

## Global Meteorological Drought Prediction Using the North American Multi-Model Ensemble

KINGTSE C. MO

*Climate Prediction Center, NOAA/NWS/NCEP, College Park, Maryland*

BRADFIELD LYON

*International Research Institute for Climate and Society, Earth Institute, Columbia University, Palisades, New York*

(Manuscript received 9 October 2014, in final form 9 February 2015)

### ABSTRACT

Precipitation forecasts from six climate models in the North American Multi-Model Ensemble (NMME) are combined with observed precipitation data to generate forecasts of the standardized precipitation index (SPI) for global land areas, and their skill was evaluated over the period 1982–2010. The skill of monthly precipitation forecasts from the NMME is also assessed. The value-added utility in using the NMME models to predict the SPI is identified by comparing the skill of its forecasts with a baseline skill based solely on the inherent persistence characteristics of the SPI itself. As expected, skill of the NMME-generated SPI forecasts depends on the season, location, and specific index considered (the 3- and 6-month SPI were evaluated). In virtually all locations and seasons, statistically significant skill is found at lead times of 1–2 months, although the skill comes largely from initial conditions. Added skill from the NMME is primarily in regions exhibiting El Niño–Southern Oscillation (ENSO) teleconnections. Knowledge of the initial drought state is critical in SPI prediction, and there are considerable differences in observed SPI values between different datasets. Root-mean-square differences between datasets can exceed typical thresholds for drought, particularly in the tropics. This is particularly problematic for precipitation products available in near-real time. Thus, in the near term, the largest advances in the global prediction of meteorological drought are obtainable from improvements in near-real-time precipitation observations for the globe. In the longer term, improvements in precipitation forecast skill from dynamical models will be essential in this effort.

### 1. Introduction

The routine generation of global seasonal climate forecasts coupled with advances in near-real-time monitoring of the global climate has now allowed for testing the feasibility of generating global drought forecasts operationally. Indeed, the development of an experimental global drought information system (GDIS) that includes both real-time monitoring and forecasts has recently been recommended by a working group of the World Climate Research Programme (WCRP 2012) and promoted in various prior forums (Pozzi et al. 2013). An essential step in developing a pragmatically useful GDIS is to demonstrate that seasonal drought forecasts are skillful at

sufficient lead times to be able to inform decision making of drought-affected activities. Evaluating such skill over land areas for the near-global domain is the main goal of this study.

In this paper, drought is viewed from a meteorological perspective and defined using the standardized precipitation index (SPI; McKee et al. 1993, 1995). Seasonal forecasts of the 3- and 6-month SPI (SPI3 and SPI6, respectively) are generated by combining an analysis of observed precipitation with monthly forecasts obtained from six coupled ocean–land–atmosphere general circulation models (CGCMs) in the North American Multi-Model Ensemble (NMME). While the focus of the paper is on the prediction of meteorological drought, the study emphasizes two fundamental constraints in generating reliable global drought predictions that will arise whether using the reported method or any other approach (e.g., land surface modeling): 1) the skill of the monthly and seasonal precipitation forecasts and 2) the

*Corresponding author address:* Kingtse C. Mo, Climate Prediction Center, NOAA/NWS/NCEP, 5830 University Research Ct., College Park, MD 20740.  
E-mail: kingtse.mo@noaa.gov

quality of the observational data needed to accurately describe the initial drought condition. The first issue will be examined by evaluating the skill of the NMME monthly and seasonal precipitation forecasts (1982–2010) as a function of location, starting date, and lead time. The second will be illustrated by presenting differences in SPI values when computed from different precipitation analyses. It will be shown that it is particularly challenging to obtain reliable global precipitation data that are both available in near-real time and have a sufficiently long history to provide fairly stable statistics. Ideally, it is desirable to have 50 or more years of historical precipitation data when computing the SPI, with at least 30 years considered a minimum value (Guttman 1999). High-spatial-resolution, satellite-derived rainfall (and soil moisture and air temperature) estimates are becoming more widely available and will play an increasingly important role in drought monitoring and prediction going forward. At present, however, such data have a fairly short historical record and, in any case, still require calibration to surface observations in some manner, which, for the global domain, presents a major challenge in its own right.

There have been some recent efforts to generate prototype seasonal drought forecasts for the globe. Yuan and Wood (2013) used seasonal precipitation forecasts from the NMME and other GCMs to examine the predictability of drought onset around the globe based on the SPI, applying an onset definition used by Mo (2011). For the global domain, they found only a modest increase in the forecast probability of onset relative to baseline expectations when using the GCM forecasts. Dutra et al. (2014a,b) generated global forecasts of the 3-, 6-, and 12-month SPI by combining seasonal precipitation reforecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF) System 4 GCM with precipitation observations from the Global Precipitation Climatology Centre (GPCC) and, alternatively, the ECMWF Interim Reanalysis. They reported on several verification metrics for the SPI forecasts for 18 regions around the globe. Using the same definition as Yuan and Wood (2013), they found that the ECMWF System 4 provides useful skill in predicting drought onset in several locations, and the skill is largely derived from El Niño–Southern Oscillation (ENSO) teleconnections. However, they also found that it is difficult to improve on “climatological” forecasts generally. Hao et al. (2014) describe the Global Integrated Drought Monitoring and Prediction System (GIDMaPS) that uses three drought indicators. The forecasting component of their system relies on a statistical approach based on an ensemble streamflow prediction (ESP) methodology. Systems to monitor drought around the globe are described in

Vicente-Serrano et al. (2010) for meteorological drought and Nijssen et al. (2014) for hydrologic and agricultural conditions. This paper builds on these previous studies by considering a multimodel ensemble framework to generate seasonal forecasts of the SPI and assessing their overall skill (i.e., not limited to onset probability) and by assessing the influence of differences in observational datasets in generating the initial condition for the SPI forecasts. The initial condition has already been shown to be a major source of seasonal forecast skill in meteorological drought prediction (Lyon et al. 2012; Quan et al. 2012).

The paper is outlined as follows. The observational and climate model datasets used and a description of the overall analysis procedures are described in section 2. Section 3 reports on the evaluation of forecast skill, while section 4 considers uncertainty in the SPI analysis associated with different observational datasets. Section 5 provides a discussion of the results and the main conclusions drawn from the study.

## 2. Data and procedures

### a. Precipitation analyses

Two monthly mean precipitation analyses (considered the precipitation observations  $P_{\text{obs}}$ ) are used in this study: GPCC, version 6 (GPCC v6; Rudolf and Schneider 2005; Becker et al. 2013), and U.S. Climate Prediction Center (CPC) Unified precipitation (Chen et al. 2002). Both datasets are based on station gauge reports that have been gridded to a  $1^{\circ}$  latitude  $\times$   $1^{\circ}$  longitude spatial resolution and cover the period from January 1950 to December 2010. The monthly CPC Unified precipitation is based on a daily analysis generated in near-real time with 1-day lag ([www.esrl.noaa.gov/psd/data/gridded/data.unified.daily.conus.html](http://www.esrl.noaa.gov/psd/data/gridded/data.unified.daily.conus.html)). The GPCC dataset is a historical analysis that uses monthly average station data with more station inputs than for the CPC analysis. The trade-off is that it is not available in near-real time (the GPCC v6 data used here end in 2010). Given the additional time available to generate the analysis, the GPCC dataset also has better quality control than the CPC Unified precipitation. It should be noted that over several regions of the globe, particularly Africa, South America, and many areas in the tropics, there are many  $1^{\circ} \times 1^{\circ}$  grid boxes in both datasets that have no station observations available, with spatial interpolation techniques used to fill out the fields. This can result in large differences in the respective precipitation analyses. The SPI3 and SPI6 were computed following the procedures of McKee et al. (1993, 1995) over the full analysis period in both datasets. Hereafter, they are labeled as SPI<sub>3</sub><sub>GPCC</sub> and SPI<sub>3</sub><sub>CPC</sub> for SPI3 and SPI<sub>6</sub><sub>GPCC</sub> and SPI<sub>6</sub><sub>CPC</sub> for SPI6.

TABLE 1. NMME models utilized.

Modeling center	Description	Ensemble size	Lead times (months)
NCEP	Climate Forecast System, version 2 (CFSv2)	16	0.5–8.5
GFDL	GFDL Climate Model, version 2.2 (GFDL CM2.2)	10	0.5–11.5
NCAR	Community Climate System Model, version 3 (CCSM3)	6	0.5–11.5
NASA	Goddard Earth Observing System Model, version 5 (GEOS-5)	10	0.5–9.5
CMC1	Third Generation Canadian Coupled Global Climate Model (CanCM3)	10	0.5–11.5
CMC2	Fourth Generation Canadian Coupled Global Climate Model (CanCM4)	10	0.5–11.5

A signal-to-noise analysis for the  $P_{\text{obs}}$  data required the use of additional monthly gridded analyses, which included the University of East Anglia Climatic Research Unit (CRU) gridded climate dataset (referred to as TS version 3.10) (Harris et al. 2014); the Global Precipitation Climatology Project (GPCP), version 2.2 (Huffman et al. 2009); the CPC Merged Analysis of Precipitation (CMAP; Xie and Arkin 1997); and the Climate Anomaly Monitoring System and OLR Precipitation Index (CAMS\_OPI; Janowiak and Xie 1999). All of these datasets were gridded to a common  $2.5^{\circ}$  latitude  $\times$   $2.5^{\circ}$  longitude resolution. The common period 1979–2005 was used for this particular calculation.

### b. NMME

Monthly mean precipitation  $P$  forecasts from 1982–2010 were obtained from the NMME archive (Kirtman et al. 2013). There are a total of six models utilized in the NMME (Table 1), of which CMC1, CMC2, GFDL, and NASA have 10 ensemble members each. The NCAR model has six ensemble members, while CFSv2 has a total of 24 members, of which only 16 ensemble members were used. The latter relates to the fact that the archived CFSv2 seasonal hindcasts were generated every 5 days from 1 January 1982 to 27 December 2010 (Saha et al. 2014), with four forecast runs made on each of these days (6 start times  $\times$  4 ensemble members = 24 runs). Including more than 16 ensemble members for CFSv2 does not increase overall forecast skill because the additional ensemble members are more than 20 days away from the start time of the forecast. The horizontal resolution for all model data from the NMME is also  $1^{\circ}$  latitude  $\times$   $1^{\circ}$  longitude.

