ABSTRACT: In addition to remote SST forcing, realistic representation of land forcing (i.e., soil moisture) over the United States is critical for a prediction of U.S. severe drought events approximately one season in advance. Using “identical twin” experiments with different land initial conditions (ICs) in the 32-yr (1979–2010) CFSv2 reforecasts (NASA GLDAS-2 reanalysis versus NCEP CFSR), sensitivity and skill of U.S. drought predictions to land ICs are evaluated. Although there is no outstanding performer between the two sets of forecasts with different land ICs, each set shows greater skill in some regions, but their locations vary with forecast lead time and season. The 1999 case study demonstrates that although a pattern of below-normal SSTs in the Pacific in the fall and winter is realistically reproduced in both reforecasts, GLDAS-2 land initial states display a stronger east–west gradient of soil moisture, particularly drier in the eastern United States and more consistent with observations, leading to warmer surface temperature anomalies over the United States. Anomalies lasting for one season are accompanied by more persistent barotropic (warm core) anomalous high pressure over CONUS, which results in better prediction skill of this drought case up to 4 months in advance in the reforecasts with GLDAS-2 land ICs. Therefore, it is essential to minimize the uncertainty of land initial states among the current land analyses for improving U.S. drought prediction on seasonal time scales.

KEYWORDS: Atmosphere-land interaction; Drought; Sea surface temperature; Soil moisture; Seasonal forecasting; Climate models

1. Introduction

Severe droughts are among the costliest natural disasters in the United States (e.g., Cook et al. 2007; Smith and Katz 2013). They cause not only huge losses to agriculture and the economy with major impacts on the food and water supplies, but also often lead to wildfires and other environmental damage. Recent U.S. severe drought events include the 2011 Texas drought (Nielsen-Gammon 2012), the 2012 Great Plains damage. Recent U.S. severe drought events include the 2011 Texas drought (Hoerling et al. 2014), and the 2012–15 California drought (Seager et al. 2015; Luo et al. 2017). Accordingly, improving the capability to predict the evolution of drought—its onset, persistence, and demise (Wu and Dirmeyer 2020)—and its severity and frequency more accurately and reliably is very important for drought early warning and mitigation (e.g., Hao et al. 2018).

There have been considerable efforts toward understanding the mechanisms and predictability of drought. The sources of meteorological drought (or precipitation deficit over a prolonged period of time) in the United States on seasonal time scales mainly result from sea surface temperature (SST) anomalies in the Pacific associated with El Niño–Southern Oscillation (ENSO) and/or the Pacific decadal oscillation (PDO), with lesser contribution from SST anomalies in the Atlantic and Indian Oceans (e.g., Hoerling and Kumar 2003; McCabe et al. 2004; Seager et al. 2005; Cook et al. 2007; Hoerling et al. 2009; Seager and Hoerling 2014; Schubert et al. 2016; Huang et al. 2019). Regional land surface states (e.g., soil moisture, snow cover, vegetation properties, etc.) also contribute to drought severity and development (e.g., Higgins et al. 1998; Schubert et al. 2007; Koster et al. 2017). In particular, positive feedbacks between land and atmosphere can exacerbate or prolong dry anomalies, playing a role in maintaining droughts (e.g., Durre et al. 2000; Fischer et al. 2007; Koster et al. 2009; Kam et al. 2014; Dirmeyer et al. 2015; Fernando et al. 2016). The role of the land surface in modulating drought has long been established in modeling studies (e.g., Dirmeyer 1994; Fennessy and Shukla 1999; Seneviratne et al. 2006; Yuan and Wood 2013; Zaitchik et al. 2013; Roundy et al. 2014; Roundy and Wood 2015).

The seasonal prediction skill of drought depends on the accurate prediction of precipitation and temperature since indicators for meteorological drought prediction are often based on seasonal prediction of precipitation and temperature (e.g., Yoon et al. 2012; Yuan and Wood 2013; Dutra et al. 2014; Mo and Lyon 2015). Current seasonal coupled forecast systems can predict major oceanic and atmospheric anomalies at useful lead times (e.g., Jin et al. 2008; Kirtman et al. 2014; Huang et al. 2017a;

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JHM-D-20-0025.s1.

Corresponding author: Chul-Su Shin, cshin3@gmu.edu

DOI: 10.1175/JHM-D-20-0025.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).
Shin et al. 2019). However, the current level of skill in forecasting drought onset, development, and demise is limited (e.g., Quan et al. 2012; Mo and Lyon 2015), possibly because other sources of predictability, such as land–atmosphere feedbacks arising from the memory in land surface states, may still be underrealized (Roundy and Wood 2015; Dirmeyer and Halder 2016).

Recently, we conducted two sets of reforecasts initialized with two different land analyses for the period of 1979–2010 (section 2; Shin et al. 2020). Since atmosphere, ocean, and sea ice initial states are identical for both sets of reforecasts, this identical-twin set of reforecasts isolates the effect of the uncertainty of the land initial states on prediction of the atmosphere and land at subseasonal and seasonal time scales. For example, Shin et al. (2020) demonstrated that estimates of soil moisture at surface layer (0–10 cm) as well as deep layer (0–100 cm) between the two land initial states are significantly different with an apparent north–south contrast for almost all seasons, leading to disparity in predicted surface air temperature between the two sets of reforecasts, particularly in regions where the magnitude of initial soil moisture difference lies in the top quintile over the land grid cells. Their results also suggest that a reduction of the uncertainty in land surface states among the current land analyses can be beneficial to improving prediction skill of surface air temperature on subseasonal time scale.

