.

### 3. $P$ downscaling and error correction

Let us consider the set of monthly mean ensemble hindcasts initialized at month  $M$ . For target year  $T_y$  and lead time  $\tau$ , the ensemble mean hindcast is labeled as  $P(T_y, \tau, M)$ , which has a coarse resolution of 250 km. We briefly review four downscaling and error correction methods below.

#### a. BI

The  $P$  anomaly from hindcasts,  $P_a(T_y, \tau, M)$  for the target year  $T_y$ , lead time  $\tau$ , and month  $M$  are defined as the departure from the model climatology averaged over the ensemble means of hindcasts for the same lead time  $\tau$  from the training period. The hindcasts at year  $T_y$  were not included in the climatology. We bilinearly interpolated  $P_a(T_y, \tau, M)$  to a 50-km grid. It should be verified against the  $P_{\text{analy}}$  anomaly at month  $\tau - 1 + M$  and year  $T_y$ .

Figure 1 shows an example of hindcast anomaly at  $\tau = 1$  for November 1988. November 1988 was a La Niña month. The  $P_{\text{analy}}$  anomaly (Fig. 1a) shows negative anomalies over the southern Great Plains and positive rainfall anomalies over the Pacific Northwest and the Ohio Valley. These are typical responses to a La Niña event. It also shows negative anomalies extending southward from Tennessee to Alabama. These were features for that year.

The BI downscaling  $P$  hindcast (Fig. 1b) captures the  $P$  signal well, but the magnitudes of anomalies are weaker in comparison with  $P_{\text{analy}}$ . In this method, anomalies are not adjusted for terrain. As noticed by Gutowski et al. (2004), a low-resolution model tends to produce  $P$  with low amplitudes. Obviously,  $P$  hindcasts downscaled with BI are too weak to be useful and cannot capture terrain-related features over the Pacific Northwest, but the BI method can be used as a building block and reference for other methods. One potential shortcoming of correcting a model's climatology is that the corrected total precipitation can be negative when the model climatology is much wetter than the  $P_{\text{analy}}$  climatology. It is noted here that dynamic forecasts are better than forecasts based on climatology or persistence (Fig. 1h), which does not capture the pattern of  $P$  anomalies.

#### b. BCSD

Before applying BCSD, all  $P$  hindcasts were bilinearly interpolated to the 50-km grid. For the month  $M$  and lead time  $\tau$ , to modify  $P(T_y, \tau, M)$  for the target year  $T_y$ , the CDF was computed at each grid point using all ensemble members with the lead time  $\tau$  based on CFS hindcasts in the training period, which does not include the target year. That is labeled as CDF (hindcast). Similarly, the CDF(analy) was obtained from the corresponding  $P_{\text{analy}}$ . At each grid point, the percentile of  $P(T_y, \tau, M)$  for the target year  $T_y$  was determined according to the CDF(hindcast). Given the same percentile, we obtained  $P$  by transferring backward to the physical space from the CDF(analy). Because the model has 420 data points for a grid point, but the  $P_{\text{analy}}$  only has 28 years of data, the  $P$  percentile sometimes falls above or below the range of the empirical Weibull percentiles. Then, an extreme value type III Weibull function is used with a minimum bound of zero for the low-precipitation case and an extreme value type I Gumbel distribution function is used for the high-precipitation case. The detailed procedures and discussions are given by Wood et al. (2002).

In comparison to BI, the BCSD method corrects both the mean and the standard deviation of the ensemble hindcasts in the normal space. The hindcasts have higher amplitudes and they are terrain adjusted because the  $P_{\text{analy}}$  were terrain adjusted (Xie et al. 2010). For November 1988, the downscaled hindcast (Fig. 1c) has higher amplitudes than Fig. 1b. Note that rainfall features over the Cascade Mountains of the Pacific Northwest are improved. Anomalies over the southern Great Plains are stronger. Because of its simplicity and robustness in correcting terrain-related features, this method has been used in many hydrologic forecast applications (Wood et al. 2002).

#### c. Schaake method

The BCSD method does not take into account the forecast skill of the hindcasts. On the contrary, both the Schaake method and the Bayesian method do. Both methods derive the conditional distribution function based on information provided by the hindcasts (Wood and Schaake 2008). The major difference between the two methods is that the Schaake method only considers the ensemble means. The Bayesian method uses information provided by all ensemble members.

The first part of the procedure is the same as the BCSD. The  $P$  ensemble hindcast for the target year and hindcasts from the training years were bilinearly interpolated to the 50-km grid. We transformed the  $P$  hindcasts in the training period to the normal probability space using the CDF

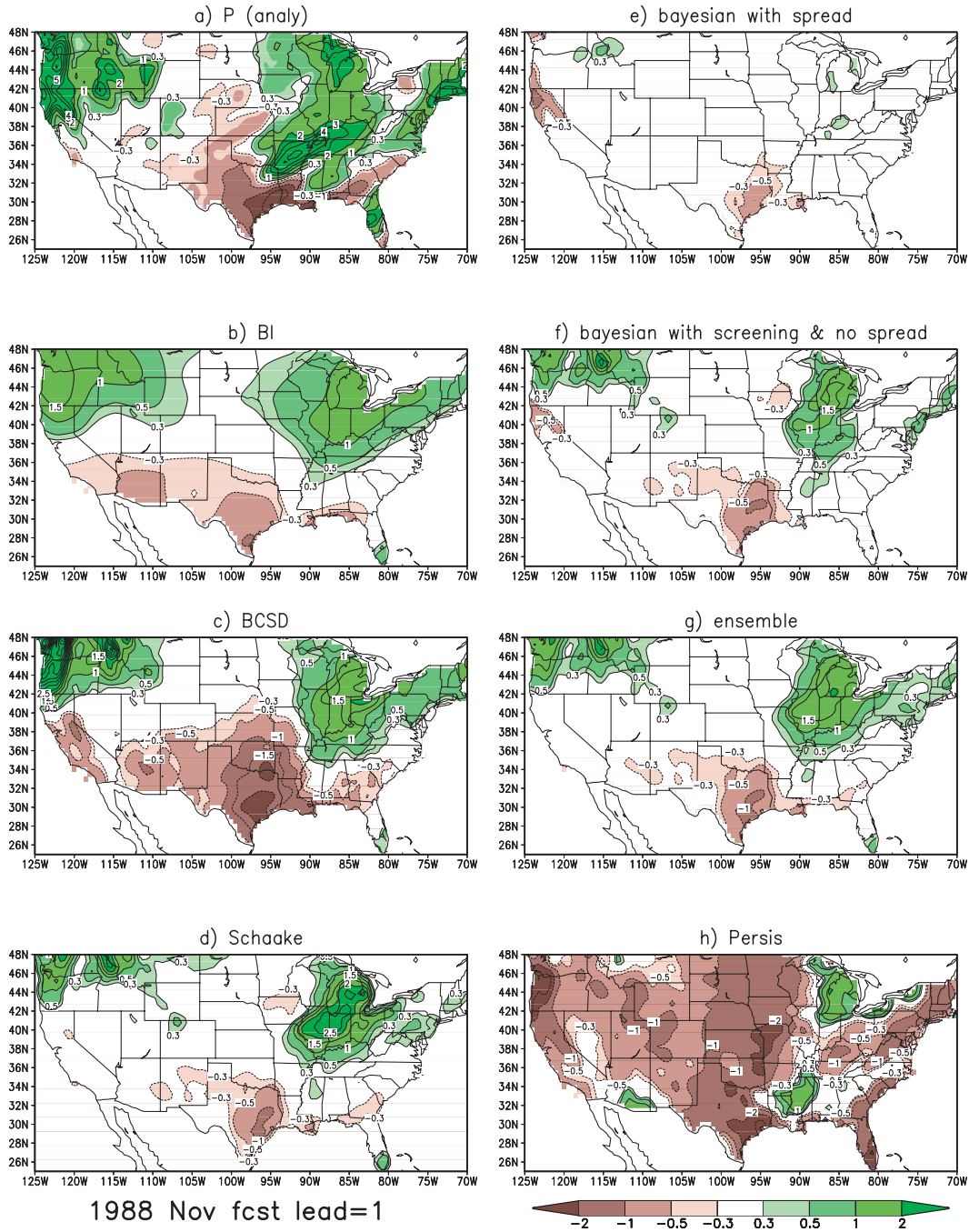


FIG. 1. (a) November 1988 precipitation anomaly from the  $P$  analyses. Contour interval is  $1 \text{ mm day}^{-1}$ . Contours  $-0.5$ – $0.2$ ,  $0.2$ , and  $0.5 \text{ mm day}^{-1}$  are added. November  $P$  forecast at 1-month lead produced by downscaling from the T62 CFS forecasts with (b) BI, (c) BCSD, (d) Schaake method, (e) Bayesian method including spread, (f) Bayesian method without spread and with screening, and (g) the multimethod ensemble mean of (c),(d), and (f), and (h) persistence. Contour interval in (b)–(g) is  $0.5 \text{ mm day}^{-1}$ . Contours  $-0.2$  and  $0.2 \text{ mm day}^{-1}$  are added. Zero contours are omitted. (h) Contour interval as in (a).

(hindcast). The same CDF(hindcast) was also used to transfer  $P(T_y, \tau, M)$  for the target year  $T_y$  to the normal space. The corresponding  $P_{\text{analy}}$  were transferred to the normal space based on CDF(analy).

In the normal space, the normal distribution with the mean  $\mu$  and variance  $\sigma^2$  is denoted as  $N(\mu, \sigma^2)$ . Let  $\bar{y}$  be the time series of the ensemble mean  $P$  hindcasts, which obey the distribution function  $N(\mu_{\bar{y}}, \sigma_{\bar{y}}^2)$ . Let  $x$  be the

corresponding  $P_{\text{analy}}$  time series, which obey the normal distribution function  $N(\mu_x, \sigma_x^2)$ . Then, the posterior conditional distribution function,  $N(\mu_p, \sigma_p^2 | \bar{y})$  for  $x$  given  $\bar{y}$  and its parameters, are determined as follows:

$$\mu_p = \mu_x + \rho(x, \bar{y}) \left( \frac{\sigma_x}{\sigma_{\bar{y}}} \right) (\bar{y}_t - \mu_{\bar{y}}) \quad (1)$$

and

$$\sigma_p^2 = (1 - \rho(x, \bar{y})^2) \sigma_x^2, \quad (2)$$

where  $\bar{y}_t$  is the ensemble mean hindcast for the target year  $T_y$  in the normal space, and  $\rho(x, \bar{y})$  is the correlation between the ensemble hindcast mean  $\bar{y}$  in the training period and the corresponding  $x$ . In doing so, this method takes into consideration the hindcast skill. If the hindcasts are unskillful, then  $\rho(x, \bar{y})$  is small and  $\mu_p$  becomes the climatology  $\mu_x$ . The percentiles of  $P$  for the target year  $T_y$  were computed based on the conditional distribution function and transferred backward to the physical space.