### c. SPI forecasts

We generated SPI forecasts for each model in the NMME separately following the procedure outlined in Yoon et al. (2012). To remove any systematic biases in the mean values of the model  $P$  forecasts, we first determined the “anomalous” values of the  $P$  forecasts computed relative to a given model’s “hindcast climatology.” To do this, we first remove one year  $T_y$  from the hindcasts for each of the NMME models individually

and denote this as the target year. The rest of the years in the hindcasts are considered the training period used to compute a model’s hindcast climatology. For example, for a given model hindcast start time of month  $M$ , we denote the monthly mean  $P$  hindcast for ensemble member  $n$ , lead time  $\tau$ , and target year as  $P(T_y, \tau, M, n)$ . The ensemble-mean  $P$  hindcast value for  $M$  and  $\tau$  is denoted as  $\bar{P}(T_y, \tau, M)$  and expressed mathematically as

$$\bar{P}(T_y, \tau, M) = \langle P(T_y, \tau, M, n) \rangle_n, \quad (1)$$

where  $\langle \cdot \rangle_n$  denotes the equally weighted value of all the ensemble members  $n$ . For the same  $M$  and  $\tau$ , we then compute a climatological hindcast averaged over all hindcasts in the training period and across all forecast ensemble members, denoting this as  $P_c(\tau, M)$ . For  $T_y$  and  $M$ , a hindcast  $P$  anomaly  $P_a(T_y, \tau, M)$  is then defined as

$$P_a(T_y, \tau, M) = \bar{P}(T_y, \tau, M) - P_c(\tau, M). \quad (2)$$

The verification of  $P_a$  values is accomplished by comparing it with the  $P_{\text{obs}}$  anomaly obtained from the GPCC or CPC datasets at  $M + \tau$  for  $T_y$ .

The procedure to generate SPI forecasts is illustrated here by considering forecasts initialized on 1 January 1993 as an example (for the CFSv2, it is based on the average of 16 ensemble members generated most closely to 1 January 1993). Therefore, the forecast at lead of 1 month is for January 1993 and is considered as the 0.5-month lead. In general, for any  $T_y$ , we first obtain the observed climatology ( $P_{\text{obs}}$ ) from  $T_y - 30$  to  $T_y - 1$ , which is the base period. For the above example,  $T_y = 1993$  and the base period is 1963–92. Note that there are no observations used after the target year. We then add the ensemble mean  $P_a$  to the observed monthly climatology  $\bar{P}_{\text{obs}}$  for each lead time,  $\tau = 1\text{--}6$  months, to obtain the monthly hindcast values of precipitation:

$$P_{\text{f cst}} = P_a + \bar{P}_{\text{obs}}. \quad (3)$$

For hindcasts with a start time of 1 January 1993, the first (sixth) lead time corresponds to January (June) 1993 so the  $\bar{P}_{\text{obs}}$  for January–June are added to the  $P_a$  forecasts

to obtain  $P_{\text{fct}}$ . These forecast precipitation values are then appended to the  $P_{\text{obs}}$  data to form a new time series, which now has 366 months total (i.e., from January 1963 to June 1993). The SPI3 and SPI6 are computed from this new time series, with the last six values representing the forecast. This was performed for each model. A multimodel, ensemble-mean SPI forecast is then obtained as the equally weighted mean of the SPI forecasts across all six models in the NMME for a given lead time. To verify these SPI forecasts, they are compared with the SPI3<sub>GPCC</sub> and SPI6<sub>GPCC</sub> observed values. Note that for forecast verification only the GPCC precipitation data are used and will be considered as representing the true observed values. In section 4, differences in the observed SPI condition will be examined across different  $P$  analyses.

#### d. Verification

We verify the SPI hindcasts for each model and the multimodel ensemble mean for each lead time (0.5–5.5 months) and each year (1982–2010). We used the Pearson correlation to assess the forecast skill. If one year is considered as having one degree of freedom, then  $N = 29$ . To be considered statistically significant at the 5% (10%) confidence level, the correlation between forecast SPI values and those in the verifying SPI<sub>GPCC</sub> data needs to be greater than 0.37 (0.31). However, since the SPI is based on accumulated monthly precipitation (here, over 3- and 6-month periods), at a given location (grid point) it is likely to exhibit substantial autocorrelation. As such, the true skill of the NMME SPI forecasts needs to be evaluated relative to the baseline autocorrelation of the index. Such a baseline is described in Lyon et al. (2012), where a methodology for generating “optimal persistence” forecasts (for the baseline case where there is no serial correlation in monthly precipitation) of the SPI based on its autocorrelation is presented. These statistical forecasts will be considered the baseline in this study and are based on the SPI<sub>GPCC</sub> data. If the NMME is providing value-added skill to the SPI forecasts, it will be manifested by temporal correlations with observations that exceed these baseline values. Under the assumption that these respective sets of forecasts are normally distributed, to assess the statistical significance of the difference between two correlations  $r_1$  and  $r_2$ , we used Fisher’s  $Z$  transformation as explained in Quan et al. (2012).

We define  $Z_i$  as

$$Z_i = \frac{1}{2} \ln \left( \frac{1 + r_i}{1 - r_i} \right) \quad (4)$$

for  $i = 1$  and 2. The transformation  $Z$  is assumed to be normally distributed with variance  $(N - 3)^{-1}$ , where  $N$  is

the number of forecasts ( $N = 29$ ). We then transformed  $r_1$  and  $r_2$  to  $Z_1$  and  $Z_2$ , with the statistical significance for the difference in correlations assessed using the  $Z$  statistics:

$$Z = \frac{Z_1 - Z_2}{\sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}}, \quad (5)$$

where  $N_1 - 3$  and  $N_2 - 3$  are the degrees of freedom for  $r_1$  and  $r_2$ , respectively. Here  $N_1 = N_2 = 29$ . Using a null hypothesis of equal correlation and a nondirectional alternative hypothesis of unequal correlation, if  $Z$  is greater than 1.96, the difference in correlations is statistically significant at the 5% confidence level.

The temporal correlation between forecast and observed values of the SPI provides an overall measure of forecast skill, one that is not limited to the case of drought alone. Therefore, we also evaluated SPI forecasts in the context of being able to detect drought identified when the index drops below a particular threshold. Here we first identified drought events as occurring when the observed SPI<sub>GPCC</sub> for a given month was  $<-1.2$ , which corresponds to the “severe drought” category D2 or higher in the U.S. Drought Monitor (Svoboda et al. 2002). There is no persistence requirement of this threshold to be classified as a drought. By adding persistence as a requirement, there will not be enough drought events to obtain reliable statistics. For a given lead time, we computed several verification measures for forecasts being below this threshold based on a contingency table approach applied at each grid point in the study domain. The entries in the table are defined as follows:  $A$  is the number of drought events that are forecast and occur,  $B$  is the number of drought events that are forecast but do not occur,  $C$  is the number of drought events that are not forecast but do occur, and  $D$  is the number of drought events that are not forecast and do not occur. The variable  $N$  is the total number of drought events analyzed from 1982 to 2010. Based on these values, the hit rate is defined as  $H = A/(A + C)$ , the false alarm rate is defined as  $\text{FAR} = B/(A + B)$ , and the threat score is defined as  $\text{TS} = A/(A + B + C)$ .

### 3. Evaluation of $P$ and SPI forecast skill

Precipitation fluctuations are the sole source of SPI variability. Therefore, before considering the SPI, we first evaluated the skill of monthly and seasonal  $P$  forecasts from the NMME. Figure 1 displays climatological values of the monthly precipitation and standard deviation in the GPCC data for reference. Figure 2 shows the ensemble-mean NMME  $P$  forecast skill

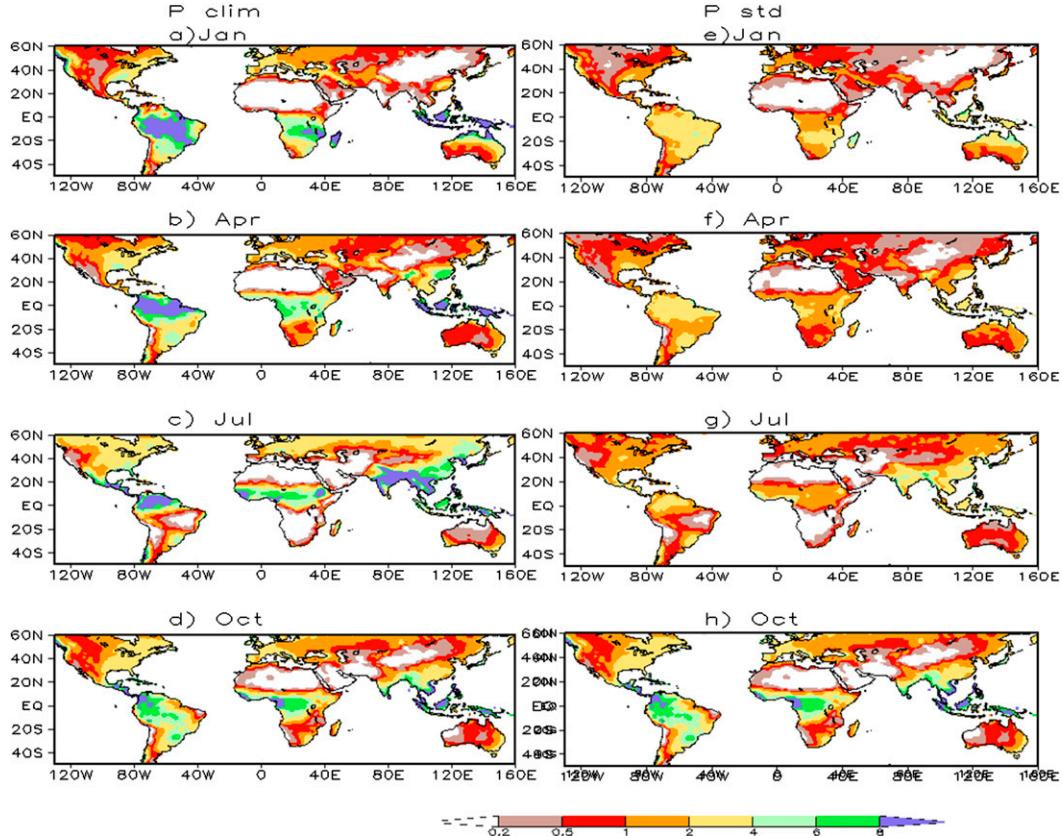


FIG. 1. The monthly mean precipitation climatology ( $\text{mm day}^{-1}$ ) from GPCC for the period 1950–2010 for (a) January, (b) April, (c) July, and (d) October. Values are indicated by the color bar. (e)–(h) As in (a)–(d), but for the std dev of monthly precipitation for the same months and base period.

measured by the Pearson correlation between the monthly mean  $P$  forecasts and the corresponding GPCC data at each lead for forecasts initialized on the first day of the month of January, April, July, and October. As expected, the forecast skill is seen to be seasonally and regionally dependent. Overall, skill is generally higher in the winter hemisphere than in the summer hemisphere, and it is generally lower for start times during the transition seasons April and October.

