A Robust Null Hypothesis for the Potential Causes of Megadrought in Western North America

Toby R. Ault, a Scott St. George, b Jason E. Smerdon, c Sloan Coats, d Justin S. Mankin, e, f Carlos M. Carrillo, a Benjamin I. Cook, e, c and Samantha Stevenson g

a Department of Earth and Atmospheric Science, Cornell University, Ithaca, New York
b Department of Geography, University of Minnesota, Twin Cities, Minneapolis, Minnesota
c Division of Ocean and Climate Physics, Lamont–Doherty Earth Observatory, Columbia University, Palisades, New York
d Climate and Large Scale Dynamics Laboratory, National Center for Atmospheric Research, Boulder, Colorado
e NASA Goddard Institute for Space Studies, New York, New York
f Department of Geography, Dartmouth College, Hanover, New Hampshire
g Bren School of Environmental Science and Management, University of California, Santa Barbara, Santa Barbara, California

(Manuscript received 7 March 2017, in final form 11 September 2017)

ABSTRACT

The western United States was affected by several megadroughts during the last 1200 years, most prominently during the Medieval Climate Anomaly (MCA; 800 to 1300 CE). A null hypothesis is developed to test the possibility that, given a sufficiently long period of time, these events are inevitable and occur purely as a consequence of internal climate variability. The null distribution of this hypothesis is populated by a linear inverse model (LIM) constructed from global sea surface temperature anomalies and self-calibrated Palmer drought severity index data for North America. Despite being trained only on seasonal data from the late twentieth century, the LIM produces megadroughts that are comparable in their duration, spatial scale, and magnitude to the most severe events of the last 12 centuries. The null hypothesis therefore cannot be rejected with much confidence when considering these features of megadrought, meaning that similar events are possible today, even without any changes to boundary conditions. In contrast, the observed clustering of megadroughts in the MCA, as well as the change in mean hydroclimate between the MCA and the 1500–2000 period, are more likely to have been caused by either external forcing or by internal climate variability not well sampled during the latter half of the twentieth century. Finally, the results demonstrate that the LIM is a viable tool for determining whether paleoclimate reconstructions events should be ascribed to external forcings or to “out of sample” climate mechanisms, or if they are consistent with the variability observed during the recent period.

1. Introduction

Megadroughts are prolonged periods of aridity typically defined by their multidecadal duration (Woodhouse and Overpeck 1998; Ault et al. 2014; Cook et al. 2016) and have been linked to the demise of several preindustrial civilizations (Benson et al. 2002, 2003; Hodell et al. 1995; Buckley et al. 2010; Stahle and Dean 2011). Mounting evidence suggests that the risk of these events is increasing throughout the southwestern United States and other subtropical dry zones due to rising temperatures and dynamic circulation changes (Seager et al. 2007b; Woodhouse et al. 2010; Ault et al. 2014, 2016; Cook et al. 2015, 2016; Kelley et al. 2015). If a megadrought were to unfold in the coming decades, it would require water management decisions to be made on scales that have no historical precedent. Understanding the risks of these events is therefore essential for supporting regional adaption strategies in the face of a changing climate. Furthermore, quantitative assessments of megadrought risk require us to disentangle the influences of external and internal climate variability in the past because future likelihoods will be governed by both of these factors (Ault et al. 2014, 2016; Cook et al. 2015).

The last millennium affords some insights into the range of forced and unforced preindustrial climate variability, yet it is only one realization of many plausible
alternative climate trajectories that could have occurred during that same time period. If a given climate forcing (e.g., enhanced solar output) made megadroughts more probable in the past, but they failed to occur due to internal climate variability, then one might erroneously conclude that the forcing was not important to the statistics of megadrought on the basis of the one realization being examined. Alternatively, if megadroughts occur without any clear external forcing, yet are nonetheless more probable during certain climate regimes, one might correctly conclude that megadroughts do not require exogenous changes to climate boundary conditions, but would miss the critical piece of information that such forcings can make megadroughts more likely. These inherent constraints of having only one realization of reality have largely prevented a consensus view from emerging on whether or not megadroughts are forced [see the review by Cook et al. (2016)].

A number of studies have suggested that megadroughts in the western United States, and especially in the southwestern United States, require external forcing mechanisms to alter regional and global circulation patterns in ways that have not been observed over the historical period (e.g., Clement et al. 1996; Graham et al. 2007, 2011; Mann et al. 2009; Seager et al. 2007a). Evidence for this hypothesis stems in part from the observation that the Medieval Climate Anomaly (MCA; 801–1300 CE) occurred during a period of slightly enhanced solar activity (e.g., Lean 2000; Vieira and Solanki 2010) and saw at least three multidecadal intervals of aridity with drought area index (DAI) values above the worst droughts of the twentieth century (Cook et al. 2004, 2016; Coats et al. 2016b). Idealized climate model experiments support this interpretation to varying degrees: when the Cane–Zebiak model (Zebiak and Cane 1987) of tropical Pacific Ocean dynamics is forced with increased solar activity, it simulates a “La Niña–like” mean state (Clement et al. 1996). Prolonged La Niña–like sea surface temperatures (SSTs), in turn, favor relatively arid conditions across the southwestern quadrant of the United States (Seager et al. 2007a; Burgman et al. 2010). In addition to external forcing factors, there is also evidence that regional feedbacks related to dust and land surface dynamics could make megadroughts more probable when the mean state of the western U.S. hydroclimate is drier than today (e.g., Cook et al. 2011, 2016).

Under the interpretations above, prolonged and widespread megadroughts arise from climate regimes that are inherently out-of-sample realizations of forced hydroclimate states not observed during the relatively short instrumental era. Analysis of unforced fully coupled general circulation model (GCM) simulations, in contrast, suggests that megadroughts could occur as a consequence of internal climate variability, if given a sufficiently long period of time. For example, Hunt (2006) argued that megadroughts should be interpreted as normal events in the unforced statistics of a long (10,000-yr) control run. Likewise, Meehl and Hu (2006) linked megadroughts in the U.S. Southwest to low-frequency modes of climate variability using a millennial-scale integration of a moderate-resolution GCM. More recently, Coats et al. (2013b, 2015a,b, 2016a,b) and Stevenson et al. (2015) argued that megadroughts occur as a result of internal variability, and hence they do not require an exogenous mechanism. Furthermore, fully coupled GCMs forced with realistic, time-evolving boundary conditions of the last millennium do not universally simulate clusters of megadroughts during the MCA (Cook et al. 2016; Coats et al. 2016a,b). This result could be interpreted as evidence that models fail to simulate forced megadroughts, or alternatively that these events occurred more frequently during the MCA due to chance (Coats et al. 2016b).