For November 1988, anomalies from the Schaake method (Fig. 1d) have the similar pattern as anomalies downscaled with BCSD, but the positive anomalies over the Ohio River valley are stronger and negative anomalies over the Southwest are weaker.

#### d. Bayesian merging

The methodology is described in detail in Luo et al. (2007). Bayes's theorem updates the probability distribution on  $P$  based on the hindcasts in the training period. Similar to the Shaake method, both hindcasts and the corresponding  $P_{\text{analy}}$  were transformed to the normal space based on the CDF(hindcasts) and CDF(analy), respectively. The ensemble members of  $P$  for the target year  $T_y$  were also transferred to the normal space based on CDF(hindcasts).

Let  $y$  be the hindcasts in the training period and  $x$  be the corresponding  $P_{\text{analy}}$  time series. The prior distribution of  $P$  labeled as  $\text{PD}(P)$  is the unconditional distribution of  $P$ . It obeys the normal distribution  $N(\mu_x, \sigma_x^2)$ , where  $\mu_x$  and  $\sigma_x^2$  are mean and variance determined from  $x$ . It is updated by information from hindcasts  $y$  as shown in Eq. (3):

$$\text{PD}(P|y) = \frac{\text{PD}(P)\text{PD}(y|P)}{\text{PD}(y)}, \quad (3)$$

where  $\text{PD}(P)$  is the prior distribution of precipitation  $P$  (i.e., the distribution without updated information),  $\text{PD}(y)$  is the unconditional (marginal) probability of hindcast time series  $y$  in the training period,  $\text{PD}(P|y)$  is

the posterior distribution of  $P$  conditioned on the hindcast  $y$  (i.e., distribution after updating based on hindcast  $y$ ), and  $\text{PD}(y|P)$  is the likelihood function, which is modeled using a linear regression between hindcast  $y$  and the corresponding  $P_{\text{analy}}$ .

The likelihood function  $\text{PD}(y|P)$  is determined by least squares fit of the ensemble mean hindcasts in the training period to the corresponding  $P_{\text{analy}}$  (Coelho et al. 2004; Luo et al. 2007):

$$\bar{y} = \alpha + \beta x + \varepsilon, \quad (4)$$

where  $\bar{y}$  is the ensemble mean of the hindcasts,  $x$  is the corresponding  $P_{\text{analy}}$ , and  $\varepsilon$  is the residual or the error of fitting. Then  $\text{PD}(y|P)$  can be written as a normal distribution with mean  $\alpha + \beta x$  and variance  $\varphi_\varepsilon$ , which is the variance of the residual  $\varepsilon$  from the linear regression

$$\text{PD}(y|P) \sim N(\alpha + \beta x, \varphi_\varepsilon). \quad (5)$$

Equation (4) does not consider the forecast spread (Coelho et al. 2004). If we assume that the skill of forecasts is inversely related to the spread among ensemble members, then  $\text{PD}(y|P)$  can be written as

$$\text{PD}(y|P) \sim N(\alpha + \beta x, \varphi_\varepsilon + \varphi_y), \quad (6)$$

where  $\varphi_y$  is the spread among the member of forecasts (Luo et al. 2007).

Then, from Bayes's theory, the posterior distribution is given by  $\text{PD}(P|y) \sim N(\mu_p, \sigma_p^2)$  with mean  $\mu_p$  and variance  $\sigma_p^2$ , which can be obtained as follows:

$$\frac{1}{\sigma_p^2} = \frac{1}{\sigma_x^2} + \frac{\beta^2}{\varphi_\varepsilon + \varphi_y}, \quad (7)$$

and

$$\frac{\mu_p}{\sigma_p^2} = \frac{\mu_x}{\sigma_x^2} + \frac{\beta^2}{\varphi_\varepsilon + \varphi_y} \left( \frac{\bar{y} - \alpha}{\beta} \right). \quad (8)$$

For the multiensemble forecasts, the Bayesian merge will adjust the distribution function by weighting the skill of the hindcasts from different models and the spread of forecasts accordingly (Luo et al. 2007). It also takes into consideration the errors of the linear fit of Eq. (4). For the unskillful forecasts, both the residual  $\varphi_\varepsilon$  and the spread  $\varphi_y$  are large, and the distribution function will be close to the prior distribution, which is determined from the  $P_{\text{analy}}$  climatology.

The spread among members of CFS hindcasts is very large because the CFS hindcasts have low skill, especially in midlatitudes over a small domain (Luo and Wood 2006).

Another reason is that initial conditions for members in the CFS hindcast ensemble are far apart—up to 30 days; thus, they may be in different weather regimes. For example, the November 1988 ensemble hindcast has the oldest member from the initial conditions on 9 October 1988 to the youngest member initialized on 3 November 1988. They are about 27 days apart. Different atmospheric and oceanic conditions can have potential influence on the hindcasts. If the spread is used, then the  $P$  anomalies after the Bayesian merging are very weak. For example, the November 1988 forecast for 1-month lead (Fig. 1e) shows a similar pattern as BI (Fig. 1b), but the magnitudes are very weak. Therefore, the spread is not used in this paper.

Instead of using the spread, we eliminate members in the ensemble that are very different from the ensemble mean before applying the Bayesian method. We computed the ensemble mean by averaging over all 15 members in the ensemble. Then, we correlated each member in the ensemble with the ensemble mean. We then selected a subset of members where the correlation is above a threshold, and updated the ensemble mean with members in the subset. The threshold correlation is 0.3, which is not large. The intention is to eliminate only those members that locate far away from the ensemble mean in the phase space. The subset usually contains 5–14 members. The screening was done for hindcasts before the Bayesian merging. Without spread and with the screening, the downscaled  $P$  hindcasts are comparable in amplitudes with the hindcasts downscaled with other methods. For the November 1988 case (Fig. 1f), amplitudes of anomalies are higher than those with spread (Fig. 1d). If we do not screen the forecasts, the skill is much lower because of low magnitudes of anomalies. In this paper, results are given for the Bayesian method without spread but with screening.

#### e. Multimethod ensemble

All three downscaling methods have advantages and disadvantages. They also show skill in different areas during different seasons. Therefore, we can form a multimethod ensemble as the equally weighted mean of hindcasts downscaled by the BCSD (Fig. 1c), Schaake (Fig. 1d), and Bayesian (Fig. 1f) methods. Figure 1g shows the example of the November 1988 hindcast with 1-month lead. It is terrain adjusted and has comparable amplitudes as hindcasts downscaled by the Bayesian procedure.

Figure 2 shows the correlation between downscaled  $P$  hindcast with the 1-month lead and the corresponding  $P_{\text{analy}}$  for hindcasts initialized from October–November and April–May conditions, respectively. If we assume that one year is  $1^\circ$  of freedom, then the correlation needs to be greater than 0.36 (0.31) to be statistically significant

at the 5% (10%) level. The forecast skill is overall low and it is regionally and seasonally dependent. For winter forecasts initialized in October–November, they are more skillful over the Southeast, New Mexico, and the north-central. November is the rainy season over the Pacific Northwest. Both the Bayesian and the BCSD show some skill there. For hindcasts initialized in April–May, the Schaake method shows large negative correlations over Kansas and eastern Colorado. That is due to the negative correlation between the  $P_{\text{analy}}$  and hindcasts. The multimethod ensemble is similar to the BCSD. There are no large negative correlations as with the Schaake method. Overall, the simple BCSD method does remarkably well. This implies that the systematic errors for the CFS forecasts are mainly terrain related.

## 4. The SPI prediction

### Method

The computation of the SPI based on McKee et al. (1993, 1995) is briefly reviewed by using SPI6 as an example. A detailed description of SPI, along with a computational algorithm, was obtained from the National Drought Mitigation Center (available online at <http://droughtmonitor.unl.edu>). The first step was to construct 6-month running  $P$  time series  $P_6$ , where  $P_6(t)$  at time  $t$  is  $P$  averaged from the time  $t - 5$  to  $t$ . The  $P_6$  time series was transformed to the normal probability space. Then, SPI6( $t$ ) was determined according to the normal distribution of the transformed dataset. The SPI at different time scales, like 3-month SPI (SPI3), can be calculated the same way by forming 3-month running means. They represent meteorological droughts on different time scales.

The method to forecast SPI from month  $M$  and year  $T_y$ , out to the lead-time  $\tau$  is outlined in Fig. 3. We explain in detail by using the November 1988 case as an example:

- 1) Obtain the monthly mean  $P_{\text{analy}}$  climatology for the base period  $T_y - 30$  to  $T_y$ . For November 1988 forecasts, the base period is from 1957 to 1987. Notice that there is no information used after the target month  $M$  and year  $T_y$ . One can easily extend the base period to a longer period. The only limitation is the record length of historical  $P_{\text{analy}}$  data.
- 2) Add the corresponding monthly mean  $P_{\text{analy}}$  climatology from the base period to the downscaled and error-corrected forecast  $P$  anomalies for lead month  $\tau = 1-9$ . For November 1988 forecasts, the first lead month is November, so the November climatology is added to the downscaled  $P$  anomaly. The last lead month is July, so the July climatology is added to the downscaled  $P$  anomaly forecast.