In Fig. 1, areas that have both low average  $P$  and standard deviation are either deserts or areas with a pronounced annual cycle that are undergoing a relative dry season. In such areas, the  $P$  forecasts are not particularly meaningful, as little or no precipitation is expected climatologically. Conversely, the global monsoons play an important role in defining the global water cycle. However, the monthly NMME  $P$  forecasts even at a lead of 0.5 months (Fig. 2) are unable to capture much of the observed variability in the monsoons of Asia, Africa, and North America (with the caveat that data coverage in these areas is sparse). Generally speaking, there is some suggestion that, at lead of 1 month, the skill

is higher for comparatively drier regions than the climatologically wettest areas. For a January start time, the ensemble mean of the NMME exhibits some skill in capturing precipitation variability over the western United States, eastern China, and portions of Europe. Statistically significant skill is identified over Central America and northeastern Brazil for January and April start times and northeastern Australia for 0.5-month leads starting in April and July. After lead of 1 month, skill of the monthly  $P$  forecast drops off sharply. The seasonal forecasts are in general more skillful than the monthly forecasts after lead of 1 month (Fig. 3). Some skill is likely contributed by ENSO. For example, skill is higher over the southern United States in January–March (JFM), Central America in JFM and October–December (OND), northeastern Brazil in April–June (AMJ), and Australia in July–September (JAS). These are regions known to have an ENSO signal, and the more seasonal skillful  $P$  forecasts contribute to the SPI3 seasonal forecasts at lead of 1 month. Large forecast errors indicated by negative correlations are not systematic and are not amenable to correction by postprocessing. Of course,

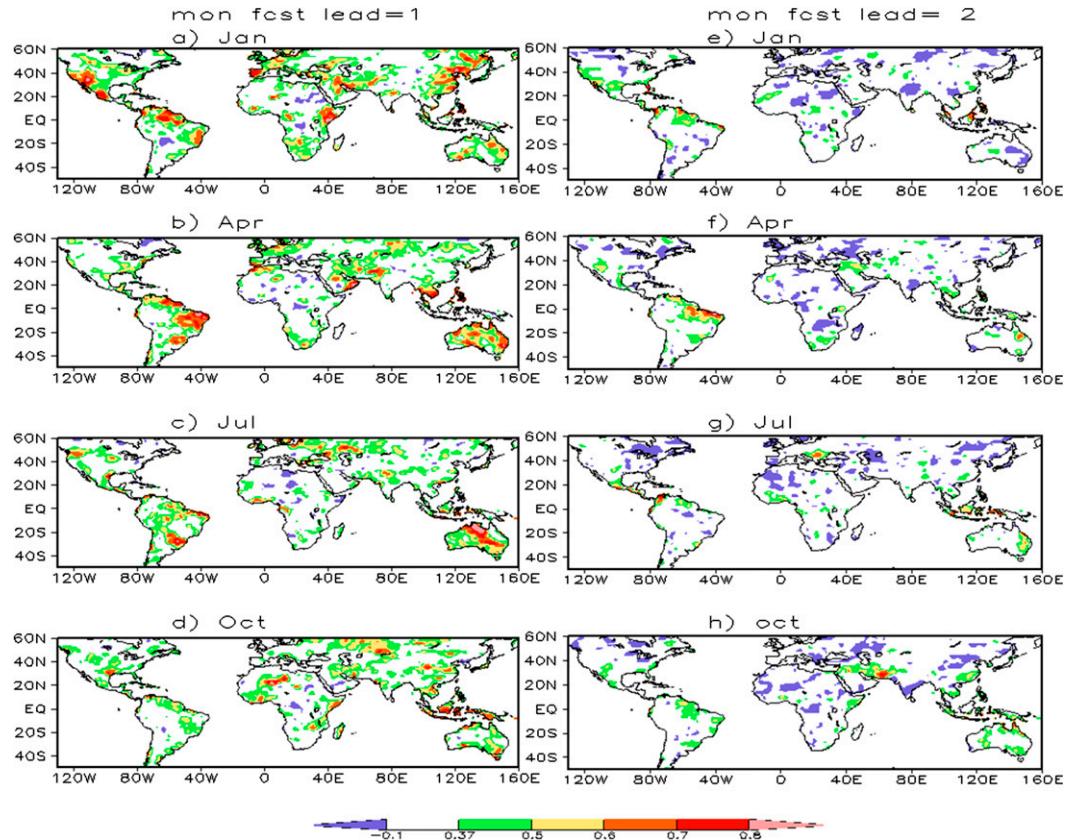


FIG. 2. NMME precipitation forecast skill as measured by the Pearson correlation with  $P_{\text{obs}}$  from GPCC for forecasts with a lead of 1 month initialized on (a) 1 Jan, (b) 1 Apr, (c) 1 Jul, and (d) 1 Oct. Values are indicated by the color bar. (e)–(h) As in (a)–(d), but for a lead of 2 months.

these unskillful  $P$  forecasts negatively impact the skill of SPI forecasts, which are considered next.

The NMME-mean SPI forecasts are typically more skillful than forecasts from any single model. As an example, we calculated the anomaly correlation coefficient

(ACC), which is the pattern correlation between forecasts and verifying SPI from  $P$  analysis (not shown) and root-mean-square error (RMSE; Fig. 4) for the SPI6 forecasts over land areas in the Northern Hemisphere ( $20^{\circ}$ – $60^{\circ}$ N), tropics ( $20^{\circ}$ S– $20^{\circ}$ N), and Southern Hemisphere

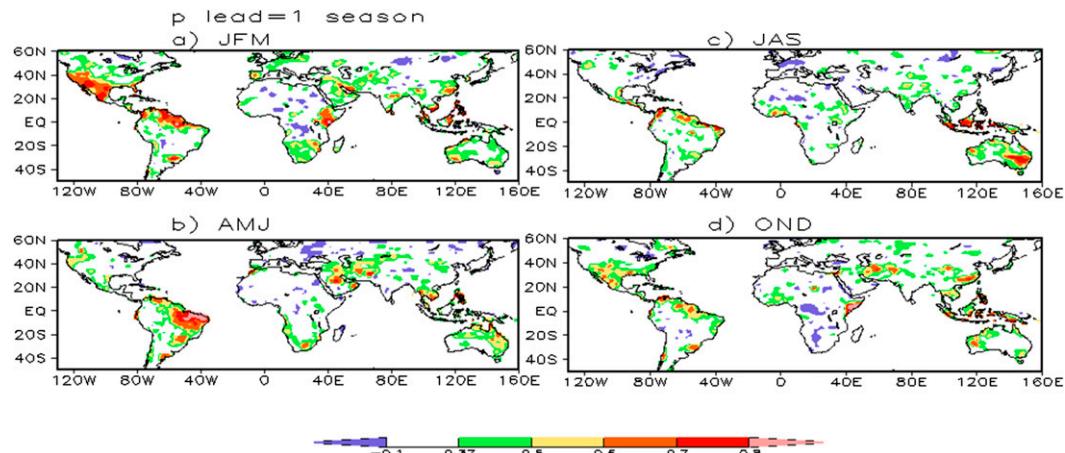


FIG. 3. NMME precipitation forecast skill as measured by the Pearson correlation with  $P_{\text{obs}}$  from GPCC for seasonal forecasts with a lead of 1 season initialized on the first of the season for (a) JFM, (b) AMJ, (c) JAS, and (d) OND. Values indicated by the color bar.

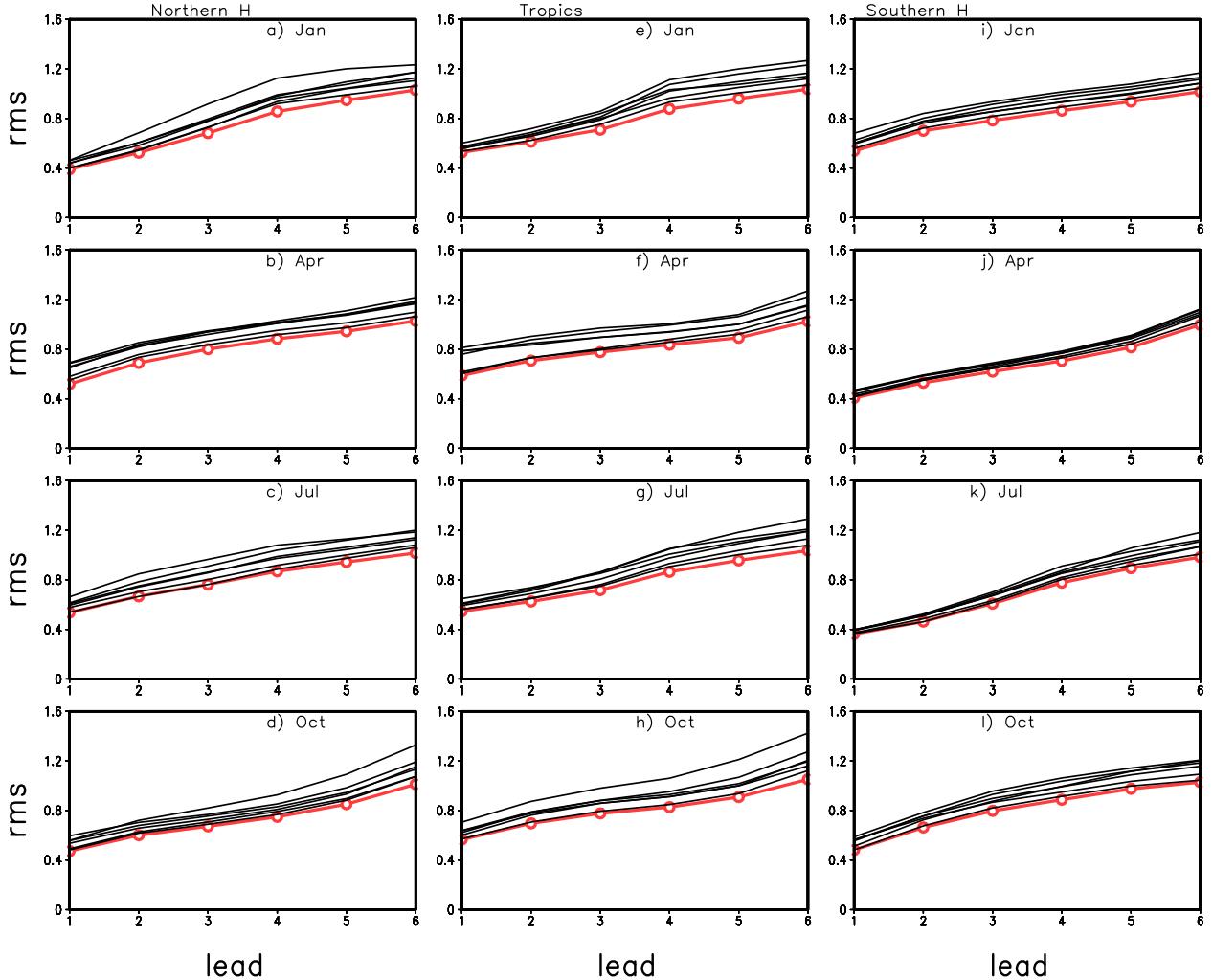


FIG. 4. RMSE of SPI6 forecasts for Northern Hemisphere ( $20^{\circ}$ – $60^{\circ}$ N) land areas as a function of lead time for forecasts initialized on (a) 1 Jan, (b) 1 Apr, (c) 1 Jul, and (d) 1 Oct for each individual model (dark line) and the NMME average (red open circles). (e)–(h) As in (a)–(d), but for the tropics ( $20^{\circ}$ S– $20^{\circ}$ N). (i)–(l) As in (a)–(d), but for the Southern Hemisphere ( $20^{\circ}$ – $60^{\circ}$ S). GPCC data used for verifications.

