AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE
10.1002/2015JD023185

Key Points:
- Seasonal drought predictability and forecast skill are positively correlated
- ENSO controls the relation between drought predictability and forecast skill
- Predictability may provide a measure for selecting models for drought forecast

Supporting Information:
- Text S1 and Figure S1

Correspondence to:
X. Yuan,
yuanxing@tea.ac.cn

Citation:
Ma, F., X. Yuan, and A. Ye (2015), Seasonal drought predictability and forecast skill over China, *J. Geophys. Res. Atmos.*, *120*, 8264–8275, doi:10.1002/2015JD023185.

Received 29 JAN 2015
Accepted 29 JUL 2015
Accepted article online 2 AUG 2015
Published online 26 AUG 2015

Seasonal drought predictability and forecast skill over China

Feng Ma[1,2], Xing Yuan[1], and Aizhong Ye[2]

[1]RCE-TEA, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China, [2]State Key Laboratory of Earth Surface Processes and Resource Ecology, College of Global Change and Earth System Science, Beijing Normal University, Beijing, China

Abstract Under a changing environment, seasonal droughts have been exacerbated with devastating impacts. However, the understanding of drought mechanism and predictability is limited. Based on the hindcasts from multiple climate models, the predictability and forecast skill for drought over China are investigated. The 3 month standardized precipitation index is used as the drought index, and the predictability is quantified by using a perfect model assumption. Ensemble hindcasts from multiple climate models are assessed individually, and the grand multimodel ensemble is also evaluated. Drought forecast skill for model ensemble mean is higher than individual ensemble members, and North American Multimodel Ensemble grand ensemble performs the best. Predictability is higher than forecast skill, indicating the room for improving drought forecast. Drought predictability and forecast skill are positively correlated in general, but they vary depending on seasons, regions, and forecast leads. Higher drought predictability and forecast skill are found over regimes where ENSO has significant impact. For the ENSO-affected regimes, both drought predictability and forecast skill in ENSO years are higher than that in neutral years. This study suggests that predictability not only provides a measure for selecting climate models for ensemble drought forecast in ENSO-affected regimes but also serves as an indicator for forecast skill especially when in situ and/or remote sensing measurements for the hindcast verifications are considered unreliable.

1. Introduction

Drought is a naturally occurring extreme climate event characterized by below-normal precipitation over a period of months to years. In addition to climate change, the demand for water has increased due to the growth of population and expansion of agricultural and industrial sectors, which exacerbates the devastating impact of drought in many parts of the world. Large-scale droughts also occur frequently in China, with an increasing trend since the 1950s [*Wang et al.*, 2011; *Zhang et al.*, 2014]. The 2013 summertime "flash" drought (with heatwave) damaged over 2×10^6 ha of crops in Guizhou and Hunan provinces in southern China [*Yuan et al.*, 2015], and the direct economic loss from droughts over China during 2014 is about 91 billion RMB (about 14.5 billion U.S. dollars; http://www.chinaam.com.cn). Therefore, reliable drought forecasting several months in advance is of primary importance for water resources and agricultural management by increasing the preparedness.

In fact, dynamical seasonal forecasting systems that are based on coupled atmosphere-ocean-land general circulation models (CGCMs) have been widely used for drought forecasting in recent years [*Luo and Wood*, 2007; *Dutra et al.*, 2012]. However, less than 30% of the global drought onsets can be detected by climate forecast models [*Yuan and Wood*, 2013]. Drought forecast skill tends to be limited, especially over the regions with less impact from El Niño–Southern Oscillation (ENSO). Therefore, we are particularly interested in question like how much better could we do for drought forecasting? Or, what is the potential predictability of drought? Basically, the predictability characterizes the "ability to be predicted" while the forecast skill characterizes "the ability to predict it" [*Boer et al.*, 2013]. Predictability is often investigated using a "perfect model" assumption in which both the forecast and the proxy observation come from the same world [*Koster et al.*, 2000; *Luo and Wood*, 2006; *Boer et al.*, 2013; *Becker et al.*, 2014]. Predictability is useful in understanding the relationships between initialization (measurement) uncertainty, parameterization uncertainty, and model result. Cross model comparisons/predictability analysis also helps in assessing the model uncertainty.

Although climate predictability and its relationship with forecast skill have been investigated in the literature [*Boer et al.*, 2013; *Holland et al.*, 2013; *Younas and Tang*, 2013], the predictability of climate extremes (e.g., drought) has received less attention. Generally, the predictability is found to be higher than the forecast skill, and the gap between them represents the room for improvement [*Younas and Tang*, 2013; *Becker et al.*, 2014]. Nevertheless, *Kumar et al.* [2014] found examples to the contrary and argued that there is not necessarily a relationship between predictability and forecast skill. The inconsistency from previous studies suggests that predictability and forecast skill may depend on the target climate models, study regions, and climate variables. Furthermore, whether the seasonal drought predictability is associated with ENSO and how do models represent such association also need careful investigation.

Droughts are generally classified into four categories, including meteorological, hydrological, agricultural, and socioeconomic droughts [*Mishra and Singh*, 2010]. Meteorological drought is defined as a lack of precipitation over a region for a period of time. Hydrological drought is related to a period with inadequate surface and subsurface water resources for a given water resource management system. Agricultural drought usually refers to a period with declining soil moisture and consequent crop failure. For quantifying meteorological drought, a number of indices have been developed, each with its own strengths and weaknesses. They include the standardized precipitation index (SPI) [*McKee et al.*, 1993, 1995], Palmer drought severity index [*Palmer*, 1965], deciles [*Gibbs and Maher*, 1967], and rainfall anomaly index [*Van Rooy*, 1965]. As recommended by the World Meteorological Organization, the most commonly used indicator of meteorological drought is the SPI [*McKee et al.*, 1993; *Hayes et al.*, 2011]. Therefore, based on the hindcasts from multiple CGCMs participating in the North American Multimodel Ensemble (NMME) project [*Kirtman et al.*, 2014], we calculate the SPI to investigate drought predictability and forecast skill over China in this paper. We focus on (1) drought predictability, forecast skill, and their relationship over China and its subregions and (2) their potential connection with the ENSO, which is considered as a major source of seasonal climate predictability [*Doblas-Reyes et al.*, 2013].

2. Data and Method

2.1. Model Forecasts and Observations

Recently, the North American Multimodel Ensemble (NMME) project has been launched for intraseasonal to interannual prediction [*Kirtman et al.*, 2014]. NMME data have been widely used for ensemble analysis and hydroclimate forecasting [*Yuan and Wood*, 2012, 2013; *Becker et al.*, 2014; *Delsole et al.*, 2014; *Mo and Lettenmaier*, 2014; *Tian et al.*, 2014; *Ma et al.*, 2015]. In this study, six models, which are being implemented to produce real-time seasonal forecasts, are used for drought analysis. They are developed at University of Miami (Rosenstiel School of Marine and Atmospheric Science/Community Climate System Model, Version 3 (CCSM3)), Geophysical Fluid Dynamics Laboratory (CM2.2), National Aeronautics and Space Administration (NASA/Goddard Earth Observing System Model, Version 5 (GEOS5)), National Centers for Environmental Prediction (NCEP/Climate Forecast System Version 2 (CFSv2)), and Canadian Meteorological Centre (CMC/CanCM3, CanCM4). The overlapping period for the hindcasts is 1982–2010, and the data sets have 1° resolution with forecast leads of 8–12 months. The number of ensemble members for each model varies from 6 to 24. Hindcasts are archived by the International Research Institute for Climate and Society (IRI) (http://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/).

