How Wet and Dry Spells Evolve across the Conterminous United States Based on 555 Years of Paleoclimate Data[✉]

MICHELLE HO

CSIRO Land and Water, Black Mountain, Canberra, Australian Capital Territory, Australia, and Columbia Water Center, Columbia University, New York, New York

UPMANU LALL

Columbia Water Center, and Department of Earth and Environmental Engineering, Columbia University, New York, New York

EDWARD R. COOK

Lamont–Doherty Earth Observatory, Columbia University, Palisades, New York

(Manuscript received 27 March 2018, in final form 5 May 2018)

ABSTRACT

Evolving patterns of droughts and wet spells in the conterminous United States (CONUS) are examined over 555 years using a tree-ring-based paleoclimate reconstruction of the modified Palmer drought severity index (PDSI). A hidden Markov model is used as an unsupervised method of classifying climate states and quantifying the temporal evolution from one state to another. Modeling temporal variability in spatial patterns of drought and wet spells provides the ability to objectively assess and simulate historical persistence and recurrence of similar patterns. The Viterbi algorithm reveals the probable sequence of states through time, enabling an examination of temporal and spatial features and associated large-scale climate forcing. Distinct patterns of sea surface temperature that are known to enhance or inhibit rainfall are associated with some states. Using the current CONUS PDSI field the model can be used to simulate the space–time PDSI pattern over the next few years, or unconditional simulations can be used to derive estimates of spatially concurrent PDSI patterns and their persistence and intensity across the CONUS.

1. Introduction

Persistent droughts and wet spells are recurrent phenomena in the conterminous United States (CONUS) that impact both natural ecosystems and developed regions (Boyer et al. 2013). Persistent droughts adversely affect agricultural production (Drabenstott and Barkema 1988; Hornbeck 2012; Wiener et al. 2016), limit electricity production (Webber 2008), strain water supply reserves (Booker 1995), drive human induced groundwater declines (Cooley et al. 2015; Ho et al. 2016; Russo and Lall 2017), and impact ecosystems (Silliman et al. 2005;

Pederson et al. 2006). Conversely, wet spells increase the probability of flooding events (Trenberth and Guillemot 1996; Mo et al. 1997) that may overwhelm critical infrastructure such as roads, water treatment facilities, electricity supply (Leavitt and Kiefer 2006), and dams (Yen and Tung 1993; Stanford University 2016), which may in turn spark cascading failures in previously protected flood plains (Pinter 2005; Ho et al. 2017). Examining drought and wet spell characteristics at a continental scale is critical to understanding how typical spatial patterns of both extremes evolve over time and across space and to inform regional planning and decision-making. In this paper we consider that there may be recurrent, large-scale spatial patterns of drought over the CONUS that evolve in response to larger-scale climate dynamics. We use a hidden Markov model (HMM) to identify these patterns and their annual transition probabilities. This potentially provides more useful

[✉] Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JCLI-D-18-0182.s1>.

Corresponding author: Michelle Ho, mh3538@columbia.edu

information into pattern classification than methods that would seek to cluster the time series or find lower-dimensional patterns (e.g., through EOFs) without also considering the temporal dynamics or memory.

Past endeavors to understand characteristics of large-scale dry and wet spells have used a suite of methods. These approaches include principal component analysis (PCA) to identify the main modes of continental-scale drought variability (Karl and Koscielny 1982; Cook et al. 1999) and location-specific trend analyses across large regions (Damberg and AghaKouchak 2014; Devineni et al. 2015; Sun and Lall 2015). Methods that simultaneously consider multiple drought characteristics, such as duration, severity, and recurrence, include Bayesian copulas (Kwon et al. 2016), severity-area-duration curves (Andreadis et al. 2005), and examination of persistence and recurrence of analogs of notable twentieth-century droughts within paleoclimate records (Fye et al. 2003).

Autoregressive models and HMMs typically have high parameter uncertainty when relatively short records are analyzed, and the underlying series have long memory. Adequately long time series are therefore needed to characterize such variability (Rodríguez-Iturbe 1969). Thyer et al. (2006) showed that at least 100–200 years of data are needed to characterize hydroclimatic variability that displays strong persistence of either dry or wet spells and even longer records are needed when persistence is weak. Ocean–atmospheric variability with periodicities that span multiple years to multiple decades such as the El Niño–Southern Oscillation (ENSO), Pacific decadal oscillation (PDO), and Atlantic multidecadal oscillation (AMO) have been shown to influence precipitation, streamflow, and drought in the CONUS region (Barlow et al. 2001; McCabe et al. 2004; Goodrich 2007). Consequently, relatively short instrumental records are a key limitation to assessing dry and wet spells in the CONUS on long (i.e., multiyear to multidecadal) time scales. Spatial information imparted by low-frequency climate signals that influence continental-scale moisture availability is not effectively used in autoregressive models. However, the HMM allows the possibility of effectively modeling the regime-like spatiotemporal behavior that one expects from the interpretation of the hydroclimatic teleconnections associated with these climate modes (Bracken et al. 2014, 2016).

Paleoclimate reconstructions offer an option for supplementing relatively short instrumental records by using proxies for climate that extend over multiple centuries. Annually resolved paleoclimate records allow for the examination of hydroclimatic variability over time scales relevant to water resource management

and planning and flood and drought risk mitigation. One such record is a reconstruction of the modified Palmer drought severity index (PDSI) across North America (Cook et al. 2010) known as the Living Blended Drought Atlas (LBDA), an update of the seminal North American Drought Atlas (NADA; Cook et al. 1999, 2007). The LBDA provides an interesting target for the development of a stochastic modeling capability for simulating spatially distributed national PDSI scenarios that can leverage the space and time information generated from the paleoclimatic reconstruction.

Our analysis contributes to understanding long-term CONUS-scale PDSI variability through the classification of spatially similar patterns of droughts and wet spells, similar to Fye et al. (2003) using the NADA. However, here an autoclassification scheme is used that models the temporal evolution of spatially distinct drought patterns given the propensity of dry and wet spells to evolve spatially to adjacent locations (Andreadis et al. 2005; Salas et al. 2005; Yoo et al. 2012; Hoerling et al. 2014). Including a temporal evolution component in the analysis enables the probabilities of transitioning from one state to another to be determined, and thus provides the basis for spatiotemporal stochastic simulation over the entire CONUS for the next 1 to 5 years or to explore the stochastic aspects of the dynamics over the period of record.

The HMM has been used extensively for downscaling precipitation from climate and weather model simulations (Bellone et al. 2000; Robertson et al. 2004; Mehrotra and Sharma 2007) and also for modeling paleoclimate time series (Prairie et al. 2008; Bracken et al. 2016). The key idea leveraged here is that multicentury records of spatially distributed CONUS climate may allow the identification of latent states that correspond to the hydroclimatic signature imparted by the underlying low-frequency modes of global climate variability. The joint consideration of both spatial and temporal evolution of CONUS dry and wet spells is motivated by multiyear and multidecadal ocean–atmosphere forcing that influences the occurrence of such events (e.g., Seager et al. 2005; McCabe et al. 2008). The climate states and resultant regional impacts identified by an HMM could then be used to explicitly model latent state transitions from year to year and conditionally simulate PDSI at each grid location of the LBDA. In this paper, we seek to diagnose the statistical properties of the resulting latent climate states and the associated spatially distinct patterns of droughts and wet spells over the past five and a half centuries using a paleoclimate reconstruction of PDSI.

2. Data

a. *The Living Blended Drought Atlas*

The LBDA is a paleoclimate reconstruction of summer (June–August) modified PDSI (Cook et al. 2010) that is spatially complete across the CONUS region from 1451 to 2005 (555 years) (see supplemental material for a more detailed description of the LBDA). The modified PDSI is a normalized measure of the balance between moisture supply and demand. The LBDA may therefore be used to assess both moisture deficits (i.e., droughts) and moisture surpluses (i.e., wet spells) and has been shown to also inform extreme precipitation over parts of the country (Steinschneider et al. 2016). Although the reconstruction is focused on the summer season, the PDSI has been shown to reflect moisture balance aggregates over approximately 12–18 months (Guttman 1998; Vicente-Serrano et al. 2010) and can therefore be representative of moisture conditions over the current and previous years (Dai et al. 2004; St. George et al. 2010).