In this study, we further quantify how prediction skill of drought in the United States responds to different land initial states on seasonal time scales and evaluate how well the two sets of reforecasts reproduce selected severe drought events that occurred during 1979–2010. We also examine the role of realistic representation of land initial states in improving drought prediction in the United States using a case study of the drought in the winter of 1999. Section 2 describes the coupled forecast system, the identical-twin experiments, and verification data. Predictive skill of precipitation indices and prediction of U.S. severe drought events for 32 years are evaluated in sections 3 and 4, respectively, followed by summary and discussions in section 5.

2. Identical-twin CFSv2 reforecast experiments (1979–2010)

CFSv2 is the current U.S. operational seasonal prediction system at the National Centers of Environmental Prediction (NCEP) and is a fully coupled climate forecast system composed of interacting atmospheric, oceanic, sea ice, and land components (Saha et al. 2014). The atmospheric model of the CFSv2 is a lower resolution version of the Global Forecast System (GFS), having a spectral horizontal resolution of T126 (equivalent to about 1° grid spacing) and 64 vertical levels in a hybrid sigma-pressure coordinate. The oceanic component is the Geophysical Fluid Dynamics Laboratory (GFDL) Modular Ocean Model version 4 (MOM4; Griffies et al. 2004). Horizontally, it has a 0.5° × 0.5° grid spacing in the poleward side of 30° with gradually increasing meridional resolution to 0.25° inside of 10° and vertically 40 levels (27 levels in the upper 400 m) with a maximum depth of about 4.5 km. The sea ice component is a three-layer global interactive dynamical sea ice model with predicted fractional ice cover and thickness (Winton 2000). The land surface component is the Noah land surface model (LSM) version 2.7.1 (Ek et al. 2003) of four soil layers with interfaces at depths of 0.1, 0.4, 1.0, and 2.0 m.

Using the revised version of CFSv2 described in Huang et al. (2015), we had recently produced 60-yr (1958–2017) ensemble reforecasts with 12-month duration initialized at the beginning of January, April, July, and October, respectively (Huang et al. 2017a, 2019). The ocean initial states for the whole period are from the instantaneous restart files of the ECMWF Ocean Reanalysis System 4 (ORA-S4) with a set of five-member ensemble assimilation runs (Balmaseda et al. 2013). After 1979, the atmosphere, land, and sea ice initial states are taken from the restart files of the NCEP CFSR (Saha et al. 2010). An ensemble reforecast of 20 members was generated by matching each of the five ocean initial states at 0000 UTC on the first day of the month with the atmospheric and land initial states at 0000 UTC of the first four days of the month while the sea ice initial state is fixed at 0000 UTC on the first day of the month for all ensemble members. More details of the initialization procedure can be found in Huang et al. (2017a). Huang et al. (2017a, 2019) also evaluated the prediction skill and predictability of the ENSO and U.S. precipitation for the period of 1958–2014 and 1958–2017, respectively.

More recently, we conducted another set of CFSv2 reforecasts by extending the reforecasts with NASA Global Land Data Assimilation System Version 2.0 (GLDAS-2; Rodell et al. 2004; Rui and Beaudoin, 2015) land initial states for the period of 1979–2010, and their 20-member ensemble is referred to hereafter as the GLDAS reforecasts. As a pair of “identical twin” experiments, we refer to the original ensemble of CFSv2 reforecasts initialized with NCEP CFSR land states for the common period of 1979–2010 as the CFSR reforecasts. The GLDAS and CFSR reforecasts form the identical-twin sets of 32-yr (1979–2010) CFSv2 reforecasts, which have exactly the same initial conditions other than land states. Therefore, they enable us to assess the effect of the uncertainty of land initial states on the seasonal prediction of global surface air temperature (Shin et al. 2020), and particularly, the U.S. drought predictions in this study.

The observed rainfall data used for verification are from the Climate Prediction Center (CPC) daily unified gauge-based analysis of precipitation at 0.25° latitude × 0.25° longitude resolution (Chen et al. 2008). The observed SST data are the global monthly Extended Reconstructed SST, version 5 (ERSSTv5; Huang et al. 2017b) on a 2° latitude × 2° longitude grid. The mean sea level pressure, 2-m air temperature, and 200-hPa geopotential height data are from the ERA-Interim reanalysis (Dee et al. 2011) while the volumetric soil moisture (0–40 cm) used for verification comes from the Soil MERGE (SMERGE) root zone soil moisture data (Crow and Tobin 2018; Tobin et al. 2017, 2019).

The statistical significance of the difference in drought forecast skill between CFSR and GLDAS reforecasts is calculated by a bootstrap resampling as follows:

1) We randomly generated two sets of 20-member ensemble means from the total 40 ensemble members (i.e., 20 members of CFSR reforecasts and 20 members of GLDAS reforecasts) without considering which ICs are used.
2) We computed the difference between the two random sets and repeated the process 1000 times.

