Evaluation of Subseasonal Drought Forecast Skill over the Coastal Western United States

Lu Su,[a] Qian Cao,[b] Shraddhanand Shukla,[c] Ming Pan,[b] and Dennis P. Lettenmaier[a]

[a] *Department of Geography, University of California, Los Angeles, Los Angeles, California*
[b] *Center for Western Weather and Water Extremes, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California*
[c] *University of California, Santa Barbara, Santa Barbara, California*

ABSTRACT: Predictions of drought onset and termination at subseasonal (from 2 weeks to 1 month) lead times could provide a foundation for more effective and proactive drought management. We used reforecasts archived in NOAA's Subseasonal Experiment (SubX) to force the Noah Multiparameterization (Noah-MP), which produced forecasts of soil moisture from which we identified drought levels D0–D4. We evaluated forecast skill of major and more modest droughts, with leads from 1 to 4 weeks, and with particular attention to drought termination and onset. We find usable drought termination and onset forecast skill at leads 1 and 2 weeks for major D0–D2 droughts and limited skill at week 3 for major D0–D1 droughts, with essentially no skill at week 4 regardless of drought severity. Furthermore, for both major and more modest droughts, we find limited skill or no skill for D3–D4 droughts. We find that skill is generally higher for drought termination than for onset for all drought events. We also find that drought prediction skill generally decreases from north to south for all drought events.

KEYWORDS: Drought; Subseasonal variability; Hydrology; Soil moisture; Forecast verification/skill

1. Introduction

Drought is among the most damaging, and least understood, of all weather and climate hazards (Pulwarty and Sivakumar 2014). Droughts are usually incremental and can span from a few weeks to decades temporally and from a few hundred to hundreds of thousands of kilometers squared spatially (Pendergrass et al. 2020). Droughts' creeping development is often neglected in the early stages, and the changes accumulate and trigger more severe direct or indirect impacts. Eventually, the unattended creeping development leads to urgent crises that are more costly to deal with (Glantz 2004). The impacts can persist even after the drought itself ends. Therefore, drought is often a "hidden" natural disaster and its risk is underestimated (UNDRR 2019; Pendergrass et al. 2020).

During the past decade, nearly all of the contiguous United States (CONUS) from Colorado to the Pacific coast has suffered from moderate to exceptional droughts (Cook et al. 2018). This includes the continuation of multiyear events (2009–11 and 2013–16) in California (Griffin and Anchukaitis 2014; Seager et al. 2015; Williams et al. 2015) and the U.S. Southwest (Delworth et al. 2015; Seager and Hoerling 2014) and the emergence of significant drought conditions across the Pacific Northwest (Oregon and Washington) in 2015 (Mote et al. 2016). Drought episodes were especially severe in the coastal western United States (including California, Oregon, and Washington). The prolonged severe droughts have stressed water resources management at the regional level (Mann and Gleick 2015; Engström et al. 2020).

As the climate warms, an argument has evolved as to whether drought duration and intensity are increasing (Christensen et al. 2007; Seneviratne et al. 2012; Pendergrass et al. 2020). If so, more foresighted responses that adopt proactive risk mitigation strategies may be necessary (Pulwarty and Verdin 2013; Wilhite et al. 2014). Drought forecast systems in this context would be especially useful (Arsenault et al. 2020; Carrão et al. 2018; Hao et al. 2018). Predictions of drought onset and termination (although evasive to date) in addition to other drought characteristics could provide a foundation for effective proactive drought management.

Seasonal climate forecast systems including the North American Multi-Model Ensemble (NMME) project (Kirtman et al. 2014; Wanders et al. 2017) consistently predicted a false wet 2015/16 winter and forecast a false signal for California drought termination. In contrast, the forecasts and reforecasts from the ECMWF and NCEP CFSv2 models, at the subseasonal-to-seasonal (S2S) (from weeks to 1–2 months) time scale, were able to predict the correct sign of precipitation anomalies (Wang et al. 2017). Wang et al. (2017) shows that what is unpredictable at the seasonal time scale can become predictable at the subseasonal time scale. Recently there has been surging interest in "flash droughts," which are characterized by their sudden onset and rapid intensification and severe impacts (Otkin et al. 2018). While many drought prediction products are updated at monthly time scales, these predictions are of limited value for flash droughts, which develop on shorter time scales (Pendergrass et al. 2020), and they are not useful in determining, for instance, whether individual storms (which

can be forecast with potentially usable accuracy at lead times from 1 week to several weeks) will terminate a drought. This further motivates the need for incorporation of S2S forecasts into drought monitoring and prediction systems. Our study aims to fill a gap in the literature on drought forecast skill to incorporate subseasonal forecasts. Like seasonal drought prediction systems, such as the NOAA Climate Prediction Center's (CPC) seasonal drought outlook, subseasonal drought forecasts derive their skill from knowledge of weather/climate information and initial hydrologic conditions (IHCs) at the onset of the forecast period (Shukla et al. 2012). While subseasonal precipitation forecast skill is generally lower than the skill of forecasts for temperature for the same location and lead time (Monhart et al. 2018; Pegion et al. 2019; Cao et al. 2021), these studies show that there nonetheless is potentially usable precipitation forecast skill to leads of 2–3 weeks. Furthermore, land surface models (LSMs) provide estimates of IHCs that are critical for drought forecasts, particularly when (as in the case of agricultural drought) soil moisture is the metric used to identify droughts (Shukla and Lettenmaier 2011; Shukla et al. 2012). In this respect, the work we report here extends this earlier work to utilize S2S forecasts, which better exploit precipitation (and hence soil moisture) forecast skill at lead times from 1 week to several weeks.

The subseasonal forecasting time scale (the terms subseasonal and subseasonal-to-seasonal are used interchangeably here) is typically defined by lead times ranging from 2 weeks to 1 (or 2) months. This is a critical lead-time window for proactive disaster mitigation efforts such as water resource management for drought mitigation (Mariotti et al. 2018; Vitart and Robertson 2018). However, research on hydrological application of forecasts has not paid much attention to subseasonal lead times until very recently due to a lack of subseasonal meteorological forecast databases (Vitart et al. 2017). Multimodel ensemble approaches have proved to be a successful tool for improving forecast quality for weather and seasonal predictions (Krishnamurti et al. 1999, 2000). They have the advantage of exploiting complementary skill from different models and allow for better estimation of forecast uncertainty (Hao et al. 2018).

As a result of joint efforts between the weather and climate communities, several subseasonal forecast databases have been developed to bridge the weather–climate prediction gap in the S2S range (Mariotti et al. 2018; Merryfield et al. 2020). These include the World Weather Research Programme (WWRP)/World Climate Research Program (WCRP) S2S Prediction Project (Vitart et al. 2017) and the NOAA/Climate Testbed Subseasonal Experiment (SubX) project (Pegion et al. 2019). Recent studies have found that the prediction skill for precipitation and the application to streamflow forecasts of the WWRP/WCRP S2S database varied among predictor combinations, catchments, and dates of prediction, and the skill is frequently less than climatology beyond 2-week lead time (Lin et al. 2018; Pan et al. 2019; Schick et al. 2019).

NOAA's SubX project is different from the WWRP/WCRP reforecasts by including both operational and research models. Furthermore, it is available in near–real time (Pegion et al. 2019). To our knowledge, little research has been done to evaluate the hydrological application of subseasonal

forecasts based on the newly developed SubX dataset. A thorough investigation of the hydrological usefulness of subseasonal drought forecasts based on the SubX dataset could form the foundation of a proactive drought management system.

SubX provides forecasts of climate variables like precipitation and temperature, but not all of them provide hydrologic variables like soil moisture and runoff. However, hydrologic forecasts based on SubX can be produced by using the SubX precipitation (and other surface variables) forecasts to drive a land surface model (see, e.g., Cao et al. 2021). Here, we drive hydrological forecasts from SubX with the Noah Multiparameterization (Noah-MP, version 4.0.1) (Niu et al. 2011). We adopted the WRF-HYDRO recommended physical options, and details are in Text S1 in the online supplemental material. Noah-MP is a state-of-the-art LSM originally intended to be the land surface scheme in numerical weather prediction (NWP) models. It is currently used for physically based, spatially distributed hydrologic simulations within the construct of NOAA's National Water Model (NWM). Noah-MP extends the capabilities of the Noah LSM (Chen et al. 1996; Chen and Dudhia 2001) and incorporates multiple options for key land–atmosphere interaction processes, such as surface water infiltration, runoff, groundwater transfer, and channel routing (Niu et al. 2007, 2011). Noah-MP has been widely used for predicting seasonal climate, weather, droughts, and floods within and beyond CONUS (Zheng et al. 2019).

Given this background, our objectives here are to examine 1) subseasonal forecast skill (at 1–4-week lead times) of drought onset and termination driven by downscaled SubX reforecasts in the coastal western United States and 2) how forecast skill for drought onset and termination vary geographically and with lead times. To achieve these objectives, we first downscaled the SubX reforecasts to a finer spatial resolution (1/16°) from their coarse native resolution (1°), in consideration of the high spatial resolution of our hydrological model. We then implemented the Noah-MP hydrology model over the coastal western United States using downscaled and bias-corrected SubX reforecasts as forcings. Based on the model outputs, we evaluated the SubX-based drought forecasts skill (all of the "forecasts" in this paper technically are reforecasts).

2. Study domain and dataset

a. Study domain

Our study domain is the coastal western United States, consisting of all of California (CA), as well as coastal Oregon (OR) and Washington (WA) (Fig. 1).

b. SubX database

We used six models from the SubX database with 30 ensemble members in total (Table 1) over the reforecast period January 1999–December 2016. The initialization interval of each model is at least once per week, and the lead time is at least 32 days. The temporal resolution of the SubX output is daily, and the raw spatial resolution is 1° × 1°. We downscaled
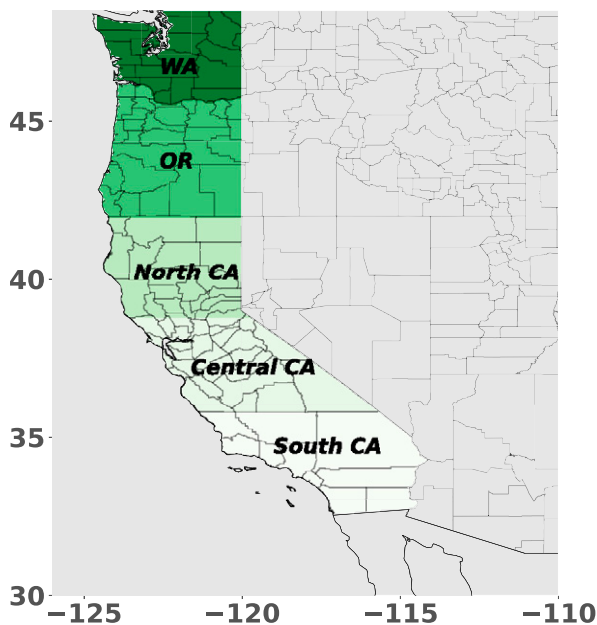
FIG. 1. Study domain, the coastal western United States.

and bias corrected the SubX output to $1/16° \times 1/16°$ as described in section 3a.