The dimensionality of the dataset was reduced from 4643 grids (Fig. S1 in the supplemental material) to 222 grids for computational efficiency and in consideration of our interest in large-scale phenomena rather than local dynamics. The $0.5^\circ \times 0.5^\circ$ latitude–longitude grid was regridded on to a $2.5^\circ \times 2.5^\circ$ grid using the mean value to spatially aggregate data. We recognize that each reconstructed gridded PDSI value in the LBDA has an associated uncertainty because of the reconstruction process. This uncertainty is temporally and spatially variable, reflecting the underlying presence or absence of tree-ring data and their representativeness for PDSI reconstruction. It would conceptually be possible to consider these uncertainties, but in this initial effort we have only used the mean values of the reconstruction as reported in the LBDA. By spatially aggregating up to the coarser grid, some of this uncertainty is reduced through a reduction of the variance under averaging, but at the expense of specificity at each location. Thus, this is useful, given our interest in the CONUS-scale spatial patterns and their evolution, rather than location specific attributes. The grid resolution evaluated here should also be sufficient for our needs given that it is effectively the same as the original NADA (Cook et al. 1999) and still finer than the 60-point grid used by Karl and Koscielny (1982) to identify regional rainfall and PDSI patterns over the CONUS.

b. *Temperature data*

Monthly SSTs were obtained from NOAA–CIRES Twentieth Century Reanalysis (version 2c) data spanning from 1851 to 2014 (NOAA 2014). The data have a

spatial resolution of approximately 200 km and are an assimilation of sea surface temperature (SST) observations. SSTs were aggregated over the December–February season to capture the season in which ENSO typically reaches a mature phase (Enfield 1992; Kousky 1993). Although this seasonal selection enables possible ENSO effects to be identified, oceanic–atmospheric phenomena that are more prevalent in other seasons are then not explicitly considered, and a detailed examination of such teleconnections is beyond the scope of this paper.

3. Methods and background

a. *HMMs for characterizing continental-scale variability*

State-space models such as HMMs and their extensions have been used as a tool to both diagnose (e.g., Foufoula-Georgiou and Lettenmaier 1987; Zucchini and Guttorp 1991; Greene et al. 2008; Pal et al. 2015; Cioffi et al. 2017) and forecast or downscale weather attributes (e.g., Robertson et al. 2004; Mehrotra and Sharma 2006; Khalil et al. 2010; Greene et al. 2015; Cioffi et al. 2016). While HMMs have been a popular tool for analyzing climate variables at a daily time step, the application of HMMs to coarser temporal resolution data has also been considered. Jackson (1975) used an HMM with two states to investigate drought impacts on annual streamflow and to inform infrastructure design and resilience. Thyer and Kuczera (2003a,b) extended the approach to multiple sites with the premise that multi-year persistence of large-scale modes of climate variability would influence annual variability. Applications using annual data have also been extended to analyzing and reconstructing annual-resolution paleoclimate data (Prairie et al. 2008; Bracken et al. 2016; Erkyihun et al. 2017). HMMs have also been applied across large spatial scales using spatially correlated gridded data to model precipitation across India (Greene et al. 2011) and the Asian summer monsoon (Yoo et al. 2010).

An HMM is a doubly stochastic construct that uses an underlying latent process to model state transitions at each time step and a second stochastic process dependent on the derived state to inform emissions or variables of interest (Baum et al. 1970). The HMM is composed of three Markov chain components and two additional observations and parameters. The three Markov chain components are

- 1) m number of states in state space $S \in [1:m]$,
- 2) the state transition matrix $\mathbf{A}$, and
- 3) π the initial state probability.

In an HMM, the above states are hidden and are derived from the following:

FIG. 1. Schematic of the HMM showing the observations (x), the modeled hidden states associated with each observed time step (t), and transition probabilities (a) between states.

- 4) **OBS** = $obs_1, obs_2, \dots, obs_t$ are the observed values, in this case reconstructed PDSI where obs_t is an n -length vector composed of all the reconstructed PDSI values across the n grids that span the CONUS region ($n = 222$ grids) at time t for $t \in [1451:2005]$.
- 5) Parameter $x_{q,k}$ is the emissions distribution at location $q \in [1:n]$ that characterizes the distribution of the observations at all n locations associated with state k in state space $S \in [1:m]$.

A schematic of the dual-level modeling construct is shown in Fig. 1, showing the states at each time step [$s(t)$] on one level with the state transition probability (a_{ij}) at each time step and the associated observations (obs_t , in this case PDSI across the CONUS) at each time on the second level. The underlying latent process is defined through m states (s) in discrete time steps (t) that are not explicitly observed and hence termed hidden (top line of Fig. 1). The latent climate states in each time step $s(t)$ are modeled as

$$s(t)|s(t-1) \sim \text{Markov}(\mathbf{A}).$$

The states follow a first-order homogeneous Markov chain where the current state is only dependent on the state in the previous time step that is governed by the transition matrix $\mathbf{A}$, which defines the probability of switching from one state to another, namely,

$$\mathbf{A} = \begin{bmatrix} a_{1,1} & \cdots & a_{1,m} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,m} \end{bmatrix},$$

where the state transition probabilities (shown in blue text in Fig. 1) $a_{ij} = p[s(t) = j | s(t-1) = i]$ and m is the selected number of climate states. The state transition probabilities are assumed stationary through time.

The second stochastic process describes the observations at each time step t conditional on the hidden state (bottom line of observable LBDA values in Fig. 1), and these are modeled at each location as $x_{q,k}$. An automatic, unsupervised state classification algorithm is used to identify the hidden state variables, and the corresponding parameters of the probability distribution of the observed variables are based on the observed values and the selected number of states. The observed variables are taken to be independent, conditional on the latent state specification in the application presented here. This assumption can be relaxed by considering spatial dependence even given the latent state. An HMM can thus be presented as the simplest dynamic Bayesian network.

To determine the m number of states to retain, we fit models using the paleoclimate PDSI record with a range of m states with $m \in [2:15]$. Eight hidden states were selected by choosing the model with the minimum Bayesian information criterion (BIC) value. This number is remarkably similar to the nine regional drought patterns identified by Karl and Koscielny (1982) and Cook et al. (1999) in their independent analyses.

The LBDA values **OBS** in each state are modeled using an “emissions” distribution associated with state k for each grid, q , as x , where

$$x_{q,t}|s_t \sim \mathcal{N}(\mu_{q,k}, \sigma_{q,k})$$

and $x_{q,t}$ are independently Gaussian distributed variables with mean $\mu_{q,k}$ and standard deviation $\sigma_{q,k}$, where both mean and standard deviation are unique to each location q for a given state k . The assumption of conditional independence of the observed values at each location that is used here has been shown to be a valid assumption for spatially sparse, widely dispersed networks of climate data (e.g., rainfall; Hughes and Guttorp 1994a), where spatial correlations are driven by large-scale forcing mechanisms rather than local topographic influences (Hughes and Guttorp 1994b). The LBDA is

highly spatially correlated, but these correlations are likely due to large-scale climate states and an HMM is therefore suitable for identifying such states. It is possible that one could consider that the spatial dependence is a function of distance and model it using a variogram, but this would require either an assumption of spatial homogeneity across the entire CONUS, which is unlikely to be valid, or the use of a spatially nonstationary variogram, which adds a significant complexity to the initial model we considered. Where both large-scale and local spatial correlation are of interest, it would be worth exploring this enhancement of our initial model. Developing such a model is deferred to a future analysis and is further discussed in [section 5](#).

The HMM analyses were conducted in R using the R package depmixS4 ([Visser and Speekenbrink 2010](#)). HMM parameters are estimated using the expectation-maximization algorithm, which involves iteratively maximizing the expected joint log-likelihood of the parameters for a predetermined number of hidden states and the given set of LBDA values.

b. Determining the state sequence and corresponding climate

The probable temporal evolution of states, and hence most likely state at each time step, may be recovered using a maximum sum algorithm. This is known as the Viterbi algorithm in HMM applications ([Viterbi 1967](#)) and is based on both the probability of the LBDA values belonging to a particular state and the Markov model state transition matrix ([Bishop 2006](#)). The algorithm determines the most likely sequence of hidden states ($s_1, \dots, s_T$), known as the Viterbi path, from time $t = 1:T$, which maximizes the sum probability of the state sequence given the initial state probability (π), state transition probability (**A**), and observed LBDA record (**x**) as follows:

$$V_{1,k} = \max[p(\mathbf{x}_1|k) \pi_k],$$

$$V_{t,k} = \max_{s \in S} (\mathbf{P}(\mathbf{x}_t|k) \cdot a_{s,k} \cdot V_{t-1,s}),$$

where $V_{t,k}$ is the probability of the most probable state sequence. If we have the Viterbi path with the highest probability of V_T , then $s_T = \max_{s \in S}(V_{T,k})$, and we can backtrack to find previous states as follows:

$$s_{t-1} = f(s_{t-1} = k | s_t, V_{t,k}).$$

4. Results

We focus on the interannual evolution of dry and wet spells on large regional scales and examine the resulting

states and significant transition probabilities ([section 4a](#)). The Viterbi path is used to assess the probabilities of occurrence, persistence, and recurrence of the various states ([section 4b](#)). The potential for improving how state persistence may be related to known climate modes is explored using a composite analysis to assess large-scale climate features associated with each state ([section 4b](#)). Some distinguishing features of PDSI variability over the past five centuries resulting from the HMM analysis are identified using basic summary statistics, calculations of state persistence, recurrence, and general trends ([section 4c](#)). An assessment of the skill in simulating PDSI using the HMM follows in [section 4d](#).

a. An HMM for identifying drought states

A model with eight states resulted in the minimum BIC. The mean LBDA values corresponding to each hidden state are shown in [Fig. 2a](#), and these reveal the different spatial patterns of relative drought or moisture surplus associated with each state. The blue color represents a positive anomaly (i.e., wet) and the red color represents a negative anomaly (i.e., dry) of LBDA values. The arrows and the associated numbers represent the annual state transition probabilities. For example, state 1 has a 0.27 probability of remaining in state 1 the next year. Similarly, state 1 will transition to states 3 and 5 with probabilities of 0.14 and 0.17, respectively. Likewise, there is a 0.22 probability of state 1 occurring the year after state 4.

