Subseasonal Ensemble Prediction of Flash Droughts over China

RUXUAN MA[a,b] AND XING YUAN[a,b]

[a] *Key Laboratory of Hydrometeorological Disaster Mechanism and Warning of Ministry of Water Resources, Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science and Technology, Nanjing, China*
[b] *School of Hydrology and Water Resources, Nanjing University of Information Science and Technology, Nanjing, China*

(Manuscript received 30 August 2022, in final form 17 February 2023, accepted 21 February 2023)

ABSTRACT: Flash droughts have been occurring frequently worldwide, which has a serious impact on food and water security. The rapid onset of flash droughts presents a challenge to the subseasonal forecast, but there is limited knowledge about their forecast skills due to the lack of appropriate identification and assessment procedures. Here, we investigate the forecast skill of flash droughts over China with lead times up to 3 weeks by using hindcast datasets from the Subseasonal-to-Seasonal Prediction (S2S) project. The flash droughts are identified by using weekly soil moisture percentiles from two S2S forecast models (ECMWF and NCEP). The comparison with reanalysis shows that ECMWF and NCEP forecast models underestimate flash drought occurrence by 5% and 19% for lead 1 week. The national mean hit rates for flash droughts are 0.22 and 0.16 for ECMWF and NCEP models for lead 1 week, and they can reach 0.29 and 0.18 over South China. The ensemble of the two models increases equitable threat score (ETS) from ECMWF and NCEP models by 8% and 40% for lead 1 week. In terms of probabilistic forecast, ECMWF has a higher Brier skill score than NCEP, especially over eastern China, which is consistent with higher temperature and precipitation forecast skill. The multimodel ensemble has the highest Brier skill score. This study suggests the importance of multimodel ensemble flash drought forecasting.

SIGNIFICANCE STATEMENT: Flash droughts have raised considerable concern, but whether they can be predicted at subseasonal time scales remains unclear. This study evaluates forecast skill of flash droughts over China based on ECMWF and NCEP hindcast data. Focusing on the historical flash drought events identified by the onset speed and duration, it is found that the ECMWF model outperformed the NCEP model with higher hit rates, lower false alarm ratios, and higher equitable threat scores, especially during the first week. However, less than 30% of the drought events can be captured in most regions by both models. An ensemble of the two models showed skill improvement against the ECMWF model for both deterministic and probabilistic forecasts.

KEYWORDS: Drought; Climate prediction; Ensembles; Forecast verification/skill; Hindcasts; Climate models

1. Introduction

The drought risk is often underestimated due to its silent propagation without sufficient early warning (Yuan and Wood 2013; Hao et al. 2017; Pendergrass et al. 2020). In addition, climate change and human intervention alter the drought characteristics, which raises a grand challenge for drought prediction in the Anthropocene (Yuan et al. 2017; Samaniego et al. 2018; Zhang et al. 2022). For instance, a type of drought with rapid onset occurs frequently worldwide in recent years, which is called "flash drought" (Hoerling et al. 2014; Yuan et al. 2015; Otkin et al. 2018; Yuan et al. 2019; Wang and Yuan 2021). Precipitation deficit is an important condition for the occurrence of flash drought (Mo and Lettenmaier 2016; Hoffmann et al. 2021), but increased evapotranspiration due to high temperature is also a critical factor for triggering the flash drought (Wang et al. 2016; Wang and Yuan 2021, 2022). Strong precipitation deficits and evapotranspiration excess increase the onset speed of flash droughts and make flash drought forecasting difficult (Pendergrass et al. 2020; Liang and Yuan 2021; Zhu and Wang 2021).

There are a number of flash drought indices in the literature (Ford and Labosier 2017; Christian et al. 2019; Yuan et al. 2019; Pendergrass et al. 2020), and many studies have indicated that soil moisture anomalies are useful for characterizing the drought onset, particularly for rapid-onset droughts (Otkin et al. 2018). In fact, soil moisture deficit is a good indicator for agricultural drought, and flash drought differs from conventional drought in that the former has a rapid onset, which can be directly characterized by the rapid decline of soil moisture or its percentile (Yuan et al. 2019). Osman et al. (2021) examined the key climate variables used in flash drought definitions, including precipitation, soil moisture, and temperature, as well as actual and potential evapotranspiration, and they found that soil moisture–based flash drought definitions captured the major events.

Flash droughts usually occur at the subseasonal time scale, with durations ranging from a few weeks to 1–2 months. Drought prediction skill at such time scale is limited due to the chaotic

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JHM-D-22-0150.s1.

Corresponding author: Xing Yuan, xyuan@nuist.edu.cn

DOI: 10.1175/JHM-D-22-0150.1

nature of the climate system and the mixture of the signals from synoptic to seasonal time scales. Different temporal and spatial scales of the atmosphere, land, and ocean processes, and their forecast skill, make the prediction a challenge (Yuan and Wood 2013; Chen et al. 2020). To improve the understanding of the sources of subseasonal to seasonal predictability as well as forecast skill, the World Weather Research Program (WWRP) and the World Climate Research Program (WCRP) jointly launched a research initiative called the Subseasonal-to-Seasonal Prediction (S2S) project (Vitart et al. 2017). The S2S datasets were used to evaluate forecast skill for extreme weather related with temperature and precipitation extremes (Luo and Wood 2006; Tian et al. 2017; Prein et al. 2022; Tuel and Martius 2021). Vitart and Robertson (2018) also showed that there is a potential to predict the occurrence and the evolution of extreme events at subseasonal time scale. As the key characteristic of flash drought is the rapid onset, whether the S2S models can capture such process needs further investigation.

Forecast skill for medium-range (usually defined as a forecast period fewer than 2 weeks) weather forecasts has been improved in recent years (Novak et al. 2014). Moreover, seasonal forecast skill for droughts has also been extensively assessed by using both dynamical ocean–atmosphere coupled models or statistical models (Yuan and Wood 2013; Ma et al. 2015; Yuan et al. 2017; Yao and Yuan 2018; Liang and Yuan 2021; Wu et al. 2021). However, the prediction of flash drought is a challenge because they occur rapidly with a strong land–atmospheric interaction (Wang and Yuan 2022), where most global climate prediction models do not necessarily represent the interaction well. Mo and Lettenmaier (2020) assessed the flash drought forecast skill based on the Global Ensemble Forecast System Reforecast v2 (GEFSv2) in the United States, and there is limited skill beyond 10 days. Moreover, they actually assessed the forecast skill of compound hot and dry extremes at pentad-scale according to their flash drought definition, without explicit requirements for the rapid onset and a drought duration of at least 15–20 days (Otkin et al. 2018; Yuan et al. 2019). Whether the S2S models provide skillful forecasts for flash drought events over China, another flash drought hotspot (Yuan et al. 2019), needs a comprehensive assessment based on multiple S2S models, with both deterministic and probabilistic metrics.

In this study, we used both the NCEP (similar to those used by Mo and Lettenmaier 2020) and the ECMWF hindcast datasets from the S2S archive to evaluate the subseasonal forecast skill of flash droughts over China. The flash drought events were identified based on the rapid decline in weekly soil moisture, and the performances of ECMWF and NCEP models were compared by using the ensemble mean results and individual ensemble members, and the skill for the combination of the two models was also assessed.

2. Data and method

a. Observations and S2S datasets

The ECMWF Reanalysis 5 (ERA5; Hersbach et al. 2020) soil moisture and the CN05.1 precipitation (Wu and Gao 2013) are used as reference data. ERA5 data are the fifth-generation of global reanalysis data released by ECMWF. Here, the 20-cm ERA5 soil moisture data with a spatial resolution of 0.5° × 0.5° are used in this study. This is because the S2S hindcast data do not have enough drought samples for the forecast verification if the deep-layer soil moisture is used for flash drought identification. The study period is 1 April–29 September 1999–2020. The CN05.1 precipitation observations are based on more than 2400 ground-based meteorological stations in China compiled by China Meteorological Administration (CMA), and they are regridded to 0.5° × 0.5° resolution.

The S2S multimodel hindcast and real-time forecast products contain data from 11 models worldwide (Vitart et al. 2017), and two of them are selected, i.e., ECMWF and NCEP. The data are at 0.5° × 0.5° resolution. The NCEP real-time prediction became operational from March 2011 with the newly developed ocean–land–atmosphere coupled model (Yuan et al. 2011), and the S2S archived the NCEP real-time forecasts from 2015. The hindcasts during 1999–2010 have more observations from ocean, land, and atmosphere to initialize the model, while the real-time forecasts may have less observations for initialization when the forecasts issued. The data period is 1999–2020 for ECMWF model and is 1999–2010 (2015–20) for hindcast (real-time forecast) for NCEP model. The ensemble sizes are 11 and 4 for ECMWF and NCEP models, respectively. To focus on the skill assessment in growing seasons (Yuan et al. 2019) based on the available S2S data, the study period is chosen as 1 April–29 September 1999–2010. For each year, we select 26 hindcast/forecast cases that are initialized every week from 1 April to 23 September (i.e., 1 April, 8 April, …, 23 September) with forecast leads up to 46 days. We use the first 21 days for weekly forecasts from lead 1 week to lead 3 weeks. The 11 ensemble members for the ECMWF model are initialized at the same date, which consist of a control ensemble member and 10 perturbed ensemble members. For the NCEP model, the four ensembles are constructed by using the hindcasts/forecasts that are initialized at 0000 UTC during the target initialization day as mentioned above, and those initialized at 0600, 1200, and 1800 UTC during the previous day.

b. Definition of flash drought

Yuan et al. (2019) developed a flash drought identification method based on the soil moisture decline rate and drought duration, which described the onset and recovery processes of flash drought events. Here we use a similar definition: a flash drought event is identified as the weekly (7 days) mean top-20-cm soil moisture decrease from above the 40th percentile down to the 20th percentile, with an average decline rate of no less than 7% for each week, and the total drought duration (from onset to recovery) should be no less than 3 weeks. The soil moisture cumulative distribution functions (CDFs) for the ERA5 reanalysis and each ensemble of S2S forecasts are constructed separately, and the soil moisture percentiles used for the flash drought analysis are calculated by using their own CDFs for reanalysis and forecast ensembles. This can remove the forecast biases implicitly.