3) The skill difference is considered statistically significant only if it is beyond the 95% range of the skill in the random sample distribution, i.e., either below the 2.5th percentile or above the 97.5th percentile (e.g., Kumar et al. 2014).

3. Evaluation of SPI3 forecast skill

From a meteorological perspective, drought in this study is defined by the standardized precipitation index (SPI; McKee et al. 1993). We have generated the 3-month SPI (SPI3) forecasts based on precipitation data from the reforecasts and observations for a 32-yr period (1979–2010). Over the contiguous United States (CONUS), observed SPI3 for each month was obtained using daily CPC unified gauge-based analysis of precipitation while forecasted SPI3 was generated from monthly forecasted precipitation, but by combination of observed and forecasted precipitation for the 1- and 2-month leads following previous studies (e.g., Mo and Lyon 2015). For the reforecasts with October initial conditions (ICs), for example, the forecasted SPI3 in October (i.e., 1-month lead) was computed with observed precipitation data in August and September combined with forecasted precipitation in October. The forecasted SPI3 at
2-month lead (i.e., November) was computed by blending observed precipitation in September with forecasted precipitation in October and November. Since all forecasted precipitation data were available in October–December, the forecasted SPI3 at 3-month lead could be defined by using forecasted precipitation data alone. Likewise, forecasted SPI3 from 4- to 12-month lead were constructed in the same fashion.

We validate SPI3 forecast skill in the two sets of reforecasts for 32 years (1979–2010). Figure 1 displays anomaly correlation maps of SPI3 for the reforecasts with October ICs at 1-, 3-, and 5-month lead (i.e., in October, December, and February) for 32 years (1979–2010) and their differences. In both the CFSR and GLDAS reforecasts at 1-month lead, almost all areas except for the California coastal region are statistically significant at the 95% confidence level according to a Student’s t test (dashed curves in the left and middle columns of Figs. 1 and 2) and area-averaged correlation coefficient values are around 0.77. Relatively high correlation skill is largely due to the contribution of observed precipitation from the previous two months (i.e., August and September) to the forecasted SPI3 at 1-month lead as explained above. As forecast lead increases, the region with statistically significant skill becomes smaller with an area-averaged correlation skill of 0.27 (0.3) in the CFSR (GLDAS) reforecasts at 3-month lead and around 0.22 in both reforecasts at 5-month lead. However, the spatial distributions still show substantial areas of skillful forecasts in these lead months. At 3-month lead (December), large areas to the south of 35°N have correlation skill above 0.4 for both reforecasts with maximum centers around 0.7 near Louisiana and Mississippi (Fig. 1b). Some of these correlations are persistent and even enhanced at 5-month lead (February) in the southwest (Texas and New Mexico) and southeast around the Florida panhandle (Fig. 1c). There are also areas of significant correlation skill in the north for both reforecasts at 3- and

![Fig. 2. As in Fig. 1, but for April ICs.](image)
5-month leads. At 3-month lead, significant correlations extend from the border areas between Colorado and Kansas northward to Nebraska, then to the eastern part of South and North Dakota (Fig. 1b). Significant correlations also appear in Idaho and Montana at 3-month lead and its northern part persists to 5-month lead (Fig. 1c) although the skill is poor there at 3-month lead (Fig. 1b). The general skill distribution is consistent with a predictable pattern of winter precipitation with north–south dipole structure over CONUS largely induced by the remote ENSO forcing (e.g., Huang et al. 2019; among many others).

Based on the two sets of reforecasts with April ICs, the prediction skill of SPI3 from 1-month lead (April) through 5-month lead (August) is shown in Fig. 2. An area-averaged correlation skill at 1-month lead is around 0.75 with very high skill greater than 0.9 (gray color) in the western United States and in parts of the southeastern United States in the both sets of reforecasts (Fig. 2a), which is comparable to that of October ICs (Fig. 1a). However, it is evident that the SPI3 prediction skill in summer degrades much faster with increasing lead time, compared to SPI3 prediction in winter when the remote ocean forcing (e.g., ENSO SST anomaly) is dominant. The SPI3 prediction in summer is challenging even at 3-month lead (June) with an area-averaged skill of 0.19, and has very limited skill over CONUS at 5-month lead (August) with an area-averaged skill of less than 0.05 (Figs. 2b,c). On the other hand, there are still substantial areas of significant correlation skill consistently appearing in both reforecasts. At 3-month lead (June), relatively high correlations appear in California and Oregon (Fig. 2b). At 5-month lead (August), significant correlations extend from Oregon to Idaho and Wyoming (Fig. 2c), which seems attributable to the enhanced prediction skill of precipitation during 1979–99 (Huang et al. 2019, their Fig. 5c).