3. Methods

a. Downscaling and bias correction

We downscaled the raw SubX output (forcings to Noah-MP) using a statistical downscaling method, bias correction and spatial downscaling (BCSD; Wood et al. 2004). We applied daily BCSD since it has been shown to be an effective approach for removing bias (e.g., Monhart et al. 2018; Baker et al. 2019; Cao et al. 2021) in atmospheric model output. By using this method, we constrained the precipitation temporal variability (wet/dry days) to be the same as in the raw data, which we view as desirable (in contrast to methods like localized constructed analogs (LOCA; Pierce et al. 2014) that attempt to reproduce realistic wet/dry sequences). We applied daily BCSD to precipitation, maximum daily temperature (Tmax), minimum daily temperature (Tmin), and wind speed following the steps in Cao et al. (2021), which can be summarized as follows: 1) we applied spatial (bilinear) interpolation of the $1° \times 1°$ daily SubX forecasts to $1/16° \times 1/16°$, and 2) we bias corrected the outputs from step 1 by each grid point using the daily empirical quantile mapping (QM) method (Wood et al. 2002; Cao et al. 2021). The training dataset we used here is the gridded observation dataset of Livneh et al. (2013) [extended to 2018 as described in Su et al. (2021)].

b. Evaluation of SubX precipitation and temperature

We evaluated SubX forecast skill for precipitation and temperature at different lead times before and after bias correction with BCSD. The skill of forecasts at S2S time scales is typically evaluated in terms of anomalies or differences from the climatology. Following Pegion et al. (2019) and Cao et al. (2021), we used the anomaly correlation coefficient (ACC; Wilks 2006). ACC provides information about how well the variability of the forecast anomalies matches the observed variability. It is calculated as the temporal correlation of anomalies at each grid cell [details of the ACC calculation procedures are as in Cao et al. (2021)]. To evaluate the performance of downscaling methods, we also compared the relative biases for both precipitation and temperature before and after the implementation of BCSD.

c. Hydrological model implementation

We implemented Noah-MP over the coastal western United States, which consists of all of CA, as well as coastal OR and WA. Noah-MP requires meteorological forcings including specific humidity, surface pressure, downward solar and longwave radiation in addition to precipitation, wind speed, air temperature. We calculated the first four variables based on the Mountain Microclimate Simulation Model (MTCLIM) algorithms [implemented as in Bohn et al. (2013), Cao et al. (2021), and Su et al. (2021)] and disaggregated the daily output to 3 hourly (Liang et al. 1994; Bennett et al. 2020).

The prediction skill of subseasonal hydrological forecasts depends on both the IHCs at the time of forecast and the accuracy of forecasts of hydrologic model forcings during the forecast period (Arnal et al. 2017; Li et al. 2009). Before we implemented Noah-MP using SubX forcings, we first ran the model using the Livneh et al. (2013) forcings for the period 1951–2016 and repeated twice. We cropped out the 1961–2016

TABLE 1. List of SubX models used in the research. The community column indicates target users for each model (SEAS for seasonal prediction community, and NWP for numerical weather prediction community).

Model	Members	Initialization day	Forecast length (days)	Community	Reference(s)
NCEP-CFSv2	4	Wednesday	45	SEAS	Saha et al. (2014)
GMAO-GEOS_V2p1	4	Varies	45	SEAS	Koster et al. (2000), Molod et al. (2012), Reichle and Liu (2014), and Rienecker et al. (2008)
RSMAS-CCSM4	3	Sunday	45	SEAS	Infanti and Kirtman (2016)
EMC-GEFS	11	Wednesday	35	NWP	Zhou et al. (2016, 2017) and Zhu et al. (2018)
ECCC-GEPS6	4	Thursday	32	NWP	Lin et al. (2016)
ESRL-FIMr1p1	4	Wednesday	32	NWP	Sun et al. (2018a,b)

TABLE 2. descriptions, and percentiles for drought categories D0–D4.

	D0	D1	D2	D3	D4
Description	Abnormally dry	Moderate drought	Severe drought	Extreme drought	Exceptional drought
Percentiles	<30	<20	<10	<5	<3

period from the second repetition to serve as a baseline run and also to provide assumed perfect IHCs at forecast initiation time for forecasts made over the period 1999–2016. The initialization interval for most SubX models is 7 days, but different models have different initiation days. We output baseline run model states for all the SubX initiation dates, and these states served as the IHCs. For each SubX ensemble member and each identified initialization, we ran Noah-MP for 28 days (4-week forecast).

To assess the hydrological model dependency and the effects of calibration, we also implemented the Variable Infiltration Capacity (VIC) model, version 4.1.2.d (Liang et al. 1994), before and after calibration (details are in Text S2 in the online supplemental material). Overall, those results show that, while there are some differences between models (Noah-MP and VIC) and VIC before and after calibration, our results are not strongly dependent on model and calibration. This is consistent with Mo et al. (2012) who found that differences in soil moisture percentiles during drought periods are modest among different LSMs.

d. Assessment of drought forecast skill

1) IDENTIFICATION OF DROUGHT EVENTS

Soil moisture is an important drought indicator, especially for agricultural droughts. We archived the total column soil moisture and calculated the soil moisture percentile (relative to that grid cell's and that week's total column soil moisture history of all the ensembles of the model) to identify drought events equivalent to D0–D4 droughts as used by the U.S. Drought Monitor (USDM; https://droughtmonitor.unl.edu/About/WhatistheUSDM.aspx) (see also Table 2).

2) EVALUATION SKILL

We evaluated the probabilistic drought forecast skill of all six SubX models using 30 ensemble members. The evaluation metrics we used include debiased Brier skill score (BSS; Weigel et al. 2007), bias score (BS), probability of detection (POD), false alarm ratio (FAR), equitable threat score (ETS), and Heidke skill score (HSS). We discuss these skill measures and our applications briefly below.

(i) BSS

The Brier skill score (Wilks 2006) is widely used to measure the mean square error of probability forecasts for binary events. It is, however, sensitive to small ensemble sizes. To overcome this issue, we used the debiased BSS, which incorporates a correction term in the denominator of the Brier score (DeFlorio et al. 2019). BSS is calculated as follows:

$$\text{BSS} = 1 - \frac{\text{BS}}{\text{BS}_{\text{ref}} + D}, \tag{1}$$

$$\text{BS} = \frac{1}{N}\sum_{i=1}^{N}(P_i - O_i)^2, \tag{2}$$

$$\text{BS}_{\text{ref}} = \frac{1}{N}\sum_{i=1}^{N}(P_{\text{clim}} - O_i)^2, \quad \text{and} \tag{3}$$

$$D = \frac{1}{M}P_{\text{clim}}(1 - P_{\text{clim}}), \tag{4}$$

where P_i is the forecast skill for drought onset/termination and is determined by the fraction of the ensemble members that predicted drought onset/termination for a single reforecast; O_i shows whether the observed drought onset/termination occurs (1 if yes; 0 if no); N is the number of reforecast droughts for the grid cell/region (varies for each grid cell/region); M is the ensemble size (30 here); and P_{clim} is the probability of the reference climatology. BSS ranges from $-\infty$ to 1. Positive values indicate that the reforecast skill is higher than the climatological forecast skill.

(ii) Contingency table

We evaluated the forecast of drought onset/termination, where a dichotomous forecast indicates whether an event will happen or not. To verify this type of forecast we start with a contingency table that shows the frequency of "yes" and "no" forecasts and occurrences (Table 3). The four combinations of forecasts (yes or no) and observations (yes or no) are

1) hit: event was forecast to occur, and it did occur;
2) miss: event was forecast not to occur, but it did occur;
3) false alarm: event was forecast to occur, but it did not occur; and
4) correct negative: event forecast was not to occur, and it did not occur.

We calculated a variety of categorical statistics from the elements in the contingency table to describe particular aspects of forecast performance, as follow.

TABLE 3. Contingency table.

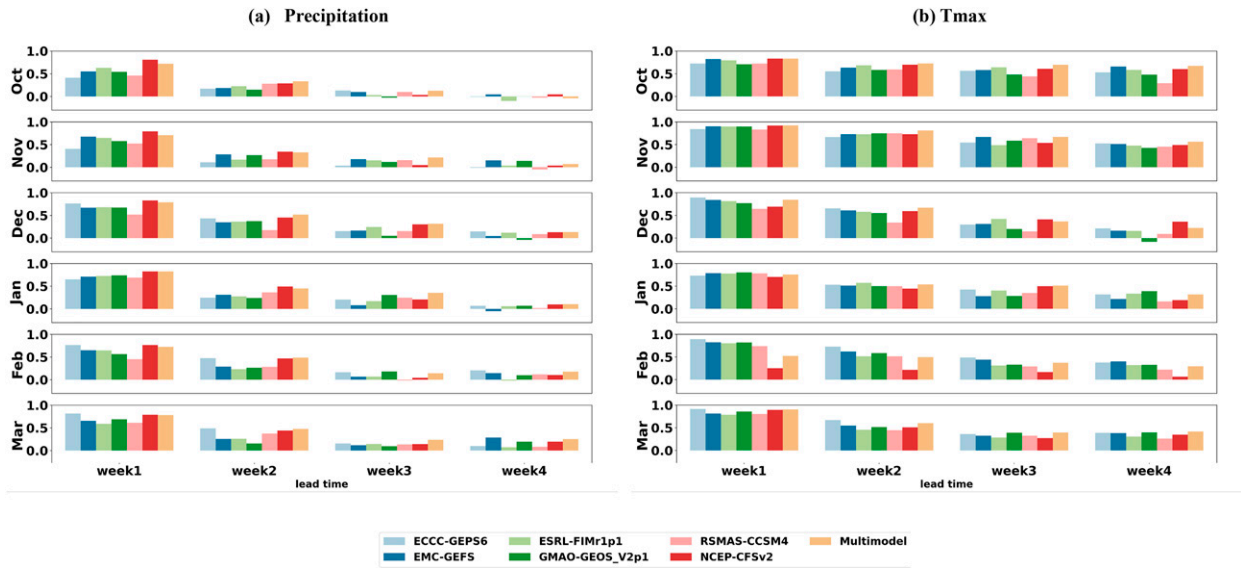|	Observed		
Forecast	Yes	No	Total
Yes	Hits	False alarms	Forecast yes
No	Misses	Correct negatives	Forecast no
Total	Observed yes	Observed no	

FIG. 2. (a) Precipitation and (b) Tmax prediction skill (as measured by the ACC) of SubX models averaged over the coastal western United States for leads 1–4 weeks (without bias correction).

(iii) Bias score (bias)

Bias score,

$$\text{bias} = \frac{\text{hits} + \text{false alarms}}{\text{hits} + \text{misses}}, \qquad (5)$$

indicates how the forecast frequency of "yes" events compared to the observed frequency of "yes" events. It ranges from 0 to ∞, with 1 being a perfect score. It indicates whether the forecast system tends to underforecast (bias < 1) or overforecast (bias > 1) events. It only measures relative frequencies

and does not measure how well the forecast corresponds to the observations.