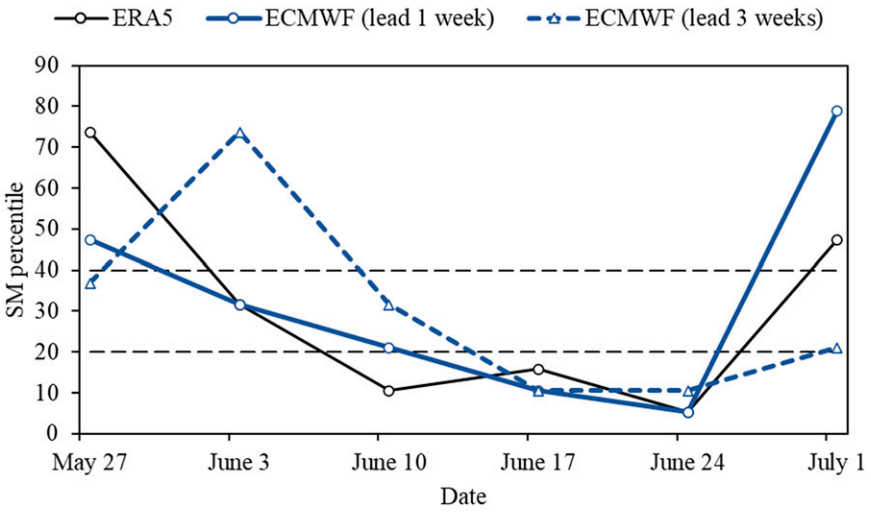

FIG. 1. Illustration for the forecast evaluation for a flash drought event. The solid black line shows the ERA5 weekly mean soil moisture (SM) percentile at the grid point (43.5°N, 124.5°E) during the 2007 flash drought. Blue solid and dashed lines show the weekly mean SM percentile from the ECMWF model at lead 1 week and lead 3 weeks, respectively. Here, the lead-1-week forecast captured the flash drought, while the lead-3-weeks forecast missed the onset week (3 Jun).

Since the 12-yr calculation for flash droughts is very short, we add the results of the previous week and the following week to calculate the CDFs.

c. *Verification measures*

Forecast skill for flash drought events can be evaluated using a 2 × 2 contingency table (Wilks 1995). Let A be the number of observed flash drought events that are forecast by the model, C be the number of observed flash drought events that are not forecast by the model, B be the number forecast flash drought events that are turn out to be false alarms, and D be the nondrought events that are both in observation and forecast, then the hit rate (HIT), false alarm ratio (FAR), and equitable threat score (ETS) can be calculated as follows:

$$\mathrm{HIT} = \frac{A}{A + C}, \tag{1}$$

$$\mathrm{FAR} = \frac{B}{A + B}, \tag{2}$$

$$\mathrm{ETS} = \frac{A - A_{\mathrm{Ref}}}{A + B + C - A_{\mathrm{Ref}}}, \tag{3}$$

$$A_{\mathrm{Ref}} = \frac{(A + B)(A + C)}{N}. \tag{4}$$

HIT indicates the probability of detection, and FAR indicates the probability of false alarm. ETS is a balanced score with an optimal value of 1.

The ensemble average soil moisture is calculated first (both for ECMWF/NCEP model, and their multimodel ensemble), and then the soil moisture percentile is calculated before identifying flash drought events. Taking HIT as an example, Fig. 1 shows that ERA5 reanalysis identified a flash drought event during 3 and 24 June 2007, the ECMWF S2S forecast model captured this event for the lead-1-week forecast, while the lead-3-week forecast failed to predict the onset of the drought on 3 June. Therefore, the lead-1-week forecast "hit" the drought, while the lead-3-weeks forecast missed the drought.

Perfect model skill indicates the ability of a model to forecast itself (Ma et al. 2015), by assuming the model is perfect and the errors mainly come from initial conditions. Taking the ECMWF model with 11 ensemble members as an example, we calculate the perfect model skill for ETS as follows: 1) we use ensemble 1 as "truth" and the mean of ensembles 2–11 as "forecast," to calculate the first ETS; 2) we use ensemble 2 as truth and the mean of ensembles 1 and 3–11 as forecast, to calculate the second ETS; 3) we repeat step 1 to obtain the third, fourth, …, and the 11th ETS, and the perfect model skill in terms of ETS is the mean of these 11 ETS values. In addition, the Brier score (BS; Wilks 1995) is also used for assessing probabilistic forecast skill. BS is essentially the mean square error of the probabilistic forecast, which is calculated as follows:

$$\mathrm{BS} = \frac{1}{N}\sum_{k=1}^{N}(y_k - o_k)^2, \tag{5}$$

where N is the number of the target (3 weeks) periods for the skill assessment [e.g., N = 26 (weeks) × 12 (years) = 312 for the NCEP model]; o represents whether flash droughts occur in the observed data, considering that the observed value is $o_k = 1$ when flash drought occurs and $o_k = 0$ when flash drought does not occur. The y_k = m1/m2 denotes the forecast

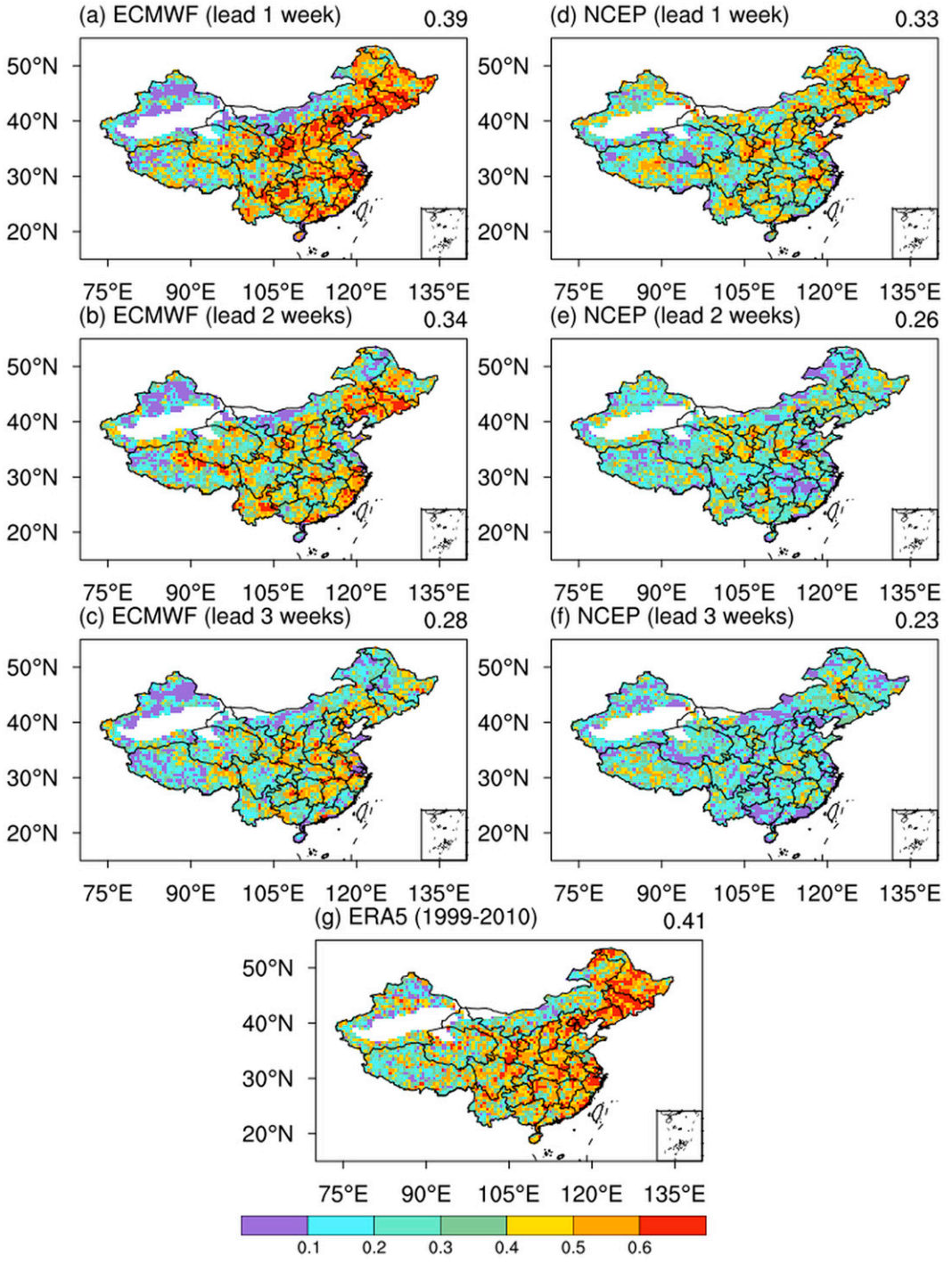

FIG. 2. Spatial patterns of the frequency of occurrence (FOC) of flash droughts (events/year) from (a)–(c) ECMWF and (d)–(f) NCEP subseasonal hindcasts during 1999–2010 with lead times of 1–3 weeks and ERA5 reanalysis during (g) 1999–2010. The flash droughts were identified based on the data during growing seasons (April–September). The spatial averaging results are shown on the upper-right corners of each panel.

probability (m1 is the number of forecasts with drought occurrence) among the ensemble members (m2) for the kth target period. For instance, for the ECMWF model, m2 = 11, and m1 varies from 0 to 11, so the y_k varies from 0 to 1 (e.g., if m1 = 5, then y_k = 5/11). A higher BS value is obtained if the accuracy of the prediction is low, and a perfect prediction has a BS value of 0. We calculate Brier skill scores (BSS) (Hamill and Juras 2006; Lee et al. 2020) as follows:

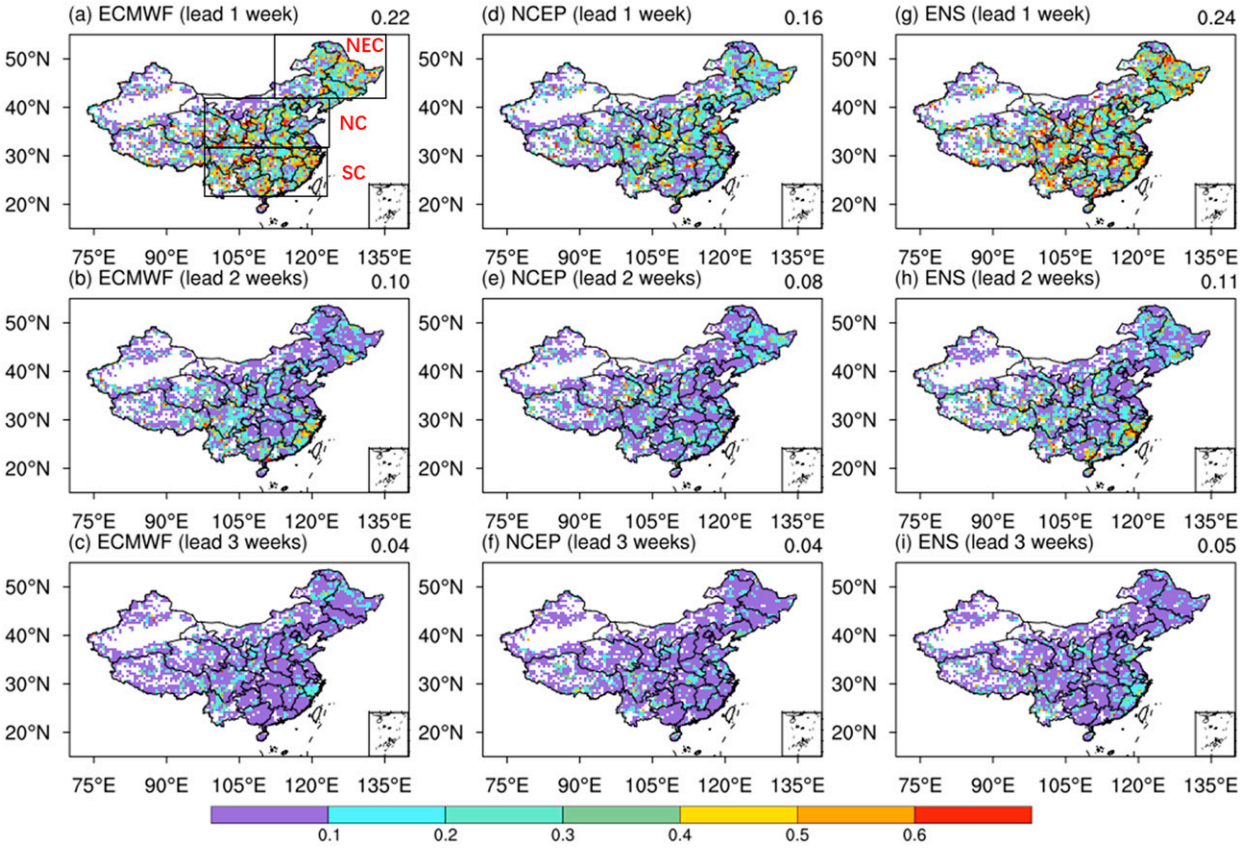

FIG. 3. (a)–(i) Hit rates for the hindcasts of flash droughts at different lead times. "ENS" represents the ensemble of ECMWF and NCEP. The boxes in (a) represent three regions: Northeast China (NEC; 42°–55°N, 112°–135°E), North China (NC; 32°–42°N, 98°–123°E), and South China (SC; 22°–32°N, 98°–122°E). The spatial averaging results are shown on the upper-right corners of each panel.

$$\mathrm{BSS} = 1 - \frac{\mathrm{BS}}{\mathrm{BS}_{\mathrm{Ref}}}, \tag{6}$$

where $\mathrm{BS}_{\mathrm{Ref}}$ is the BS for the climatological forecast consists of observed soil moisture for the target forecast weeks from different years, but excluding the target forecast year. A positive BSS value indicates a skillful forecast, while a negative value indicates a forecast worse than the climatological forecast.

3. Results

a. *Prediction of the climatology of flash drought events*

Similar to Mo and Lettenmaier (2020), we calculated the frequency of occurrence (FOC) of the flash drought to evaluate whether forecast models can capture the climatology during 1999–2010. The FOC is the average number of flash drought occurrences per year. In dry areas, small soil moisture variability may prevent a reasonable evaluation of forecast skills. Therefore, areas with annual mean precipitation below 100 mm were ignored in this study. Figure 2g shows that eastern China has more flash droughts than western China. Note that the spatial distribution is a little different from Yuan et al. (2019) due to different study period [1999–2020 in this study, while 1961–2005 in Yuan et al. (2019)] and different soil depths for drought identification [20 cm in this study, while 100 cm in Yuan et al. (2019)].

The FOC for flash droughts from the ECMWF hindcasts (Figs. 2a–c) from lead 1 week (day 1–7) to lead 3 weeks (day 15–21) captures the spatial pattern, although the underestimation is significant at long leads. The spatial correlations of ECMWF-predicted FOC for leads 1–3 weeks are 0.28, 0.21, and 0.16, respectively. The NCEP model has worse performance than the ECMWF model for reproducing the FOC pattern, with spatial correlations of 0.14, 0.11, and 0.08 for leads 1–3 weeks. The observed national mean FOC during 1999–2010 is 0.41 (Fig. 2g), while the ECMWF model-predicted national mean FOCs are 0.39, 0.34, and 0.28 for the leads of 1–3 weeks (Figs. 2a–c), with underestimations of 5%, 17%, and 32% respectively. For the NCEP model-predicted national mean FOCs are 0.33, 0.26, and 0.23 for the leads of 1–3 weeks (Figs. 2d–f), with underestimations of 19%, 36%, and 43% respectively.

b. *Forecast skill for individual flash drought events*

In this study, we used HIT, FAR, and ETS to assess forecast skill for individual flash drought events. In areas with small FOC, insufficient flash drought samples may prohibit a

TABLE 1. The regional average results of the flash drought hindcast skill with different lead times. NEC, NC, and SC represent Northeast China, North China, and South China, respectively. The hindcast period is 1999–2010 for ECMWF and NCEP models, respectively. "ENS" represents the ensemble of ECMWF and NCEP.

	HIT			FAR			ETS		
	Lead 1	Lead 2	Lead 3	Lead 1	Lead 2	Lead 3	Lead 1	Lead 2	Lead 3
ECMWF									
NEC	0.23	0.06	0.04	0.59	0.65	0.71	0.13	0.03	0.02
NC	0.22	0.09	0.04	0.66	0.73	0.76	0.13	0.05	0.02
SC	0.29	0.14	0.04	0.62	0.69	0.75	0.17	0.08	0.02
China	0.22	0.10	0.04	0.65	0.71	0.75	0.13	0.05	0.02
NCEP									
NEC	0.18	0.06	0.02	0.75	0.76	0.81	0.10	0.04	0.01
NC	0.19	0.09	0.04	0.75	0.79	0.83	0.11	0.05	0.02
SC	0.18	0.08	0.04	0.73	0.78	0.82	0.11	0.05	0.02
China	0.16	0.08	0.04	0.79	0.82	0.84	0.10	0.05	0.02
ENS									
NEC	0.27	0.08	0.03	0.63	0.68	0.73	0.16	0.05	0.01
NC	0.24	0.09	0.04	0.68	0.73	0.73	0.14	0.05	0.02
SC	0.29	0.15	0.06	0.62	0.70	0.73	0.18	0.08	0.03
China	0.24	0.11	0.05	0.67	0.72	0.75	0.14	0.06	0.02

reasonable evaluation of the forecast skill. Therefore, areas with FOC below 10% were ignored in the study. We focus on forecast skill in three regions of China: Northeast China, North China, and South China (Fig. 3a).

Figure 3 shows the hit rates for flash drought from ECMWF and NCEP models as well as their combinations "ENS" at different lead times, and Table 1 lists the national and regional mean statistics. At lead 1 week, hit rates of the ECMWF model