Our results also demonstrate some differences in prediction skill between the two reforecasts (right column of Figs. 1 and 2 as well as Figs. S1 and S2 in the online supplemental material). Note that contours in Figs. 1 and 2 (black dots in Figs. S1 and S2) indicate the statistical significance of the difference based on a bootstrap method (1000 resampling) as explained in section 2. In fact, one set of reforecasts shows better skill than the other in some regions, although their locations vary with lead time and season. For example, the GLDAS reforecasts with October ICs have higher skill in California and Nevada at 1-month lead, in the northern Great Plains, the western half of Washington State, from Missouri to Michigan, and the states of Arkansas and Tennessee at 3-month lead, and in eastern Washington and from Mississippi to Tennessee at 5-month lead, compared to the CFSR reforecasts. The latter shows better skill in northern Florida and southern Georgia at 1-month lead, in New Mexico and northwestern Texas at 3-month lead, and in Nebraska and the northeastern United States at 5-month lead (right column of Fig. 1 and Fig. S1). On the other hand, while the GLDAS reforecasts with April ICs exhibit relatively higher skill in the western coastal regions at 1-month lead, in the northeastern United States at 3-month lead than the CFSR reforecasts, the latter shows better skill in Texas, Utah, and North Dakota at 1-month lead, in the southeastern United States at 3-month lead, and in the northern Great Plains at 5-month lead (right column of Fig. 2 and Fig. S2).

We compare more quantitatively SPI3 prediction skill between the CFSR and GLDAS reforecasts initialized with October and April ICs by displaying the percentage of grid cells over CONUS with statistically significant correlation skill (i.e., passing the 95% confidence level using a Student’s t test) for 1979–2010 for (a) October ICs and (b) April ICs. The abscissa is the lead month from 1 to 12 months. Black bars are the total percentage of CONUS grid cells with significant skill in either the CFSR or GLDAS reforecasts, representing an upper boundary of skillful SPI3 prediction achievable by CFSv2 reforecasts with the two different land initial states. Red bars indicate the percentage of grid cells where the CFSR reforecasts show better skill than the GLDAS reforecasts while blue bars denote the percentage of grid points where the latter outperforms the former.
CFSv2 reforecasts with the two different land initial states. They are decomposed into two color bars; Red bars indicate the percentage of grid cells where the CFSR reforecasts show better skill than the GLDAS reforecasts, while blue bars denote the percentage of grid cells where the GLDAS reforecasts outperform the CFSR reforecasts. The black bars in Fig. 3a are at 40% or greater up to 7-month lead, suggesting that SPI3 prediction skill for October ICs is reasonably good up to two seasons, but the skill for April ICs drops quickly after one season with a rebound in boreal winter (black bars in Fig. 3b).
Reforecasts with July ICs perform similarly to April ICs (not shown). One interesting common feature between October and April IC reforecasts is that the GLDAS reforecasts tend to show higher percentages than the CFSR reforecasts at short leads (i.e., the blue bars are greater than the red bars), while they are the opposite at long leads (Fig. 3). This is also found in the reforecasts with January and July ICs (not shown). We further computed the percentage of grid cells over CONUS where the difference in correlation skill between CFSR and GLDAS reforecasts is statistically significant based on a bootstrap method with 1000 resamples and decomposed it into the two colors in the same manner (Fig. S3). Although the percentages of each color bar are overall much reduced with a maximum of about 15%, compared to those in Fig. 3, the common features described above is confirmed.

### TABLE 1. A list of 38 selected U.S. severe drought events during 1979–2010. Bold font indicates predicted cases while italic font with an asterisk denotes missed cases.

<table>
<thead>
<tr>
<th>Target month (3-month lead)</th>
<th>December (October ICs)</th>
<th>March (January ICs)</th>
<th>June (April ICs)</th>
<th>September (July ICs)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Condition</strong></td>
<td><em>Percentage of grid cells with SPI3 &lt; −1.2 that are greater than each month mean + 0.75σ</em>, but using 0.5σ for September (σ is standard deviation for each target month)*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Years</strong></td>
<td><strong>1980</strong></td>
<td><strong>1981</strong></td>
<td><strong>1985</strong></td>
<td><strong>1988</strong></td>
</tr>
<tr>
<td></td>
<td>13.9%</td>
<td>14.7%</td>
<td>13.8%</td>
<td>12.5%</td>
</tr>
<tr>
<td>1995</td>
<td>1996</td>
<td>1992*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td>2001</td>
<td>1998</td>
<td>1999*</td>
<td></td>
</tr>
<tr>
<td>2001*</td>
<td>2003</td>
<td>2005</td>
<td>2002*</td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td>2006</td>
<td>2006</td>
<td>2003</td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td>2009</td>
<td>2005*</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total 38 events</strong></td>
<td>9 events</td>
<td>10 events</td>
<td>9 events</td>
<td>10 events</td>
</tr>
</tbody>
</table>

**FIG. 5.** Spatial distribution of observed and predicted SPI3 for the two successfully predicted severe drought events over CONUS in (a) June 1992 and (b) December 2010. The number in the top-right corner of panels indicates pattern correlation coefficient of observed and predicted SPI3 in the CFSR and GLDAS reforecasts, respectively.
We also examine the ability of the SPI3 prediction to capture severe drought conditions. Severe drought in this study is defined at each grid point as occurring when SPI3 is below $-1.2$. As an example, Fig. 4 displays the equitable threat score (ETS) of severe drought at 2-month lead for each starting month in the two sets of reforecasts and their difference. Since ETS represents the skill of forecasts relative to chance, the accuracy of forecast becomes higher as ETS is closer to one while for a random forecast if ETS is zero. The spatial pattern of the ETS looks similar between the CFSR and GLDAS reforecasts (left and center columns of Fig. 4) but regions with relatively high ETS varies with season. For example, high ETS commonly appears over the central and southeastern United States in February (January ICs), over the southeastern United States and the states of Virginia and Maryland in May (April ICs), and over the northern United States and Texas in August (July ICs) and November (October ICs).