The GCM-based studies above are inherently limited by the rarity of megadroughts even in long control simulations, and by structural biases in GCM moisture fields (Ault et al. 2012; Flato et al. 2013) and teleconnections (Coats et al. 2013a). Even the largest ensembles (e.g., Kay et al. 2014; Otto-Bliesner et al. 2016) only include a few dozen realizations. If megadroughts tend to occur at a rate of 1–2 events per millennium [as originally proposed by Woodhouse and Overpeck (1998) and supported analytically by Ault et al. (2014)], then independent millennial-scale realizations still yield a relatively small sample size to assess megadrought statistics [as does a 10,000-yr control run as in Hunt (2006)]. Likewise, GCMs are known to have systematic biases both in their mean precipitation (Flato et al. 2013), as well as their tropical Pacific variability (Guilyardi et al. 2009; Ault et al. 2013b). Together, these biases imply that megadrought statistics estimated from GCMs are an imperfect approximation of the “true” limiting behavior of the system.

Here we develop an empirical approach for ruling out the possibility that megadroughts could arise from unforced, stationary statistics. Specifically, we use a linear inverse model (LIM) framework to emulate the observable spatiotemporal patterns inherent to the climate system over the instrumental period, when ocean–atmosphere data are widespread and of high quality. In climate studies, LIMs have been used to characterize whether or not the recent clustering of central Pacific El Niño events (also called “Modokis”) is anomalous in the context of El Niño–Southern Oscillation (ENSO) dynamics as they evolve naturally over the course of
centuries (Newman et al. 2011). It has also been applied to test the significance of the substantial low-frequency (decadal to centennial) variability in the tropical Pacific (Ault et al. 2013b) and to evaluate significance of “peaks” in the power spectrum of Pacific decadal oscillation (PDO) reconstructions (Newman et al. 2016).

Our null hypothesis is that megadrought characteristics in the western United States are consistent with those of a stochastically forced, linearly damped system with stationary statistics. That is, over any given millennium, the interannual variability in global SSTs and western U.S. (WUS) hydroclimate, combined with spatial and temporal sources of autocorrelation, will produce megadrought statistics that are indistinguishable from reconstructions. We use an LIM to populate the null distribution associated with this hypothesis, which in turn permits us to test the significance of the following characteristics of WUS megadrought (Fig. 1).

1) Events were widespread, encompassing between 20% and 40% of the domain at their peak (Fig. 1a; Cook et al. 2004).
2) The worst event in the southwestern United States occurred during the late 12th century and was as severe as the interannual droughts of the twenty-first century, but persisted for decades (Cook et al. 2004, 2016). This event can be clearly identified in the 35-yr running mean of southwestern U.S. Palmer drought severity index (PDSI) in Fig. 1b (red circle).
3) A “cluster” comprising five megadroughts occurred during a 500-yr period corresponding to the MCA (gray vertical bars in Fig. 1; Coats et al. 2016b; Cook et al. 2016).
4) And, finally, the mean hydroclimatic state of the Southwest was drier on average during the MCA than any other subsequent period of time (red line in Fig. 1b; Cook et al. 2004, 2016; Graham et al. 2007; Coats et al. 2016b).

The list above serves as a basis for developing megadrought test statistics, the significance of which can then be evaluated against the LIM’s distribution of these same quantities. For example, if LIM realizations regularly reproduce the spatial scale, magnitude, and clustering of megadroughts, as well as changes in the long-term mean values of reconstructed PDSI, then we would not be able to reject the null hypothesis. Failing to reject this null hypothesis would imply that megadrought occurrence does not require external forcing mechanisms or climate behaviors that are any different from those observed during the twentieth century (e.g., a multidecadal warm phase of the Atlantic multidecadal oscillation; Feng et al. 2008; Oglesby et al. 2012; Coats et al. 2016b; Wang et al. 2017). Rejecting the null hypothesis, on the other hand, would imply that the processes responsible for generating multidecadal periods of aridity in the western U.S. hydroclimate originate from either external forcings (e.g., solar variability,
explosive volcanism, or slow-varying orbital changes), from low-frequency climate processes that are not well sampled over the twentieth century, or from nonlinear coupled processes and feedbacks at multidecadal and longer time scales. Accordingly, applying our null hypothesis to the list of western U.S. hydroclimate characteristics from the last 1200 years will help clarify which aspects of megadrought are fundamentally unusual, versus merely unexpected.

Finally, for clarity and consistency with previous studies, we make a distinction between hydroclimate of the entire western United States (WUS; 32°–50°N, 125°–105°W), and the southwestern (SWUS) portion of that domain (32°–41°N, 124°–105°W). The dual foci of this study therefore permits analysis of both the spatial scale of megadroughts across the WUS and their severity and duration in the SWUS, the latter features having received somewhat more attention in the literature than the former. Moreover, among the leading modes of variability in WUS hydroclimate is a “dipole” pattern with dry and wet anomalies of opposite sign between the Pacific Northwest and the SWUS (e.g., Ropelewski and Halpert 1986; Cook et al. 2017), which further motivates analysis of both domains because megadroughts conditions in the SWUS are not necessarily guaranteed to be representative of WUS-wide conditions.

2. Linear inverse modeling

A linear inverse model presents a simple method to emulate the statistical characteristics of certain dynamical systems (Penland and Matrosova 1994; Penland and Sardeshmukh 1995). In addition to the examples cited in the previous section, the LIM approach has been used to make seasonal predictions of ENSO (Penland and Sardeshmukh 1995; Newman et al. 2011), to quantify predictability of the PDO (Newman 2007; Alexander et al. 2008; Newman et al. 2016), to characterize tropical Pacific ocean–atmosphere coupling across time scales (Newman et al. 2009), and to evaluate decadal forecasts made by fully coupled global climate models (Newman 2013). This suite of studies demonstrates that seasonal to decadal SST variability is well approximated by an LIM (Penland and Sardeshmukh 1995; Newman 2007, 2013; Alexander et al. 2008; Smirnov et al. 2014). Because SSTs, and especially tropical Pacific SSTs, exert a major influence on WUS hydroclimate (Ropelewski and Halpert 1986, 1996; Redmond and Koch 1991; Brown and Comrie 2004), we extend the LIM approach to emulate drought statistics in the WUS and the influence of SSTs on this region across seasonal to centennial time scales.