The observed daily precipitation data were obtained from China Meteorological Information Center. It was resampled from 2474 meteorological observation stations in China by using the optimum interpolation method [see *Xie et al.*, 2007] based on "climate background fields." Quality control was carried out to ensure that data meet requirements. In this study, the monthly data sets at 0.5° resolution are regridded to 1°to match the resolution of NMME hindcast data. The sea surface temperature (SST) in the Niño 3.4 region (5°N–5°S, 120°W–170°W) for the verification is from ERSSTv3b data [*Smith et al.*, 2008].

To assess seasonal drought forecasting, a 3 month standardized precipitation index (SPI3) is calculated from both model hindcasts and observations as the drought index [*McKee et al.*, 1993]. The frequency distribution of precipitation is described using a two-parameter gamma probability density function (Text S1 in the supporting information). Given an SPI3 time series, the drought condition is defined as the SPI3 below a specified threshold (−0.8 in this study according to *Svoboda et al.* [2002]).

2.2. Verification Measures

One of the measures for assessing predictability and forecast skill is the anomaly correlation (AC) [*Wilks*, 1995]. The anomaly correlation is defined as

$$\mathrm{AC} = \frac{\sum_s \sum_j X'(s,j,m,l) Y'(s,j,m,l)}{\left[\sum_s \sum_j X'(s,j,m,l)^2 \cdot \sum_s \sum_j Y'(s,j,m,l)^2\right]^{1/2}}, \tag{1}$$

where $X'(s, j, m, l)$ is the SPI3/drought forecast, and $Y'(s, j, m, l)$ is the SPI3/drought verification; s, j, m, l are space (grid point), year (1982–2010), target season, and forecast lead (only for model forecasts), respectively. Here we focus on the forecasts during lead-0 and lead-1 seasons for SPI3 and lead-0 seasons for drought. For example, the lead-0 drought forecast starting from February is February–April; the lead-1 SPI3 forecast starting from February is March–May.

To define the SPI3/drought predictability, AC of one model's ensemble mean SPI3/drought (based on N-1 members, X in this case) against the remaining member (the proxy observation, Y in this case) is calculated first, then this procedure is repeated N times by switching the remaining member, and the average AC is used as a measure of predictability. The predictability suggests the model's capability in predicting itself. In addition, the predictability across models, i.e., the AC of one model's ensemble mean (based on all N members, X in this case) verified against individual members of another model (Y in this case), is also briefly mentioned. The AC of the SPI3/drought ensemble mean of an individual model against the observed SPI3/drought values is used as measures of the forecast skill.

For AC, zero means no correlation between X and Y, higher than zero means positive correlation between them. In this study, we use t test [*Wilks*, 1995] to test whether the correlation is statistically significant. Under the null hypothesis H_0, $\rho = 0$, and

$$t = \frac{(r-\rho)\sqrt{n-2}}{\sqrt{1-r^2}}, \tag{2}$$

where r is the AC, n is sample size, and t is the test statistics. A threshold of AC can be computed given a significance level.

In addition, the Brier Score (BS) [*Wilks*, 1995] that is widely implemented in assessing the probabilistic drought forecasting [*Yuan et al.*, 2013] is also used in this study for predictability and skill analysis. The BS is essentially the mean square error of the probability forecasts, considering that the observation is $o_1 = 1$ if the drought occurs and that the observation is $o_2 = 0$ if the drought does not occur. It is computed as

$$\mathrm{BS} = \frac{1}{n}\sum_{k=1}^{n}(y_k - o_k)^2, \tag{3}$$

where y_k is the forecast probabilities and the index k denotes the number of forecast-event pairs. Less accurate forecasts receive higher BS, with perfect forecast exhibiting BS = 0. For BS of drought predictability, the o_k is the ensemble mean of each model instead.

Not all forecasts for droughts come true, and not all observed droughts can be predicted. Commonly, a forecast for drought event can be summed up using a 2×2 contingency table [*Wilks*, 1995]. Possible outcomes are counted: drought is forecast and observed (a), drought is forecast but is not observed (b), observed drought that is not forecast (c), and correct negative (d). Three metrics are used here, including the false alarm rate (FAR), hit rate (HR), and Equitable Threat Score (ETS), which are computed as

$$\mathrm{FAR} = \frac{b}{a+b}, \mathrm{HR} = \frac{a}{a+c}, \mathrm{ETS} = \frac{a-r}{a+b+c-r}, \tag{4}$$

where r is random forecasts as the reference and $r = \frac{(a+c)(a+b)}{a+b+c+d}$. HR and ETS of one would mean perfect forecast. Similar to the AC analysis above, for the drought predictability based on FAR, HR, and ETS, the forecast and observation are replaced by forecasts from each ensemble member and the ensemble mean, respectively.

Table 1. SPI3 Predictability and Forecast Skill Over China, Averaged Over the 12 Lead-1 Seasons[a]

	CanCM3	CanCM4	CCSM3	CM2.2	GEOS5	CFSv2	Obs (Ens AC)
CanCM3 Ens	**0.13**	0.10	0.04	0.08	0.04	0.04	*0.05*
CanCM4 Ens	0.10	**0.19**	0.05	0.10	0.05	0.05	*0.07*
CCSM3 Ens	0.03	0.05	**0.10**	0.04	0.02	0.02	*0.03*
CM2.2 Ens	0.08	0.10	0.04	**0.17**	0.06	0.05	*0.09*
GEOS5 Ens	0.04	0.06	0.03	0.07	**0.13**	0.05	*0.05*
CFSv2 Ens	0.05	0.07	0.02	0.07	0.05	**0.18**	*0.07*
Singmem AC	0.02	0.03	0.02	0.04	0.02	0.02	***NMME = 0.12***

[a]The bold diagonal shows the SPI3 predictability (anomaly correlation (AC)) within single models. The off-diagonal elements show the predictability across models. As the Ens is the prediction and the single member is the verification, each row reads horizontally. For example, the row labeled "CanCM3 Ens" shows predictability for itself, and predictability verified by CanCM4, CCSM3, CM2.2, GEOS5, and CFSv2, and so on. Italic column: forecast skill (AC) of the model ensemble mean verified against observations; The cell in bold and italic: forecast skill (AC) of the NMME 6-model grand mean verified against observations. Bottom row: average skill (AC) of a single model member, verified against observations.

2.3. ENSO Phase

To determine the ENSO phase of a particular winter (summer), we divide the entire 28 years from 1982 to 2009 into three categories: El Niño December-January-February (DJFs) (June-July-August (JJAs)), La Niña DJFs (JJAs), and neutral DJFs (JJAs). According to the Niño 3.4 index [*Smith et al.*, 2008], seven El Niño DJFs (1982–1983, 1986–1987, 1987–1988, 1991–1992, 1994–1995, 1997–1998, and 2002–2003), five La Niña DJFs (1984–1985, 1988–1989, 1998–1999, 1999–2000, and 2007-2008), seven El Niño JJAs (i.e., 1982, 1987, 1991, 1997, 2002, 2004, and 2009), and five La Niña JJAs (i.e., 1985, 1988, 1998, 1999, and 2000) are selected. The remaining years for winter (summer) are the neutral DJFs (JJAs).

