

## Paleoclimate Constraints on the Spatiotemporal Character of Past and Future Droughts

S. COATS,<sup>a,b,c</sup> J. E. SMERDON,<sup>d</sup> S. STEVENSON,<sup>e</sup> J. T. FASULLO,<sup>b</sup> B. OTTO-BLIESNER,<sup>b</sup> AND T. R. AULT<sup>f</sup>

<sup>a</sup> *Department of Earth Sciences, University of Hawai'i at Mānoa, Honolulu, Hawaii;* <sup>b</sup> *Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, Boulder, Colorado;* <sup>c</sup> *Woods Hole Oceanographic Institution, Woods Hole, Massachusetts;* <sup>d</sup> *Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York;* <sup>e</sup> *Bren School of Environmental Science and Management, University of California, Santa Barbara, Santa Barbara, California;* <sup>f</sup> *Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, New York*

(Manuscript received 3 January 2020, in final form 12 July 2020)

**ABSTRACT:** Machine-learning-based methods that identify drought in three-dimensional space–time are applied to climate model simulations and tree-ring-based reconstructions of hydroclimate over the Northern Hemisphere extratropics for the past 1000 years, as well as twenty-first-century projections. Analyzing reconstructed and simulated drought in this context provides a paleoclimate constraint on the spatiotemporal characteristics of simulated droughts. Climate models project that there will be large increases in the persistence and severity of droughts over the coming century, but with little change in their spatial extent. Nevertheless, climate models exhibit biases in the spatiotemporal characteristics of persistent and severe droughts over parts of the Northern Hemisphere. We use the paleoclimate record and results from a linear inverse modeling-based framework to conclude that climate models underestimate the range of potential future hydroclimate states. Complicating this picture, however, are divergent changes in the characteristics of persistent and severe droughts when quantified using different hydroclimate metrics. Collectively our results imply that these divergent responses and the aforementioned biases must be better understood if we are to increase confidence in future hydroclimate projections. Importantly, the novel framework presented herein can be applied to other climate features to robustly describe their spatiotemporal characteristics and provide constraints on future changes to those characteristics.

**KEYWORDS:** Drought; Climate change; Paleoclimate; Climate models; Climate variability; Other artificial intelligence/machine learning

### 1. Introduction

Projecting regional hydroclimate over the coming century necessitates the use of coupled general circulation models (CGCMs). Such projections are of utmost importance to society's ability to plan for the risks associated with climate change. A central question is thus whether CGCMs provide useful information for this purpose, which requires that they realistically simulate responses to anthropogenic forcing and natural climate variability. While the need to accurately represent natural climate variability can be less clear it is nevertheless critical for regional hydroclimate projections in places like the Northern Hemisphere (NH) extratropics, where the magnitude of natural hydroclimate variability is large on the decadal-to-centennial time scales typical of future projections (Stine 1994; Cook et al. 2016a; see also below). Where natural variability is large, it has the potential to significantly mitigate or exacerbate the response to anthropogenic forcing, and therefore will have implications for societally relevant hydroclimatic risks. Importantly, the instrumental interval is too short to properly characterize the longest of these time scales of variability. A confident validation of the ability of CGCMs to project extratropical regional hydroclimate thus requires a longer record of climate against which to compare.

One such record comes from tree rings, which have been used to produce annually resolved and spatially complete (with

up to 0.5° spatial resolution) reconstructions of hydroclimate covering at least the last 1000 years (Cook et al. 2007, 2010, 2016a; E. Cook et al. 2015). Over the North American Southwest (NASW; 32°–40°N, 125°–105°W), tree-ring reconstructions suggest that decadal-scale droughts have occurred regularly over the last millennium (e.g., Cook et al. 2016a) and that there is centennial-scale variability in the characteristics of these features. For instance, the five most persistent and severe droughts of the last millennium all occurred during the Medieval Climate Anomaly (MCA; Coats et al. 2016a,b; Ault et al. 2018), a period in which the spatial extent of drought may have been greater (Cook et al. 2014b; Coats et al. 2015b) and the seasonality of drought may have been different (Coats et al. 2015c) compared to our present climate. Importantly, CGCMs can capture many of the aforementioned characteristics of these droughts, providing confidence that they can quantify the contribution of natural variability to future hydroclimate change (Coats et al. 2013, 2015a; Stevenson et al. 2015, 2018).

While persistent and severe droughts in the NASW in simulations of the last millennium appear to result predominantly from internal variability (Coats et al. 2013, 2015a; Stevenson et al. 2015, 2018), which may be consistent with the real world (Coats et al. 2016a,b; Ault et al. 2018; Steiger et al. 2019), the hydroclimate response to anthropogenic forcing in the NASW reduces the contribution of natural variability to future drought risk (Cook et al. 2014a; Deser et al. 2014; Ault et al. 2016; Lehner et al. 2018). Specifically, increases in evaporative demand drive hydroclimate conditions in the region to nearly permanent drought conditions, regardless of the large intermodel spread in

Corresponding author: S. Coats, scoats@hawaii.edu

DOI: 10.1175/JCLI-D-20-0004.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) ([www.ametsoc.org/PUBSReuseLicenses](http://www.ametsoc.org/PUBSReuseLicenses)).

projected precipitation changes in the region (Cook et al. 2015x; Ault et al. 2016). This change may result, at least in part, from the choice of hydroclimate metric used for analysis (Milly and Dunne 2016; Swann et al. 2016; Berg et al. 2017; Scheff et al. 2017; Mankin et al. 2017, 2018; Berg and Sheffield 2018; Yang et al. 2019) and may not occur globally, because evaporative demand increases are smaller or compensated by increased precipitation in other regions [e.g., Mongolia (Hessl et al. 2018) and parts of the Amazon basin (Duffy et al. 2015)]. The potential regionality and metric dependence of future drought risk suggests a need to expand upon previous analyses.

To do so, we use a set of tree-ring-based PDSI reconstructions covering 58% of land area in the NH extratropics at 2.5° resolution (Cook et al. 2007, 2010, 2016a; E. Cook et al. 2015), forced transient simulations of the last millennium from state-of-the-art CGCMs, and novel machine-learning-based techniques that allow for drought identification and characterization in three dimensions (latitude, longitude, and time; Fu et al. 2012), to analyze hydroclimate variability continuously over the last millennium and into the future, with a focus on persistent and severe droughts in the NH extratropics regardless of their location. While it is expected that for some regions and variables, future drought projections will be dominated by the response to anthropogenic forcing (e.g., the nearly permanent drought conditions in the NASW; B. Cook et al. 2015), for regions and variables with a weaker anthropogenically forced response, and before the full emergence of that response, accurately projecting future drought also requires that CGCMs simulate realistic natural variability. For instance, if a period like the MCA occurs in the future because of internal or naturally forced variability, drought risk will be much higher than if the climate system is on a different trajectory. Capturing the full range of potential outcomes is necessary to provide confident risk projections (e.g., Coats and Mankin 2016), and thus using CGCMs to project risk requires that they capture such variability.

Our analysis expands on previous research in three regards: 1) Droughts are analyzed over the majority of the NH extratropics; 2) the ability of CGCMs to simulate persistent and severe droughts is evaluated within the same methodological framework in which future projections are analyzed; and 3) the spatial and temporal characteristics of these droughts are *simultaneously* analyzed within one methodological framework. The fundamental questions that are addressed are these: 1) Can CGCMs simulate persistent and severe droughts across the NH extratropics that are consistent with the paleoclimate record? 2) If so, do CGCMs suggest that the characteristics of these droughts will change in the future? Critically, our novel and comprehensive approach to assessing future drought risk provides new insights into the physical mechanisms underlying persistent and severe droughts and the ability of state-of-the-art CGCMs to reproduce them.

## 2. Methods

### a. Paleoclimate record

We employ the North American Drought Atlas (NADA; Cook et al. 2007, 2016a), Monsoon Asia Drought Atlas (MADA; Cook et al. 2010; Baek et al. 2017), and Old World

Drought Atlas (OWDA; E. Cook et al. 2015), which collectively comprise tree-ring-based reconstructions of the June–August (JJA) average Palmer drought severity index (PDSI; Palmer 1965) on at least a 2.5° latitude–longitude grid for 58% of the terrestrial NH going back to at least 1000 CE. The drought atlases (e.g., the 0.5° resolution NADA) are regridded to a 1.9° × 2.5° latitude–longitude grid common to the model simulations from the National Center for Atmospheric Research (NCAR). Although the combined drought atlases are a reconstruction of hydroclimate with associated uncertainties, they have been demonstrated to robustly reflect patterns of hydroclimate variability on seasonal to multidecadal time scales and subregional to multiregional spatial scales (Baek et al. 2017), and thus provide a useful ground truth for comparison with CGCM simulations. It is important to note, however, that fewer tree-ring records are available back in time for all three Drought Atlas reconstructions. As such, individual grid point reconstructions will share a greater amount of the same spatial information back in time, particularly outside of western North America and the Mediterranean and prior to 1500 CE. While it is not possible to determine exactly how this uncertainty will propagate through the methodology employed herein, the spatial extent of reconstructed droughts may be greater during the early parts of the last millennium because of sampling density issues. Hereinafter the combined drought atlases will be referred to as the paleoclimate record.

### b. Model simulations

For comparison to the paleoclimate record, only CGCMs with continuous forced transient simulations spanning the last millennium (850–1849 CE), historical (1850–2005 CE), and future [2006–2100 CE, using representative concentration pathway (RCP) 8.5] intervals are analyzed. The output of the simulations must additionally include all variables necessary to calculate Penman–Monteith (Penman 1948) potential evapotranspiration (PET) for use in the PDSI calculations. Only six simulations, performed with three unique CGCMs, meet these analysis requirements: four from the NCAR Community Earth System Model 1 (CESM1) Last Millennium Ensemble (LME) project (hereinafter CESM; Otto-Bliesner et al. 2016), one from the NCAR Community Climate System Model 4 (CCSM4, hereinafter simply CCSM; Landrum et al. 2013), and one from the Institute Pierre Simon Laplace (IPSL) CM5A low-resolution (LR) Earth system model (hereinafter IPSL). While there are 13 CESM LME simulations available, only the four that are run continuously for the last millennium through the end of the twenty-first century are used herein. The CCSM and IPSL simulations are from phase 3 of the Paleoclimate Modeling Intercomparison Project (PMIP3), a part of phase 5 of the Coupled Model Intercomparison Project (CMIP5; Taylor et al. 2012). For the future projections, 20 additional simulations from CMIP5 CGCMs that did not produce last millennium simulations meeting our analysis requirements are analyzed using the historical and future (RCP8.5) intervals (Table 1). The RCP8.5 emissions scenario was chosen for the future projections to maximize the climate change signal. All model outputs are monthly and regridded to a 1.9° × 2.5° latitude–longitude grid common to the model simulations from the NCAR.

TABLE 1. CGCMs and simulations analyzed herein.

CGCM	Historical/future	Last millennium
ACCESS1.0	×1	—
CCSM4	×6	×1
CNRM-CM5	×3	—
CanESM2	×5	—
GISS-E2-R	×1	—
IPSL-CM5A-LR	×1	×1
MIROC-ESM-CHEM	×1	—
MIROC-ESM	×1	—
NorESM1-M	×1	—
NorESM1-ME	×1	—
INM-CM4	×1	—
CESM1	×4	×4

For all simulations, JJA average PDSI is calculated following Cook et al. (2014a) using the 1931–90 CE period as the standardization interval (as was used in the NADA). While JJA average PDSI is technically a metric of NH summer drought, PDSI has roughly annual persistence and thus reflects an integrated signal of drought across seasons. The JJA PDSI should thus be interpreted as characteristic of interannual and longer time scale drought and the same analyses computed with annually averaged PDSI produce similar results (not shown). The integrated nature of JJA PDSI is also why it is compared to annual averages of other hydroclimate variables (see below). Hereinafter, use of the term PDSI will refer to JJA average PDSI. The PET used to calculate PDSI comes from the Penman–Monteith formulation for a reference crop with constant stomatal conductance. While it is clear that neglecting the impact of anthropogenic greenhouse gasses on stomatal conductance can make offline metrics like PDSI unreliable when estimating the characteristics of future drought (Milly and Dunne 2016; Swann et al. 2016; Berg et al. 2017; Scheff et al. 2017; Mankin et al. 2017, 2018; Berg and Sheffield 2018; Yang et al. 2019), the use of this metric is necessary here as it allows for a direct comparison of the CGCMs with the paleoclimate record. Nevertheless, annually averaged soil moisture (SM) over the top 30 cm and 2 m of the soil column and precipitation minus evaporation ( $P - E$ ) are additionally analyzed to assess the consistency of results using offline metrics, such as PDSI, with those using online metrics (i.e., prognostic model variables like SM, precipitation, and evaporation). To provide a climate dynamical understanding of biases that impact the simulation of hydroclimate, surface temperature for the three CGCMs that ran simulations of the last millennium is used to calculate indices of the Atlantic multidecadal oscillation [AMO; following Trenberth and Shea (2006)], the Pacific decadal oscillation [PDO; following Mantua et al. (1997)], and El Niño–Southern Oscillation (ENSO; using the Niño-3.4 index; surface temperature average over 5°S–5°N, 120°–170°W).

### c. Drought identification algorithm

Droughts are identified in space–time using a method based on Markov random fields (MRFs). In climate science the Markov property is typically considered between adjacent time

steps at a single spatial location. In this case, the MRF means that the Markov property is considered between adjacent time steps and adjacent spatial locations (Fu et al. 2012). The drought identification problem is formulated as finding the most likely configuration of a binary MRF, where a space–time grid point is given a value of one if it is in a drought state (the converse, or normal state, is assigned a value of zero). Superficially, this characterization can be accomplished by assigning a space–time grid point with a value of PDSI lower (drier) than a specified drought threshold to a drought state independently of the characteristics of all other grid points. However, such an approach would require additional methods to identify unique drought features—disparate drought features are likely to be connected by randomly (i.e., not climatically connected) dry space–time grid points. The MRF-based drought identification leverages the known spatiotemporal consistency of drought (Herrera-Estrada et al. 2017) by considering a drought threshold while simultaneously encouraging neighboring grid points in both space and time to take the same state. This maximum a posteriori (MAP) inference results in identified droughts that are contiguous and distinct “clouds” of ones that traverse space–time.