( $20^{\circ}$ – $60^{\circ}$ S) as a function of lead time. Conclusions from RMSE and ACC are the same. The NMME SPI6 forecasts consistently show lower RMSE (or higher ACC) values than for the individual model members. The RMSE values are also typically lower over the winter hemisphere than the summer hemisphere, although in addition to skill of the  $P$  forecasts, the effect of the seasonal cycle of climatological  $P$  on the persistence of the SPI may be at least partially responsible for this behavior (Lyon et al. 2012). On average, the RMSE values across these regions reach 1 between leads of 3 and 4 months, which means very little skill at longer leads. Although not reflected in the area averages in Fig. 4, at short leads forecasts for several tropical locations show lower RMSE (and greater skill) because the models are able to capture ENSO teleconnections with  $P$  there (Kirtman et al. 2013).

The higher skill of the NMME relative to its component models is an expected result given the larger ensemble size and possible complimentary skill arising from the diversity of models (Kirtman et al. 2013). As a simple examination of the latter contribution, the area-averaged correlation<sup>1</sup> ( $60^{\circ}$ S– $60^{\circ}$ N, land areas only) between the CFSv2 seasonal  $P$  forecasts and observations (GPCC) was compared with a mini-NMME consisting of four models (CMC1, CMC2, GFDL, and NASA). The CFSv2 and mini-NMME each used a total of 16 ensemble members,

<sup>1</sup> The area-averaged correlation was computed by squaring gridpoint correlation values while retaining their sign, multiplying by the cosine of latitude, averaging across all grid points, and then taking the square root of the results.

the latter using four ensemble members for each of the four models in the mini-NMME. Six overlapping 3-month seasons were evaluated using forecasts initialized in October. The results (Fig. 5a) indicate that, with the exception of the first season lead (where skill is comparable), the mini-NMME forecast shows greater skill than the CFSv2. Since the number of ensemble members is the same, this suggests the existence of complimentary skill in the individual mini-NMME models. To examine this further, for the same forecast start and lead times, we used 10 ensemble members for the CFSv2 and 10 each for the four models in the mini-NMME (i.e., a total of 40 ensemble members for the latter). The area-averaged correlation with observations was again computed, but now for each model individually and then averaged across the models. If complimentary skill exists across the mini-NMME models, we would expect the average skill of the four individual models to be less than the skill of the ensemble average forecast of the four models (Fig. 5a). For a lead of 1 month, the CFSv2 in fact shows (Fig. 5b) greater skill than the average of the four models in the mini-NMME. For longer leads, the skill of the CFSv2 and average of the mini-NMME component models is roughly equal, but still less than that of the ensemble average forecasts in the mini-NMME. While these are small samples and the statistical significance of the results was not attempted, they again suggest that complimentary skill across the NMME models does add to the total skill of the NMME forecasts.

Since the NMME mean has the highest skill, in what follows we only present the skill for the multimodel, ensemble-mean SPI forecasts, correlating them with the observed SPI<sub>3GPCC</sub> (Fig. 6) and SPI<sub>6GPCC</sub> (Fig. 7) over 1982–2010. The lack of skill in  $P$  forecasts clearly influences the skill of the SPI forecasts at longer leads in both figures, but for short lead times the SPI<sub>6</sub> and SPI<sub>3</sub> forecasts are overall quite skillful. For example, SPI<sub>6</sub> forecasts at virtually all locations have correlation values above 0.7 at a lead of 1 month, which is highly statistically significant. However, the main reason for these high correlations is that forecasts of SPI<sub>6</sub> at a lead of 1 month are based on 5 months of observed  $P$  and only 1 month of forecast  $P$  values, which do show some skill in several locations as well (see Fig. 2). Here, it suffices to say that it is not a surprising outcome that the forecasts of SPI<sub>3</sub> at a lead of 1 month and SPI<sub>6</sub> at a lead of 3 months generally have a similar level of skill. Overall, regardless of its source, the skill of the SPI<sub>3</sub> and SPI<sub>6</sub> forecasts in Figs. 6 and 7 is statistically significant in many regions of the globe, with spatial variations in skill dependent on the start time of the forecast. Regions with lower skill in SPI forecasts generally occur for start times that are near the start of the rainy season. For example, SPI<sub>6</sub> forecasts

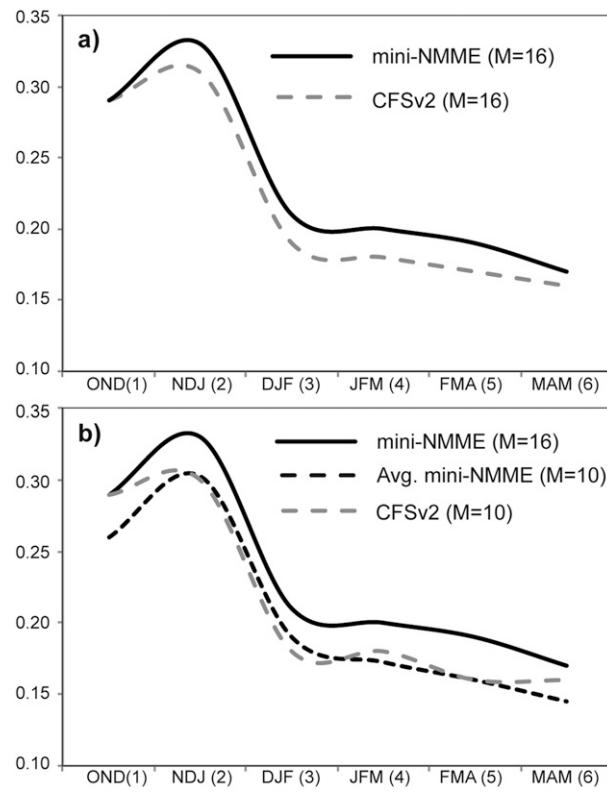


FIG. 5. (a) Area-averaged correlation (land areas only, 60°S–60°N) between seasonal rainfall forecasts from CFSv2 (16 ensemble members) with a 16-member mini-NMME (4 models  $\times$  4 ensemble members per model) and GPCC observations for six overlapping, 3-month seasons (1982–2010). (b) As in (a), but for the mini-NMME (solid black line) with the average correlation of the four individual models (dashed black line) and CFSv2 (gray dashed line), each model using 10 ensemble members. All forecasts were initialized in October. The parentheses contain the number of ensemble members for the models, and the lead time of the forecast season is on the abscissa.

initialized on 1 April at a lead of 3 (5) months, which should be verified against SPI<sub>GPCC</sub> in June (August) of the monsoon season, have no statistically significant skill over most of China, India, and north-central Africa. Seasonal forecasts of SPI<sub>3</sub> (Fig. 8) at a lead of 1 season exhibit skill over such areas as the southern and south-western United States in JFM and OND, much of South America in AMJ and JAS, southern Africa in JFM, and Australia in OND. These are areas that have an ENSO signal. Obviously, skillful NMME  $P$  forecasts will increase the skill of the SPI forecasts.

The evaluation of SPI forecast skill relative to baseline skill is given in Fig. 9, which shows the difference in correlation between the NMME SPI<sub>6</sub> forecasts and the baseline SPI<sub>6</sub> forecasts based on optimal persistence (Lyon et al. 2012). The optimal persistence only has the information of the seasonal cycle of precipitation and the initial conditions so it can be viewed as the

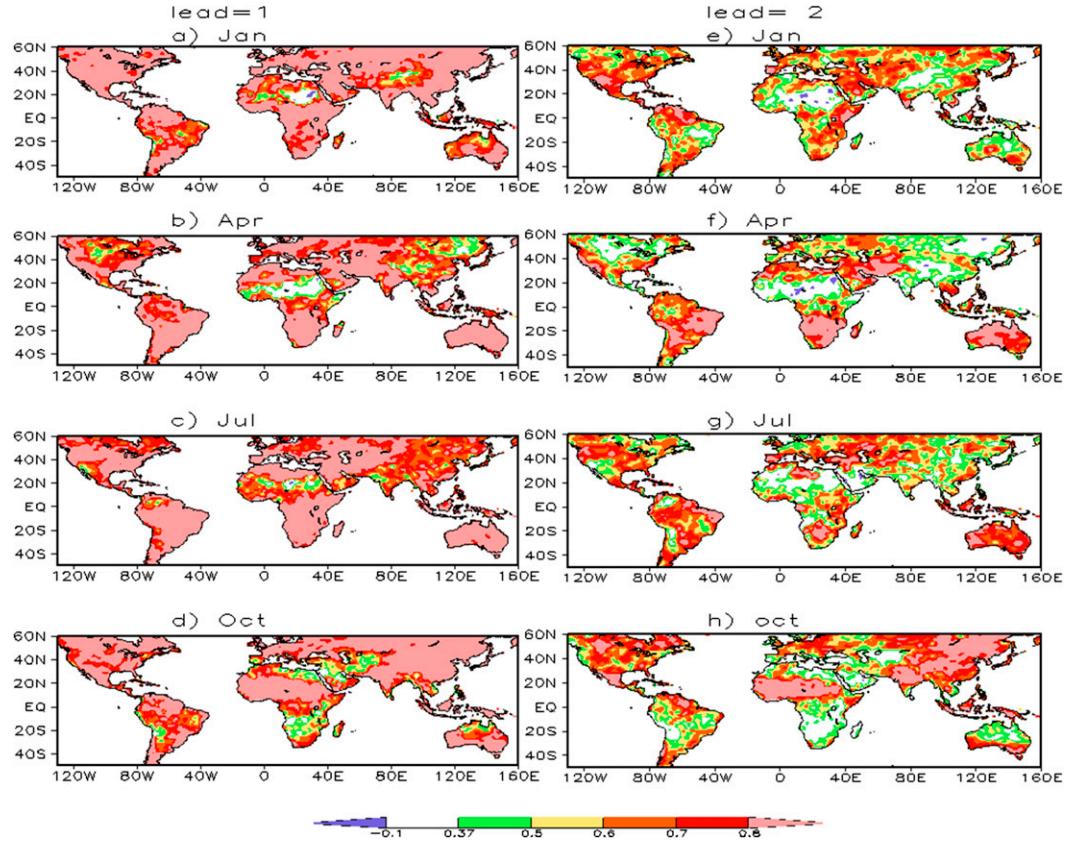


FIG. 6. As in Fig. 2, but for SPI3.

unconditional, inherent persistence of SPI. The differences between the NMME-based forecasts and the baseline forecasts will indicate whether there is additional skill obtained from the dynamical model  $P$  forecasts. For most areas, the NMME forecasts have higher skill than the baseline (orange color), but the differences often are not statistically significant at the 10% level based on the Fisher's Z test (red color). For SPI3, there are more areas where the skill of NMME-based forecasts is higher than the baseline forecasts at a lead of 2 months when persistence loses value. Overall, our results are consistent with Dutra et al. (2014a,b), namely, that it is difficult to improve on SPI forecasts that are based on climatology and persistence.