(iv) Probability of detection (also known as hit rate)

Probability of detection,

$$\text{POD} = \frac{\text{hits}}{\text{hits} + \text{misses}}, \qquad (6)$$

tells us what fraction of the observed "yes" events were correctly forecast. It ranges from 0 to 1, with 1 being a perfect
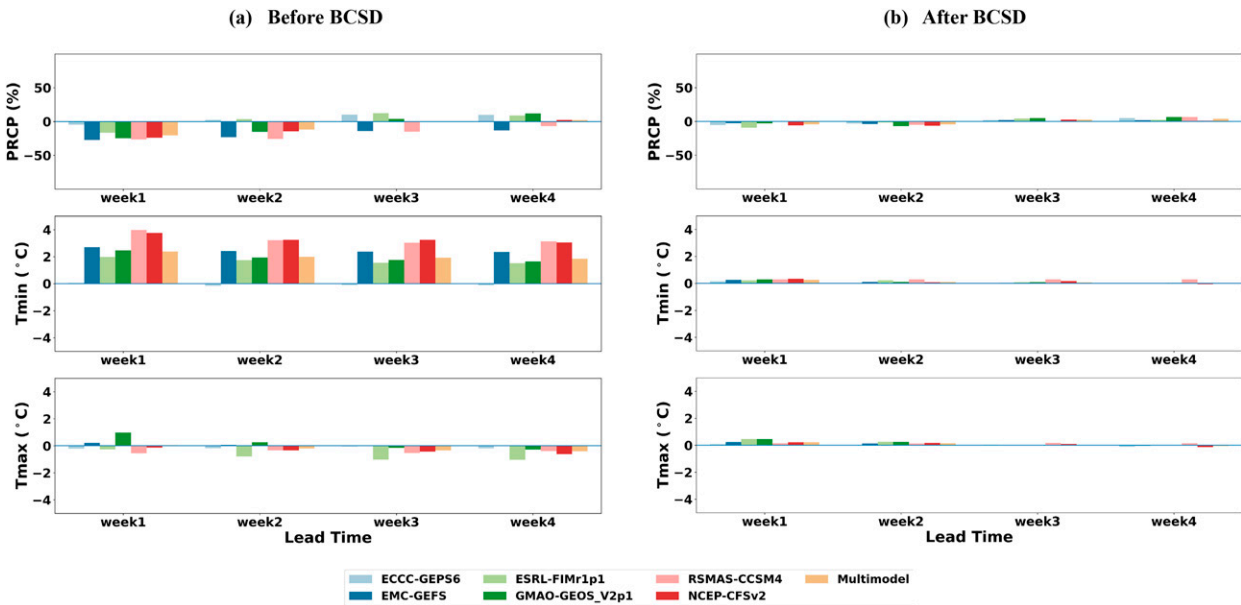

FIG. 3. (top) Precipitation, (middle) Tmin, and (bottom) Tmax bias of SubX models averaged over representative basins and over October–March (a) before and (b) after bias correction.
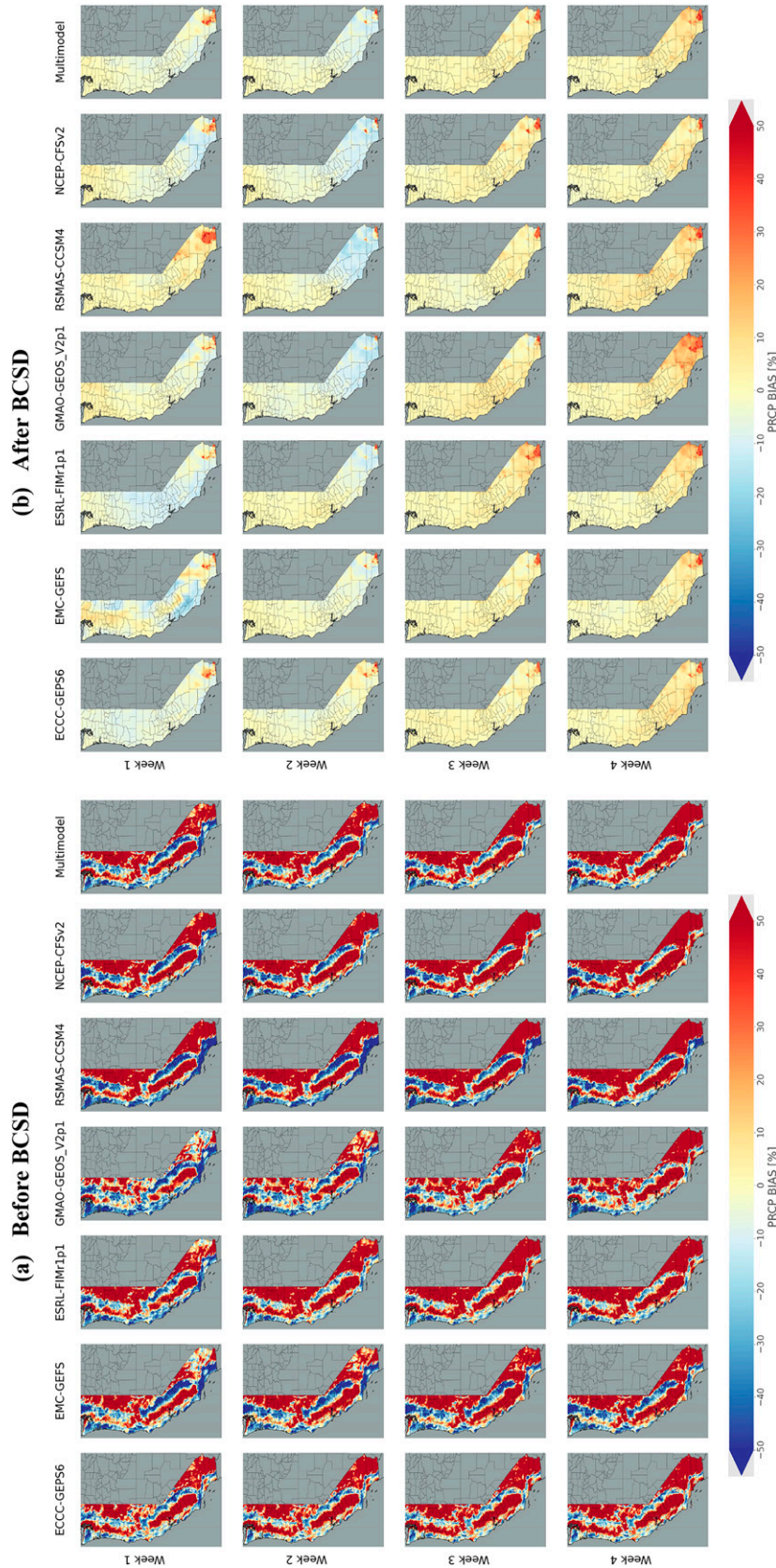
Fig. 4. Spatial distribution of precipitation bias of SubX models over October–March (a) before and (b) after bias correction.
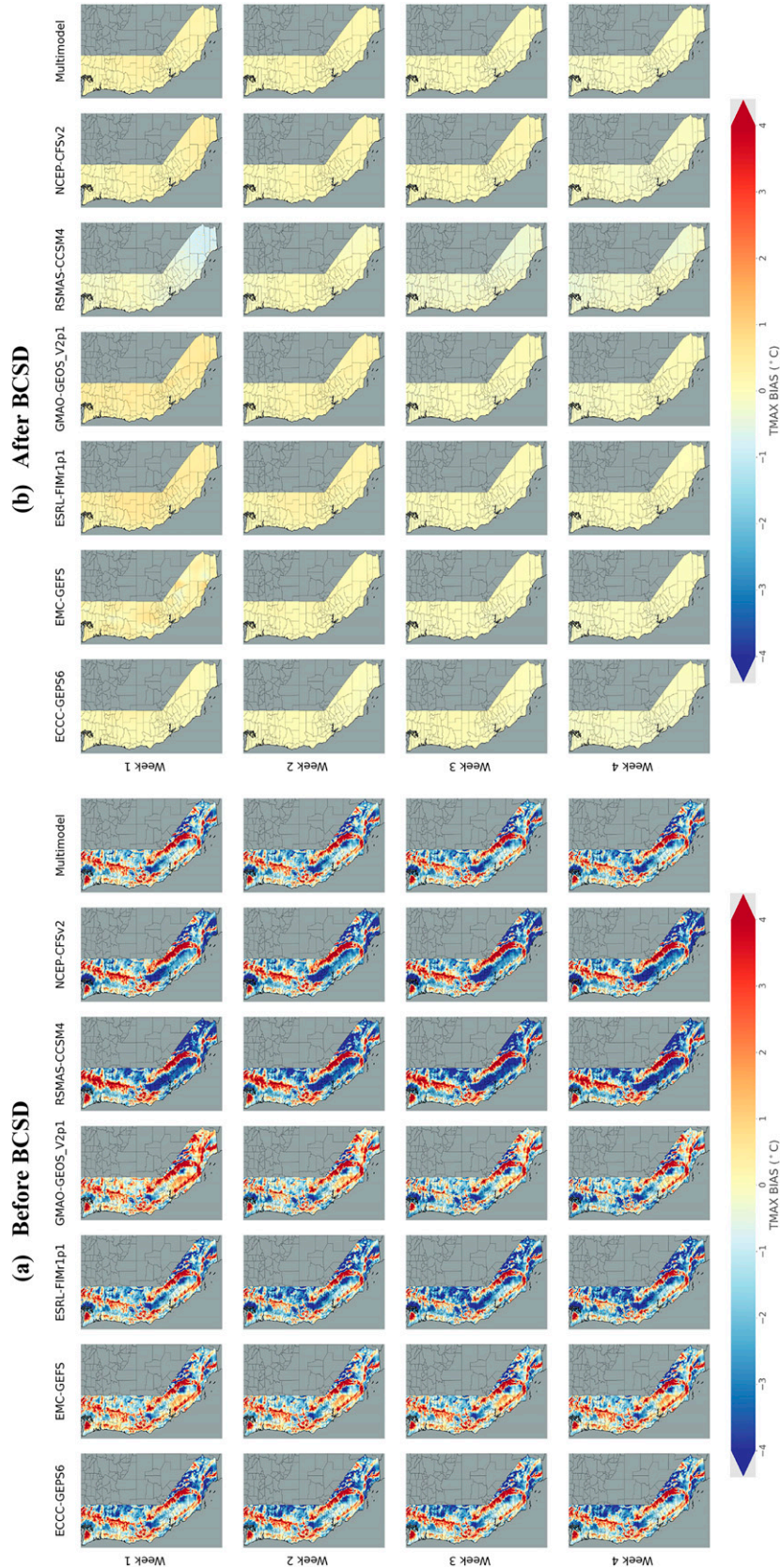
FIG. 5. As in Fig. 4, but for Tmax bias.

score. POD is sensitive to the climatological frequency of the event and is most informative for rare events.

(v) False alarm ratio

FAR,

$$\text{FAR} = \frac{\text{false alarms}}{\text{hits} + \text{false alarms}}, \tag{7}$$

gives the fraction of predicted "yes" events that actually did not occur (i.e., were false alarms). It ranges from 0 to 1, with 0 being a perfect score. FAR is sensitive to false alarms but ignores misses and should be used in conjunction with POD (above).