3. Results

The model performance is assessed in two aspects: (1) the predictability and forecast skill using the entire 28 year SPI3 records and (2) only if an extreme event (drought) occurs (either in the observation or the forecast).

Table 1 presents the results for SPI3 predictability and forecast skill at lead-1 seasons, based on all gridded data in China. Each AC is calculated for one season, and then averaged over all 12 target seasons. For each season, there are 28 year hindcasts in a large domain (955 grids over China), which results in a sample size (n) of 26740, so a correlation of 0.03 is statistically significant. SPI3 forecast skill for the model ensemble mean ranges from 0.03 to 0.09. They are small values, but are significant ($P < 0.05$) given large spatial samples. The range of forecast skill for the single ensemble members is 0.02–0.04, unsurprisingly, lower than the skill for ensemble means. In addition, the models (i.e., CM2.2 and CanCM4) with highest ensemble mean forecast skill also have the highest single-member skill. The forecast skill for NMME 6-model grand ensemble is 0.12, which is higher than any individual models.

The SPI3 predictability within a single model varies from 0.13 to 0.18. This is higher than the forecast skill, indicating that there is room for improving the SPI3 forecast. The models with higher predictability also have higher forecast skill (e.g., CanCM4, CFSv2, and CMC2.2). The predictability values across different models vary from 0.02 to 0.10, which are closer to the values of actual forecast skill. It can be believed that models predict themselves better than they predict other models (or reality) [*Becker et al.*, 2014]. SPI3 predictability and forecast skill are low for CCSM3, and the CCSM3 is also poorly predicted by other models. The NMME/CCSM4 is being used to produce hindcasts and real-time forecasts, and it is expected to increase the skill with better physics and data assimilation. CanCM3 and CanCM4 can predict each other well, but it may suggest similarity between the models developed at the same institution [*Yuan and Wood*, 2012].

Furthermore, the results of SPI3 predictability and forecast skill for lead-0 seasons are shown in Table 2. Unlike the results at lead-1 seasons, the GEOS5 and CFSv2 models have low potential predictability but with high forecast skill at lead-0 seasons. In fact, most models are initialized on the date closest to the beginning of the month, while CFSv2 and GEOS5 (part of ensemble members) are initialized every 5 days [*Kirtman et al.*, 2014]. The latter realization method leads to more reliable and skillful ensembles [*Yuan and Wood*, 2013] especially at shorter forecast lead. However, the ensemble spread of the latter method is usually higher than

Table 2. SPI3 Predictability and Forecast Skill Over China, Averaged Over the 12 Lead-0 Seasons[a]

	CanCM3	CanCM4	CCSM3	CM2.2	GEOS5	CFSv2	Obs (Ens AC)
CanCM3 Ens	**0.35**	0.26	0.04	0.11	0.10	0.08	*0.12*
CanCM4 Ens	0.25	**0.39**	0.04	0.13	0.12	0.09	*0.14*
CCSM3 Ens	0.05	0.04	**0.11**	0.04	0.03	0.03	*0.05*
CM2.2 Ens	0.12	0.15	0.04	**0.31**	0.10	0.08	*0.13*
GEOS5 Ens	0.13	0.15	0.03	0.11	**0.24**	0.11	*0.15*
CFSv2 Ens	0.12	0.14	0.03	0.11	0.14	**0.24**	*0.16*
Singmem AC	0.07	0.08	0.02	0.06	0.07	0.05	***NMME = 0.22***

[a]The same as Table 1 but for SPI3 over lead-0 seasons.

the former; thus, the predictability is lower. By excluding CFSv2 and GEOS5, models with higher drought predictability also have higher forecast skill at lead-0 seasons. To testify the sensitivity and ensure consistency, the results of SPI6-based predictability and forecast skill over China are also analyzed, and they are consistent with those based on the SPI3 at the same lead (not shown).

To further understand the relationship between predictability and forecast skill, as well as their seasonal differences, the results for each model and each season are shown in Figure 1. Models perform differently in a single season. The average SPI3 predictability and forecast skill at lead-1 seasons are highest during winter (0.17 and 0.08) and lowest during autumn (0.14 and 0.05) (Figure 1a). In general, predictability is higher than forecast skill, with only a few exceptions (Figures 1a and 1b). The Pearson's correlation coefficients *R* between SPI3 predictability and forecast skill are greater than zero during each individual season and all four seasons, suggesting a positive relationship between SPI3 predictability and forecast skill (Figures 1a and 1b). The Pearson's correlation coefficients *R* between SPI3 predictability and forecast skill at lead-1 (lead-0) seasons are 0.59 (0.73), 0.49 (0.42), and 0.43 (0.48) during autumn, summer, and spring, respectively, which are statistically significant with $P < 0.1$. At lead-1 seasons, the SPI3 predictability and forecast skill are the highest in winter due to the influence of ENSO, but it is not clear why the relationship between predictability and forecast skill appears to be the weakest (Figure 1a).

Previous studies [e.g., *Trenberth and Caron*, 2000] have revealed that seasonal forecast predictability originates from ocean, and the strongest signal is ENSO. Therefore, we are interested in how ENSO influences drought index (SPI3) in terms of predictability and forecast skill over China. The correlation analysis is carried out over 17 river basins (Figure S1 in the supporting information) in China, where the basins have been divided according to their hydroclimate characteristics [*Lang et al.*, 2014]. During December-January-February (DJF), higher

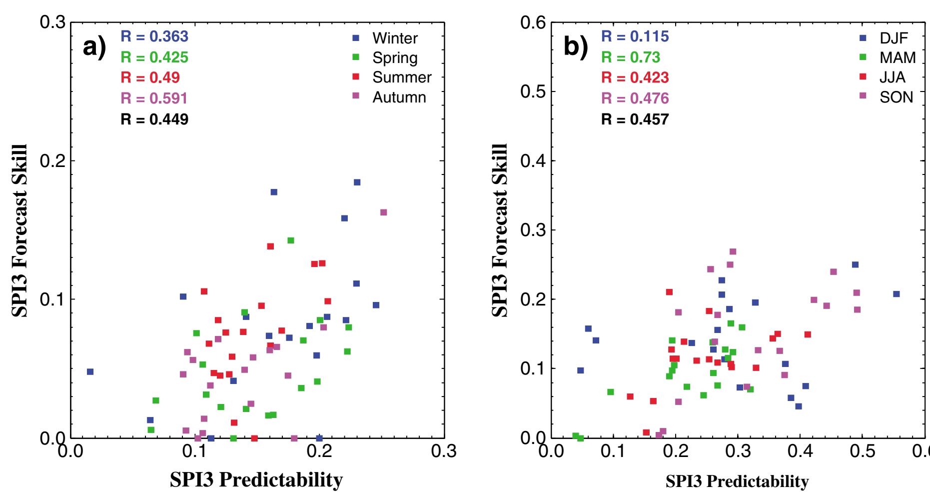

Figure 1. SPI3 predictability (*x* axis) versus forecast skill (*y* axis) in terms of Anomaly Correlation (AC) at (a) lead-1 seasons and (b) lead-0 seasons. Colors represent seasons: blue for boreal winter seasons (November–January, December–February, and January–March), green for spring seasons (February-April, March-May, and April–June), red for summer seasons (May–July, June–August, and July–September), and orange for autumn seasons (August–October, September–November, and October–December).