The MAP inference is equivalent to the following:

$$x^* \in \operatorname{argmax}_{x \in \{0,1\}} \left\{ \sum_{u \in E} f_u(x_u) + \sum_{(u,v) \in V} f_{uv}(x_u, x_v) \right\},$$

where  $u \in E$  are all space–time grid points, and  $(u, v) \in V$  are all pairs of neighboring grid points in space. For a given year  $t$  at a grid point  $u$ , there are three potential functions. The first takes the form of the log-likelihood of two Gaussian distributions, one for if a space–time grid point is assigned to a drought state and one for a normal state (see example in Fig. 2b):

$$f_u^t(x_u^t = 1) = \log N(y_u^t | \mu_u^{<\text{threshold}}, \sigma_u^2),$$

$$f_u^t(x_u^t = 0) = \log N(y_u^t | \mu_u^{>\text{threshold}}, \sigma_u^2),$$

where  $\mu_u^{<\text{threshold}}$  is the mean PDSI of all years with PDSI less than one standard deviation below the mean of PDSI at grid point  $u$  between 1850 and 1929 CE,  $\mu_u^{>\text{threshold}}$  is the mean PDSI for all other years, and  $\sigma_u^2$  is the standard deviation of PDSI at grid point  $u$  between 1850 and 1929 CE. This potential function encourages space–time grid points with anomalously dry conditions to be assigned to a drought state.

The second and third potential functions are of nearly the same form:

$$f_u^{t-1,t}(x_u^{t-1}, x_u^t) = \{C_2 > 0, \quad \text{if } x_u^{t-1} = x_u^t,$$

which encourages the same spatial grid point in subsequent years to take the same state, and

$$f_{uv}^t(x_u^t, x_v^t) = \{C_1 > 0, \quad \text{if } x_u^t = x_v^t,$$

which encourages neighboring grid points in space to take the same value in the same year. The parameters  $C_2$  and  $C_1$  are user-defined, with larger values encouraging greater temporal and spatial consistency, respectively. The neighborhood

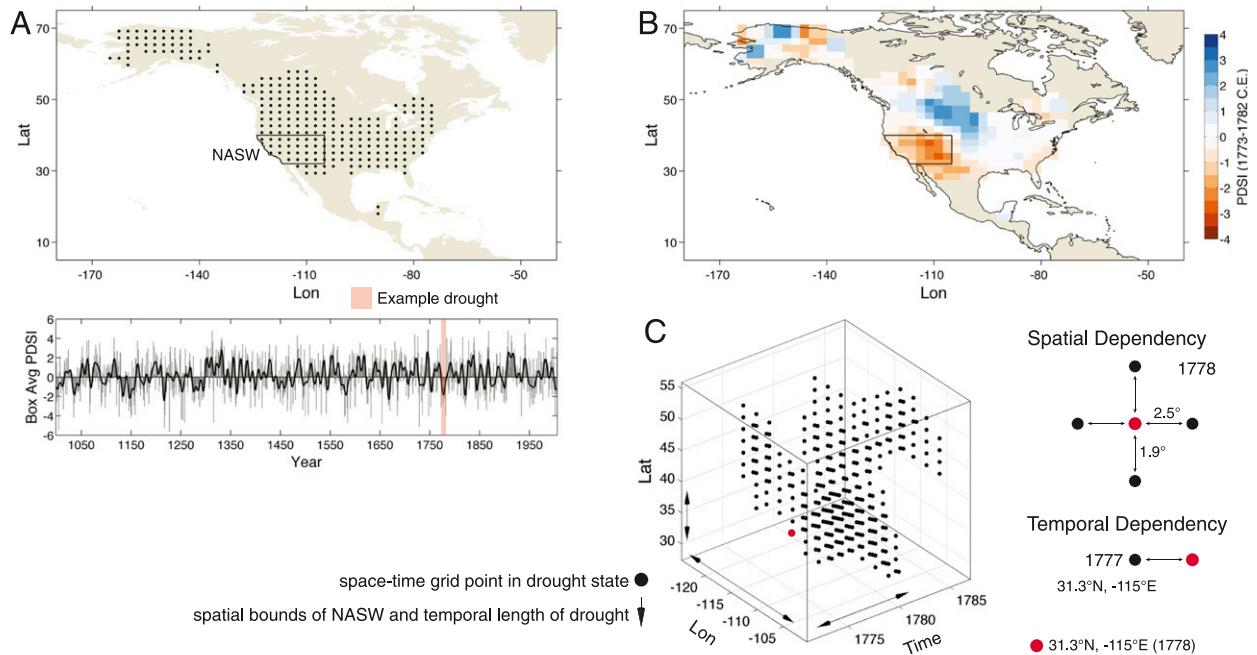


FIG. 1. (a) An example of a traditional approach (and statistical technique) for analyzing drought. (top) The spatial grid corresponding to the NADA is shown with the NASW ( $32^{\circ}$  to  $40^{\circ}$ N,  $-125^{\circ}$  to  $-105^{\circ}$ E) boundaries superimposed. (bottom) The average of reconstructed PDSI over the NASW is shown as the time series with an example drought identified by the orange shading (interannual variability are the gray bars and the 10-yr low-pass filtered values, using a 10-point Butterworth filter, are shown as the thick black line). (b) A composite of reconstructed PDSI during the example drought, a traditional approach to analyzing the spatial characteristics of drought. (c) Three-dimensional grid showing only those grid points assigned a drought state by the drought identification algorithm (section 2)—this drought will also be used in Figs. 2 and 5. This grid fully encompasses the example drought and the NASW region (arrows) but is larger in both space and time. The dependency structure (neighborhood structure) of the MRF from the drought identification algorithm is shown for a randomly chosen space–time grid point in the example drought (highlighted in red).

structure in Fig. 1c shows the space and time adjacency (dependency) for the MRF.

The term  $x^*$  is the assignment (ones and zeros assigned to space–time grid points) that maximizes the sum of these three potential functions across all space–time grid points. Figure 2 provides an example of the impact of the first potential function on this sum for a randomly chosen space–time grid point in the example domain (Fig. 1c). The functional form of the two log-likelihood equations is such that there is a linearly increasing “penalty” (the black line, or difference between the equations) for assigning a space–time grid point drier than the drought threshold to a normal state. The drought threshold herein is one standard deviation below the mean PDSI at a spatial grid point between 1850 and 1929 CE. At the randomly chosen space–time grid point ( $38.8^{\circ}$ N,  $-115^{\circ}$ E in 1778 CE), the difference in the log-likelihood equations is  $-2.0$ , a large penalty because its  $-3.52$  PDSI is well below the drought threshold ( $-2.2$  PDSI). Nevertheless, consider the possibility that the four neighboring spatial grid points (those to the north, south, east, and west; Fig. 1c) and the same spatial grid point in the previous year (1777 CE) are anomalously wet (positive PDSI). In that case, there would still be a high probability that the space–time grid point will be assigned to a normal state despite the large penalty. Specifically, if the space–time grid point and its neighbors are all assigned to a normal state then

the sum of the potential functions increases by 4 times  $C_1$  (as there are four spatial neighbors all in a normal state) plus  $C_2$  (as the previous year was assigned a normal state). Reasonable values of these parameters ( $C_2$  and  $C_1$ ) range from approximately 0.3 to 1.5 and thus this assignment would produce a large increase in the sum of the potential functions.

The MAP estimation problem involves balancing these local dependencies, as quantified by the potential functions, but on the scale of the full dataset. Solving such a problem explicitly is not possible because it would require  $2^N$  distinct calculations, where  $N$  is the number of space–time grid points, in order to compare the sum of the potential functions for every possible combination of ones and zeros in the space–time grid. Proximal methods are thus implemented using the algorithm of Fu et al. (2012), who have shown that MRF methods with a linear programming relaxation, where the binary classification requirement is relaxed to allow for the assignment of fractional values between 0 and 1, are capable of efficiently identifying major droughts over the instrumental interval. The parameters used herein exactly follow Fu et al. (2012) with the exception of  $C_1 = 0.5$  and  $C_2 = 0.5$ , which were chosen using sensitivity tests based on the ability of the algorithm to identify known droughts over the instrumental interval. In particular, a smaller value of  $C_1$  was required because Fu et al. (2012) identified droughts using precipitation, which has less temporal

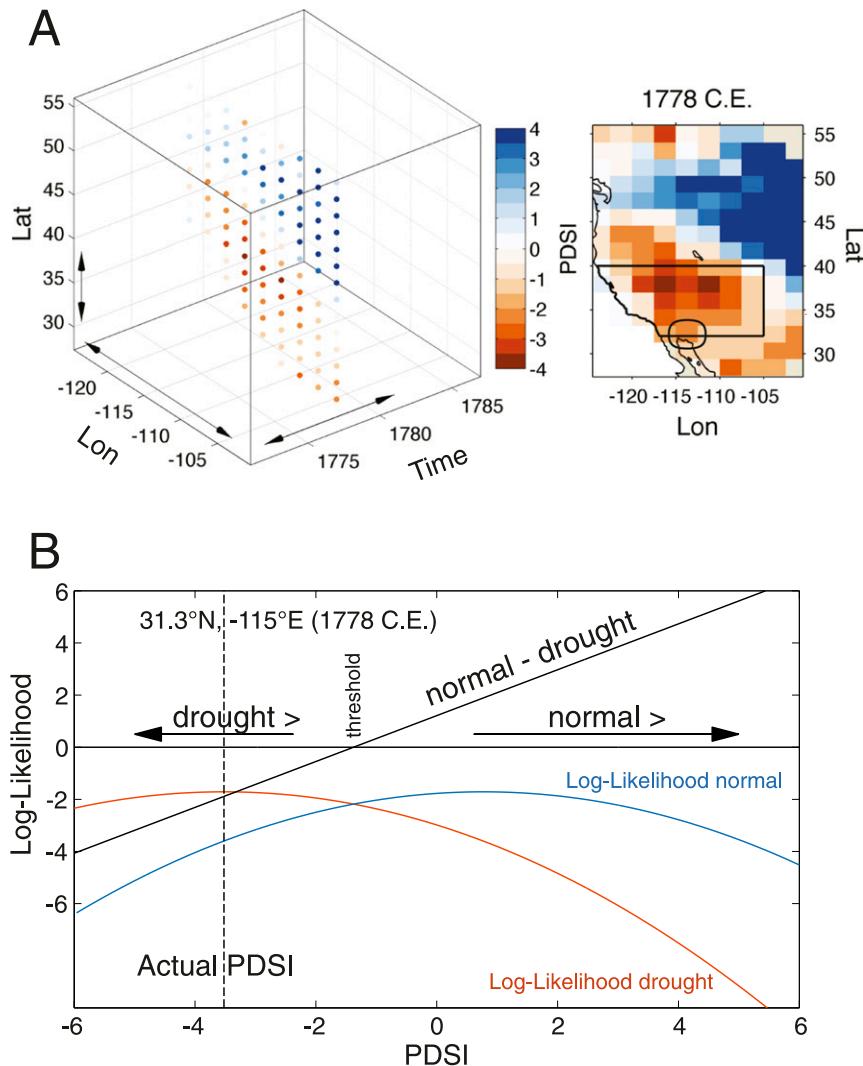


FIG. 2. (a) A (left) three-dimensional and (right) two-dimensional view of the year 1778 CE with colors indicating the PDSI values at each spatial grid point. (b) A visualization of how the first potential function encourages space–time grid points with more negative PDSI values (anomalously dry conditions) to be assigned a drought state. The functional form of the two log-likelihood equations for the circled grid point [from the year 1778 CE in (a), with the location being  $31.3^{\circ}\text{N}$ ,  $-115^{\circ}\text{E}$ ] are shown with the difference in these plotted as the solid black line. For PDSI values more negative than the zero crossing of the solid black line the potential function encourages the spatial grid point in a given year to be assigned a drought state. The actual value of PDSI for the year 1778 CE at this spatial grid point is shown as the dashed line with the much larger value of the log-likelihood of drought than normal suggesting that the potential function will strongly encourage the assignment of a drought state.

persistence than PDSI. With regard to the linear programming methodology, the Bregman divergence is set as the KL divergence, alpha is set to 1, beta is set to 2, and 500 is set as the maximum number of iterations (Fu et al. 2012).