To examine the ability of the SPI6 forecasts to specifically capture severe drought conditions, we computed the hit rate, false alarm rate, and threat score for selected lead times (Fig. 10) across the period 1982–2010. As mentioned previously, for this calculation drought is defined at each grid point as occurring when the SPI6 is below  $-1.2$ . Persistence of this threshold value is not a criterion here because there will not be enough events to obtain meaningful statistics. We have pooled together all seasons in generating Fig. 10.

Looking across all land areas, it shows that the forecasts with a lead of 1 month have the highest skill, as expected. Overall, the hit rates are generally above 0.6 and the threat scores are above 0.4 for the SPI6 forecasts evaluated at a lead of 1 month. Except for the Asian monsoon regions and northern Africa, the threat scores for SPI6 are above 0.5 at 1-month lead, with a false alarm rate generally averaging about 0.3 and below. The large false alarm rate over the African and Mongolian deserts is not particularly meaningful because of the climatologically low mean and variance of  $P$  there. Overall, the SPI6 forecasts at a lead of 1 month do a good job in capturing severe droughts, that is, classified as D2 class or higher by the U.S. Drought Monitor (Svoboda et al. 2002), with the skill dropping off sharply after a lead of 1 month. For SPI6 at a lead of 2 months (and SPI3 at a lead of 1 month), the skill is lower and the forecasts are able to capture severe drought conditions only over selected areas. For the western and the southeastern United States, Central America, Europe, central Africa, and eastern Australia, the hit rate is about 0.5–0.6 and the threat scores are between 0.3 and 0.4. There is little skill exhibited by these metrics for SPI6 at a lead of 3 months even in areas where the temporal correlations between

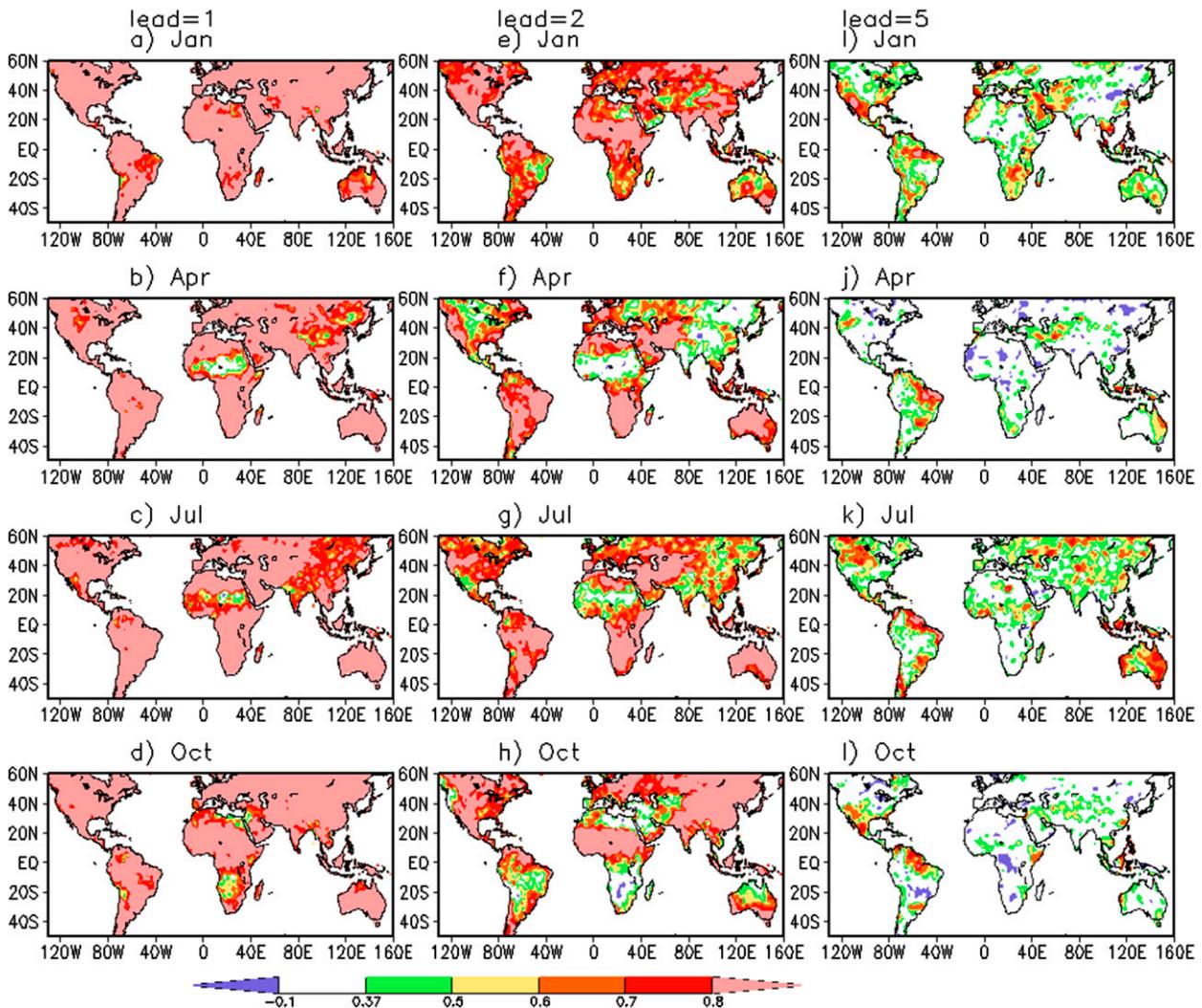


FIG. 7. NMME SPI6 forecast skill as measured by the Pearson correlation with observations from GPCC (i.e.,  $\text{SPI6}_{\text{GPCC}}$ ) for forecasts with a lead of 1 month initialized on (a) 1 Jan, (b) 1 Apr, (c) 1 Jul, and (d) 1 Oct. Values are indicated by the color bar. (e)–(h) As in (a)–(d), but for a lead of 3 months. (i)–(l) As in (a)–(d), but for a lead of 5 months. GPCC data are used for initial conditions.

forecast and observed SPI6 values is rather high (i.e., with correlation above 0.7; see Fig. 7). These results indicate that forecasting the severity of meteorological drought on seasonal time scales remains quite challenging.

#### 4. Forecast uncertainties arising from the $P$ analysis

In addition to having skillful  $P$  forecasts, a second fundamental challenge to generate reliable meteorological drought forecasts for the globe is associated with uncertainties in the  $P$  analysis used. This is especially problematic given that the  $P$  analysis needs to be made available in near-real time in order to be used in either generating a drought indicator forecast, such as the SPI, or in forcing a land surface model. In several locations of the globe, real-time station reports of  $P$  (and other variables) are not

available in near-real time, particularly over historically data-sparse areas over much of Africa and South and Central America. In such data-sparse regions, the  $P$  analysis is unlikely to accurately represent the true observed conditions, which is also a function of the analysis method used and temporal and spatial variations in the number of station reports available for any particular month.

Figure 11 illustrates the influence of the particular  $P_{\text{obs}}$  dataset used on the skill of the SPI6 forecasts. The figure shows the difference in skill of SPI6 forecasts from the NMME as measured by the Pearson correlation when using  $P$  analyses from GPCC and CPC to represent the observed conditions. The respective forecasts are both evaluated against  $\text{SPI6}_{\text{GPCC}}$ . Only areas where the difference in the temporal correlation is statistically significant at the 5% level are shaded in the figure. At

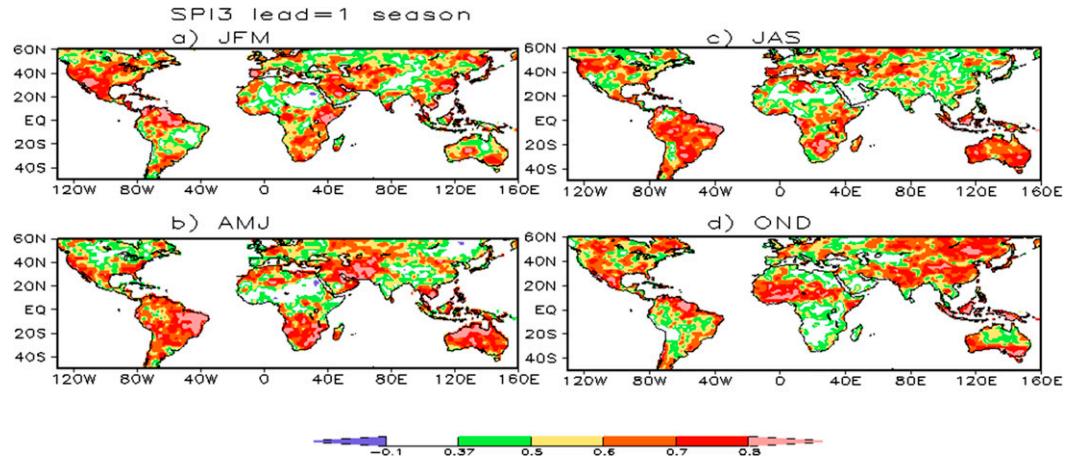


FIG. 8. As in Fig. 3, but for SPI3.

a lead of 1 month, the SPI6 has 5 months of  $P_{\text{obs}}$  and 1 month of forecast precipitation, so differences in the SPI6 forecasts at this lead essentially measure the influence of the two  $P_{\text{obs}}$  analyses on the SPI calculation. The figure shows that the differences are large over data-sparse areas such as western and central South America, central Africa, and Russia, where the correlation (skill) differences are above 0.3 in many locations. Thus, differences in the particular  $P_{\text{obs}}$  data used in describing the initial drought state results in large uncertainties in the SPI forecasts in many land areas. As the lead time increases, the difference in the correlations decreases for the two different  $P_{\text{obs}}$  datasets because the calculation of SPI6 contains fewer months of observations and a greater number of forecast  $P$  values.

To further quantify the uncertainties arising from the  $P_{\text{obs}}$  analyses when computing the SPI, we computed the root-mean-square (RMS) difference between the observed SPI3 and SPI6 when calculated using the GPCC versus CPC data over the period from January 1979 to December 2010. The results are shown in Fig. 12, where the largest differences are seen mainly in tropical land areas where the observational reports are sparse. For both SPI3 and SPI6, the differences over South America north of 35°S, central Africa, and the Asian monsoon areas and Maritime Continent are larger than 1.2. If  $-1.2$  is the threshold for severe drought, then the uncertainties are too large for drought classification. The situation is even more dire when considering that the expected RMS difference between random pairs of values taken from two independent, normally distributed variables having unit variance (as the SPI does) is expected to be  $\sqrt{2} \approx 1.4$ . In some locations, the RMS difference approaches this value. The data situation is also typically much worse for near-real-time data than for historical datasets. To highlight this, Fig. 12c shows

the data counts per day averaged over the year 2010. For places such as the United States, China, and Europe, there is fairly good spatial coverage, but for most of Central America, Africa, and even Australia, station reports are seen to be too few to have a reliable  $P_{\text{obs}}$  analysis based on them.