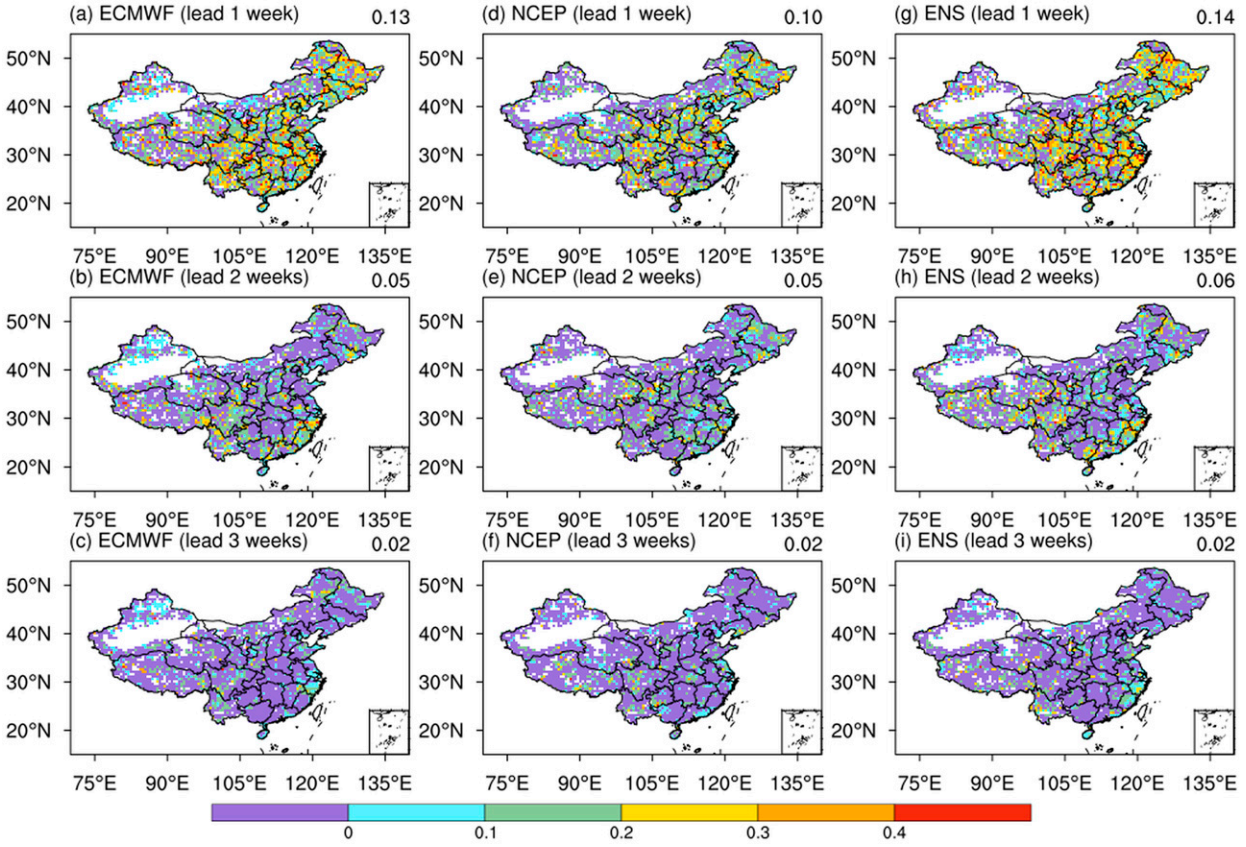

FIG. 4. As in Fig. 3, but for equitable threat score (ETS).

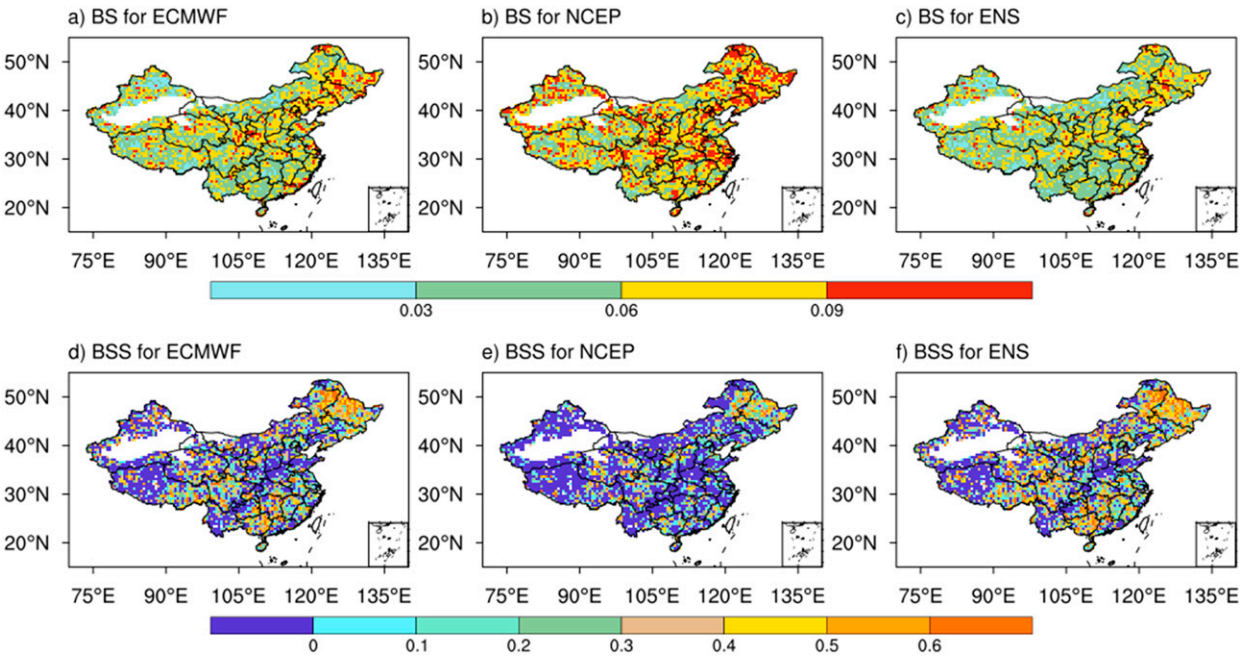

FIG. 5. Probabilistic forecast skill for ECMWF, NCEP, and ENS at 1-week lead time. (a)–(c) Brier score (BS) and (d)–(f) Brier skill score (BSS).

is highest over South China, followed by Northeast China and North China. While for the NCEP model, the hit rates at lead 1 week are highest over North China, followed by Northeast China and South China. The better forecast skill in the Northeast China and South China may be related to the better performance of FOC. With the increase in lead times, the hit rates decrease. On average, the ECMWF model has higher hit rates than the NCEP model (Figs. 3a–f). The results of the ENS are better than both the ECMWF and NCEP models (Figs. 3g–i), with increased hit rates for the national mean results of 9%–50%, 10%–38%, and 25% for the lead times of 1–3 weeks, respectively. Figure S1 in the online supplemental material shows the FAR for individual flash drought events. Eastern China has lower FAR than western China, which is associated with higher FOC over eastern China than that over western China. The FAR increases over lead times, and the NCEP model has higher FAR than the ECMWF model. This suggests that the reliability of the ECMWF model is higher than the NCEP model for flash drought forecasts, and the reliability decreases over lead times. Table 1 shows that although the ENS increases the detectability (increases the hit rate), it does not necessarily increase the reliability (decreases the false alarm ratio).

Figure 4 shows the ETS scores for individual flash droughts over each grid cell, with the positive values indicating skillful forecasts. The ECMWF has a good performance at lead 1 week, with 62% of the grid cells showing skillful forecasts (Fig. 4a) and a national average ETS of 0.13 (Table 1). However, the skill drops quickly at lead 2 weeks and lead 3 weeks, with only 35% and 18% of the grid cells showing positive ETS (Figs. 4b,c). The situation is even worse for the NCEP model, where 53%, 31%, and 15% of the grid cells show skillful forecasts at lead times of 1–3 weeks, respectively (Figs. 4d–f). The multimodel ensemble shows significant skill improvement at lead 1 week (Fig. 4g), where the ETS increases by 8%–40% (Table 1). Due to the quick skill declines of the ECMWF and NCEP models at the lead times of 2–3 weeks (Figs. 4b,c,e,f), their combination only shows improvement over limited regions (Figs. 4h,i).

Figure 5 shows the probabilistic forecast results at lead 1 week for flash droughts. Similar to the deterministic forecast skill (Fig. 4), the ECMWF model has a higher probabilistic forecast skill than the NCEP model especially over eastern China (Figs. 5d–f). Specifically, there are 57% of the grid cells with positive BSS (skillful forecast) for the ECMWF model at lead 1 week (Fig. 5d), while there are only 41% of the grid cells for the NCEP model (Fig. 5e). Although the differences in the ensemble members between ECMWF and NCEP models, multimodel ensemble show a higher BSS than ECMWF (Fig. 5f), which 62% of the grid cells for the ENS.

To understand the subseasonal forecast skill of flash droughts, we calculated the correlations between model predictions and observations for both temperature and precipitation. Figures 6a and 6d show that the correlations of temperature for ECMWF and NCEP predictions at lead 1 week are above 0.95 in most areas of China. However, the NCEP model has lower temperature correlation than the ECMWF model at lead 3 weeks (Figs. 6c,f). Figure 7a shows that the correlations of precipitation for ECMWF prediction at lead 1 week are above

