ENVIRONMENTAL RESEARCH
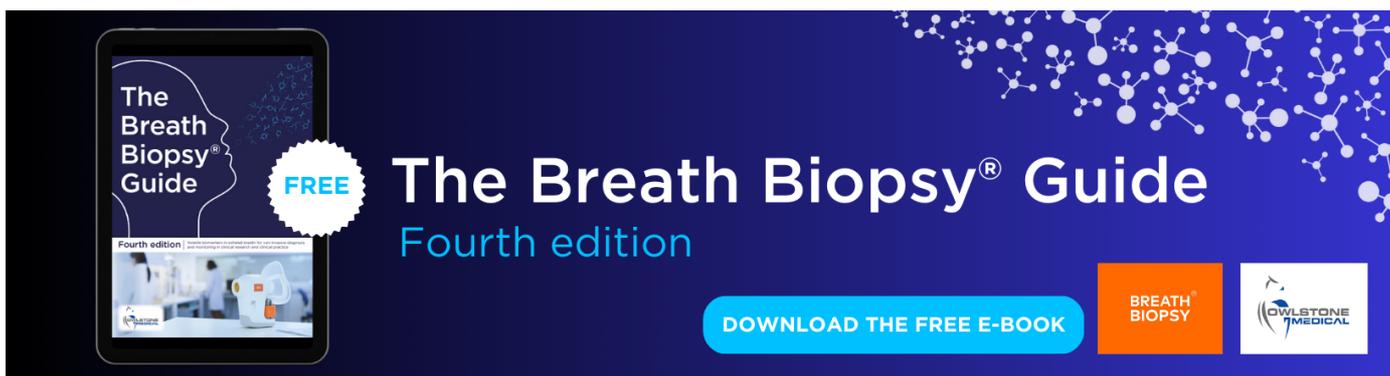## LETTERS

LETTER • OPEN ACCESS

Potential caveats in land surface model evaluations using the US drought monitor: roles of base periods and drought indicators

To cite this article: Hailan Wang *et al* 2022 *Environ. Res. Lett.* **17** 014011

View the article online for updates and enhancements.

ENVIRONMENTAL RESEARCH
LETTERS

LETTER

Potential caveats in land surface model evaluations using the US drought monitor: roles of base periods and drought indicators

Hailan Wang[1,*] , Li Xu[1,2], Mimi Hughes[3], Muthuvel Chelliah[1], David G DeWitt[1], Brian A Fuchs[4] and Darren L Jackson[3,5]

1 NOAA/NWS Climate Prediction Center, College Park, MD 20740, United States of America
2 Innovim LLC, Greenbelt, MD 20770, United States of America
3 NOAA/ESRL Physical Sciences Laboratory, Boulder, CO 80305, United States of America
4 National Drought Mitigation Center, Lincoln, NE 68583, United States of America
5 Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO 80309, United States of America
* Author to whom any correspondence should be addressed.

E-mail: Hailan.Wang@noaa.gov

Abstract

The US drought monitor (USDM) has been widely used as an observational reference for evaluating land surface model (LSM) simulation of drought. This study investigates potential caveats in such evaluation when the USDM and LSMs use different base periods and drought indices to identify drought. The retrospective national water model (NWM) v2.0 simulation (1993–2018) was used to exemplify the evaluation, supplemented by North American land data assimilation system phase 2 (NLDAS-2). Over their common period (2000–2018), in distinct contrast with the USDM which shows high drought occurrence (>50%) in the western half of the continental US (CONUS) and the southeastern US with low occurrence (<30%) elsewhere, the NWM and NLDAS-2 based on soil moisture percentiles (SMPs) consistently show higher drought occurrence (30%–40%) in the central and southeastern US than the rest of the CONUS. Much of the differences between the LSMs and USDM, particularly the strong LSM underestimation of drought occurrence in the western and southeastern US, are not attributed to the LSM deficiencies, but rather the lack of long-term drought in the LSM simulations due to their relatively short lengths. Specifically, the USDM integrates drought indices with century-long periods of record, which enables it to capture both short-term (<6 months) drought and long-term (⩾6 months) drought, whereas the relatively short retrospective simulations of the LSMs allows them to adequately capture short-term drought but not long-term drought. In addition, the USDM integrates many drought indices whereas the NWM results are solely based on the SMP, further adding to the inconsistency. The high occurrence of long-term drought in the western and southeastern US in the USDM is further found to be driven collectively by the post-2000 long-term warm sea surface temperature (SST) trend, cold Pacific decadal oscillation and warm Atlantic multi-decadal oscillation, all of which are typical leading patterns of global SST variability that can induce drought conditions in the western, central, and southeastern US. Our findings highlight the effects of the above caveats and suggest that LSM evaluation should stay qualitative when the caveats are considerable.

1. Introduction

Land surface models (LSMs) have been commonly used as an objective tool for operational drought monitoring (e.g. soil moisture) (Wood *et al* 2015). Land analyses, produced by driving LSMs with observation-based meteorological forcings (e.g. North American land data assimilation system, phase 2—NLDAS-2, Xia *et al* 2012a, 2012b), provide data continuous in both time and space and thus facilitate land surface monitoring. The LSMs simulate the exchange of water and energy fluxes at the Earth's

land surface and have been used as an essential tool to understand, simulate, and predict the land surface and its role within the Earth system. The LSMs are nevertheless subject to limitations in their representations (e.g. parameterization schemes) of land physical processes and the quality of the input meteorological forcing data. Their evaluation using observations is thus essential for the LSM development and the assessment of land products they produce.

The evaluation of LSMs in simulating drought requires independent observation-based data (e.g. for soil moisture) as the references. Such data include ground observations, satellite observations, and the US drought monitor (USDM, http://droughtmonitor.unl.edu). These data have their respective strengths and weaknesses (Ford and Quiring 2019). The ground observations are based on *in-situ* measurements and typically represent ground truth. They are nevertheless point observations with limited spatial and temporal coverage; some of them are also subject to random and systematic measurement errors. The satellite observations are available at the global scale. Their data quality is subject to coverage gaps due to satellite orbits, capabilities of satellite sensors in detecting land surface properties, as well as performance of calibration algorithms. The satellite data records are also relatively short, typically ranging from several years to a few decades (Beck *et al* 2021), which makes them difficult for monitoring long-term drought. As an alternative, the USDM has been commonly used to evaluate LSM simulations of drought, including both drought statistics and individual drought events (e.g. Su *et al* 2021, Mocko *et al* 2021). The USDM starts from 2000. It is an operational weekly map that shows the location and intensity of drought across the US. It uses five drought categories based on percentile ranking, consisting of $D0$ (abnormally dry, 21%–30%), $D1$ (moderate drought, 11%–20%), $D2$ (severe drought, 6%–10%), $D3$ (extreme drought, 3%–5%) and $D4$ (exceptional drought, 0%–2%). The drought categories are determined based on a combination of both objective and subjective expert assessments, where the 'convergence of evidence' approach is used to integrate short- and long-term drought indicators based on precipitation, temperature, soil moisture, streamflows and reservoir levels, runoff, snow water equivalent, and regional drought impacts. The number of the input indices and indicators has changed over time, which increased from ~5 to 6 in early USDM years to several dozen today. It is also worth noting that these input drought indicators and indices are computed using their respective datasets and unique periods of record, which vary from century long for divisional precipitation measurements (1895–present) to a few decades for satellite-based observations. The USDM website (https://droughtmonitor.unl.edu/About/WhatistheUSDM.aspx) provides reference information on converting individual drought indices (e.g.

soil moisture percentiles—SMPs) to the USDM drought categories, which enables a direct quantitative comparison between the USDM and LSM simulations of drought based on a single drought index.

