



## Water Resources Research

### RESEARCH ARTICLE

10.1002/2016WR019632

#### Key Points:

- A new reconstruction of streamflow spanning the conterminous United States region is developed
- Hierarchical clustering and wavelet analysis is used to characterize 555 years of regional and temporal streamflow patterns
- Spatial differences in CONUS streamflow are opportunities for reducing national drought risk

#### Supporting Information:

- Supporting Information S1

#### Correspondence to:

M. Ho,  
mh3538@columbia.edu

#### Citation:

Ho, M., U. Lall, X. Sun, and E. R. Cook (2017), Multiscale temporal variability and regional patterns in 555 years of conterminous U.S. streamflow, *Water Resour. Res.*, 53, 3047–3066,  
doi:10.1002/2016WR019632.

Received 10 AUG 2016

Accepted 12 MAR 2017

Accepted article online 16 MAR 2017

Published online 13 APR 2017

## Multiscale temporal variability and regional patterns in 555 years of conterminous U.S. streamflow

Michelle Ho<sup>1</sup> , Upmanu Lall<sup>1,2</sup> , Xun Sun<sup>1,3</sup> , and Edward R. Cook<sup>4</sup> 

<sup>1</sup>Columbia Water Center, Columbia University, New York, New York, USA, <sup>2</sup>Department of Earth and Environmental Engineering, Columbia University, New York, New York, USA, <sup>3</sup>Key Laboratory of Geographic Information Science (Ministry of Education), East China Normal University, Shanghai, 200241, China, <sup>4</sup>Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York, USA

**Abstract** The development of paleoclimate streamflow reconstructions in the conterminous United States (CONUS) has provided water resource managers with improved insights into multidecadal and centennial scale variability that cannot be reliably detected using shorter instrumental records. Paleoclimate streamflow reconstructions have largely focused on individual catchments limiting the ability to quantify variability across the CONUS. The Living Blended Drought Atlas (LBDA), a spatially and temporally complete 555 year long paleoclimate record of summer drought across the CONUS, provides an opportunity to reconstruct and characterize streamflow variability at a continental scale. We explore the validity of the first paleo-reconstructions of streamflow that span the CONUS informed by the LBDA targeting a set of U.S. Geological Survey streamflow sites. The reconstructions are skillful under cross validation across most of the country, but the variance explained is generally low. Spatial and temporal structures of streamflow variability are analyzed using hierarchical clustering, principal component analysis, and wavelet analyses. Nine spatially coherent clusters are identified. The reconstructions show signals of contemporary droughts such as the Dust Bowl (1930s) and 1950s droughts. Decadal-scale variability was detected in the late 1900s in the western U.S., however, similar modes of temporal variability were rarely present prior to the 1950s. The twentieth century featured longer wet spells and shorter dry spells compared with the preceding 450 years. Streamflows in the Pacific Northwest and Northeast are negatively correlated with the central U.S. suggesting the potential to mitigate some drought impacts by balancing economic activities and insurance pools across these regions during major droughts.

**Plain Language Summary** The development of paleoclimate streamflow reconstructions in the conterminous United States (CONUS) has provided water resource managers with improved insights into multidecadal and centennial scale variability that cannot be reliably detected using shorter instrumental records. However, most reconstructions to date have focused on streamflow at a catchment scale. Here we use the Living Blended Drought Atlas (LBDA), a spatially and temporally complete 555 year long paleoclimate record of summer drought across the CONUS, to reconstruct and characterize streamflow variability at a continental scale. The reconstructions show signals of contemporary droughts such as the Dust Bowl (1930s) and 1950s droughts. Decadal-scale variability was detected in the late 1900s in the western U.S., however, similar modes of temporal variability were rarely present prior to the 1950s. The twentieth century featured longer wet spells and shorter dry spells compared with the preceding 450 years. Streamflows in the Pacific Northwest and Northeast are negatively correlated with the central U.S. suggesting the potential to mitigate some drought impacts by balancing economic activities and insurance pools across these regions during major droughts.

### 1. Introduction

Observations of streamflow in the conterminous United States (CONUS) exhibit variability on interannual through to multidecadal timescales. Recent changes in streamflow availability and water demands, including prolonged droughts, increased competition for water resources through regional population increases, and the push to expand economic activities, have prompted the search for improved knowledge of natural streamflow variability. One approach has been to develop streamflow reconstructions using paleoclimate proxy records [Woodhouse *et al.*, 2016]. Paleohydrologic records have been used to place recent

hydrological extremes in the context of climate variability and improve estimates of flood and drought risk [e.g., Stockton and Jacoby, 1976]. Paleohydrologic records have likewise informed evaluations of water availability for in-stream water withdrawals, optimization of reservoir operations, assessments of water-dependent economic activities, and improving the accuracy of streamflow forecasts over longer lead times [e.g., Woodhouse and Lukas, 2006b; Carrier et al., 2013; Devineni et al., 2013].

Paleoclimate reconstructions of streamflow in the CONUS have, to date, been primarily focused on catchment-scale variability [Woodhouse et al., 2002] with the exception of a small number of interbasin analyses [e.g., Meko and Woodhouse, 2005; Meko et al., 2012; Carrier et al., 2013]. Catchment-specific studies have included streamflow reconstructions in the Colorado River Basin [Stockton and Jacoby, 1976; Woodhouse et al., 2006a; Woodhouse and Lukas, 2006a, 2006b], watersheds supplying major centers such as New York City [Pederson et al., 2012; Devineni et al., 2013], and regions experiencing increased pressures on water supplies such as Phoenix, Arizona [Smith and Stockton, 1981; Hirschboeck and Meko, 2005]. However, there is a dearth of streamflow reconstructions over much of the eastern CONUS [Woodhouse et al., 2006b]. A national-scale perspective would inform the national availability of water and identify implications for the water-energy-food nexus as well as ecosystems services. A consideration of these nexus questions can elevate the interest from watershed to national policy questions, as questions for facility siting for energy production, which crops to grow where, and how to maintain ecosystems services come to the fore, while assuring public water supplies in rapidly growing or declining areas. A national perspective on water-dependent economic risks (e.g., electricity production [Diehl and Harris, 2014], irrigated agriculture [Iglesias et al., 2003], and water supply [Rajagopalan et al., 2009; Gleick, 2016]) and its spatiotemporal variation is needed for future investment and resource allocation decisions.

Cataloging the characteristics of year-to-year variations in streamflow across the CONUS is required in order to understand linkages in flow regimes between basins and regions across time [Probst and Tardy, 1987; Dettinger and Diaz, 2000]. Streamflow across the CONUS has previously been analyzed using instrumental records of streamflow. One such study was by Lins [1985] who used gauged streamflow records from 1931 to 1978 to produce a gridded streamflow database. The main modes of streamflow variability were then extracted using orthogonal (in this case unrotated and varimax) principle component analysis (PCA). Lins [1985] found that patterns of CONUS streamflow variability were comparable with CONUS-scale analyses of precipitation [Walsh and Mostek, 1980] and drought variability [Karl and Koscielny, 1982].

Some notable characteristics of CONUS streamflow include nationwide variability that showed anomalously high streamflow in the 1940s and 1970s [Lins, 1985]. In contrast, anomalously low streamflow over much of the CONUS region was identified in the 1930s, approximating the Dust Bowl Drought [Gregory, 1991; Schubert et al., 2004], and in 1953–1956, approximating the 1950s drought [Fye et al., 2003; Wiener et al., 2016]. However, the Pacific Northwest region showed a notable wet spell from 1948 to 1956 contrasting the hydrologic drought that occurred over the remainder of the nation. Lins [1985] also noted the pattern of opposing streamflow anomalies between the Northwest and Southwest later confirmed in studies of Pacific Ocean teleconnections in the western U.S. [e.g., Redmond and Koch, 1991; McCabe and Dettinger, 1999]. Lins [1985] found less interannual variability in the Northeast, Great Plains, and Midwest regions. However, negative streamflow anomalies were still notable from 1962 to 1966 in the Northeast, corresponding with the drought noted by Namias [1966], followed by an extended wet spell from the 1970s until present [Pederson et al., 2012]. In addition, anomalous negative streamflow in the Great Plains and Midwest coincide with the Dust Bowl and 1950s droughts. Interestingly, historic streamflow anomalies in the Rio Grande area around south-west Texas are typically opposite in sign to streamflow anomalies in the Great Plains region [Busby, 1964; Nace and Pluhowski, 1965; Lins, 1985].

In addition to identifying regional streamflow variability, assessments of annual streamflow characteristics in the CONUS have also included investigations of streamflow trends. In general, there have been few locations that have shown trends in annual maximum streamflows, while low to moderate streamflows (e.g., annual minimum and median) have largely increased (i.e., decrease in hydrological drought), with the exception of the Pacific Northwest and Southeast [Lins and Slack, 1999, 2005]. McCabe and Wolock [2002] identified a step change increase in low to moderate streamflow around the 1970s coinciding with increased precipitation, that has likewise shown increases in heavy rainfall events across much of the CONUS region [Karl and Knight, 1998]. These studies have elucidated key characteristics of streamflow variability across the CONUS since the 1930s. However, it has been demonstrated that CONUS streamflow is

influenced by large-scale modes of climate variability that persist on decadal and multidecadal scales [e.g., Barlow *et al.*, 2001; Enfield *et al.*, 2001; Mo *et al.*, 2009; Nigam *et al.*, 2011] and the ability to identify long-term variability in streamflow is limited by the length of instrumental records.