States 1 and 8 appear to represent congruent patterns of wet spells and drought conditions, respectively, in the region west of the Mississippi River concentrated around Utah. Likewise, states 2 and 7 show opposing wet and dry patterns, but these patterns are centered near the U.S.–Mexico border and feature the often-observed climate divide between northern and southern California (e.g., [Dettinger et al. 1998](#); [Andrews et al. 2004](#)), with the Pacific Northwest region showing an opposite sign to the remainder of the country. These pairs of states with opposing patterns represent results similar to those obtained using PCA ([Karl and Koscielny 1982](#); [Cook et al. 1999](#)), particularly when a small number of principal components are retained (see [Fig. S2](#) in the supplemental material for results using four principal components).

The transition probability significance levels (matrix **A** shown in Table S1 in the supplemental material) were calculated using 10 000 bootstrap samples of the Viterbi path to determine the distribution of transition probabilities expected by chance. Transition probabilities that were found to be significant at the 90th percentile level using a one-sided test are shown in [Fig. 2a](#). Each state showed persistence from year to year, with the

FIG. 2. (a) Mean LBDA value in each hidden state. “D” is the percentage of the area where mean LBDA values are less than or equal to -2 and “W” is the percentage of the area where mean LBDA values are greater than or equal to 2 . Arrows and gray values show transition probabilities that are significant at the 90th percentile probability based on a bootstrap confidence test. (b) Likely sequence of hidden states, or Viterbi path, determined using the Viterbi algorithm and results from the HMM shown in block colors.

probability of remaining in the same state at the next time step (i.e., $a_{k,k}$) generally being highest, with the exception of states 5 and 8. State 5 is most likely to transition to state 2 (21.5%) or state 7 (20%) than to itself. States 2 and 7 are two states with spatially congruent patterns of LBDA. This result thus helps identify how state 5 may be expected to shift to either a wet or dry state in the next year. Similarly, state 8 is more likely to transition to state 6 (25.4%), which results in drought conditions shifting from the U.S. West to the northern High Plains.

As one can see from Fig. 2, the persistence in the same state is typically dominant, and just a few transitions dominate the statistically significant probabilities in the transition matrix (see Table S1 in the supplemental material), indicating that the model is able to capture characteristics of persistent droughts and wet spells. This sparsity of the transition matrix is consequently highly informative regarding the potential movement of the drought over time. Consider the operational application of the HMM: given the state identified for the current year, across the CONUS, one can easily identify the most likely states the drought may transition to in the following year. The HMM may then be used to simulate values of PDSI, x_{t+j} for time steps of $j = 1 \dots J$.

The probable temporal evolution of drought and wet spells can be discerned by the Viterbi path (Fig. 2b). During the 1950s drought, the state sequence is 3, 7, 7, 8, and 7 for years 1951–56, respectively (observed LBDA patterns can be seen in Fig. S3 in the supplemental material). State 3 largely represents neutral conditions across the CONUS, with marginally dry conditions in the Southwest and Florida combined with marginally wetter conditions elsewhere similar to the spatial pattern of the LBDA in 1951. Drought in years identified as state 7 was primarily concentrated in and around Texas. In contrast, drought in 1955 was widespread, and this year was assigned to state 8. The recurrence of states 7 and 8 suggest a stable forcing mechanism may drive these patterns of drought. The mean LBDA values in state 8 are comparable with the drought patterns obtained by Cook et al. (2011) when their stochastic model of CONUS drought was forced by equatorial Pacific SSTs. The drought pattern in state 7 is likewise similar to the pattern produced by Cook et al. (2011) when both equatorial and North Pacific SSTs are considered.

The Viterbi path appears to also capture the 1930s Dust Bowl drought with the most widespread severe years of drought (1934 and 1939) assigned to state 8. Years where LBDA values in California were largely positive (1936, 1937, 1940) were assigned to state 6, where the most severe drought is concentrated in the central northern region of the country. Years where the

LBDA values were positive (i.e., wetter conditions) across a large portion of the U.S. Midwest (1935 and 1938) were assigned to state 3. State 6 does not show distinct parallels with drought patterns forced by SSTs (Cook et al. 2011) and suggests that other mechanisms may be driving the prolonged drought (Cook et al. 2009). It should be noted that the Viterbi path is the most probable state sequence based on the state transition probabilities and the probability of the observed values fitting the emissions distribution of a given state. While the Viterbi path is a useful tool for visualizing the state occurrences through time, there is, however, uncertainty associated with the state classification that has been addressed in studies elsewhere (Hernando et al. 2005; Novoa et al. 2018).

While large-scale droughts appear to be well captured by the HMM, droughts with a smaller spatial scale are less robust. For example, the relatively regionally confined drought from 1962 to 1966 in the Northeast (Namias 1966; Barlow et al. 2001) appears prominently in the LBDA record. However, these years are assigned to states 1, 7, 5, 1, and 5, respectively. Mean LBDA values in these states show mildly dry (state 5 and 7) or mildly wet conditions (state 1) in the Northeast. This representation of the Northeast drought is consistent with Seager et al.'s (2012) observation that internal climate variability, likely unconnected with a larger CONUS drought pattern, likely caused the Northeast drought. The larger-scale features of LBDA variability across the remainder of the CONUS, such as the anomalous wet conditions in the northern High Plains in 1962 and 1965 and Pacific Northwest in 1963 (Namias 1966), do, however, feature in the corresponding state.

b. Large-scale SST patterns associated with each state

A composite analysis was used to assess large-scale SST patterns associated with each state over ~150 years with the analysis period constrained by the SST data. Standardized SST from the preceding ENSO season (defined here as December–February) were used to assess whether ENSO-like variability is associated with any of the states. The lagged comparison was selected as PDSI has a long memory at time scales of around 12–18 months (Vicente-Serrano et al. 2010) and summer PDSI can therefore be representative of moisture conditions from the previous year (Dai et al. 2004; St. George et al. 2010).

SST anomalies associated with states 1, 2, 7, and 8 manifest as anomalous temperatures west of South America, reminiscent of the Pacific warm and cool tongue characteristic of ENSO variability (Fig. 3) and correspond with the findings of Cook et al. (2011). Anomalous warm SSTs in the Pacific Ocean are

FIG. 3. Anomalies of standardized December–February surface temperature (1851–2004) composited by the following summer’s (JJA) state, as defined by the HMM.

associated with state 1, which is characterized by wetter conditions across much of the CONUS concentrated in the West. The anomalous SSTs in the equatorial Pacific mimic an El Niño–like pattern but with a latitudinally broad warm band. A similar pattern of anomalously warm SSTs is associated with state 2, but stronger anomalies are seen in the Niño-3.4 region (defined as 5°S – 5°N and 170°W – 120°W ; Trenberth 1997) in comparison to state 1 composites. Anomalously warm SSTs in the Niño-3.4 region are associated with wetter conditions in the Southwest and drier conditions in Oregon and Northern California (Redmond and Koch 1991). In addition, ENSO impacts are enhanced by negative PDO (warmer SSTs in the central North Pacific Ocean), which is somewhat evident in state 2 (McCabe and Dettinger 1999).

The opposing La Niña–like cool tongue pattern has been shown to suppress precipitation in the Great Plains (Seager et al. 2003; Livneh and Hoerling 2016), resulting in a spatial drought pattern similar to state 7 (Cole et al. 2002). Anomalously cool SSTs along the equatorial Pacific are linked with occurrences of megadroughts in the west and southwest CONUS (Coats et al. 2016) and likely contribute to drought conditions associated with states 6, 7, and 8. However, weak SST anomalies for states 6 and 8 suggest that other mechanisms influence drought in these states (Cole and Cook 1998), such as variability in the North Pacific and Atlantic Ocean in different seasons and over different time scales. Some of these SST patterns are evident despite the seasonally restricted analysis period (December–February SSTs) used here.