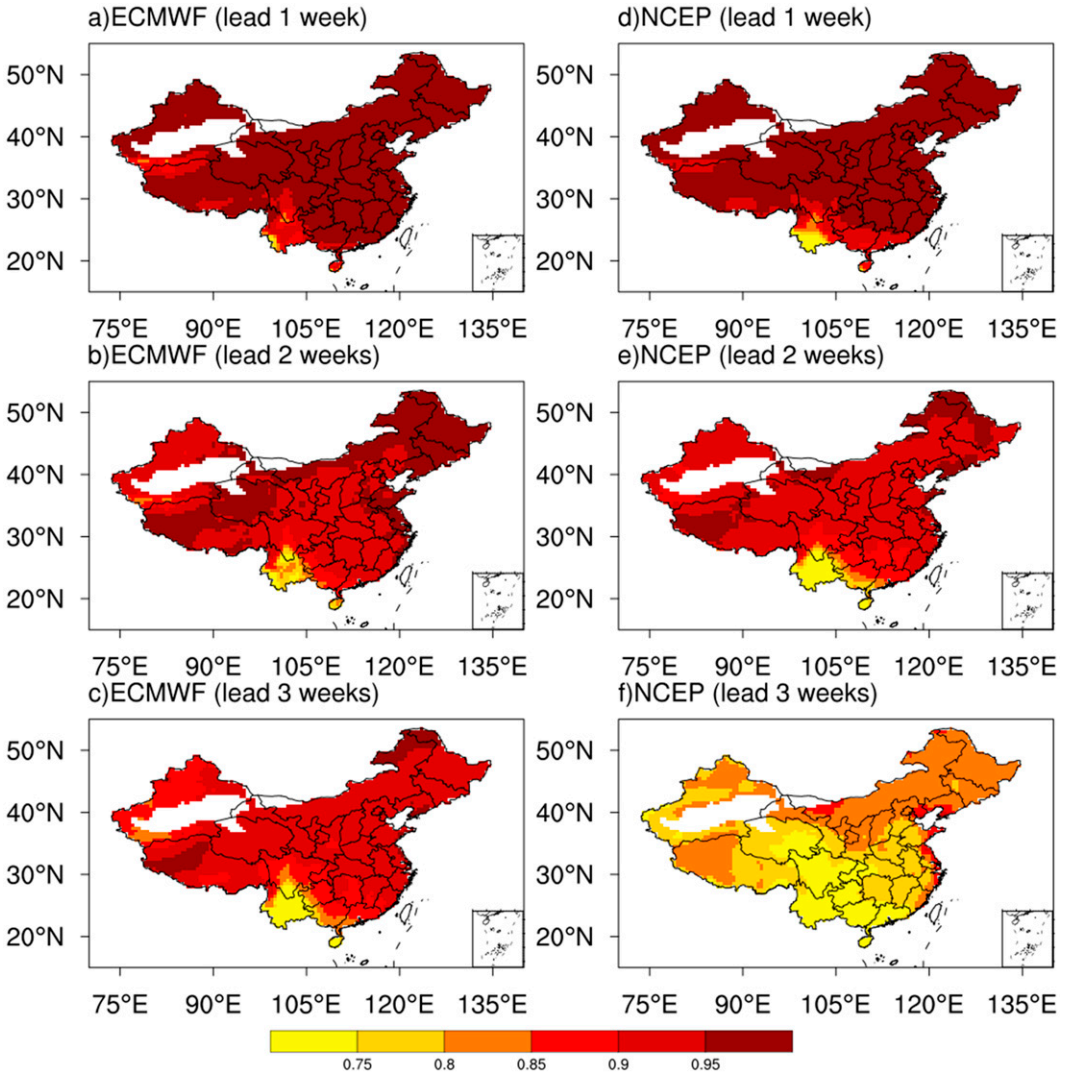

FIG. 6. Temperature correlation for the ECMWF and NCEP hindcasts at different lead times.

0.7 in most areas of eastern China, which is consistent with regional high skill in flash drought forecast (Figs. 4a and 5c). The precipitation correlations decrease significantly at leads of 2–3 weeks (Figs. 7b,c), which can explain the sharp declines in flash drought forecast skill at long leads. The NCEP model has lower precipitation correlation than the ECMWF model especially at the first 2 weeks (Figs. 7d,e), so the former has lower flash drought forecast skill than the latter. The precipitation correlations over the Qinghai–Tibetan Plateau region are higher than other regions, and they do not decline quickly over leads for both the ECMWF and NCEP models (Fig. 7). Whether the S2S models do have higher precipitation predictive skill over the plateau needs further investigation as the observation gauges are limited. The eastern part of the plateau has a large number of flash droughts (Fig. 2), and the good drought forecast skill is consistent with the good precipitation forecast skill over this region. While the flash droughts are very few over the western part of the plateau, there is no obvious connection between the drought forecast skill and precipitation forecast skill.

To explore the potential forecast skill for flash droughts, we calculated the metrics by using a "perfect model" assumption (Luo and Wood 2006; Ma et al. 2018). We computed the potential forecast skill of flash droughts in terms of HIT (Fig. S2), FAR (Fig. S3), and ETS (Fig. 8). After eliminating the model error, the potential hit rates for the ECMWF model can reach 0.5 in most areas at lead 1 week (Fig. S2a), which are much higher the actual hit rates (Fig. 3a). Moreover, the potential false alarm ratios are below 0.6, and the potential ETS values are above 0.3 for most regions (Fig. S3a and Fig. 7a). This suggests a room for model improvement for a better flash drought forecast skill. The results for the NCEP model are similar, although they are worse than those of the ECMWF model. The potential forecast skill represents the ability for model to predict itself, which suggests that the NCEP model's ensemble

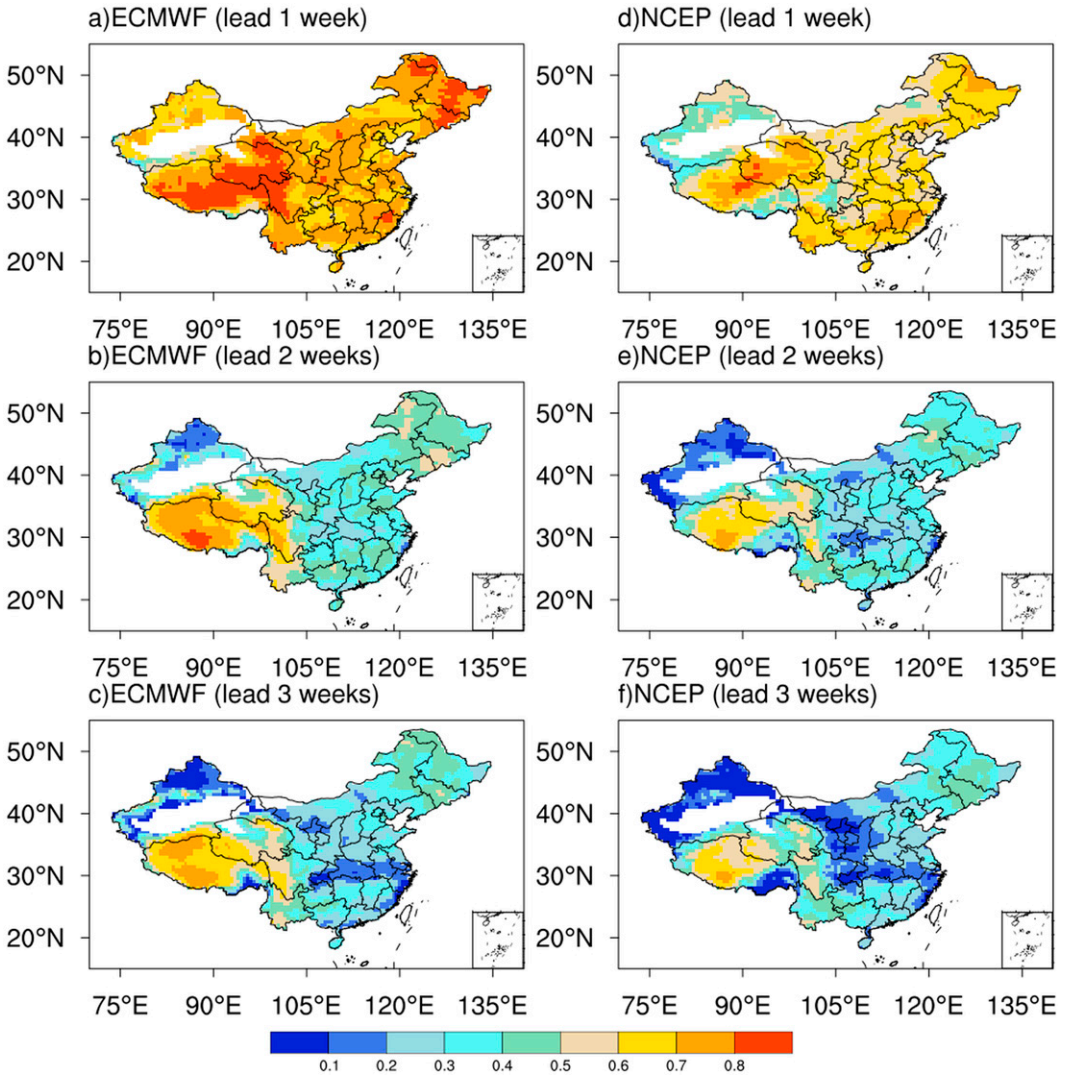

FIG. 7. Precipitation correlation for the ECMWF and NCEP hindcasts at different lead times.

members are more different from each other than those for the ECMWF model. Although the potential forecast skill shows promising results for the first two weeks, the results for the third week are not satisfactory (Fig. 8, Figs. S2 and S3), suggesting the challenge to forecast flash droughts at long leads, even with a perfect model assumption. This study suggests the importance of multimodel ensemble flash drought forecasting. This suggests room for model improvement for a better flash drought forecast skill, at least for the ECMWF model. The results for the NCEP model are similar, although they are worse than those of the ECMWF model. In fact, ECMWF model generates ensembles by perturbing the initial conditions, while the NCEP model generates ensembles by using forecasts at different lead times (e.g., 0-, 6-, 12-, and 18-h lead). The latter method may generate ensembles with a wider spread than the former, but the latter method is proved to be reliable for seasonal drought ensemble forecasting (Yuan and Wood 2013).

c. *Sensitivity of the results to sample size and ensemble members*

Given that the ECMWF model has 22-yr hindcasts (1999–2020) while the NCEP model only has 12-yr hindcasts (1999–2010), we calculated the results for the ECMWF model by using 22-yr hindcasts (1999–2020). Table 2 shows that the improvement of study years in the ECMWF model improves drought hindcast skills. In addition, we extended the sample of the NCEP model from 12 years (1999–2010) to 18 years (1999–2010 and 2015–20) by incorporating the real-time forecast data. Figure 9 and Table 3 show that the hindcast skills are similar to the real-time forecast skills for the NCEP model. We also used different ensemble members for the NCEP model, from 4 to 28 members with different initialization times. Figure 8 and Table 3 show that increasing ensemble members does not improve the skill at lead 1 week, but does improve the skill at lead 2 weeks and lead 3 weeks. Specifically, using 16 ensemble members results in highest HIT and