(vi) Equitable threat score (also known as Gilbert skill score)

ETS,

$$\text{ETS} = \frac{\text{hits} - \text{hits}_{\text{random}}}{\text{hits} + \text{misses} + \text{false alarms} - \text{hits}_{\text{random}}}, \tag{8}$$

where

$$\text{hits}_{\text{random}} = \frac{(\text{hits} + \text{misses})(\text{hits} + \text{false alarms})}{\text{total}}. \tag{9}$$

measures the fraction of observed events that were correctly predicted, adjusted for hits associated with random chance (e.g., it is easier to correctly forecast precipitation occurrence in a wet climate than in a dry climate). It ranges from $-1/3$ to 1; 0 indicates no skill, and 1 is a perfect score. ETS is often used in the verification of precipitation in NWP models because its "equitability" allows scores to be compared more fairly across different regimes.

(vii) Heidke skill score (also known as Cohen's k)

HSS,

$$\text{HSS} = \frac{(\text{hits} + \text{correct negative}) - (\text{expected correct})_{\text{random}}}{N - (\text{expected correct})_{\text{random}}}, \tag{10}$$

where

$$(\text{expected correct})_{\text{random}} = \frac{1}{N}(A + B),$$

$$A = (\text{hits} + \text{misses})(\text{hits} + \text{false alarms}),$$

$$B = (\text{correct negatives} + \text{misses})(\text{correct negatives} + \text{false alarms}), \quad \text{and}$$

$$N = \text{hits} + \text{misses} + \text{false alarms} + \text{correct negatives},$$

measures the fraction of correct forecasts after eliminating those forecasts that could be correct purely because of random chance. It ranges from -1 to 1; 0 indicates no skill, and 1 is a perfect score. HSS is used in NOAA's Climate Prediction Center (https://www.cpc.ncep.noaa.gov/products/predictions/90day/skill_exp.html).
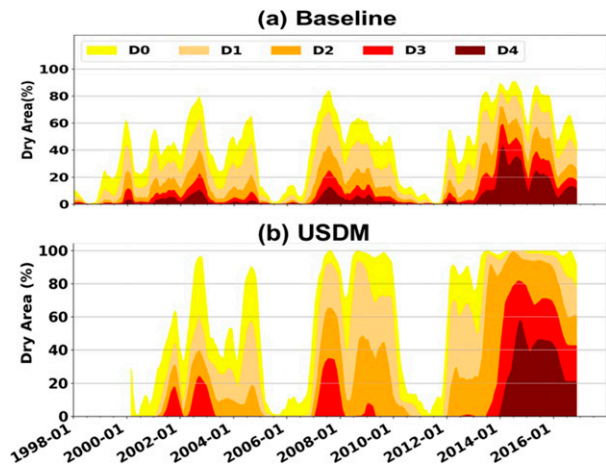

FIG. 6. California drought (D0–D4) area time series for different drought levels from (a) baseline [driven by Livneh et al. (2013) forcing] and (b) USDM.

4. Results

a. Evaluation of SubX reforecasts

1) PRECIPITATION AND TEMPERATURE SKILL

We examined the precipitation and temperature skill of the individual SubX models (raw data, 1° resolution), as well as the multimodel ensemble mean (denoted as "Multimodel"), at lead times of 1–4 weeks averaged over the coastal western United States for each month during the October–March period separately (see Fig. 2). We chose to focus our evaluation on the cool season months October–March as precipitation is generally much lower over most of our domain in the warm season. Figure 2a shows that precipitation skill (as measured by ACC) drops rapidly by approximately 40% after week 1. Almost all models have positive ACC in all months, but by week 3, some models show almost zero ACC in certain months. Among individual models, NCEP-CFSv2 performs best in weeks 1–2, with skill similar to Multimodel. However, the model performance at longer lead times varies by months.

Figure 2b shows the forecast skill for temperature (the pattern for Tmin is similar, so we only show Tmax here). The temperature of SubX models individually as well as their multimodel mean shows statistically significant (different from zero) skill for all lead times in most conditions. Similar to precipitation, Tmax skill drops quickly after week 1. Tmax shows higher skill than precipitation for all leads and shows fewer negative ACC values in weeks 3–4. Overall, multimodel shows consistently statistically significant ACC across all lead times for both precipitation and temperature. The precipitation and temperature skill we found is consistent with previous studies of SubX (Cao et al 2021; DeAngelis et al. 2020).

2) PERFORMANCE OF DAILY BCSD

The difference in precipitation and temperature skill (as measured by ACC) before and after applying daily BCSD is
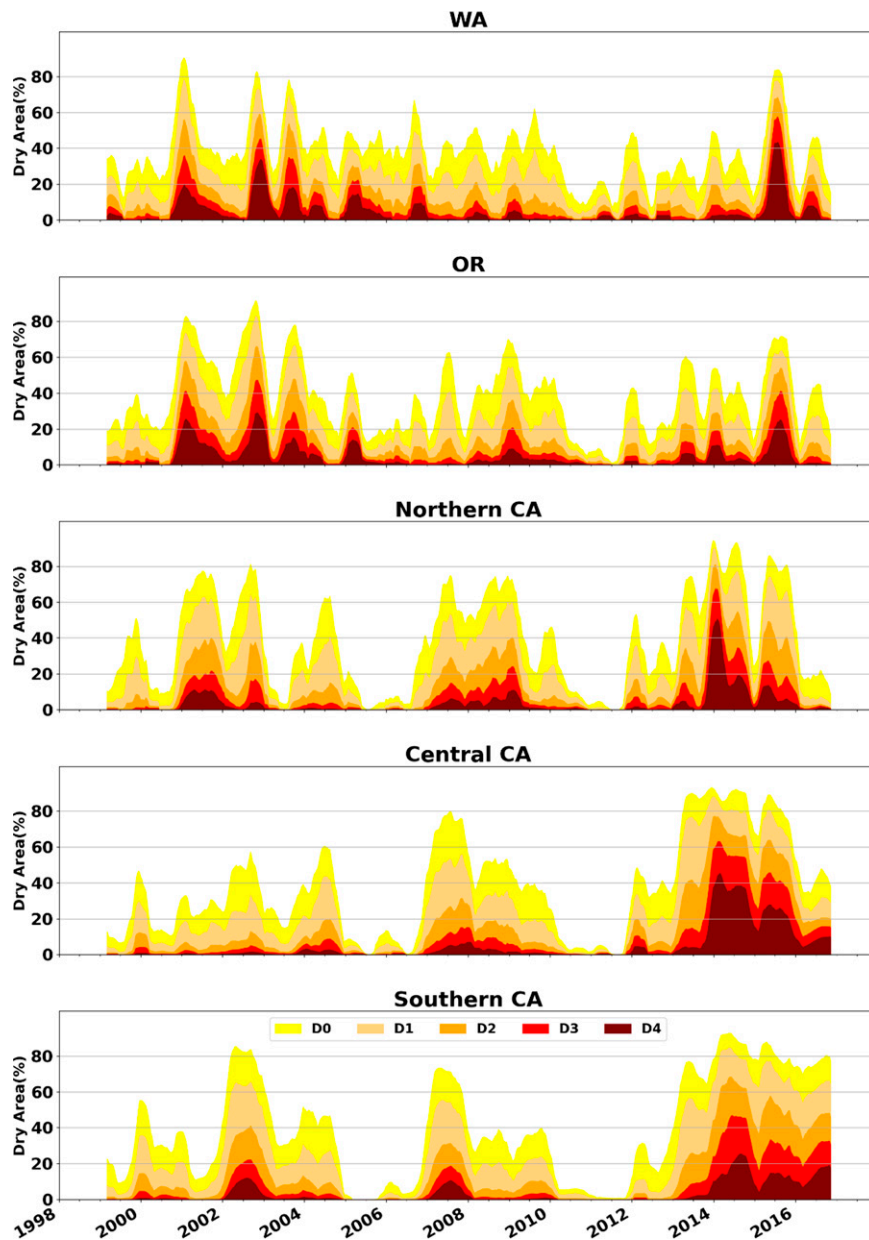
FIG. 7. Baseline drought area time series for different drought levels for five subregions (coastal Washington, coastal Oregon, northern California, central California, and southern California from north to south).

small. This meets our expectation since the QM is performed in a lead-time-dependent manner. Figures 3–5 show the average relative bias [(model − observation)/observation; %] for precipitation forecasts and bias (model − observation) for temperature forecasts before and after applying daily BCSD, averaged over October–March. Before applying daily BCSD, the absolute relative biases of precipitation were up to 80% across models and over weeks 1–4. They were reduced to below 6% after applying BCSD. The biases in temperature were also reduced from up to 3.5°C to below 0.5°C after applying BCSD (Fig. 3). The bias maps before and after BCSD also

show that the biases were essentially removed after applying BCSD (Figs. 4 and 5).