This study has two related but distinct goals. First, it aims to perform an in-depth investigation on the LSM evaluation using the USDM. While the differences between the USDM and LSM-based objective drought indices have been discussed in the past (e.g. Wood *et al* 2015, Mocko *et al* 2021), a systematic investigation of potential caveats therein is lacking. The study will address this by systematically studying the fidelity in such evaluation, identifying potential caveats, and assessing their effects. Second, in the context of the above evaluation, the study investigates the causes of spatial distribution of drought occurrence in the USDM, and reveals why relatively short LSM simulations have difficulties in reproducing the spatial distribution.

2. Methods

2.1. LSM evaluation using the USDM

To exemplify the LSM evaluation using the USDM, the national water model (NWM) v2.0 retrospective simulation (1993–2018) (https://water.noaa.gov/about/nwm) was used, supplemented by NLDAS-2 (1979–present). The NWM is an operational hydrologic modeling framework built on Weather Research and Forecasting-Hydro. It produces hydrologic guidance at a very fine spatial and temporal scale (e.g. 1 km and 3 h for soil moisture) covering the continental US (CONUS). The NWM v2.0 retrospective simulation was produced by driving the NWM v2.0 with the NLDAS-2 hourly meteorological forcings. NLDAS-2 is an operational multi-model land modeling system run uncoupled to the atmosphere covering the CONUS. It consists of four LSMs, among which Mosaic, Noah and variable infiltration capacity (VIC) were used in this study. The data are in 1/8° grid spacing and hourly from 2 January 1979 to the present with a 4 d latency.

To facilitate the USDM and LSM comparison, all the data were spatially interpolated onto NLDAS-2's 1/8° spatial grid. The USDM data (2000–present) are weekly and are obtained by rasterizing the USDM shapefiles. To match the USDM, for each of the USDM map release dates, the daily top 1 m SMPs in the NWM and NLDAS-2 were computed relative to the 1993–2018 period using a 15 d moving window of daily soil moisture values, where the use of the same base period for the NWM and NLDAS-2 facilitates a fair LSM intercomparison. The SMPs were subsequently converted to $D0$–$D4$ drought categories following the USDM recommendations (e.g. 21%–30% for $D0$).

The evaluation of the NWM v2.0 and NLDAS-2 using the USDM focused on their common period (2000–2018). For a given grid box, a drought is

defined to occur when the grid box has drought category of D0, D1, D2, D3 or D4. The evaluation metrics for drought statistics are frequency of drought occurrence and metrics based on a contingency table (hit, miss, false alarm, correct negative), where the former is defined as the percentage of number of weeks in drought, and the latter includes probability of detection, false alarm ratio (FAR), and critical success index. The frequency of occurrence for all drought conditions is further separated into those for short-term (<6 months) drought and long-term (⩾6 months) drought, where the drought duration of 6 months is used for the separation following the USDM conventions. Such evaluations were performed for D0–D4 as well as higher drought categories (e.g. D2–D4). Since the choices of drought categories affect the magnitude of drought statistics (e.g. frequency of occurrence) but not the qualitative picture and conclusions we draw (not shown), we only show the evaluation results for D0–D4.

It is worth noting that the evaluation of the NWM and NLDAS-2 using the USDM is complicated by their use of different base periods and drought indicators to identify drought anomalies. Specifically, the drought indicators and indices used in the USDM production have differing periods of historical record, which vary from century long for divisional precipitation measurements to a few decades for drought indices based on satellite observations. By comparison, the NWM and NLDAS-2 use 1993–2018 as the base period to compute SMPs. Further, the USDM is obtained by subjectively integrating a number of drought indicators, whereas the LSM results shown here use a single index—SMP—to indicate drought conditions.

2.2. Long-term VIC land surface analysis

To assess the separate effects of using different base periods and drought indicators on estimating drought statistics, we analyzed daily soil moisture from a century-long (1915–2011) land surface dataset for the CONUS. The dataset was generated by driving the VIC hydrologic model v4.1.2c with a companion set of observed daily meteorological forcings (Livneh *et al* 2013).

2.3. Diagnosis of spatial distribution of USDM drought occurrence
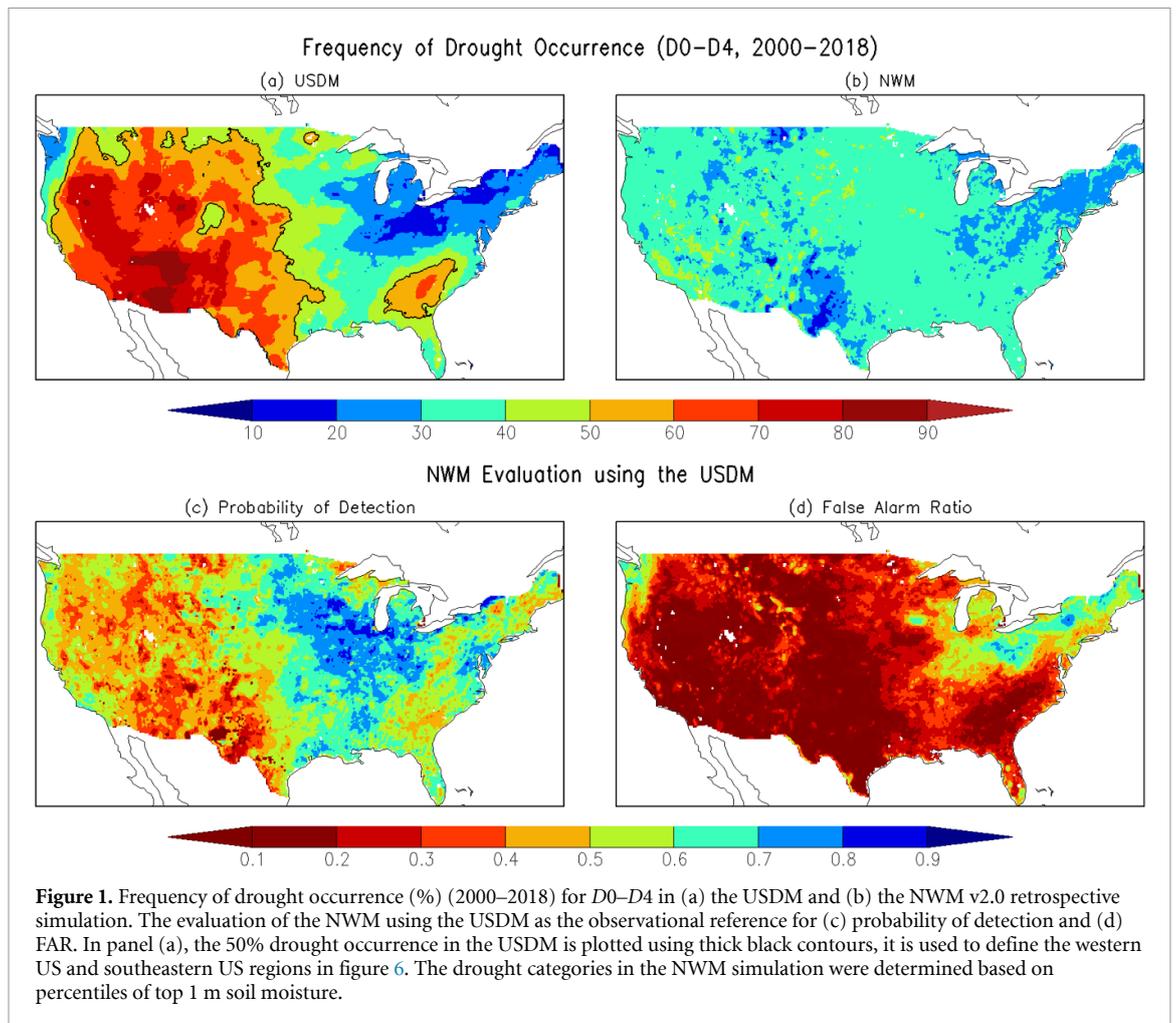
To investigate the causes of spatial distribution of USDM drought occurrence, we studied the effects of climate variations on decadal and longer timescales by performing an integrated data diagnosis using long-term observations and atmospheric model intercomparison project (AMIP)-style simulations. The observations analyzed include the global precipitation climatology centre (GPCC) precipitation (1891–2019, Schneider *et al* 2011), the national aeronautics and space administration, goddard institute for space studies (NASA GISS) surface temperature