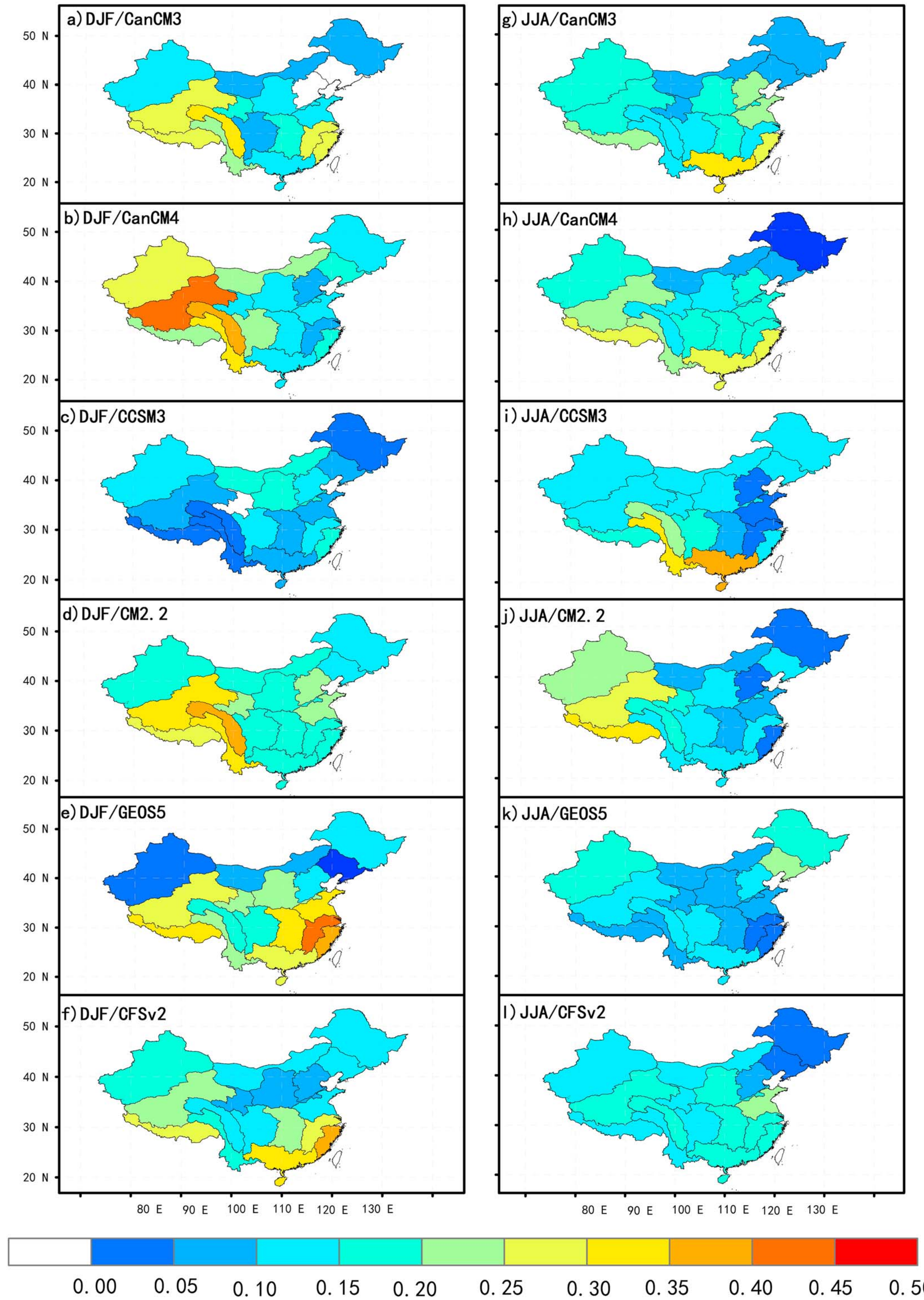

Figure 2. SPI3 predictability in terms of anomaly correlation (AC) at lead-1 seasons for 17 hydroclimate regions in China during DJF and JJA.

predictability is found over southwest, Pearl River or southeast rivers for most models (Figures 2a–2f), while higher forecast skill of most models is located at eastern or southeastern China (Figures 3a–3f). In addition, the correlations between SPI3 and Nino 3.4 SST anomaly over southeastern China are also high for both observation (Figure 4a) and most climate forecast models (e.g., CanCM3, CCSM3, GEOS5, and CFSv2 in