As was noted above, the drought threshold value of PDSI is one standard deviation below the mean PDSI at each spatial grid point between 1850 and 1929 CE. This time interval was chosen as it is common to all CGCM simulations and the paleoclimate record and should not contain a large anthropogenic

signal. The latter point is important because emergence of an anthropogenic signal during this interval would change the mean of PDSI and thus inflate or deflate the drought threshold. This has the potential to bias the results, in particular, by altering the drought identification such that drought characteristics over the last millennium are a function of the anthropogenic signal, and thus not purely natural. Figure 3 demonstrates that the 1850–1929 CE interval does not contain a large anthropogenic signal. Specifically, between 1850 and 1929 CE the 81-yr trend

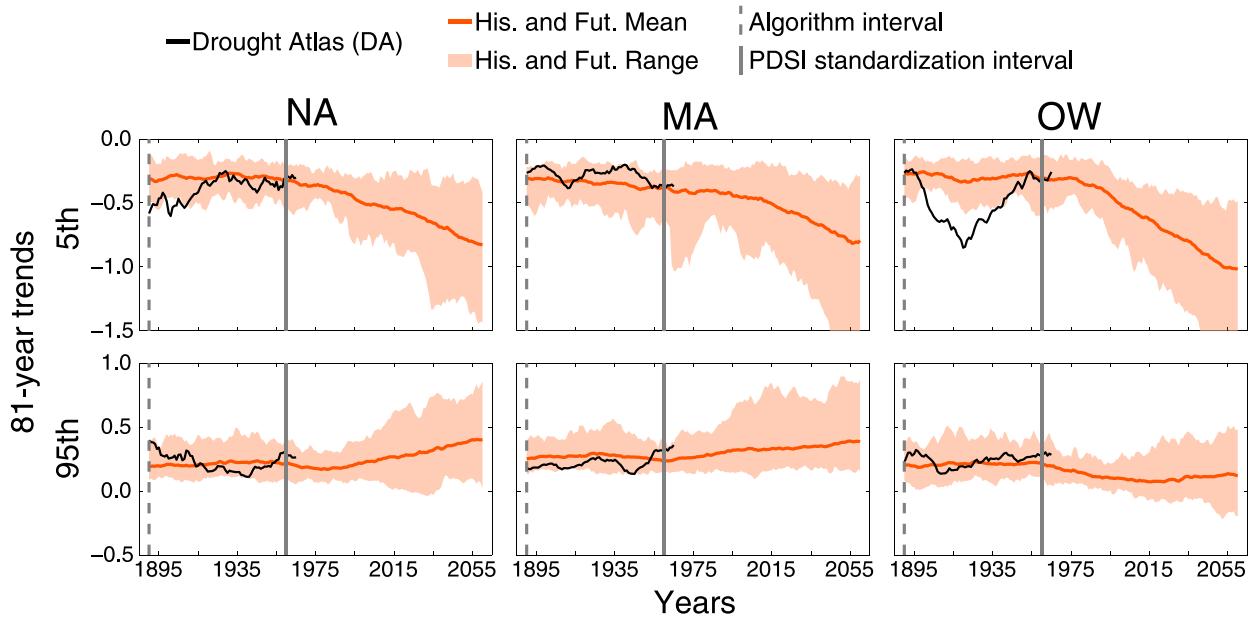


FIG. 3. The (top) 5th and (bottom) 95th percentiles of spatial grid point trends in each region for a sliding 81-yr interval (length of the standardization interval for PDSI and the interval used to define the drought threshold PDSI value for the drought identification algorithm). For instance, the values at the year 1895 represent the trends for the period 1855–1935 CE. The timing of the actual standardization interval for PDSI and the interval used to define the drought threshold PDSI value for the drought identification algorithm are plotted as black and gray dashed lines, respectively. Values are plotted for both the paleoclimate record and the mean and range of the 20 CMIP5 and 6 PMIP3 last millennium simulations. All trends are estimated from a linear least squares fit.

in PDSI at each spatial grid point was calculated and the 5th and 95th percentile of the distribution of these values is plotted in Fig. 3 for each interval and region. The largest (95th percentile) and smallest (5th percentile) values are approximately  $\pm 0.5$  PDSI, and these values are stationary through time (the lines are flat) until the intervals centered after 1990 CE. This suggests that any interval centered before 1990 CE is not at risk of having a large anthropogenic signal (as would be evidenced by systematic trends). More generally, results are not highly sensitive to the choice of this interval or any of the aforementioned parameters.

#### d. Drought definition and characteristics

The drought identification algorithm is applied between 1000 and 2000 CE for the paleoclimate record, between 1000 and 2100 CE for the CGCMs with simulations of the last millennium, and between 1900 and 2100 CE for the additional CMIP5 simulations. While the CGCM simulations of the last millennium actually extend to 850 CE, the analyses are limited to 1000 CE because of the temporal limitations of the paleoclimate record (the MADA and OWDA only extend to 1000 CE). Droughts in each of the individual drought atlas regions [North America (NA), Old World (OW), and Monsoon Asia (MA)] are identified separately (the fixed spatial domains corresponding to each region are shown in Fig. 4a).

As noted in the introduction, to accurately project drought risk requires that CGCMs simulate these features as well as centennial-scale natural variability in their characteristics. Likewise, characterizing natural variability helps to provide

confidence that future drought characteristics (i.e., those over the next century) are distinct from those over the last millennium. Given this motivation, we require robust statistics of the characteristics of persistent and severe droughts and how these characteristics vary on centennial time scales. To do this, after drought identification we determine which droughts occur within each 100-yr interval. Specifically, we take the year associated with all space–time grid points in each identified drought and calculate the median of these values. If the median year falls within the start and end year of a 100-yr interval, then the drought occurs within that interval. The 10 most severe droughts within each 100-yr interval are chosen for further analysis, with severity defined as the sum of PDSI over all space–time grid points in each identified drought (Fig. 5a). The identification of 10 droughts for each 100-yr interval allows for an assessment of the statistics of their characteristics, preventing outliers from dominating the results.

Four additional metrics are defined to describe the total spatiotemporal extent and location of each drought. The first three metrics are termed spatiotemporal extent, spatial extent, and temporal extent and the definitions are outlined in Figs. 5a–c. Spatiotemporal extent is the total number of space–time grid points in each identified drought (Fig. 5a), spatial extent is the maximum number of spatial grid points in any year of the drought (Fig. 5b), and temporal extent is the maximum number of years that any spatial grid point is in a drought state during the drought (Fig. 5c). Drought location is defined as the area-weighted average of the spatial locations (latitude and longitude) of the space–time grid points assigned to a drought state in each identified drought. Hereinafter the

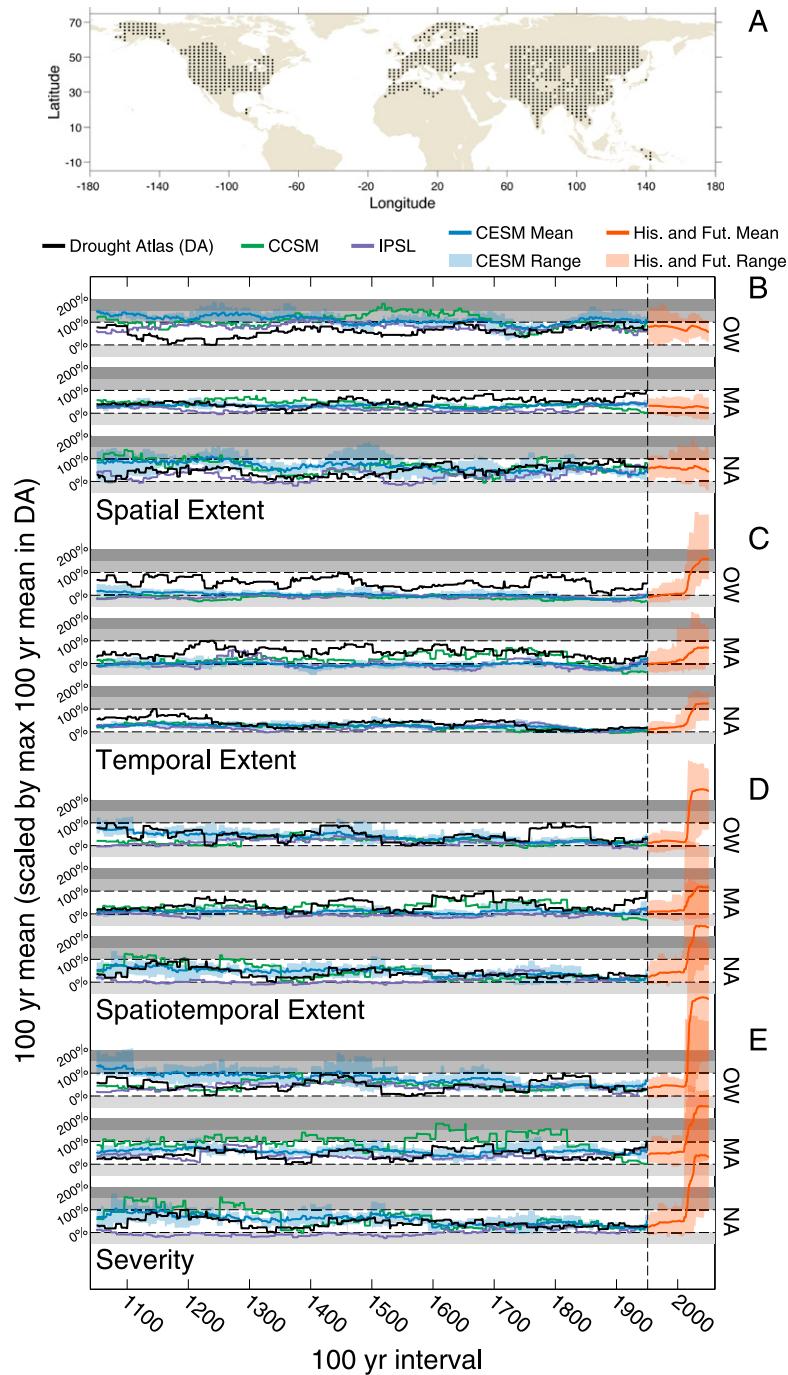


FIG. 4. (a) Fixed grid corresponding to the NA, MA, and OW, with each black dot being a spatial grid point. This fixed grid is used for both the paleoclimate record and CGCM simulations in all analyses. Mean (b) spatial extent, (c) temporal extent, (d) spatiotemporal extent, and (e) severity of the top 10 most severe droughts in each region for each 100-yr interval. The dashed black lines represent the full range in these values for the paleoclimate record between 1000 and 2000 CE, with the solid black line representing the actual time histories. The gray shaded regions denote 50%–100%, 100%–150%, and 150%–200% of this full range. Each color represents a different last millennium simulation. For the four CSM last millennium simulations the blue line is the mean and the blue shaded region spans the maximum and minimum values. The orange line is the mean and the orange shaded region spans the maximum and minimum values of the 20 CMIP5 historical and 6 PMIP3 last millennium simulations from the historical interval into the future.

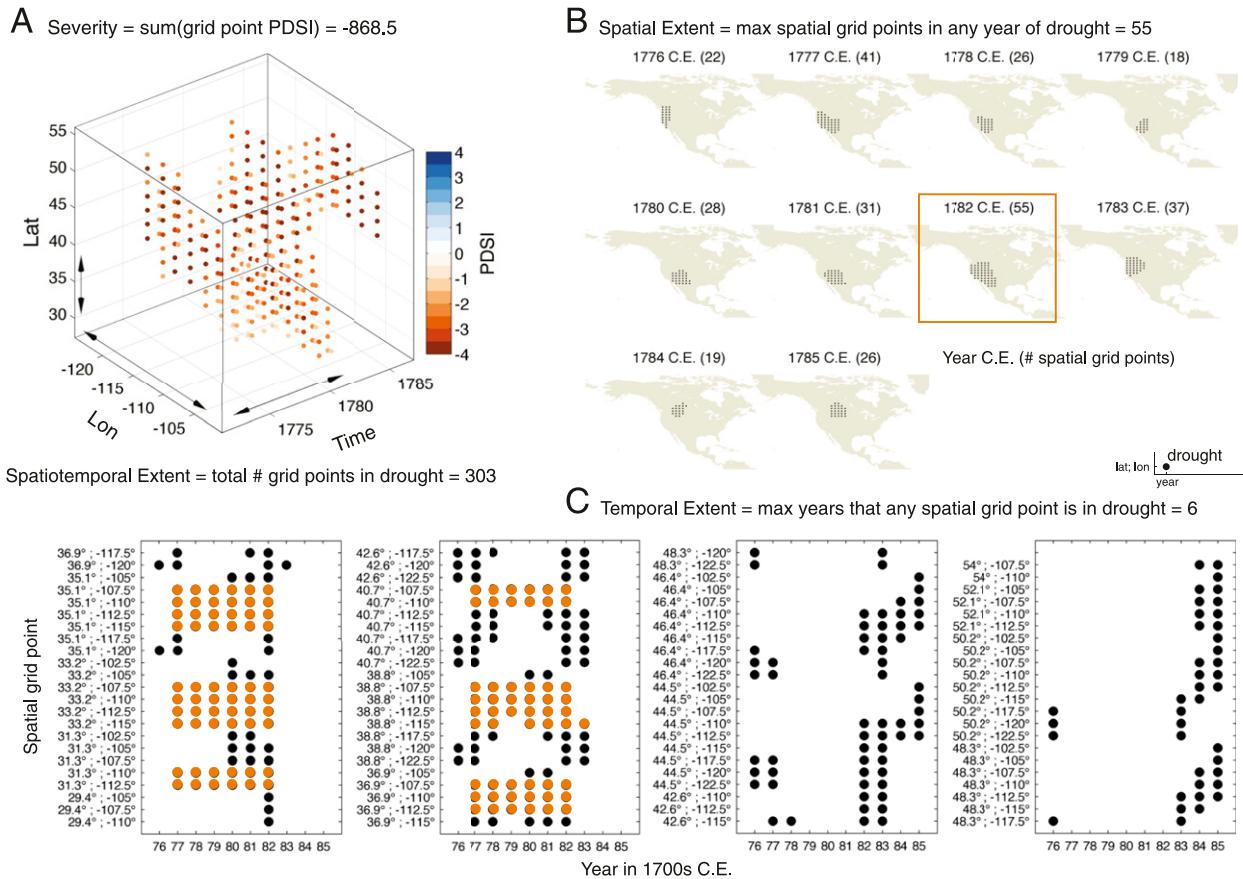


FIG. 5. (a) Space–time grid points assigned a drought state by the drought identification algorithm with each grid point colored corresponding to its PDSI value. The severity and spatiotemporal extent metrics and definitions are shown for the identified drought. (b) Spatial extent and definition using the identified drought as an example. The orange box highlights the year 1782 CE with 55 spatial grid points in drought conditions, which is the spatial extent value for the identified drought. (c) Temporal extent and definition using the identified drought as an example. The spatial grid points highlighted in orange are those with 6 years in a drought state, which is the temporal extent value for the identified drought.

four metrics will be referred to as the drought characteristics. Results are not highly sensitive to the choice of characteristics, drought definition (i.e., 10 most severe droughts for each 100-yr interval), or the length of the analyzed intervals (i.e. 50-yr vs 100-yr intervals).

#### e. Method validation

Figure 6 compares the composite of PDSI and the fraction of years at each spatial grid point assigned to a drought state by the drought identification algorithm for a randomly chosen drought among the 20 most severe in each of the three regions between 1000 and 2000 CE. The qualitative similarity between these two fields provides a validation that the drought identification algorithm successfully identifies persistent and severe droughts. This is the case for both the paleoclimate record and the CGCM simulations. The demonstrated similarities are not dependent on the selected droughts: the same qualitative similarity exists over many different random drought selections. This result is also consistent with previous research suggesting that the drought identification algorithm successfully identifies

the most prominent droughts in observations of precipitation (Fu et al. 2012), which is also true of PDSI over the instrumental interval (not shown). Collectively, these experiments suggest that the drought identification algorithm provides a consistent framework for identifying and analyzing persistent and severe droughts over the last millennium and into the future.