A pattern of SST variability similar to AMO is seen in composites for states 2 (a weak pattern) and 7 with

anomalies in the northern Atlantic Ocean that extend east toward Africa and into the tropics (Nigam et al. 2011). Warm Atlantic Ocean SSTs are associated with anomalously low precipitation across much of the CONUS, resulting in patterns of drought similar to states 7 and 8. AMO and Pacific Ocean variability potentially differentiate the drought patterns for these states. Composited SSTs for state 7 show a distinct horseshoe shape of cool waters in the northeast and tropical Pacific, with warmer waters in the Atlantic. In contrast, state 8 SST composites show a weak signal in the northern Pacific. The widespread anomalously warm Atlantic Ocean SSTs and drought pattern associated with state 8 more closely resemble modeled results by Sutton and Hodson (2005), who focused on AMO variability, and findings by McCabe et al. (2004), who considered AMO impacts independent of PDO.

A lack of moisture transport from the Gulf of Mexico has also been identified as a contributing mechanism for drought in the central United States (Hoerling et al. 2014). Conversely, extreme precipitation in the southern Mississippi region often results from moisture transport from the Gulf (Leipper and Volgenau 1972). Cool waters in the Gulf of Mexico for state 5 coincide with dry conditions in the Midwest centered around the Mississippi River. In contrast, state 3 composites feature anomalously warm SSTs in and wetter conditions over the Mississippi River region.

The composite analysis shows that indices representing SST variability in the Pacific and Atlantic Oceans and Gulf of Mexico could be used to inform state transitions probabilities. However, these would require a credible reconstruction or simulation model for these

fields. In future applications, we propose to explore a multilevel stochastic model for the SSTs and the LBDA using appropriate data records for each field.

c. Assessing drought characteristics

The Climate Prediction Center's definitions of drought ($\text{PDSI} \leq -2$) and unusually moist conditions ($\text{PDSI} \geq 2$; [Climate Prediction Center 2015](#)) were used to determine the proportion of the country that was dry (D) or wet (W) for each mean state (values shown in [Fig. 2a](#)). A larger portion of the CONUS is in drought in the driest state (state 8) compared with unusually moist areas in the wettest state (state 1), and this result is likewise seen in the congruent states 2 and 7. The marginal probability of occurrence of a state (first column of Table S1 of the supplemental material that also shows the number of years classed as each state) showed that wetter states were more likely to occur over the 555-yr record compared with drier states. The combined results of spatial impact and marginal probability of occurrence support findings based on the instrumental record that while dry states may occur slightly less frequently than wetter states ([Diaz 1983](#)), they are experienced over a larger proportion of the CONUS ([Kangas and Brown 2007](#)).

The higher probability of a wet state occurring is even more pronounced over the twentieth century (see [Fig. 2](#) and Table S1 in the supplemental material). For example, 87 years are associated with state 1 over the 555-yr record, with 27 of these years occurring in the time since 1901. The marginal probability of state 1 occurring therefore increases from 0.16, when the entire record is considered, to 0.26 for the most recent period. To further understand how state occurrences have changed over the past 555 years, a local regression was fitted to each binary time series representing the occurrence of each state using the Locfit package in R (see [Fig. S4](#) in the supplemental material). Shorter smoothing windows indicate high decadal-scale variability in state occurrence. However, regressions using longer windows suggest that there is an increasing occurrence of wetter states (states 1, 2, and 3) and a slight decrease in the occurrence of predominantly dry states. The higher occurrence of drier states in the earlier portion of the record, as defined using the HMM, corroborates with identification of a "megadrought epoch" in the paleoclimate PDSI ending at the start of the seventeenth century ([Cook et al. 2016](#)).

The persistence of states was examined using the residence time in each state. States with the highest probability of remaining in the same state (i.e., $a_{k,k}$) had the longest mean persistence (e.g., state 6 in Table S2 in the supplemental material). While the maximum marginal persistence of a state was only 5 years, it can be

seen from [Fig. 2b](#) that the sequencing of different states that are predominantly dry (or wet) can lead to prolonged periods of drought (or longer wet spells). An example of such sequencing occurs in the 1450s, where the repeated occurrence of states 5, 7, and 8 reflect a near-decadal drought previously documented by [Cook et al. \(2007\)](#) and [Stahle et al. \(2007\)](#). Persistent wet spells composed of different states are likewise evident. One example is from 1825 to 1840 (16 years), which was dominated by the occurrence of states 1 and 2, apart from four years that were classified as either state 3 or 4. Similarly, 12 out of 17 years from 1905 to 1921 were classified as state 1 or 2. The climate around this period, dominated by wetter states, was used to inform the allocation of water between seven western states in the 1922 Colorado River compact. The wetter period at the start of the twentieth century led to an overestimation of available water resources and overallocation of water in the following decades ([Woodhouse et al. 2005](#); [Grafton et al. 2011](#)).

The duration before the recurrence of a state was also calculated (Table S2 in the supplemental material). The maximum recurrence intervals for states 1 and 8 were 63 and 62 years, respectively. Recall that these states are basically reflections of each other and have the largest areal coverage of the extreme ($|\text{LBDA}| > 2$) wet or dry pattern.

d. Conditional simulation skill using the HMM

An initial assessment of the HMM conditional simulations for simulating droughts was made by hindcasting over the duration of the LBDA record. Simulations were made using only the HMM parameters and an input of the state at time t for $t \in [1:(555 - \text{lead time})]$. The state sequence between time t and time $t + \text{lead time}$ was generated and the hindcast PDSI values were simulated using the emissions distribution at each grid location q , associated with the simulated state k at time $t + \text{lead time}$. For each initial condition (i.e., the state identified in the Viterbi path at time t), 1000 simulations were generated for the lead time of interest. An assessment of the potential simulation skill was made using correlations between the LBDA and the median of the simulated PDSI values for each grid. Lead times of 1–5 years were tested (see [Fig. S5](#) in the supplemental material), and the results for a lead time of 1 year are shown in [Fig. 4](#).

Correlations between the median of the hindcasted PDSI range and the LBDA were significant at the 95th percentile across most of the country for a lead time of 1 year, with the exception of the southeastern United States and the West Coast ([Fig. 4a](#)). The root-mean-square of the median hindcasts over 555 years was less

FIG. 4. (a) Correlation and (b) root-mean-square error between median of annually hindcast PDSI for lead time of 1 year and LBDA value.

than 0.21 across the CONUS (Fig. 4b). However, lead times of 3 years or more resulted in anticorrelated hindcasts. The reduced skill over longer lead times indicates that the model does not accurately simulate persistence despite the ability to diagnose key persistent events in the LBDA record. We therefore examined the extent to which a first-order Markov chain assumption was able to adequately quantify persistence by comparing the state persistence and recurrence probabilities as revealed by the Viterbi path with those obtained by Markov recursion. Recall that the Viterbi path is derived using both the transition matrix and observed values, while state sequences generated in the hindcasting process were dependent only on the state transition probabilities (i.e., Markov recursion).

The probabilities of a state persisting over t years based on either the Viterbi path or Markov recursion were found to be similar (see Table S3 in the supplemental material). However, recurrence probabilities differed when comparing the Viterbi path to Markov recursion probabilities (see Table S4 in the supplemental material). The difference in recurrence probability suggests that a first-order HMM does not fully capture long-term temporal evolutions of droughts and wet spells. This feature of the HMM could potentially be improved by using an exogenous covariate such as ENSO to alter the transition matrix at each time step within a nonhomogeneous HMM (NHMM) framework. In theory, ENSO would operate to increase the probability of transitioning to a wetter state during an El Niño or conversely to a dry state during a La Niña. Note that this extension would require a credible model for future

ENSO forecasts in practice. Another alternative would be to consider a wavelet HMM that better captures low-frequency variability. Work in these directions is being explored.

5. Discussion and conclusions

a. Drought impacts and analysis

Improving the ability to adequately prepare for, respond to, and mitigate regional- and continental-scale drought and flood risks requires skillful estimates of their likely development and evolution in both space and time. Dix and Hunt (1995) observed that accurate simulations of droughts in GCMs are limited by the chaotic nature of hydroclimatic variables, such as precipitation, wind, and temperature, and by the complexity of hydroclimatic processes. In contrast, stochastic analyses are limited by the use of assumptions of independence between severity, duration, and or spatial extents in order to simplify analyses (Yoo et al. 2012), with recent studies making attempts to address these limitations (Andreadis et al. 2005; Salas et al. 2005).