In order to facilitate a national assessment of multicentennial streamflow variability, we present a 555 year-long paleoclimate reconstruction of annual streamflow at locations across the CONUS (further described in section 2.4). Methods of characterizing continental-scale streamflow variability are many and varied [Poff *et al.*, 2006]. Here we identify long-term characteristics of CONUS-wide streamflow using multicentury CONUS streamflow by first grouping streamflows with similar patterns of variability using hierarchical clustering applied in the time domain to standardized streamflow. The temporal variability and frequency characteristics of regional streamflow as organized by cluster are then investigated.

## 2. Data and Methods

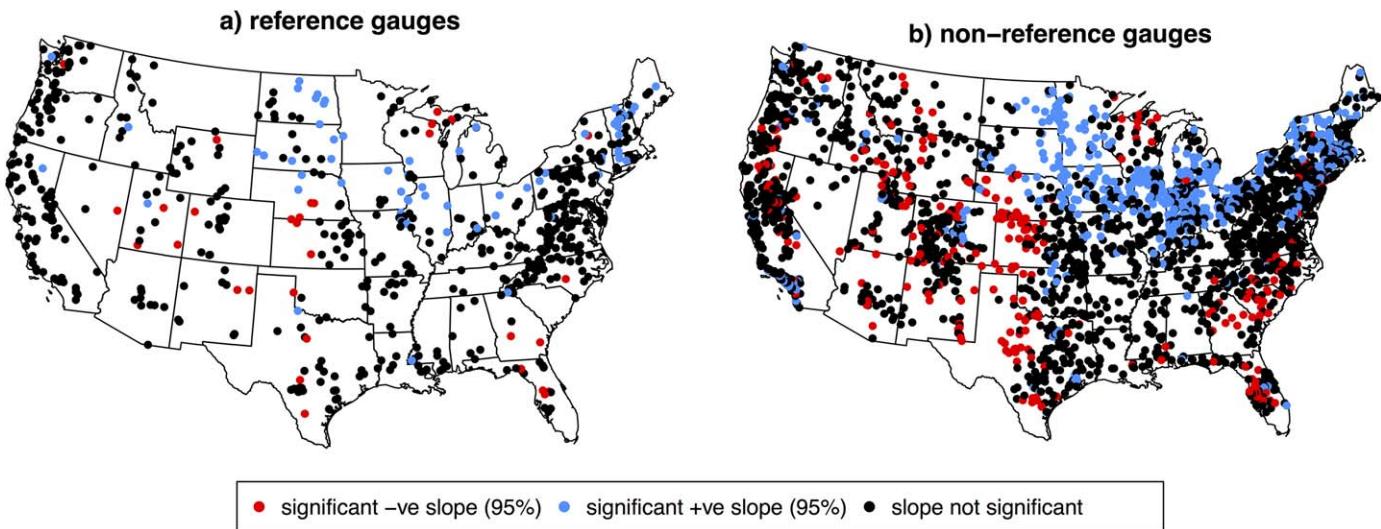
### 2.1. Streamflow Data

Streamflow records from the U.S. Geological Survey (USGS) [2011] were screened for length and completeness using a criterion of a minimum record length of 40 years and less than 5% missing monthly data resulting in the selection of 627 reference gauges. USGS reference gauges [Falcone *et al.*, 2010] represent relatively unregulated flows with minimal catchment modification as detailed in U.S. Geological Survey [2012]. In addition to reference gauges, nonreference gauges, which do not meet the USGS requirement for undisturbed catchments and unregulated streamflows, were also considered and an additional 2597 nonreference gauges were included in the analysis that met the same criteria of minimum record length and completeness. Station details are documented, along with the resulting regional streamflow reconstructions, on the NOAA paleoclimate database (see Acknowledgment for details). For the nonreference gauges, consumptive water withdrawals in developed catchments alter the annual streamflow record resulting in a reconstructed streamflow that is unlikely to reflect the actual prehistoric streamflow. Information regarding the dates at which the nonreference streamflow become regulated or altered by changes in catchment characteristics and the degree to which streamflow is altered is not readily available. However, the streamflow reconstructions may represent the streamflow that would have occurred under the climate conditions described by the Living Blended Drought Atlas and may still preserve interannual and multidecadal variability, which are of interest. While the use of nonreference gauges may be problematic, the nonreference gauges will be retained in the analysis in order to achieve a better spatial coverage of streamflow variability. Consideration of the differences between reference and nonreference gauges will be given in presenting the reconstruction results.

Streamflow was aggregated to an annual water year (October–September) as opposed to the calendar year aggregation used in Ho *et al.* [2016] as a water year was found to have the strongest relation to the LBDA across the CONUS (see Figure S1 in the supporting information). A Kolmogorov-Smirnov (K-S) normality test was performed on annual (water year) log-transformed streamflow (see Figure S2 in supporting material). The K-S test showed that the assumption of lognormally distributed streamflow could not be rejected at the 95% significance level in over 99% and 95% of reference and nonreference streamflow gauges, respectively. The Sen's slope estimator [Theil, 1992] was also calculated for all gauges to determine whether streamflow showed significant monotonic trends (Figure 1ii) based on the median of pairwise slopes. Of the 627 reference gauges and 2597 nonreference gauges, 96 (15.3%) and 861 (33.2%) of streamflow sites, respectively, showed a significant slope at the 95% level. For both reference and nonreference gauges, there were approximately twice as many gauges that showed a significant positive slope compared with a negative slope. Most gauges showing significant trends were located in the Midwest around the Great Lakes and around the 100°W meridian from North Dakota through to Texas.

### 2.2. Paleoclimate Data

An updated version of the Cook *et al.* [2010] Living Blended Drought Atlas (LBDA) was used, which itself was an updated version of the seminal North American Drought Atlas [Cook *et al.*, 1999, 2004]. The LBDA is a tree-ring-informed paleoclimate reconstruction of the summer (June–August) modified Palmer Drought Severity Index (PDSI) from Heim *et al.* [2007]. The modified PDSI was an improvement on the original PDSI [Palmer, 1965] to address issues of continuity in PDSI values when transitioning between drought states and wet spells [Heddinghaus and Sabol, 1991]. PDSI is a locally normalized soil moisture metric based on atmospheric moisture supply and natural surface demand [Dai *et al.*, 2004]. Negative values indicate anomalously

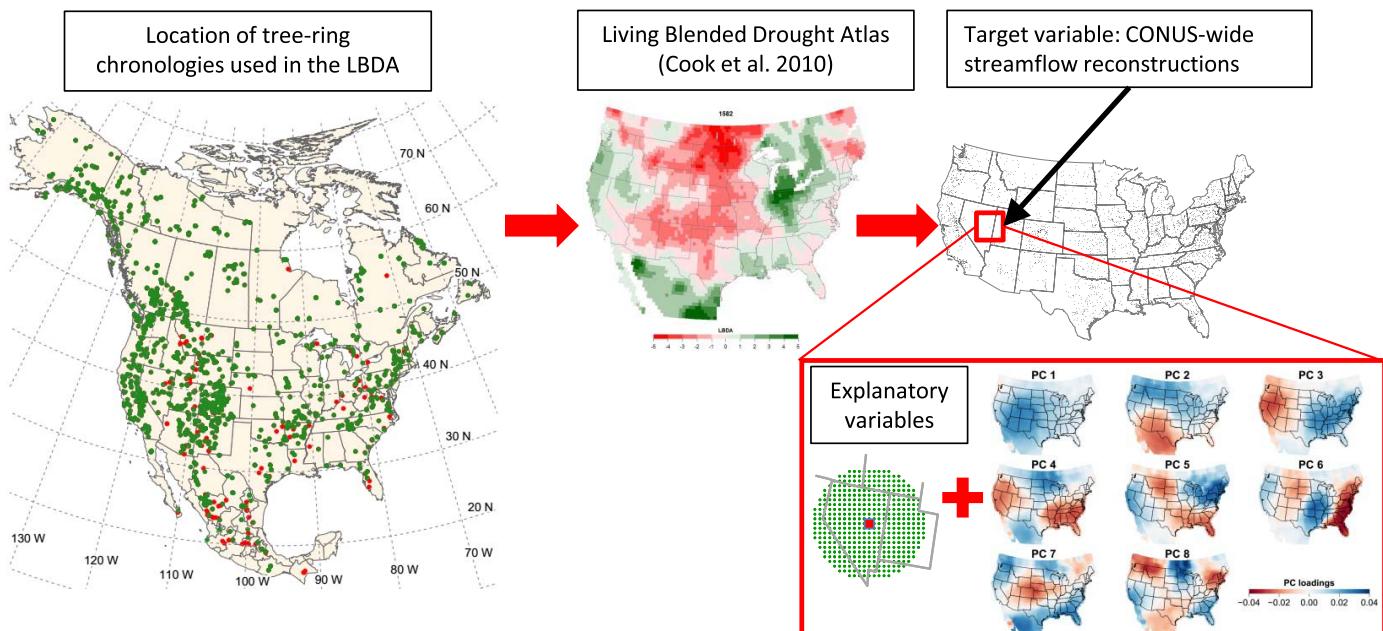


**Figure 1.** Significant Sen slopes showing trends. Results are separated into (a) reference and (b) nonreference gauges. Of the reference gauges, 31 gauges showed a significant negative slope, while 65 gauges showed a significant positive slope. For the nonreference gauges, 297 gauges showed a significant negative slope, while 564 stations showed a significant positive slope.

dry conditions and positive values indicate anomalously wet conditions [Palmer, 1965]. The LBDA paleoclimate reconstruction of the modified PDSI over the CONUS region is spatially complete from 1451 to 2005. The updated LBDA includes an additional 99 tree-ring chronologies (see red dots in Figure 2, locations of tree-ring chronologies used in the LBDA), resulting in total of 1944 tree-ring chronologies. The additional chronologies improved the spatial resolution of the LBDA and enabled spatially complete reconstructions of the PDSI over the CONUS region to be extended back in time.