b. Hydrologic model evaluation

We examined model performance of the baseline run, forced by the Livneh et al. (2013) data with hourly disaggregation. We evaluated California drought area history for various drought levels (D0–D4 drought based on USDM) in comparison with the USDM. The drought area time series in baseline run and USDM are highly consistent with correlation coefficients ranging from about 0.8 for D0 to 0.6 for D4 (Fig. 6).
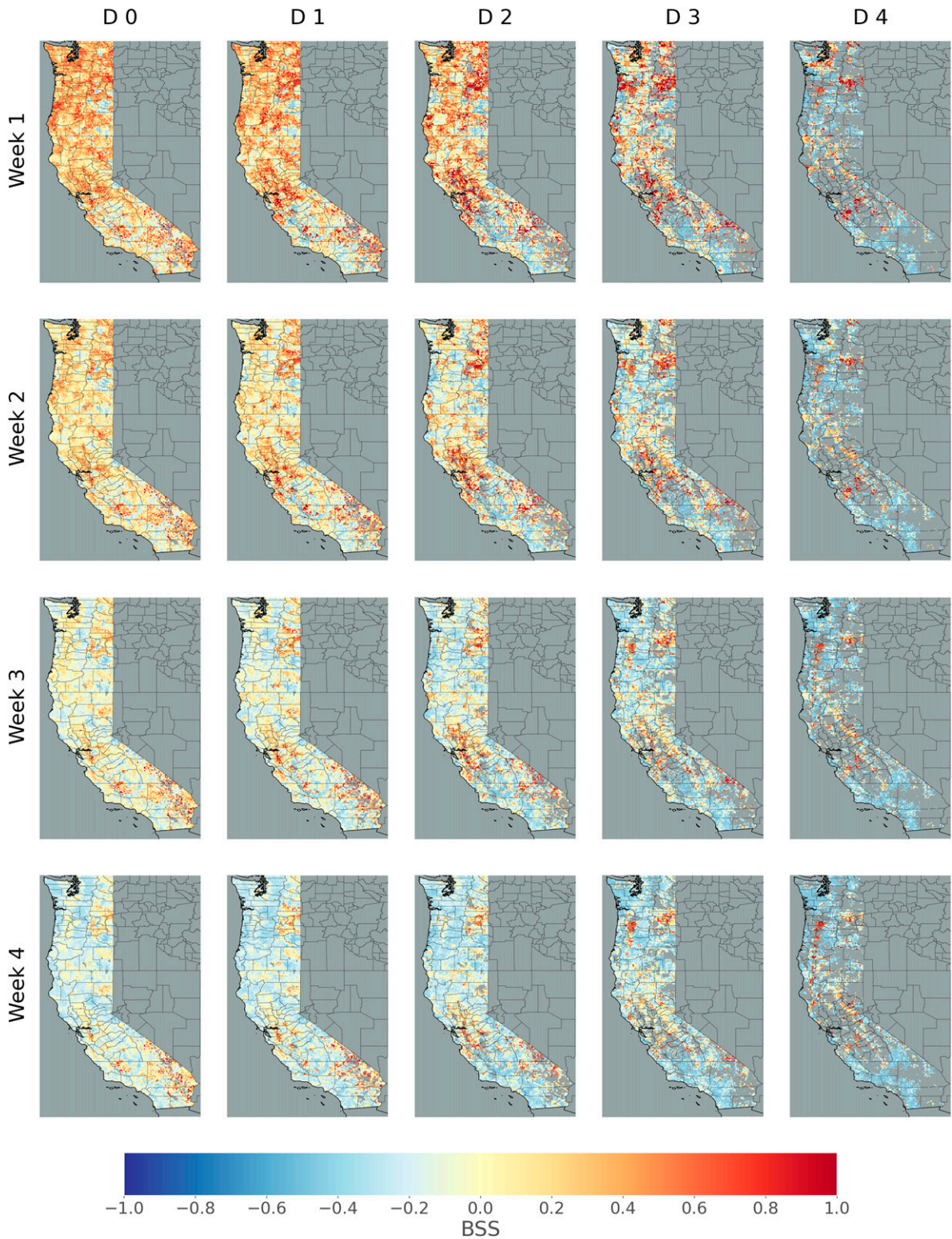
FIG. 8. SubX-based debiased BSS for lead weeks 1–4 for drought termination. The columns show results for drought levels D0–D4; the rows show leads from week 1 to 4. Blank areas denote no drought at this level in this location.
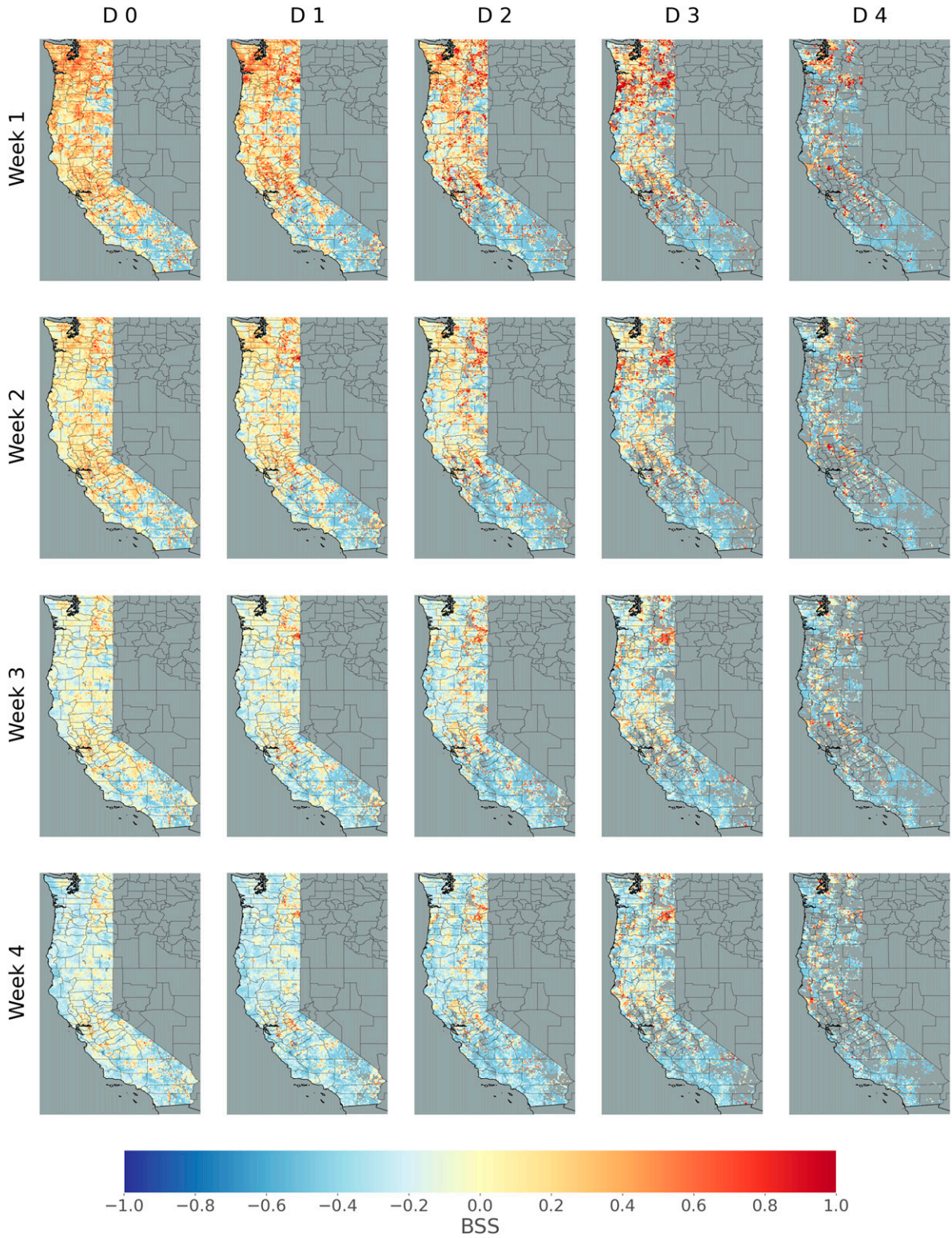
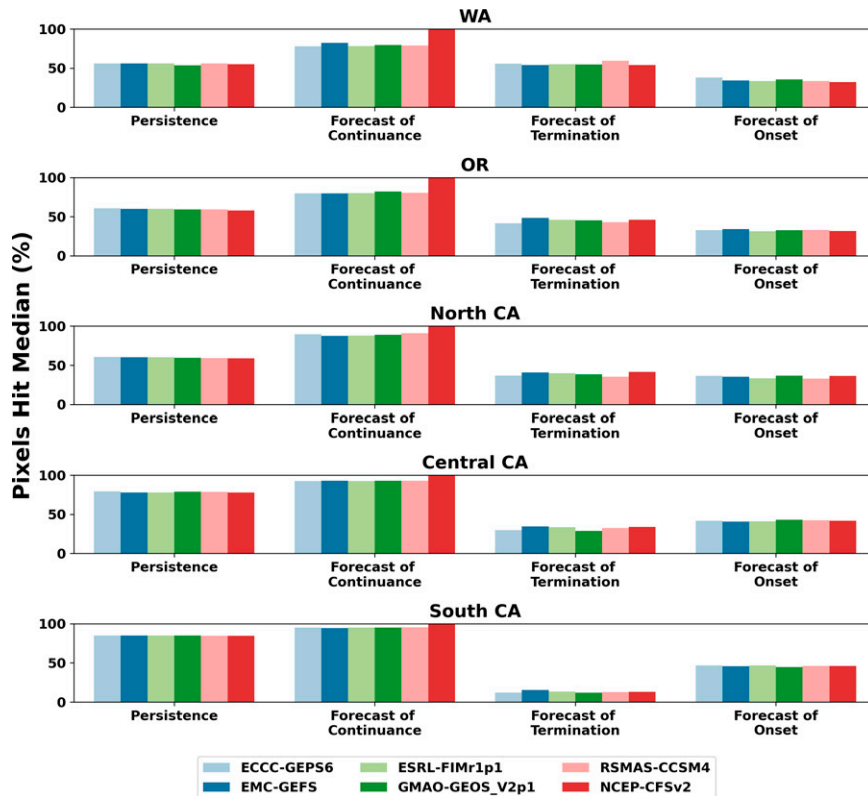FIG. 9. As in Fig. 8, but for drought onset.

Fɪɢ. 10. Drought persistence, continuance, termination, and onset forecast skill for D1 drought at 2-week lead time by subregions and by SubX models.

We further compared the drought area time series for different drought levels in five subregions (coastal Washington, coastal Oregon, northern California, central California, and southern California from north to south; see Fig. 7). We found that drought duration becomes longer and that drought spatial coverage becomes larger from north to south. There are more small drought events in the north, and the droughts in the south are more prolonged.

It is important to note that our results are from the Noah-MP model with the Livneh forcing as the truth. Use of observed soil moisture was not feasible because soil moisture observations are sparsely distributed and in most cases are only available for a decade or so at most. We nonetheless argue that use of model output soil moisture is plausible based on our past work and work of others. For instance, Su et al. (2021) compared the Livneh et al. (2013) forced Noah-MP simulated soil moisture with observed soil moisture from USDA/NRCS SCAN (Soil Climate Analysis Network) across CONUS. Their results showed in general that the spatial patterns of abnormally low soil moisture in the Noah-MP model constructions are similar to those in the observations. Furthermore, as shown in Text S2 in the online supplemental material and noted in section 3c, our comparison here of Noah-MP soil moisture with VIC soil moisture yielded similar results. We might, alternatively, have used soil from one of several coupled land–atmosphere reanalyses, e.g., ERA-5 (ECMWF 2017). ERA-5 soil moisture was found to have the highest skill among reanalysis products when compared with in situ observations of soil moisture by Alessi et al. (2022) and Li et al. (2020). However, it was less accurate than soil moisture produced by the LSM-based North American Land Data Assimilation System (NLDAS) and, in particular, the Noah LSM (Xia et al. 2012; Alessi et al. 2022). We opted therefore not to use reanalysis soil moisture (e.g., ERA5) in consideration of the above studies and also because of issues of root-zone soil moisture discontinuities at the transition points of some of the ERA5 production streams (Hersbach et al. 2020).

c. Assessment of drought forecast skill

Figure 8 shows the SubX-based BSS values for major drought termination at lead weeks 1–4. Here we define major droughts at the gridcell level as 1) the drought period is greater than 50 days and 2) the drought event is separated by at least 30 days from any other drought. The drought termination and onset forecast is defined as a hit when the date that was forecast and the observed date fall within a 1-week window. We found that drought termination skill is highest for D0 drought and lead week 1. Here we show median results of the 30 ensembles. At lead week 1, we see widespread high skill (BSS score higher than 0.4–0.5) for droughts D0–D2 (except for southern CA for D2; Fig. 8). The skill drops to negative for D3 in large parts of southern and central CA and part of OR. The decreasing skill spreads farther in CA and OR for D4. At lead week 2, the skills for D0–D2 are still relatively