The right column of Fig. 4 and Fig. S4 shows that one set of reforecasts is not always better than the other. Moreover, ETS difference between the two sets of reforecasts in some regions is quite large. The ETS of the GLDAS reforecasts is much greater than that of the CFSR reforecasts, especially over Northern California and southern Texas in February; in Nevada, Georgia, and South Carolina and parts of Kansas, Iowa, and Missouri in May; and the states of Washington and Oregon in November [dark pink areas with contour (black dots) in the right column of Fig. 4 (Fig. S4)]. In contrast, the ETS of the CFSR reforecasts is higher than that of the GLDAS reforecasts over the midwestern United States in February, parts of Mississippi and Tennessee in May, southern Texas in August, and New Mexico in November [blue areas with contour (black dots) in the right column of Fig. 4 (Fig. S4)]. Especially, qualitative difference seems most visible in November between the two reforecasts when the ETS of the GLDAS reforecasts are apparently larger than that of the CFSR reforecasts over most CONUS. Overall, these comparisons demonstrate that the prediction skill of U.S. drought in CFSv2 is sensitive to the land initial states with regard to indices of categorical forecasts as well as anomaly correlations.


To investigate the impact of land initial state uncertainty on the prediction skill of U.S. drought, we selected a total of 38 severe drought events in March, June, September, and December for the period of 1979–2010 (32 years). They were determined by the condition that percentage of grid points over CONUS with an observed SPI3 $< -1.2$ (i.e., severe...
drought) is greater than the value of the 32-yr mean $+ 0.75 \times$ the standard deviation for the given month. Considering that CONUS experiences a wet season from July to September, we used $0.5 \times$ the standard deviation for September. The corresponding criteria of the percentage and selected year of the events for each target month are presented in Table 1. The target months correspond to 3-month lead from each starting month, and we focus on drought prediction at 3-month lead because it is the shortest lead at which predicted SPI3 is constructed entirely with model forecast precipitation.

If the pattern correlation coefficient (PCC) over CONUS between observed and predicted SPI3 for a given event at the
target month is above 0.4 (below 0.2) in both CFSR and GLDAS reforecasts, that event was considered as a predicted (missed) case. Among the selected 38 events, we obtained nine predicted cases and nine missed cases in Table 1. There are nine other events with moderate prediction skill \(0.3 < \text{PCC} < 0.4\) and eleven events with limited skill \(0.2 < \text{PCC} < 0.3\), although they were not specified in Table 1. More predicted cases (bold font) appear in December and March while more missed cases (italic font with an asterisk) are listed in June and September, implying that U.S. severe droughts may be more predictable during winter and spring.

As examples, Figs. 5 and 6 display the spatial distribution of observed and predicted SPI3 for two predicted and missed cases, respectively. For the case of June 1992, both the CFSR and GLDAS reforecasts realistically reproduce severe drought conditions centered in the central United States and wet conditions across the southern United States including Arizona, New Mexico and Texas (Fig. 5a). Likewise, the two sets of reforecasts are capable of capturing the observed contrast of wet conditions in the western and northeastern United States and dry in the Great Plains and southeastern United States in the case of December 2010, although they generally overestimated drought severity (Fig. 5b). The numbers in the top-right corner of each panel denote PCC between observed and predicted SPI3, which are all above 0.5. For the case of June 1988, on the other hand, both sets of reforecasts totally fail to predict the severe drought condition that covers almost half of CONUS including the midwestern and northeastern United States and over the southern United States from eastern Oklahoma/Texas to Tennessee/Alabama (Fig. 6a). For December 2001, observed severe drought conditions in the eastern United States are mostly missed in both the CFSR and GLDAS reforecasts while a false alarm of severe drought over the southwestern United States is given by the both reforecasts, as well as an erroneous belt of large positive SPI3 extends from Kansas to Michigan (Fig. 6b). The consistency of the predicted spatial patterns between the two reforecasts.
in this case is noticeable, which may be related to response to some external forcing.

Among the nine predicted cases, we are most interested in the case of December 1999 (IC: October 1999) wherein the CFSR and GLDAS reforecasts give quite different predictions (Fig. 7). It is noted that regions with brown and dark brown color (SPI3 < −1.2) experience severe drought. The observed droughts over CONUS seem to have two branches; one is extending from north to south cross the central-eastern United States while the other is located in the western United States (left column of Fig. 7a). The former had already gained strength by October 1999, peaked in November, then gradually weakened and retreated northward in the next two months. The latter was initiated near Nevada, Idaho, and Utah in October and then enhanced substantially in November and December while extending southward. It sustained in the southwest United States in January 2000.