### 2.3. Reconstructing Streamflow Using the LBDA

We extend the approach applied to reconstructing streamflow in the Missouri River Basin [Ho et al., 2016] with a few updates. An expanded network of streamflow gauges as described in section 2.1 was used



**Figure 2.** Location of (left) tree-ring chronologies used to reconstruct the Palmer Drought Severity Index known as the (middle) LBDA and LBDA inputs of grids within a 450 km radius and the first eight PCs of CONUS-wide LBDA (red box) used to reconstruct streamflow at each gauge in the CONUS. Both national-scale and regional-scale features of the LBDA are thus used as candidates for the streamflow reconstruction at a given gage.

covering the CONUS region. In addition, we used an updated version of the LBDA as described in section 2.2 that extends further back in time. We utilize a spatially and temporally complete gridded paleoclimate reconstruction of PDSI from tree-ring chronologies (i.e., the LBDA) to reconstruct streamflow. Our approach aims to exploit the similarities between streamflow and PDSI as both variables represent different filters on the underlying precipitation and temperature variations, and common factors in these filters could result in a relationship between PDSI and streamflow. Thus, the LBDA reconstruction of PDSI from tree-rings could potentially provide an effective representation of the streamflow, at least in locations where the common signals are strong. The approach to reconstructing CONUS streamflow is therefore based on the preservation of climate signals in tree-rings used to reconstruct PDSI, which is then used to reconstruct streamflow (see flowchart of modeling approach in Figure 2). While the direct use of moisture-sensitive tree-ring chronologies to reconstruct streamflow is more widespread, the approach of using PDSI is not entirely novel given previous attempts to reconstruct streamflow from drought indices [e.g., *Graham and Hughes, 2007; Adams et al., 2015; Coulthard et al., 2016*].

The streamflow covariates were LBDA data from grids within a 450 km radius (the same radius used to produce the LBDA directly from tree-rings) of the target streamflow gauge and the first eight principal components (PCs) of CONUS-wide LBDA variability, as per the covariate selection in *Ho et al. [2016]* (see schematic of explanatory variables in Figure 2). The covariate selection results in several hundred LBDA inputs. A dimension reduction of the explanatory variables was achieved by taking the first canonical variate of the LBDA inputs calculated using regularized canonical correlation analysis (rCCA) to maximize the correlation between a rotation of the LBDA predictors and the target streamflow. The first canonical variate of the LBDA inputs was used as the streamflow predictor at each gauge. The rCCA procedure used to reduce the dimensionality of the LBDA inputs is briefly sketched below, followed by the streamflow estimation model that uses the resulting canonical variate.

Canonical correlation analysis (CCA) is a method of linearly transforming two vector variables to canonical form to maximize the correlation between the variables [*Hotelling, 1936*] as follows: consider two sets of random variables represented by matrices  $\mathbf{X}$  and  $\mathbf{Y}$ .  $\mathbf{X}$  is a  $n \times p$  matrix where  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_p]$  containing  $n$  observations at  $p$  different locations, while  $\mathbf{Y}$  is a  $n \times q$  matrix where  $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_q]$  containing  $n$  observations at  $q$  different locations. Both  $\mathbf{X}$  and  $\mathbf{Y}$  have finite variance matrices represented by  $\Sigma_{XX}$  and  $\Sigma_{YY}$ , respectively.  $\Sigma_{XY}$  is the covariance matrix between  $\mathbf{X}$  and  $\mathbf{Y}$  while  $\Sigma_{YX}$  is the covariance matrix between  $\mathbf{Y}$  and  $\mathbf{X}$ .

Here  $\mathbf{X}$  consists of the  $p$  LBDA inputs specific to each streamflow gauge (see Figure 2) and  $\mathbf{Y}$  is the log-transformed streamflow at the gauge. CCA involves rotating the coordinate axes of both  $\mathbf{X}$  and  $\mathbf{Y}$  to new coordinate systems in order to maximize the correlation between the rotated components of  $\mathbf{X}$  and  $\mathbf{Y}$ ,  $\mathbf{U} = \mathbf{X} \times \alpha$  and  $\mathbf{V} = \mathbf{Y} \times \gamma$ , respectively.  $\mathbf{U}$  and  $\mathbf{V}$  yield the first pair of canonical variates, while  $\alpha$  and  $\gamma$  are the vectors of canonical weights of length  $p$  and  $q$ , respectively. An eigen decomposition of  $\Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}$  and  $\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}$  is used to calculate successive pairs of canonical variates and the resulting  $\min(p, q)$  eigenvalues are common to both. The canonical weights,  $\alpha$  and  $\gamma$ , which are used to transform  $\mathbf{X}$  and  $\mathbf{Y}$ , are given by the eigenvectors of  $\Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}$  and  $\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}$ , respectively. Here the weight for  $\mathbf{Y}$  (streamflow) is 1 and the weights for  $\mathbf{X}$  (LBDA) are given by the first eigenvector of  $\Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}$ .

The CCA process was regularized [*Vinod, 1976*] to address the issue of having a large number of predictors (i.e., large  $p$ ) relative to the number of observations (i.e., number of LBDA observations overlapping the streamflow record). Regularization is a smoothing process similar to the technique of ridge regression [*De Bie and De Moor, 2003*]. In regularized CCA (rCCA), a regularization parameter ( $\lambda$ ), or a “roughness penalty,” with values ranging between 0 and 1, is introduced to the variance matrices converting  $\Sigma_{YY}$  and  $\Sigma_{XX}$  to  $\Sigma_{YY} + \lambda_y\mathbf{I}$  and  $\Sigma_{XX} + \lambda_x\mathbf{I}$ , respectively [*Leurgans et al., 1993*]. A suitable value of  $\lambda$  was calculated using a leave-one-out cross validation score that aimed to maximize the correlation between the transformed data sets of LBDA inputs ( $\mathbf{U}$ ) and streamflow ( $\mathbf{V}$ ). In calculating  $\lambda$ , the degrees of freedom (as defined by *Dijkstra [2014]*) was limited to a maximum of  $n - 10$ . A maximum  $\lambda$  value of one was set if the criteria for the maximum degrees of freedom could not be achieved. rCCA was executed in R using the R package “CCA” by *González et al. [2008]* and is freely available from the Comprehensive R Archive Network (CRAN <https://cran.r-project.org/>).

To reconstruct the CONUS streamflow, a generalized linear model (GLM) with a log link was used to model the relationship between streamflow and the first canonical variate of the LBDA inputs. The GLM was specified as follows:

$$g(\mu) = \mathbf{U}\beta \quad (1)$$

where  $\mathbf{U} = \mathbf{X} \times \alpha$  is the first canonical variate of LBDA inputs, unique to each streamflow station;  $\beta$  is the coefficient;  $g$  is a lognormal function; and  $\mu$  is a vector of mean streamflow that is lognormally distributed at each gauge  $i$  at time  $t$  ( $\mu_{it}$ ) with variance  $\sigma_i^2$  as follows:

$$y_{it} \sim LN(\mu_{it}, \sigma_i^2) \quad (2)$$

where  $y_{it}$  is the streamflow at gauge  $i$  and time  $t$ . The parameters of the generalized linear model were fit using maximum likelihood within the `glm` script in R, an open source package.

Verification metrics of variance explained were calculated for each reconstruction, while cross validation was used to calculate the reduction of error (RE) and coefficient of efficiency (CE) equivalent to the Nash-Sutcliff efficiency test where the metrics are normalized using the mean of the calibrated and cross validated data, respectively. That is

$$RE = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_c)^2} \quad (3)$$

and

$$CE = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_v)^2} \quad (4)$$

where  $y_i$  and  $\hat{y}_i$  are the observed and modeled streamflows in the verification period, while  $\bar{y}_c$  and  $\bar{y}_v$  are the mean of the observations in the calibration and verification period, respectively.

The cross validation statistics were calculated on the log-transformed reconstructions using 100 randomly selected leave-k-out samples where  $k$  was equal to 10% of the observations at a streamflow gauge. The model was then fit on the remaining 90% of data, while the k-observations were used for verification resulting in a different set of values used for  $y_i, \hat{y}_i$ , and  $\bar{y}_v$  for each of the 100 resamples. This new reconstruction of streamflow spanning the CONUS region for each of the regional clusters (described in section 2.4) is available on the NOAA paleoclimate database (see Acknowledgment for details) along with the verification metrics.

#### 2.4. Approach for Quantifying Space-Time Streamflow Variability for the CONUS

We explored the spatiotemporal structure in the resulting 555 year long reconstructions of streamflow at all 3224 sites. Specifically, we were interested in seeing if groups of sites that have similar temporal variation at interannual and longer time scales could be identified; whether there were distinct time scales of variation associated with each group; and whether these temporal patterns were associated with known low frequency climate modes. This was achieved using the following steps.

1. Clustering streamflow based on a selected attribute using a hierarchical clustering algorithm.
2. Dimension reduction of streamflow within a cluster using PCA, on the correlation matrix, to extract the main mode of variability (given by the first PC) within the cluster.
3. Applying a wavelet transform on the first PC to identify periods where cyclical temporal variability is significant.