Using an LIM to generate null distributions for the statistics of widespread megadrought is similar to using an AR(1) model to evaluate the significance of behavior in a scalar time series (e.g., Bartlett 1978). The key difference is that the AR(1) model applies only to a univariate setting (\(x_t\)), with variations governed by an autocorrelation term (\(\rho\)) and stochastic white noise (\(\eta\)), such that \(x_t = \rho x_{t-1} + \eta\); it therefore cannot incorporate information about both the spatial and temporal scales of autocorrelation. An LIM, in contrast, is appropriate for simulating the variability produced by any multivariate system that evolves according to the following Langevin equation for an Ornstein–Uhlenbeck process (Penland and Sardeshmukh 1995):

\[
\frac{dX}{dt} = LX + \zeta,
\]

where \(X\) is an \(m\)-dimensional state vector and \(\zeta\) is white-noise forcing term, which itself may have spatial (or temporal) structure, but whose realizations are uncorrelated from one time step to the next. The term \(L\) is an \(m \times m\) linear deterministic feedback matrix, which includes information about how stochastic excitations propagate (linearly) through the system. If the system is indeed approximately linear, the eigenvalues of \(L\) must have negative real parts; otherwise, it would contain modes that grow exponentially and violate our assumptions about stable linear statistics.

The deterministic linear feedback matrix, \(L\), can be estimated empirically from data as

\[
L = n^{-1} \ln \{C_nC_0^{-1}\}.
\]

Here, the term \(n\) is a time lag parameter, \(C_n\) is the lag covariance matrix (at lag \(n\)), and \(C_0\) is the instantaneous covariance matrix. To the extent that the system is approximately linear, the choice of \(n\) does not critically alter the structure of \(L\) because, by definition, the most probable state vector (\(X\)) at some lead time \((t+n)\) is given by Penland and Sardeshmukh (1995) as

\[
X_{t+n} = \exp(Lt)X_t.
\]

In a typical application of an LIM, the noise term (\(\zeta\)) is not known, but instead its covariance matrix (\(Q = \langle \zeta \zeta^T \rangle dt\)) is estimated as a residual using the fluctuation–dissipation theorem (Penland and Matrosova 1994):

\[
\frac{dC_0}{dt} = LC_0 + C_0L^T + Q.
\]

Further information on the assumptions, limitations, and details of implementing an LIM for any generalized system with approximately linear dynamics can be found in Penland and Sardeshmukh (1995) and Penland and Matrosova (1994). Importantly, LIMs partition variability into an approximately linear part, represented...
by \( L \), and a noise forcing part \( Q \). The variables contributing to the noise term \( Q \) might exhibit nonlinear behavior at short time scales, and so an LIM approach is most appropriate for situations where there is temporal averaging inherent to the system of interest \( X \). Additionally, the time series included in \( X \) are usually smoothed to emphasize the approximately linear sources of variability in the system (Penland and Sardeshmukh 1995). This averaging (and smoothing), in turn, causes multiple realizations of the nonlinear phenomena to be integrated into each time step, and hence exhibit Gaussian statistics via the central limit theorem (Yuval and Hsieh 2002). This assumption is demonstrably valid for monthly SST anomalies, as shown by many previous studies across weekly to multidecadal time scales (e.g., Penland and Sardeshmukh 1995; Newman et al. 2009, 2011; Newman 2013). As with SST, the drought indices of interest here integrate monthly moisture supply and demand, and hence their variations are well suited to be simulated by an LIM approach.

Given the similarity between Eq. (1) and the equation for univariate red noise, LIM output can be thought of as “multivariate red noise” (Newman et al. 2011; Ault et al. 2013b). Unlike univariate red noise, which smoothly increases in amplitude from short to long time scales (Bartlett 1978), multivariate red noise can be used to generate time series characterized by strong spectral peaks, as is the case in the tropical Pacific. These peaks arise from the linear dynamics of the system, even when it is forced by stochastic white noise, and hence yield an important baseline for variations generated in regions affected by phenomena dominated by quasi-periodic variability (e.g., ENSO).

3. A linear inverse model of western U.S. drought

a. Data

We use global SST anomalies and PDSI values to construct an LIM of the surface ocean and WUS hydroclimate. Monthly SST anomalies from 60°S to 75°N originate from the Kaplan et al. (1998) dataset. This product is generated at a native spatial resolution of \( 5° \times 5° \) (500 km) lat/lon using a statistical procedure that estimates missing data through time from spatial empirical orthogonal functions (EOFs), statistical interpolation, and Kalman filtering (Kaplan et al. 1998). Although data are available as far back as 1856, we only use data from 1950–2000 for training the LIM. It is also detrended to minimize the contribution of anthropogenic forcing over that period, and smoothed prior to estimating \( L \) (see section 3b).

Hydroclimate data for the WUS are obtained from the “self-calibrating” PDSI dataset developed and published by Sheffield et al. (2006). This version of the PDSI (termed scPDSIpm) was generated using the Penman–Montieth method for calculating potential evapotranspiration, and the University of East Anglia’s (UEA) Climate Research Unit’s (CRU) TS3.10 precipitation data product as input, as well as near-surface pressure, humidity, and radiation fields derived from reanalysis and remote sensing data (Sheffield et al. 2006). For comparison of the widespread multidecadal megadrought intervals identified by Cook et al. (2004), we restricted our use of this scPDSIpm database to the WUS region (32°–50°N; 125°–105°W). Again, the period from January 1950 through December 2000 was used to train the LIM. Sensitivities of our LIM to the underlying data are presented at the end of the section 4.

b. Data preparation

Following prior LIM studies (Penland and Sardeshmukh 1995; Newman et al. 2011), principal components (PCs) were used to reduce the dimensionality of each field, and a 3-month moving average was applied to smooth each time series prior to computing the PCs. The first 11 PCs were retained from global SST [as in Newman (2013)] and jointly account for 60% of the total variance in that field, although the RMSE between the filtered and raw data, as well as the total fraction of shared variability, can be locally much higher (Fig. 2). While the smoothing has the benefit of stabilizing our estimate of \( L \) (i.e., making it less sensitive to noise), an obvious limitation is that the LIM cannot reproduce 100% of the variance at all locations (Fig. 2). This limitation is not a serious one for our applications because, locally, at least 90% of the variability in the tropical Pacific and the WUS is captured by the LIM. In other applications, however, such as seasonal prediction, this limitation could present problems if we needed high-accuracy LIM forecasts of future hydroclimate states.

The SST portion of the LIM preserves information about the growth and decay structures of tropical Pacific ENSO variability. Hence, it produces realistic interannual variability and Niño-3.4 power spectra, both of which are critical for emulating hydroclimate statistics in the WUS. The characteristics of the LIM were not particularly sensitive to this choice, and a somewhat larger or smaller number of PCs could have been retained to yield similar results.

In selecting the number of PCs to retain from the eigenvector decomposition of scPDSIpm, we required the total amount of variance in the field be comparable to the full range of the raw dataset (otherwise, any inferences about the statistics of megadrought in the LIM would have limited relevance to paleoclimate records of these events). We therefore retained the leading 22 PCs from scPDSIpm, which accounted for 90.8% of the raw (unsmoothed) variance of the WUS domain.
The final step in constructing \( \mathbf{X} \) entails normalizing the PCs from SST and scPDSIpm. We required that \( \| \text{SST} \| \gg \| \text{scPDSI} \| \) so that variability in the scPDSIpm would not appreciably affect the structure of \( \mathbf{L} \). Physically, this would be interpreted to mean that SST variability and regional sources of atmospheric noise can force scPDSIpm in the LIM, but scPDSIpm cannot affect SST. By requiring \( \text{SST} = 1 \) and \( \| \text{scPDSI} \| = 10^{-3} \), we found that the spatiotemporal structure of the SST patterns in \( \mathbf{L} \) were insensitive to scPDSIpm.