analysis (1880–present, Lenssen *et al* 2019), and the national oceanic and atmospheric administration (NOAA) extended reconstructed sea surface temperature (SST) (1854–present, Huang *et al* 2017). To investigate the contributions from SST and atmospheric internal variability, we also analyzed three long-term AMIP-style simulations forced with observed SST and time-varying external radiative forcings, produced respectively using the national center for atmospheric research (NCAR) community atmosphere model version 5 (CAM5, 1900–2019, 40 ensembles), the geophysical fluid dynamics laboratory, atmospheric model version 3 (GFDL AM3) (1870–2014, 17 ensembles), and the national aeronautics and space administration, goddard earth observing system model, version 5 (NASA GEOS-5) (1871–2014, 12 ensembles) (Murray *et al* 2020). The analysis of the AMIP simulations focused on the CAM5 as it is available through 2019 and has data for more variables (e.g. soil moisture) and ensembles available. In the analysis, the 20th century (1900–1999) was used as the base period to identify drought, as it not only provides adequate samples but also is sufficiently long to average out much of the effects of natural decadal-to-multidecadal variabilities (e.g. Pacific decadal oscillation (PDO), Atlantic multidecadal oscillation (AMO)). The period 2000–2018 was contrasted with the 20th century base period to examine their differences in mean climate and drought statistics, with a focus on the effects of the post-2000 SST anomalies. Since long-term soil moisture observations are unavailable, much of the investigation focused on precipitation and temperature as they are the key meteorological drivers for drought.

3. Results

3.1. LSM evaluation using the USDM

Figure 1(a) shows the frequency of drought occurrence over the CONUS in the USDM for 2000–2018. The most striking features are the considerably more frequent drought occurrence (>50%) in much of the western US, the Great Plains and southeastern US than elsewhere in the CONUS, consistent with past studies (e.g. Chen *et al* 2019). In the western US, the frequency of drought occurrence is particularly high (>70%) in Arizona, New Mexico and most of Utah. By comparison, the Midwest, northeast and coastal Pacific Northwest had <20% drought occurrence. In distinct contrast with the USDM, the NWM (figure 1(b)) shows a spatially more homogeneous distribution, with moderately more droughts occurring in the Great Plains, southeastern US, and California than the rest of the CONUS. Much of the high drought occurrence in the western and southeastern US in the USDM is almost entirely missing in the NWM simulation. As a result, when using the USDM as the observational reference, the NWM has rather low detection rate (<0.5) in the western
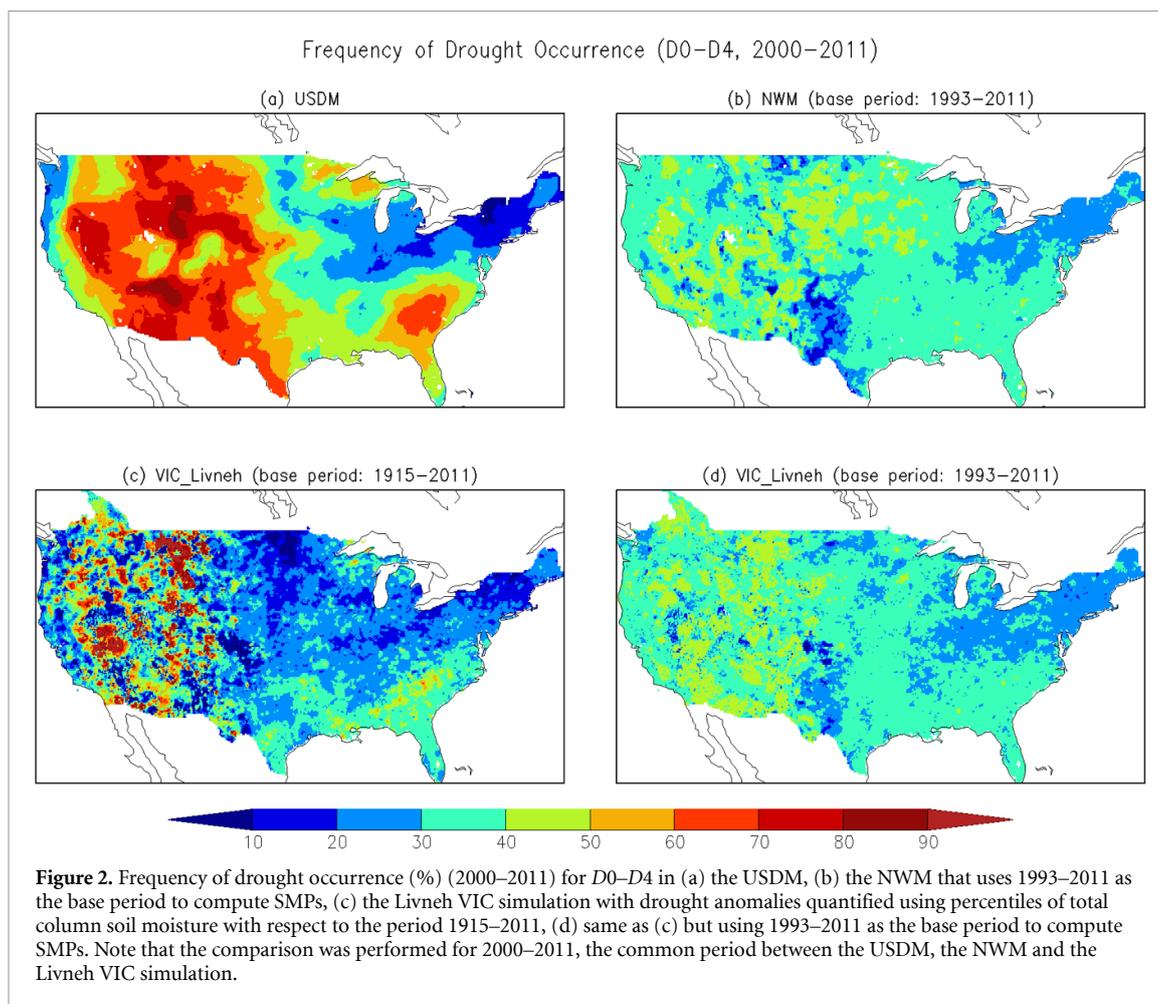
Figure 1. Frequency of drought occurrence (%) (2000–2018) for *D*0–*D*4 in (a) the USDM and (b) the NWM v2.0 retrospective simulation. The evaluation of the NWM using the USDM as the observational reference for (c) probability of detection and (d) FAR. In panel (a), the 50% drought occurrence in the USDM is plotted using thick black contours, it is used to define the western US and southeastern US regions in figure 6. The drought categories in the NWM simulation were determined based on percentiles of top 1 m soil moisture.