An iterative clustering method called hierarchical clustering was used. Hierarchical clustering has the benefit of not requiring a predetermined number of clusters to be specified. In the agglomerative hierarchical clustering algorithm used here, each object, in this case streamflow at one gauge, is first assigned its own cluster. At each iterative step, the two most similar clusters as measured by the Euclidian distance are joined and this process continues until there is a single cluster. Dissimilarity between clusters is measured at each stage using Ward's criterion that minimizes the intracluster variance and aims to group objects into compact spherical clusters [Ward, 1963; Murtagh and Legendre, 2014].

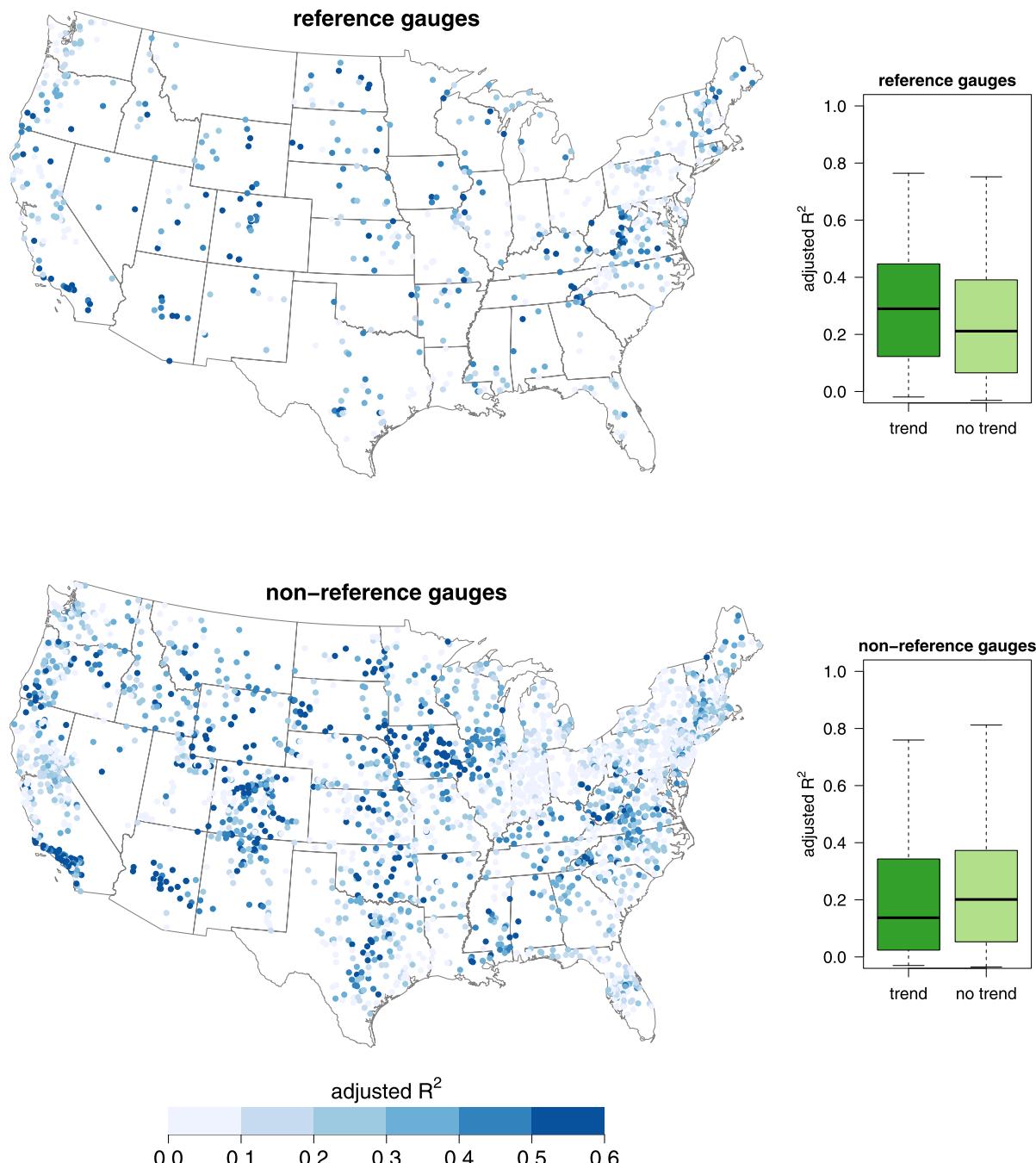
Determining an appropriate number of clusters for a data set is subjective and may be guided using a number of metrics for quantifying the clusters' intracluster variance (compactness), the degree to which objects outside a cluster are similar to those in a given cluster (connectivity) and the difference between clusters (separation). Typically, as the number of clusters increase objects within a cluster become more compact but the degree of separation between clusters is reduced. The Silhouette width and the Dunn index are two metrics that simultaneously measure both compactness and separation. The Silhouette width is the degree of confidence in a particular clustering assignment. Well-clustered observations have a value near 1 and poorly clustered observations have values near -1. The Dunn index is the ratio of the smallest distance between observations in different clusters (separation) to the largest intra cluster distance (compactness). The use of two or three complementary measures enables a rigorous approach to estimating an appropriate number of clusters [Handl *et al.*, 2005]. The connectivity, Dunn index, and Silhouette width were calculated in R using a package called cLValid [Brock *et al.*, 2008] and cluster [Maechler *et al.*, 2015]. In selecting a suitable number of clusters, we wish to minimize the connectivity between clusters while maximizing the Dunn index and Silhouette widths.

Hierarchical clustering was performed on standardized annual streamflow reconstructions (standardized to achieve a mean of zero and standard deviation of one). The purpose of clustering by normalized streamflow was to enable streamflow sites to be grouped based on similar patterns of departures from the mean that indicate regional droughts or regional wet spells. Each cluster of streamflow was further analyzed by extracting the main mode of variability through time represented by the first principal component (PC1). Significant temporal modes of variability were identified using a wavelet transform on PC1 of each cluster using a Morlet mother wavelet and padded with zeros to improve the stability of the wavelet transform. In order to verify the findings and identify significant periodicities, a multitaper method (MTM) was also applied to the first PC of each cluster using three tapers and a bandwidth of 2 years. Significant periodicities were identified using a 95% significance level against red noise. MTM analysis was carried out using KSpectra [Ghil *et al.*, 2002].

### 3. Five Centuries of Reconstructed CONUS Streamflow

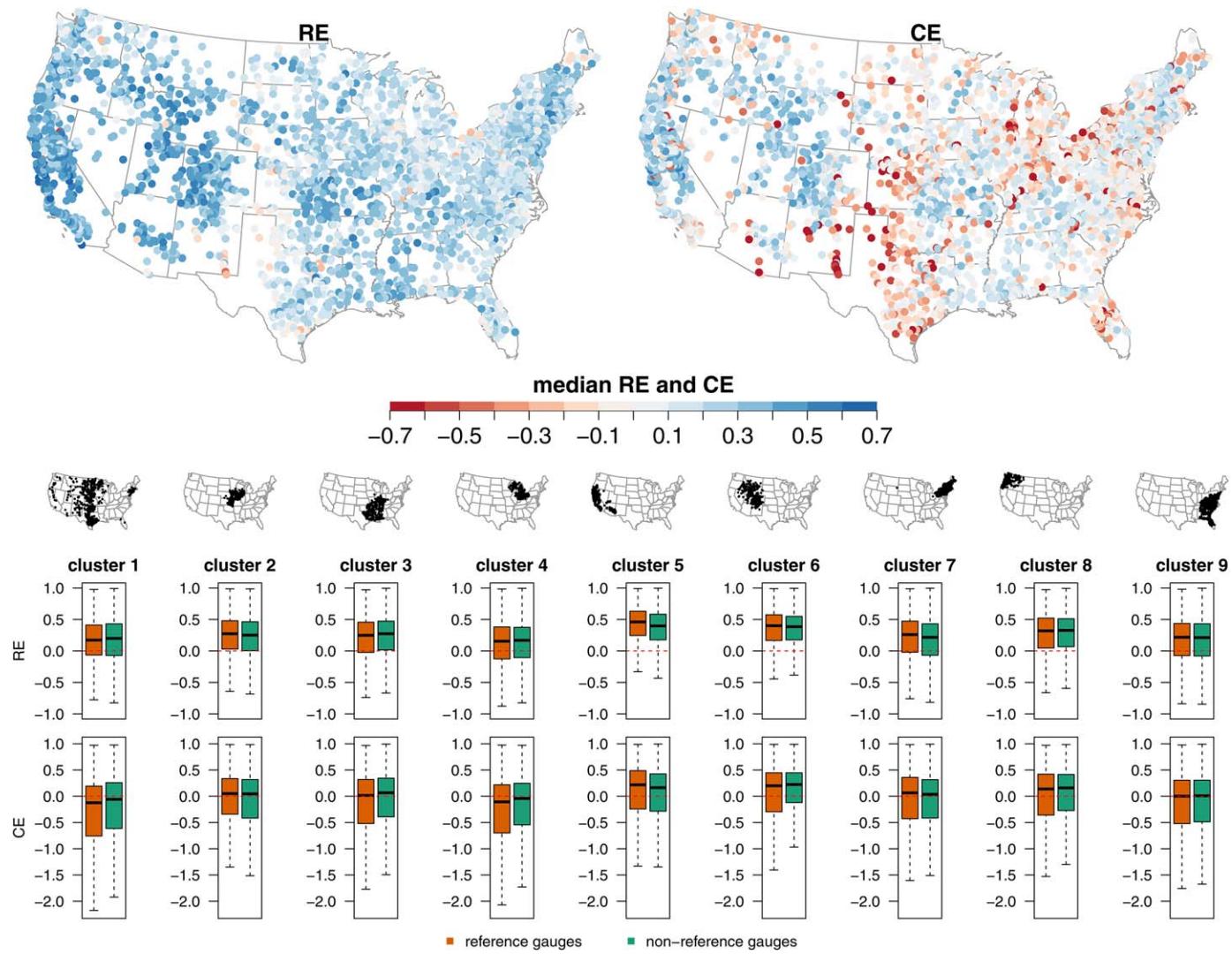
The reconstruction of streamflow across the CONUS used the first canonical variate of the selected LBDA inputs. The rCCA process did not noticeably favor either regional (grids surrounding the streamflow gauge) or national (PCs of CONUS-wide LBDA) covariates. The resulting model of streamflow showed good fit in southern California, the southern Intermountain region (around Colorado) and areas in the Cornbelt region (west of the Great Lakes) and through the mid-east coast (Figure 3, reference gauges are shown in the top and nonreference gauges below). However, in many regions across the CONUS, the streamflow reconstructions did not perform well as measured by degree of variance explained. Of the 3224 streamflow gauges, 1654 (51%) gauges had an adjusted  $R^2$  value of less than 0.2. A two-sided  $t$  test between the distributions of adjusted  $R^2$  values for reference and nonreference gauges showed a small but significant difference in the mean values of 0.028. In addition, there was no conclusive impact on the variance explained when considering streamflows with or without significant monotonic trends (see boxplots of adjusted  $R^2$  for gauges with/without a significant trend in Figure 3) with higher adjusted  $R^2$  values for references gauges with trends and lower adjusted  $R^2$  values for nonreference gauges with trends. However, adjusted  $R^2$  values for reference gauges showing trends was approximately 50% larger than the adjusted  $R^2$  of nonreference gauges displaying trends (a difference of 0.093), suggesting that trends in the nonreference gauges are likely disconnected from climate signals that influence natural streamflow, tree-ring growth, and the LBDA.