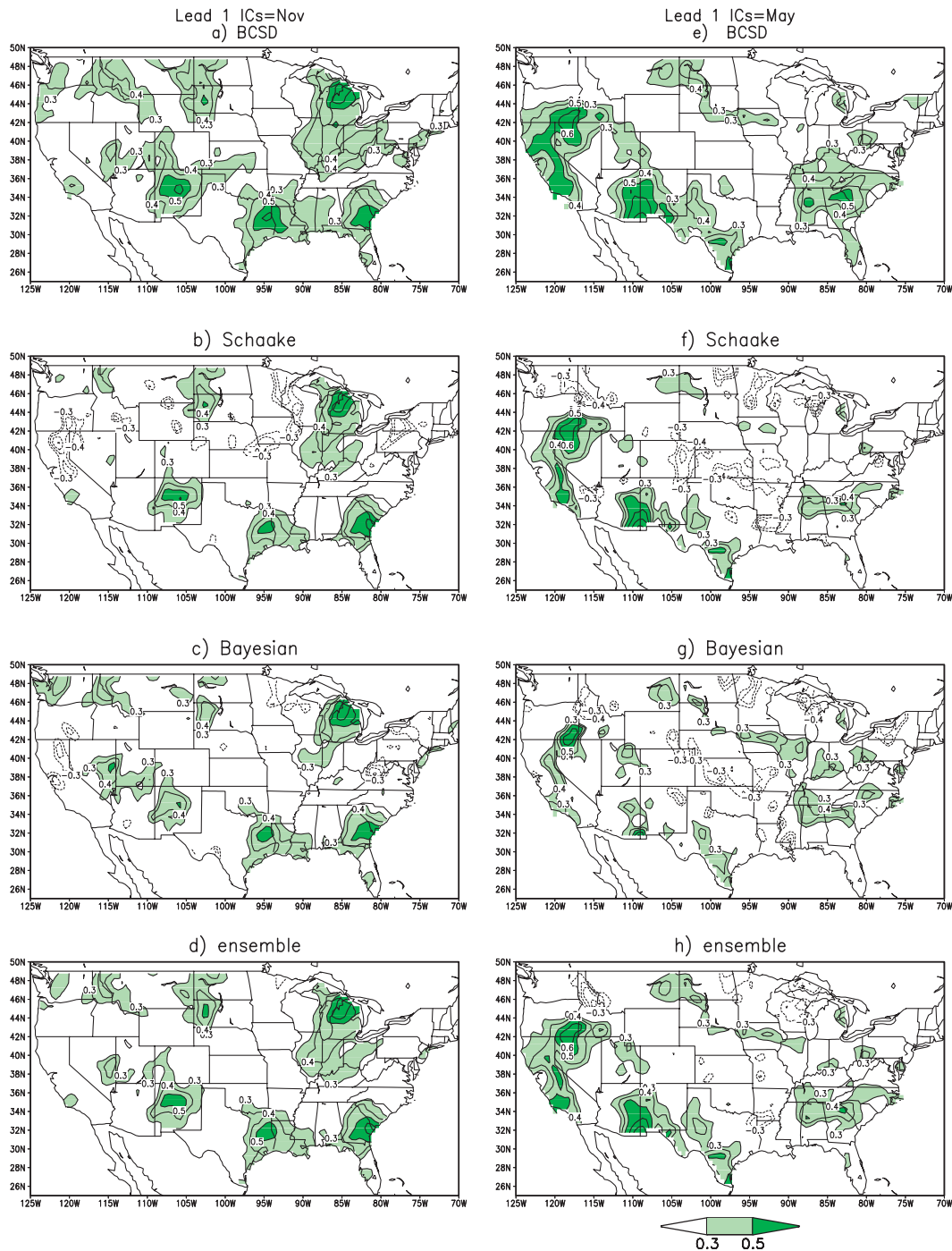


FIG. 2. Correlation between the  $P$  hindcasts with 1-month lead initialized in October–November and the verifying  $P_{\text{analy}}$  for (a)  $P$  hindcasts produced by the BCS D method, (b) Schaake method, (c) Bayesian method, (d) multi-method ensemble, and (e)–(h) as in (a)–(d), but for hindcasts initialized in April–May. Contour interval is 0.1. Values less than 0.3 are omitted.

3) The forecast time series is then appended to the  $P_{\text{analy}}$  time series to form the extended  $P$  time series, which covers the period from year  $T_y - 30$  month  $M$  to year  $T_y$  month  $M - 1$ . For November 1988, the new time

series consists of the  $P_{\text{analy}}$  time series from November 1958 to October 1988 for the first 360 months and the forecast time series for month 361 to month 369. The length of the combined  $P$  time series is 369 months.



## Forecast Procedure

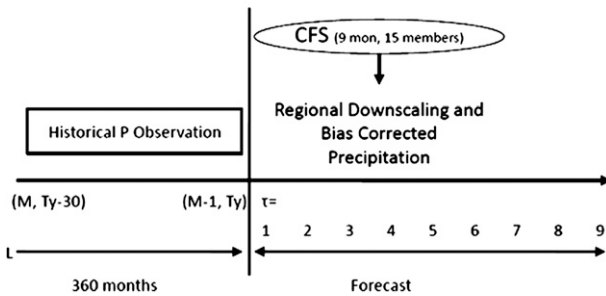


FIG. 3. A schematic diagram of SPI forecast procedure.

#### 4) Compute SPI from this new extended time series.

The important point is that  $P_{\text{analy}}$  after year  $T_y$  is not used because the information will not be available in the actual forecast situation. By definition, the 1-month lead SPI6 has  $P$  mean, which contains 5 months of  $P_{\text{analy}}$  and a 1-month forecast. Therefore, for the first few months' lead, the SPI6 forecasts are expected to be skillful. This can be considered as an analogy to the hydrologic prediction of soil moisture. They are known to be heavily influenced by initial conditions (Li et al. 2009).

Figure 4 shows the anomaly correlation (AC) and root-mean-square error (RMSE) of SPI6 hindcasts averaged over the contiguous United States as a function of the lead time for hindcasts initialized in four different seasons. The skill is seasonally dependent. When SPI6 at a given time is below (above) the threshold  $-0.8$  ( $0.8$ ), it is considered as an extreme dry (wet) event. Therefore, a seasonal forecast for extreme events is skillful if RMSE is below  $0.8$ . The corresponding AC is roughly  $0.6$ . If we use this criterion, then SPI6 forecasts are on average skillful up to the lead time of 4 months. The winter forecasts are more skillful than in summer in general. There is no statistically significant difference in skill for forecasts downscaled with different methods for the first 4 months when forecasts are skillful. After 4 months, the forecasts become unskillful, so differences in skill among different downscaling methods do not matter much. Relatively speaking, the multimethod ensemble mean of three methods has higher skill. There is a steep drop in forecast skill after 5 months for SPI6. In computing SPI6, the treated  $P$  forecasts are appended to  $P_{\text{analy}}$  time series to form a new extended  $P$  time series (Fig. 3). For the lead time longer than 5 months, this extended time series contains only  $P$  forecasts and it does not contain  $P_{\text{analy}}$ . The  $P$  forecasts for the lead time longer than 2 months are unskillful, which causes the SPI6 skill to drop.

To show the forecast skill for SPI6 from  $P$  downscaled with different methods in regional details, the RMSE for forecasts initialized in October–November is given in Fig. 5. This set of forecasts is chosen because they have the highest skill in comparison with forecasts initialized in other months (Fig. 4). Skills are comparable for these three sets. They all show that forecasts are more skillful over the central United States. Over the central United States, winter is not a rainy season, so low RMSE is expected. For winter, the rainy season starts over the Pacific Northwest and moves southward to California as the season progresses. The forecast skill is overall lower over the West Coast. After 3 months, forecasts over the West Coast and the Southeast become less skillful.

Forecast skill from the Bayesian method is seasonally dependent. Figure 6 shows the RMSE between SPI6 hindcasts and the corresponding verifying SPI6 calculated from the  $P_{\text{analy}}$  time series for different lead times and different initial seasons from 1981 to 2008. The forecasts verified in winter months are more skillful and the spring forecasts initialized in April–May have the least skill. The errors often appear in the rainy season. For example, the rainy season for the West is from November to March. Forecasts initialized in October–November show large errors located over the West Coast. The summer rainy season over the Great Plains is from May to September, while the West coast is overall dry. The forecast errors are largest over the Great Plains. Errors are smaller over the West Coast in summer. The overall area where the forecasts have the most skill is the Southeast, where rainfall occurs during all seasons.

The first few months of the SPI6 forecasts are skillful because the extended time series of  $P$  contains historical  $P_{\text{analy}}$ , as shown in Fig. 3. For SPI on shorter time scales, the forecasts deteriorate faster because the extended  $P$  time series contains fewer months of  $P_{\text{analy}}$  and more of forecasts. Figure 7 shows RSME for SPI3. Again, forecast skill is shown to be seasonally dependent. The most (least) skill forecast is found when the forecast is initialized in November–December (April–May). Overall, the 1- and 2-month lead forecasts are skillful for SPI3.