Our state vector is therefore defined as

\[
\mathbf{X} = \begin{bmatrix} T_0 \\ D \end{bmatrix},
\]

where \( T_0 \) and \( D \) are the filtered SST and scPDSIpm sets of PCs, respectively. The state vector defines the instantaneous covariance matrix: \( \mathbf{C}_0 = \langle \mathbf{X}^T \mathbf{X} \rangle \), where the angled brackets indicate averaging. The matrix \( \mathbf{C}_0 \) will have diagonal sections that are the covariances of the individual fields (i.e., \( C_{T_0} = \langle T_0^T T_0 \rangle \), and \( C_D = \langle D^T D \rangle \)), and off-diagonal sections that are the covariances between SST and scPDSIpm fields:

\[
\mathbf{C}_0 = \begin{bmatrix} C_{T_0} & C_{T_0,D} \\ C_{D,T_0} & C_D \end{bmatrix}.
\]

Finally, there is a well-known trend in global SSTs over the twentieth century, and possible (albeit more uncertain) trends in PDSI over the same period (e.g., Sheffield and Wood 2008; Sheffield et al. 2012; Dai 2011). Trends can influence the structure of \( \mathbf{L} \) because they will appear as weakly damped modes with decorrelation time scales proportional to the length of the
training interval (1950 to 2000 in this case). This component, in turn, could impart ultralow-frequency variability on the dynamical system that would be unrepresentative of its unforced statistics. We took several steps to guard against this possibility. First, SST and scPDSIpm fields were both linearly detrended prior to analysis. Second, patterns and time series of the PCs obtained from both fields were examined to ensure they did not include any “residual” trend. Third, the eigenvalues of L were inspected to make sure their decay time scales were relatively short (seasonal to interannual). And fourth, by focusing on the interval of 1950–2000, and especially by calibrating our LIM on the instantaneous and lag 3-month covariance matrices, we minimize the effects of secular and long-term variations because seasonal and interannual sources of variability dominate the fluctuations at these time scales.

The LIM was run for 12000 months (1000 yr) a total of 1000 times [using the formula provided by Penland and Sardeshmukh (1995), Eqs. (15a) and (15b) therein]. The PC time series generated from each of these stochastic realizations were projected onto the original subset of spatial loading vectors to generate spatially complete fields of each variable. Drought and tropical Pacific statistics (e.g., the DAI, spatial correlations, and power spectra) were then computed from this output. Although output from the LIM is available at monthly resolution, the PDSI field was averaged over June–August (JJA) for comparison with paleoclimate reconstructions and LIM output by calculating the 35-yr running average of PDSI as $PDSI_{35}$. The specific choice of 35 years for the time averaging is arbitrary; we adopt it here following earlier studies because it is simple and reliably identifies megadroughts as $-0.5\sigma$ anomalies in $PDSI_{35}$ preindustrial times in the WUS (Ault et al. 2014, 2016; Cook et al. 2016). It also satisfies earlier, more qualitative definitions of megadroughts as being periods of aridity as severe as the worst droughts of the twentieth century, but much longer lasting (e.g., Woodhouse and Overpeck 1998; Acuna-Soto et al. 2002), and it is amenable to analytical treatment [see the supplementary material of Ault et al. (2016)]. This quantity ($PDSI_{35}$) is computed from reconstructed JJA PDSI in the NADA and compared to the LIM output of JJA PDSI.

d. Megadrought test statistics

We identify megadroughts in the NADA and in the LIM based on the four features of these events described in the introduction (and summarized in Table 1). Megadroughts have been widespread, of comparable magnitude to the twentieth century in their severity, and clustered during the MCA. There has also been a shift in mean hydroclimatic state of 0.47 PDSI units since the MCA (Coats et al. 2016b), from relatively more arid to more humid modern conditions. Each of these characteristics are expressed as the quantitative test statistics described below:

- **Spatial scale.** We generate a multidecadal DAI from reconstructions and LIM output by calculating the fraction of WUS grid points with $PDSI_{35}$ at or below $-1$ PDSI units. To distinguish this quantity from the DAI employed by Cook et al. (2004), which first computed the fraction of the domain below $-1$ PDSI units, then smoothed the resulting time series, we term the DAI employed here “DAI$_{35}$” to emphasize its multidecadal dependence. Standard deviations of the reconstructed PDSI are close to 2 across much of the WUS, so that the threshold used to identify events ($-1$) is generally consistent with earlier studies that used a $-0.5\sigma$ threshold to identify events in time series normalized to unit variance over the historical period (e.g., Cook et al. 2004, 2016; Ault et al. 2014, 2016). The DAI$_{35}$ may therefore be interpreted as the fraction of the WUS experiencing megadrought lasting
35 years or longer at a given time. It could theoretically vary between zero (no grid points in megadrought) and 100% (the entire WUS in megadrought). LIM output is used to compute the DAi35, which is then compared against the NADA’s DAi35 (see below).

- **Magnitude.** The significance of multidecadal megadrought magnitude is assessed by comparing LIM output to the worst event in the mean (domainwide) SWUS PDSI35 between 801 and 2000 CE. This event reached a peak amplitude from 1124 to 1159 CE of −1.04 PDSI units on average (red circle in Fig. 1b).

- **Clustering.** Five events of relatively low amplitude (−0.5 PDSI units, or approximately −0.25σ) and long duration (35 yr or more) occurred during a 500-yr period of the MCA (Fig. 1b). We tally the total number of events of equivalent amplitude (−0.25σ) in each possible consecutive 500-yr interval of each LIM realization to evaluate the significance of the clustering observed in the MCA (applying this same criterion to the reconstructed PDSI identifies the five low-amplitude megadroughts of the MCA; Fig. 1).

- **Mean shift.** Mean PDSI values of the MCA were 0.47 PDSI units lower than they were during 1500–2000 CE. Differences between the first and last 500 years of each LIM realization are compared against this value.

e. **Evaluating significance**

We generate empirical p values for each test statistic above by comparing its reconstructed value against the distribution of those same quantities generated by the LIM. The p value is therefore the empirical probability that a given megadrought test statistic was drawn from the LIM’s distribution, and is defined as

\[
p = \frac{1}{\text{Total number of LIM realization}} \times \left[ 1 - \left( \frac{\text{LIM events less severe than those in NADA}}{\text{Total number of LIM realization}} \right) \right]. \tag{7}
\]

Empirical p values defined in this way will be close to zero for megadrought characteristics that are unlikely to have been drawn from the null distribution of the LIM, and larger (closer to unity) when associated with events for which the null hypothesis cannot be rejected.