#### f. Linear inverse modeling

Future drought risk will be determined by the hydroclimate response to anthropogenic forcing and natural hydroclimate variability, with the characteristics of the variability itself potentially changing in response to anthropogenic forcing. Of particular interest is natural variability on the decadal-to-centennial time scales over which hydroclimate is typically projected by CGCMs. There are multiple approaches for testing whether CGCMs, because of their biases, are likely to over- or underestimate variability on these time scales (over- or underdispersive). One is to run large initial condition ensembles with a single CGCM (e.g., the NCAR CESM large

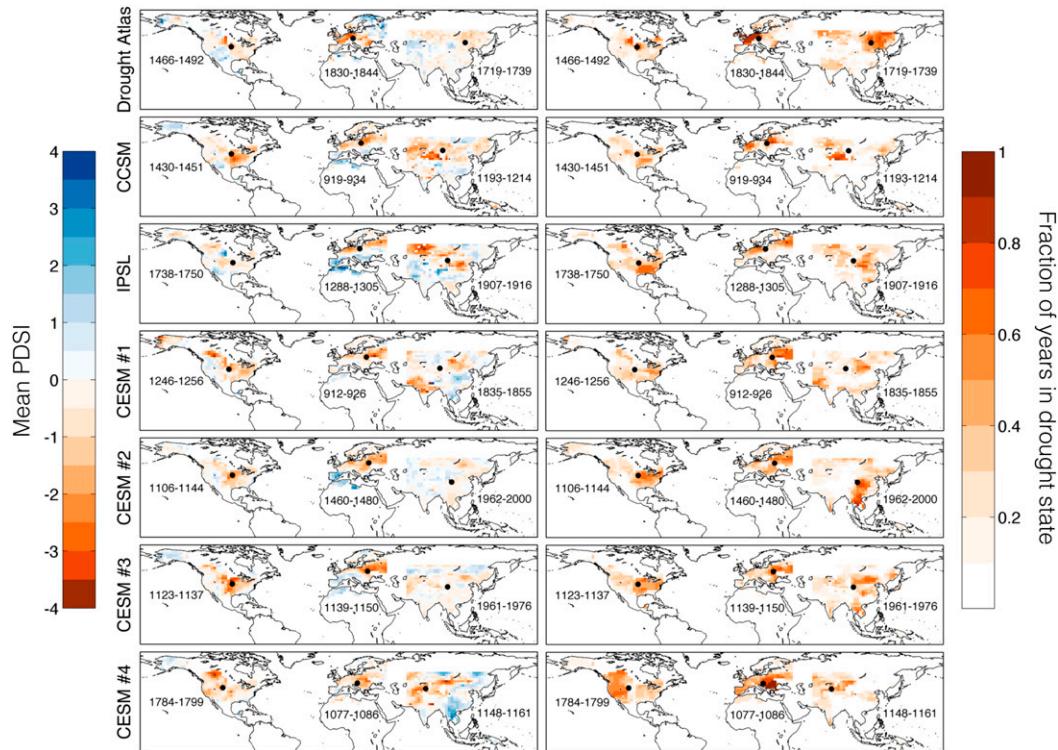


FIG. 6. For a randomly chosen drought among the 20 most severe in each of the three regions between 1000 and 2000 CE, the composite of PDSI and the fraction of years at each space–time grid point assigned to a drought state. This was done for the six last millennium simulations and the paleoclimate record. Inset years next to each region show the time period that is covered by the composite. The black dots mark the area-weighted average drought location of the randomly chosen drought for each region, defined as described in section 2c.

ensemble; Kay et al. 2015) to characterize the range of trajectories of the CGCM-simulated climate system for a single time interval (e.g., the instrumental interval). A second approach is to run a single CGCM for a long interval of time (e.g., the last millennium) such that there are sufficient samples by which to characterize longer time scales of variability [e.g., the PMIP3 last millennium ensemble simulations used herein and the preindustrial control simulations used in Karnauskas et al. (2012)]. Both approaches are limited by computational expense, with only a few modeling centers choosing to produce such simulations, and the first approach is additionally limited by the single observed-climate trajectory (the instrumental interval) against which to compare the CGCM simulations. For the modeling centers that did produce last millennium through future simulations we will use the second approach and compare the variability in the CGCMs to that of the paleoclimate record. Nevertheless, because of variable and simulation availability this only allows for the analysis of three CGCMs, two of which come from the same modeling center (NCAR). To test the additional CMIP CGCMs and associated simulations, we produce an ensemble of surrogate climate trajectories using a linear inverse model (LIM) following Ault et al. (2018), but applied to the CMIP5 CGCMs and including the full NA, OW, and MA regions. Analyzing the variability in these surrogate climate trajectories is equivalent to the first

approach, although limited in a number of ways relative to the use of actual simulations from CGCMs. Nevertheless, the same LIM-based techniques can be applied to observations to produce a range of climate trajectories consistent with the instrumental interval, which can then be compared to the CGCM LIM results. Importantly, none of these approaches test how CGCM biases impact the hydroclimate response to anthropogenic forcing—although these results can indirectly inform those impacts.

The full details of the LIM approach can be found in Ault et al. (2018), but we will briefly describe the method, with a focus on the unique details of the approach used herein. Using an LIM to produce surrogate climate trajectories is similar to modeling a time series as red noise to evaluate statistical significance against autocorrelation. An LIM, however, allows for the simultaneous treatment of both spatial and temporal autocorrelation and produces spatiotemporal outputs (in this case latitude, longitude, and time). Functionally an LIM has the form

$$\frac{d\mathbf{X}}{dt} = \mathbf{L}\mathbf{X} + \zeta,$$

where  $\mathbf{X}$  is the state vector,  $\mathbf{L}$  is the deterministic feedback matrix, and  $\zeta$  is the white noise forcing term. The inputs to an

LIM are temporally smoothed gridded spatiotemporal data, with the temporal smoothing required because the LIM is by construction linear. In this case, and more generally, the smoothing window is chosen to be three months, with the rationale that the climate system is approximately linear on greater than seasonal time scales up to the centennial time scales being analyzed herein.

To construct the LIM (i.e., to form  $\mathbf{X}$ ) we use linearly detrended monthly sea surface temperature (SST) anomalies globally (60°S–80°N), regridded to a common 5° × 5° latitude–longitude grid, and linearly detrended monthly PDSI over the NA, OW, and MA regions (see grids in Fig. 4) from 1951 to 2000 CE. This was done for all CGCMs (and each individual simulation) listed in Table 1, and for observations with the observation-based PDSI data from Sheffield et al. (2006) and SST data from Kaplan et al. (1998). As in Ault et al. (2018), principal components were used to decrease the dimensionality of the LIM inputs and thus to increase computational efficiency. The first 17 EOFs were retained from the monthly SST anomalies (a range of 58% to 76% variance explained across the datasets) and 22 EOFs from the monthly PDSI, with the EOFs calculated independently for each of the three regions (a range of 85% to 93%, 86% to 94%, and 75% to 83% variance explained across the datasets for NA, OW, and MA, respectively). As in Ault et al. (2018), the PDSI portion of  $\mathbf{X}$  was down-weighted by three orders of magnitude such that SSTs can impact PDSI in  $\mathbf{L}$  but PDSI cannot impact SSTs, which assumes that SST variability can drive hydroclimate variability but not the other way around. All additional choices used to construct the LIM are consistent with those in Ault et al. (2018); further details for calculating  $\mathbf{L}$  and integrating the LIM can be found therein. We run each LIM 100 times over 100 years and then analyze the range of possible 100-yr climate trajectories. Each iteration produces a spatiotemporal hydroclimate trajectory that is consistent with the stationary linear lagged covariance statistics of the simulated or observed climate system from 1951–2000 CE. Drought identification in, and analysis of, these surrogate climate trajectories follows the methods in the previous subsections.

### 3. Results

#### a. Drought characteristics in the paleoclimate record

To validate that CGCMs realistically simulate persistent and severe droughts, simulations are compared to the paleoclimate record over a common period (1000–2000 CE) with the expectation that there is broad consistency in both the average characteristics of these droughts and their natural variability on centennial time scales. Consistency in average drought characteristics is represented in Figs. 4b–e by the time histories of drought characteristics in CGCMs falling between the two dashed lines and particularly being centered between these lines, which represent the range of natural variability on centennial time scales in the paleoclimate record. Falling below (above) the bottom (top) line for a drought characteristic indicates that the CGCM has a smaller (larger) value of that drought characteristic than the paleoclimate record. Consistency in natural variability on centennial time scales is

represented by the CGCM simulations fully traversing the space between the two dashed lines (but not extending well beyond this space). If the actual time histories of drought characteristics in the simulations are in phase it could suggest that the variability on centennial time scales is externally forced, although it may also arise from randomness given the small number of analyzed models. If not, either internal variability is larger than the variability imposed by external forcing in the CGCMs, or the individual CGCMs are responding differently to external forcing. Additionally, if the time histories are consistent between the CGCMs and the paleoclimate record then it is possible that the real-world natural variability on centennial time scales is externally forced *and* the CGCMs are realistically responding to external forcing. Nevertheless, this consistency could also arise from randomness given the large number of analyzed years.

In the following subsections we describe the drought characteristics in the CGCMs and the paleoclimate record. Focus will be on the spatiotemporal extent and severity of droughts, with spatial extent and temporal extent being used to better understand the spatiotemporal extent results. For instance, droughts can be large because they are temporally persistent or spatially extensive, making it important to distinguish between these characteristics. The results in the following subsection are all based on results in Figs. 4b–e unless otherwise noted.

#### 1) DROUGHT SEVERITY

For drought severity there is broad consistency in the average drought characteristics and centennial-scale natural variability in these characteristics between the CGCMs and the paleoclimate record, although there are some differences across models and regions. The IPSL simulation in Fig. 4e, for instance, has droughts that are not as severe on average as the paleoclimate record over NA (purple line is largely over the dashed line representing the paleoclimatic minimum), with essentially no natural variability on centennial time scales. This is not the case, however, over the MA and OW regions where both average drought severity and its centennial-scale natural variability are more realistic, although the latter is still muted. Over all three regions, the CCSM and CESM simulations have average drought severity that is consistent with the paleoclimate record and large natural variability on centennial time scales. Drought severity is also largely consistent between the two CGCMs, which may not be surprising given that they are from the same NCAR lineage and driven by the same external forcing reconstructions.

Of particular interest is that some of the natural variability on centennial time scales is in phase across the CGCMs. For instance, all three CGCMs show high drought severity between 1400 and 1500 CE over OW, which is broadly consistent with the paleoclimate record (Fig. 4e). The two NCAR CGCMs, including all four individual CESM simulations, show a minimum in drought severity over NA at the end of the paleoclimate record (~1900 CE). Likewise, these CGCMs show a maximum in drought severity over NA at the beginning of the last millennium. While this could suggest a role for external forcing in driving drought severity during all three

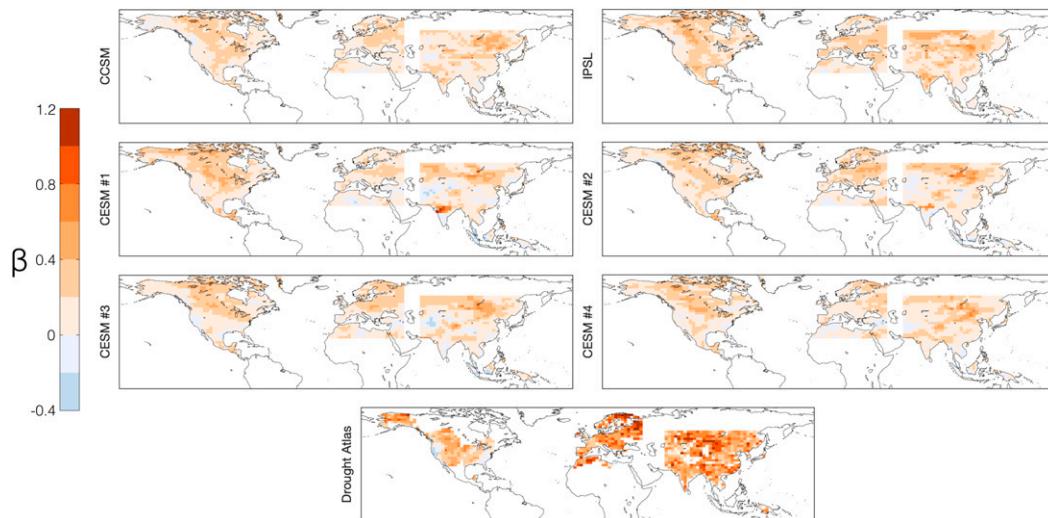


FIG. 7. The scaling exponent of spectral density ( $\beta$ ) calculated between 1000–1849 CE at each spatial grid point for the paleoclimate record and the six last millennium simulations. The value of  $\beta$  is calculated using linear least squares in log power–log frequency space on spectra estimated using Thomson’s multitaper method (Thomson 1982; see also section 2).

periods in the simulations, it may also arise simply from chance.