Figure 8 gives the Hovmuller diagram (time–latitude plot) for SPI6 forecasts at the 2- and 4-month lead time forecasts initialized in April–May and October–November and averaged over the latitude band ( $32^{\circ}$ – $36^{\circ}$ N), respectively. The SPI6 forecasts are based on the multimodel ensemble downscaled  $P$ . The SPI6 forecasts are shown together with the verifying SPI6 from the corresponding  $P_{\text{analy}}$  (colored). For lead month 2, forecasts capture most dry and wet events with slightly lower amplitudes. For example, 2-month lead forecasts verified for June (Fig. 8a) capture dry conditions from 1986 to 1989 and over 2004–08 from  $80^{\circ}$  to  $90^{\circ}$ W. It also

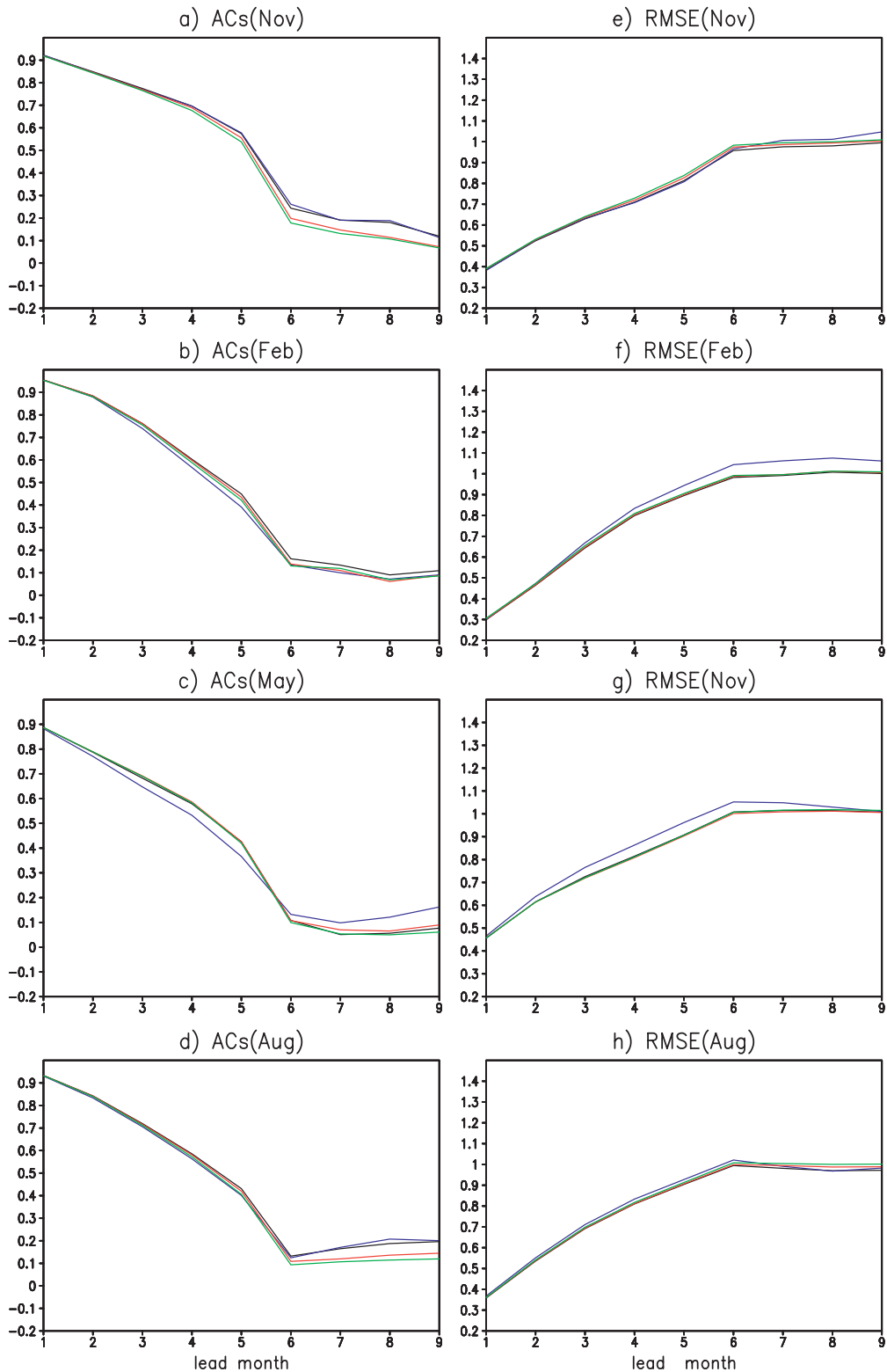


FIG. 4. Anomaly correlation of SPI6 as a function of the lead time for initial conditions in (a) October–November, (b) January–February, (c) April–May, and (d) July–August. SPI6 is derived from  $P$  hindcasts produced by downscaling with the BCSd (blue), Schaake (red), Bayesian (green), and multimethod ensemble (black). (e)–(h) As in (a)–(d), but for RMSE.

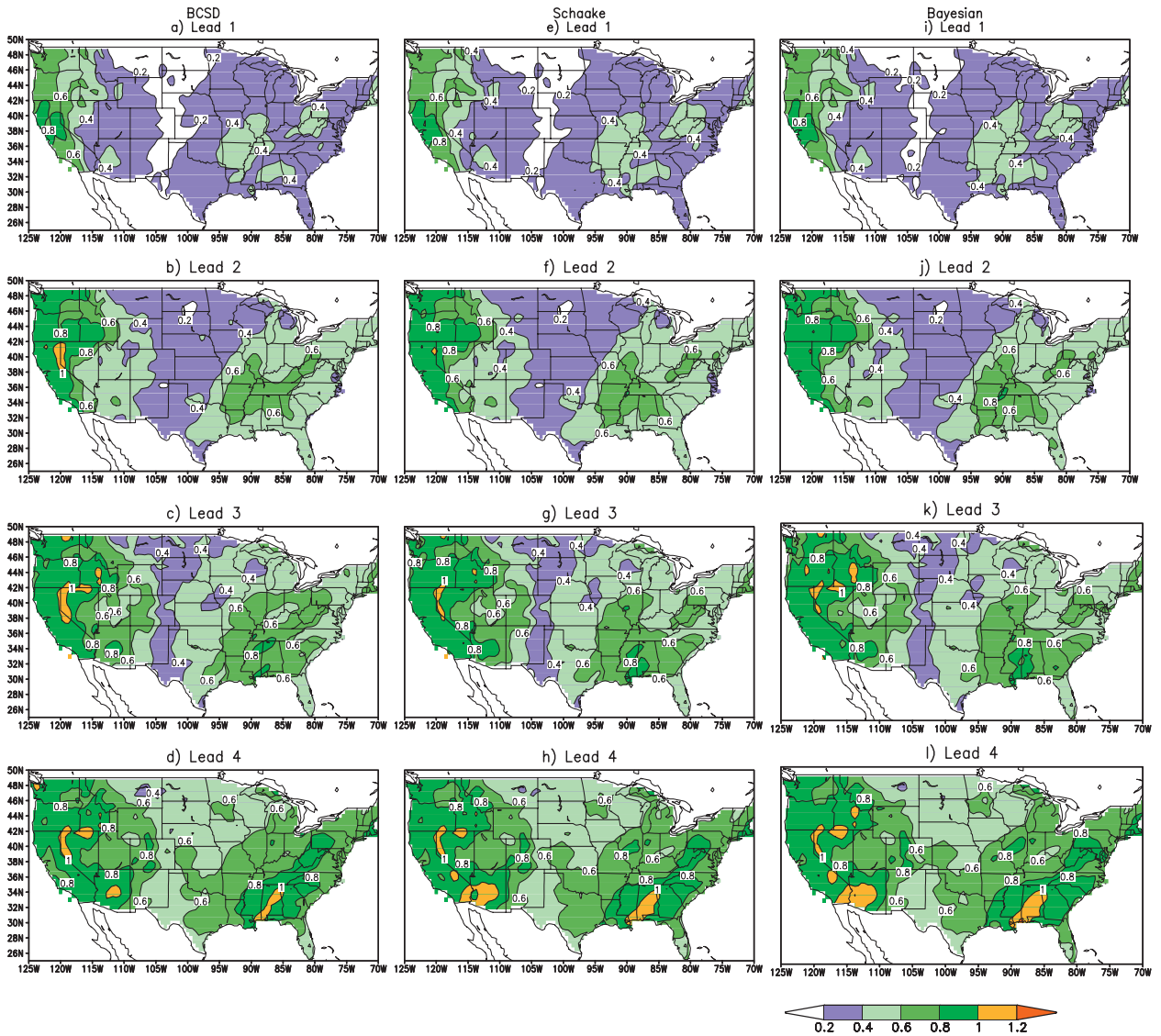


FIG. 5. RMSE between the SPI6 forecast initialized from October to November and the verifying SPI6 derived from the  $P_{\text{analy}}$  with  $P$  produced by BCSD for the lead time of (a) 1, (b) 2, (c) 3, and (d) 4 months. Values are indicated by the color bar. (e)–(h) As in (a)–(d), but with  $P$  produced by the Schaake method, and (i)–(l) as in (a)–(d), but for  $P$  produced by the Bayesian method.

captures the wet spells in 1990 over the southern plains and the 1992 wet events over the Southwest. The 4-month lead forecasts are still able to indicate drought and wet spells, but intensity is lower than in  $P_{\text{analy}}$ . Similar assessments are also true for forecasts initialized in other months.

**5. Forecasts from the NCEP RSM and the CFS T382 model**

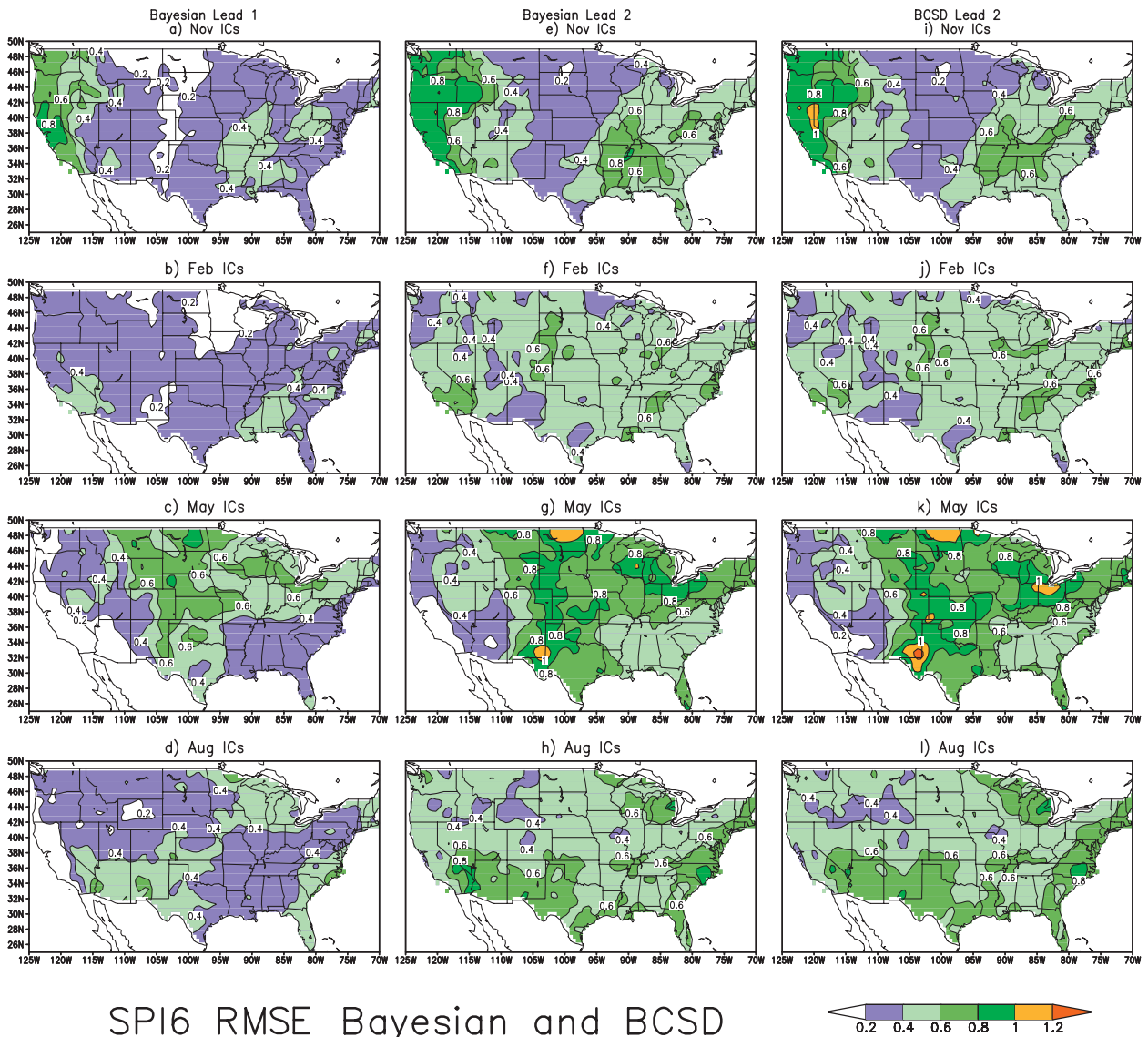
In this section, we compare hindcasts from statistical downscaling with dynamic downscaling with the NCEP RSM and hindcasts from a T382 model directly. Only