For example, the most widespread event from 801 to 2000 CE reached peak spatial extent (maximum DAi35) over the period from 1130 to 1164 CE (Fig. 1a), with 40.8% of the domain experiencing megadrought conditions (PDSI35 below −1.0 PDSI units). We use the LIM to test if events of this scale are significant by computing a “fraction of time” test statistic (f*) over which megadrought conditions prevailed across some portion of the WUS domain. In the NADA, a total of 3 years (1/400 of all the years from 801 to 2000 CE) saw DAi35 values above 40%. The test statistic characterizing such an event at the DAi35 40% level would therefore be f* = 1/400. The values of f* can likewise be calculated for any other DAi35 threshold.

Output from the 1000 LIM realizations is used to generate the null distribution of f*. For each run, the fraction of years above various DAi35 values is computed, yielding an empirical probability density function (PDF) of f* based on the entire LIM ensemble. NADA-based values of f* are then ranked against this empirical distribution to establish p values, and hence confidence limits for rejecting the null hypothesis. For example, if a reconstructed value of f* falls outside the range of 95% of the LIM-generated data for that combination of area and fraction of time, we would reject the null hypothesis at the 95% confidence limit.

As with our significance testing of NADA-based values of f* for the entire WUS domain, we evaluate each of the three other test statistics for the SWUS: the magnitude of SWUS PDSI35 minima (m*), clustering of (low-amplitude) megadroughts during a 500-yr period (c*), and the change in mean PDSI (Δμ*). Because the
value of $m^*$ is negative ($-0.62\sigma$), we evaluate significance by comparing the distribution of the LIM with minimum $\text{PDSI}_{35}$ values below this threshold. In contrast, the number of low-amplitude events ($-0.25\sigma$ anomalies in $\text{PDSI}_{35}$) that occur during a given 500-yr period ($e^*$) is a positive integer. Significance is therefore assigned by tallying the number of LIM realizations with as many or more low-amplitude events over any continuous 500-yr period. Finally, if the shift in the 500-yr mean of the MCA was generated merely by chance (following our null hypothesis), then its sign is necessarily arbitrary. Significance is therefore evaluated against the absolute value of all differences between the first and last 500 years of all realizations.

4. Results

Using an LIM that represents global SST variability and drought in the western United States, we generated 1000 synthetic PDSI and SST realizations that each span one millennium. The LIM reproduces realistic spatial correlation patterns between the tropical Pacific and WUS hydroclimate (Fig. 3). Correlations patterns between the January–March (JFM) Niño-3.4 time series (average SSTs over 5°S–5°N, 170°–120°W) and June–August (JJA) scPDSIpm (Sheffield et al. 2006) are very similar in both the observations and the LIM output (Fig. 3).

Power spectra of the Niño-3.4 time series generated by the LIM exhibit a broad spectral peak over the interannual time scales associated with ENSO (gray shading, Fig. 4). Indeed, the range of LIM power spectra encompasses three different observational versions of Niño-3.4 (Fig. 4): Kaplan SST (Kaplan et al. 1998), ERSSTv3b (Smith et al. 2008), and HadSST1 (Rayner et al. 2003), which verifies that this version of the LIM generates realistic interannual variability [a result also discussed in Newman (2013) and Penland and Sardeshmukh (1995), among others]. Although one observational product (ERSSTv3b) exhibits spectral densities above the LIM at near-centennial time horizons, this low-frequency variability probably reflects long-term trends in the dataset that are not robust across the tropical Pacific (e.g., Solomon and Newman 2011; Emile-Geay et al. 2013a,b).
To help visualize the output generated by our LIM of global SSTs and western hydroclimate, we illustrate the timing and spatial structure of megadroughts in just one LIM realization (Figs. 5 and 6). In this run, the maximum WUS DAI35 and minimum SWUS PDSI35 values coincide, but this does not occur in all realizations, and high DAI35 values are not always driven by the same SST patterns as low SWUS PDSI35 values. During the first widespread event (culminating at year 54), DAI35 values reached 27.1% (Fig. 5a). Other comparable periods are also present in this run (e.g., years 230–270, 310–390, and 950–970). Here we focus on the year 54 event (vertical dashed line) to distinguish its SST pattern from that of the most severe event (Fig. 6). SST anomalies were cool almost everywhere, and most notably in the northern Atlantic (Fig. 5b). Negative values are apparent in much of the central portion of the domain, with wet anomalies occurring in the Northwest and along much of the Pacific coast (Fig. 5c). The most severe event in the Southwest reached a value of \(-0.72\sigma\) (red circle in Fig. 6; LIM year 946). Average SST anomalies in the LIM during that time were close to neutral in the Indian Ocean, cool in the tropical Pacific, and warm in the western Pacific and North Atlantic (Fig. 6b). Negative PDSI values (between \(-1\) and \(-2\)) are found throughout most of the entire Southwest, and most especially in the central part of the domain. Modest positive values are present across the northern regions of the WUS (Fig. 6c).

DAI35 time series from individual LIM realizations encompass a similar range as the tree-ring-based reconstructions (Fig. 7). Maps of the most widespread events (maximum DAI35) likewise depict prolonged events commensurate with observations in their spatial scale and severity (Fig. 8). Several of these realizations have at least one event with DAI35 values greater than or equal to the maximum reconstructed value (40.8% of WUS in the NADA). There are also instances of widespread megadroughts that are even more spatially extensive than have been observed (e.g., run 2). In total, 35.4% of the runs included at least one event as widespread as the 1130–1164 megadrought in the NADA.

Average SST anomalies from the 35 years associated with each run’s maximum DAI35 are reminiscent of the canonical patterns associated with La Niña and the “cold phase” of the PDO with cool SSTs in the Indian Ocean, equatorial Pacific, and along the western edges of North and South America (Fig. 9). Warm SST anomalies are found in the northwestern Pacific and North Atlantic. The amplitudes of the averaged anomalies,
however, are much weaker than during La Niña or PDO cold phases (individual events can have stronger anomalies relative to composites). The westward extent of the cold anomaly along the equator is also less zonal than it is during normal La Niña events.

Global average SST anomalies associated with the lowest values of $P_{D35}$ from each run (Fig. 10) differ from those connected with maximum $DAI_{35}$ (Fig. 9). In particular, cool SST anomalies along the equatorial Pacific extend less westward in the maximum $DAI_{35}$ pattern than they do at times of minimum $P_{D35}$ (cf. Figs. 9 and 10). Moreover, SST anomalies associated with SWUS megadrought have a more distinctively La Niña–like pattern, as compared with the anomalies associated with widespread drought.