In October and November 1999 (1- and 2-month leads), both sets of reforecasts realistically capture the drought conditions in observations (Figs. 7a,b). In particular, the GLDAS reforecasts well reproduce the severe drought conditions over the western United States, central United States, and Texas in November at 2-month lead (Fig. 7b). The observed drought conditions in the southwestern United States and from eastern Texas to Kentucky for December 1999 and January 2000 are also reasonably reproduced in the GLDAS reforecasts. However, the CFSR reforecasts show a much quicker demise of the central-eastern branch of the drought and feature unrealistically wet conditions in the northern United States, especially at 4-month lead (Figs. 7c,d). Furthermore, the other center of the drought was not maintained in the southwest but shifted eastward in December and January while drought conditions were generated around the Florida panhandle. This forms a wet north-dry south structure over a large part of the United States, which is quite different from the drought distribution in the observations and the GLDAS reforecasts. As a result, the GLDAS reforecasts maintain high PCC above 0.7 up to 4-month lead whereas PCC for the CFSR reforecasts quickly decreases from 3-month lead and drops to 0.14 at 4-month lead.3

As described in section 2, only land initial states are different in the identical-twin experiments. In the remainder of the paper, we concentrate on an examination of how two different land initial states result in the discrepancies in prediction of this particular drought event.

One of the important factors for U.S. drought prediction is a remote ocean forcing, i.e., SST anomalies in the Pacific and Atlantic Oceans (e.g., Seager and Hoerling 2014; Schubert et al. 2016). Observed SST anomalies in the tropical Pacific display below normal temperatures with less than −1.5°C at its minimum in October 1999 that intensified with the westward shifted minimum in the winter of 1999–2000 (left panels of Fig. 8). The negative phase of PDO is also apparent in the North Pacific and its anomalous warm center located between 30° and 45°N moves eastward in time. This combination of PDO and ENSO favors drought conditions in the central and western United States (Hu and Huang 2009). These features (i.e., the developing La Niña in the tropical Pacific and the negative PDO phase in the North Pacific) are realistically reproduced in both sets of reforecasts, although the minimum center of predicted SST anomalies is shifted eastward along the Equator and the centers of predicted warm and cold SST anomalies in the North Pacific are less organized, especially in the CFSR reforecasts, compared to those in observations (center and right panels of Fig. 8). In comparison, the drought patterns in January 2000 predicted by the CFSR reforecasts (middle panel of Fig. 7d) is more typical of the ENSO-induced response (e.g., Huang et al. 2019) than the GLDAS reforecasts (right panel of Fig. 7d).

We argue that different land–atmosphere interactions triggered by the differences in land initial state account for a part

---

3 It is pointed out that blending of observed precipitation accounts for the higher-resolution structures in the 1- and 2-month lead SPI3 forecasts, compared to 3-month and onward lead SPI3 forecasts.
of the different drought predictions between the CFSR and GLDAS reforecasts. Land surface memory relevant to sub-seasonal to seasonal prediction is typically defined based on anomalies of soil moisture. Figure 9 displays anomalies of volumetric soil moisture within the top 40-cm layer in the two land initial states as well as in observations (SMERGE). GLDAS-2 land initial states show a stronger east–west gradient of soil moisture anomalies than CFSR land initial states, with wetter than normal conditions over a broad diagonal band from Arizona toward North Dakota and Minnesota, while drier than normal over the southeastern and central United States. It seems that the former overestimates dry conditions over the eastern United States.

Anomalously dry conditions in the observations are gradually enhanced and peak in November 1999, especially over the central United States (left panels of Fig. 10), but the CFSR reforecasts have anomalies of 0–40-cm soil moisture that are reduced in magnitude over the eastern United States as the lead month increases (center panels of Fig. 10). In contrast, drier initial soil moisture anomalies over the eastern United States in the GLDAS reforecasts persist for about one season, leading to land forcing that is more compatible with observations compared to that of the CFSR reforecasts (right panels of Fig. 10). On the other hand, anomalously dry conditions over the southwestern United States in the observations are continuously enhanced until December 1999 (left panels of Fig. 10). The CFSR reforecasts with anomalously drier initial soil moisture there (Fig. 9) seem to reproduce them better than the GLDAS reforecasts, whereas the magnitude of soil moisture anomalies in the Intermountain West in the latter looks closer to the observed one than that of the former (Fig. 10).

It is well known that surface air temperature interacts closely with soil moisture (e.g., Koster et al. 2009; Shin et al. 2020) and snow cover as well (e.g., Xu and Dirmeyer 2011; Shin et al. 2020). Figure 11 shows the observed and predicted 2-m air temperature anomalies. In October 1999, both the CFSR and GLDAS reforecasts reasonably well capture warm (cold) anomalies of observed surface air temperature over the western United States (eastern Canada), although they predict colder surface air temperature than observations over western

![Figure 10](image-url)
Canada (Fig. 11a). It demonstrates that the GLDAS reforecasts realistically reproduce the observed spatial pattern of warm temperature anomalies over almost the entire North America at 2- and 3-month leads (i.e., November and December 1999), but the CFSR reforecasts predict a seesaw pattern with warm anomalies in the southern United States and cold anomalies in the northern United States and Canada (Figs. 11b,c). It is noted that the spatial pattern of 2-m air temperature in the CFSR reforecasts resembles more typical of the La Niña–induced response.