and southeastern US, and has relatively high detection rate (>0.7) in regions where both the USDM and NWM show lower drought occurrence (e.g. Midwest) (figure 1(c)). The FAR is overall low (<0.2), particularly in the western, central and southeastern US; the exceptions are in the northeastern US and coastal Washington state where the FAR exceeds 0.5 (figure 1(d)). The three NLDAS-2 LSMs (figure S1 (available online at stacks.iop.org/ERL/17/014011/mmedia)) in general well resemble the NWM for the above drought statistics.

3.2. Roles of base period and drought indicators
The distinct differences between the LSMs and USDM (figures 1 and S1), particularly the rather low probability of detection in the western and southeastern US in the LSMs, raise the question of whether they reflect the true performance of the LSMs or are due to the use of different base periods and drought indicators between the LSMs and USDM in estimating drought. To investigate this, we analyzed the century-long VIC land analysis (Livneh *et al* 2013), while keeping in mind that it uses a LSM and input meteorological forcings different from the NWM and NLDAS-2. To assess the dependence of drought anomalies on the length of base periods, SMPs were computed using two base periods: a century-long 1915–2011 and a

short 1993–2011. When the short base period (1993–2011) is used, the Livneh VIC simulation (figure 2(d)) broadly agrees with the NWM (figure 2(b)) in that their spatial distribution is in general homogeneous across the CONUS. Such homogeneous distribution is expected, as the period (2000–2011) over which the SMP-based drought occurrence was obtained is not too different from the base period (1993–2011) used to define soil moisture anomalies. In fact, when the two periods are identical, one would expect a perfect spatially homogenous distribution by design. In contrast, when the century-long base period (1915–2011) is used, the Livneh VIC simulation captures considerably more droughts in the western and southeastern US (cf figures 2(c) and (d)) and shows lower drought occurrence in the northeastern US, yielding a better agreement with the USDM (figure 2(a)). The Livneh VIC simulation, however, still considerably differs from the USDM (cf figures 2(a) and (c)), displaying underestimations across parts of the western, central and southeastern US. These differences could be due in part to the differences in the drought indicators they use and in part to the VIC performance in these geographical regions. Despite these differences, figure 2 suggests that the choices of base periods and drought indicators matter for drought monitoring. In particular, when LSMs use

Figure 2. Frequency of drought occurrence (%) (2000–2011) for *D*0–*D*4 in (a) the USDM, (b) the NWM that uses 1993–2011 as the base period to compute SMPs, (c) the Livneh VIC simulation with drought anomalies quantified using percentiles of total column soil moisture with respect to the period 1915–2011, (d) same as (c) but using 1993–2011 as the base period to compute SMPs. Note that the comparison was performed for 2000–2011, the common period between the USDM, the NWM and the Livneh VIC simulation.

relatively short base periods to identify drought, the inter-model differences (e.g. figure 2(b) vs 2(d)) in drought statistics are substantially smaller than their differences from the USDM (cf figures 2(b) and (d) with figure 2(a)); much of the differences between the LSMs and the USDM are attributed to their differing aspects of drought identifications rather than the LSM deficiencies.

The causes of the differences between the LSMs and USDM were further investigated by decomposing their drought anomalies into short-term and long-term ones and comparing their statistics (figures 3(a)–(d) and S2) with those of the total (figures 1(a)–(b) and S1). Such decomposition method can effectively separate drought events of different durations (Andreadis *et al* 2005, Sheffield and Wood 2008). Focusing on the NWM, the comparison clearly shows that the frequent drought occurrence in the western and southeastern US in the USDM is contributed by long-term drought. By comparison, the NWM shows weak indications of long-term drought in the western US (<0.3) and little indications elsewhere. The USDM and NWM have a considerably better agreement for short-term drought, with the NWM showing slightly higher occurrences in the central US, parts of the southeastern US and the Pacific Northwest. Their differences mainly occur

in the northeastern US, where the NWM (<0.3) has higher drought occurrence than the USDM (<0.2). The three NLDAS-2 LSMs broadly agree with the NWM (figure S2). Figure 3(f) further examines the ratio of the NWM drought occurrence to that in the USDM for each of the US hydrologic unit code (HUC)-2 regions. The NWM considerably underdetects long-term drought in the USDM for all regions, with the ratio ranging from ∼30% in the western US to ∼50% in the HUC-2 region 9 (Souris-Red-Rainy). For short-term drought, the NWM overestimates the USDM in both the western and eastern US (e.g. by 50% in HUC-2 region 1 New England) while underestimating the USDM in the central US.

Taken collectively, the above results suggest that much of the differences between the LSMs and USDM are traced to the lack of long-term drought in the LSMs due to their relatively short simulations. Their use of different drought indicators to quantify drought further adds to the inconsistency.