Standard reconstruction metrics of RE and CE were also calculated (Figure 4) where values greater than zero represent reconstructions that are more informative than the climatology of calibrated and validated streamflow, respectively [Wilson *et al.*, 2010]. The distributions of RE and CE values for regional clusters (clusters are determined in section 4.1) separated into RE and CE values for reference and nonreference gauges are also shown. RE and CE values were largely positive across the CONUS region. Lower values of both RE and CE including some negative values were found for the south and southwestern regions and around the 100°W meridian. The presence or absence of a significant trend in instrumental streamflow had no consistent impact on the value of either RE or CE across the reconstructions and the distinction between these streamflows is therefore not shown.



**Figure 3.** Adjusted  $R^2$  for annual (October–September) streamflow for (top) reference and (bottom) nonreference gauges fit using a Gaussian generalized linear model with a log link and the first canonical variate of LBDA inputs comprised of LBDA grids within a 450 km radius and the first eight PCs of CONUS LBDA. Distribution of adjusted  $R^2$  values are shown for reference and nonreference gauges for gauges with/without a trend, significant at the 95% level, in instrumental streamflow.

Weaker RE and negative CE values around North and South Dakota correspond to negative CE values in the LBDA reconstruction (shown in Figure 5), suggesting that PDSI values reconstructed with poorly fitting models or insufficient covariates impact the quality of streamflow reconstructions based on the LBDA. However, in locations such as Texas and Florida, LBDA CE values are largely positive (Figure 5) and RE and CE values for the streamflow reconstruction are small or negative, indicating a weaker model fit. The regions around Texas and Florida also correspond with locations that display a significant trend in streamflow (as seen in the significance of the Sen's slope estimator in Figure 1ii). These trends suggest that streamflow at these



**Figure 4.** (top) Median RE and CE values (calculated on log streamflow). (bottom) box plots of median RE and CE values for each cluster—clusters are defined in section 4.1.

locations may be influenced by variables to which LBDA is not sensitive. In the semiarid regions around Texas, New Mexico, and Arizona, base streamflow is often outflow from hydraulically connected shallow groundwater systems [DuMars and Minier, 2004] that are impacted by remote, but hydraulically connected, changes in groundwater extractions used to supply urban and irrigated agricultural water demands [Stromberg *et al.*, 2005]. Similarly, streamflows in some areas of Florida are likewise connected with aquifer systems and groundwater depletions have resulted in reduced streamflow [Hammett, 1990]. Streamflows in the southwest are further complicated by the impacts of changes in vegetation, land use, and land cover on the rainfall-runoff relationship. For example, range restoration has been linked to increased infiltration and reduced runoff [Wilcox *et al.*, 2008], while runoff from excess irrigation has contributed to increased streamflows [Scanlon *et al.*, 2005]. In addition, there is a paucity of tree-ring chronologies in west Texas and localized effects on streamflow, such as groundwater extraction and vegetation changes, are therefore unlikely to be detected in the growth of distant trees. Consequently, these factors may also affect reconstructions done directly through the use of tree-ring chronologies in these regions.

The validity of these CONUS-wide streamflow reconstructions is further discussed in section 5 where qualitative comparisons are made with existing catchment-specific reconstructions of streamflow. We present further analysis of the spatial and temporal variability of streamflow in section 4.

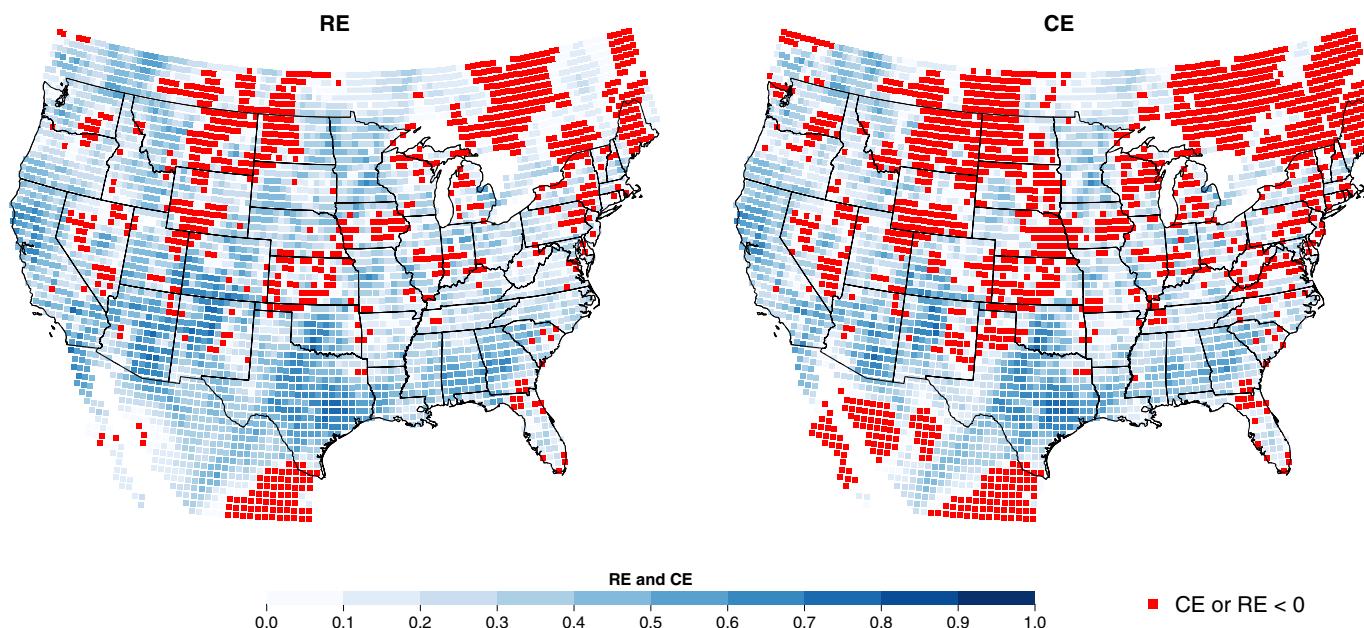


Figure 5. RE and CE values for LBDA reconstructions.