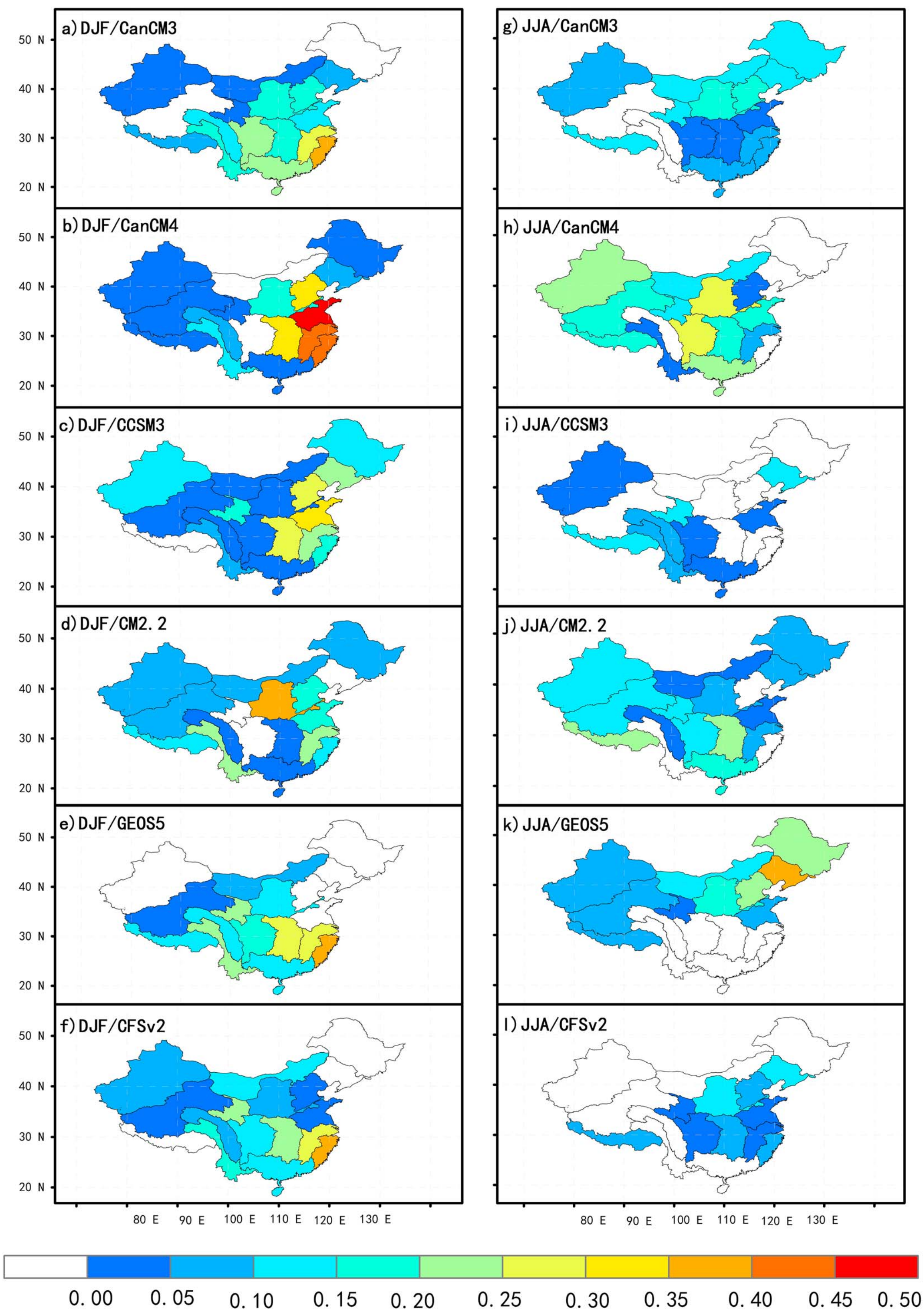

Figure 3. SPI3 forecast skill in terms of anomaly correlation (AC) at lead-1 seasons for 17 hydroclimate regions in China during DJF and JJA.

Figures 4b, 4d, 4f, and 4g). This phenomenon suggests that precipitation over southeastern China is largely affected by ENSO during winter time, and climate models capture such precipitation-ENSO SST relationship quite well. The predictability and forecast skill are high in CanCM4, GEOS5, and CFSv2 during winter (Figures 2b, 2e, and 2f and Figures 3b, 3e, and 3f). The correlations between SPI3 and ENSO SST are also strong in these

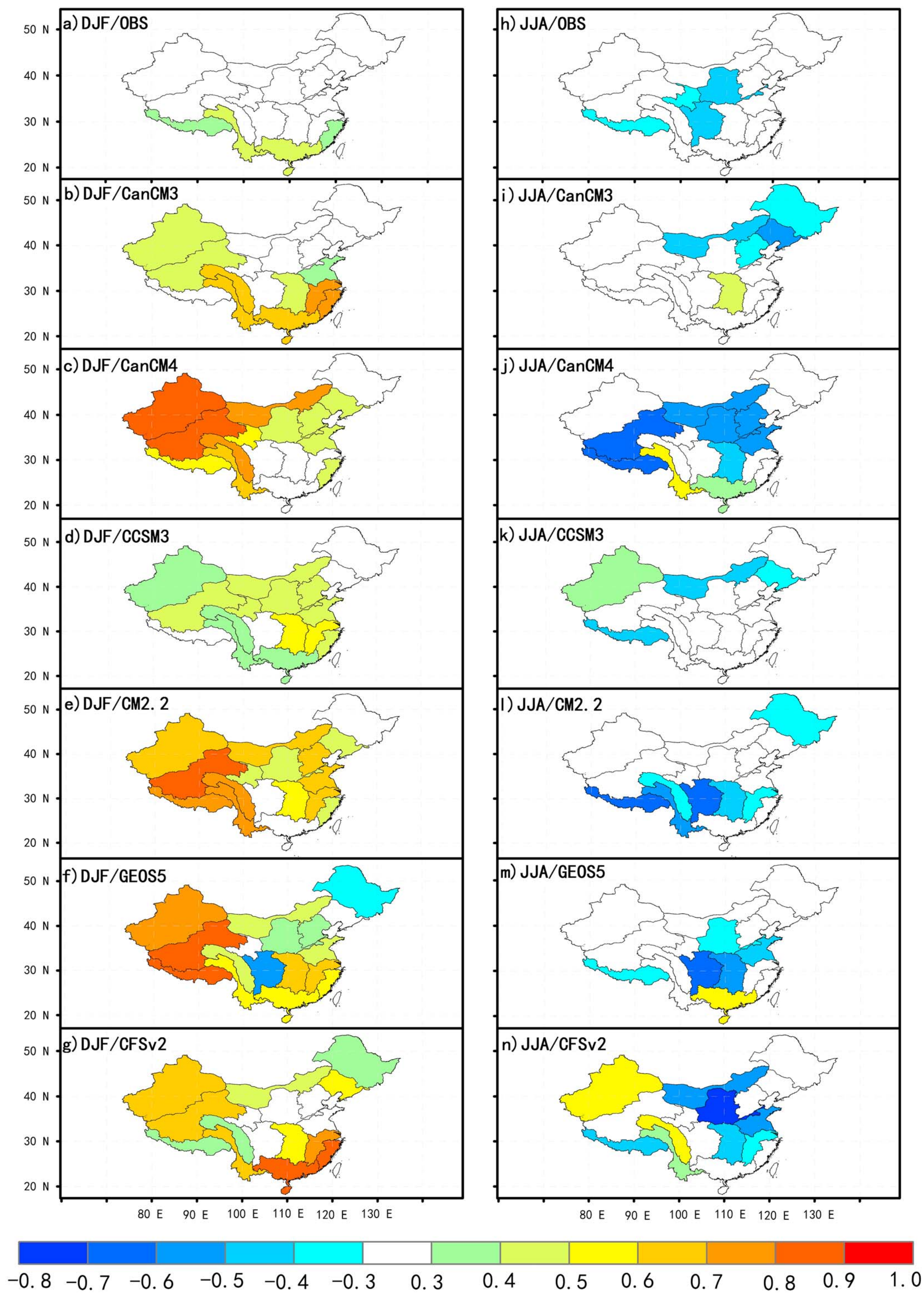

Figure 4. (a–n) The spatial distributions of correlations between regional mean SPI3 and Niño 3.4 SST anomaly from observations and NMME seasonal climate forecasts at lead-1 seasons during DJF and JJA. Values between −0.3 and 0.3 are insignificant and masked out.

models (Figures 4c, 4f, and 4g). This indicates that drought predictability and forecast skill over ENSO-affected regions during winter are highly dependent on whether the climate model is able to capture the ENSO teleconnection, and how strong the ENSO SST and regional precipitation is coupled in the climate model. For June-July-August (JJA), drought predictability and forecast skill vary among models and regions (Figures 2g–2l and