3.3. Causes of spatial distribution of drought occurrence in the USDM

We next investigated the causes of the spatial distribution of drought occurrence in the USDM, particularly the higher occurrence in the Intermountain West, Great Plains and southeastern US. Since a number of
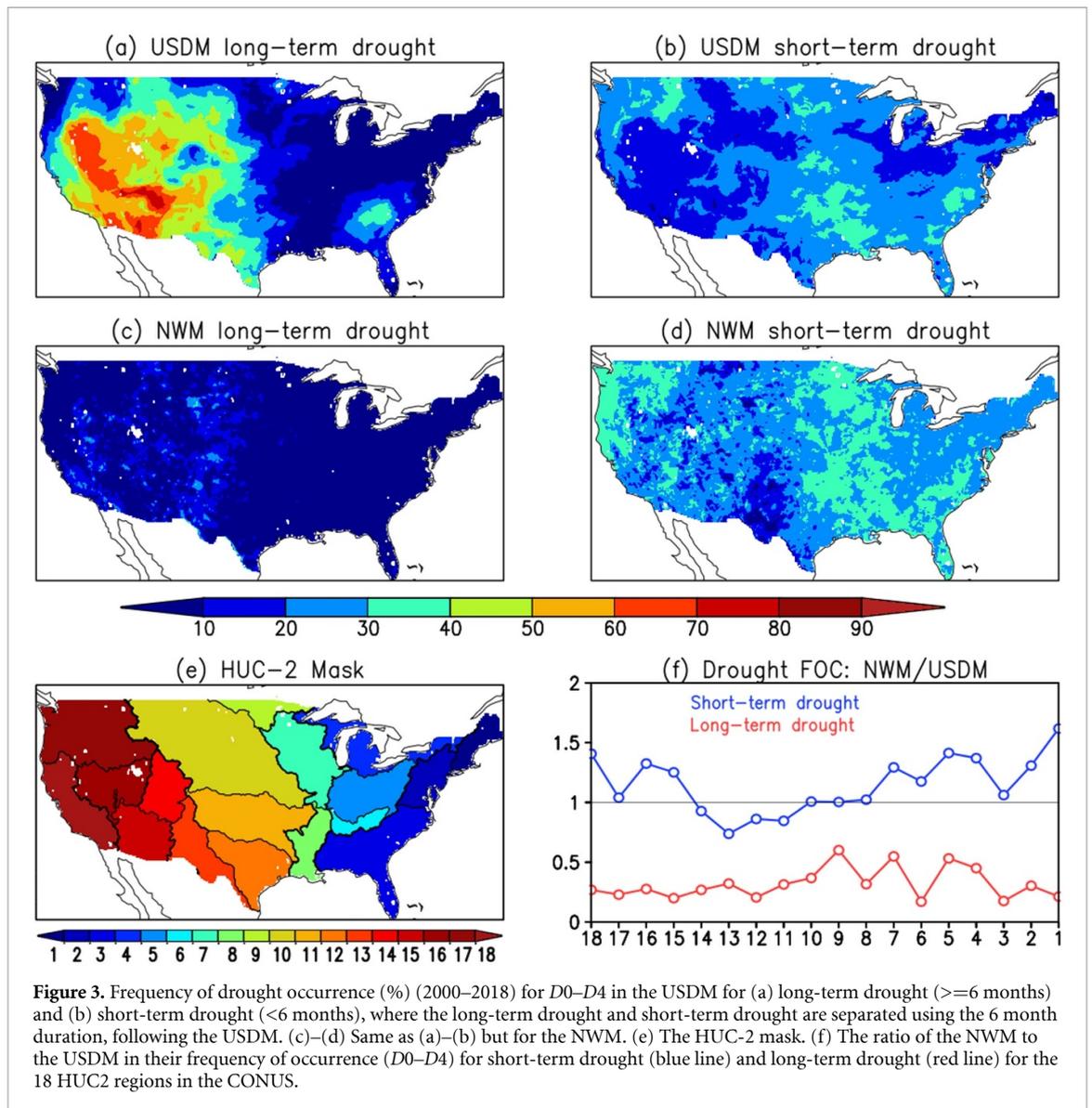
Figure 3. Frequency of drought occurrence (%) (2000–2018) for *D*0–*D*4 in the USDM for (a) long-term drought (>=6 months) and (b) short-term drought (<6 months), where the long-term drought and short-term drought are separated using the 6 month duration, following the USDM. (c)–(d) Same as (a)–(b) but for the NWM. (e) The HUC-2 mask. (f) The ratio of the NWM to the USDM in their frequency of occurrence (*D*0–*D*4) for short-term drought (blue line) and long-term drought (red line) for the 18 HUC2 regions in the CONUS.

input drought indices and indicators (e.g. precipitation) in the USDM use long base periods to identify drought, the USDM drought occurrence likely reflects the effects of post-2000 anthropogenic and natural climate variations on decadal and longer timescales (e.g. Williams *et al* 2015, 2020, Berg and Hall 2017, USGCRP 2018, Xiao *et al* 2018). To investigate this, we first examined the mean differences between the post-2000 period and the 20th century for precipitation and temperature, the immediate meteorological drivers for drought. Not surprisingly, precipitation (figure 4(a)) shows a remarkable spatial resemblance to the USDM drought occurrence (figure 1(a)). Relative to the 20th century, the post-2000 precipitation decreased in the western and southeastern US where the USDM shows higher drought occurrence, and increased in the Midwest and northeastern US where the USDM has lower drought occurrence. The elevated surface warming is prominent across much of the CONUS, ranging from 0.4 K in the southeastern US to 1.2 K in the southwestern US. The

spatial resemblance between the mean precipitation changes (figure 4(a)) and the USDM drought occurrence (figure 1(a)) suggests the dominant role of precipitation deficits in driving drought, consistent with past studies (e.g. Livneh and Hoerling 2016, Luo *et al* 2017, Koster *et al* 2019).