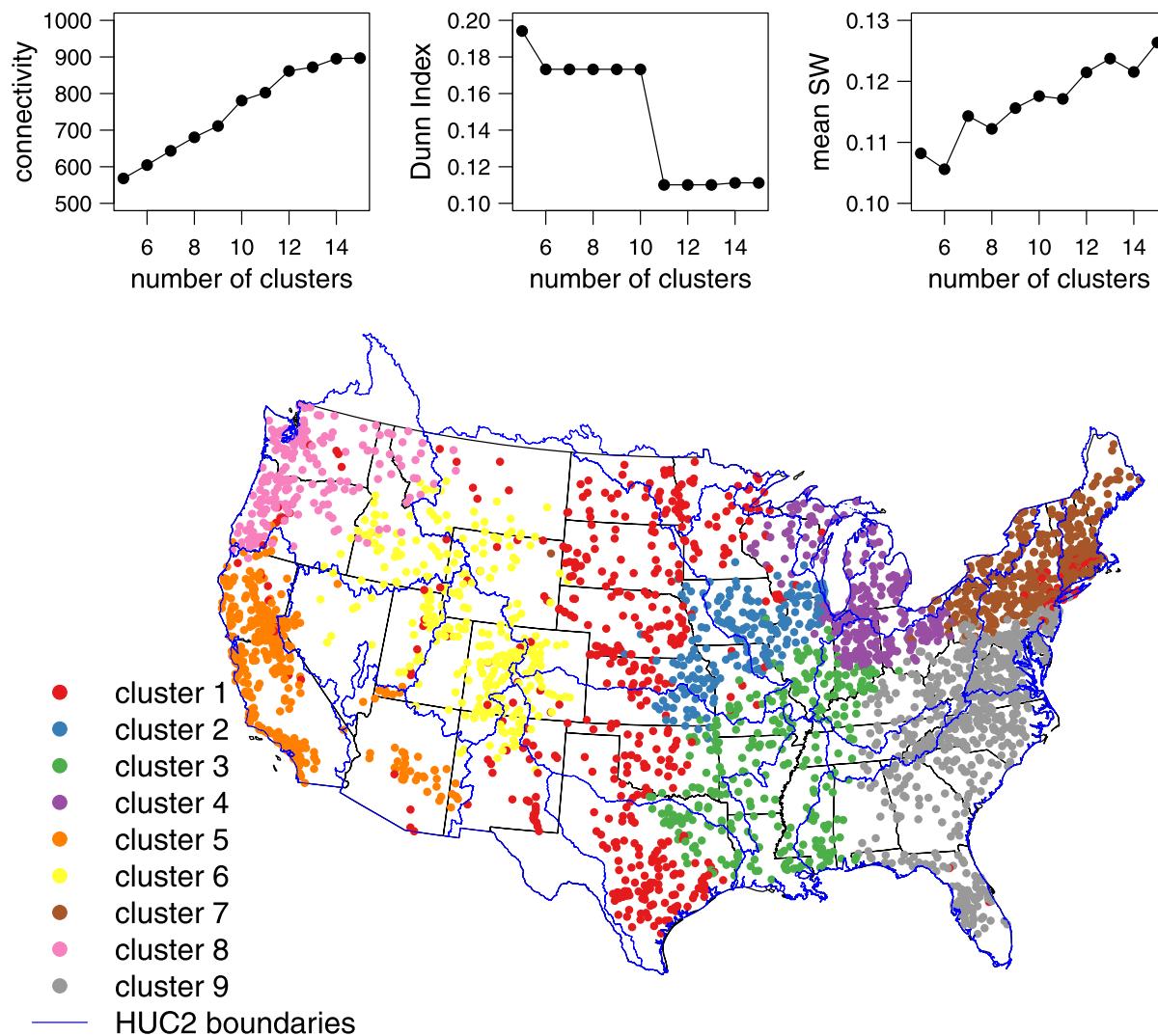
#### 4. Spatial and Temporal Variability of CONUS Streamflow: 1451–2005

##### 4.1. Hierarchical Clustering of Standardized Annual Paleoclimate Streamflow

Spatial and temporal variability of CONUS streamflow was assessed by first clustering streamflow gauges based on the similarity of standardized streamflow (mean of zero and standard deviation of 1) to identify groups of streamflow with similar patterns of anomalies. In order to determine an appropriate number of clusters to retain, clustering validation statistics were calculated for the 555 year reconstructed record for the 3224 stations using between 3 and 15 clusters. The differences in measures of connectivity, the Dunn index, and Silhouette index for the different number of clusters are shown in Figure 6. A grouping of streamflow sites into nine clusters was selected to achieve a balance between minimizing the connectivity between clusters while maximizing the Dunn index and the mean Silhouette width (SW) (Silhouette width values for each streamflow site are shown in Figure S3 of the supporting information).

Clustering by standardized annual streamflow produces clustered regions that loosely approximate topographic divides, topographical features, latitude and longitudinal lines, differences in regional climatology, precipitation seasonality, and runoff characteristics that contribute to annual runoff. For example, cluster 5 approximates California and the Lower Colorado Basins as represented by the USGS' hydrological unit code (HUC) for major geographic regions (HUC2). In contrast, clusters 4 and 6 appear to be grouped around topographical features with cluster 6 located along the Continental Divide and cluster 4 located around the western Great Lakes. Some cluster regions match the drought regions found using varimax principal component analysis in *Karl and Koscielny* [1982] and *Cook et al.* [1999]. For example, clusters 7, 8, and 9 match the Northeast, Northwest, and Southeast regions shown in *Karl and Koscielny* [1982], respectively. However, a more detailed examination of the hydrometeorology and climatology that underlie the clustering divisions is beyond the scope of this study but warrants further investigation.

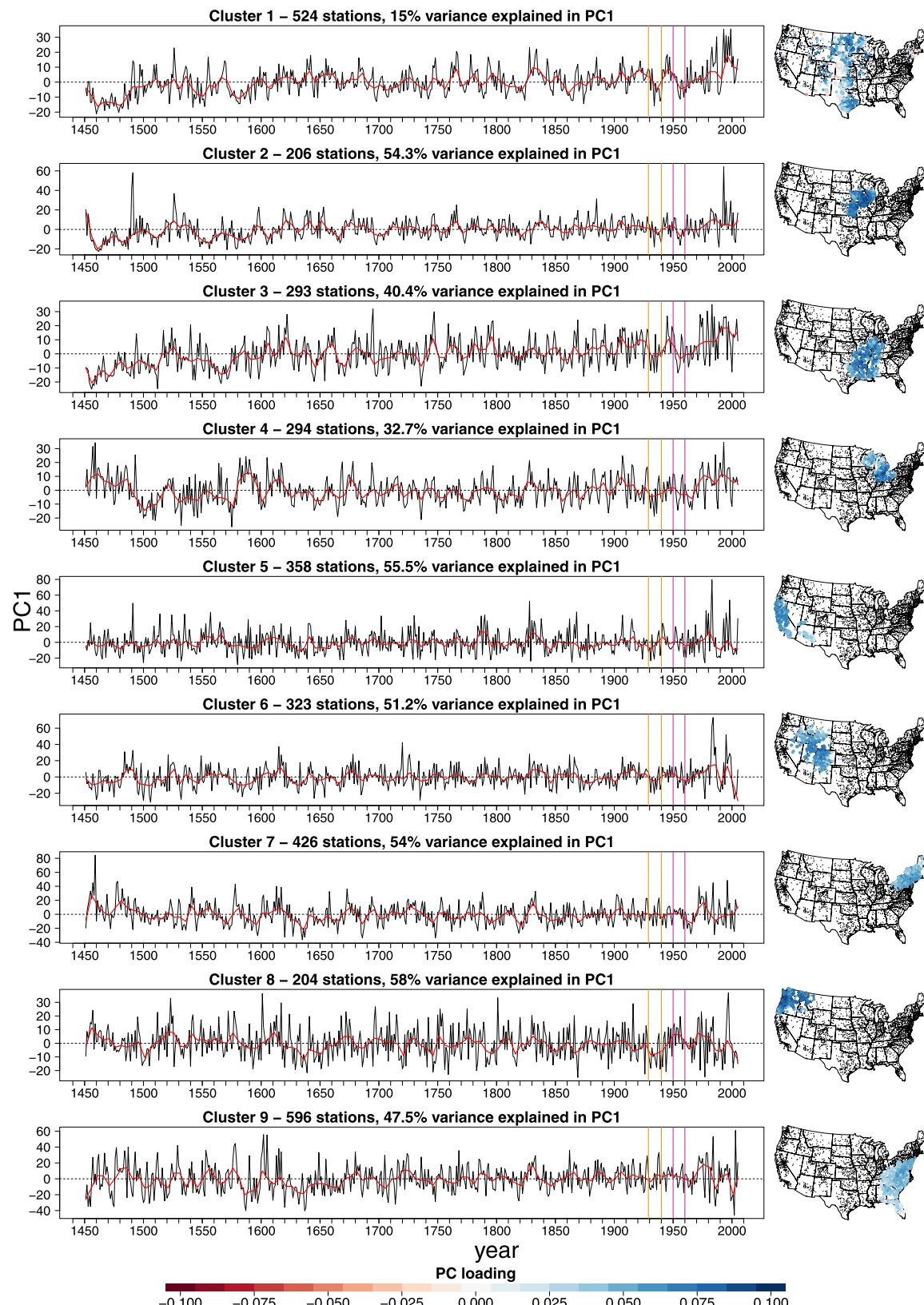
A principal component analysis was performed on streamflow within each cluster to extract the main mode of variability (Figure 7). The first PC of most clusters explained between 40 and 58% of total streamflow variability for streamflow clusters with a close spatial distribution. However, less variance was captured in the first PC of streamflow clusters with a broader spatial coverage (e.g., cluster 1 spread across the 100°W longitude line and cluster 4 with gauges on the east and west of the Great Lakes). The first PC of cluster 1 explains only 15% of variability indicating that streamflow characteristics within this cluster are not homogeneous. The first PCs of clusters 1–3 (North Dakota to Texas, Cornbelt, and southern Mississippi River Basin regions), 5 (California), 6 (Intermountain), and 8 (Pacific Northwest region) show signatures of the 1930s



**Figure 6.** Measures of cluster connectivity, Dunn index, and mean Silhouette width for 3–15 clusters and the grouping of streamflow sites into nine clusters. Dark blue lines delineate major river basin boundaries as classified by USGS (HUC2 regions).