Distributions of our test statistic for widespread WUS megadrought ($f^*$) from LIM output establish confidence limits for various reconstructed $DAI_{35}$ thresholds (Fig. 11). In this figure, each panel shows the LIM-based PDF of time ($f^*$) spent at or above each $DAI_{35}$ threshold in 5% increments. The vertical dashed line marks the fraction of time in the NADA spent at or above these thresholds; the gray shaded area of each PDF identifies the portion of LIM realizations that have a lower fraction of years than the NADA at each threshold. For example, in the NADA 1072 out of 1200 years saw $DAI_{35}$ values of at least 5% ($f^* = 0.893$). Of the 1000 LIM realizations, 81.3% simulated a smaller fraction of time below this threshold than in the NADA, meaning that the probability ($p$ value) that the observed value at least as large as $f^*$ was drawn from the null distribution is 0.187 ($p = 1 - 0.813/1000$).

The shape of the distributions for $f^*$ is clearly more skewed for higher $DAI_{35}$ thresholds because most of the realizations only simulate a small fraction of years with $DAI_{35}$ values greater than 35% (Fig. 11). The LIM also produces a small number of events (not shown) with even greater values (i.e., the $DAI_{35}$ thresholds of 45% and 50% both include a few nonzero values when considering the full 1000-member ensemble).

None of the reconstructed $DAI_{35}$ values are significant at the 95% confidence limit (Fig. 12). In this case, we applied a one-sided significance test because values of $f^*$ can only be positive, and we are interested in the fraction of LIM realizations with more time spent below various $DAI_{35}$ thresholds. A significance level value of

![Figure 5](https://example.com/f5.png)
0.3 (y axis of Fig. 12) means that 3/10 of all LIM realizations simulated less time at a given DAI35 level than in the NADA, while 7/10 spent more time at that threshold. Even the most widespread reconstructed event (1130–1164 CE), when approximately 40.8% of the WUS was affected by megadrought, is not significant at the 90% confidence limit \((p \text{ value } = 0.354)\).

We assign significance to the other test statistics \((m^*, c^*, \text{ and } D_m^*)\) as follows:

- **Magnitude.** During the worst SWUS drought, tree-ring reconstructed PDSI35 reached \(-0.62\sigma (m^* = -0.62\sigma)\). Out of the 1000 LIM realizations, 34.8% of the runs included events as severe or worse (more negative) than this period (Fig. 13), while the other runs did not include such high magnitude events. The null hypothesis can only be rejected at about the 65% confidence limit \((p \text{ value } = 0.348)\) for the magnitude of the worst megadroughts in the SWUS over the last 1200 years.

- **Clustering.** Five low-amplitude \((\text{PDSI}_{35} \leq -0.25\sigma)\) megadroughts occurred during the MCA (Fig. I in Coats et al. 2016b), which is more than occurred during most 500-yr periods of the LIM (Fig. 14). In total, 2.1% of LIM runs had 500-yr periods with as many or more megadroughts, suggesting that the null hypothesis for megadrought clusters \((c^*)\) can be rejected at the 90% confidence level \((p \text{ value } = 0.021)\).

- **Mean shift.** The long-term difference in the MCA mean versus the 1500–2000 CE mean (0.47 PDSI units) is highly significant against the null hypothesis \((p \text{ value } < 0.001)\). Out of the 1000 LIM realizations, none had higher magnitude 500-yr mean shifts, hence the reconstructed values are significant at the 99.9% confidence limit or higher \((p \text{ value } < 0.001)\).

Our results were robust with respect to a number of other alternative assumptions made as part of constructing of the LIM, identifying megadrought, and evaluating significance. For example, other calibration periods (e.g., 1980–2010) yielded structurally similar linear deterministic matrices, meaning that estimating \(L\) is not particularly sensitive to the 1950–2000 interval used here. Likewise, the scPDSIpm product we included in the LIM uses precipitation from the CRU TS3.10 data product [for details, see Sheffield et al. (2006)].
are 15 alternative PDSI products, however, including ones with different underlying precipitation inputs and approaches to estimating evapotranspiration (Sheffield et al. 2006). Use of the other 15 versions of PDSI did not appreciably change the structure of the LIM or the confidence limits we established. Furthermore, the scPDSIpm differs from its predecessor, the PDSI, in that the local surface storage coefficient is estimated empirically from data, rather than prescribed as a spatially invariant constant. Use of the original PDSI did not critically alter the nature of variability in the LIM, nor did use of alternative SST datasets (e.g., ERSSTv3b; Smith et al. 2008).

There was not any strong dependence of the LIM on the details of the detrending; use of a spline fit, high-pass filter, or linear fit all yielded structurally similar LIMs. We did find, however, that failing to detrend the data had a modest impact on the significance levels, in that it made more prolonged and more widespread events somewhat more common in the LIM. The likely explanation for this particular sensitivity is that the global warming pattern becomes included in the LIM as a “least damped” mode that grows and decays on multi-decadal time scales, imparting additional low-frequency variability. The limitation of detrending is, of course, that there is some possibility that we have inadvertently removed unforced low-frequency climate variability. If this is the case, rejecting the null hypothesis would be even more difficult, as it would raise the confidence limits even further. Our results should therefore be interpreted as being slightly “conservative” because of our use of detrended data in constructing the LIM. Note, however, that the tree-ring data are also detrended, so there is also a possibility that the true amplitude of climate fluctuations is greater in reality than inferred from the reconstruction (see discussion). In either case, the results presented here are at least consistent, in that both the LIM and the tree-ring data have had long-term trends removed.

5. Discussion

Several key features of Pacific SST variability and its impact on WUS hydroclimate can be reproduced by the relatively simple, low-order linear dynamical system represented by the LIM. For example, the correlation patterns between the Niño-3.4 time series and JJA PDSI are nearly identical between observations and the LIM.
Likewise, the LIM simulates spectral densities in the central Pacific Ocean that encompass the range of observational estimates (Fig. 4). Both of these features are important for verifying that the LIM’s underlying statistics in space and time are representative of observations over the short instrumental period, and hence relevant to the longer time scales of megadrought.

The LIM generates synthetic megadroughts, but this outcome was not guaranteed from its formulation because we used only seasonal data from the late twentieth century (an interval when such events have not been observed). We also emphasize the LIM did not include any information from tree-ring reconstructions of past droughts in the western United States. Megadroughts in the LIM are therefore to be regarded as an emergent feature of the low-order dynamical system we have investigated. This finding could be further used as a basis for predicting an optimal forcing pattern (or combinations of patterns) that would elevate the risk of widespread megadrought if it were to appear across ocean basins this century.