The challenge of developing reliable  $P$  analyses in near-real time is further demonstrated in Fig. 13, which shows the signal-to-noise ratio (ratio – 1) for a set of four historical  $P_{\text{obs}}$  analyses (from CRU, GPCC, CMAP, and GPCP) and three  $P_{\text{obs}}$  analyses available in near-real time [CPC Unified, Precipitation Reconstruction over Land (PREC/L), and CAMS\_OPI]. The figure shows this ratio computed for monthly observations for January over the period 1979–2005. This ratio is seen to be much higher for the historical datasets, having values larger than 2–3 for most land areas except north-central Africa, Peru, and portions of the western Amazon and central Asia. For the near-real-time data, the ratio values are lower, even in places such as the United States, Europe, and Australia. There are several locations where the ratio is less than unity (blue shading in the figure), indicating that the noise is greater than the signal in these areas. The lack of precipitation data inputs also influences the quality of the Global Land Data Assimilation System (GLDAS) because precipitation is the major forcing in land surface models used to obtain soil moisture and other derived variables (Nijssen et al. 2014). Overall, these results clearly illustrate a major constraint currently faced in making global forecasts of drought indicators: in many locations, the quality of the required observational data is insufficient for the task. This is true not only for the SPI prediction and any other meteorological drought indicators, but also for forcings used to drive land surface models. The development of improved observational precipitation analyses that are

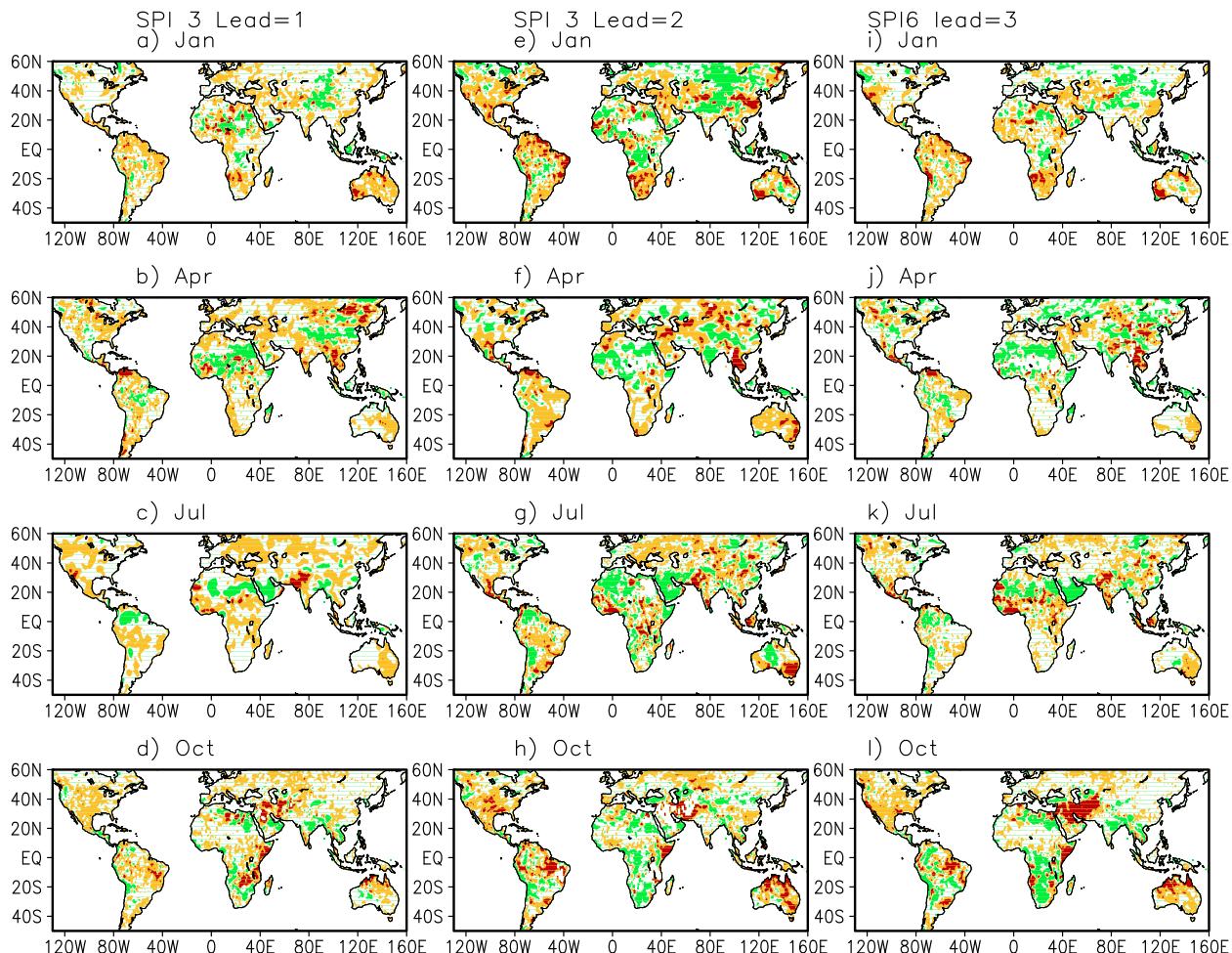


FIG. 9. Difference in forecast skill (Pearson correlation) between the NMME-based SPI3 forecasts and baseline SPI3 forecasts at a lead of 1 month for forecasts initialized on (a) 1 Jan, (b) 1 Apr, (c) 1 Jul, and (d) 1 Oct. Green shading indicates that the NMME forecasts are less skillful than the baseline forecasts. Orange shading indicates that the NMME forecasts are more skillful but not at a statistically significant level, with red indicating a statistically significant improvement at the 10% confidence level. Values less than 0.05 are omitted. (e)–(h) As in (a)–(d), but for a lead of 2 months. (i)–(l) As in (a)–(d), but for SPI6 at a lead of 3 months.

both available in near-real time and have sufficiently long historical archives is a paramount need for the development of a GDIS.

## 5. Discussion and conclusions

In this paper, we evaluated the skill of global SPI predictions that incorporate  $P$  forecasts from the NMME. This is viewed as an essential step in moving toward a forecast component to a Global Drought Information System (GDIS). The skill of the SPI3 and SPI6 forecasts was found to be seasonally and regionally dependent, but overall, SPI3 predictions at a lead of 1 month and SPI6 at leads of 1–3 months are found to have “useful” skill (temporal correlations with observations  $>0.6$ ) in many areas. The difference in skill between the

NMME SPI forecasts and a baseline forecast based on the persistence characteristics of the SPI, while positive in many areas, is largely statistically insignificant. This indicates that the skill of the SPI predictions made using the NMME comes largely from the initial conditions. The situation is similar to hydroclimate forecasts of soil moisture and runoff obtained by driving a land surface model such as the hydrologic Variable Infiltration Capacity model (VIC) using forcing derived from  $P$  and air temperature forecasts from climate models (Shukla and Lettenmaier 2011) or from the NMME (Mo and Lettenmaier 2014b). There is little difference in the skill of NMME-based soil moisture and runoff forecasts and forecasts based on an ensemble streamflow prediction (ESP) approach in which the forecast skill comes entirely from the initial conditions. Here we show that the

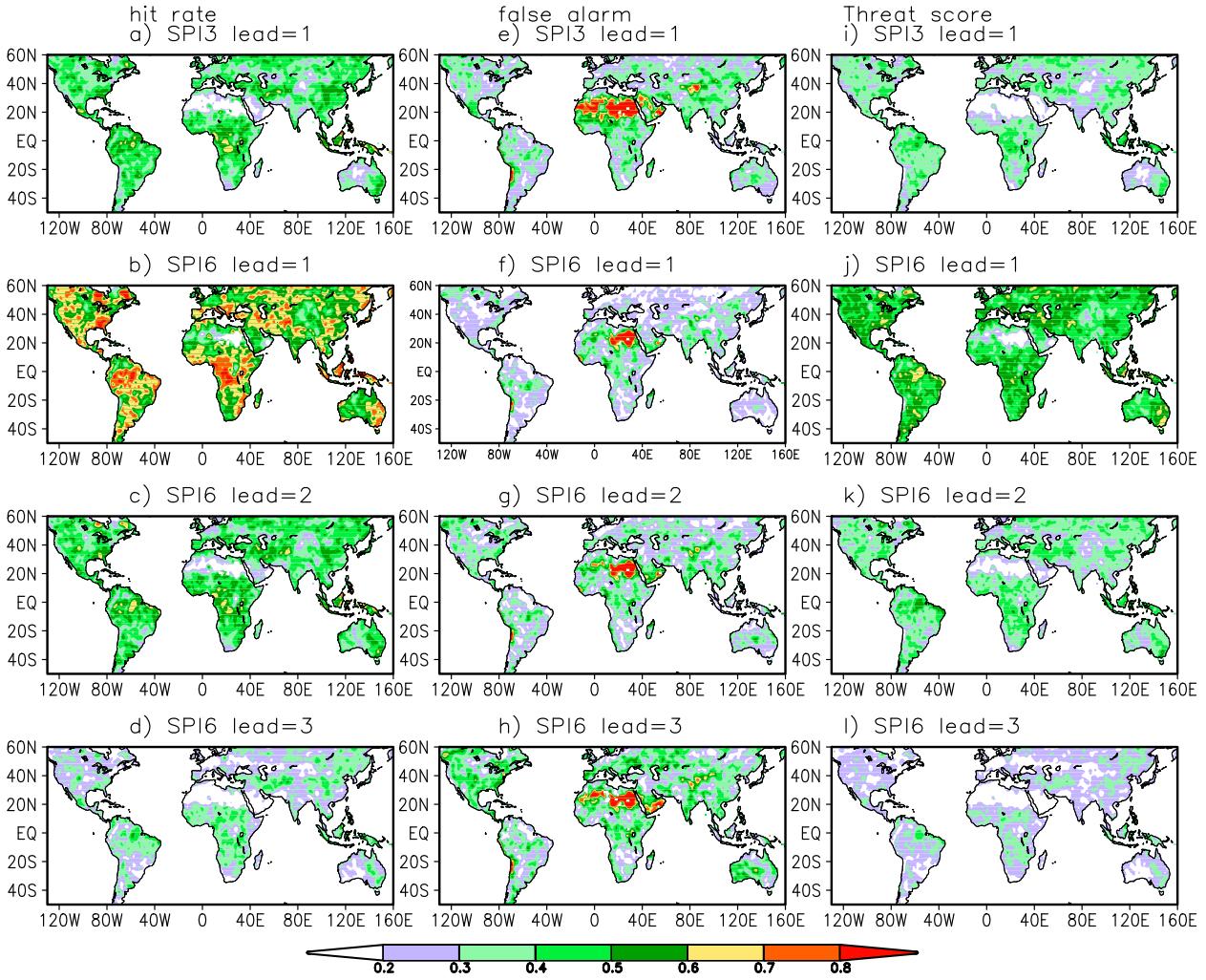


FIG. 10. Hit rate for NMME predictions of drought (defined as SPI values  $< -1.2$ ) for (a) SPI3 forecasts at a lead of 1 month verified against the SPI3<sub>GPCC</sub> and (b) SPI6 forecasts at a lead of 1 month verified against the SPI6<sub>GPCC</sub>. (c) As in (b), but for SPI6 forecasts at a lead of 2 months. (d) As in (b), but for SPI6 forecasts at a lead of 3 months. Values are indicated by the color bar. (e)–(h) As in (a)–(d), but for FAR. (i)–(l) As in (a)–(d), but for threat score.