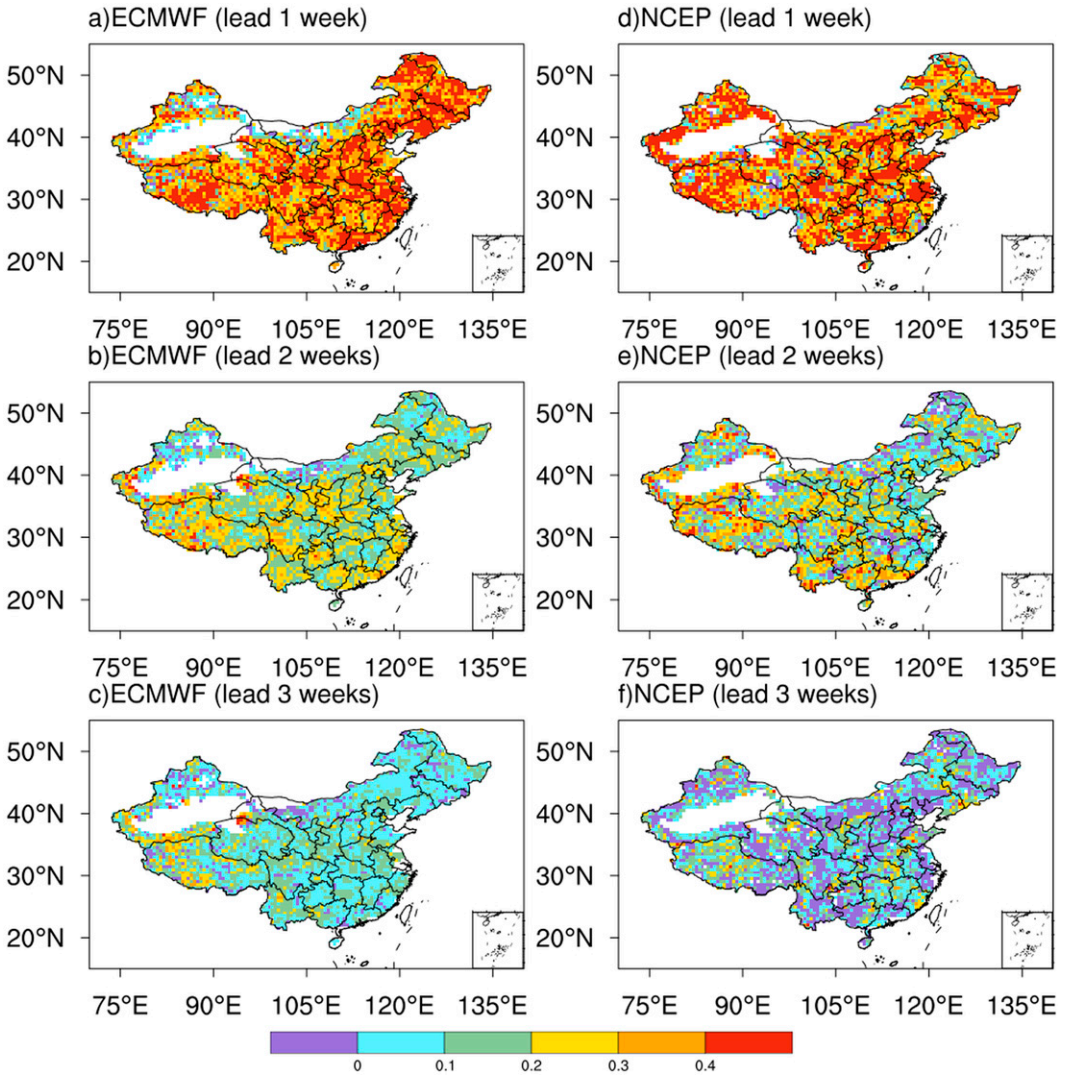

FIG. 8. Potential forecast skill of flash droughts in terms of ETS for (a)–(c) ECMWF and (d)–(f) NCEP at different lead times.

ETS, and lowest FAR. Beyond that, there is no obvious improvement. Table 3 shows that NCEP model with 16 ensemble members has higher ETS than ECMWF model at lead 3 weeks, although the former still has lower ETS than the latter at lead 1 week. Therefore, the sample size and ensemble members cannot explain the skill difference between the ECMWF and NCEP models at short lead, but the NCEP model with sufficient ensemble members can have higher skill than the ECMWF model at long lead. In fact, Yuan and Wood (2013) found the NCEP model is the most reliable model among the North American Multimodel Ensemble (NMME) climate models for forecasting droughts at seasonal time scale, although the hit rate is not the highest. The ensemble members for the ECMWF model start from the same date, while the ensemble members for the NCEP model start from different date. It is possible that the two ensemble generation methods have advantages at short and long leads, respectively.

TABLE 2. National mean results for flash droughts hindcast skill for the ECMWF model for the hindcast period 1999–2020.

	HIT	FAR	ETS
Lead 1 week	0.24	0.62	0.14
Lead 2 weeks	0.11	0.67	0.06
Lead 3 weeks	0.04	0.74	0.02

d. Sensitivity of the results to the calculation of percentiles

Given that flash drought is an extreme event that has a very low probability of occurrence (Yuan et al. 2019), an ideal forecast ensemble requires a reasonable spread to cover the extreme event. Figure 2 shows that even the ECMWF model

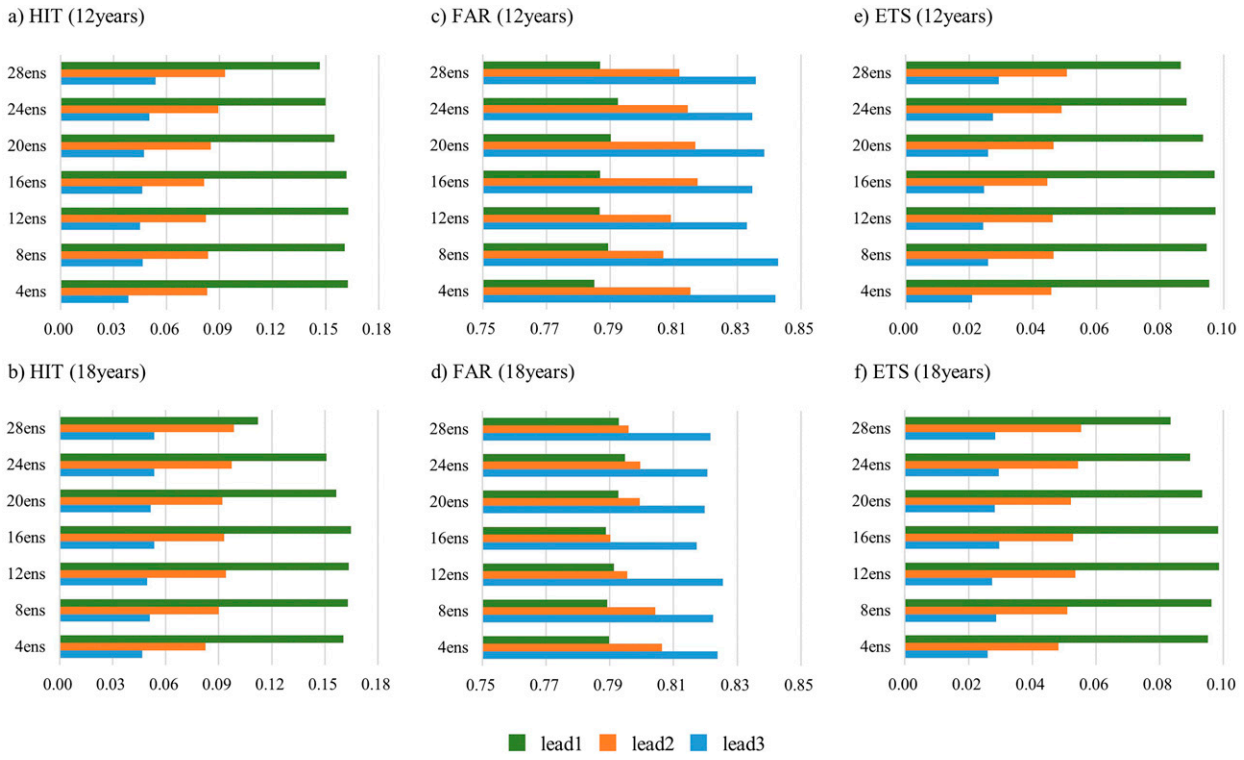

FIG. 9. (a),(b) HIT; (c),(d) FAR; and (e),(f) ETS for NCEP hindcasts of flash drought at lead 1–3 weeks with different ensemble sized and different hindcast/forecast years. Here, 12 years represent the hindcast period 1999–2010, and 18 years represent the combination of hindcast (1999–2010) and real-time forecast (2015–20) periods; 4ens, … , 28ens refer to the ensembles consisting of different initialization times. For example, 8ens represents the 8 ensembles started from 0, 6, 12, … , 42 h before the target initialization date.

with better performance underestimates the FOC, therefore, we plot the results for each ensemble member to check the ensemble spread. Figure 10 shows that the FOC patterns from the control ensemble member (Fig. 10a) and 10 perturbed ensemble members (Figs. 10b–k) are similar, and national mean FOCs are lower than that for ensemble mean (Fig. 10l). These results are based on the percentiles that are calculated from different cumulative distribution functions (CDFs) for different ensemble members. We also collect all hindcasts from all ensemble members to construct a unified CDF and calculate the corresponding percentiles and FOCs of flash droughts based on the unified CDF, but the difference among individual ensemble members does not change much (Fig. 11). The results also suggest that the spread of ECMWF ensemble might be too small to capture the flash droughts.

4. Concluding remarks

In this study, we assessed the subseasonal forecast skill for flash droughts based on the two S2S models. The conclusions are as follows:

1) The S2S hindcasts can capture the spatial distributions of flash drought frequency, especially for the ECMWF model. While the ECMWF and NCEP models only underestimate 5% and 19% of the national mean climatology of flash drought occurrence at lead 1 week, the underestimations increase significantly at leads of 2–3 weeks.
2) Averaged over China, the ECMWF and NCEP models can predict 22% and 16% of the observed flash drought events at 0.5° resolution at lead 1 week. The national mean detectability decreases to 4% at lead 3 weeks, suggesting a grand challenge for predicting flash droughts at subseasonal time scale. The ensemble of the two models increases equitable threat score from ECMWF and NCEP by 8% and 40% for lead 1 week, respectively. In terms of probabilistic forecast, ECMWF also has higher Brier skill score than NCEP especially over eastern China, and ENS has highest Brier score.
3) The perfect forecast skill of flash droughts is higher than the forecast skill for both the ECMWF and NCEP models, suggesting there is room for model improvement. Specifically, the ECMWF model has higher perfect forecast skill than the NCEP model. The assessment of precipitation forecast skill also help to explain the model difference, where the flash drought forecast skill is highly correlated with the temperature and precipitation forecast skill if sufficient drought samples are available during the assessment period.