We used the LIM to evaluate our null hypothesis, namely that widespread WUS megadroughts occur fundamentally from stochastic forcing. To test the significance of reconstructed events against this null hypothesis, we compared the fraction of years spent below each DAI$_{35}$ threshold ($f^*$; see section 3e) in the NADA to the LIM’s distribution at those same levels. This comparison, in turn, permits us to establish confidence limits for rejecting the null hypothesis across reconstructed DAI$_{35}$ thresholds. For instance, the maximum DAI$_{35}$ value of the NADA is 40.8%, and just 3 years over the
last 1200 years were spent at or above the 40% threshold. The fraction of time megadrought conditions prevailed in the NADA at this level is therefore 3/1200 (0.0025). While 64.6% of the LIM realizations did not yield DAI35 values this high, 35.4% of the runs include megadroughts that are equally (or more) widespread (Fig. 12). We may therefore only reject our null hypothesis at the nominal 64% confidence level. Similar considerations apply to other lower thresholds, none of which may be said to be significant at the 95% confidence limit.

As with the spatial scale of reconstructed WUS droughts, the null hypothesis cannot be rejected with very much confidence for the magnitude ($m^*$) of the worst droughts in the SWUS. That is, about a third (34.8%) of the LIM realizations generated events as severe as, or worse than, the worst drought in the SWUS. Hence, such events should be considered relatively normal occurrences in the hydroclimatic statistics of the region on millennial time scales.

Two features of SWUS megadrought are more significant than the scale ($f^*$) and severity ($m^*$) of past events. First, the clustering ($c^*$) of low-amplitude megadroughts during the MCA is unlikely to have occurred from chance alone ($p$ value = 0.021), and hence we reject the null hypothesis for this characteristic of SWUS hydroclimate with fairly high (>95%) confidence. We can be even more confident in rejecting the null hypothesis for the shift in 500-yr mean reconstructed PDSI values ($p$ value < 0.001), implying that such long-term changes are extremely unlikely to be driven by chance alone.

What does accepting the null hypothesis for two of the four megadrought test statistics, and rejecting it for the other two, mean for the nature of hydroclimate in the WUS and SWUS? First and foremost, our results help...
clarify what aspects of megadrought occurrence in the past are truly exceptional given what we can observe in the modern system. Clusters of megadroughts, and especially changes in long-term means, are unlikely to be generated by stationary statistics and stochastic forcing. Their existence in the NADA therefore requires further examination. One possibility is that their odds of occurrence were externally forced during the MCA [e.g., as discussed extensively in Cook et al. (2016)], either by slightly enhanced summer insolation (Lean 2000) or by coupled feedbacks between the land surface, vegetation, and atmosphere (Cook et al. 2011). Alternatively, a multicentury warm phase of the Atlantic multidecadal oscillation (AMO) (Coats et al. 2016b; Wang et al. 2017) could have shifted the mean state of hydroclimate in the SWUS and increased megadrought risk (Feng et al. 2008; Oglesby et al. 2012). Such persistence itself would be unusual in the context of seasonal dynamics of global SST over the late twentieth century, and could require either ocean–atmosphere coupling as originally proposed by Bjerknes (1969) or the influences of external forcings [including “quiet” periods of relatively low variability in forcings as described by Bradley et al. (2016)]. It is clear that a smaller fraction of total SST variance is captured by the LIM in the Atlantic than in the Pacific (Fig. 2). Consequently we could be under-sampling AMO variability in construction of the LIM due to the short time scale of the observational record. If the AMO is responsible for megadrought clustering during the MCA, then we cannot completely rule out the possibility of megadroughts being internally generated because our LIM does not fully characterize variability in the North Atlantic.

The change in the long-term mean of reconstructed PDSI is the most statistically significant feature of SWUS hydroclimate. If it is climatic in origin (Cook et al. 2004; Coats et al. 2016b), it too likely arises from either nonlinear or externally forced sources of variability. If it is not driven by climate, then it could arise from phenomena related to the proxy itself (e.g., biological

![Figure 11. LIM-based empirical distributions of widespread megadrought for various DAI35 thresholds. In each panel, the ordinate is the fraction of all years spent below a given DAI35 threshold (marked at the top of each panel in bold). The shaded portion of the PDF is the fraction of LIM realizations with a proportionally lower number of years spent below each threshold than the NADA, while the white shading is the remaining portion of the PDF. The vertical dotted line is fraction of time (f) in the NADA spent at or below each DAI35 threshold. The y axis is normalized so that the number of realizations (N) at each value of f is divided by the total number of realizations (r).](http://journals.ametsoc.org/doi/pdf/10.1175/JCLI-D-17-0154.1)
growth trends not completely removed through de-trending; Fritts 1976) or the methodology used to develop the reconstruction. For example, reconstructed PDSI values of the NADA are associated with calibration and measurement uncertainties, and the effect of these uncertainties on our rejection or acceptance of the null hypothesis cannot be directly addressed using the NADA. In contrast, more recent drought atlases (Cook et al. 2010, 2015) have been constructed using methods that explicitly generate an ensemble of probable reconstructions over the respective domains, and a similar approach could be incorporated to derive an ensemble of WUS reconstructions to compare against the LIM. Further work should therefore test ensembles of drought atlas reconstructions against our LIM-based significance levels, although we have focused here exclusively on the available expected value reconstructions over the NADA domain.

Notwithstanding the inherent methodological limitations of reconstructing past climates, the differences in the multicentury mean PDSI values between the modern period and the MCA are extremely unlikely to have arisen from stationary statistics. The LIM therefore provides important insight into the details of the data

---

**FIG. 12.** Empirical confidence limits for rejecting the null hypothesis that widespread megadroughts of the last millennium are caused by SST variability, combined with spatial and temporal autocorrelation. For each DAI₃₅ threshold (abscissa), the fraction of years in the NADA found at that value is compared against the total fraction of years in all LIM realizations. The line with open circles traces the significance limits for the amount of time spent at each DAI₃₅ threshold. The 50% (black), 66% (dashed), 90% (dash-dotted), and 95% (dotted) confidence limits are also shown for reference.

**FIG. 13.** PDF of normalized minimum PDSI₃₅ (m*), averaged over the SWUS. Black vertical bars indicate realizations with runs that produced minimum values below the NADA (−0.63σ).
that require further investigation even if they are related to the proxy or reconstruction methodology.

Under the changing climates of the present day, it is often stated that individual extreme weather events cannot be unambiguously attributed to climate change, but that the occurrence of more frequent extremes reflects the net effects of anthropogenic activity on the statistics of the climate system (e.g., Trenberth et al. 2015). A similar principle applies to megadroughts during the MCA: such events should occur naturally even without any changes to the boundary conditions (e.g., solar, orbital, or land cover) of WUS hydroclimate. However, their frequency from 801 to 1300 CE suggests that exogenous factors may have played a role in making megadroughts more probable during that time.