NMME  $P$  forecast skill drops off sharply after a lead of 1 month, although seasonal  $P$  forecasts do exhibit statistically significant skill in many regions of the globe, particularly in regions influenced by ENSO. Overall, to improve drought forecasts based on the SPI (or soil moisture, or runoff from land surface models), improvements are needed in  $P$  forecasts beyond lead times of 1 month.

Additional challenges to SPI (or land surface model based) forecasts are the uncertainties inherent in the observed  $P$  analyses. For the historical period considered here (1982–2010), the  $P$  station reports were generally adequate over the United States, Europe, and Australia. In near-real time, however, the number of station reports included in the analyses drops off substantially. An example is given in Fig. 12c, which shows

the data counts per day for 2010 in near-real time. Over Africa, South America, and the tropics, there is a marked lack of reports. This clearly will have an impact on the quality of  $P_{\text{obs}}$  analysis. RMS differences in SPI3 and SPI6 values computed from two different  $P_{\text{obs}}$  analyses (Fig. 12) were shown to exceed 1.2 in many areas, particularly in the tropics. Such uncertainties are too large to reliably identify the initial drought state (e.g., SPI values of less than  $-1.2$  are commonly used to denote severe drought). The uncertainties in the  $P_{\text{obs}}$  analysis also influence soil moisture, runoff, and evaporation derived from land surface models, as variations in precipitation are the major source of variability in these quantities (Mo et al. 2012). As one way to address uncertainties in the  $P_{\text{obs}}$  analyses, Dutra et al. (2014a,b) have suggested to use ECMWF short-range forecasts in

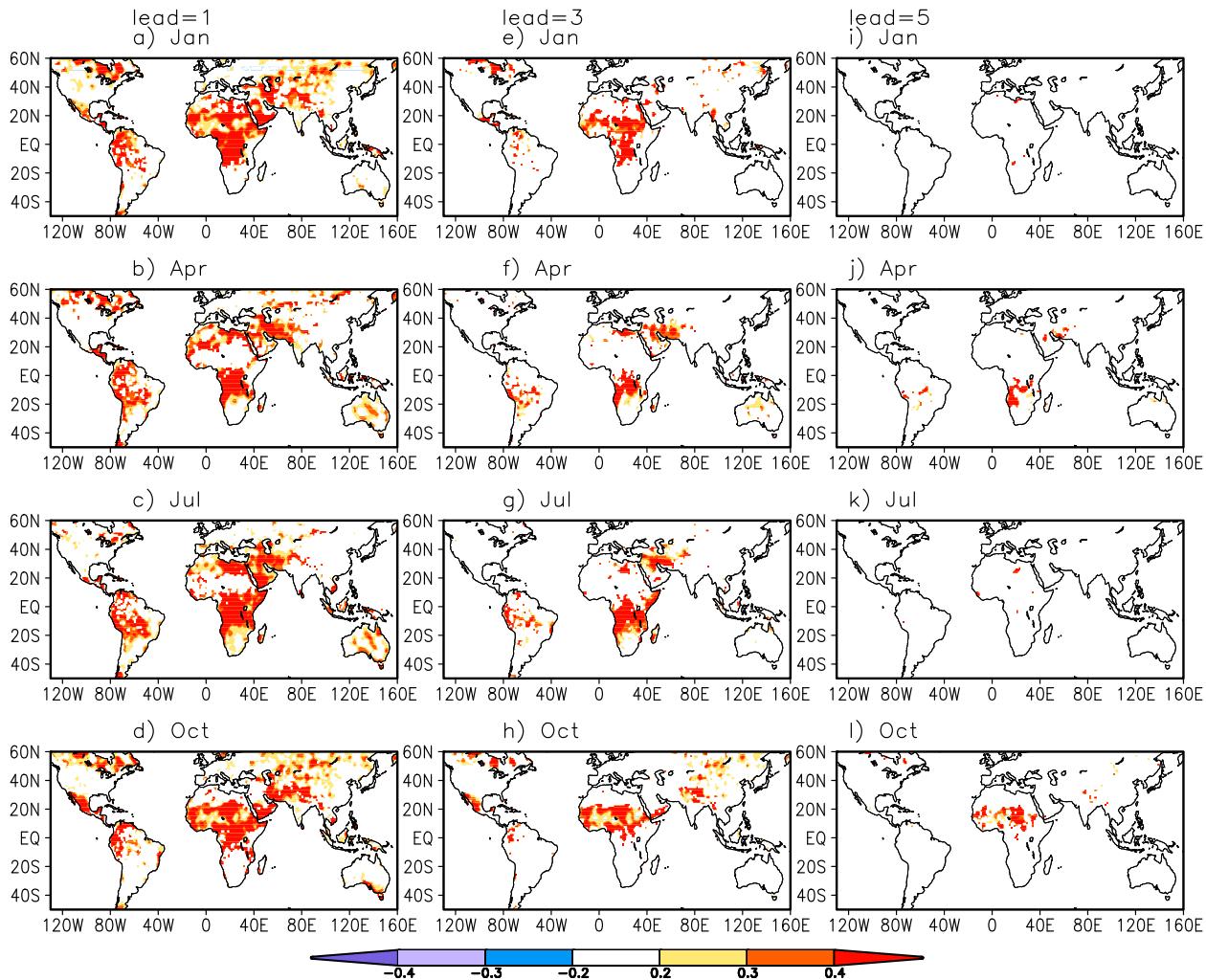


FIG. 11. Difference in forecast skill (Pearson correlation) when SPI6 predictions are based on NMME  $P$  forecasts combined with  $P_{\text{obs}}$  from GPCC and when they use  $P_{\text{obs}}$  from CPC data. SPI6<sub>GPCC</sub> data are used for verification in both cases. Differences in skill are shown for monthly forecasts at a lead of 1 month initialized on (a) 1 Jan, (b) 1 Apr, (c) 1 Jul, and (d) 1 Oct. (e)–(h) As in (a)–(d), but for forecasts at a lead of 3 months. (i)–(l) As in (a)–(d), but for forecasts at a lead of 5 months. These differences arise solely from dissimilarities in the precipitation analyses used.

combination with the GPCC data for monitoring. One possibility to improve real-time precipitation analysis is to use satellite-based observations with careful calibration to assure consistency (e.g., AghaKouchak and Nakhjiri 2012). Given uncertainties in the underlying analyses, Mo and Lettenmaier (2014a) proposed the use of multiple analyses to form a mean index for drought monitoring while also addressing its uncertainties. The latter, probabilistic method quantifies the level of uncertainty to better inform users.

Another approach is to use a probabilistic multivariate methodology where the SPI is used in combination with soil moisture or runoff indices (Hao et al. 2014). Similar to the multimodel ensemble, the multivariate approach can lead to higher skill because different

variables may have higher skill at different locations: the diversity of information can provide complimentary skill. Unfortunately, not all NMME forecasts initialized surface variables such as soil moisture and snow water equivalent properly. Because the skill of the soil moisture forecasts comes largely from initial conditions, the low skill of the soil moisture forecasts will not improve forecast skill.

Overall, to support the forecast component of GDIS, this study finds there is some modest skill derived from the NMME in making global forecasts of the SPI and that the persistence characteristics of the index allow for skillful predictions out to several months in their own right. Thus, in regions and seasons where the NMME does not provide skillful  $P$  forecasts, a statistical

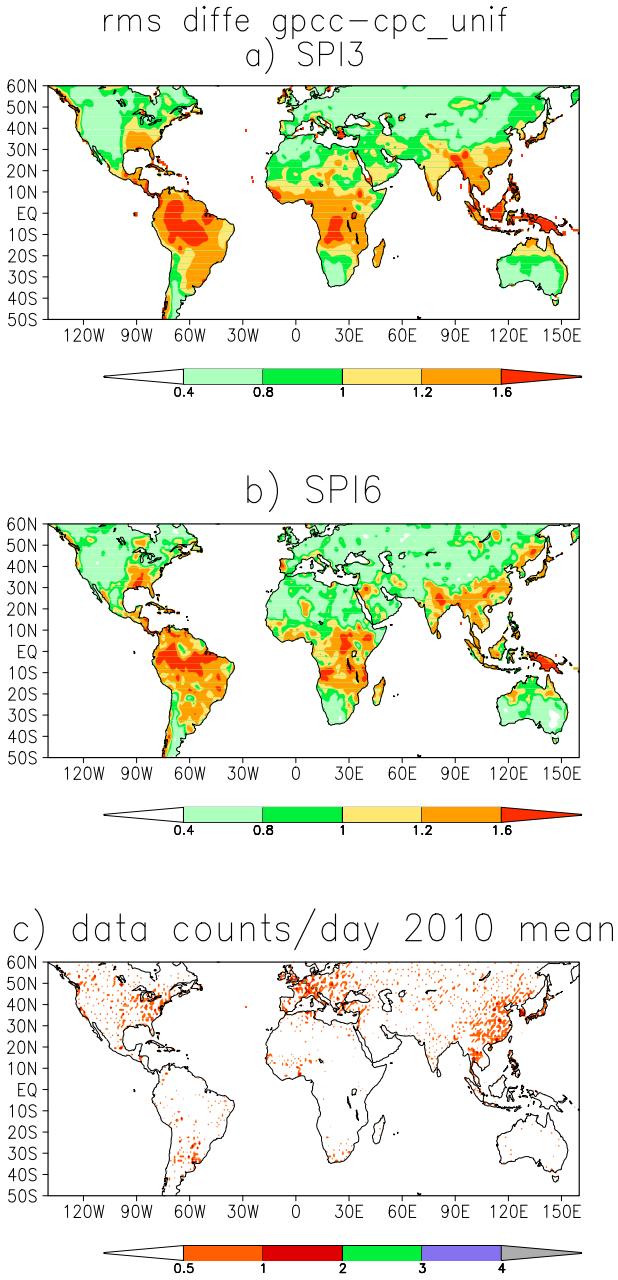


FIG. 12. The RMS difference between monthly, observed values of the SPI computed using the GPCC and CPC precipitation analyses over the period (from January 1979 to December 2010). Values (dimensionless) are given by the color bar. Differences between (a) SPI3<sub>GPCC</sub> and SPI3<sub>CPC</sub> or (b) SPI6<sub>GPCC</sub> and SPI6<sub>CPC</sub>. (c) Input data counts ( $\text{day}^{-1}$ ) for  $1^\circ \times 1^\circ$  grid boxes averaged from 1 Jan to 31 Dec 2010.

approach can be used as a fallback forecast, as described for the North American domain in Lyon et al. (2012) and Quan et al. (2012). This provides more detail to drought predictions than is typically available when climate signals such as ENSO are weak (Pulwarty and Sivakumar