Dust Bowl drought (orange vertical lines in Figure 7 mark the period 1929–1940). The 1950s drought is also reproduced in the reconstructed streamflows with PC1 of cluster 1 (approximating the region from North Dakota south to Texas), cluster 2 (Cornbelt region), and cluster 3 (including sections of Texas extending north-north-east into Illinois, loosely termed here as the southern Mississippi region) showing a distinct reduction in streamflow around the 1950s drought (pink lines in Figure 7 mark the years 1950 and 1960). The epicenter of the 1950s drought was located in Texas and the mid-west primarily impacting agricultural production [Nace and Pluhowski, 1965] and prompted the release of emergency relief funds to a total of over \$2 billion (present day dollars) [Wilhite, 1983]. Cluster 7 (Northeast region) also experienced a notable drought in the early 1960s [Namias, 1966], followed by a pluvial beginning in the early 1970s [Pederson et al., 2012] that is reproduced in the first PC. The first PC of each cluster separated into reference or nonreference gauges (see Figure S4 in the supporting information) resulted in comparable time series, with a notable exception of a distinct trend in the nonreference gauges in cluster 1 after 1950.

The CONUS streamflow reconstructions show periods of increased streamflow in the early 1900s for clusters 5 and 6 (Californian and Intermountain regions). The early 1900s are often cited as a period that was particularly wet. Water allocations for the Colorado River Basin detailed in the 1922 Colorado River compact were based on this wet period resulting in the over allocation of flows in the following decades [Woodhouse and Lukas, 2006b; Grafton et al., 2010]. Other notable periods of increased streamflow include the 1970s in the



**Figure 7.** The first principal component (PC) of each streamflow cluster (black time series) with a 10 year locally weighted smoothing scatter plot fitted (red time series). The loading patterns for PC1 for each cluster are shown in the right. Loadings are all of the same sign, representing a positive weighted average across sites in each cluster (with the exception of some stations in cluster 1). The number of stations in each cluster and the variance explained by the first PC of each cluster are shown in the title above each graph.

**Table 1.** Longest Drought and Wet Spell in Each Regional Cluster

Cluster	Longest Drought (No. of Years)	Longest Wet Spell (No. of Years)
1. North Dakota to Texas region	1451–1490 (40)	1913–1931 (19)
2. Cornbelt region	1544–1592 (49)	1975–2005 (31)
3. Southern Mississippi region	1451–1514 (64)	1969–2005 (37)
4. Western Great Lakes	1492–1515 (24)	1452–1488 (37)
5. Californian region	1574–1598 (25)	1549–1579 (21)
6. Intermountain region	1452–1480 (29)	1969–1988 (20)
7. Northeast	1625–1640 (16)	1473–1496 (24)
8. Pacific North-West region	1621–1654 (34)	1453–1472 (20)
9. Southeast	1620–1668 (49)	1969–1983 (15)

southeast (cluster 9), attributed to a combination of warm phase AMO and La Niña events [Tootle *et al.*, 2005], and the early 1950s in Washington (see cluster 8 in Figure 7), which was in contrast with much of the remainder of the CONUS that was experiencing severe drought.

The identification of modern dry and wet spells of the twentieth century provides some assurance (in addition to cross validation statistics shown in section 3) of the fidelity of the CONUS streamflow reconstructions and the ability to use these reconstructions to investigate streamflow variability prior to the twentieth century. The occurrences of low streamflow prior to the twentieth century are seen in PC1 of each cluster. Clusters 1–6 (all of CONUS with the exception of the Pacific Northwest and eastern region) all show decreases in streamflow around the mid to late 1500s corresponding with periods of persistent drought discussed in other climate reconstructions [Stahle *et al.*, 2007]. In contrast, cluster 8 (Pacific Northwest) shows periods of low or below average streamflow and below average decadal streamflow from approximately 1620 to 1670. While the magnitude of these low annual streamflows are less severe than the record 1977 drought in the region [Glantz, 1982] (driest October–August and driest winter period on record), the 50 years of low streamflow were persistent between 1620 and 1670 and were not interrupted by more than 3 consecutive years of above average streamflow.

A threshold analysis was used to quantitatively assess the persistence of droughts and wet spells. A threshold value of mean  $\pm 0.5$  standard deviation of the decadally smoothed streamflow was used to define states of droughts and wet spells and the number of consecutive years in either state was used to quantify the persistence of the event. The maximum duration of droughts and wet spells and their occurrence is shown in Table 1 for each cluster. Using this threshold definition of drought, the 1950s drought was only notable in cluster 1 and persisted for 6 years between 1956 and 1961. The early twentieth century is considered to be a relatively wet period across the CONUS. The threshold analysis identified persistent wet spells in the first three decades of the twentieth century in all clusters spanning between 4 and 27 years. Shorter wet spells were closely grouped together and droughts were largely absent during this period.

Many of the longest wet spells occurred during the twentieth century. Conversely, the longest droughts during the twentieth century in each cluster ranged from 1 to 16 years. In contrast, the longest droughts in each cluster, as informed by the streamflow reconstruction, range between 16 and 64 years. The prevalence of longer wet spells and shorter dry spells during the most recent century likely biases our understanding of streamflow variability when based on instrumental records alone.

In order to assess the spatial coherency of streamflow variability across the U.S., correlations between the first PC of each cluster were calculated. Correlations between the first PC of each cluster showed that most regions are significantly correlated at the 99% significance level (with the exception of correlations in red italics in Table 2). Interestingly, negative correlations exist between cluster 8 (Pacific Northwest) and clusters 1–3 that cover the Midwest region. This suggests that the Pacific Northwest likely experiences anomalously high streamflows during near-CONUS-wide droughts, similar to the 1950s drought, and vice versa. Cluster 7 (Northeast) is similarly negatively correlated with clusters 1 and 2. The dichotomy in anomalous streamflow in these regions presents a possibility of exploiting these differences by relocating, where possible, similarly vulnerable economic activities in these regions to buffer national drought impacts.

#### 4.2. Temporal Modes of Streamflow Variability Using Wavelet Transforms

In order to assess the strength of frequency signals in the streamflow, a wavelet analysis was conducted using a wavelet transform applied to the first PC of each cluster executed in R using code modified from

**Table 2.** Correlations Between First PC of the Nine Clusters<sup>a</sup>

Cluster	1	2	3	4	5	6	7	8	9
1	1	0.67	0.49	<i>0.07</i>	0.26	0.46	-0.19	-0.13	<i>0.01</i>
2		1	0.43	0.11	0.29	0.39	-0.21	-0.12	<i>-0.09</i>
3			1	0.35	0.15	0.33	-0.03	-0.18	0.17
4				1	0.26	0.28	0.39	0.18	0.19
5					1	0.63	0.17	0.26	0.25
6						1	0.13	0.21	0.32
7							1	0.45	0.61
8								1	0.29
9									1

<sup>a</sup>Correlations in italics are the only results that are not significant at the 99% significance level.

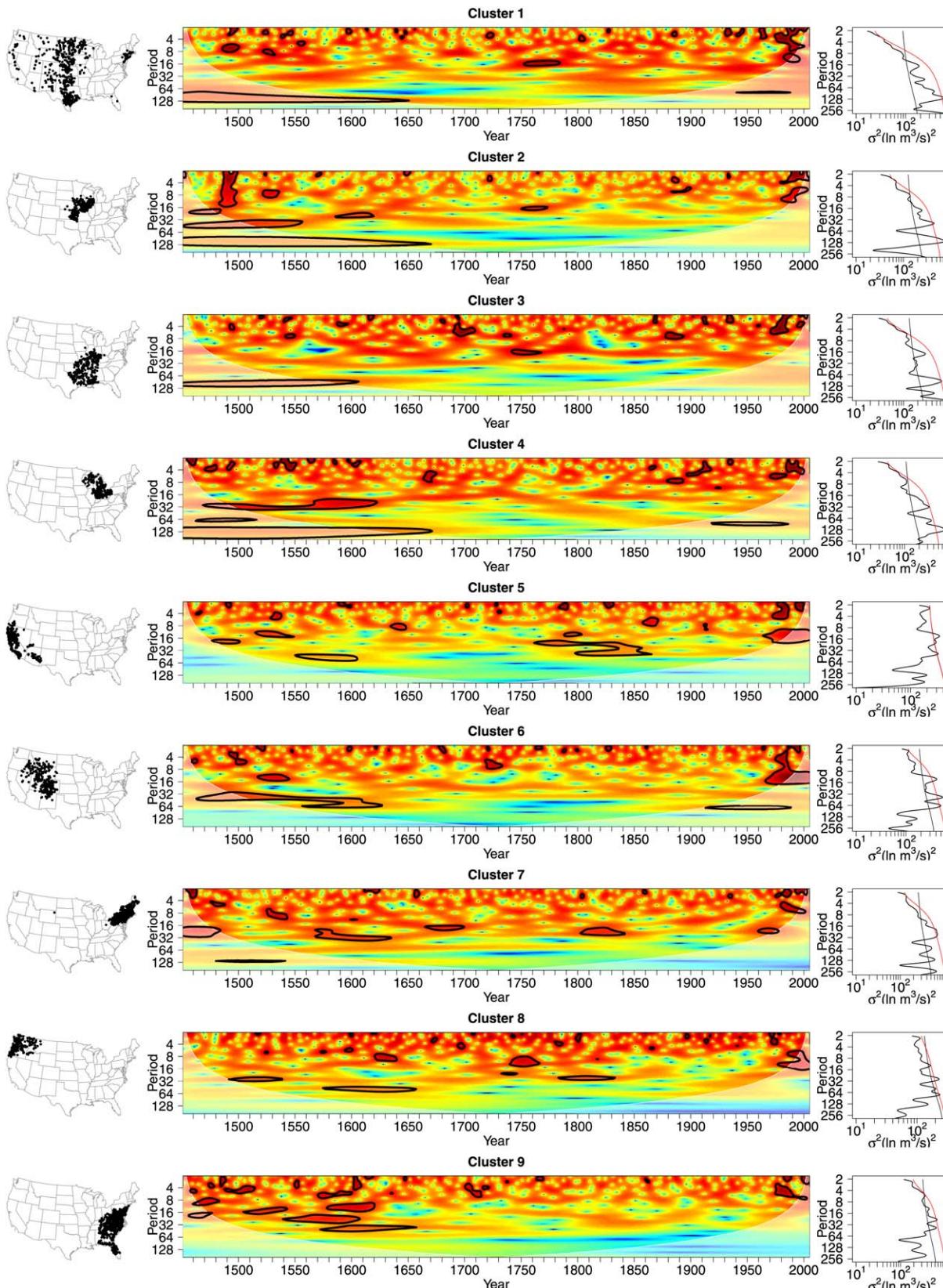
*Torrence and Compo* [1995] and results are shown in Figure 8. In the plot of the wavelet spectra (center figures in Figure 8), a solid black line shows the wavelet power significant at a 95% confidence level against white noise, while the ghosted regions show the cone of influence where edge effects may become important. A global wavelet spectrum was calculated by taking the time-averaged wavelet power spectrum. The global wavelet plots also show the time-average power at which white and red noise spectrums are significant at the 95% level. The MTM results are shown in supporting information Figure S5 and significant periodicities are tabulated in supporting information Table S1. Significant periodicities around 4–8 years, which are in the ENSO time scale, can be seen in the more recent decades in all clusters suggesting the occurrence of a common forcing mechanism. However, similar periodicities are only seen sporadically over the past 555 years around the late 1400s in cluster 2 (Cornbelt region), around 1700 in cluster 3 (southern Mississippi region), the late 1600s and around 1900 in cluster 4 (western Great Lakes region), around 1500 and 1720 for cluster 6 (Intermountain region), and around the 1500s and 1600s in cluster 9 (Southeast).