Figure 12 shows evolution of snow depth anomalies in observation and the two sets of reforecasts from October to December 1999. Anomalies of snow depth between CFSR and GLDAS reforecasts at 1-month lead are similar to each other with nearly no change from their climatology, especially over CONUS and southern Canada, which are consistent with the observations (Fig. 12a). At 2-month lead (November 1999), GLDAS reforecasts show negative anomalies of snow depth over the northern United States and southern Canada as shown in observation, whereas positive anomalies of snow depth appear in the CFSR reforecasts (Fig. 12b). The decreased snow depth in observations for December 1999 is well captured in the GLDAS reforecasts but is opposite in the CFSR reforecasts (Fig. 12c). Shin et al. (2020) found that in the CFSv2 reforecasts with October ICs, surface air temperature, and underlying soil moisture in middle latitudes of North America are negatively coupled until mid-November (about 6-week lead) with statistical significance based on a Student’s t test (their Fig. 8). As a consequence, we argue that drier soil moisture over the central-eastern United States in the GLDAS reforecasts may account for warmer surface air temperature there by mid-November, and positive feedback between warmer air temperature and decreased snow cover plays a role in enhanced surface warming in the central United States from November 1999. The land-atmosphere feedback of soil moisture–surface temperature–snow in the GLDAS reforecasts makes the spatial patterns of anomalous soil moisture, 2-m temperature, and snow depth more consistent with

4 It is noted that 32-yr climatologies of snow depth between CFSR and GLDAS reforecasts with October ICs have little difference over North America for the first 3-month leads.

FIG. 11. As in Fig. 8, but for monthly 2-m temperature anomalies (°C).
those in observations, compared to the CFSR reforecasts (Figs. 10–12).

Warm anomalies lasting for more than one month over North America are accompanied by persistent barotropic (warm core) anomalous high pressure during this drought event. For observations, centers of anomalous high pressure at both the surface (shading in Fig. 13) and upper atmosphere (i.e., at 200 hPa, contours in Fig. 13) are located over the western United States in October 1999, associated with the warm anomalies of surface air temperature there (left panels of Figs. 11a and 13a). The anomalous high pressure centers move slightly northeastward in November 1999, but remain over the northern United States collocated with the maximum warm temperature anomalies. Also, both surface low pressure and an upper-level trough are located across the northeastern Pacific, western Canada and Alaska (left panel of Figs. 11b and 13b). For December 1999, barotropic anomalous high pressure is centered in the northeastern Pacific and a northeast–southwest contrast of anomalous surface pressure is apparent in observations (left panel of Fig. 13c). The evolution of anomalous pressure at the surface and upper atmosphere in observations is much better reproduced in the GLDAS reforecasts (right panels of Fig. 13) than in the CFSR reforecasts (center panels of Fig. 13), especially at 2- and 3-month leads (i.e., November and December 1999). Although ERA-Interim estimates of the mean sea level pressure, 2-m air temperature, snow depth, and 200-hPa geopotential height are being discussed as observations in Figs. 11–13, it should be noted that they are observation-based reanalysis (not observation per se) and may contain the models’ influence in reanalysis products. However, these fields are usually adequate, and are widely used as verification data.

More realistic representation of the anomalous pressure systems in the GLDAS reforecasts results in better predictions of monthly mean rainfall anomalies compared to the CFSR reforecasts (Fig. 14). Both the CFSR and GLDAS reforecasts seem to capture below-normal (above-normal) rainfall anomalies over the Great Plains (in Florida) at 1-month lead (October 1999), but the GLDAS reforecasts are more similar to observations in terms of the spatial pattern and magnitude of rainfall anomalies. At 2-month lead (November 1999), while the CFSR reforecasts show a typical pattern of anomalous

Fig. 12. As in Fig. 8, but for monthly snow depth anomalies (cm).
rainfall over the United States in the winter of developing La Niñas as seen in many previous studies (e.g., Huang et al. 2019), the GLDAS reforecasts reasonably well predict suppressed rainfall over the central United States associated with anomalous high pressure, as well as enhanced rainfall over the northwestern coastal regions related to the anomalous low pressure located in the northeastern Pacific (Figs. 13b and 14b). Also, below-normal rainfall anomalies over California and Oregon in the GLDAS reforecasts are in good agreement with those in observations at 3-month lead (December 1999) (Fig. 14c). On the other hand, the CFSR reforecasts predict overly enhanced rainfall in the northeastern United States (Fig. 14c), which leads to unrealistically wet conditions in predicted SPI3 over the northern United States in the center panels of Figs. 7c and 7d, although they seem to show better prediction of suppressed rainfall in the southeastern United States than the GLDAS reforecasts.

Our case study, therefore, suggests that more realistic representation of the land initial states results in better realization of land–atmosphere coupling that can play a critical role in improving drought predictions.