(a) Best condition
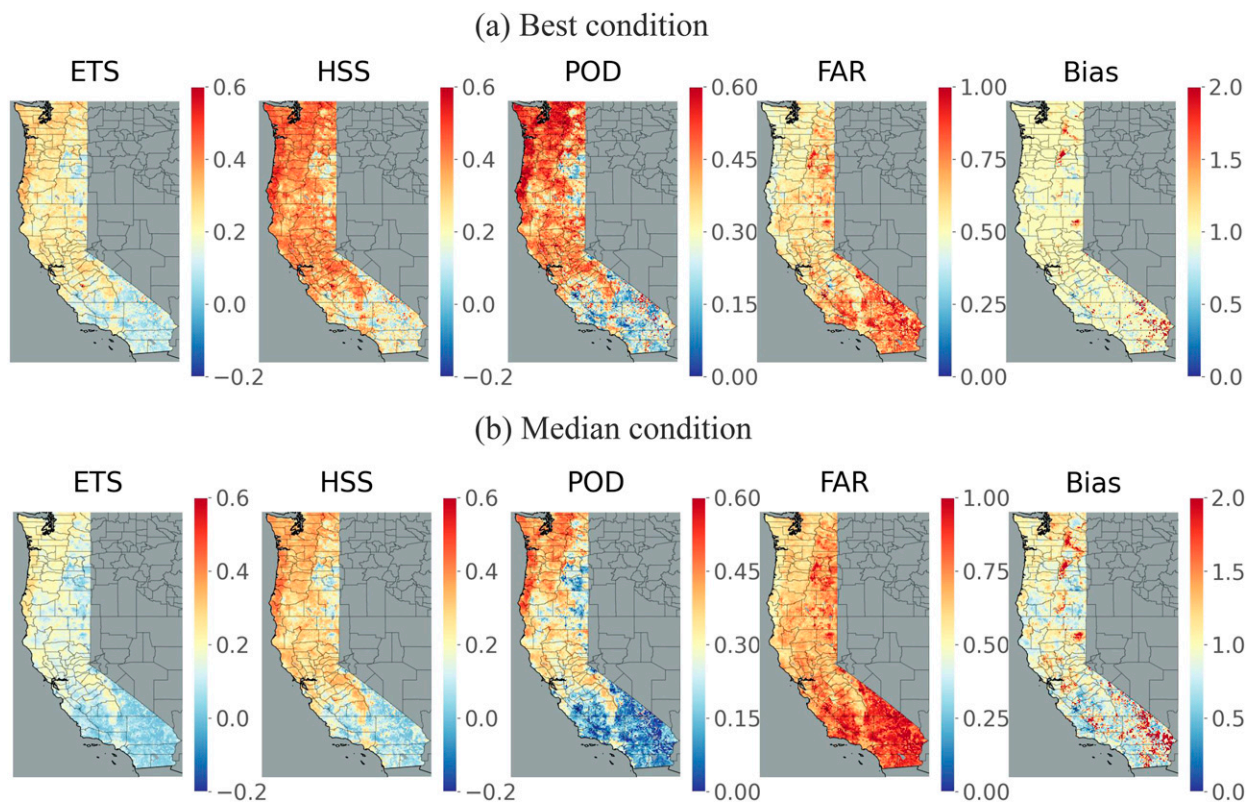

(b) Median condition

FIG. 11. ETS, HSS, POD, FAR, and bias score for drought termination in (a) best condition and (b) median condition across all ensembles at 2-week lead time.

high (BSS score around 0.2–0.3 for the most part, except for southern CA for D2). We see more widespread negative skill in D2–D4 when compared with that at week 1. At lead week 3, there is some limited skill for D0–D2. At week 4, most of our study domain shows no skill for D0–D4 (except a small part of inland southern CA and WA). Overall, the skill decreases as the drought severity increases and also as the lead time increases. From a spatial perspective, skill decreases from north to south. Figure 9 shows the SubX-based BSS values for drought onset at lead weeks 1–4. We see usable onset skill in lead weeks 1 and 2 for droughts D0–D2 over most of WA, OR northern, and central CA. Overall, onset skill is a little lower than termination skill. The onset skill also decreases with drought severity and lead time and decreases from north to south. To reduce noise spatially, we averaged the soil moisture for the subregions shown in Fig. S5 in the online supplemental material and assessed the drought forecast skills at different subregions (see Text S3 in the online supplemental material for details). The skills at the subregion level are generally consistent with what we found from grid-cell-based skills.

The drought forecast skill is highly related to precipitation forecast skill. Li et al. (2021) found a similar degradation pattern of SubX precipitation forecast skill from north to south over the coastal western United States for most of the models and at all lead times (weeks). This might explain the north to south decreasing drought forecast skill we found here.

Atmospheric rivers (ARs) play a critical role as a common cause of the end of droughts on the West Coast (Dettinger 2013). The high skill of drought termination at lead weeks 3–4 in southern CA and WA might be related to the high AR forecast skill in these regions. DeFlorio et al. (2019) found isolated positive skill over these locations at weeks 3–4 lead for strong AR activities in some of the SubX models.

Figure 10 shows forecast POD for major D1 drought continuance, termination, and onset for the five subregions and for different models at 2-week lead time. We summarized the POD (hit rates) based on the percent detection at the grid cell level. A forecast of drought continuance is counted as hit when the drought remains through the predicted period. The forecast of continuance is evaluated relative to persistence, defined as drought conditions assumed to persist through the period (if there is no drought in the beginning, then it is assumed no drought in the end; if there is drought in the beginning, then it is assumed drought in the end). The figure shows that skill for forecasts of continuance is consistently high in all regions and across all models. Skill for forecasts of termination is higher in the north than in the south. Except for forecasts of termination in WA, which have skill comparable to persistence, all other regions' onset and termination forecast skill are lower than persistence. We see very low forecast termination skill and very high continuance skill in southern CA. The reason might be that 1) the precipitation forecast skill in southern CA is comparatively lower (Fig. 4), which leads to
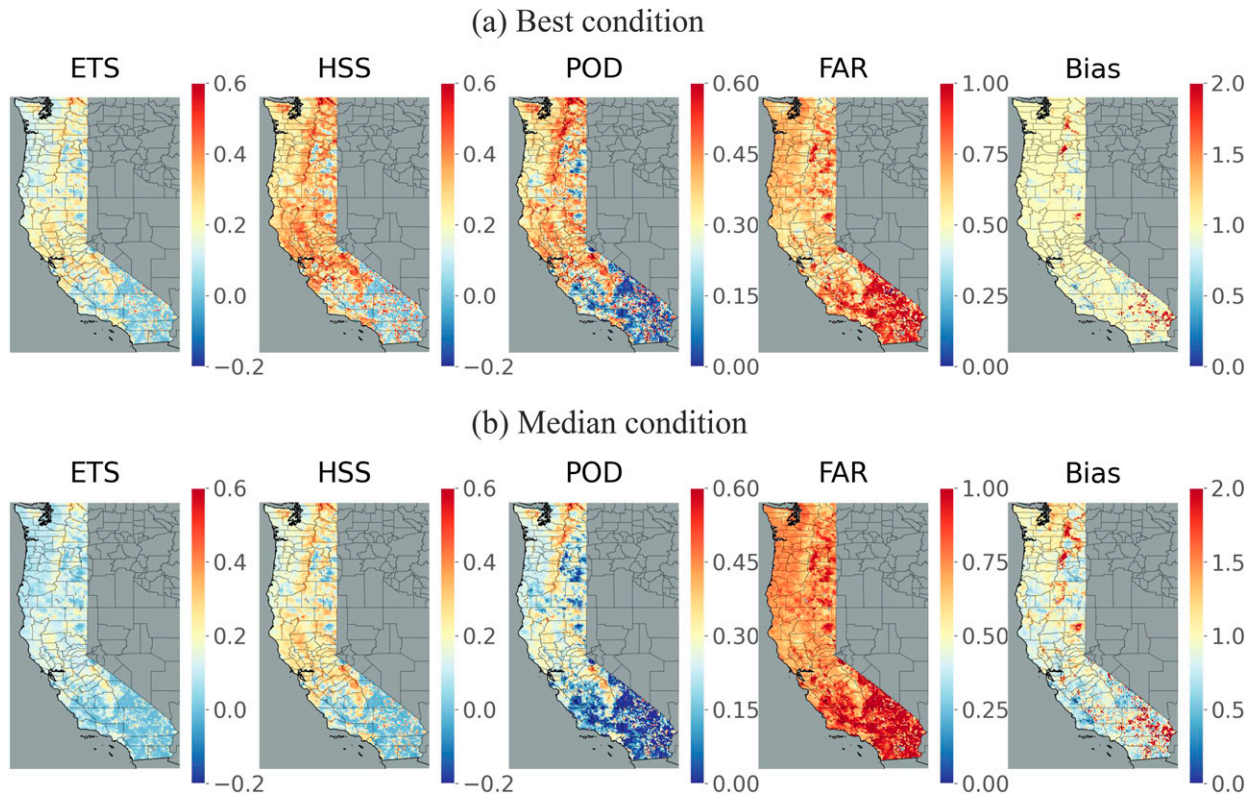
FIG. 12. As in Fig. 8, but for drought onset.

lower soil hydrological forecast skill, and 2) drought events in southern CA are very prolonged and the drought event pool is small particularly during the SubX time period. Fewer events give a false prediction more weight in the calculation of POD, and this may reduce apparent drought termination skill.

The previous analyses all examined major droughts. We also want to know if the patterns for major droughts are similar to those for more modest drought events. Thus, we also examined all drought events without restrictions on drought length. We calculated the ETS, HSS, POD, FAR, and bias score for drought termination, at gridcell scale at 2-week lead time (Fig. 11). Using all 30 ensembles, we evaluated the best condition and the median condition among all ensemble members. For ETS, HSS, and POD, positive values indicate skill. ETS for drought termination is ~0.3 in coastal WA and OR and southern and central CA in the best condition and ~0.2 in the median condition. HSS and POD are as high as high ~0.4–0.6 in the above locations in the best condition and ~0.2–0.3 in the median condition. These metrics all show the lowest skill in southern CA. FAR results show higher false alarms in the south (especially southern CA) and lower in the north. The bias score is almost 1 in most of our study area in the best condition, indicating almost no bias in this case. We see scattered high bias (overforecast, mostly in inland southern CA and inland WA) and low bias (underforecast, mostly in CA and OR) in the median condition. In summary, all metrics show the same general trend as for major droughts: higher

skill in the north and lower in the south. We repeated the same procedure for drought onset (Fig. 12) and found similar patterns from north to south; however, the overall forecast skill for onset is lower than for drought termination.

5. Conclusions

We examined the performance of SubX-driven forecasts of droughts in the coastal western United States with leads from 1 to 4 weeks. We first evaluated SubX reforecasts of precipitation and temperature. Our findings with respect to SubX precipitation and temperature skill are similar to previous studies (e.g., Cao et al. 2021). After statistical downscaling and bias correction of the forcings, we ran the Noah-MP LSM over the domain for the period 1999–2016. We then evaluated skill of SubX-based drought forecasts with a focus on drought termination and onset by using a variety of metrics. We evaluated both major droughts and more modest events.

Based on our analysis, we found usable drought termination and onset forecast skill at week 1 and 2 leads for major D0–D2 droughts; we found limited skill at week 3 for major D0–D1 droughts and essentially no skill at week 4. Drought prediction skill generally declines with increasing drought severity. We found that the skill is generally higher on termination than for onset for both major and all drought events. We also found that drought prediction skill generally increases from south to north for both major and all drought events.