hindcasts initialized in April–May are available from the high-resolution T382 experiments.

*a. NCEP RSM*

The 50-km RSM was nested in the CFS hindcasts to produce downscaled precipitation. The boundary conditions from the CFS were given every 6 h. The RSM has the same physics package and the similar dynamical core as the CFS. The initial conditions were taken from the CFS hindcasts initialized from the period 28 April–3 May each year in this study.

For May  $P$  forecasts, the correlation between the  $P$  anomalies and the corresponding  $P_{\text{analy}}$  indicates that



## SPI6 RMSE Bayesian and BCSD

FIG. 6. RMSE between the 1-month lead SPI6 forecasts initialized in (a) October–November, (b) January–February, (c) April–May, and (d) July–August, and the verifying SPI6 derived from the  $P$  analysis. The SPI6 forecasts were derived from the  $P$  forecasts produced by downscaling with Bayesian method. Contour interval is given by the color bar. (e)–(h) As in (a)–(d), but for 2-month lead, and (i)–(l) as in (e)–(h), but for  $P$  forecasts downsampled using BCSD.

the RSM downsampled hindcasts (Fig. 9a) are overall more skillful than the 15-member ensemble downsampled with BCSD (Fig. 9b) except in the Southeast. It is noted that the initial conditions contribute to the forecast skill for the first lead month because the RSM members are the youngest members in the ensemble initialized at the end of April and early May. If the RSM downsampled forecasts are compared with the same five youngest members in the ensemble downsampled with BCSD, then skill is more comparable (Fig. 9c). For lead times longer than 1 month, when the influence of initial conditions diminishes, the differences between the dynamic downsampled and the

statistical downsampled hindcasts are small. The differences in skill for the SPI6 forecasts are also small. The overall RMSE shows that the RSM forecasts (dark crosses) have comparable skill with the same 5-member ensemble downsampled with BCSD (blue line) for the first month, but the 15-member ensemble downsampled by BCSD is slightly more skillful after the first month (green line).

If the only difference between the regional model and the global model is resolution, then downsampled  $P$  does not always have an advantage in comparison with simple statistical downscaling. The RSM does improve

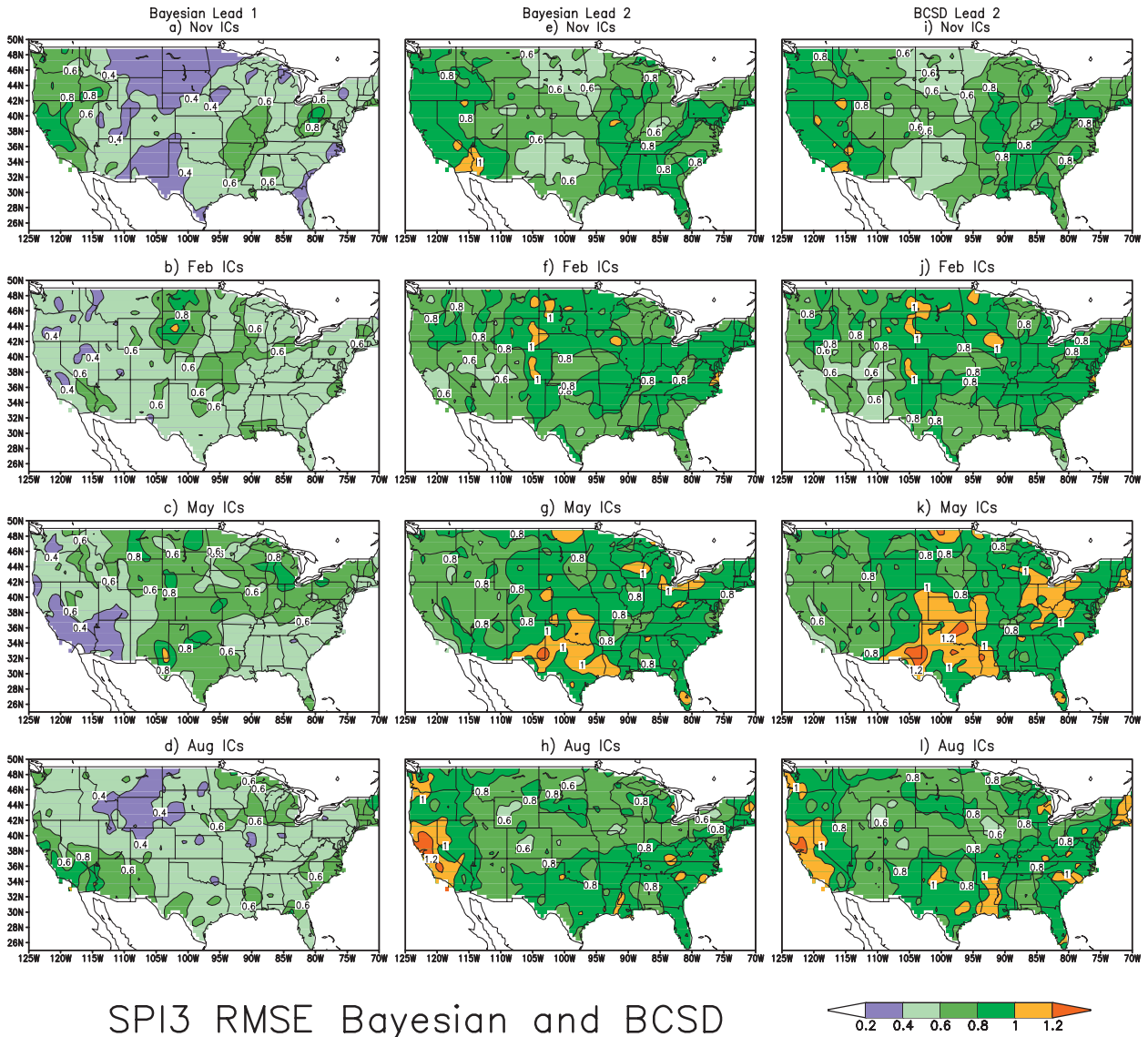


FIG. 7. As in Fig. 6, but for SPI3 forecasts.

climatology because it can resolve terrain-related features (Kim et al. 2002). However, it does not improve interannual variability of anomalies enough to make a difference. Our case is limited to the summer forecasts with April–May initial conditions. During summer, convective processes are important to forecast rainfall over the central United States, while tropical storms will bring rainfall over the East Coast and the Southeast. The dynamic downscaling has been demonstrated to improve skill over the Pacific Northwest (Leung et al. 2003) during its rainy winter season because regional models can more realistically capture the directional dependencies of the wintertime flow during El Niño–Southern Oscillation (ENSO) events and their

interactions with the complex topography of the western United States.

*b. CFS T382 model*

For the CFS T382 forecasts, there are five members initialized from 19 to 23 April each year for the period 1982–2008. The overall skill is slightly worse than the 15-member ensemble CFS forecasts downscaled with BCSD (Figs. 10a,b). However, it is comparable in skill with the 5-member ensemble with the same initial conditions (Fig. 10c). This indicates again for the first month that the initial conditions are important. The interesting point is that the forecasts from T382 are