The decreasing trend in drought severity through the last millennium in the NCAR CGCMs in Fig. 4e is also broadly consistent with the paleoclimate record, where the time period between approximately 900–1300 CE has been termed the MCA, in part, because of a clustering of megadroughts over the NASW (e.g., Coats et al. 2016a,b; Ault et al. 2018; Steiger et al. 2019). Nevertheless, previous studies have found no preference for the simulation of megadroughts over the NASW during the MCA and little role for external forcing in driving NASW hydroclimate over the last millennium in CGCMs (Coats et al. 2013, 2015a; Stevenson et al. 2015), although recently a weak role for radiative forcing during the MCA has been identified by Steiger et al. (2019). This discrepancy may result from intermodel differences (e.g., high drought severity during the MCA is not found in the IPSL simulation) or from the small number of model simulations analyzed herein, or may be due to the analysis of subregional average hydroclimate time series in previous studies. During the MCA, droughts are more severe in CCSM and CESM over NA because they are larger (Fig. 4d), with the increase in spatiotemporal extent driven by increases in spatial extent (Fig. 4b). The average location of droughts in these simulations does shift southwestward during the MCA (not shown), but the droughts are increasing in spatial extent and not temporal extent (Fig. 4c). The corresponding severity increase is thus not apparent when analyzing hydroclimate time series averaged for the NASW, as much of the increase in spatiotemporal extent occurs outside of the region. All of these changes are true of the paleoclimate record but the droughts also increase in temporal extent. For this reason the corresponding severity increase *is* apparent when analyzing hydroclimate time series averaged for the NASW in the paleoclimate record.

## 2) SPATIOTEMPORAL EXTENT OF DROUGHTS

The interpretation of both average spatiotemporal extent of droughts and centennial-scale natural variability in this characteristic are largely the same as for drought severity, with the exception of the spatiotemporal extent of droughts over MA, where droughts are too small and do not exhibit enough centennial-scale natural variability in any simulations but CCSM (Fig. 4d). There are, however, clear regional biases in the temporal extent (Fig. 4c) and spatial extent (Fig. 4b) of droughts in CGCMs. These are most prevalent over OW where droughts are too short and too spatially extensive in CGCMs as compared to the paleoclimate record, suggesting that CGCMs simulate droughts with realistic spatiotemporal extent for the wrong reasons. Over MA, biases in the spatiotemporal extent of droughts appear to result from simulating droughts that are too short and not spatially extensive enough, although the biases in the temporal extent of droughts are qualitatively more severe. Importantly, the IPSL simulation does not include land use–land cover changes, which could play a part in these biases. Nevertheless, the bias in IPSL looks similar to that in the NCAR CGCMs, suggesting the possibility of a common bias source.

## 3) ATTRIBUTING BIASES IN DROUGHT CHARACTERISTICS

It is important to better understand the origin of the biases in the spatiotemporal extent of droughts (Figs. 4b–d). Focus in this subsection will be limited to biases in average drought characteristics despite clear projection of these biases onto the ability of CGCMs to simulate realistic centennial-scale natural variability (e.g., simulating droughts that are on average too small tends to coincide with not simulating enough centennial-scale natural variability in the spatiotemporal extent of droughts).

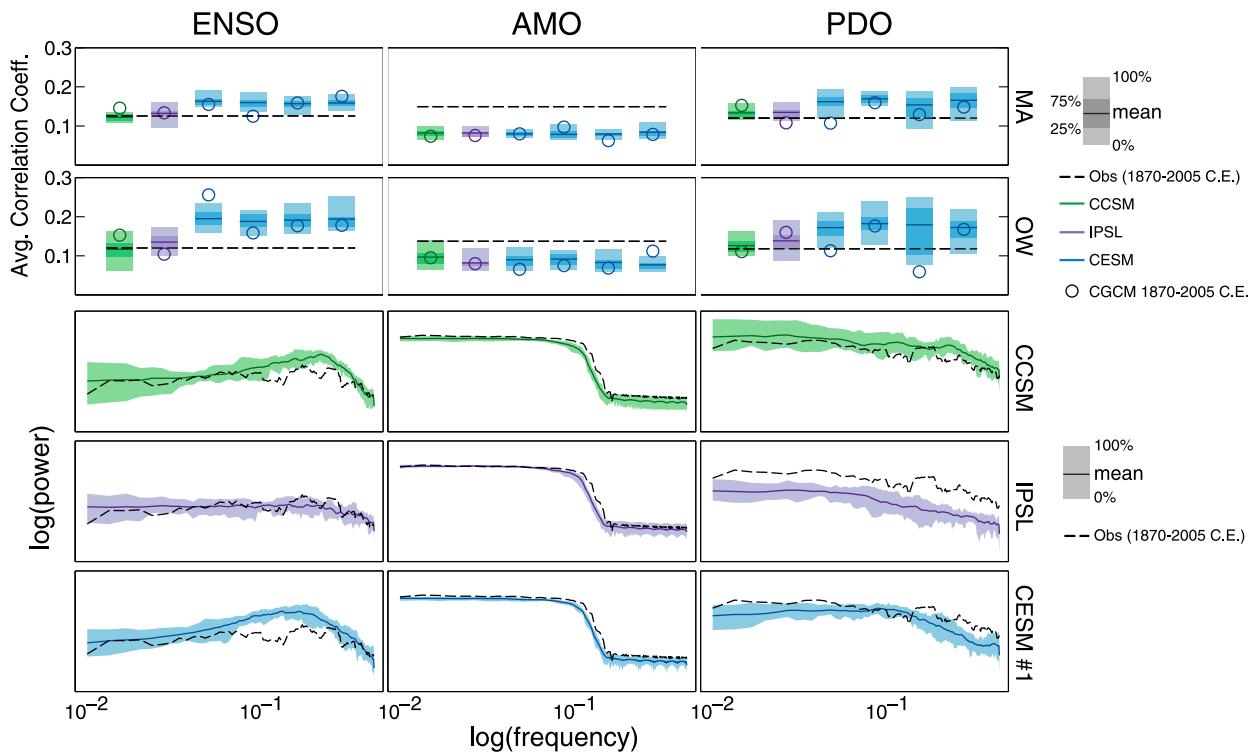


FIG. 8. (top two rows) Average absolute value of correlation coefficients between the (left) ENSO, (center) AMO, and (right) PDO and spatial grid point PDSI. These are calculated over the MA and OW regions for every 136-yr interval between 1000–2005 CE in the last millennium simulations (length of the instrumental interval used for comparison), with the box plots representing the mean, 25th–75th percentiles, and full range. The dashed line is the same metric between 1870 and 2005 CE using the paleoclimate record and ENSO, AMO, and PDO indices calculated from HadISST (Rayner et al. 2003). (bottom three rows) Spectra of the ENSO, AMO, and PDO for every 136-yr interval between 1000 and 2005 CE in the last millennium simulations (mean and range) and for HadISST between 1870–2005 CE. Spectra are estimated using the Thomson’s multitaper method (Thomson 1982). The ENSO index is the average of SSTs over the Niño-3.4 region (5°S–5°N, 90°–150°W), the AMO index is calculated following Trenberth and Shea (2006), and the PDO index is calculated following Mantua et al. (1997). All indices are annually averaged. For the spectra, only one simulation is shown for CESM as the results are nearly identical across the four simulations.

Figure 7 shows the scaling exponent of spectral density, or  $\beta$ , calculated at each space–time grid point in the paleoclimate record and the CGCMs. The term  $\beta$  is a measure of the proportion of variance in low frequencies relative to high frequencies in a time series, with positive  $\beta$  values indicating that low frequencies contribute more to the total variance (e.g., Huybers and Curry 2006; Ault et al. 2013). To calculate  $\beta$ , the power spectral density (PSD) of PDSI at each spatial grid point (Fig. 4a) between 1000 and 1849 CE in the CGCMs and paleoclimate record was calculated using the Thomson’s multitaper method (Thomson 1982). The PSD values corresponding to the lowest 1% of frequencies were then removed, with the remaining PSD values and their frequencies base-10 log transformed. The log-transformed PSD values were binned and averaged into equal log frequency bins before calculating a linear least squares fit. The slope of this fit is defined as  $\beta$ . To test the sensitivity of  $\beta$  to the length of the analysis interval, the calculation was additionally performed for sliding 100-yr intervals between 1000 and 1849 CE. Results are largely consistent between any 100-yr interval and those for the full analysis interval, indicating relatively stationary temporal

estimates of  $\beta$  and therefore justifying an analysis based on the full interval.

The value of  $\beta$  is smaller in CGCMs than in the paleoclimatic record in all six simulations over OW and MA, consistent with simulating droughts that are too short. This does not appear to be a general feature of CGCMs, as the  $\beta$  values are consistent

TABLE 2. Correlation between 1000 and 1849 CE of PDSI averaged over the West (32°N–34°N, 10°W–0°), East (36°–41°N, 20°–37°E), and MidEast (30°–34°N, 33°–47°E) boxes from Cook et al. (2016b). Values are shown for the paleoclimate record and the six simulations of the last millennium.

	West and East	West and MidEast	East and MidEast
Drought Atlas	0.22	−0.01	−0.12
CCSM	0.45	0.32	0.43
IPSL	0.44	0.24	0.39
CESM 1	0.36	0.31	0.40
CESM 2	0.37	0.15	0.34
CESM 3	0.45	0.31	0.46
CESM 4	0.37	0.18	0.36

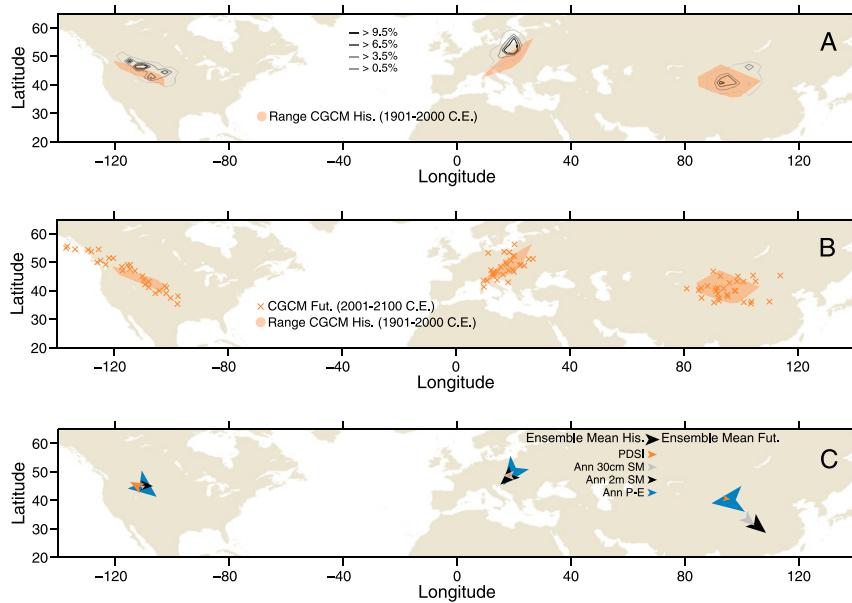


FIG. 9. (a) Area-weighted average drought location in each region for the historical interval (1901–2000 CE), defined as described in section 2c. The orange shaded region represents the geographical range in this metric for the historical interval in the 20 CMIP5 and 6 PMIP3 last millennium simulations. For comparison, the contours represent the area-weighted average drought location for all 100-yr intervals in the paleoclimate record between 1000 and 2000 CE. (b) Area-weighted average drought location in each region for the historical (1901–2000 CE) and future (2001–2100 CE) intervals. The orange shaded region represents the geographical range in this metric for the historical interval in the 26 CMIP5 simulations and the crosses represent each individual simulation for the future interval. (c) For average drought locations computed using PDSI, annually averaged SM integrated over the top 30 cm and 2 m (orange, gray, and black, respectively), and annually averaged  $P - E$  (blue): the base of the arrow (center indent) represents the ensemble mean (20 CMIP5 and 6 PMIP3 last millennium simulations) of the area-weighted average drought location for the historical interval and the tip of the arrow represents the ensemble mean for the future interval.

(and the temporal extent of droughts more consistent) over NA. This is important because it has been suggested that the paleoclimate record may contain nonclimatic, and thus erroneous, low-frequency variability [e.g., Franke et al. 2013; however, the opposite has also been suggested (e.g., Cook et al. 2016a; Smerdon et al. 2016)]. While that may still be the case, the comparisons in Fig. 7 suggest that the paleoclimate record and CGCMs can have similar low-frequency variability.

To better understand the potential dynamical origins of these  $\beta$  differences, Fig. 8 characterizes the dominant modes of NH atmosphere–ocean variability and their teleconnections to hydroclimate over OW and MA. NA is not shown because of the relatively similar low-frequency variability as compared to the paleoclimate record. The hydroclimate impacts of ENSO, a relatively high-frequency mode of variability, are exaggerated in CGCMs over OW and MA and the mode itself appears to have too much high-frequency variability compared to observations (e.g., Bellenger et al. 2014; Otto-Bliesner et al. 2016; Stevenson et al. 2016). Both biases are consistent with simulating droughts that are too short over these regions (Stevenson et al. 2018). Likewise, the much lower-frequency AMO has

hydroclimate impacts that are weaker than in observations. The PDO, which is also associated with relatively low frequencies, has a connection to hydroclimate over OW and MA that is too strong in most CGCMs, while having weaker variability that is biased toward higher frequencies than the observed PDO. Although the latter comparisons are limited by the length of the instrumental interval, for which there are few degrees of freedom for the AMO and PDO, these biases are all consistent with simulating droughts that are too short. Interestingly, simulating exaggerated hydroclimate impacts of the Pacific over MA may also help to explain why CGCMs simulate droughts that are not spatially extensive enough, as the ENSO and PDO teleconnections in this region are a northwest-to-southeast-tilted tripole (not shown) that would limit the total number of space–time grid points in drought conditions when under the Pacific’s influence. Over OW, these exaggerated Pacific hydroclimate impacts would act in the opposite sense, as the teleconnections are a dipole with only weak correlations over the northern grid points and correlations that are too strong and homogenous over the Mediterranean (not shown). Indeed, correlation coefficients