Point analysis of drought using time series records has a long history with applications to both short instrumental and long paleoclimate records. Its utility has been well recognized by the large literature on synthetic streamflow and rainfall simulations. The use of continental data with the HMM builds on this legacy by enhancing the information contained in the point time series with large-scale regional information, using an automatic grouping and identification of regimes that evolve in a similar way and hence may relate to common

climate mechanisms. The inclusion of continental-scale information that considers spatial and temporal variability may then improve the ability to assess the probability of drought inception, persistence, or termination, all of which are important for local and regional resource managers. Moreover, the ability to identify the probability of where the drought or wet spell may move to is of interest to national resource and emergency managers. The continental scale of the United States makes this a particularly attractive application.

The approach could also be applied using other drought metrics, depending on the application of interest. In this case, a paleoclimate reconstruction of the modified PDSI was used to represent drought variability. While issues with the fidelity of the PDSI in representing drought have been extensively documented (e.g., Alley 1984; Guttman et al. 1992), it is unreasonable to claim that a particular drought index is unanimously superior to all others (Heim 2002).

b. Model limitations, improvements, and potential extensions

A limitation to modeling persistence using HMMs or Markov models is that persistence is characterized by a geometric decay with fixed state transition probabilities. Higher-order Markov models lead to an explosion of the number of parameters to be estimated, especially as the number of states considered increases. Consequently, a BIC-like criteria for model selection may usually not admit such a model as “optimal.” Extensions of HMMs such as NHMMs that use an exogenous variable to condition the state transition probabilities are potentially more skillful in capturing persistence, and the state transition probabilities are subsequently variable through time. The composite analysis in section 4b indicates that the occurrence of some of these states coincide with distinct SST patterns associated with ENSO, PDO, AMO, and Gulf of Mexico variability. Including information on these ocean–atmosphere drivers or other climate variables is somewhat inhibited by a dearth of high-resolution paleoclimate records that are informed by proxies independent of those used to construct the LBDA (Wilson et al. 2010). An alternative to using proxy-based climate reconstructions could be to use numerical models (e.g., Claussen et al. 2002; Goosse et al. 2010; Schmidt et al. 2011; Taylor et al. 2012) to reproduce SST patterns or temperature over the past 500 years. However, such an approach may compound the biases and uncertainties across both types of models. We are currently considering a wavelet-HMM development that would consider the application of the HMM to objectively selected wavelet coefficient time series corresponding to temporal frequencies that are

identified as statistically significant. This approach blends elements of the HMM with the wavelet autoregressive moving average modeling approach presented in Kwon et al. (2007).

A source of uncertainty in the analysis stems from the LBDA data, as they are based on a decreasing number of tree-ring chronologies further back in time. The changing number of chronologies likely impact the variability in the resulting PDSI reconstruction and subsequent state classification and emissions distribution. Maximum entropy bootstrapped ensembles and bootstrapping on regression residuals can be used to quantify the uncertainty in the LBDA (Wahl et al. 2012). A future study could use these ensembles to refit the HMM and Viterbi algorithm and quantify the uncertainties in transition probabilities and latent state identification associated with the LBDA.

Given the high dimensionality of the LBDA, an assumption of spatial independence was employed here. The high dimensionality of the dataset resulted in a degenerate matrix when the system was modeled using a multivariate distribution. Alternatives for modeling spatial dependence, without the use of a large covariance matrix, exist and could potentially be incorporated into a HMM framework. These approaches include Chow–Liu trees (Chow and Liu 1968), copulas (Genest and Favre 2007; Lall et al. 2016), and non-stationary spatial variograms (Banerjee et al. 2004), among others. Appropriately modeling the spatial dependence would improve both the characterization of HMM states as well as future simulation applications.

Acknowledgments. We wish to thank David Farnham, Soojun Kim, Ipsita Kumar, Indrani Pal, and Xun Sun for their helpful discussions. We would also like to thank the Editor in Chief, Timothy DelSole, and three anonymous reviewers for providing guidance and informative feedback and suggestions. This work is funded by NSF Awards 1360446 and 1404188. Lamont–Doherty Contribution Number 8227. Support for the Twentieth Century Reanalysis Project version 2c dataset is provided by the U.S. Department of Energy, Office of Science Biological and Environmental Research, and by the National Oceanic and Atmospheric Administration Climate Program Office.

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).
- Andreadis, K. M., E. A. Clark, A. W. Wood, A. F. Hamlet, and D. P. Lettenmaier, 2005: Twentieth-century drought in the conterminous United States. *J. Hydrometeorol.*, **6**, 985–1001, <https://doi.org/10.1175/JHM450.1>.