Our results suggest that the LIM is a viable tool to establish baselines (or expectations) when evaluating long model simulations or suites of paleoclimate reconstructions. For example, it is not entirely clear if GCMs simulate megadroughts at realistic rates with or without externally varying boundary conditions (e.g., Ault et al. 2014; Coats et al. 2013b). A LIM could be trained on model output, and the PDF of megadrought occurrence derived from it could be compared to the observation-based LIM we used. Such an analysis would therefore not be limited by the size of millennial-scale

![Figure 14](image1.png)

**Fig. 14.** Distribution of maximum number of low-amplitude megadroughts ($c^*$) to have occurred over a 500-yr period in each LIM realization. Black vertical bars identify the fraction of runs with 500 year periods that include an equal or greater number of events than occurred during the MCA (801–1300 CE), as inferred from tree-ring reconstructions.

![Figure 15](image2.png)

**Fig. 15.** PDF of SWUS shifts in 500-yr means, expressed in PDSI unit differences ($D_{\mu}$). The bars are white because all runs had smaller amplitude differences than inferred from the NADA for the SWUS between 801–1300 CE and 1500–2000 CE (this difference is marked by an asterisk on the x axis).
forcing mechanisms, or exceptionally persistent climate hypothesis, respectively. Accordingly, these characteristics nominally, and highly, significant against the null hypothesis change in the 500-yr mean since the MCA are megadroughts, the clustering of these anomalies and the magnitude of the most severe SWUS megadroughts.

6. Conclusions and implications

We constructed a simple, low-order empirical model (an LIM) of SST and WUS hydroclimate to evaluate the odds that megadroughts occur by chance—that is, arising from observable spatiotemporal modes of late-twentieth-century climate variability, stochastic forcing, and sufficient time. Despite being trained only on seasonal data from the late twentieth century, the LIM successfully reproduces realistic power spectra in the tropical Pacific and correlation patterns between the Niño-3.4 region and WUS hydroclimate. It also produces megadroughts that are comparable in their duration, spatial scale, and magnitude as the most severe events of the last 12 centuries (801–2000 CE). For example, the most widespread megadrought in the NADA lasted from between 1130 and 1164 CE and reached DAI35 values of approximately 40%. About 35% of LIM realizations included events of this scale or worse. Because the LIM is an empirical model of dynamics with fundamentally stationary statistics, we cannot reject the null hypothesis that even the most widespread megadroughts are internally generated. Similar considerations apply to other smaller DAI35 thresholds, and to the magnitude of the most severe SWUS megadroughts.

In contrast to the spatial scale and magnitude of megadroughts, the clustering of these anomalies and the change in the 500-yr mean since the MCA are nominally, and highly, significant against the null hypothesis, respectively. Accordingly, these characteristics of the NADA may require nonlinear processes, external forcing mechanisms, or exceptionally persistent climate states that have not been observed during the twentieth century [e.g., a multicentury warm phase of the AMO, as discussed by Coats et al. (2016b)]. These results help clarify which aspects of megadrought are unusual but unremarkable in the context of the observable system (widespread intervals and high magnitude events), and what characteristics are truly unusual and require further investigation (clustering and the lower mean of the MCA).

In summary:

- The most widespread megadrought in the last millennium occurred between 1130 and 1164 CE and encompassed 40% of the WUS domain. About 35% of the LIM output simulates megadroughts that are equally (or more) widespread (Fig. 12), meaning we cannot reject the null hypothesis of internal variability as causative at the 90% or 95% confidence limit ($p$ value $= 0.646$).

- The worst drought in the reconstructed SWUS PDSI average was $-0.62\sigma$. Because 34.8% of LIM runs had PDSI35 values below this level, we cannot reject the null hypothesis for the severity of the worst megadrought in the NADA ($p$ value $= 0.652$).

- During the MCA five low-amplitude megadroughts (PDSI35 $\leq -0.25\sigma$) occurred over a 500-yr period; only 2.1% of LIM runs had 500-yr periods with as many or more megadroughts, suggesting that we can reject the null hypothesis at the 95% confidence limit ($p$ value $= 0.021$).

- Mean PDSI values from 801 to 1300 CE were 0.47 PDSI units lower than from 1500 to 2000 CE. All of the LIM runs had smaller mean shifts between their first and last 500 years, meaning we can reject the null hypothesis with a high degree of confidence ($p$ value $< 0.001$).

Finally, our results hold a number of key implications for stakeholder and research communities alike. First, events as arid as and even more arid than the most widespread drought of the last 1200 years are possible, even without any changes to the boundary conditions. Given their spatial extent, these events would require large-scale efforts to manage water across much of the western United States, as well as to protect ecosystems and agriculture. A LIM could be used to support water resource management strategies because it permits us to generate a large number of statistically plausible outcomes even without climate change. These scenarios could, in turn, be used by policy makers and water resource managers to prepare for possible future water stress. Second, given that the LIM simulates events that are slightly worse than those in the NADA, and given that the NADA is based on tree-ring data that are

GCM ensembles because the LIM would empirically fill out the PDF of megadrought from a limited set of runs. Likewise, an LIM could be adapted to simulate the risk of widespread megadrought, building on the work of Ault et al. (2014) and Cook et al. (2015), both of which focused on single time series methods of estimating risk. Finally, and perhaps most importantly, this effort should motivate wider adoption of the LIM framework in paleoclimate significance testing. In high-resolution paleoclimatology, researchers frequently observe low-frequency variability in time series or networks of paleoclimate archives, and it is often unclear whether such fluctuations require external forcings (e.g., volcanic, solar, or land surface processes), “out of sample” climate mechanisms, or if they are simply consistent with the variability observed during the late twentieth century. Adapting an LIM to these settings would shed new light on the importance of internal variability and forced climate change across a broad range of reconstructed phenomena.

Ault et al. (2014) and Cook et al. (2015), both of which GCM ensembles because the LIM would empirically fill

ANU 2018 AULT ET AL. 21

1 JANUARY 2018 A U L T E T A L.
themselves detrended, prewhitened, and then “re-reddened” [see the supplementary material in Cook et al. (2004)], it is possible that there is more low-frequency variability inherent to the climate system than would be inferred from the NADA alone [see also Ault et al. (2013a)]. Even if this is the case, our results clearly show that we cannot rule out the possibility that such variability is internally generated by the climate system itself. Future widespread megadroughts are therefore likely to emerge as a combination of this internal low-frequency variability and long-term forced changes.

Acknowledgments. This material is based on work partially supported by NSF EaSM2 (Earth System Model 2) Grants AGS1243125 and AGS1243204 and NSF Grants AGS1602564, AGS1401400, and AGS1602512. BI Cook is supported by the NASA Modeling, Analysis, and Prediction program. LDEO Contribution No. 8165.

REFERENCES