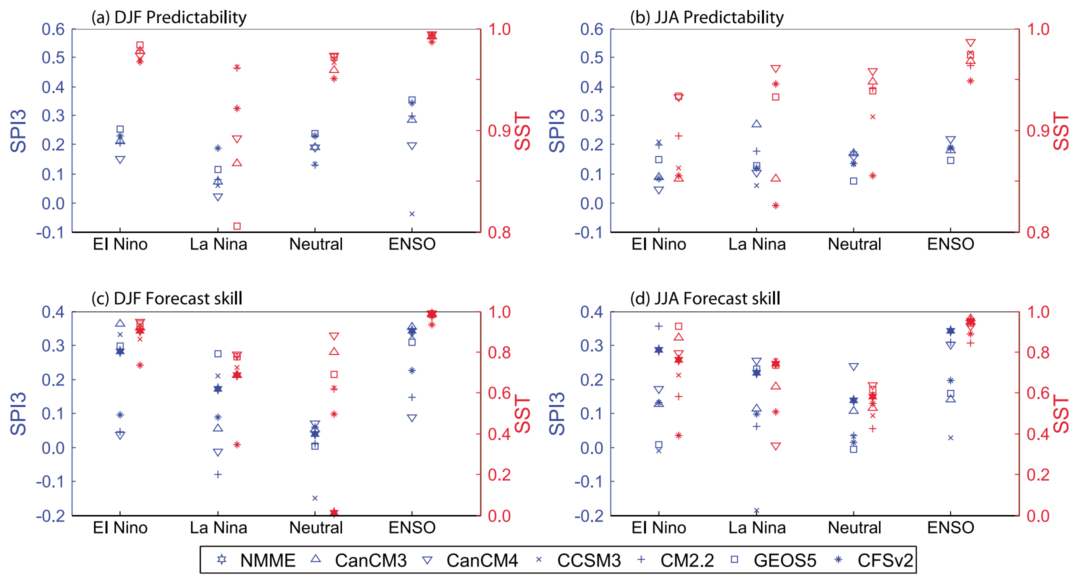

Figure 5. The (a and b) predictability and (c and d) forecast skill in terms of anomaly correlation (AC) for the SPI3 and SST in different ENSO phases over ENSO-affected river basins during DJF (Figures 5a and 5c) and JJA (Figures 5b and 5d) . The blue color refers to SPI3, and the red color refers to SST.

Figures 3g–3l). CanCM3, CanCM4, and CCSM3 have high predictability over Pearl River (Figures 2a–2c), and GEOS5 and CFSv2 models show low predictability over most regions (Figures 2e–2f). Most models (except CCSM3) have high forecast skill over Yellow River (Figures 3g–3l). The relationship between SPI3 and ENSO SST is also weak over most regions during summer (Figure 4h–4n). As compared with winter, summer time drought in China is more complicated, and the ENSO signal is mixed with other oceanic signals (e.g., Indian Ocean monsoon) as well as mesoscale land-atmosphere coupling processes. In addition, there seems to be a stronger relationship between ENSO and SPI3 in the models than in the actual observations (Figure 4), which may be due to stronger ocean-atmosphere coupling in these models. Strong coupling seems to be a common deficiency across many CGCMs including the NMME models. The interaction between ocean and atmosphere is complicated and nonlinear in reality, while the models may have certain linearized assumptions to make the ENSO-precipitation relationship more deterministic and to increase the predictability and decrease the chaos. On one hand, we would like to make use of the ENSO-rainfall relationship for higher predictability. On the other hand, we lose nonlinear information. This finding is consistent with *Yuan and Wood* [2013] on global drought onset assessment, where they speculated that current NMME models need to strengthen other processes (e.g., land-atmosphere coupling) to avoid overrepresentation of the ENSO SST-precipitation coupling.

To diagnose the interannual variation of the results, the SPI3 predictability and forecast skill conditional on different ENSO phases [*Li et al.*, 2014] over the ENSO-affected regions are calculated. It has been found that ENSO has significant influence over southern China during DJF (Figure 4a), and over Yellow River, upstream Yangtze River, and southern Tibet Plateau during JJA (Figure 4h). Therefore, statistics over these ENSO-affected regions during DJF and JJA conditional on different ENSO phases are calculated and analyzed. Predictability for ENSO SST is especially high in six models during most of ENSO phases (Figures 5a and 5b). It is obvious that ENSO SST predictability in ENSO phase is higher than that in neutral phase, so is the SPI3 predictability (except CCSM3; Figures 5a and 5b) over ENSO-affected regions mentioned above. Forecast skill for ENSO SST is lower than predictability but is still high in ENSO phases (Figures 5c and 5d). The forecast skill for ENSO SST and SPI3 in ENSO phase is apparently higher than that in neutral phase (except CCSM3 during JJA; Figures 5c and 5d). Current CGCMs show good capability in predicting SST in most ENSO phases. In addition, Figure 5 shows that good predictability and forecast skill of drought over ENSO-affected regimes are associated with good predictions of ENSO-related SST patterns and their teleconnections, and the phenomenon is more obvious in DJF than JJA.

In fact, SPI3 carries not only information about dry spells but also wet spells. Therefore, special attention is also paid to the analysis of predictability and forecast skill during drought conditions (i.e., SPI3 < -0.8). The

Table 3. Drought (SPI3 < −0.8) Predictability and Forecast Skill Over China, Averaged Over the 12 Lead-0 Seasons[a]

	CanCM3	CanCM4	CCSM3	CM2.2	GEOS5	CFSv2	Obs (Ens AC)
CanCM3 Ens	**0.32**	0.23	0.05	0.11	0.09	0.07	*0.07*
CanCM4 Ens	0.22	**0.36**	0.03	0.12	0.10	0.08	*0.07*
CCSM3 Ens	0.06	0.04	**0.11**	0.05	0.02	0.02	*0.04*
CM2.2 Ens	0.12	0.13	0.04	**0.31**	0.08	0.06	*0.06*
GEOS5 Ens	0.11	0.12	0.02	0.09	**0.22**	0.09	*0.08*
CFSv2 Ens	0.10	0.11	0.03	0.09	0.11	**0.23**	*0.06*
Singmem AC	0.03	0.04	0.02	0.03	0.03	0.02	***NMME = 0.10***

[a]The same as Table 1 but for drought conditions (SPI3 < −0.8) over lead-0 seasons.

results for drought forecasting at lead-0 seasons are shown in Table 3. The reason for focusing on the lead-0 season only is that the drought forecast skill at lead-1 is too low to pass significant test (not shown). Table 3 shows that the drought predictability is close to SPI3 predictability, but drought forecast skill is much lower than the SPI3 at lead-0 seasons. This suggests a larger room for improving the drought forecasts than the SPI3 forecasts. Similarly, the CFSv2 and GEOS5 models also have low drought predictability but high forecast skill in terms of deterministic forecasting (AC), probably due to the same reason (different ensemble generation method) as the SPI3. The forecast skill for NMME 6-model grand ensemble is 0.1, which is also the highest as compared with individual climate models. By excluding CFSv2 and GEOS5, models with higher predictability have higher forecast skill. Again, this is also consistent with the results for SPI3.

The relationship between drought predictability and forecast skill and their seasonal differences are also shown in Figure 6. The average drought predictability and forecast skill at lead-0 seasons are highest during autumn (0.30 and 0.08) and lowest during spring (0.21 and 0.03). Drought predictability is higher than forecast skill. These results are consistent with those for SPI3 at the same lead (Figure 1b). There are positive relationships between drought predictability and forecast skill, but they are not significant. Except for the winter seasons, the correlations between predictability and forecast skill for drought are generally lower than that for the SPI3 (Figure 1b and Figure 6). This may indicate the limitation for current CGCMs: the climate models can represent the relationship between potential and actual forecast skill through intrinsic physical processes (e.g., the ENSO SST-precipitation coupling) to some extent, but less robust to fully account for the processes for extremes (e.g., drought) in this regard.