- Andrews, E. D., R. C. Antweiler, P. Neiman, and F. M. Ralph, 2004: Influence of ENSO on flood frequency along the California coast. *J. Climate*, **17**, 337–348, [https://doi.org/10.1175/1520-0442\(2004\)017<0337:IOEOF>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<0337:IOEOF>2.0.CO;2).
- Banerjee, S., B. P. Carlin, and A. E. Gelfand, 2004: *Hierarchical Modeling and Analysis for Spatial Data*. CRC Monographs on Statistics and Applied Probability, Vol. 101, CRC Press, 474 pp.
- Barlow, M., S. Nigam, and E. H. Berbery, 2001: ENSO, Pacific decadal variability, and U.S. summertime precipitation, drought, and stream flow. *J. Climate*, **14**, 2105–2128, [https://doi.org/10.1175/1520-0442\(2001\)014<2105:EPDVAU>2.0.CO;2](https://doi.org/10.1175/1520-0442(2001)014<2105:EPDVAU>2.0.CO;2).
- Baum, L. E., T. Petrie, G. Soules, and N. Weiss, 1970: A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Stat.*, **41**, 164–171, <https://doi.org/10.1214/aoms/1177697196>.
- Bellone, E., J. P. Hughes, and P. Guttorp, 2000: A hidden Markov model for downscaling synoptic atmospheric patterns to precipitation amounts. *Climate Res.*, **15**, 1–12, <http://www.jstor.org/stable/24867296>.
- Bishop, C. M., 2006: *Pattern Recognition and Machine Learning*. Springer, 738 pp.
- Booker, J. F., 1995: Hydrologic and economic impacts of drought under alternative policy responses. *J. Amer. Water Resour. Assoc.*, **31**, 889–906, <https://doi.org/10.1111/j.1752-1688.1995.tb03409.x>.
- Boyer, J. S., and Coauthors, 2013: The U.S. drought of 2012 in perspective: A call to action. *Global Food Secur.*, **2**, 139–143, <https://doi.org/10.1016/j.gfs.2013.08.002>.
- Bracken, C., B. Rajagopalan, and E. Zagona, 2014: A hidden Markov model combined with climate indices for multi-decadal streamflow simulation. *Water Resour. Res.*, **50**, 7836–7846, <https://doi.org/10.1002/2014WR015567>.
- , —, and C. Woodhouse, 2016: A Bayesian hierarchical nonhomogeneous hidden Markov model for multisite streamflow reconstructions. *Water Resour. Res.*, **52**, 7837–7850, <https://doi.org/10.1002/2016WR018887>.
- Chow, C., and C. Liu, 1968: Approximating discrete probability distributions with dependence trees. *IEEE Trans. Inf. Theory*, **14**, 462–467, <https://doi.org/10.1109/TIT.1968.1054142>.
- Cioffi, F., F. Conticello, and U. Lall, 2016: Projecting changes in Tanzania rainfall for the 21st century. *Int. J. Climatol.*, **36**, 4297–4314, <https://doi.org/10.1002/joc.4632>.
- , —, —, L. Marotta, and V. Telesca, 2017: Large scale climate and rainfall seasonality in a Mediterranean area: Insights from a non-homogeneous Markov model applied to the Agro-Pontino plain. *Hydrol. Processes*, **31**, 668–686, <https://doi.org/10.1002/hyp.11061>.
- Claussen, M., and Coauthors, 2002: Earth system models of intermediate complexity: Closing the gap in the spectrum of climate system models. *Climate Dyn.*, **18**, 579–586, <https://doi.org/10.1007/s00382-001-0200-1>.
- Climate Prediction Center, 2015: Drought monitoring. NOAA, http://www.cpc.ncep.noaa.gov/products/monitoring_and_data/drought.shtml.
- Coats, S., J. E. Smerdon, K. B. Karnauskas, and R. Seager, 2016: The improbable but unexceptional occurrence of megadrought clustering in the American West during the Medieval Climate Anomaly. *Environ. Res. Lett.*, **11**, 74025, <https://doi.org/10.1088/1748-9326/11/7/074025>.
- Cole, J. E., and E. R. Cook, 1998: The changing relationship between ENSO variability and moisture balance in the continental United States. *Geophys. Res. Lett.*, **25**, 4529–4532, <https://doi.org/10.1029/1998GL900145>.
- , J. T. Overpeck, and E. R. Cook, 2002: Multiyear La Niña events and persistent drought. *Geophys. Res. Lett.*, **29**, 1647, <https://doi.org/10.1029/2001GL013561>.
- Cook, B. I., R. L. Miller, and R. Seager, 2009: Amplification of the North American “Dust Bowl” drought through human-induced land degradation. *Proc. Natl. Acad. Sci. USA*, **106**, 4997–5001, <https://doi.org/10.1073/pnas.0810200106>.
- , E. R. Cook, K. J. Anchukaitis, R. Seager, and R. L. Miller, 2011: Forced and unforced variability of twentieth century North American droughts and pluvials. *Climate Dyn.*, **37**, 1097–1110, <https://doi.org/10.1007/s00382-010-0897-9>.
- , and Coauthors, 2016: The paleoclimate context and future trajectory of extreme summer hydroclimate in eastern Australia. *J. Geophys. Res. Atmos.*, **121**, 12 820–12 838, <https://doi.org/10.1002/2016JD024892>.
- Cook, E. R., D. M. Meko, D. W. Stahle, and M. K. Cleaveland, 1999: Drought reconstructions for the continental United States. *J. Climate*, **12**, 1145–1162, [https://doi.org/10.1175/1520-0442\(1999\)012<1145:DRFTCU>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<1145:DRFTCU>2.0.CO;2).
- , 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>.
- , —, R. R. Heim, R. S. Vose, C. Herweijer, and C. Woodhouse, 2010: Megadroughts in North America: Placing IPCC projections of hydroclimatic change in a long-term palaeoclimate context. *J. Quart. Sci.*, **25**, 48–61, <https://doi.org/10.1002/jqs.1303>.
- Cooley, H., K. Donnelly, R. Phurisamban, and M. Subramanian, 2015: Impacts of California’s ongoing drought: Agriculture. Pacific Institute Doc., 30 pp., <http://pacinst.org/app/uploads/2015/08/ImpactsOnCaliforniaDrought-Ag.pdf>.
- Dai, A., K. E. Trenberth, and T. Qian, 2004: A global dataset of Palmer drought severity index for 1870–2002: Relationship with soil moisture and effects of surface warming. *J. Hydrometeorol.*, **5**, 1117–1130, <https://doi.org/10.1175/JHM-386.1>.
- Damberg, L., and A. AghaKouchak, 2014: Global trends and patterns of drought from space. *Theor. Appl. Climatol.*, **117**, 441–448, <https://doi.org/10.1007/s00704-013-1019-5>.
- Dettinger, M. D., D. R. Cayan, H. F. Diaz, and D. M. Meko, 1998: North–south precipitation patterns in western North America on interannual-to-decadal timescales. *J. Climate*, **11**, 3095–3111, [https://doi.org/10.1175/1520-0442\(1998\)011<3095:NSPPIW>2.0.CO;2](https://doi.org/10.1175/1520-0442(1998)011<3095:NSPPIW>2.0.CO;2).
- Devineni, N., U. Lall, E. Etienne, D. Shi, and C. Xi, 2015: America’s water risk: Current demand and climate variability. *Geophys. Res. Lett.*, **42**, 2285–2293, <https://doi.org/10.1002/2015GL063487>.
- Diaz, H. F., 1983: Drought in the United States. *J. Climate Appl. Meteorol.*, **22**, 3–16, [https://doi.org/10.1175/1520-0450\(1983\)022<0003:DITUS>2.0.CO;2](https://doi.org/10.1175/1520-0450(1983)022<0003:DITUS>2.0.CO;2).
- Dix, M. R., and B. G. Hunt, 1995: Chaotic influences and the problem of deterministic seasonal predictions. *Int. J. Climatol.*, **15**, 729–752, <https://doi.org/10.1002/joc.3370150703>.
- Drabenstott, M., and A. Barkema, 1988: U.S. agriculture shrugs off the drought. *Economic Review*, Vol. 73, No. 10, Federal Reserve Bank of Kansas City, Kansas City, MO, 32–50, <https://www.kansascityfed.org/~media/files/publicat/econrev/econrevarchive/1988/4q88drab.pdf>.
- Enfield, D. B., 1992: Historical and prehistorical overview of El Niño/Southern Oscillation. *El Niño: Historical and Paleoclimatic Aspects of the Southern Oscillation*, H. F. Diaz and V. Markgraf, Eds., Cambridge University Press, 476 pp.