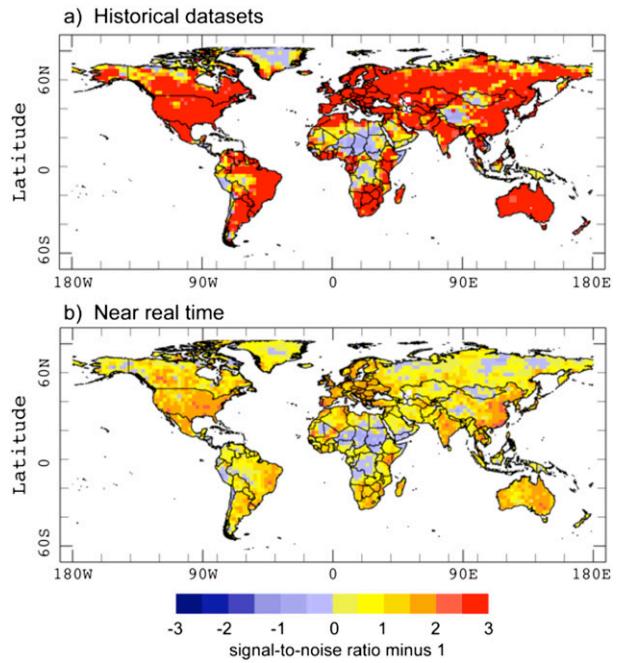


FIG. 13. Signal-to-noise ratio (ratio – 1) computed for January 1979–2005 for (a) four historical precipitation analyses (CRU, GPCC, CMAP, GPCP) and (b) three precipitation analyses available in near-real time (PREC/L, CAMS\_OPI, and CPC Unified).

2014). An advantage of using the SPI for global drought monitoring, prediction, and risk assessment is that it is already used in many countries around the globe and it is a drought index endorsed by the World Meteorological Organization (Hayes et al. 2011). Different flavors of the SPI that emphasize different time scales of variation are also desirable for early warning (Hayes et al. 2005). However, the quality of the underlying data used for drought monitoring in near-real time is seen as the largest current obstacle to more reliable drought predictions on the global scale. This issue has been recognized for some time (Pulwarty and Sivakumar 2014; Wilhite et al. 2000), and advances in drought prediction on the global scale will likely come as much from improvements in observational datasets (including inputs from satellites) as from improvements in climate model predictions in the near future. In the longer term, improved skill of monthly and seasonal precipitation forecasts will be essential.

**Acknowledgments.** The authors thank the reviewers for their helpful comments on an earlier draft of the manuscript. This work was supported by a grant from the National Oceanic and Atmospheric Administration (NOAA MAPP Award NA12OAR4310088) to IRI and GC12-351b to CPC, which is gratefully acknowledged.

## REFERENCES

- AghaKouchak, A., and N. Nakhjiri, 2012: A near real-time satellite-based global drought climate data record. *Environ. Res. Lett.*, **7**, 044037, doi:10.1088/1748-9326/7/4/044037.
- Becker, A., P. Finger, A. Meyer-Christoffer, B. Rudolf, K. Schamm, U. Schneider, and M. Ziese, 2013: A description of the global land surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901–present. *Earth Syst. Sci. Data*, **5**, 77–99, doi:10.5194/essd-5-71-2013.
- Chen, M., P. Xie, J. E. Janowiak, and P. A. Arkin, 2002: Global land precipitation. A 50-yr monthly analysis based on gauge observations. *J. Hydrometeor.*, **3**, 249–266, doi:10.1175/1525-7541(2002)003<0249:GLPAYM>2.0.CO;2.
- Dutra, E., F. Wetterhall, F. Di Giuseppe, G. Naumann, P. Barbosa, J. Vogt, W. Pozzi, and F. Pappenberger, 2014a: Global meteorological drought—Part 1: Probabilistic monitoring. *Hydrol. Earth Syst. Sci. Discuss.*, **11**, 889–917, doi:10.5194/hessd-11-889-2014.
- , and Coauthors, 2014b: Global meteorological drought—Part 2: Seasonal forecasts. *Hydrol. Earth Syst. Sci. Discuss.*, **11**, 919–944, doi:10.5194/hessd-11-919-2014.
- Guttman, N. B., 1999: Accepting the Standardized Precipitation Index: A calculation algorithm. *J. Amer. Water Resour. Assoc.*, **35**, 311–322, doi:10.1111/j.1752-1688.1999.tb03592.x.
- Hao, Z., A. AghaKouchak, N. Nakhjiri, and A. Farahmand, 2014: Global integrated drought monitoring and prediction system. *Sci. Data*, **1**, 140001, doi:10.1038/sdata.2014.1.
- Harris, I., P. D. Jones, T. J. Osborn, and D. H. Lister, 2014: Updated high-resolution grids of monthly climatic observations—The CRU TS3.10 dataset. *Int. J. Climatol.*, **34**, 623–642, doi:10.1002/joc.3711.
- Hayes, M., M. Svoboda, D. LeComte, K. Redmond, and P. Pasteris, 2005: Drought monitoring: New tools for the 21st century. *Drought and Water Crises: Science, Technology, and Management Issues*, D. A. Wilhite, Ed., CRC Press, 53–69.
- , —, N. Wall, and M. Widhalm, 2011: The Lincoln Declaration on Drought Indices. *Bull. Amer. Meteor. Soc.*, **92**, 485–488, doi:10.1175/2010BAMS3103.1.
- Huffman, G. J., R. F. Adler, D. T. Bolvin, and G. Gu, 2009: Improving the global precipitation record: GPCC version 2.1. *Geophys. Res. Lett.*, **36**, L17808, doi:10.1029/2009GL040000.
- Janowiak, J. E., and P. Xie, 1999: CAMS-OPI: A global satellite–rain gauge merged product for real-time precipitation monitoring applications. *J. Climate*, **12**, 3335–3342, doi:10.1175/1520-0442(1999)012<3335:COAGSR>2.0.CO;2.
- Kirtman, B., and Coauthors, 2013: The North American Multimodel Ensemble: Phase-1 seasonal-to-interannual prediction; Phase-2 toward developing intraseasonal prediction. *Bull. Amer. Meteor. Soc.*, **95**, 585–601, doi:10.1175/BAMS-D-12-00050.1.
- Lyon, B., M. A. Bell, M. K. Tippett, A. Kumar, M. P. Hoerling, X.-W. Quan, and H. Wang, 2012: Baseline probabilities for the seasonal prediction of meteorological drought. *J. Appl. Meteor. Climatol.*, **51**, 1222–1237, doi:10.1175/JAMC-D-11-0132.1.
- 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.
- Mo, K. C., 2011: Drought onset and recovery over the United States. *J. Geophys. Res.*, **116**, D20106, doi:10.1029/2011JD016168.
- , and D. P. Lettenmaier, 2014a: Objective drought classification using multiple land surface models. *J. Hydrometeor.*, **15**, 990–1010, doi:10.1175/JHM-D-13-071.1.
- , and —, 2014b: Hydrologic prediction over the conterminous United States using the National Multi-Model Ensemble. *J. Hydrometeor.*, **15**, 1457–1472, doi:10.1175/JHM-D-13-0197.1.
- , L. Chen, S. Shukla, T. Bohn, and P. D. Lettenmaier, 2012: Uncertainties in North American Land Data Assimilation systems over the contiguous United States. *J. Hydrometeor.*, **13**, 996–1009, doi:10.1175/JHM-D-11-0132.1.
- Nijssen, B., and Coauthors, 2014: A prototype global drought information system based on multiple land surface models. *J. Hydrometeor.*, **15**, 1661–1676, doi:10.1175/JHM-D-13-090.1.
- Pozzi, W., and Coauthors, 2013: Toward global drought early warning capability: Expanding international cooperation for the development of a framework for monitoring and forecasting. *Bull. Amer. Meteor. Soc.*, **94**, 776–785, doi:10.1175/BAMS-D-11-00176.1.
- Pulwarty, R. S., and M. Sivakumar, 2014: Information systems in a changing climate: Early warnings and drought risk management. *Wea. Climate Extremes*, **3**, 14–21, doi:10.1016/j.wace.2014.03.005.
- Quan, X.-W., M. P. Hoerling, B. Lyon, A. Kumar, M. A. Bell, M. K. Tippett, and H. Wang, 2012: Prospects for dynamical prediction of meteorological drought. *J. Appl. Meteor. Climatol.*, **51**, 1238–1252, doi:10.1175/JAMC-D-11-0194.1.
- Rudolf, B., and U. Schneider, 2005: Calculation of gridded precipitation data for global land surface using in situ gauge observations. *Proc. Second Workshop of the Int. Precipitation Working Group*, Monterey, CA, Naval Research Laboratory, 231–247.
- Saha, S., and Coauthors, 2014: The NCEP Climate Forecast System version 2. *J. Climate*, **27**, 2185–2208, doi:10.1175/JCLI-D-12-00823.1.
- Shukla, S., and D. P. Lettenmaier, 2011: Seasonal hydrologic prediction in the United States: Understanding the role of initial conditions and seasonal climate forecast skill. *Hydro. Earth Syst. Sci.*, **15**, 3529–3538, doi:10.5194/hess-15-3529-2011.
- Svoboda, M., and Coauthors, 2002: The Drought Monitor. *Bull. Amer. Meteor. Soc.*, **83**, 1181–1190, doi:10.1175/1520-0477(2002)083<1181:TDM>2.3.CO;2.
- Vicente-Serrano, S. M., S. Beguería, and J. I. López-Moreno, 2010: A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index. *J. Climate*, **23**, 1696–1718, doi:10.1175/2009JCLI2909.1.
- WCRP, 2012: WCRP Global Drought Information System (GDIS) Workshop. Final Rep., 17 pp. [Available online at [http://eprints.soton.ac.uk/342410/1/GDIS\\_Report\\_final.pdf](http://eprints.soton.ac.uk/342410/1/GDIS_Report_final.pdf).]
- Wilhite, D. A., M. V. K. Sivakumar, and D. A. Wood, Eds., 2000: Early warning systems for drought preparedness and drought management. WMO/TD-1037, WMO, Geneva, Switzerland, 185 pp. [Available online at [www.wamis.org/agm/pubs/agm2/agm02.pdf](http://www.wamis.org/agm/pubs/agm2/agm02.pdf).]
- Xie, P., and P. A. Arkin, 1997: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates and numerical model outputs. *Bull. Amer. Meteor. Soc.*, **78**, 2539–2558, doi:10.1175/1520-0477(1997)078<2539:GPAYMA>2.0.CO;2.
- Yoon, J. H., K. C. Mo, and E. F. Wood, 2012: Dynamic model based seasonal prediction of meteorological drought over the contiguous United States. *J. Hydrometeor.*, **13**, 463–482, doi:10.1175/JHM-D-11-038.1.
- Yuan, X., and E. F. Wood, 2013: Multimodel seasonal forecasting of global drought onset. *Geophys. Res. Lett.*, **40**, 4900–4905, doi:10.1002/grl.50949.