The above mean changes in precipitation and temperature (figures 4(a) and (b)) presumably result from the effect of climate variations on decadal and longer timescales, a considerable amount of which is reflected in oceanic low-boundary conditions (e.g. SST). Figure 4(c) shows that, relative to the 20th century, the post-2000 period had an overall SST warming, with stronger warming (>0.6 K) in the Indian Ocean, western Pacific, much of the tropical and North Atlantic, and weaker warming (<0.4 K) in the central and eastern Pacific. Much of these mean SST changes are due to the combined contributions from the global warming trend, negative PDO, and positive AMO (figure 4(d)). These three SST modes are among the leading modes of global SST variability
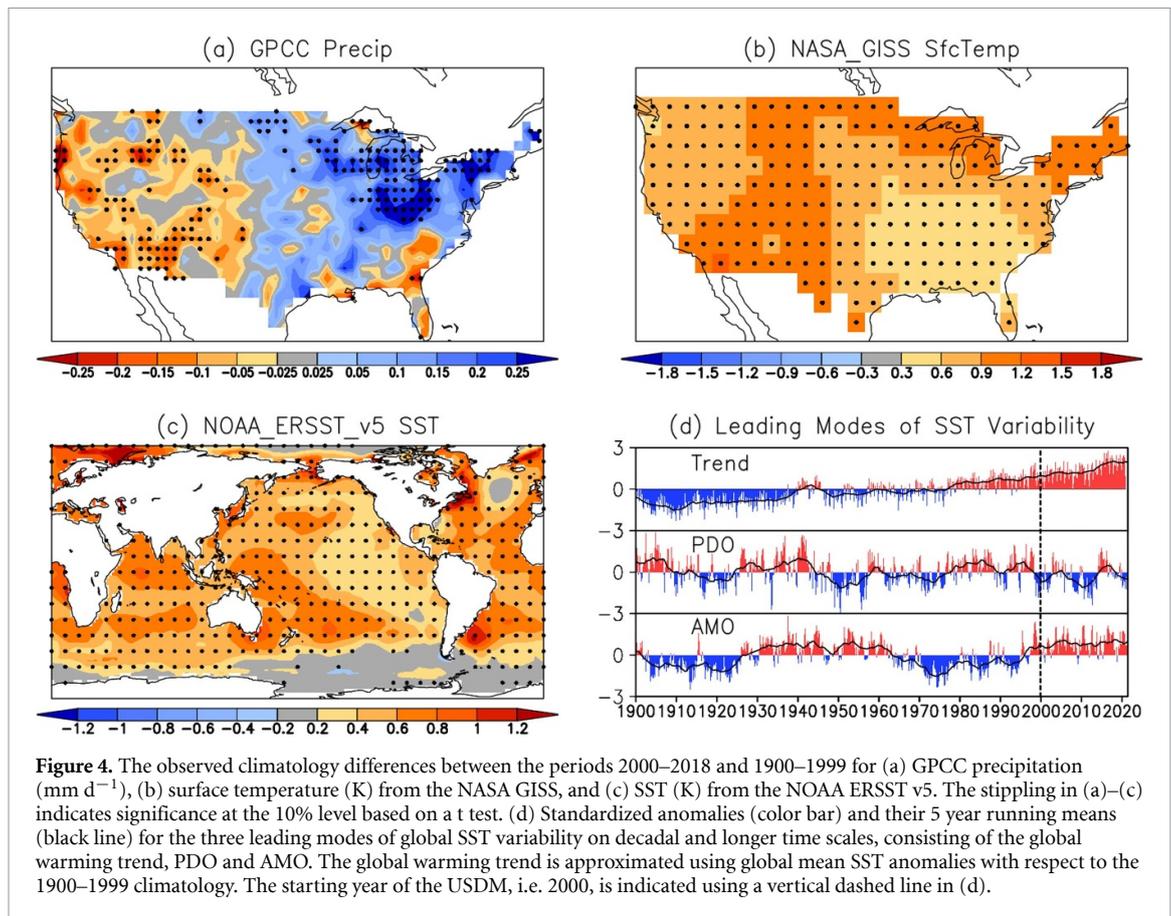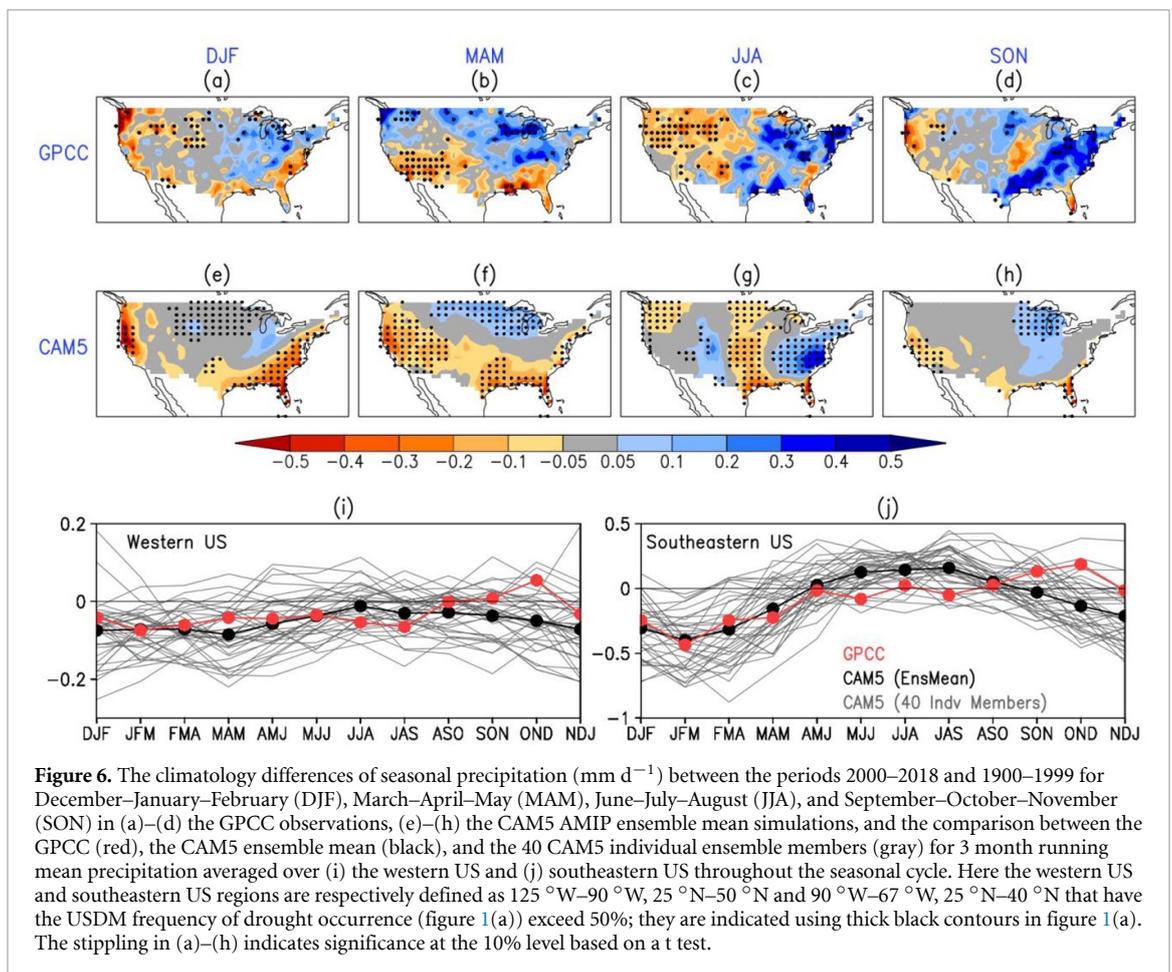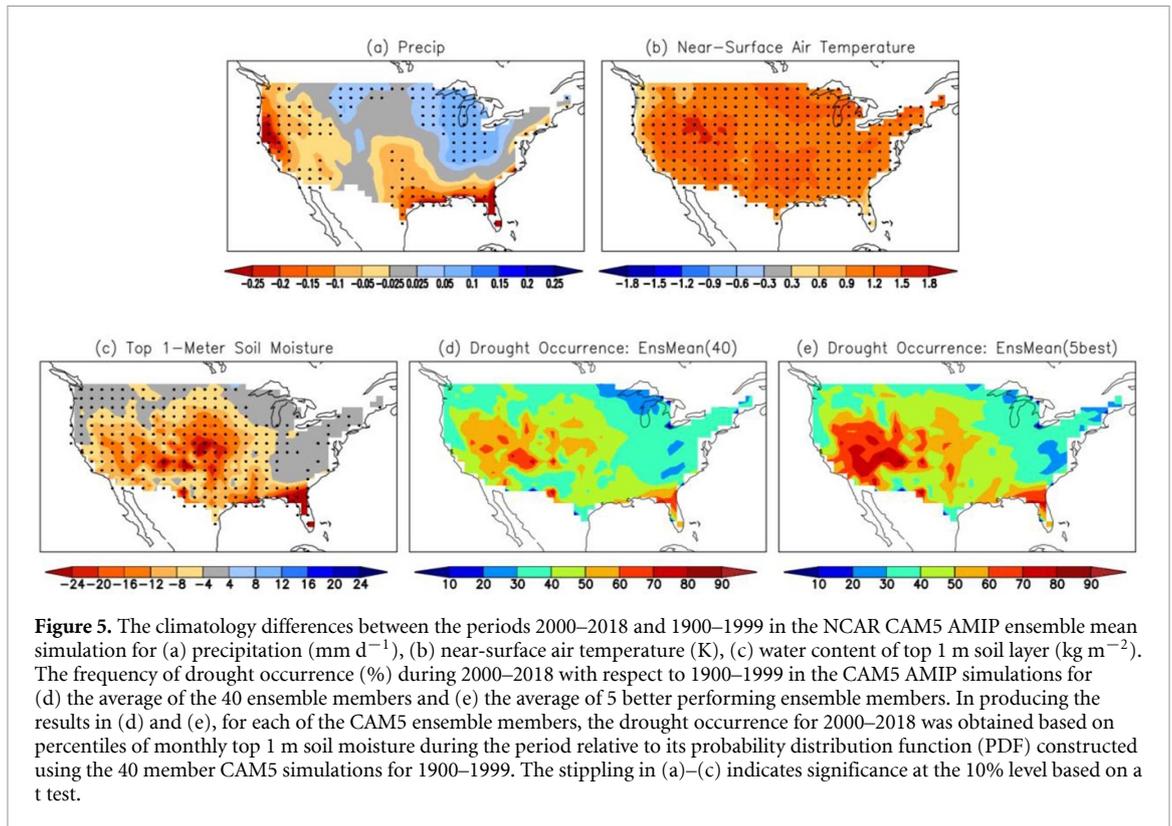
Figure 4. The observed climatology differences between the periods 2000–2018 and 1900–1999 for (a) GPCC precipitation (mm d⁻¹), (b) surface temperature (K) from the NASA GISS, and (c) SST (K) from the NOAA ERSST v5. The stippling in (a)–(c) indicates significance at the 10% level based on a t test. (d) Standardized anomalies (color bar) and their 5 year running means (black line) for the three leading modes of global SST variability on decadal and longer time scales, consisting of the global warming trend, PDO and AMO. The global warming trend is approximated using global mean SST anomalies with respect to the 1900–1999 climatology. The starting year of the USDM, i.e. 2000, is indicated using a vertical dashed line in (d).