TABLE 3. The national mean results of the flash drought hindcast skill for the NCEP model with different ensemble sizes and hindcast/forecast periods. Here, 12 years represents the hindcast period 1999–2010, and 18 years represents the combination of hindcast (1999–2010) and real-time forecast (2015–20) periods; 4ens, …, 28ens refer to the ensembles consisting of different initialization times. For example, 8ens represents the 8 ensembles started from 0, 6, 12, …, 42 h before the target initialization date.

	HIT			FAR			ETS		
	Lead 1	Lead 2	Lead 3	Lead 1	Lead 2	Lead 3	Lead 1	Lead 2	Lead 3
NCEP (12 years)									
4ens	0.16	0.08	0.04	0.79	0.82	0.84	0.10	0.05	0.02
8ens	0.16	0.08	0.05	0.79	0.81	0.84	0.09	0.05	0.03
12ens	0.16	0.08	0.05	0.79	0.81	0.83	0.10	0.05	0.02
16ens	0.16	0.08	0.05	0.79	0.82	0.83	0.10	0.04	0.02
20ens	0.16	0.08	0.05	0.79	0.82	0.84	0.09	0.05	0.03
24ens	0.15	0.09	0.05	0.79	0.81	0.83	0.09	0.05	0.03
28ens	0.15	0.09	0.05	0.79	0.81	0.84	0.09	0.05	0.03
NCEP (18 years)									
4ens	0.16	0.08	0.05	0.79	0.81	0.82	0.10	0.05	0.03
8ens	0.16	0.09	0.05	0.79	0.80	0.82	0.10	0.05	0.03
12ens	0.16	0.09	0.05	0.79	0.80	0.83	0.10	0.05	0.03
16ens	0.16	0.09	0.05	0.79	0.79	0.82	0.10	0.05	0.03
20ens	0.16	0.09	0.05	0.79	0.80	0.82	0.09	0.05	0.03
24ens	0.15	0.10	0.05	0.79	0.80	0.82	0.09	0.05	0.03
28ens	0.11	0.10	0.05	0.79	0.80	0.82	0.08	0.06	0.03

The assessments of subseasonal ensemble forecast skill for flash droughts based on two S2S models suggest that the raw soil moisture prediction product is useful during the first week, but the skill is marginal at long leads. Therefore, the S2S soil moisture product might be helpful for flash drought monitoring or watching, while the skillful forecasts at long leads might need additional efforts such as assimilating the information of land surface memory by combining with advanced land surface models (Yuan et al. 2018), correcting precipitation and/or soil moisture predictions by using artificial

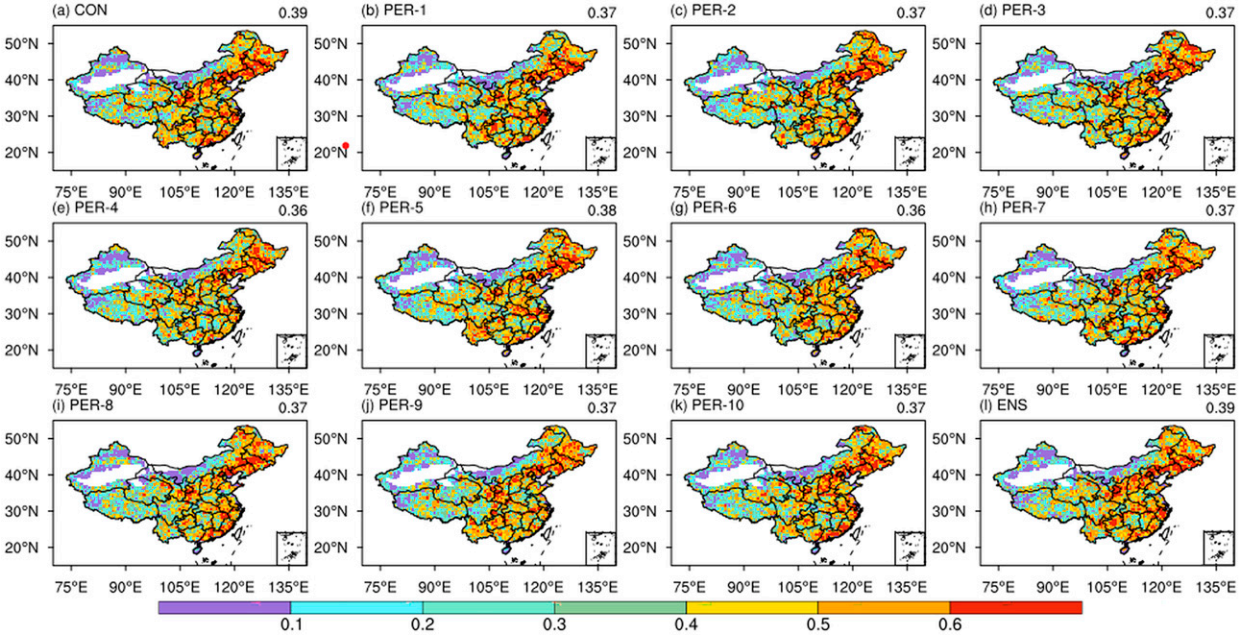

FIG. 10. Spatial patterns of the frequency of occurrence (FOC) of flash droughts (events/year) from (a) ECMWF control hindcast, (b)–(k) 10 perturbed hindcasts, and (l) ensemble mean hindcast during 1999–2020 with lead times of 1 week. The flash droughts were identified based on the data during growing seasons (April–September). The spatial averaging results are shown on the upper-right corners of each panel. Each ensemble member in (a)–(k) and the ensemble mean in (l) has their own cumulative distribution functions (CDFs) for the calculations of percentiles and flash droughts.

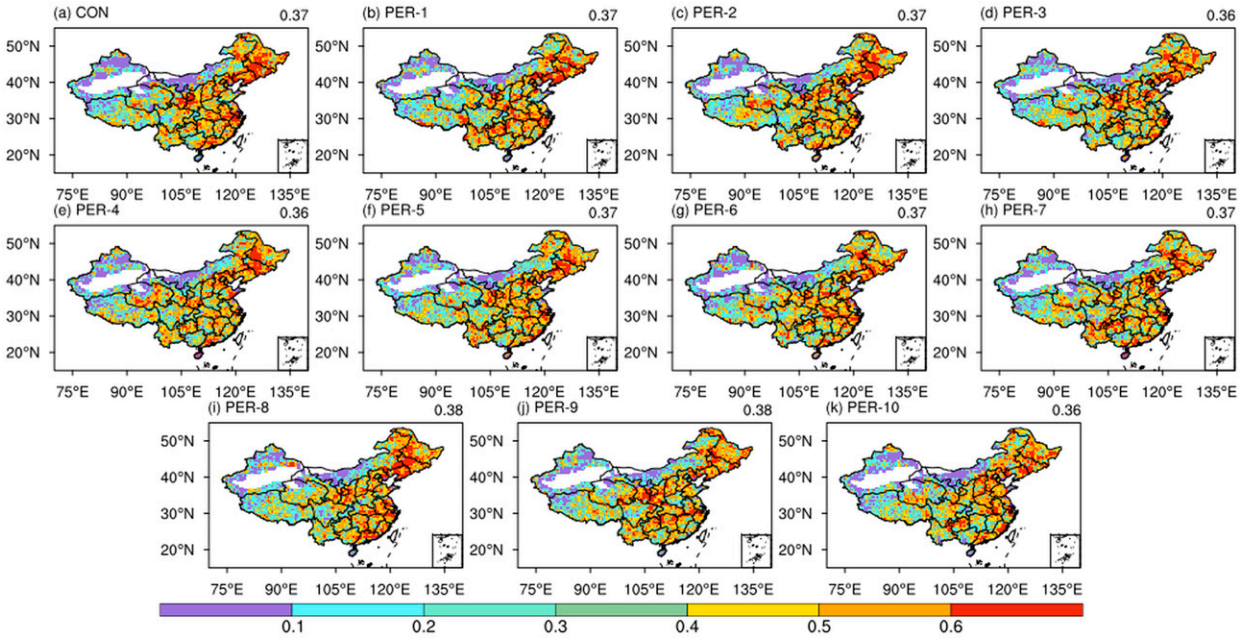

FIG. 11. As in Fig. 10, but for the performance of individual ensemble members using the percentiles calculated from a unified CDF with all ensemble members.

intelligence (Liu et al. 2022), or linking specific flash drought events with subseasonal climate modes (Tian et al. 2021).

Acknowledgments. This work was supported by National Key R&D Program of China (2022YFC3002803), National Natural Science Foundation of China (41875105), and Natural Science Foundation of Jiangsu Province for Distinguished Young Scholars (BK20211540), and the Major Science and Technology Program of the Ministry of Water Resources of China (SKS-2022019).

Data availability statement. The hourly soil moisture datasets from ERA5 reanalysis are publicly available for 1999–2020: https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=form. The daily precipitation datasets from CN05.1 reanalysis are publicly available for 1999–2020: http://ccrc.iap.ac.cn/resource/detail?id=228. The soil moisture and precipitation datasets from S2S are publicly available at 1999–2020 for ECMWF and 1999–2010 for NCEP: https://apps.ecmwf.int/datasets/data/s2s-realtime-instantaneous-accum-ecmf/levtype=sfc/type=pf/.

REFERENCES

Chen, L. G., A. Hartman, B. Pugh, J. Gottschalck, and D. Miskus, 2020: Real-time prediction of areas susceptible to flash drought development. *Atmosphere*, **11**, 1114, https://doi.org/10.3390/atmos11101114.