S2S forecasting of meteorological and hydrologic variables is an active research topic that is attracting significant attention from both the research community (Vitart et al. 2017; Vitart and Robertson 2018; DeFlorio et al. 2019; Pan et al. 2019; Zhu et al. 2018) and the applications and stakeholders' communities (including the public health, agriculture, and emergency management and response sectors, along with water resource management; e.g., White et al. 2017, 2022; Robertson et al. 2020). We acknowledge, however, that S2S forecasting is still a maturing area. The drought forecast skill (in onset and termination) that we find is highly dependent on precipitation forecast skill. Precipitation forecast products with finer resolution and higher skill likely will improve drought forecast skill. Future studies could extend our work to more extreme events like floods and explore the usefulness of including higher resolution of forecast products. Exploiting of the large-scale climate drivers might also benefit by identifying additional sources of skill [e.g., El Niño–Southern Oscillation (ENSO), the Madden–Julian Oscillation (MJO), and North Atlantic Oscillation (NAO)] (DeFlorio et al. 2019; White et al. 2022). Employing artificial intelligence and machine learning techniques (e.g., Chapman et al. 2019; Bouaziz et al. 2021; Qian et al. 2021) may have the potential to improve S2S prediction skill.

Data availability statement. NCEP-CFSv2 output was obtained online (https://cfs.ncep.noaa.gov/cfsv2/downloads.html), and output for the other models were obtained from the IRI data library (http://iridl.ldeo.columbia.edu/SOURCES/.Models/.SubX/). The LSMs' simulation results are available online (https://doi.org/10.6084/m9.figshare.21508047.v1).

REFERENCES

Alessi, M. J., D. A. Herrera, C. P. Evans, A. T. DeGaetano, and T. R. Ault, 2022: Soil moisture conditions determine land–atmosphere coupling and drought risk in the northeastern United States. *J. Geophys. Res. Atmos.*, **127**, e2021JD034740, https://doi.org/10.1029/2021JD034740.

Arnal, L., A. W. Wood, E. Stephens, H. L. Cloke, and F. Pappenberger, 2017: An efficient approach for estimating streamflow forecast skill elasticity. *J. Hydrometeor.*, **18**, 1715–1729, https://doi.org/10.1175/JHM-D-16-0259.1.

Arsenault, K. R., and Coauthors, 2020: The NASA hydrological forecast system for food and water security applications. *Bull. Amer. Meteor. Soc.*, **101**, E1007–E1025, https://doi.org/10.1175/BAMS-D-18-0264.1.

Baker, S. A., A. W. Wood, and B. Rajagopalan, 2019: Developing subseasonal to seasonal climate forecast products for hydrology and water management. *J. Amer. Water Resour. Assoc.*, **55**, 1024–1037, https://doi.org/10.1111/1752-1688.12746.

Bennett, A. R., J. J. Hamman, and B. Nijssen, 2020: MetSim: A Python package for estimation and disaggregation of meteorological data. *J. Open Source Software*, **5**, 2042, https://doi.org/10.21105/joss.02042.

Bohn, T. J., B. Livneh, J. W. Oyler, S. W. Running, B. Nijssen, and D. P. Lettenmaier, 2013: Global evaluation of MTCLIM and related algorithms for forcing of ecological and hydrological models. *Agric. For. Meteor.*, **176**, 38–49, https://doi.org/10.1016/j.agrformet.2013.03.003.

Bouaziz, M., E. Medhioub, and E. Csaplovisc, 2021: A machine learning model for drought tracking and forecasting using remote precipitation data and a standardized precipitation index from arid regions. *J. Arid Environ.*, **189**, 104478, https://doi.org/10.1016/j.jaridenv.2021.104478.

Cao, Q., S. Shukla, M. J. DeFlorio, F. M. Ralph, and D. P. Lettenmaier, 2021: Evaluation of the subseasonal forecast skill of floods associated with atmospheric rivers in coastal western U.S. watersheds. *J. Hydrometeor.*, **22**, 1535–1552, https://doi.org/10.1175/JHM-D-20-0219.1.

Carrão, H., G. Naumann, E. Dutra, C. Lavaysse, and P. Barbosa, 2018: Seasonal drought forecasting for Latin America using the ECMWF S4 forecast system. *Climate*, **6**, 48, https://doi.org/10.3390/cli6020048.

Chapman, W., A. C. Subramanian, L. Delle Monache, S. P. Xie, and F. M. Ralph, 2019: Improving atmospheric river forecasts with machine learning. *Geophys. Res. Lett.*, **46**, 10 627–10 635, https://doi.org/10.1029/2019GL083662.

Chen, F., and J. Dudhia, 2001: Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Mon. Wea. Rev.*, **129**, 569–585, https://doi.org/10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2.

——, and Coauthors, 1996: Modeling of land-surface evaporation by four schemes and comparison with FIFE observations. *J. Geophys. Res.*, **101**, 7251–7268, https://doi.org/10.1029/95JD02165.

Christensen, J., and Coauthors, 2007: Regional climate projections. *Climate Change 2007: The Physical Science Basis*, S. Solomon et al., Eds., Cambridge University Press, 847–940.

Cook, B. I., A. P. Williams, J. S. Mankin, R. Seager, J. E. Smerdon, and D. Singh, 2018: Revisiting the leading drivers of Pacific coastal drought variability in the contiguous United States. *J. Climate*, **31**, 25–43, https://doi.org/10.1175/JCLI-D-17-0172.1.

DeAngelis, A. M., H. Wang, R. D. Koster, S. D. Schubert, Y. Chang, and J. Marshak, 2020: Prediction skill of the 2012 U.S. Great Plains flash drought in subseasonal experiment (SubX) models. *J. Climate*, **33**, 6229–6253, https://doi.org/10.1175/JCLI-D-19-0863.1.

DeFlorio, M. J., and Coauthors, 2019: Experimental subseasonal-to-seasonal (S2S) forecasting of atmospheric rivers over the western United States. *J. Geophys. Res. Atmos.*, **124**, 11 242–11 265, https://doi.org/10.1029/2019JD031200.

Delworth, T. L., F. Zeng, A. Rosati, G. A. Vecchi, and A. T. Wittenberg, 2015: A link between the hiatus in global warming and North American drought. *J. Climate*, **28**, 3834–3845, https://doi.org/10.1175/JCLI-D-14-00616.1.

Dettinger, M. D., 2013: Atmospheric rivers as drought busters on the U.S. West Coast. *J. Hydrometeor.*, **14**, 1721–1732, https://doi.org/10.1175/JHM-D-13-02.1.

Engström, J., K. Jafarzadegan, and H. Moradkhani, 2020: Drought vulnerability in the United States: An integrated assessment. *Water*, **12**, 2033, https://doi.org/10.3390/w12072033.

ECMWF, 2017: ERA5 reanalysis. RDA at NCAR, accessed 10 February 2021, https://doi.org/10.5065/D6X34W69.

Glantz, M. H., 2004: Early warning systems: Do's and don'ts. Usable Science 8 Workshop Rep., 76 pp., https://ilankelman.org/glantz/Glantz2003Shanghai.pdf.

Griffin, D., and K. J. Anchukaitis, 2014: How unusual is the 2012–2014 California drought? *Geophys. Res. Lett.*, **41**, 9017–9023, https://doi.org/10.1002/2014GL062433.

Hao, Z., V. P. Singh, and Y. Xia, 2018: Seasonal drought prediction: Advances, challenges, and future prospects. *Rev. Geophys.*, **56**, 108–141, https://doi.org/10.1002/2016RG000549.

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

Infanti, J. M., and B. P. Kirtman, 2016: Prediction and predictability of land and atmosphere initialized CCSM4 climate forecasts over North America. *J. Geophys. Res. Atmos.*, **121**, 12 690–12 701, https://doi.org/10.1002/2016JD024932.

Kirtman, B. P., and Coauthors, 2014: The North American Multimodel Ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bull. Amer. Meteor. Soc.*, **95**, 585–601, https://doi.org/10.1175/BAMS-D-12-00050.1.

Koster, R. D., M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar, 2000: A catchment-based approach to modeling land surface processes in a general circulation model: 1. Model structure. *J. Geophys. Res.*, **105**, 24 809–24 822, https://doi.org/10.1029/2000JD900327.

Krishnamurti, T. N., C. M. Kishtawal, T. E. Larow, D. R. Bachiochi, Z. Zhang, C. E. Williford, S. Gadgil, and S. Surendran, 1999: Improved weather and seasonal climate forecasts from multimodel superensemble. *Science*, **285**, 1548–1550, https://doi.org/10.1126/science.285.5433.1548.

——, ——, Z. Zhang, T. Larow, D. Bachiochi, E. Williford, S. Gadgil, and S. Surendran, 2000: Multimodel ensemble forecasts for weather and seasonal climate. *J. Climate*, **13**, 4196–4216, https://doi.org/10.1175/1520-0442(2000)013<4196:MEFFWA>2.0.CO;2.

Li, H., L. Luo, E. F. Wood, and J. Schaake, 2009: The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting. *J. Geophys. Res.*, **114**, D04114, https://doi.org/10.1029/2008JD010969.

Li, M., P. Wu, and Z. Ma, 2020: A comprehensive evaluation of soil moisture and soil temperature from third-generation atmospheric and land reanalysis data sets. *Int. J. Climatol.*, **40**, 5744–5766, https://doi.org/10.1002/joc.6549.

Li, Y., D. Tian, and H. Medina, 2021: Multimodel subseasonal precipitation forecasts over the contiguous United States: Skill assessment and statistical postprocessing. *J. Hydrometeor.*, **22**, 2581–2600, https://doi.org/10.1175/JHM-D-21-0029.1.

Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges, 1994: A simple hydrologically based model of land surface water and energy fluxes for GSMs. *J. Geophys. Res.*, **99**, 14 415–14 428, https://doi.org/10.1029/94JD00483.

Lin, H., N. Gagnon, S. Beauregard, R. Muncaster, M. Markovic, B. Denis, and M. Charron, 2016: GEPS-based monthly prediction at the Canadian Meteorological Centre. *Mon. Wea. Rev.*, **144**, 4867–4883, https://doi.org/10.1175/MWR-D-16-0138.1.

——, R. Mo, F. Vitart, and C. Stan, 2018: Eastern Canada flooding 2017 and its subseasonal predictions. *Atmos.–Ocean*, **57**, 195–207, https://doi.org/10.1080/07055900.2018.1547679.

Livneh, B., E. A. Rosenberg, C. Lin, B. Nijssen, V. Mishra, K. M. Andreadis, E. P. Maurer, and D. P. Lettenmaier, 2013: A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: Update and extensions. *J. Climate*, **26**, 9384–9392, https://doi.org/10.1175/JCLI-D-12-00508.1.

Mann, M. E., and P. H. Gleick, 2015: Climate change and California drought in the 21st century. *Proc. Natl. Acad. Sci. USA*, **112**, 3858–3859, https://doi.org/10.1073/pnas.1503667112.

Mariotti, A., P. M. Ruti, and M. Rixen, 2018: Progress in subseasonal to seasonal prediction through a joint weather and climate community effort. *npj Climate Atmos. Sci.*, **1**, 4, https://doi.org/10.1038/s41612-018-0014-z.

Merryfield, W. J., and Coauthors, 2020: Current and emerging developments in subseasonal to decadal prediction. *Bull. Amer. Meteor. Soc.*, **101**, E869–E896, https://doi.org/10.1175/BAMS-D-19-0037.1.