(Schubert *et al* 2009). Their phases, as occurred in the post-2000 period, are well-known leading SST patterns that can induce drought conditions over the western, central and southeastern US via atmospheric teleconnections (e.g. Mo *et al* 2009, Kushnir *et al* 2010, Schubert *et al* 2009, Wang *et al* 2010).

To investigate the roles of the post-2000 global SST changes and atmospheric internal variability, we turned to the century-long (1900–2019) 1° 40 ensemble CAM5 AMIP simulations (Murray *et al* 2020). The AMIP ensemble mean highlights the SST-forced signal whereas the ensemble spread reflects the unforced variability generated by processes internal to the atmosphere. The precipitation differences in the CAM5 ensemble mean broadly agree with those of the GPCC observations. Consistent with the GPCC (figure 4(a)), the CAM5 ensemble mean (figure 5(a)) shows dry responses in the western, southern central and southeastern US and wet anomalies in the Midwest and nearby regions. Strong warming responses span across the entire CONUS, with the peak warming (∼1.5 K) occurring in parts of the western US (figure 5(b)). Associated with the changes in precipitation and temperature, the CAM5 ensemble mean soil moisture considerably dries in the western, southern central and southeastern US (figure 5(c)). As a result, relative to the 20th century, the recent decades (2000–2018) in the CAM5 ensemble mean have >40% of drought occurrence

across much of the western and southeastern US, with lower occurrence elsewhere (figure 5(d)). Select individual ensemble members have higher drought occurrence (>50%) in the western and southeastern US (figure 5(e)), in better agreement with the USDM (figure 1(a)). The above results, particularly the spatial resemblance between the CAM5 ensemble mean and the observations for precipitation and drought occurrence, suggest that the mean SST changes acted as a main driver.

Given the prominent role of accumulated precipitation deficits in driving drought and that their spatial distribution and controlling physical processes vary from season to season, figure 6 further examines the mean precipitation changes by focusing on their variations with season. The CAM5 ensemble mean simulation was compared with the GPCC to infer the role of SST. During winter and spring, consistent with the GPCC (figures 6(a)–(d)), the CAM5 ensemble mean (figures 6(e)–(h)) produces dry anomalies in the western and southeastern US and wet anomalies in the Midwest and nearby regions, much of which resemble the effect of a cold PDO (Newman *et al* 2016). During summer and fall, the CAM5 ensemble mean agrees with the GPCC in dry anomalies in the western and southeastern coastal states; there are however considerable differences in the central and eastern US, where the observed anomalies could be due to processes unrelated to the mean

Figure 5. The climatology differences between the periods 2000–2018 and 1900–1999 in the NCAR CAM5 AMIP ensemble mean simulation for (a) precipitation (mm d^{-1}), (b) near-surface air temperature (K), (c) water content of top 1 m soil layer (kg m^{-2}). The frequency of drought occurrence (%) during 2000–2018 with respect to 1900–1999 in the CAM5 AMIP simulations for (d) the average of the 40 ensemble members and (e) the average of 5 better performing ensemble members. In producing the results in (d) and (e), for each of the CAM5 ensemble members, the drought occurrence for 2000–2018 was obtained based on percentiles of monthly top 1 m soil moisture during the period relative to its probability distribution function (PDF) constructed using the 40 member CAM5 simulations for 1900–1999. The stippling in (a)–(c) indicates significance at the 10% level based on a t test.

Figure 6. The climatology differences of seasonal precipitation (mm d^{-1}) between the periods 2000–2018 and 1900–1999 for December–January–February (DJF), March–April–May (MAM), June–July–August (JJA), and September–October–November (SON) in (a)–(d) the GPCC observations, (e)–(h) the CAM5 AMIP ensemble mean simulations, and the comparison between the GPCC (red), the CAM5 ensemble mean (black), and the 40 CAM5 individual ensemble members (gray) for 3 month running mean precipitation averaged over (i) the western US and (j) southeastern US throughout the seasonal cycle. Here the western US and southeastern US regions are respectively defined as 125 °W–90 °W, 25 °N–50 °N and 90 °W–67 °W, 25 °N–40 °N that have the USDM frequency of drought occurrence (figure 1(a)) exceed 50%; they are indicated using thick black contours in figure 1(a). The stippling in (a)–(h) indicates significance at the 10% level based on a t test.

SST changes. In particular, the observed substantial precipitation increases in the southern-central and eastern US during fall are likely due to the recently intensified western North Atlantic Subtropical High, which enhanced moisture transports into the region and led to substantial increases in intensity of extreme precipitation events (Bishop *et al* 2019). Focusing on the western US and southeastern US (figures 6(i) and (j)), throughout the seasonal cycle, the CAM5 ensemble mean in general agrees with the GPCC for all seasons except in late fall when the GPCC shows a modest wet anomaly whereas the ensemble mean displays a dry anomaly. The GPCC nevertheless falls within the CAM5 ensemble spread, suggesting that the CAM5 does a reasonable job in capturing the observed precipitation changes. The above further substantiates that a considerable portion of the observed precipitation deficits in the western and southeastern US (figure 4(a)) are driven by the global SST changes, the accumulated effects of which subsequently leads to considerably drier soil (figure 5(c)) and more frequent drought occurrence (figures 5(d) and (e)) in these regions. The CAM5-based results on the role of SST for the mean changes of precipitation and temperature are broadly supported by the GFDL AM3 and NASA GEOS-5 (cf figures S3 and S4 with figures 5 and 6).

4. Conclusions

This study performed two related but distinct investigations. It systematically investigated potential caveats in using the USDM to evaluate LSM simulations of drought, and in the context of the evaluation, looked into the causes of high USDM drought occurrence in the western and southeastern US. The NWM v2.0 retrospective simulation (1993–2018) and NLDAS-2 were used to exemplify the evaluation.

The evaluation shows that the LSM simulations strongly underestimate the USDM drought occurrence in the western and southeastern US, where the probability of detection is less than 0.5. Much of the LSM underestimation, however, is not attributed to LSM deficiencies. Rather, it occurs because such evaluation does not represent a fair comparison. Specifically, the LSMs and USDM utilize different base periods and drought indicators to identify and quantify drought. The USDM integrates a number of drought indicators and indices that have long periods of record dating back to the early 20th century or earlier, which enables it to capture both short-term drought and long-term drought. By comparison, the relatively short base period (1993–2018) of the NWM v2.0 and NLDAS-2 allows them to adequately detect short-term droughts but not long-term ones. This leads to the strong LSM underestimation of drought occurrences in the western US and southeastern US where long-term droughts are abundant. Moreover, the USDM integrates numerous

drought indicators and indices that represent various drought types whereas the LSMs in this study utilize a single index—SMP—and focus on agricultural drought, further adding to the inconsistency. The effects of the caveats can be substantial in that the differences between the USDM and LSMs can be considerably larger than the inter-LSM differences. This study stresses the importance of considering the above potential caveats when evaluating LSMs using the USDM, and suggests that the evaluation should stay qualitative when the aforementioned caveats are considerable. The study also suggests that when using the USDM to quantitatively evaluate LSM simulations of relatively short lengths, one may consider focusing on components that both data have, e.g. short-term drought.