- Erkyihun, S. T., E. Zagona, and B. Rajagopalan, 2017: Wavelet and hidden Markov-based stochastic simulation methods comparison on Colorado River streamflow. *J. Hydrol. Eng.*, **22**, 04017033, [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001538](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001538).
- Foufoula-Georgiou, E., and D. P. Lettenmaier, 1987: A Markov renewal model for rainfall occurrences. *Water Resour. Res.*, **23**, 875–884, <https://doi.org/10.1029/WR023i005p00875>.
- Fye, F. K., D. W. Stahle, and E. R. Cook, 2003: Paleoclimatic analogs to twentieth-century moisture regimes across the United States. *Bull. Amer. Meteor. Soc.*, **84**, 901–909, <https://doi.org/10.1175/BAMS-84-7-901>.
- Genest, C., and A.-C. Favre, 2007: Everything you always wanted to know about copula modeling but were afraid to ask. *J. Hydrol. Eng.*, **12**, 347–368, [https://doi.org/10.1061/\(ASCE\)1084-0699\(2007\)12:4\(347\)](https://doi.org/10.1061/(ASCE)1084-0699(2007)12:4(347)).
- Goodrich, G. B., 2007: Influence of the Pacific decadal oscillation on winter precipitation and drought during years of neutral ENSO in the western United States. *Wea. Forecasting*, **22**, 116–124, <https://doi.org/10.1175/WAF983.1>.
- Goosse, H., E. Cressin, A. de Montety, M. E. Mann, H. Renssen, and A. Timmermann, 2010: Reconstructing surface temperature changes over the past 600 years using climate model simulations with data assimilation. *J. Geophys. Res.*, **115**, D09108, <https://doi.org/10.1029/2009JD012737>.
- Grafton, R. Q., G. D. Libecap, E. C. Edwards, F. J. O'Brien, and C. Landry, 2011: A comparative assessment of water markets: Insights from the Murray-Darling basin of Australia and the western USA. *Water Policy*, **14**, 175–193.
- Greene, A. M., A. W. Robertson, and S. Kirshner, 2008: Analysis of Indian monsoon daily rainfall on subseasonal to multidecadal time-scales using a hidden Markov model. *Quart. J. Roy. Meteor. Soc.*, **134**, 875–887, <https://doi.org/10.1002/qj.254>.
- , —, P. Smyth, and S. Triglia, 2011: Downscaling projections of Indian monsoon rainfall using a non-homogeneous hidden Markov model. *Quart. J. Roy. Meteor. Soc.*, **137**, 347–359, <https://doi.org/10.1002/qj.788>.
- , T. Holsclaw, A. W. Robertson, and P. Smyth, 2015: A Bayesian multivariate nonhomogeneous Markov model. *Machine Learning and Data Mining Approaches to Climate Science*, V. Lakshmanan et al., Eds., Springer, 61–69, https://doi.org/10.1007/978-3-319-17220-0_6.
- Guttman, N. B., 1998: Comparing the Palmer drought index and the Standardized Precipitation Index. *J. Amer. Water Resour. Assoc.*, **34**, 113–121, <https://doi.org/10.1111/j.1752-1688.1998.tb05964.x>.
- , J. R. Wallis, and J. R. M. Hosking, 1992: Spatial comparability of the Palmer drought severity index. *J. Amer. Water Resour. Assoc.*, **28**, 1111–1119, <https://doi.org/10.1111/j.1752-1688.1992.tb04022.x>.
- Heim, R. R., 2002: A review of twentieth-century drought indices used in the United States. *Bull. Amer. Meteor. Soc.*, **83**, 1149–1165, <https://doi.org/10.1175/1520-0477-83.8.1149>.
- Hernando, D., V. Crespi, and G. Cybenko, 2005: Efficient computation of the hidden Markov model entropy for a given observation sequence. *IEEE Trans. Inf. Theory*, **51**, 2681–2685, <https://doi.org/10.1109/TIT.2005.850223>.
- Ho, M., V. Parthasarathy, E. Etienne, T. A. Russo, N. Devineni, and U. Lall, 2016: America's water: Agricultural water demands and the response of groundwater. *Geophys. Res. Lett.*, **43**, 7546–7555, <https://doi.org/10.1002/2016GL069797>.
- , U. Lall, M. Allaire, N. Devineni, H.-H. Kwon, I. Pal, D. Raff, and D. Wegner, 2017: The future role of dams in the United States of America. *Water Resour. Res.*, **53**, 982–998, <https://doi.org/10.1002/2016WR019905>.
- Hoerling, M., J. Eischeid, A. Kumar, R. Leung, A. Mariotti, K. Mo, S. Schubert, and R. Seager, 2014: Causes and predictability of the 2012 Great Plains drought. *Bull. Amer. Meteor. Soc.*, **95**, 269–282, <https://doi.org/10.1175/BAMS-D-13-00055.1>.
- Hornbeck, R., 2012: The enduring impact of the American Dust Bowl: Short-and long-run adjustments to environmental catastrophe. *Amer. Econ. Rev.*, **102**, 1477–1507, <https://doi.org/10.1257/aer.102.4.1477>.
- Hughes, J. P., and P. Guttorp, 1994a: A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena. *Water Resour. Res.*, **30**, 1535–1546, <https://doi.org/10.1029/93WR02983>.
- , and —, 1994b: Incorporating spatial dependence and atmospheric data in a model of precipitation. *J. Appl. Meteor.*, **33**, 1503–1515, [https://doi.org/10.1175/1520-0450\(1994\)033<1503:ISDAAD>2.0.CO;2](https://doi.org/10.1175/1520-0450(1994)033<1503:ISDAAD>2.0.CO;2).
- Jackson, B. B., 1975: Markov mixture models for drought lengths. *Water Resour. Res.*, **11**, 64–74, <https://doi.org/10.1029/WR011i001p00064>.
- Kangas, R. S., and T. J. Brown, 2007: Characteristics of US drought and pluvials from a high-resolution spatial dataset. *Int. J. Climatol.*, **27**, 1303–1325, <https://doi.org/10.1002/joc.1473>.
- Karl, T. R., and A. J. Koscielny, 1982: Drought in the United States: 1895–1981. *J. Climatol.*, **2**, 313–329, <https://doi.org/10.1002/joc.3370020402>.
- Khalil, A. F., H.-H. Kwon, U. Lall, and Y. H. Kaheil, 2010: Predictive downscaling based on non-homogeneous hidden Markov models. *Hydrol. Sci. J.*, **55**, 333–350, <https://doi.org/10.1080/02626661003780342>.
- Kousky, V. E., 1993: The global climate of December 1991–February 1992: Mature-phase warm (ENSO) episode conditions develop. *J. Climate*, **6**, 1639–1655, [https://doi.org/10.1175/1520-0442\(1993\)006<1639:TGCODM>2.0.CO;2](https://doi.org/10.1175/1520-0442(1993)006<1639:TGCODM>2.0.CO;2).
- Kwon, H.-H., U. Lall, and A. F. Khalil, 2007: Stochastic simulation model for nonstationary time series using an autoregressive wavelet decomposition: Applications to rainfall and temperature. *Water Resour. Res.*, **43**, W05407, <https://doi.org/10.1029/2006WR005258>.
- , —, and S.-J. Kim, 2016: The unusual 2013–2015 drought in South Korea in the context of a multicentury precipitation record: Inferences from a nonstationary, multivariate, Bayesian copula model. *Geophys. Res. Lett.*, **43**, 8534–8544, <https://doi.org/10.1002/2016GL070270>.
- Lall, U., N. Devineni, and Y. Kaheil, 2016: An empirical, non-parametric simulator for multivariate random variables with differing marginal densities and nonlinear dependence with hydroclimatic applications. *Risk Anal.*, **36**, 57–73, <https://doi.org/10.1111/risa.12432>.
- Leavitt, W. M., and J. J. Kiefer, 2006: Infrastructure interdependency and the creation of a normal disaster: The case of Hurricane Katrina and the city of New Orleans. *Public Works Manage. Policy*, **10**, 306–314, <https://doi.org/10.1177/1087724X06289055>.
- Leipper, D. F., and D. Volgenau, 1972: Hurricane heat potential of the Gulf of Mexico. *J. Phys. Oceanogr.*, **2**, 218–224, [https://doi.org/10.1175/1520-0485\(1972\)002<0218:HHPOTG>2.0.CO;2](https://doi.org/10.1175/1520-0485(1972)002<0218:HHPOTG>2.0.CO;2).
- Livneh, B., and M. P. Hoerling, 2016: The physics of drought in the U.S. central Great Plains. *J. Climate*, **29**, 6783–6804, <https://doi.org/10.1175/JCLI-D-15-0697.1>.
- McCabe, G. J., and M. D. Dettinger, 1999: Decadal variations in the strength of ENSO teleconnections with precipitation in the western United States. *Int. J. Climatol.*, **19**, 1399–1410,