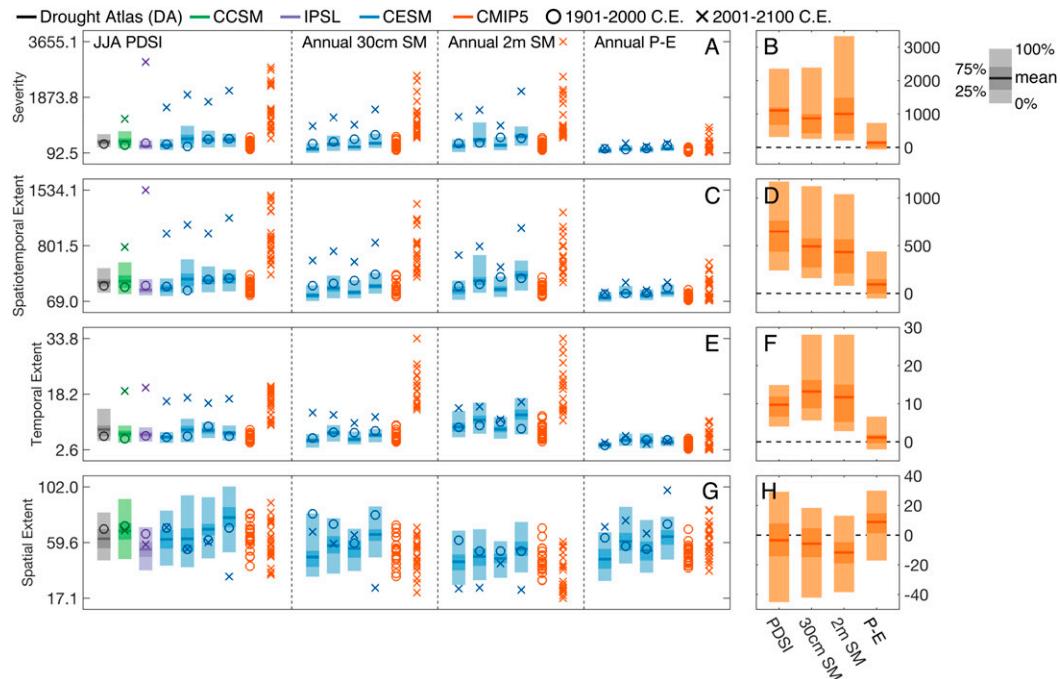


FIG. 11. All results are for the NA region. The box plots represent the average, 25th–75th percentiles, and full range of (a) drought severity, (c) spatiotemporal extent, (e) spatial extent, and (g) temporal extent for all 100-yr intervals between 1000 and 1900 CE in the last millennium simulations and the paleoclimate record (as in Fig. 4). The circle and cross symbols represent the drought severity, spatiotemporal extent, temporal extent, and spatial extent for the historical (1901–2000 CE) and future (2001–2100 CE) intervals. For comparison to the PDSI results the drought severity, spatiotemporal extent, temporal extent, and spatial extent are also plotted using annually averaged SM integrated over the top 30 cm or 2 m and  $P - E$ , subject to data availability. For each hydroclimate metric the droughts are identified for each region using the same drought identification algorithm (see section 2c), with all methodological choices being the same (e.g., the standardization interval, potential functions). (b),(d),(f),(h) The range in drought severity, spatiotemporal extent, temporal extent, and spatial extent, respectively, from the historical (1901–2000 CE) to future (2001–2100 CE) intervals for each hydroclimate metric. The box plots represent the average, 25th–75th percentiles, and full range for each characteristic. The plots (vertical axis) show (a),(b) the sum of standardized departures, (c),(d) the number of spatiotemporal grid points, (e),(f) the number of temporal grid points, and (g),(h) the number of spatial grid points.

between hydroclimate variability in commonly defined regions over OW are too large and positive in all three CGCMs as compared to the paleoclimate record (Table 2). This suggests that hydroclimate variability is too homogeneous (i.e., in phase) in CGCMs as compared to the paleoclimate record, particularly over the drought-prone Mediterranean region (e.g., Cook et al. 2016b), and this may help to explain the tendency for CGCMs to simulate droughts that are too spatially extensive over OW.

#### b. Drought characteristics in the historical interval and the future

We now turn to the characteristics of droughts in the historical interval (1901–2000 CE) and the future. Recent research suggests that hydroclimate in CGCMs is consistent with the paleoclimate record over much of the last millennium, but that CGCMs overestimate the impact of rising temperatures on hydroclimate over the past century (Ljungqvist et al. 2016). Within the framework presented herein, the drought characteristics in CGCMs during the historical interval have not emerged from centennial-scale natural variability (Fig. 4).

In the future, CGCMs suggest that droughts over all three regions become more severe than at any time over the last millennium (Fig. 4e). By the end of the century, droughts are nearly 400% as severe as those in the paleoclimate record in some regions (e.g., OW), consistent with a shift in hydroclimate over large parts of these regions toward permanent drought conditions (B. Cook et al. 2015; Ault et al. 2016). It is worth noting that the “step” in drought severity after the 100-yr interval centered on 2000 CE marks the first 100-yr interval containing a drought that persists through the end of the simulation (2100 CE). These droughts are a consequence of hydroclimate over a subregion drying in response to anthropogenic forcing such that all subsequent years at those spatial grid points are assigned to a drought state.

The increase in severity in Fig. 4e is driven, in part, by an increase in the spatiotemporal extent of droughts (Fig. 4d). Nevertheless, this increase is not large enough to explain the increased severity, suggesting that future droughts will be more severe than past droughts of equivalent spatiotemporal extent. Interestingly, the increase in spatiotemporal extent is driven

largely by an increase in the temporal extent of droughts (Fig. 4c), with no change in their spatial extent in the future (Fig. 4b). While this may seem counterintuitive, droughts over the last millennium already extend over a large portion of the spatial domain and because of the constraints of ocean and imposed boundaries (e.g., between OW and MA) these droughts have little room to increase in space. By contrast, drought persistence is not limited in the future, except by the end of the simulations.

Surprisingly, there is little change in drought locations in the historical interval (1901–2000 CE; Fig. 9a) or the future (Fig. 9b), which may also result from the aforementioned arguments on the spatial extent of these features—the most persistent and severe droughts may already be so spatially extensive that there is little room for their location to shift. Nevertheless, there does appear to be a shift in future drought locations over NA, with these features shifting northwest relative to the historical interval. In particular, there is a much higher occurrence of droughts in the Alaska portion of the domain in the future (not shown). This result could be due to the use of PDSI as the hydroclimate metric, as PDSI can struggle to represent variability at high-latitude locations like Alaska (e.g., Alley 1984). It may also result from the fixed spatial grid that is used herein, as the fixed spatial grid is missing much of southern and central Canada (Fig. 4a). Nevertheless, a more complete fixed spatial grid corresponding to the tree-ring record in 1500 (Fig. 10a) also suggests that there is a northward (though less westward) shift in future drought locations over NA (not shown). Importantly, all results for the future and historical interval are consistent for this more complete fixed spatial grid (Figs. 10b–e), as well as shorter time intervals (e.g., 50-yr instead of 100-yr intervals; not shown). Note that the missing spatial grid points and the large number of future droughts over Alaska explain the presence of average drought locations in some models over the Pacific. Specifically, if the locations of the 10 most persistent and severe droughts in the future are largely split between Alaska and the western and central continental United States, the average drought location can occur over the ocean.

### c. Consistency of results across hydroclimate metrics

Importantly, discrepancies between drought characteristics and hydroclimate change as measured by different hydroclimate metrics are expected as they are measuring different climatic phenomena:  $P - E$  should approximately measure runoff, while PDSI is a soil moisture metric. Nevertheless, PDSI has well-documented issues when used as a metric of future hydroclimate change, and while it is necessary to use PDSI for a direct comparison of the CGCMs with the paleoclimate record, these issues are expected to impact some of the results presented herein. To specifically test whether the results presented in the previous subsection are consistent across hydroclimate metrics, the characteristics of droughts defined using different hydroclimate metrics are compared in Fig. 11. For each hydroclimate metric the droughts in Fig. 11 are identified for each region using the same drought identification algorithm (see section 2c), with all methodological choices being the same (e.g., the standardization interval, potential functions, spatiotemporal grid points, etc.); nevertheless, the individual droughts being identified for each

metric will be different. Because the metrics displayed in Fig. 11 are either an integration of standardized departures from the mean of a common interval (e.g., severity) or an emergent feature of droughts described in a common geographical framework (e.g., spatiotemporal extent), all results are directly comparable across the different hydroclimate metrics. For simplicity, only results for NA are presented in Fig. 11, as the results are largely consistent with those derived for the OW and MA regions. Nevertheless, there are regional differences that will be described qualitatively in the following subsection.

The results for PDSI (Figs. 4b–e) are consistent with those for SM over NA and OW, with large increases in drought severity (Fig. 11b) and spatiotemporal extent (Fig. 11d) in the future and little change in the spatial extent of these features (Fig. 11h). Shallow (30 cm) and deep (2 m) SM changes in the future over these regions are largely consistent, at least as measured by the severity of droughts (Fig. 11b). Future changes in PDSI thus appear to be consistent with those for SM, not just at shallow depths as has been suggested previously (e.g., Cheng et al. 2016; Ault et al. 2016; Berg et al. 2017; Swann et al. 2016).

Interestingly, however, the results for PDSI are not consistent with SM over MA, particularly in the CESM simulations where drought severity for SM does not increase in the future (not shown). More generally, the CMIP simulations suggests bimodality across CGCMs with respect to changes in future drought severity for SM over MA—approximately half of CGCMs suggest large increases in severity at 2-m depth, while increases occur across nearly all CGCMs for PDSI. Diagnosing the reasons for this bimodality is difficult given the available CGCM outputs, and the lack of ensembles for each CGCM. Nevertheless, the four CESM simulations allow for a qualitative diagnosis of PDSI and SM decoupling over MA within a single CGCM. In CESM the biggest differences between future PDSI and SM are in the northeastern portion of the MA region (not shown). In that subregion there is a large increase in snowfall from December–March, which is partially compensated by decreases in April and May. We hypothesize that this increase in total winter snowfall produces a delayed recharge of summer SM that can compensate for increased atmospheric demand for moisture. PDSI does not capture snowfall, and thus does not have this delayed recharge effect, and by consequence may overestimate the impact of future increases in atmospheric moisture demand over the northeastern subregion of MA.

For  $P - E$ , on average, there are increases in future drought severity across all three regions, largely driven by the spatiotemporal extent of these features, with the exception of the CESM simulations over OW. Nevertheless, these increases are much smaller than for PDSI or SM (Fig. 11b), although they do rise above the range of centennial-scale natural variability (the last millennium) in the CESM simulations over NA (Fig. 11a) and MA. Interestingly, unlike for PDSI and SM, future increases in the spatiotemporal extent of  $P - E$  droughts over NA (Fig. 11d) include a role for increases in the spatial extent of droughts (Fig. 11h), although the spatiotemporal extent increases are much smaller overall for  $P - E$  than for the other hydroclimate metrics.

Perhaps the biggest difference between the hydroclimate metrics is the change in drought locations from the historical interval into the future (Fig. 9c). Of particular interest is the change in drought locations for  $P - E$ , which are much larger than those for PDSI or SM. Likewise, while they tend to be in the same direction as for PDSI and SM over OW (westward) and for SM over NA (eastward), they are actually in the opposite direction over MA (westward vs eastward). As was noted previously, drought location changes for PDSI are generally small, with the exception of over NA where there is a large increase in the occurrence of droughts in the Alaska portion of the domain. This same shift in drought locations does not occur for SM, further suggesting that it results from issues with PDSI as a hydroclimate metric at high latitudes (e.g., Alley 1984). More generally, while the future changes to drought characteristics are consistent between PDSI and SM, this is less true for changes in the locations of these features.

For SM, the drought location changes are in the same direction at both depths, but are larger for 2-m than for 30-cm SM. To better understand this discrepancy, we analyze the spatial pattern of end-of-century changes (2081–2100 CE averages relative to preindustrial) in moisture supply (precipitation) and demand (PET) as compared to SM in the four CESM simulations (not shown). The discrepancy is consistent with 2-m SM responding to moisture supply changes more closely than 30-cm SM (e.g., Cheng et al. 2016), while the surface soil moisture layers respond more closely to moisture demand. The actual drought location changes, however, involve a complex interplay between subregional changes in moisture supply and demand. For instance, over MA drying is projected for the northwest and southeast subregions. The drying in the northwest is dominated by moisture demand (large PET increases), leading to greater and more spatially extensive drying for 30-cm SM. The drying in the southeast is dominated by moisture supply (large precipitation decreases), leading to greater drying for 2-m SM. For both 30-cm and 2-m SM, drought locations are shifting toward the southeast, but for 30-cm SM there is also an increase in the occurrence of droughts in the northwest. This makes the total drought location changes, an average of all drought locations, less southeast shifted for 30-cm SM. Similar arguments can be made for the drought location changes for 30-cm and 2-m SM over the NA and OW regions.

#### *d. Will CGCM biases cause systematic biases in drought risk projections?*

The risk of a future drought will necessarily involve anthropogenically forced changes *and* natural variability. While anthropogenic forcing can itself change natural variability and thus its contribution to future drought risk, it is still useful to validate that CGCMs simulate the full range of natural variability in the absence of this forcing. With this in mind, a comparison of the paleoclimate record to CGCMs, suggests that they can either vastly underestimate (IPSL over NA) or slightly overestimate (CCSM over NA) centennial-scale natural variability in the characteristics of persistent and severe drought (Figs. 4b–e). Nevertheless, only a limited number of CGCMs were available for this comparison, precluding confident statements on systematic behavior across CGCMs. In the following subsection we focus specifically on

natural variability over NA because of the clear overestimate (CCSM) and underestimate (IPSL) split across CGCMs.

To better understand if CGCMs systematically over or underestimate centennial-scale natural variability in drought characteristics we use the LIM-based framework to produce 100-member ensembles of surrogate climate trajectories for each CGCM simulation. The range of drought characteristics for each 100-member ensemble are shown in Fig. 12. While CGCMs can both under- or overestimate the range in drought severity as compared to a similar observation-based ensemble of the instrumental interval, the majority of CGCMs underestimate these ranges and thus are likely to underestimate natural contributions to drought risk (22 of 25 LIMs have a narrower range, shown as the bar width in Fig. 12, as compared to the observation-based ensemble). This result is consistent with previous research suggesting that hydroclimate in CGCMs may be underdispersive as compared to the paleoclimate record [Ault et al. 2013; Parsons et al. 2017; however, it has been argued that the paleoclimate record may be biased; see Franke et al. (2013)] but these statements can now be expanded to droughts specifically as compared to observations, and with spatiotemporal context.