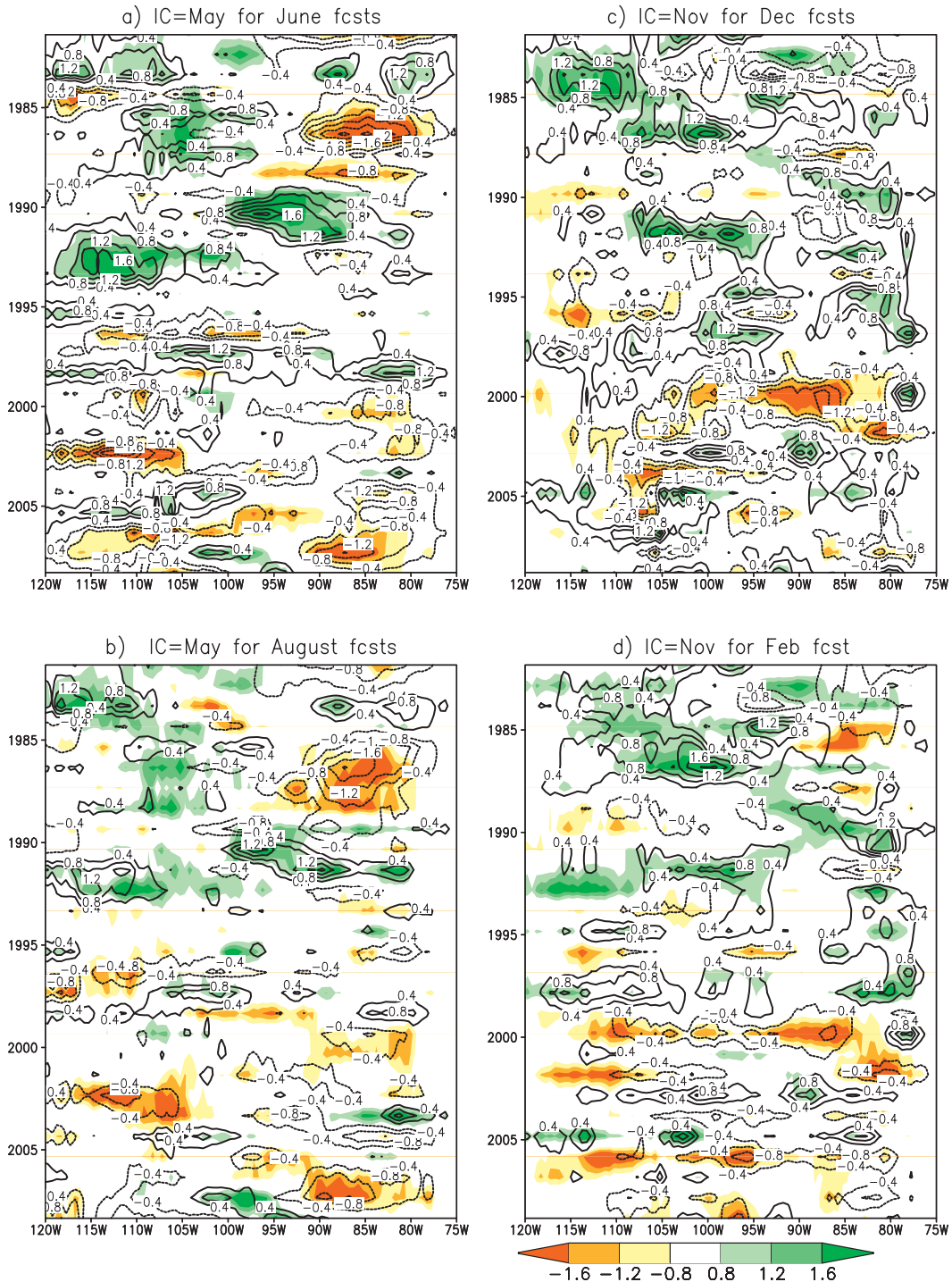


FIG. 8. (a) Longitude–time plots of SPI6 for 2-month lead time from the multimethod ensemble mean  $P$  forecasts initialized in April–May averaged from  $32^{\circ}$  to  $36^{\circ}\text{N}$  (contoured) and the verifying SPI6 derived from the  $P$  analyses (colored). The verifying SPI6 values are given by the color bar. The forecasts are contoured every 0.4. Zero contours are omitted. (b) As in (a), but for 4-month lead SPI6 forecasts; (c) as in (a) but for forecasts initialized in October–November; and (d) as in (c), but for 4-month lead SPI6 forecasts.

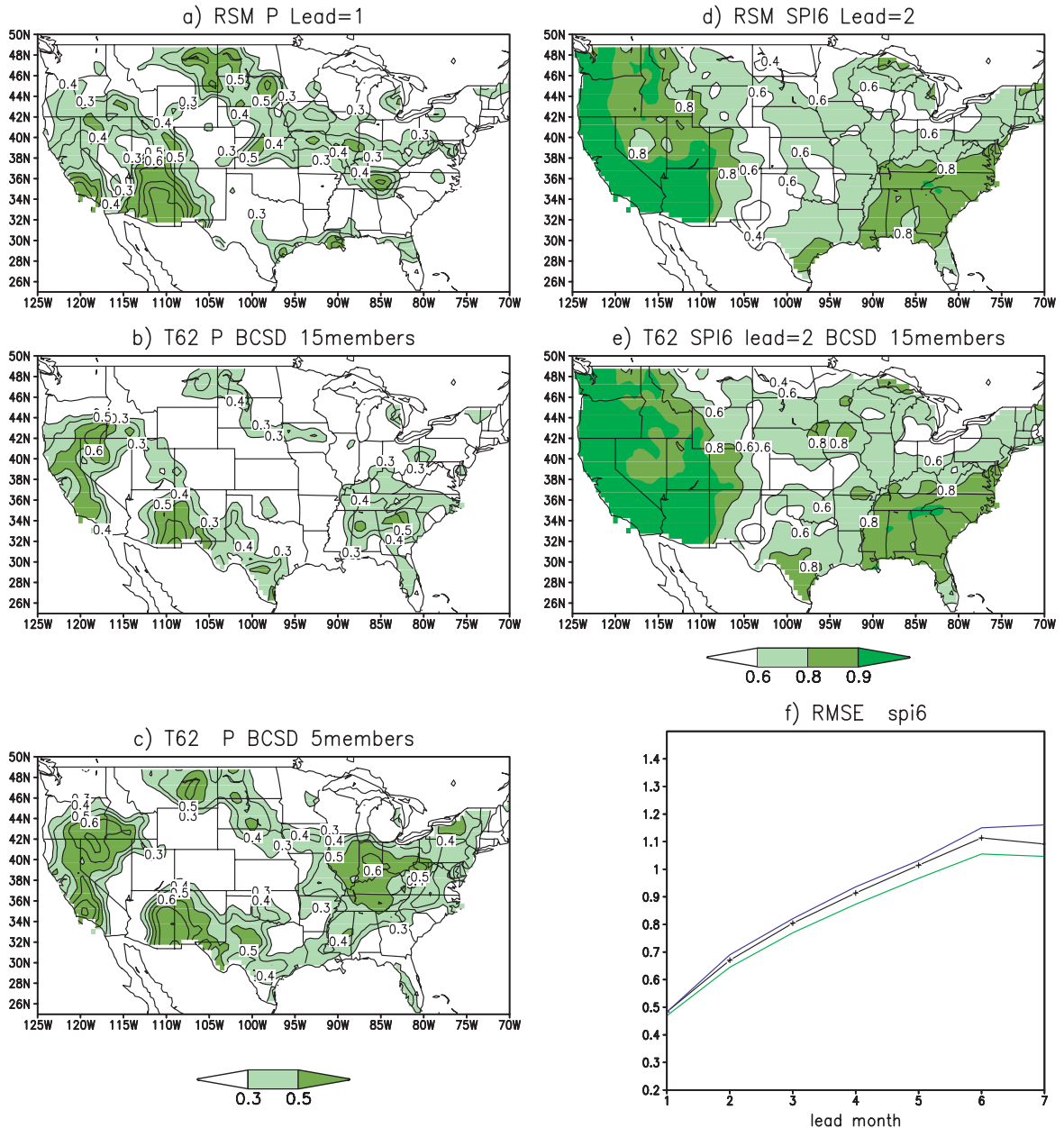


FIG. 9. (a) Correlation between 1-month lead  $P$  forecasts from the NCEP RSM 5-member ensemble and the corresponding  $P_{\text{analy}}$  for forecasts initialized in 28 Apr–3 May each year. Contour interval is 0.1. Values less than 0.3 are omitted. (b) As in (a), but for the T62 CFS  $P$  forecasts from the 15-member ensemble downsampled based on BCSD; (c) as in (b), but for the 5 members initialized at 29–30 Apr and 1–3 May; (d) as in (a), but for 2-month lead SPI6 derived from the RSM forecasts; (e) as in (d), but for the SPI6 forecasts derived from (b); and (f) RMSE for SPI6 forecasts averaged over the contiguous United States for 1982–2007. SPI6 were derived from  $P$  from the RSM (black crosses), the 15-member ensemble downsampled T62 CFS forecasts initialized from April to May using the BCSD method (green), and from the 5-member T62 forecasts initialized from late April and early May using the BCSD method (blue).

worse than the RSM (Fig. 10a). For SPI6, the T382 forecasts are comparable in skill with the 5-member ensemble T62 forecasts with the same initial conditions after downscaling with BCSD. However, they are not as skillful as the 15-member CFS forecasts after BCSD downscaling after the first month.

Although the CFS can capture sea surface temperature anomalies (SSTAs) of ENSO in the tropical Pacific Ocean, it has limited skill in  $P$  forecasts over the midlatitude regions (Saha et al. 2006). The CFS model can capture the impact of ENSO on  $P$  only when the initial conditions already contain the ENSO signal. This experimental