- [https://doi.org/10.1002/\(SICI\)1097-0088\(19991115\)19:13<1399::AID-JOC457>3.0.CO;2-A](https://doi.org/10.1002/(SICI)1097-0088(19991115)19:13<1399::AID-JOC457>3.0.CO;2-A).
- , M. A. Palecki, and J. L. Betancourt, 2004: Pacific and Atlantic Ocean influences on multidecadal drought frequency in the United States. *Proc. Natl. Acad. Sci. USA*, **101**, 4136–4141, <https://doi.org/10.1073/pnas.0306738101>.
- , J. L. Betancourt, S. T. Gray, M. A. Palecki, and H. G. Hidalgo, 2008: Associations of multi-decadal sea-surface temperature variability with US drought. *Quat. Int.*, **188**, 31–40, <https://doi.org/10.1016/j.quaint.2007.07.001>.
- Mehrotra, R., and A. Sharma, 2006: A nonparametric stochastic downscaling framework for daily rainfall at multiple locations. *J. Geophys. Res.*, **111**, D15101, <https://doi.org/10.1029/2005JD006637>.
- , and —, 2007: A semi-parametric model for stochastic generation of multi-site daily rainfall exhibiting low-frequency variability. *J. Hydrol.*, **335**, 180–193, <https://doi.org/10.1016/j.jhydrol.2006.11.011>.
- Mo, K. C., J. N. Paegle, and R. W. Higgins, 1997: Atmospheric processes associated with summer floods and droughts in the central United States. *J. Climate*, **10**, 3028–3046, [https://doi.org/10.1175/1520-0442\(1997\)010<3028:APAWSF>2.0.CO;2](https://doi.org/10.1175/1520-0442(1997)010<3028:APAWSF>2.0.CO;2).
- Namias, J., 1966: Nature and possible causes of the northeastern United States drought during 1962–65. *Mon. Wea. Rev.*, **94**, 543–554, [https://doi.org/10.1175/1520-0493\(1966\)094<0543:NAPCOT>2.3.CO;2](https://doi.org/10.1175/1520-0493(1966)094<0543:NAPCOT>2.3.CO;2).
- NOAA, 2014: Twentieth Century Reanalysis (V2c) data. NOAA, https://www.esrl.noaa.gov/psd/data/gridded/data.20thC_ReanV2c.html.
- Nigam, S., B. Guan, and A. Ruiz-Barradas, 2011: Key role of the Atlantic Multidecadal Oscillation in 20th century drought and wet periods over the Great Plains. *Geophys. Res. Lett.*, **38**, L16713, <https://doi.org/10.1029/2011GL048650>.
- Novoa, J., J. Fredes, V. Poblete, and N. B. Yoma, 2018: Uncertainty weighting and propagation in DNN–HMM-based speech recognition. *Comput. Speech Lang.*, **47**, 30–46, <https://doi.org/10.1016/j.csl.2017.06.005>.
- Pal, I., A. W. Robertson, U. Lall, and M. A. Cane, 2015: Modeling winter rainfall in Northwest India using a hidden Markov model: Understanding occurrence of different states and their dynamical connections. *Climate Dyn.*, **44**, 1003–1015, <https://doi.org/10.1007/s00382-014-2178-5>.
- Pederson, G. T., S. T. Gray, D. B. Fagre, and L. J. Graumlich, 2006: Long-duration drought variability and impacts on ecosystem services: A case study from Glacier National Park, Montana. *Earth Interact.*, **10**, 1–28, <https://doi.org/10.1175/EI153.1>.
- Pinter, N., 2005: One step forward, two steps back on U.S. floodplains. *Science*, **308**, 207–208, <https://doi.org/10.1126/science.1108411>.
- Prairie, J., K. Nowak, B. Rajagopalan, U. Lall, and T. Fulp, 2008: A stochastic nonparametric approach for streamflow generation combining observational and paleoreconstructed data. *Water Resour. Res.*, **44**, W06423, <https://doi.org/10.1029/2007WR006684>.
- Redmond, K. T., and R. W. Koch, 1991: Surface climate and streamflow variability in the western United States and their relationship to large-scale circulation indices. *Water Resour. Res.*, **27**, 2381–2399, <https://doi.org/10.1029/91WR00690>.
- Robertson, A. W., S. Kirshner, and P. Smyth, 2004: Downscaling of daily rainfall occurrence over northeast Brazil using a hidden Markov model. *J. Climate*, **17**, 4407–4424, <https://doi.org/10.1175/JCLI-3216.1>.
- Rodríguez-Iturbe, I., 1969: Applications of the theory of runs to hydrology. *Water Resour. Res.*, **5**, 1422–1426, <https://doi.org/10.1029/WR005i006p01422>.
- Russo, T. A., and U. Lall, 2017: Depletion and response of deep groundwater to climate-induced pumping variability. *Nat. Geosci.*, **10**, 105–108, <https://doi.org/10.1038/ngeo2883>.
- Salas, J. D., C. Fu, A. Cancelliere, D. Dustin, D. Bode, A. Pineda, and E. Vincent, 2005: Characterizing the severity and risk of drought in the Poudre River, Colorado. *J. Water Resour. Plann. Manage.*, **131**, 383–393, [https://doi.org/10.1061/\(ASCE\)0733-9496\(2005\)131:5\(383\)](https://doi.org/10.1061/(ASCE)0733-9496(2005)131:5(383)).
- Schmidt, G. A., and Coauthors, 2011: Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.0). *Geosci. Model Dev.*, **4**, 33–45, <https://doi.org/10.5194/gmd-4-33-2011>.
- Seager, R., N. Harnik, Y. Kushnir, W. Robinson, and J. Miller, 2003: Mechanisms of hemispherically symmetric climate variability. *J. Climate*, **16**, 2960–2978, [https://doi.org/10.1175/1520-0442\(2003\)016<2960:MOHSCV>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<2960:MOHSCV>2.0.CO;2).
- , Y. Kushnir, C. Herweijer, N. Naik, and J. Velez, 2005: Modeling of tropical forcing of persistent droughts and pluvials over western North America: 1856–2000. *J. Climate*, **18**, 4065–4088, <https://doi.org/10.1175/JCLI3522.1>.
- , N. Pederson, Y. Kushnir, J. Nakamura, and S. Jurburg, 2012: The 1960s drought and the subsequent shift to a wetter climate in the Catskill Mountains region of the New York City watershed. *J. Climate*, **25**, 6721–6742, <https://doi.org/10.1175/JCLI-D-11-00518.1>.
- Silliman, B. R., J. van de Koppel, M. D. Bertness, L. E. Stanton, and I. A. Mendelssohn, 2005: Drought, snails, and large-scale die-off of southern U.S. salt marshes. *Science*, **310**, 1803–1806, <https://doi.org/10.1126/science.1118229>.
- St. George, S., D. M. Meko, and E. R. Cook, 2010: The seasonality of precipitation signals embedded within the North American drought atlas. *Holocene*, **20**, 983–988, <https://doi.org/10.1177/0959683610365937>.
- Stahle, D. W., F. K. Fye, E. R. Cook, and R. D. Griffin, 2007: Tree-ring reconstructed megadroughts over North America since A.D. 1300. *Climatic Change*, **83**, 133–149, <https://doi.org/10.1007/s10584-006-9171-x>.
- Stanford University, 2016: National Performance of Dams Program. Stanford University, <http://npdp.stanford.edu/>.
- Steinschneider, S., M. Ho, E. R. Cook, and U. Lall, 2016: Can PDSI inform extreme precipitation?: An exploration with a 500 year long paleoclimate reconstruction over the U.S. *Water Resour. Res.*, **52**, 3866–3880, <https://doi.org/10.1002/2016WR018712>.
- Sun, X., and U. Lall, 2015: Spatially coherent trends of annual maximum daily precipitation in the United States. *Geophys. Res. Lett.*, **42**, 9781–9789, <https://doi.org/10.1002/2015GL066483>.
- Sutton, R. T., and D. L. R. Hodson, 2005: Atlantic Ocean forcing of North American and European summer climate. *Science*, **309**, 115–118, <https://doi.org/10.1126/science.1109496>.
- 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>.
- Thyer, M., and G. Kuczera, 2003a: A hidden Markov model for modelling long-term persistence in multi-site rainfall time series 1. Model calibration using a Bayesian approach. *J. Hydrol.*, **275**, 12–26, [https://doi.org/10.1016/S0022-1694\(02\)00412-2](https://doi.org/10.1016/S0022-1694(02)00412-2).
- , and —, 2003b: A hidden Markov model for modelling long-term persistence in multi-site rainfall time series. 2. Real data analysis. *J. Hydrol.*, **275**, 27–48, [https://doi.org/10.1016/S0022-1694\(02\)00411-0](https://doi.org/10.1016/S0022-1694(02)00411-0).
- , A. J. Frost, and G. Kuczera, 2006: Parameter estimation and model identification for stochastic models of annual hydrological

- data: Is the observed record long enough? *J. Hydrol.*, **330**, 313–328, <https://doi.org/10.1016/j.jhydrol.2006.03.029>.
- Trenberth, K. E., 1997: The definition of El Niño. *Bull. Amer. Meteor. Soc.*, **78**, 2771–2778, [https://doi.org/10.1175/1520-0477\(1997\)078<2771:TDOENO>2.0.CO;2](https://doi.org/10.1175/1520-0477(1997)078<2771:TDOENO>2.0.CO;2).
- , and C. J. Guillemot, 1996: Physical processes involved in the 1988 drought and 1993 floods in North America. *J. Climate*, **9**, 1288–1298, [https://doi.org/10.1175/1520-0442\(1996\)009<1288:PPIITD>2.0.CO;2](https://doi.org/10.1175/1520-0442(1996)009<1288:PPIITD>2.0.CO;2).
- Vicente-Serrano, S. M., S. Beguería, J. I. López-Moreno, M. Angulo, and A. El Kenawy, 2010: A new global 0.5° gridded dataset (1901–2006) of a multiscalar drought index: Comparison with current drought index datasets based on the Palmer drought severity index. *J. Hydrometeorol.*, **11**, 1033–1043, <https://doi.org/10.1175/2010JHM1224.1>.
- Visser, I., and M. Speekenbrink, 2010: depmixS4: An R package for hidden Markov models. *J. Stat. Software*, **36**, 1–21, <https://doi.org/10.18637/jss.v036.i07>.
- Viterbi, A., 1967: Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Trans. Inf. Theory*, **13**, 260–269, <https://doi.org/10.1109/TIT.1967.1054010>.
- Wahl, E. R., E. Cook, H. F. Diaz, and D. M. Meko, 2012: Reconstructing precipitation from temperature and drought-index reconstructions in western North America. *2012 Fall Meeting*, San Francisco, CA, Amer. Geophys. Union, Abstract PP21B-1989.
- Webber, M. E., 2008: Catch-22: Water vs. energy. *Sci. Amer.*, **18**, 34–41, <https://doi.org/10.1038/scientificamericanearth0908-34>.
- Wiener, J. D., R. S. Pulwarty, and D. Ware, 2016: Bite without bark: How the socioeconomic context of the 1950s U.S. drought minimized responses to a multiyear extreme climate event. *Wea. Climate Extremes*, **11**, 80–94, <https://doi.org/10.1016/j.wace.2015.11.007>.
- Wilson, R., E. Cook, R. D'Arrigo, N. Riedwyl, M. N. Evans, A. Tudhope, and R. Allan, 2010: Reconstructing ENSO: The influence of method, proxy data, climate forcing and teleconnections. *J. Quat. Sci.*, **25**, 62–78, <https://doi.org/10.1002/jqs.1297>.
- Woodhouse, C. A., K. E. Kunkel, D. R. Easterling, and E. R. Cook, 2005: The twentieth-century pluvial in the western United States. *Geophys. Res. Lett.*, **32**, L07701, <https://doi.org/10.1029/2005GL022413>.
- Yen, B. C., and Y. K. Tung, 1993: *Reliability and Uncertainty Analyses in Hydraulic Design*. American Society of Civil Engineers, 305 pp.
- Yoo, J., H.-H. Kwon, T.-W. Kim, and J.-H. Ahn, 2012: Drought frequency analysis using cluster analysis and bivariate probability distribution. *J. Hydrol.*, **420–421**, 102–111, <https://doi.org/10.1016/j.jhydrol.2011.11.046>.
- Yoo, J. H., A. W. Robertson, and I.-S. Kang, 2010: Analysis of intraseasonal and interannual variability of the Asian summer monsoon using a hidden Markov model. *J. Climate*, **23**, 5498–5516, <https://doi.org/10.1175/2010JCLI3473.1>.
- Zucchini, W., and P. Guttorp, 1991: A hidden Markov model for space-time precipitation. *Water Resour. Res.*, **27**, 1917–1923, <https://doi.org/10.1029/91WR01403>.