Interestingly, however, LIMs trained on different simulations from the same CGCM can fall on either side of this split—the LIMs trained on CESM simulations 3 and 4 have drought severity ranges that are over- and underestimated relative of the range of drought severity from the observation-based LIM. This apparent discrepancy arises because the coupled atmosphere–ocean variability during the LIM training interval (1951–2000 CE) is spurious in CESM simulations 3 and 4. In CESM simulation 3 there is very weak variability in the tropical Pacific over the 50-yr LIM training interval, including a near total absence of El Niño events, with the opposite being true of CESM simulation 4. The ENSO teleconnection to hydroclimate in CESM is strong, with a dipolar pattern over NA that limits the total number of space–time grid points in drought conditions when under the Pacific’s influence. The weak variability in the tropical Pacific in simulation 3 during the LIM training interval leads to incomplete sampling of this teleconnection. Without the tropical Pacific influence on hydroclimate in the LIM, the associated surrogate climate trajectories contain droughts with greater spatial extent and thus spatiotemporal extent and severity. These results reinforce the importance of running large ensembles and long simulations with CGCMs (e.g., Coats and Mankin 2016), as training statistical models on a single climate trajectory can result in unrepresentative behavior and thus erroneous conclusions.

## 4. Discussion

CGCMs struggle to simulate some characteristics of persistent and severe droughts as compared to the paleoclimate record, particularly over the OW and MA regions. While this may limit confidence in future projections of drought risk, by the end of the century the characteristics of droughts are dominated by the hydroclimate response to anthropogenic forcing. More specifically, droughts become at least 300% more severe by the end of the twenty-first century than at any time over the last millennium, changes so large that they render

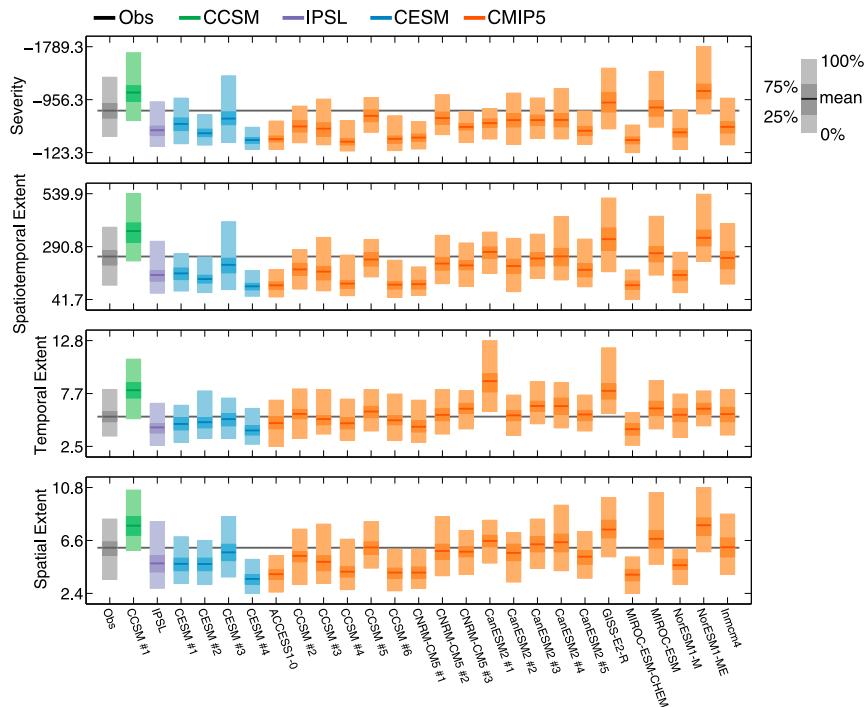


FIG. 12. All results are for the NA region. The box plots represent the average, 25th–75th percentiles, and full range of drought severity, spatiotemporal extent, temporal extent, and spatial extent based on 100 ensemble members of an LIM trained on PDSI and global SSTs for the 1951–2000 period from each simulation or the Sheffield et al. (2006) PDSI dataset and global SSTs from the Kaplan et al. (1998) dataset (section 2). The black boxplots are for the observations, and the green and purple for the CCSM and IPSL last millennium simulations, respectively. The blue colors are for the four CESM last millennium simulations and the orange for the 20 CMIP5 historical simulations.

natural (forced and internal) variability unimportant to the occurrence of future droughts. Therefore, if CGCMs realistically simulate the hydroclimate response to anthropogenic forcing, long-term drought risk projections can be viewed with more confidence. Of relevance to this, and in contrast to previous studies (e.g., Ljungqvist et al. 2016), the results presented herein suggest that CGCMs do not overestimate the response of hydroclimate to anthropogenic forcing over the past century—at least with regard to the characteristics of persistent and severe droughts. This is critical, because if CGCMs are already overestimating the hydroclimate response to anthropogenic forcing, the dire future projections are also likely to be overestimated (Ljungqvist et al. 2016). We find no evidence that anthropogenic forcing of the characteristics of droughts has emerged from centennial-scale natural variability in CGCMs, consistent with the still weak, though nonnegligible, influence of anthropogenic forcing on persistent and severe droughts in the real world (e.g., Williams et al. 2015). Recent research suggests that global “fingerprints” of anthropogenic forcing are now detectable in hydroclimate, but consistent with results presented herein this signal cannot yet be detected regionally (Marvel et al. 2019).

Projections of large increases in drought severity across all three regions are consistent with previous studies of future

drought risk over individual subregions like the NASW (B. Cook et al. 2015; Ault et al. 2016), suggesting that drought risk increases are a ubiquitous feature of future hydroclimate over the NH extratropics. There are important caveats to this interpretation, however, including the use of PDSI as a hydroclimate metric. While this offline metric is consistent with some online metrics (with the exception of changes in drought locations), the aforementioned changes are most relevant to SM (at 30 cm and 2 m). For instance, the response of runoff and vegetation to anthropogenic forcing can be large and in opposition to those in SM (e.g., Scheff et al. 2017; Mankin et al. 2017, 2018), in part due to the impact of increasing levels of atmospheric carbon dioxide on vegetation physiology. This is apparent in future changes to the characteristics of  $P - E$  droughts. While  $P - E$  droughts become more severe in the future, the projected changes are much smaller than for PDSI or SM. It is clear from these results that both hydroclimate and drought must be rigorously defined in future projections, as the associated uncertainties depend on these definitions. Of particular interest is whether CGCMs realistically simulate vegetation responses to anthropogenic forcing (Smith et al. 2016; Mankin et al. 2017, 2018; Trugman et al. 2018), an understanding that will further untangle the various hydroclimatic components and how they will change in the future.

Nevertheless, the response of  $P - E$  to anthropogenic forcing is not sufficiently large as to render natural variability unimportant to future climate change. Realistically projecting drought risk using this and related metrics, like runoff, requires that CGCMs realistically simulate natural variability, because it is a more important component of future drought risk. Therefore, when using these hydroclimate metrics, biases in the ability of CGCMs to simulate natural variability will limit confidence in drought risk projections. In this regard, CGCMs could be under- or overdispersive in representing the full range of potential future hydroclimate states. With the assumption that natural PDSI variability is relevant to these metrics, which should be true even if the PDSI response to anthropogenic forcing is not, we present two frameworks for determining if under- or overdispersiveness is more likely. Based on the paleoclimate record and last-millennium simulations, it is clear that CGCMs struggle to simulate the characteristics of droughts over OW (and to a lesser extent MA). This causes systematic underdispersiveness in the temporal extent of droughts over OW and MA, and the spatiotemporal extent of droughts over MA—with these biases related to modes of NH atmosphere–ocean variability and their impacts on regional hydroclimate. Over NA, CGCMs are more successful at simulating the characteristics of droughts. However, an LIM-based framework applied to a multimodel ensemble of CGCMs suggests that they are generally underdispersive when it comes to drought severity over this region. Importantly, it is not clear if underdispersiveness results from biases in the simulation of internal variability, the response to natural forcing, or the imposed natural forcings themselves. In the latter case, the results presented herein may not limit confidence in drought risk projections.

Importantly, all of the aforementioned interpretations also apply to projections of PDSI and SM over the next few decades (before the greatest anthropogenic impacts on hydroclimate) or if real-world anthropogenic emissions are reduced relative to the RCP8.5 scenario that is analyzed herein (see, e.g., Coats and Mankin 2016; Lehner et al. 2017). Regardless of these caveats, the end-of-century changes in drought severity projected by CGCMs are troubling, particularly for societal risks that track changes in near-surface SM (e.g., wildfire; Westerling and Bryant 2008; Williams and Abatzoglou 2016). Confidence in risk projections of other societally relevant hydroclimatic components will require additional research similar to what we present herein, that seeks to better understand the ability of CGCMs to simulate natural variability on the decadal-to-centennial time scales most relevant to future projections.

## 5. Conclusions

Here we use a novel framework for analyzing spatiotemporally contiguous climatic events to provide critical insights into how droughts change in response to anthropogenic forcing in both space and time and how climate model biases impact these projections. Importantly, this framework can be applied to other climatic features to provide similar insights. The specific insights that we have provided are the following:

- We find no evidence that anthropogenic forcing of the regional characteristics of droughts emerges from natural variability over the past century (1901–2000).
- Despite this, large increases in drought severity, as measured by soil moisture metrics, are projected for the twenty-first century over the Northern Hemisphere extratropics.
- These increases are dominated by changes in the spatiotemporal extent of droughts via their temporal extent, with little increase in their spatial extent.
- There is little change in the location of droughts in the future for offline soil moisture metrics, but there are changes for some online soil moisture metrics.
- Small changes in drought severity, as measured by precipitation minus evaporation, are projected for the coming century. There are, however, large changes in the locations of these droughts.
- Biases in the ability of climate models to simulate natural variability undermine confidence in drought risk projections for: 1) hydroclimate metrics with smaller magnitude responses to anthropogenic forcing, and 2) for all hydroclimate metrics over the next few decades.
- These biases lead climate models to be underdispersive when attempting to simulate the full range of future hydroclimate states.

*Acknowledgments.* This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement 1852977. JAF was also supported by the Regional and Global Model Analysis (RGMA) component of the Earth and Environmental System Modeling Program of the U.S. Department of Energy's Office of Biological & Environmental Research (BER) via National Science Foundation IA 1844590. JS was supported in part by the U.S. National Science Foundation through Grants AGS-1602920 and AGS-1805490, and by the National Oceanic and Atmospheric Administration by Grant NA20OAR4310425. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (listed in Table 1) for producing and making available their model output. For CMIP, the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portal. We thank the editor and two reviewers for comments that greatly improved the quality of this manuscript. This is SOEST Publication No. 11116 and LDEO Publication No. 8450.

## REFERENCES

- Alley, W. M., 1984: The Palmer drought severity index: Limitations and assumptions. *J. Climate Appl. Meteor.*, **23**, 1100–1109, [https://doi.org/10.1175/1520-0450\(1984\)023<1100:TPDSIL>2.0.CO;2](https://doi.org/10.1175/1520-0450(1984)023<1100:TPDSIL>2.0.CO;2).
- Ault, T. R., J. E. Cole, J. T. Overpeck, G. T. Pederson, S. St. George, B. Otto-Bliessner, C. A. Woodhouse, and C. Deser, 2013: The continuum of hydroclimate variability in western North America during the last millennium. *J. Climate*, **26**, 5863–5878, <https://doi.org/10.1175/JCLI-D-11-00732.1>.
- , J. S. Mankin, B. I. Cook, and J. E. Smerdon, 2016: Relative impacts of mitigation, temperature, and precipitation on 21st-