Figure 7 shows the HR, FAR, and ETS values of predictability and forecast skill for drought events in six models over China, averaged over 12 seasons. The values of HR for both predictability and forecast skill are higher than FAR, indicating higher drought detections than misses. The range of HR for drought forecast skill is from 0.2 to 0.3, lower than those for drought predictability (0.35–0.5). The FAR for drought forecast skill vary from 0.20 to 0.25, higher than those for drought predictability. The ETS for six models are positive, indicating that skill of six models for drought prediction is higher than random forecast. The GEOS5 and CFSv2 models have lower drought predictability (lower HR and ETS and higher FAR) and higher forecast skill (higher HR and ETS and lower FAR), which is contrary to other models. The CFSv2 model has the highest hit rate and lowest false alarm rate, indicating more accurate drought detections. The CCSM3 model has worse drought

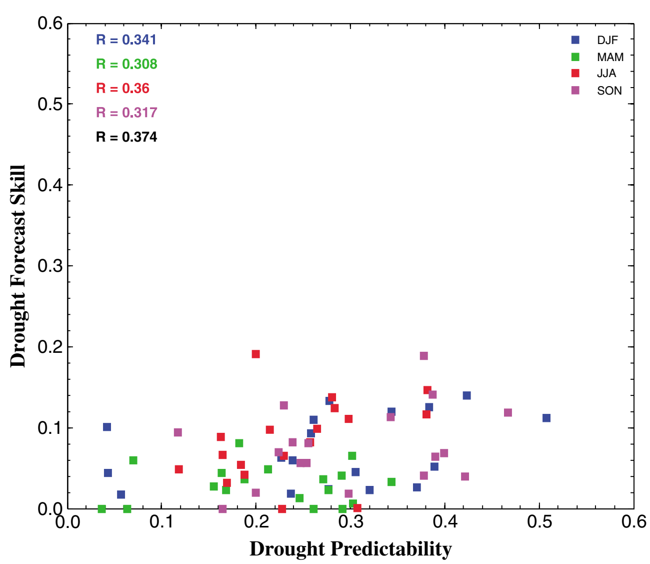

Figure 6. The same as Figure 1 but for drought (SPI3 < −0.8) predictability (*x* axis) versus forecast skill (*y* axis) at lead-0 seasons.

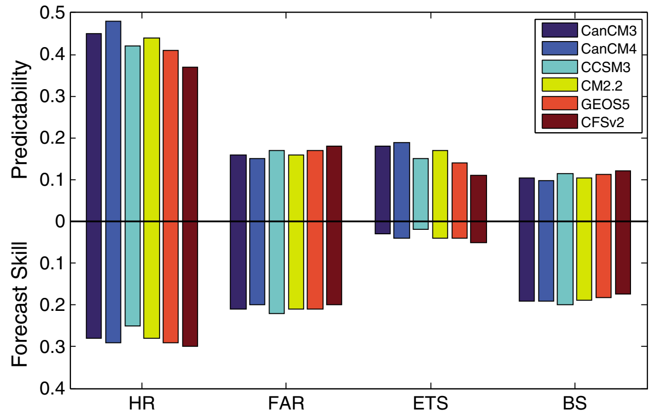

Figure 7. Hit rate (HR), false alarm rate (FAR), Equitable Threat Score (ETS), Brier Score (BS) for predictability and forecast skill for drought events (SPI3 < -0.8) at lead-0 seasons.

detections. The performances of other models fall between the CFSv2 and CCSM3. In addition, the CFSv2 model also has the highest ETS (0.05), exhibiting better drought forecasts.

The ensemble verification for drought is also carried out. Figure 7 displays the drought predictability and forecast skill in terms of Brier Score (BS). The range of BS for predictability is 0.09–0.13, better than the forecast skill (0.17–0.20), indicating that drought predictability is higher than forecast skill in terms of probabilistic forecasting. Excluding GEOS5 and CFSv2, the models (e. g. CanCM3, CanCM4, and CM2.2) with higher predictability (smaller BS) also have higher forecast skill (smaller BS). The results are consistent with the deterministic forecast analysis above.

4. Concluding Remarks

In this paper, based on 28 year (1982–2009) hindcasts from multiple North American Multi-Model Ensemble (NMME) climate forecast models, predictability and forecast skill for drought and their relationship are investigated over China. Metrics for skill assessment include the anomaly correlation (AC), Brier score (BS), hit rate (HR), false alarm rate (FAR), and Equitable Threat Score (ETS). Drought forecast skill for model ensemble mean is higher than the single-ensemble members, and NMME grand ensemble performs the best. Drought predictability is higher than the forecast skill in terms of deterministic and probabilistic forecasting, suggesting that there is room for improving the drought forecast. Models predict themselves better than they predict other models and observations. The skill of models predicting themselves is related to the method used to create the ensemble. There are positive correlations between drought predictability and actual forecast skill; i.e., the models with higher potential predictability also show higher forecast skill. These positive correlations hold for different seasons especially during autumn, and even for longer forecast leads. At lead-0 seasons, the positive correlation can be obtained by excluding the models (GEOS5 and CFSv2) with ensembles generated from different forecast leading times.

At seasonal time scale, ENSO has a substantial impact on drought over China. The impact is more significant over southeastern China during DJF, and over Yellow River, upstream Yangtze River, and southwestern China during JJA. Predictability and forecast skill for drought is higher over these regions as a result of the models' capability in capturing the ENSO teleconnection. For these ENSO-affected regions, both predictability and forecast skill for drought index and ENSO SST in ENSO phases are higher than that in neutral phases.

In this study, it is also found that the relationship between drought predictability and forecast skill varies depending on different climate regimes. For example, the relationship is more significant over southern China during winter, and more significant over southwestern China and downstream Yangtze River during summer (not shown). They are primarily ENSO-affected regimes (Figures 4a and 4h), which indicates that ENSO teleconnection controls the drought predictability and forecast skill in NMME seasonal climate forecasting models, given that they have a prediction system that can capture ENSO.

Predictability can be used as an indicator for hindcast skill, when in situ and/or remote sensing measurements for the application and location at hand are considered unreliable/useless. For instance, for the northwestern China where in situ measurements are sparse and uncertain, the model results indeed become a valuable proxy for observations; i.e., the predictability can be a measure of model performance when the verification data are unreliable. Forecast skill can also be assessed based on predictability in the forecasts and the relationship between predictability and forecast skill in the hindcasts. Further analysis by using Phase-2 of NMME with atmospheric pressure level data would be useful for exploring drought predictability and its physical mechanism in details.

Acknowledgments
This work was supported by the Thousand Talents Program for Distinguished Young Scholars, the Natural Science Foundation of China (41475093), and the Special Fund for Public Welfare Industry (GYHY201506001). We would like to thank three anonymous reviewers for their helpful comments and IRI (http://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/) for making the NMME forecast information available.