Christian, J. I., J. B. Basara, J. A. Otkin, E. D. Hunt, R. A. Wakefield, P. X. Flanagan, and X. Xiao, 2019: A methodology for flash drought identification: Application of flash drought frequency across the United States. *J. Hydrometeor.*, **20**, 833–846, https://doi.org/10.1175/JHM-D-18-0198.1.

Ford, T. W., and C. F. Labosier, 2017: Meteorological conditions associated with the onset of flash drought in the eastern United States. *Agric. For. Meteor.*, **247**, 414–423, https://doi.org/10.1016/j.agrformet.2017.08.031.

Hamill, T. M., and J. Juras, 2006: Measuring forecast skill: Is it real skill or is it the varying climatology? *Quart. J. Roy. Meteor. Soc.*, **132**, 2905–2923, https://doi.org/10.1256/qj.06.25.

Hao, Z., X. Yuan, Y. Xia, F. Hao, and V. P. Singh, 2017: An overview of drought monitoring and prediction systems at regional and global scales. *Bull. Amer. Meteor. Soc.*, **98**, 1879–1896, https://doi.org/10.1175/BAMS-D-15-00149.1.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Hoerling, M., J. Eischeid, A. Kumar, R. Leung, A. Mariotti, K. Mo, S. Schubert, and R. Seager, 2014: Causes and predictability of the 2012 Great Plains Drought. *Bull. Amer. Meteor. Soc.*, **95**, 269–282, https://doi.org/10.1175/BAMS-D-13-00055.1.

Hoffmann, D., A. J. E. Gallant, and M. Hobbins, 2021: Flash drought in CMIP5 models. *J. Hydrometeor.*, **22**, 1439–1454, https://doi.org/10.1175/JHM-D-20-0262.1.

Lee, C.-Y., S. J. Camargo, F. Vitart, A. H. Sobel, J. Camp, S. Wang, M. K. Tippett, and Q. Yang, 2020: Subseasonal predictions of tropical cyclone occurrence and ACE in the S2S dataset. *Wea. Forecasting*, **35**, 921–938, https://doi.org/10.1175/WAF-D-19-0217.1.

Liang, M. L., and X. Yuan, 2021: Critical role of soil moisture memory in predicting the 2012 central United States flash drought. *Front. Earth Sci.*, **9**, 615969, https://doi.org/10.3389/feart.2021.615969.

Liu, J., X. Yuan, J. Zeng, Y. Jiao, Y. Li, L. Zhong, and L. Yao, 2022: Ensemble streamflow forecasting over a cascade reservoir

catchment with integrated hydrometeorological modeling and machine learning. *Hydrol. Earth Syst. Sci.*, **26**, 265–278, https://doi.org/10.5194/hess-26-265-2022.

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, https://doi.org/10.1029/2005GL025292.

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

——, L. Luo, A. Ye, and Q. Duan, 2018: Seasonal drought predictability and forecast skill in the semi-arid endorheic Heihe River basin in northwestern China. *Hydrol. Earth Syst. Sci.*, **22**, 5697–5709, https://doi.org/10.5194/hess-22-5697-2018.

Mo, K. C., and D. P. Lettenmaier, 2016: Precipitation deficit flash droughts over the United States. *J. Hydrometeor.*, **17**, 1169–1184, https://doi.org/10.1175/JHM-D-15-0158.1.

——, and ——, 2020: Prediction of flash droughts over the United States. *J. Hydrometeor.*, **21**, 1793–1810, https://doi.org/10.1175/JHM-D-19-0221.1.

Novak, D. R., C. Bailey, K. F. Brill, P. Burke, W. A. Hogsett, R. Rausch, and M. Schichtel, 2014: Precipitation and temperature forecast performance at the Weather Prediction Center. *Wea. Forecasting*, **29**, 489–504, https://doi.org/10.1175/WAF-D-13-00066.1.

Osman, M., B. F. Zaitchik, H. S. Badr, J. I. Christian, T. Tadesse, J. A. Otkin, and M. C. Anderson, 2021: Flash drought onset over the contiguous United States: Sensitivity of inventories and trends to quantitative definitions. *Hydrol. Earth Syst. Sci.*, **25**, 565–581, https://doi.org/10.5194/hess-25-565-2021.

Otkin, J. A., M. Svoboda, E. D. Hunt, T. W. Ford, M. C. Andersoson, C. Hain, and J. B. Basara, 2018: Flash droughts a review and assessment of the challenges imposed by rapid-onset droughts in the United States. *Bull. Amer. Meteor. Soc.*, **99**, 911–919, https://doi.org/10.1175/BAMS-D-17-0149.1.

Pendergrass, A. G., and Coauthors, 2020: Flash droughts present a new challenge for subseasonal-to-seasonal prediction. *Nat. Climate Change*, **10**, 191–199, https://doi.org/10.1038/s41558-020-0709-0.

Prein, A. F., E. Towler, M. Ge, D. Llewellyn, S. Baker, S. Tighi, and L. Barrett, 2022: Sub-seasonal predictability of North American monsoon precipitation. *Geophys. Res. Lett.*, **49**, e2021GL095602, https://doi.org/10.1029/2021GL095602.

Samaniego, L., and Coauthors, 2018: Anthropogenic warming exacerbates European soil moisture droughts. *Nat. Climate Change*, **8**, 421–426, https://doi.org/10.1038/s41558-018-0138-5.

Tian, D., E. F. Wood, and X. Yuan, 2017: CFSv2-based sub-seasonal precipitation and temperature forecast skill over the contiguous United States. *Hydrol. Earth Syst. Sci.*, **21**, 1477–1490, https://doi.org/10.5194/hess-21-1477-2017.

Tian, F., N. P. Klingaman, and B. Dong, 2021: The driving processes of concurrent hot and dry extreme events in china. *J. Climate*, **34**, 1809–1824, https://doi.org/10.1175/JCLI-D-19-0760.1.

Tuel, A., and O. Martius, 2021: A global perspective on the sub-seasonal clustering of precipitation extremes. *Wea. Climate Extremes*, **33**, 100348, https://doi.org/10.1016/j.wace.2021.100348.

Vitart, F., and A. W. Robertson, 2018: The Sub-Seasonal to Seasonal Prediction Project (S2S) and the prediction of extreme events. *npj Climate Atmos. Sci.*, **1**, 3, https://doi.org/10.1038/s41612-018-0013-0.

——, and Coauthors, 2017: The Subseasonal to Seasonal Prediction (S2S) project database. *Bull. Amer. Meteor. Soc.*, **98**, 163–173, https://doi.org/10.1175/BAMS-D-16-0017.1.

Wang, L., X. Yuan, Z. Xie, P. Wu, and Y. Li, 2016: Increasing flash droughts over China during the recent global warming hiatus. *Sci. Rep.*, **6**, 30571, https://doi.org/10.1038/srep30571.

Wang, Y., and X. Yuan, 2021: Anthropogenic speeding up of South China flash droughts as exemplified by the 2019 summer-autumn transition season. *Geophys. Res. Lett.*, **48**, e2020GL091901, https://doi.org/10.1029/2020GL091901.

——, and ——, 2022: Land-atmosphere coupling speeds up flash drought onset. *Sci. Total Environ.*, **851**, 158109, https://doi.org/10.1016/j.scitotenv.2022.158109.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences: An Introduction*. Academic Press, 467 pp.

Wu, H., X. Su, V. P. Singh, K. Feng, and J. Niu, 2021: Agricultural drought prediction based on conditional distributions of vine copulas. *Water Resour. Res.*, **57**, e2021WR029562, https://doi.org/10.1029/2021WR029562.

Wu, J., and X. Gao, 2013: A gridded daily observation dataset over China region and comparison with the other datasets. *Acta Geophys.*, **56**, 1102–1111, https://doi.org/10.6038/cjg20130406.

Yao, M.-N., and X. Yuan, 2018: Evaluation of summer drought ensemble prediction over the Yellow River basin. *Atmos. Ocean. Sci. Lett.*, **11**, 314–321, https://doi.org/10.1080/16742834.2018.1484253.

Yuan, X., and E. F. Wood, 2013: Multimodel seasonal forecasting of global drought onset. *Geophys. Res. Lett.*, **40**, 4900–4905, https://doi.org/10.1002/grl.50949.

——, ——, L. F. Luo, and M. Pan, 2011: A first look at Climate Forecast System version 2 (CFSv2) for hydrological seasonal prediction. *Geophys. Res. Lett.*, **38**, L13402, https://doi.org/10.1029/2011GL047792.

——, 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, https://doi.org/10.1002/2015GL064125.

——, M. Zhang, L. Wang, and T. Zhou, 2017: Understanding and seasonal forecasting of hydrological drought in the Anthropocene. *Hydrol. Earth Syst. Sci.*, **21**, 5477–5492, https://doi.org/10.5194/hess-21-5477-2017.

——, P. Ji, L. Wang, X. Liang, K. Yang, A. Ye, Z. Su, and J. Wen, 2018: High-resolution land surface modeling of hydrological changes over the Sanjiangyuan Region in the eastern Tibetan Plateau: 1. Model development and evaluation. *J. Adv. Model. Earth Syst.*, **10**, 2806–2828, https://doi.org/10.1029/2018MS001412.

——, L. Wang, P. Wu, P. Ji, J. Sheffield, and M. Zhang, 2019: Anthropogenic shift towards higher risk of flash drought over China. *Nat. Commun.*, **10**, 4661, https://doi.org/10.1038/s41467-019-12692-7.

Zhang, M., X. Yuan, J. A. Otkin, and P. Ji, 2022: Climate warming outweighs vegetation greening in intensifying flash droughts over China. *Environ. Res. Lett.*, **17**, 054041, https://doi.org/10.1088/1748-9326/ac69fb.

Zhu, Q., and Y. Wang, 2021: The diagnosis about spatiotemporal characteristics and driving factors of flash drought and its prediction over typical humid and semiarid basins in China. *J. Hydrometeor.*, **22**, 2783–2798, https://doi.org/10.1175/JHM-D-21-0062.1.