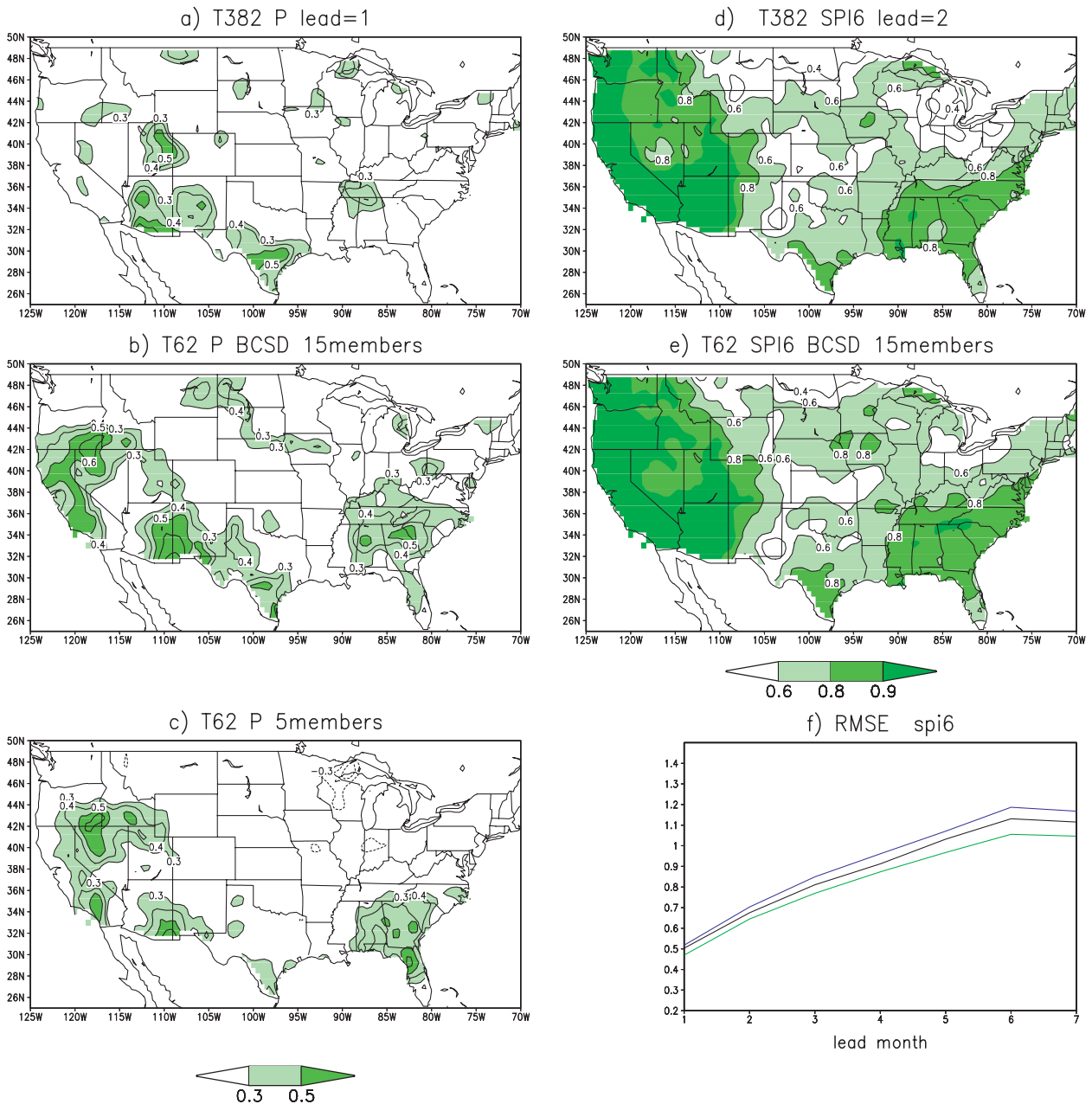


FIG. 10. (a) Correlation between  $P$  1-month lead forecasts from the high-resolution T382 model with a 5-member ensemble initialized from 19 to 23 Apr and the corresponding  $P_{\text{analy}}$ . Contour interval is 0.1. Values less than 0.3 are omitted. (b) As in (a), but for the T62 CFS  $P$  forecasts from 15-member ensemble downscaled based on BCSD; (c) as in (b), but for five members initialized from 19 to 23 April; (d) as in (a), but for 2-month lead SPI6 derived from the T382 forecasts; (e) as in (d), but for the SPI6 forecasts derived from (b); and (f) RMSE for SPI6 forecasts averaged over the contiguous United States for 1982–2007. SPI6 were derived from  $P$  from the T382 (black), the 15-member ensemble downscaled T62 CFS forecasts using the BCSD method (green), and from the 5-member T62 forecasts initialized from 19–23 April downscaled with the BCSD method (blue).

forecast of CFS T382 version is able to produce reasonable forecast of tropical storm activity over the major basins. However, it still has room for improvement in predicting summer rainfall over the contiguous United States.

## 6. Conclusions

The standardized precipitation index (SPI) has been used routinely to classify meteorological drought by national and international forecast centers and authorities.



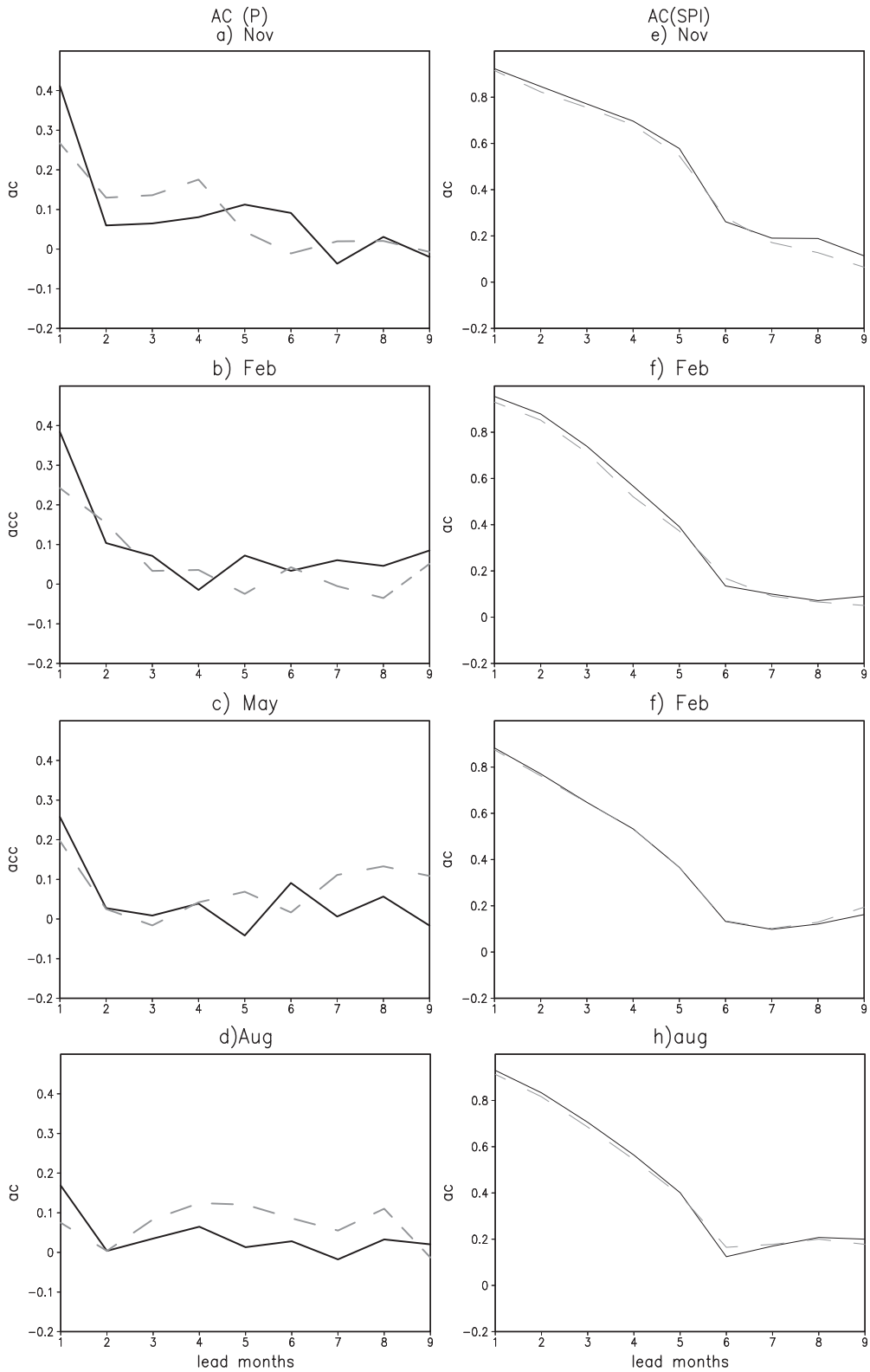


FIG. 11. Anomaly correlation of  $P$  hindcasts over the contiguous United States after the BCSD downscaling as a function of the lead time for initial conditions in (a) October–November, (b) January–February, (c) April–May, and (d) July–August for hindcasts from the CFSv2 (black) and the CFSred. (e)–(h) As in (a)–(d), but for SPI6 derived from  $P$  hindcasts.

Based on the  $P$  seasonal forecasts from the NCEP CFS, a method is developed to forecast SPIs to predict meteorological drought over the United States. Before predicting SPI, the  $P$  forecasts from a coarse-resolution global model, CFS, are downscaled and bias corrected to a regional 50-km grid. Five different methods of statistical downscaling and error correction are tested, which are 1) bilinear interpolation (BI), 2) bias correction and spatial downscaling based on the probability distribution functions (BCSD), 3) the Schaake method (Schaake et al. 2007), 4) the Bayesian method used by the Princeton University group and EMC/NCEP (Luo and Wood 2008), and 5) the multimethod ensemble. We tested cases with initial conditions in October–November, January–February, April–May, and July–August. Because drought means persistent dry conditions, we are interested in 6-month SPI (SPI6) and short-term drought measured by 3-month SPI (SPI3).

The skill is regionally and seasonally dependent. Overall, the 6-month SPI is skillful out to 3–4 months. For the first 3-month lead times, there is no statistically significant difference among different downscaling methods. After 4 months, forecasts from the multimethod ensemble have slightly lower RMSE. However, the skill is very low after 3-month lead times and it does not make any difference in reality. Different downscaling methods produce different  $P$  anomalies. However, the differences in them are not large enough to make a difference in SPI prediction. For the SPI prediction, skill comes from  $P_{\text{analy}}$  appended to the  $P$  forecasts.

For  $P$  downscaling, the simple BCSD method does well. This implies that the systematic errors of the CFS forecasts are in both the mean and the standard deviation largely caused by the detailed features, which cannot be resolved by a coarse-resolution GCM such as steep terrains. The Schaake method does well only if the CFS forecasts are skillful and the correlation is high. The occasional large negative correlation between the hindcasts and the corresponding  $P_{\text{analy}}$  degrades forecast skill because this is not a systematic feature of forecasts. The Bayesian method does not stand out in this comparison because the CFS forecasts have low skill and large ensemble spread. Even though the spread is not included here, the method still tends to move toward climatology. The multimethod ensemble has the best skill because forecasts with different downscaling methods show skill in different locations, so the average has an advantage.

The NCEP RSM and the CFS T382 model have comparable skill with statistical BCSD downscaling. For the first month of the forecast, how to construct the ensemble is more important than which downscaling methods are used. In other words, for the first month when the initial conditions are dominant, the ensemble with five youngest members has the best skill. After the first month, the

ensemble with all 15 members has a slight edge. The importance of initial conditions suggests the need to have a better observational system and better data assimilation schemes.