References

Becker, E., H. van den Dool, and Q. Zhang (2014), Predictability and forecast skill in NMME, *J. Clim.*, *27*, 5891–5906.
Boer, G. J., V. V. Kharin, and W. J. Merryfield (2013), Decadal predictability and forecast skill, *Clim. Dyn.*, *41*, 1817–1833, doi:10.1007/s00382-013-1705-0.
DelSole, T., J. Nattala, and M. K. Tippett (2014), Skill improvement from increased ensemble size and model diversity, *Geophys. Res. Lett.*, *41*, 7331–7342, doi:10.1002/2014GL060133.
Doblas-Reyes, F. J., J. García-Serrano, F. Lienert, A. P. Biescas, and L. R. L. Rodrigues (2013), Seasonal climate predictability and forecasting: Status and prospects, *WIREs Clim. Change*, *4*, 245–268, doi:10.1002/wcc.217.
Dutra, E., L. Magnusson, F. Wetterhall, H. L. Cloke, G. Balsamo, S. Boussetta, and F. Pappenberger (2012), The 2010–2011 drought in the Horn of Africa in ECMWF reanalysis and seasonal forecast products, *Int. J. Climatol.*, *33*, 1720–1729, doi:10.1002/joc.3545.
Gibbs, W. J., and J. V. Maher (1967), *Rainfall Deciles as Drought Indicators*, Bur. of Meteorol., Melbourne, Australia.
Hayes, M., M. Svoboda, N. Wall, and M. Widhalm (2011), The Lincoln declaration on drought indices: Universal meteorological drought index recommended, *Bull. Am. Meteorol. Soc.*, *92*, 485–488.
Holland, M. M., E. Blanchard-Wrigglesworth, J. Kay, and S. Vavrus (2013), Initial-value predictability of Antarctic sea ice in the Community Climate System Model 3, *Geophys. Res. Lett.*, *40*, 2121–2124, doi:10.1002/grl.50410.
Kirtman, B. P., et al. (2014), The North American Multi-Model Ensemble (NMME): Phase-1 seasonal to interannual prediction, phase-2 toward developing intra-seasonal prediction, *Bull. Am. Meteorol. Soc.*, *95*, 585–601.
Koster, R. D., M. J. Suarez, and M. Heiser (2000), Variance and predictability of precipitation at seasonal-to-interannual timescales, *J. Hydrometeorol.*, *1*, 26–46.
Kumar, A., P. Peng, and M. Chen (2014), Is there a relationship between potential and actual skill?, *Mon. Weather Rev.*, *142*, 2220–2227.
Lang, Y., A. Ye, W. Gong, C. Miao, Z. Di, J. Xu, Y. Liu, L. Luo, and Q. Duan (2014), Evaluating skill of seasonal precipitation and temperature predictions of NCEP CFSv2 forecasts over 17 hydroclimatic regions in China, *J. Hydrometeorol.*, *15*, 1546–1559.
Li, C., R. Lu, and B. Dong (2014), Predictability of the western North Pacific summer climate associated with different ENSO phases by ENSEMBLES multi-model seasonal forecasts, *Clim. Dyn.*, *43*, 1829–1845, doi:10.1007/s00382-013-2010-7.
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.
Luo, L., and E. F. Wood (2007), Monitoring and predicting the 2007 U.S. drought, *Geophys. Res. Lett.*, *34*, L22702, doi:10.1029/2007GL031673.
Ma, F., A. Ye, X. Deng, Z. Zhou, X. Liu, Q. Duan, J. Xu, C. Miao, Z. Di, and W. Gong (2015), Evaluating the skill of NMME seasonal precipitation ensemble predictions for 17 hydroclimatic regions in continental China, *Int. J. Climatol.*, doi:10.1002/joc.4333.
McKee, T. B., N. J. Doesken, and J. Kleist (1993), The relationship of drought frequency and duration to time scales, paper presented at 8th Conference on Applied Climatology, Am. Meteorol. Soc., Anaheim, Calif.
McKee, T. B., N. J. Doesken, J. Kleist (1995), Drought monitoring with multiple time scales, paper presented at 9th Conference on Applied Climatology, Am. Meteorol. Soc., Dallas, Tex.
Mishra, A. K., and V. P. Singh (2010), A review of drought concepts, *J. Hydrol.*, *391*, 202–216.
Mo, K., and D. P. Lettenmaier (2014), Hydrologic prediction over the conterminous United States using the National Multi-Model Ensemble, *J. Hydrometeorol.*, *15*, 1457–1472.
Palmer, W. C. (1965), *Meteorologic Drought*, U.S. Dep. of Commer., Weather Bur., U.S. Weather Bur., Washington, D. C.
Smith, T. M., R. W. Reynolds, T. C. Peterson, and J. Lawrimore (2008), Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880–2006), *J. Clim.*, *21*, 2283–2296, doi:10.1175/2007JCLI2100.1.
Svoboda, M., et al. (2002), The drought monitor, *Bull. Am. Meteorol. Soc.*, *83*, 1181–1190.
Tian, D., C. J. Martinez, W. D. Graham, and S. Hwang (2014), Statistical downscaling multimodel forecasts for seasonal precipitation and surface temperature over the southeastern United States, *J. Clim.*, *27*, 8384–8411.
Trenberth, K. E., and J. M. Caron (2000), The Southern Oscillation revisited: Sea level pressures, surface temperatures and precipitation, *J. Clim.*, *13*, 4358–4365, doi:10.1175/1520-0442(2000)013<4358:TSORSL>2.0.CO;2.
Van Rooy, M. P. (1965), A rainfall anomaly index independent of time and space, *Notos*, *14*, 43.
Wang, A., D. P. Lettenmaier, and J. Sheffield (2011), Soil moisture drought in China, 1950–2006, *J. Clim.*, *24*, 3257–3271, doi:10.1175/2011JCLI3733.1.
Wilks, D. S. (1995), *Statistical Methods in the Atmospheric Sciences: An Introduction*, 467 pp., Academic Press, San Diego, Calif.
Xie, P., M. Chen, S. Yang, A. Yatagai, T. Hayasaka, Y. Fukushima, and C. Liu (2007), A gauge-based analysis of daily precipitation over East Asia, *J. Hydrometeorol.*, *8*, 607–626.
Younas, W., and Y. Tang (2013), PNA predictability at various time scales, *J. Clim.*, *26*, 9090–9114, doi:10.1175/JCLI-D-12-00609.1.
Yuan, X., and E. F. Wood (2012), On the clustering of climate models in ensemble seasonal forecasting, *Geophys. Res. Lett.*, *39*, L18701, doi:10.1029/2012GL052735.
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.
Yuan, X., E. F. Wood, N. W. Chaney, J. Sheffield, J. Kam, M. Liang, and K. Guan (2013), Probabilistic seasonal forecasting of African drought by dynamical models, *J. Hydrometeorol.*, *14*, 1706–1720, doi:10.1175/JHM-D-13-054.1.
Yuan, X., Z. Ma, M. Pan, and C. Shi (2015), Microwave remote sensing of short-term droughts during crop growing seasons, *Geophys. Res. Lett.*, *42*, 4394–4401, doi:10.1002/2015GL064125.
Zhang, W., F.-F. Jin, and A. Turner (2014), Increasing autumn drought over southern China associated with ENSO regime shift, *Geophys. Res. Lett.*, *41*, 4020–4026, doi:10.1002/2014GL060130.