Mo, K. C., L.-C. Chen, S. Shukla, T. J. Bohn, and D. P. Lettenmaier, 2012: Uncertainties in North American land data assimilation systems over the contiguous United States. *J. Hydrometeor.*, **13**, 996–1009, https://doi.org/10.1175/JHM-D-11-0132.1.

Molod, A., L. Takacs, M. Suarez, J. Bacmeister, I.-S. Song, and A. Eichmann, 2012: The GEOS-5 atmospheric general circulation model: Mean climate and development from MERRA to Fortuna. NASA Tech. Memo. NASA/TM-2012-104606, Vol. 28, 115 pp., https://gmao.gsfc.nasa.gov/pubs/docs/tm28.pdf.

Monhart, S., C. Spirig, J. Bhend, K. Bogner, C. Schär, and M. A. Liniger, 2018: Skill of subseasonal forecasts in Europe: Effect of bias correction and downscaling using surface observations. *J. Geophys. Res. Atmos.*, **123**, 7999–8016, https://doi.org/10.1029/2017JD027923.

Mote, P. W., and Coauthors, 2016: Perspectives on the causes of exceptionally low 2015 snowpack in the western United States. *Geophys. Res. Lett.*, **43**, 10 980–10 988, https://doi.org/10.1002/2016GL069965.

Niu, G.-Y., Z.-L. Yang, R. E. Dickinson, L. E. Gulden, and H. Su, 2007: Development of a simple groundwater model for use in climate models and evaluation with gravity recovery and climate experiment data. *J. Geophys. Res.*, **112**, D07103, https://doi.org/10.1029/2006JD007522.

——, and Coauthors, 2011: The community Noah land surface model with multiparameterization options (Noah MP): 1. Model description and evaluation with local scale measurements. *J. Geophys. Res.*, **116**, D12109, https://doi.org/10.1029/2010JD015139.

Otkin, J. A., M. Svoboda, E. D. Hunt, T. W. Ford, M. C. Anderson, 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.

Pan, B., K. Hsu, A. AghaKouchak, S. Sorooshian, and W. Higgins, 2019: Precipitation prediction skill for the West Coast United States: From short to extended range. *J. Climate*, **32**, 161–182, https://doi.org/10.1175/JCLI-D-18-0355.1.

Pegion, K., and Coauthors, 2019: The Subseasonal Experiment (SubX): A multimodel subseasonal prediction experiment. *Bull. Amer. Meteor. Soc.*, **100**, 2043–2060, https://doi.org/10.1175/BAMS-D-18-0270.1.

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.

Pierce, D. W., D. R. Cayan, and B. L. Thrasher, 2014: Statistical downscaling using localized constructed analogs (LOCA). *J. Hydrometeor.*, **15**, 2558–2585, https://doi.org/10.1175/JHM-D-14-0082.1.

Pulwarty, R. S., and J. P. Verdin, 2013: Crafting integrated early warning information systems: The case of drought. *Measuring Vulnerability to Natural Hazards: Towards Disaster Resilient Societies*, 2nd ed. J. Birkmann, Ed., United Nations University Press, 124–147.

——, and M. V. K. Sivakumar, 2014: Information systems in a changing climate: Early warnings and drought risk management. *Wea. Climate Extremes*, **3**, 14–21, https://doi.org/10.1016/j.wace.2014.03.005.

Qian, Q., X. Jia, H. Lin, and R. Zhang, 2021: Seasonal forecast of nonmonsoonal winter precipitation over the Eurasian continent using machine-learning models. *J. Climate*, **34**, 7113–7129, https://doi.org/10.1175/JCLI-D-21-0113.1.

Reichle, R., and Q. Liu, 2014: Observation-corrected precipitation estimates in GEOS-5. NASA Tech. Memo. NASA/TM-2014-104606, Vol. 35, 18 pp., https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/ 20150000725.pdf.

Rienecker, M. M., and Coauthors, 2008: The GEOS-5 Data Assimilation System—Documentation of versions 5.0.1, 5.1.0, and 5.2.0. NASA Tech. Memo. NASA/TM-2008-104606, Vol. 27, 97 pp., https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20120011955.pdf.

Robertson, A. W., F. Vitart, and S. J. Camargo, 2020: Subseasonal to seasonal prediction of weather to climate with application to tropical cyclones. *J. Geophys. Res. Atmos.*, **125**, e2018JD029375, https://doi.org/10.1029/2018JD029375.

Saha, S., and Coauthors, 2014: The NCEP Climate Forecast System version 2. *J. Climate*, **27**, 2185–2208, https://doi.org/10.1175/JCLI-D-12-00823.1.

Schick, S., O. Rössler, and R. Weingartner, 2019: An evaluation of model output statistics for subseasonal streamflow forecasting in European catchments. *J. Hydrometeor.*, **20**, 1399–1416, https://doi.org/10.1175/JHM-D-18-0195.1.

Seager, R., and M. Hoerling, 2014: Atmosphere and ocean origins of North American droughts. *J. Climate*, **27**, 4581–4606, https://doi.org/10.1175/JCLI-D-13-00329.1.

——, ——, S. Schubert, H. Wang, B. Lyon, A. Kumar, J. Nakamura, and N. Henderson, 2015: Causes and predictability of the 2011–2014 California drought. NOAA Drought Task Force Rep., 40 pp., https://cpo.noaa.gov/Portals/0/Docs/MAPP/TaskForces/DTF/california_drought_report.pdf.

Seneviratne, S. I., and Coauthors, 2012: Changes in climate extremes and their impacts on the natural physical environment. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*, C. B. Field et al., Eds., Cambridge University Press, 109–230.

Shukla, S., and D. P. Lettenmaier, 2011: Seasonal hydrologic prediction in the United States: Understanding the role of initial hydrologic conditions and seasonal climate forecast skill. *Hydrol. Earth Syst. Sci.*, **15**, 3529–3538, https://doi.org/10.5194/hess-15-3529-2011.

——, N. Voisin, and D. P. Lettenmaier, 2012: Value of medium range weather forecasts in the improvement of seasonal hydrologic prediction skill. *Hydrol. Earth Syst. Sci.*, **16**, 2825–2838, https://doi.org/10.5194/hess-16-2825-2012.

Su, L., Q. Cao, M. Xiao, D. M. Mocko, M. Barlage, D. Li, C. D. Peters-Lidard, and D. P. Lettenmaier, 2021: Drought

variability over the conterminous United States for the past century. *J. Hydrometeor.*, **22**, 1153–1168, https://doi.org/10.1175/JHM-D-20-0158.1.

Sun, S., R. Bleck, S. G. Benjamin, B. W. Green, and G. A. Grell, 2018a: Subseasonal forecasting with an icosahedral, vertically quasi-Lagrangian coupled model. Part I: Model overview and evaluation of systematic errors. *Mon. Wea. Rev.*, **146**, 1601–1617, https://doi.org/10.1175/MWR-D-18-0006.1.

——, B. W. Green, R. Bleck, and S. G. Benjamin, 2018b: Subseasonal forecasting with an icosahedral, vertically quasi-lagrangian coupled model. Part II: Probabilistic and deterministic forecast skill. *Mon. Wea. Rev.*, **146**, 1619–1639, https://doi.org/10.1175/MWR-D-18-0007.1.

UNDRR, 2019: *Global Assessment Report on Disaster Risk Reduction*. UNDRR, 472 pp.

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 (S2S) prediction project database. *Bull. Amer. Meteor. Soc.*, **98**, 163–173, https://doi.org/10.1175/BAMS-D-16-0017.1.

Wanders, N., and Coauthors, 2017: Forecasting the hydroclimatic signature of the 2015/16 El Niño event on the western United States. *J. Hydrometeor.*, **18**, 177–186, https://doi.org/10.1175/JHM-D-16-0230.1.

Wang, S., A. Anichowski, M. K. Tippett, and A. H. Sobel, 2017: Seasonal noise versus subseasonal signal: Forecasts of California precipitation during the unusual winters of 2015–2016 and 2016–2017. *Geophys. Res. Lett.*, **44**, 9513–9520, https://doi.org/10.1002/2017GL075052.

Weigel, A. P., M. A. Liniger, and C. Appenzeller, 2007: The discrete Brier and ranked probability skill scores. *Mon. Wea. Rev.*, **135**, 118–124, https://doi.org/10.1175/MWR3280.1.

White, C. J., and Coauthors, 2017: Potential applications of subseasonal-to-seasonal (S2S) predictions. *Meteor. Appl.*, **24**, 315–325, https://doi.org/10.1002/met.1654.

——, and Coauthors, 2022: Advances in the application and utility of subseasonal-to-seasonal predictions. *Bull. Amer. Meteor. Soc.*, **103**, E1448–E1472, https://doi.org/10.1175/BAMS-D-20-0224.1.

Wilhite, D. A., M. V. K. Sivakumar, and R. Pulwarty, 2014: Managing drought risk in a changing climate: The role of national drought policy. *Wea. Climate Extremes*, **3**, 4–13, https://doi.org/10.1016/j.wace.2014.01.002.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. International Geophysics Series, Vol. 100, Academic Press, 648 pp.

Williams, A. P., R. Seager, J. T. Abatzoglou, B. I. Cook, J. E. Smerdon, and E. R. Cook, 2015: Contribution of anthropogenic warming to California drought during 2012–2014. *Geophys. Res. Lett.*, **42**, 6819–6828, https://doi.org/10.1002/2015GL064924.

Wood, A. W., E. P. Maurer, A. Kumar, and D. P. Lettenmaier, 2002: Long-range experimental hydrologic forecasting for the eastern United States. *J. Geophys. Res.*, **107**, 4429, https://doi.org/10.1029/2001JD000659.

——, L. R. Leung, V. Sridhar, and D. P. Lettenmaier, 2004: Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Climatic Change*, **62**, 189–216, https://doi.org/10.1023/B:CLIM.0000013685.99609.9e.

Xia, Y., and Coauthors, 2012: Continental-scale water and energy flux analysis and validation for the North American land data assimilation system project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *J. Geophys. Res.*, **117**, D03109, https://doi.org/10.1029/2011JD016048.

Zheng, H., Z.-L. Yang, P. Lin, J. Wei, W.-Y. Wu, L. Li, L. Zhao, and S. Wang, 2019: On the sensitivity of the precipitation partitioning into evapotranspiration and runoff in land surface parameterizations. *Water Resour. Res.*, **55**, 95–111, https://doi.org/10.1029/2017WR022236.

Zhou, X., Y. Zhu, D. Hou, and D. Kleist, 2016: A comparison of perturbations from an ensemble transform and an ensemble Kalman filter for the NCEP Global Ensemble Forecast System. *Wea. Forecasting*, **31**, 2057–2074, https://doi.org/10.1175/WAF-D-16-0109.1.

——, ——, ——, Y. Luo, J. Peng, and R. Wobus, 2017: Performance of the new NCEP Global Ensemble Forecast System in a parallel experiment. *Wea. Forecasting*, **32**, 1989–2004, https://doi.org/10.1175/WAF-D-17-0023.1.

Zhu, Y., and Coauthors, 2018: Toward the improvement of subseasonal prediction in the National Centers for Environmental Prediction Global Ensemble Forecast System. *J. Geophys. Res. Atmos.*, **123**, 6732–6745, https://doi.org/10.1029/2018JD028506.