5. Summary and discussion

The NCEP CFSv2 ensemble reforecasts initialized with two land analyses are conducted for the period of 1979–2010. The two observation-based land initial states are adapted from the NCEP CFSR and the NASA GLDAS-2 analysis, whereas atmosphere, ocean, and ice initial states are identical for both reforecasts, forming a pair of identical-twin experiments. The ensemble mean of CFSv2 reforecasts with CFSR (GLDAS-2) land ICs is referred to as the CFSR (GLDAS) reforecasts, and they have 12-month duration starting from early January, April, July, and October. This study compares correlation skill of 3-month SPI (SPI3) and indices of categorical forecasts (e.g., equitable threat score) over CONUS between the CFSR and GLDAS reforecasts to quantify how prediction skill of U.S. drought is responsive to different land initial states on seasonal time scales. It is evident that SPI3 prediction skill over CONUS in summer degrades...
much faster with increasing lead time compared to SPI3 predictions in winter when the remote oceanic forcing (e.g., ENSO SST anomaly) is dominant. Overall, there is no outstanding performer for all locations, seasons, and lead times between the forecasts with different land ICs. Instead, one set of reforecasts shows better skill than the other in some regions, but their locations vary in lead time and season. For all starting months, the GLDAS reforecasts tend to show a higher percentage of grid cells over CONUS with statistically significant skill than the CFSR reforecasts at short leads (about up to 3-month lead), while at long leads (after one season), the CFSR reforecasts have higher percentage of grid cells with significant skill than the GLDAS reforecasts (Fig. 3 and Fig. S3). This is interesting and further work is required to find what caused this feature. For example, we may need to 1) evaluate soil moisture forecast skill as a function of lead month, 2) investigate possibly different characteristics of snow accumulation from fall to winter and of snow melting from winter to spring, and 3) examine possible differences in remote ocean forcing at long lead between the two sets of reforecast.

We selected a total of 38 severe drought events at given target months corresponding to at 3-month lead from each starting month for 1979–2010, which were determined by the criteria that the percentage of grid cells over the CONUS with the observed SPI3 $<-1.2$ is greater than a threshold for each target month as summarized in Table 1. Among the selected 38 events, we find 9 predicted and 9 missed cases, as well as 9 (11) events with moderate (limited) prediction skill. More well-predicted cases appear in December and March while more missed cases are listed in June and September, implying that U.S. severe droughts are more predictable in winter and spring.

Land surface feedbacks appear to interact with the remote ocean drivers, which can combine to help maintain the drought (e.g., Yang et al. 2001; Hong and Kalnay 2002; Schubert et al. 2004). In addition to remote SST forcing, realistic representation of land states (viz., soil moisture) over CONUS is critical for prediction of U.S. severe drought events approximately one season in advance. In our detailed case study of a severe drought event occurring in the winter of 1999, a pattern of below normal SST (i.e., developing La Niña) in the tropical Pacific and a negative PDO phase in the North Pacific is realistically reproduced in both sets of reforecasts. However, GLDAS-2 land initial states show stronger east–west gradient.
of soil moisture with drier in the eastern United States than CFSR ones, which tends to persist for about one season, leading to more compatible land forcing with observation compared to that of the CFSR reforecasts. The winter 1999 drought was especially large in spatial extent, including much of the southern and southwestern United States, which may be a factor in the contribution of land surface feedbacks that are normally thought to be locally weak across much of the midlatitudes during winter.

As Shin et al. (2020) confirmed, soil moisture interacts closely with surface air temperature in both the CFSR and GLDAS reforecasts. Coupling of drier soil moisture and warmer surface air temperature centered in the central United States for the GLDAS reforecasts makes the spatial patterns of anomalous soil moisture and 2-m temperature in 1999 more consistent with those in observations. Positive feedback of increased 2-m temperature and reduced snow cover results in warm anomalies lasting for more than a month over North America, which are accompanied by persistent barotropic (warm core) anomalous high pressure during this drought event. More realistic anomalous pressure simulation in the GLDAS reforecasts results in better prediction skill of this drought case at up to 4-month lead times. Our results suggest that reduction of the uncertainty of land surface properties among the current land analyses can play an important role in improving U.S. severe drought prediction on seasonal time scales.

Acknowledgments. This work is supported by the NOAA MAPP drought project (NA17OAR4310144). We acknowledge the Extreme Science and Engineering Discovery Environment (XSEDE) for providing the computational resources for the two sets of 32-yr (1979–2010) CFSv2 reforecasts. We also thank the editor and the anonymous reviewers for their constructive comments and suggestions.

Data availability statement. The two sets of CFSv2 reforecasts were performed on the Texas Advance Computing Center (TACC’s) high-performance computing platform (Stampede and Stampede2) and model output data are stored on the TACC’s long-term tape archival storage system (Ranch). The CPC daily unified gauge-based analysis of precipitation were retrieved from https://climatedataguide.ucar.edu/climate-data/cpc-unified-gauge-based-analysis-global-daily-precipitation. The global monthly extended reconstructed SSTs, version 5 (ERSSTv5), were retrieved from https://www.ncdc.noaa.gov/data-access/oceanographic-data/extended-reconstructed-sea-surface-temperature-ersst-v5. The ERA-Interim reanalysis data were retrieved from https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim. The SMERGE 0.40 cm root zone soil moisture (doi:10.5067/PAVQY1KHTMUT) were retrieved from the Earthdata Search Client (EDSC), https://search.earthdata.nasa.gov/search?q=SMERGE_RWZSM0_40CM.

REFERENCES


Crow, W., and K. Tobin, 2018: Smerge-Noah-CCI root zone soil moisture 0-40 cm L4 daily 0.125 × 0.125 degree V1.0. Goddard Earth Sciences Data and Information Services Center (GES DISC), accessed June 2020, https://doi.org/10.5067/NRIWAMBNN6D.