- century megadrought risk in the American Southwest. *Sci. Adv.*, **2**, e1600873, <https://doi.org/10.1126/sciadv.1600873>.
- , S. St. George, J. E. Smerdon, S. Coats, J. S. Mankin, C. M. Carillo, B. I. Cook, and S. Stevenson, 2018: A robust null hypothesis for the potential causes of megadrought in western North America. *J. Climate*, **31**, 3–24, <https://doi.org/10.1175/JCLI-D-17-0154.1>.
- Baek, S. H., J. E. Smerdon, S. Coats, A. P. Williams, B. I. Cook, E. R. Cook, and R. Seager, 2017: Precipitation, temperature, and teleconnection signals across the combined North American, Monsoon Asia, and Old World drought atlases. *J. Climate*, **30**, 7141–7155, <https://doi.org/10.1175/JCLI-D-16-0766.1>.
- Bellenger, H., É. Guilyardi, J. Leloup, M. Lengaigne, and J. Vialard, 2014: ENSO representation in climate models: From CMIP3 to CMIP5. *Climate Dyn.*, **42**, 1999–2018, <https://doi.org/10.1007/s00382-013-1783-z>.
- Berg, A., and J. Sheffield, 2018: Climate change and drought: The soil moisture perspective. *Curr. Climate Change Rep.*, **4**, 180–191, <https://doi.org/10.1007/s40641-018-0095-0>.
- , —, and P. C. Milly, 2017: Divergent surface and total soil moisture projections under global warming. *Geophys. Res. Lett.*, **44**, 236–244, <https://doi.org/10.1002/2016GL071921>.
- Cheng, L., M. Hoerling, A. AghaKouchak, B. Livneh, X. W. Quan, and J. Eischeid, 2016: How has human-induced climate change affected California drought risk? *J. Climate*, **29**, 111–120, <https://doi.org/10.1175/JCLI-D-15-0260.1>.
- Coats, S., and J. S. Mankin, 2016: The challenge of accurately quantifying future megadrought risk in the American Southwest. *Geophys. Res. Lett.*, **43**, 9225–9233, <https://doi.org/10.1002/2016GL070445>.
- , J. E. Smerdon, R. Seager, B. I. Cook, and J. F. González-Rouco, 2013: Megadroughts in southwestern North America in ECHO-G millennial simulations and their comparison to proxy drought reconstructions. *J. Climate*, **26**, 7635–7649, <https://doi.org/10.1175/JCLI-D-12-00603.1>.
- , —, B. I. Cook, and R. Seager, 2015a: Are simulated megadroughts in the North American Southwest forced? *J. Climate*, **28**, 124–142, <https://doi.org/10.1175/JCLI-D-14-00071.1>.
- , B. I. Cook, J. E. Smerdon, and R. Seager, 2015b: North American pancontinental droughts in model simulations of the last millennium. *J. Climate*, **28**, 2025–2043, <https://doi.org/10.1175/JCLI-D-14-00634.1>.
- , J. E. Smerdon, R. Seager, D. Griffin, and B. I. Cook, 2015c: Winter-to-summer precipitation phasing in southwestern North America: A multicentury perspective from paleoclimatic model–data comparisons. *J. Geophys. Res. Atmos.*, **120**, 8052–8064, <https://doi.org/10.1002/2015JD023085>.
- , —, K. B. Karnauskas, and R. Seager, 2016a: The improbable but unexceptional occurrence of megadrought clustering in the American West during the Medieval Climate Anomaly. *Environ. Res. Lett.*, **11**, 074025, <https://doi.org/10.1088/1748-9326/11/7/074025>.
- , —, B. I. Cook, R. Seager, E. R. Cook, and K. J. Anchukaitis, 2016b: Internal ocean–atmosphere variability drives megadroughts in Western North America. *Geophys. Res. Lett.*, **43**, 9886–9894, <https://doi.org/10.1002/2016GL070105>.
- Cook, B. I., J. E. Smerdon, R. Seager, and S. Coats, 2014a: Global warming and 21st century drying. *Climate Dyn.*, **43**, 2607–2627, <https://doi.org/10.1007/s00382-014-2075-y>.
- , —, —, and E. R. Cook, 2014b: Pan-continental droughts in North America over the last millennium. *J. Climate*, **27**, 383–397, <https://doi.org/10.1175/JCLI-D-13-00100.1>.
- , T. R. Ault, and J. E. Smerdon, 2015: Unprecedented 21st century drought risk in the American Southwest and Central Plains. *Sci. Adv.*, **1**, e1400082, <https://doi.org/10.1126/sciadv.1400082>.
- , E. R. Cook, J. E. Smerdon, R. Seager, A. P. Williams, S. Coats, D. W. Stahle, and J. V. Díaz, 2016a: North American megadroughts in the common era: Reconstructions and simulations. *Wiley Interdiscip. Rev.: Climate Change*, **7**, 411–432, <https://doi.org/10.1002/wcc.394>.
- , K. J. Anchukaitis, R. Touchan, D. M. Meko, and E. R. Cook, 2016b: Spatiotemporal drought variability in the Mediterranean over the last 900 years. *J. Geophys. Res. Atmos.*, **121**, 2060–2074, <https://doi.org/10.1002/2015jd023929>.
- Cook, E. R., R. Seager, M. A. Cane, and D. W. Stahle, 2007: North American drought: Reconstructions, causes, and consequences. *Earth-Sci. Rev.*, **81**, 93–134, <https://doi.org/10.1016/j.earscirev.2006.12.002>.
- , K. J. Anchukaitis, B. M. Buckley, R. D. D’Arrigo, G. C. Jacoby, and W. E. Wright, 2010: Asian monsoon failure and megadrought during the last millennium. *Science*, **328**, 486–489, <https://doi.org/10.1126/science.1185188>.
- , and Coauthors, 2015: Old World megadroughts and pluvials during the common era. *Sci. Adv.*, **1**, e1500561, <https://doi.org/10.1126/sciadv.1500561>.
- Deser, C., A. S. Phillips, M. A. Alexander, and B. V. Smoliak, 2014: Projecting North American climate over the next 50 years: Uncertainty due to internal variability. *J. Climate*, **27**, 2271–2296, <https://doi.org/10.1175/JCLI-D-13-00451.1>.
- Duffy, P. B., P. Brando, G. P. Asner, and C. B. Field, 2015: Projections of future meteorological drought and wet periods in the Amazon. *Proc. Natl. Acad. Sci. USA*, **112**, 13 172–13 177, <https://doi.org/10.1073/pnas.1421010112>.
- Franke, J., D. Frank, C. C. Raible, J. Esper, and S. Brönnimann, 2013: Spectral biases in tree-ring climate proxies. *Nat. Climate Change*, **3**, 360–364, <https://doi.org/10.1038/nclimate1816>.
- Fu, Q., A. Banerjee, S. Liess, and P. K. Snyder, 2012: Drought detection of the last century: An MRF-based approach. *Proc. 2012 SIAM Int. Conf. on Data Mining*, Anaheim, CA, Society for Industrial and Applied Mathematics, 24–34.
- Herrera-Estrada, J. E., Y. Satoh, and J. Sheffield, 2017: Spatiotemporal dynamics of global drought. *Geophys. Res. Lett.*, **44**, 2254–2263, <https://doi.org/10.1002/2016GL071768>.
- Hessl, A. E., and Coauthors, 2018: Past and future drought in Mongolia. *Sci. Adv.*, **4**, e1701832, <https://doi.org/10.1126/sciadv.1701832>.
- Huybers, P., and W. Curry, 2006: Links between annual, Milankovitch and continuum temperature variability. *Nature*, **441**, 329–332, <https://doi.org/10.1038/nature04745>.
- Kaplan, A., M. A. Cane, Y. Kushnir, A. C. Clement, M. B. Blumenthal, and B. Rajagopalan, 1998: Analyses of global sea surface temperature 1856–1991. *J. Geophys. Res.*, **103**, 18 567–18 589, <https://doi.org/10.1029/97JC01736>.
- Karnauskas, K. B., J. E. Smerdon, R. Seager, and J. F. González-Rouco, 2012: A Pacific centennial oscillation predicted by coupled GCMs. *J. Climate*, **25**, 5943–5961, <https://doi.org/10.1175/JCLI-D-11-00421.1>.
- Kay, J. E., and Coauthors, 2015: The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bull. Amer. Meteor. Soc.*, **96**, 1333–1349, <https://doi.org/10.1175/BAMS-D-13-00255.1>.
- Landrum, L., B. L. Otto-Bliesner, E. R. Wahl, A. Conley, P. J. Lawrence, N. Rosenbloom, and H. Teng, 2013: Last millennium climate and its variability in CCSM4. *J. Climate*, **26**, 1085–1111, <https://doi.org/10.1175/JCLI-D-11-00326.1>.

- Lehner, F., S. Coats, T. F. Stocker, A. G. Pendergrass, B. M. Sanderson, C. C. Raible, and J. E. Smerdon, 2017: Projected drought risk in 1.5°C and 2°C warmer climates. *Geophys. Res. Lett.*, **44**, 7419–7428, <https://doi.org/10.1002/2017GL074117>.
- , C. Deser, I. R. Simpson, and L. Terray, 2018: Attributing the US Southwest's recent shift into drier conditions. *Geophys. Res. Lett.*, **45**, 6251–6261, <https://doi.org/10.1029/2018GL078312>.
- Ljungqvist, F. C., P. J. Krusic, H. S. Sundqvist, E. Zorita, G. Brattström, and D. Frank, 2016: Northern Hemisphere hydroclimate variability over the past twelve centuries. *Nature*, **532**, 94–98, <https://doi.org/10.1038/nature17418>.
- Mankin, J. S., J. E. Smerdon, B. I. Cook, A. P. Williams, and R. Seager, 2017: The curious case of projected twenty-first-century drying but greening in the American West. *J. Climate*, **30**, 8689–8710, <https://doi.org/10.1175/JCLI-D-17-0213.1>.
- , R. Seager, J. E. Smerdon, B. I. Cook, A. P. Williams, and R. M. Horton, 2018: Blue water trade-offs with vegetation in a CO<sub>2</sub>-enriched climate. *Geophys. Res. Lett.*, **45**, 3115–3125, <https://doi.org/10.1002/2018GL077051>.
- Mantua, N. J., S. R. Hare, Y. Zhang, J. M. Wallace, and R. C. Francis, 1997: A Pacific interdecadal climate oscillation with impacts on salmon production. *Bull. Amer. Meteor. Soc.*, **78**, 1069–1079, [https://doi.org/10.1175/1520-0477\(1997\)078<1069:APICOW>2.0.CO;2](https://doi.org/10.1175/1520-0477(1997)078<1069:APICOW>2.0.CO;2).
- Marvel, K., B. I. Cook, C. J. Bonfils, P. J. Durack, J. E. Smerdon, and A. P. Williams, 2019: Twentieth-century hydroclimate changes consistent with human influence. *Nature*, **569**, 59–65, <https://doi.org/10.1038/s41586-019-1149-8>.
- Milly, P. C., and K. A. Dunne, 2016: Potential evapotranspiration and continental drying. *Nat. Climate Change*, **6**, 946–949, <https://doi.org/10.1038/nclimate3046>.
- Otto-Bliesner, B. L., and Coauthors, 2016: Climate variability and change since 850 CE: An ensemble approach with the Community Earth System Model. *Bull. Amer. Meteor. Soc.*, **97**, 735–754, <https://doi.org/10.1175/BAMS-D-14-00233.1>.
- Palmer, W. C., 1965: *Meteorological Drought*. U.S. Department of Commerce, Weather Bureau, 65 pp.
- Parsons, L. A., G. R. Loope, J. T. Overpeck, T. R. Ault, R. Stouffer, and J. E. Cole, 2017: Temperature and precipitation variance in CMIP5 simulations and paleoclimate records of the last millennium. *J. Climate*, **30**, 8885–8912, <https://doi.org/10.1175/JCLI-D-16-0863.1>.
- Penman, H. L., 1948: Natural evaporation from open water, bare soil and grass. *Proc. Roy. Soc. London*, **193A**, 120–145, <https://doi.org/10.1098/rspa.1948.0037>.
- Rayner, N. A. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, <https://doi.org/10.1029/2002JD002670>.
- Scheff, J., R. Seager, H. Liu, and S. Coats, 2017: Are glacials dry? Consequences for paleoclimatology and for greenhouse warming. *J. Climate*, **30**, 6593–6609, <https://doi.org/10.1175/JCLI-D-16-0854.1>.
- Sheffield, J., G. Goteti, and E. F. Wood, 2006: Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *J. Climate*, **19**, 3088–3111, <https://doi.org/10.1175/JCLI3790.1>.
- Smerdon, J. E., S. Coats, and T. R. Ault, 2016: Model-dependent spatial skill in pseudoproxy experiments testing climate field reconstruction methods for the Common Era. *Climate Dyn.*, **46**, 1921–1942, <https://doi.org/10.1007/s00382-015-2684-0>.
- Smith, W. K., S. C. Reed, C. C. Cleveland, A. P. Ballantyne, W. R. Anderegg, W. R. Wieder, Y. Y. Liu, and S. W. Running, 2016: Large divergence of satellite and Earth system model estimates of global terrestrial CO<sub>2</sub> fertilization. *Nat. Climate Change*, **6**, 306–310, <https://doi.org/10.1038/nclimate2879>.
- Steiger, N. J., J. E. Smerdon, B. I. Cook, R. Seager, A. P. Williams, and E. R. Cook, 2019: Oceanic and radiative forcing of medieval megadroughts in the American Southwest. *Sci. Adv.*, **5**, eaax0087, <https://doi.org/10.1126/sciadv.aax0087>.
- Stevenson, S., A. Timmermann, Y. Chikamoto, S. Langford, and P. DiNezio, 2015: Stochastically generated North American megadroughts. *J. Climate*, **28**, 1865–1880, <https://doi.org/10.1175/JCLI-D-13-00689.1>.
- , B. Otto-Bliesner, J. Fasullo, and E. Brady, 2016: “El Niño like” hydroclimate responses to last millennium volcanic eruptions. *J. Climate*, **29**, 2907–2921, <https://doi.org/10.1175/JCLI-D-15-0239.1>.
- , and Coauthors, 2018: Climate variability, volcanic forcing, and last millennium hydroclimate extremes. *J. Climate*, **31**, 4309–4327, <https://doi.org/10.1175/JCLI-D-17-0407.1>.
- Stine, S., 1994: Extreme and persistent drought in California and Patagonia during mediaeval time. *Nature*, **369**, 546–549, <https://doi.org/10.1038/369546a0>.
- Swann, A. L., F. M. Hoffman, C. D. Koven, and J. T. Randerson, 2016: Plant responses to increasing CO<sub>2</sub> reduce estimates of climate impacts on drought severity. *Proc. Natl. Acad. Sci. USA*, **113**, 10 019–10 024, <https://doi.org/10.1073/pnas.1604581113>.
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, <https://doi.org/10.1175/BAMS-D-11-00094.1>.
- Thomson, D. J., 1982: Spectrum estimation and harmonic analysis. *Proc. IEEE*, **70**, 1055–1096, <https://doi.org/10.1109/PROC.1982.12433>.
- Trenberth, K. E., and D. J. Shea, 2006: Atlantic hurricanes and natural variability in 2005. *Geophys. Res. Lett.*, **33**, L12704, <https://doi.org/10.1029/2006GL026894>.
- Trugman, A. T., D. Medvigy, J. S. Mankin, and W. R. L. Anderegg, 2018: Soil moisture stress as a major driver of carbon cycle uncertainty. *Geophys. Res. Lett.*, **45**, 6495–6503, <https://doi.org/10.1029/2018GL078131>.
- Westerling, A. L., and B. P. Bryant, 2008: Climate change and wildfire in California. *Climatic Change*, **87**, 231–249, <https://doi.org/10.1007/s10584-007-9363-z>.
- Williams, A. P., and J. T. Abatzoglou, 2016: Recent advances and remaining uncertainties in resolving past and future climate effects on global fire activity. *Curr. Climate Change Rep.*, **2**, 1–14, <https://doi.org/10.1007/s40641-016-0031-0>.
- , 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>.
- , and Coauthors, 2020: Large contribution from anthropogenic warming to an emerging North American megadrought. *Science*, **368**, 314–318, <https://doi.org/10.1126/science.aaz9600>.
- Yang, Y., M. L. Roderick, S. Zhang, T. R. McVicar, and R. J. Donohue, 2019: Hydrologic implications of vegetation response to elevated CO<sub>2</sub> in climate projections. *Nat. Climate Change*, **9**, 44–48, <https://doi.org/10.1038/s41558-018-0361-0>.