In the above context, we further investigated why the USDM has more drought occurrence in the western and southeastern US than elsewhere. Such spatial distribution of drought occurrence is found to result from the post-2000 precipitation changes relative to the 20th century, which consist of reductions in the western and southeastern US and increases elsewhere. The elevated surface temperatures played a secondary contributing role. A considerable portion of these post-2000 precipitation and temperature changes are driven by the concurrent global SST changes contributed by the global warming trend, negative PDO and positive AMO, all of which are the well-known leading patterns of global SST variability that can induce drought conditions in the western, central, and southeastern US. It can be inferred that when the PDO and AMO reverse their phases in the future, they can act to offset the warming and drying effects of the long-term warming trend, reducing drought occurrence in the western, central and southeastern US.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

ORCID iD

Hailan Wang ⬥ https://orcid.org/0000-0003-1041-0233

References

Andreadis K M, Clark E A, Wood A W, Hamlet A F and Lettenmaier D P 2005 Twentieth-century drought in the conterminous United States *J. Hydrometeorol.* **6** 985–1001

Beck H E *et al* 2021 Evaluation of 18 satellite- and model-based soil moisture products using *in situ* measurements from 826 sensors *Hydrol. Earth Syst. Sci.* **25** 17–40

Berg N and Hall A 2017 Anthropogenic warming impacts on California snowpack during drought *Geophys. Res. Lett.* **44** 2511–8

Bishop D A, Williams A P, Seager R, Fiore A M, Cook B I, Mankin J S, Singh D, Smerdon J E and Rao M P 2019 Investigating the causes of increased twentieth-century fall precipitation over the southeastern United States *J. Clim.* **32** 575–90

Chen L G, Gottschalck J, Hartman A, Miskus D, Tinker R and Artusa A 2019 Flash drought characteristics based on US drought monitor *Atmosphere* **10** 498

Ford T W and Quiring S M 2019 Comparison of contemporary *in situ*, model, and satellite remote sensing soil moisture with a focus on drought monitoring *Water Resour. Res.* **55** 1565–82

Huang B, Thorne P W, Banzon V F, Boyer T, Chepurin G, Lawrimore J H, Menne M J, Smith T M, Vose R S and Zhang H-M 2017 *Extended Reconstructed Sea Surface Temperature, Version 5 (ERSSTv5): Upgrades, Validations, and Intercomparisons* **30** 8179–205

Koster R, Schubert S, Wang H, Mahanama S and DeAngelis A 2019 Flash drought as captured by reanalysis data: disentangling the contributions of precipitation deficit and excess evapotranspiration *J. Hydrometeorol.* **20** 1241–58

Kushnir Y, Seager R, Ting M, Naik N and Nakamura J 2010 Mechanisms of tropical Atlantic SST influence on North American precipitation variability *J. Clim.* **23** 5610–28

Lenssen N, Schmidt G, Hansen J, Menne M, Persin A, Ruedy R and Zyss D 2019 Improvements in the GISTEMP uncertainty model *J. Geophys. Res. Atmos.* **124** 6307–26

Livneh B and Hoerling M P 2016 The physics of drought in the US central great plains *J. Clim.* **29** 6783–804

Livneh B, Rosenberg E A, Lin C, Nijssen B, Mishra V, Andreadis K M, Maurer E P and Lettenmaier D P 2013 A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: update and extensions *J. Clim.* **26** 9384–92

Luo L, Apps D, Arcand S, Xu H, Pan M and Hoerling M 2017 Contribution of temperature and precipitation anomalies to the California drought during 2012–2015 *Geophys. Res. Lett.* **44** 3184–92

Mo K C, Schemm J-K E and Yoo S-H 2009 Influence of ENSO and the Atlantic multidecadal oscillation on drought over the United States *J. Clim.* **22** 5962–82

Mocko D M, Kumar S V, Peters-Lidard C D and Wang S 2021 Assimilation of vegetation conditions improves the representation of drought over agricultural areas *J. Hydrometeorol.* **22** 1085–98

Murray D *et al* 2020 Facility for weather and climate assessments (FACTS): a community resource for assessing weather and climate variability *Bull. Am. Meteorol. Soc.* **101** E1214–24

Newman M *et al* 2016 The Pacific decadal oscillation, revisited *J. Clim.* **29** 4399–427

Schneider U, Becker A, Finger P, Meyer-Christoffer A, Rudolf B and Ziese M 2011 GPCC full data reanalysis version 6.0 at 0.5°: monthly land-surface precipitation from rain-gauges built on GTS-based and historic data (https://doi.org/10.5676/DWD_GPCC/FD_M_V7_050)

Schubert S D *et al* 2009 A USCLIVAR project to assess and compare the responses of global climate models to drought-related SST forcing patterns: overview and results *J. Clim.* **22** 5251–72

Sheffield J and Wood E F 2008 Global trends and variability in soil moisture and drought characteristics, 1950–2000, from observation-driven simulations of the terrestrial hydrologic cycle *J. Clim.* **21** 432–58

Su L, Cao Q, Xiao M, Mocko D M, Barlage M, Li D, Peters-Lidard C D and Lettenmaier D P 2021 Drought variability over the conterminous United States for the past century *J. Hydrometeorol.* **22** 1153–68

USGCRP 2018 *Impacts, Risks, and Adaptation in the United States: Fourth National Climate Assessment* vol II, ed D R Reidmiller, C W Avery, D R Easterling, K E Kunkel, K L M Lewis, T K Maycock and B C Stewart (Washington, DC: US Global Change Research Program) p 1515

Wang H, Schubert S D, Suarez M J and Koster R D 2010 The physical mechanisms by which the leading patterns of SST variability impact US precipitation *J. Clim.* **23** 1815–36

Williams A P, Cook E R, Smerdon J E, Cook B I, Abatzoglou J T, Bolles K, Baek S H, Badger A M and Livneh B 2020 Large contribution from anthropogenic warming to an emerging North American megadrought *Science* **368** 314–8

Williams A P, Seager R, Abatzoglou J T, Cook B I, Smerdon J E and Cook E R 2015 Contribution of anthropogenic warming to California drought during 2012–2014 *Geophys. Res. Lett.* **42** 6819–28

Wood E F, Schubert S D, Wood A W, Peters-Lidard C D, Mo K C, Mariotti A and Pulwarty R S 2015 Prospects for advancing drought understanding, monitoring, and prediction *J. Hydrometeorol.* **16** 1636–57

Xia Y *et al* 2012b Continental-scale water and energy flux analysis and validation for North American land data assimilation system project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow *J. Geophys. Res. Atmos.* **117** D03110

Xia Y *et al* 2012a 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. Atmos.* **117** D03109

Xiao M, Udall B and Lettenmaier D 2018 On the causes of declining Colorado River streamflows *Water Resour. Res.* **54** 6739–56