The SPI prediction has a few advantages. It is a simple index-based approach that can be applied to station data at locations where the historical observations are available. The skill of the SPI forecasts comes from both CFS forecast and the recent history of  $P_{\text{analy}}$ . SPI prediction is a useful tool for drought forecasters to consider because of its robustness and simplicity.

The hindcasts started in 1981. The record is too short to examine the influence of decadal SST modes such as the Pacific decadal oscillation (PDO) and the Atlantic multidecadal oscillation (AMO) on forecast skill. The model is able to capture the impact of ENSO on  $P$  over the contiguous United States if the starting month is already in an ENSO episode, but the model is not able to capture the onset or demise of ENSO. Because forecast skill of the SPI in part comes from  $P_{\text{analy}}$ , the impact of SST modes on SPI may not be as large as the influence on  $P$ . This can be an area for the future research.

The new set of the NCEP Climate Forecast System version 2 (CFSv2; Saha et al. 2010) became in operation in March 2011. It has a high-resolution atmospheric model (T126) coupled with the MOM4 ocean model. The CFSv2 forecasts start from the CFS reanalysis (CFSRR; Saha et al. 2010) with consistent oceanic and atmospheric conditions, while the CFS hindcasts obtain initial conditions from separate oceanic and atmospheric assimilation systems (R2). The configuration of CFSv2 is also different from CFS. There are 9-month forecasts initialized from every fifth day and run for four cycles of that day beginning from 1 January each year. We compared the  $P$  hindcasts after the BCSD downscaling and the corresponding SPI6 prediction between two versions of CFS hindcasts from 1982 to 2008. There are eight members in the CFSv2 ensemble. The skill does not differ much from the 16-member ensemble except in the first month, where the 8-member ensemble has slightly higher skill. The new CFSv2 has higher skill for the first month because they have consistent initial conditions. The anomaly correlation for  $P$  increases from 0.05 in summer (August) to 0.1 in winter (Figs. 11a–d), which is consistent with Yuan et al. (2011). Unfortunately, the improvement is limited to the first month. After that, the skills for the two CFS models are comparable. Because the skill is so low after the first month, they do not make any difference in reality. The improvement in  $P$  is not significant enough to improve skill in SPI6 (Figs. 11e–h).

It is worthwhile to mention another recent effort: the Multi-RCM Ensemble Downscaling of Global Seasonal Forecast (MRED) project. The MRED project utilizes

a large number of regional climate models to produce downscaled CFS forecasts initialized at November–December. As shown in our study, systematic errors of the CFS are mainly caused by its spatially coarse resolution and steep terrain. The comparison between the regional model downscaling and statistical downscaling is focused here on boreal summer. In contrast, the target of the MRED project is the Northern Hemisphere winter over the western United States. The western region receives a large amount of rainfall and snow in the winter and the region is largely forced by large-scale atmospheric circulation influenced by orographic (terrain related) features. Therefore, regional models can provide additional forecast skill.

*Acknowledgments.* This work was done while J.-H. Yoon was at the Climate Prediction Center, NOAA/NWS and the Cooperative Center for Climate and Satellites (CICS), Earth System Science Interdisciplinary Center (ESSIC), University of Maryland, College Park. We would like to acknowledge support from the NOAA (TRACS) Grants GC09-505 and GC09-552. Comments from anonymous reviewers were helpful in improving the manuscript. We also thank Dr. Po-Lun Ma at PNNL for his internal review and constructive comments. Also, this research is partially supported by NOAA Climate Prediction Programs for the Americas (CPPA) to PNNL. PNNL is operated for the U.S. Department of Energy by Battelle Memorial Institute under Contract DE-AC06-76RPL1830.

#### REFERENCES

- Barros, A. P., and G. J. Bowden, 2008: Toward long-lead operational forecasts of drought: An experimental study in the Murray-Darling River Basin. *J. Hydrol.*, **357**, 349–367.
- Behringer, D. W., M. Ji, and A. Leetmaa, 1998: An improved coupled model for ENSO prediction and implications for ocean initialization. Part I: The ocean data assimilation system. *Mon. Wea. Rev.*, **126**, 1013–1021.
- Coelho, C. A. S., S. Pezzulli, M. Balmaseda, E. J. Doblas-Reyes, and D. B. Stephenson, 2004: Forecast calibration and combination: A simple Bayesian approach for ENSO. *J. Climate*, **17**, 1504–1516.
- Gutowski, W. J., F. Otieno, R. W. Arritt, E. S. Takle, and Z. Pan, 2004: Diagnosis and attribution of a seasonal precipitation deficit in a U.S. regional climate simulations. *J. Hydrometeorol.*, **5**, 230–242.
- Hayes, M. J., M. D. Svoboda, D. A. Wilhite, and O. V. Vanyarkho, 1999: Monitoring the 1996 drought using the standardized precipitation index. *Bull. Amer. Meteor. Soc.*, **80**, 429–438.
- , O. V. Wilhelm, and C. L. Knutson, 2004: Reducing drought risk: Bridging theory and practice. *Nat. Hazards Rev.*, **5**, 106–113.
- Hidalgo, H. G., M. D. Dettinger, and D. R. Cayán, 2007: Downscaling with constructed analogues: Daily precipitation and temperature fields over the United States. PIER Final Project Rep. CEC-500-2007-123, 48 pp.
- Juang, H.-M. H., and M. Kanamitsu, 1994: The NMC nested regional spectral model. *Mon. Wea. Rev.*, **122**, 3–26.
- , S. Y. Hong, and M. Kanamitsu, 1997: The NCEP regional spectral model: An update. *Bull. Amer. Meteor. Soc.*, **78**, 2125–2143.
- 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.
- Keyantash, J., and J. A. Dracup, 2002: The quantification of drought: An evaluation of drought indices. *Bull. Amer. Meteor. Soc.*, **83**, 1167–1180.
- Kim, J., T.-K. Kim, R. W. Arritt, and N. L. Miller, 2002: Impacts of increase atmospheric CO<sub>2</sub> on the hydroclimate of the western United States. *J. Climate*, **15**, 1926–1942.
- Leung, L. R., Y. Qian, X. Bian, and A. Hunt, 2003: Hydroclimate of the western United States based on observations and regional climate simulation of 1981–2000. Part II: Mesoscale ENSO anomalies. *J. Climate*, **16**, 1912–1928.
- 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.
- Long, L. N., J.-K. E. Schemm, and S. Saha, 2008: Assessment of the CFS on the predictability of the North American Monsoon. NOAA, 8 pp. [Available online at [http://www.nws.noaa.gov/ost/climate/STIP/CTB-COLA/Lindsey\\_013108.pdf](http://www.nws.noaa.gov/ost/climate/STIP/CTB-COLA/Lindsey_013108.pdf).]
- Lott, J. N., and T. F. Ross, 2006: Tracking and evaluating U.S. billion dollar weather disasters, 1980–2005. Preprints, *AMS Forum: Environmental Risk and Impacts on Society: Successes and Challenges*, Atlanta, GA, Amer. Meteor. Soc., 1.2. [Available online at [http://ams.confex.com/ams/Annual2006/techprogram/paper\\_100686.htm](http://ams.confex.com/ams/Annual2006/techprogram/paper_100686.htm).]
- Luo, L., and E. F. Wood, 2006: Assessing the idealized predictability of precipitation and temperature in the NCEP Climate Forecast System. *Geophys. Res. Lett.*, **33**, L04708, doi:10.1029/2005GL025292.
- , 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.
- Mahrt, L., and H.-L. Pan, 1984: A two-layer model of soil hydrology. *Bound.-Layer Meteorol.*, **29**, 1–20.
- Maurer, E. P., and H. G. Hidalgo, 2008: Utility of daily vs. monthly large-scale climate data: An intercomparison of two downscaling methods. *Hydrol. Earth Syst. Sci.*, **12**, 551–563.
- McKee, T. B., N. J. Doesken, and J. Kleist, 1993: The relationship of drought frequency and duration to time scales. Preprints, *Eighth Conf. on Applied Climatology*, Anaheim, CA, Amer. Meteor. Soc., 179–184.
- , —, and —, 1995: Drought monitoring with multiple time scale. Preprints, *Ninth Conf. on Applied Climatology*, Dallas, TX, Amer. Meteor. Soc., 233–236.
- Moorathi, S., H.-L. Pan, and P. Caplan, 2001: Changes to the 2001 NCEP operational MRW/AVN global analysis/forecast system. NWS Tech. Procedures Bulletin 484, 14 pp. [Available online at <http://www.nws.noaa.gov/om/tpb/484.htm>.]
- Moreira, E. E., C. A. Coelho, A. A. Paulo, L. S. Pereira, and J. T. Mexia, 2008: SPI-based drought category prediction using loglinear models. *J. Hydrol.*, **354**, 116–130.
- Pacanowski, R. C., and S. M. Griffies, cited 1998: MOM 3.0 manual. 708 pp. [Available online at <http://www.gfdl.noaa.gov/>]

- cms-filessystem-action/model\_development/ocean/mom3\_manual.pdf.]
- 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. *Hydro. Earth Sci. Discuss.*, **4**, 655–717.
- Wilby, R. L., and T. M. L. Wigley, 1997: Downscaling general circulation model output: A review of methods and limitations. *Prog. Phys. Geogr.*, **21**, 530–548.
- Wood, A. W., and D. P. Lettenmaier, 2006: A test bed for new seasonal hydrologic forecasting approaches in the western United States. *Bull. Amer. Meteor. Soc.*, **87**, 1699–1712.
- , and J. Schaake, 2008: Correcting errors in streamflow forecast ensemble mean and spread. *J. Hydrometeor.*, **9**, 132–148.
- , 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.
- Xie, P. P., M. Chen, and W. Shi, 2010: CPC unified gauge-based analysis of global daily precipitation. Preprints, *24th Conf. on Hydrology*, Atlanta, GA, Amer. Meteor. Soc., 2.3A. [Available online at [http://ams.confex.com/ams/90annual/techprogram/paper\\_163676.htm](http://ams.confex.com/ams/90annual/techprogram/paper_163676.htm).]
- Yuan, X., E. F. Wood, 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.

Copyright of Journal of Hydrometeorology is the property of American Meteorological Society and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.