Persistent decadal-scale signals are largely absent in the first PC of each cluster within the last century with the exception of clusters 5, 6, and 8 (California, Intermountain region, and Pacific Northwest regions), which all show significant periodicities around 8–16 year cycles in the late 1900s. The results for the Intermountain region correspond with previous findings of multidecadal variability around the 1600s and in the twentieth century [Woodhouse *et al.*, 2006a]. Multidecadal modes of variability emerge in the late 1400s through to the early 1600s in most clusters with little decadal-scale activity in the mid-1600s though to the mid-1900s. Multidecadal periodicities were also found to be significant using MTM spectral analysis in all clusters except cluster 1 (see Figure S5 and Table S1 in supporting information).

## 5. Discussion

The reconstruction of streamflow in the CONUS using paleoclimate records has become a popular tool for examining long-term streamflow variability since the discovery of climate sensitive trees throughout much of the region. Using characteristics of long-term streamflow variability informed by paleoclimate streamflow reconstructions to test water management practices and assess future system vulnerabilities has also gained wider acceptance in recent decades [Phillips *et al.*, 2009; Woodhouse *et al.*, 2016]. Most studies have focused on either a single river catchment (e.g., seminal work on the upper Colorado River Basin by Stockton and Jacoby [1976]) or smaller tributaries and headwater regions [e.g., Cleaveland, 2000; Watson *et al.*, 2009]. To the best of our knowledge, this is the first study that attempts to reconstruct streamflow across the CONUS region.

The focus on a continental scale may not reproduce details that could be obtained with a targeted streamflow or catchment-scale reconstruction. However, some initial comparisons (not presented here) for clusters 5 and 6 (Californian and Intermountain regions, respectively) showed that our regional results were significantly correlated with previous catchment and gauge-specific streamflow reconstruction results (obtained from Woodhouse *et al.* [2002]). While catchment-specific reconstructions are able to inform catchment-specific water management practices, such reconstructions are yet to be developed in many catchments in the eastern and southern CONUS. Our regional reconstructions therefore provide water resource managers with preliminary streamflow reconstructions that may be improved with future reconstructions that target specific catchments or riverine systems. The results presented here can initially be used to estimate regional



**Figure 8.** Streamflow stations in the (left) cluster, (middle) wavelet spectra, and (right) global wavelet spectrum of the first PC of each cluster. Black lines on the wavelet spectrum show regions that are significant at the 95% confidence level against a white noise null hypothesis. Ghosted regions show where edge effects may become important. Black and red lines on the global wavelet plots show the 95% significance levels for the white and red noise spectrum, respectively. Periods at which the global spectrum has higher power than these levels may represent frequencies of interest for each cluster.

streamflow characteristics, such as temporal variability in streamflow and probable persistence of dry and wet spells, and may be used as an initial assessment of long-term water resource vulnerabilities. We emphasize that our results should be interpreted at a regional and continental scale. Subsequently, only the first PC of reconstructed streamflow for each of the nine clusters are available through the NOAA paleoclimate archives to ensure that these results are not used in lieu of more targeted streamflow reconstructions that are already available [e.g., *Woodhouse et al.*, 2002].

In developing our approach to reconstructing streamflow from the tree-ring-based LBDA, we are utilizing the fact that both variables amalgamate signals in the hydrologic cycle [*Piechota and Dracup*, 1996] and are derivatives of a set of (unspecified) climate variables to which both streamflow and PDSI are sensitive (e.g., precipitation amount, temperature, and evaporation). Although variability is lost in each step of reconstructing PDSI from tree-rings and in reconstructing streamflow from the LBDA, it is possible that the reconstructed LBDA may have removed noise irrelevant to hydrological variability rendering this paleoclimate reconstruction of PDSI particularly suitable for informing streamflow.

Furthermore, although streamflow gauges are a point measurement, the variable is in fact a filter of numerous other processes and influences including, but not limited to, climate variables such as rainfall, snowmelt, run off, and evaporation, and nonclimate variables such as catchment area, topography, land use, and groundwater interactions. Apart from streamflow, neither climate variables nor nonclimate variables are explicitly considered in our model, with the exception of the consideration of catchment land and water use through our separation of results for reference and nonreference gauges. However, streamflow records should represent the influence of the suite of contributing variables including catchment characteristics that would be indirectly considered when LBDA inputs are weighted using rCCA.

In order to provide a parsimonious analysis of CONUS-wide streamflow variability over the past 555 years, some degree of aggregation was required. This was achieved using hierarchical clustering and principal component analysis to investigate paleoclimate reconstructions of streamflow across the CONUS from 1451 to 2005. Overall, the first PC of each cluster captured distinct streamflow events previously identified in other studies. The CONUS-wide streamflow reconstruction often extended the temporal span of earlier studies whose tree-ring chronologies contribute to the LBDA. Key features that were replicated include above average streamflow in the twentieth century, with the exception of the dustbowl and 1950s droughts, and the extended drought from mid to late sixteenth century [*Stahle et al.*, 2007; *Margolis et al.*, 2011] throughout much of the CONUS region. As previously observed, the mid to late sixteenth century drought period did not correspond with distinct years of low streamflow, rather a persistence of below average streamflow over multiple decades [*Watson et al.*, 2009]. Cluster 5 shows the occurrence of above average streamflow in the twentieth century. These events coincide with data records and documented events such as the widespread flooding of the Great Salt Lake region in Utah in the 1980s [*DeRose et al.*, 2015] and the wet period leading up to the development of the 1922 Colorado Compact and subsequent over allocations in later years [*Woodhouse et al.*, 2006a]. Interestingly, although streamflow stations in Arizona are not heavily weighted in cluster 5, the high streamflow period in the mid-1780s to mid-1790s observed by *Smith and Stockton* [1981] are featured in PC 1 of cluster 5, suggesting a more widespread occurrence of the wet spell. A more widespread wet spell during this period is likely given that an updated version of the Salt and Verde Rivers in Arizona using a more targeted set of tree-rings resulted in a relatively damped wet spell during this period [*Hirschboeck and Meko*, 2005].

However, some features of streamflow variability that were captured in previous catchment-specific studies are not detectable in the analysis presented here as only the first PC of each cluster was examined in detail. These include, as examples:

1. reconstructions of drought in the 1770s in the Rio Grande [*Woodhouse et al.*, 2012] that are not detected in the first PC of cluster 1;
2. an extended drought period early 1700s in the Weber River, Utah [*Bekker et al.*, 2014], is not featured in cluster 6; and
3. a pluvial at the start of the 1600s in the Colorado River Basin [*Gray et al.*, 2003] that does not feature prominently in cluster 6.

For many of these examples, the drought or wet spell covers a relatively small region compared with the cluster extents. For example, the Rio Grande region belongs to cluster 1 that stretches from North Dakota to

Texas. Rio Grande streamflow variability therefore may not be a large feature of the first PC of cluster 1. Fine scale features such as the seventeenth century pluvial in Ashley Creek in north-east Utah do not feature strongly in cluster 6 and this is likely due to the smoothing of streamflow across a large region by only examining the first PC in addition to differences in streamflow between subbasins [Gray *et al.*, 2003; Devineni *et al.*, 2013]. Future analysis could involve further verifying the streamflow reconstructions against catchment-specific reconstructions to test the spatial robustness and limitations of using LBDA to reconstruct streamflow across the CONUS.

Reference streamflow gauges exhibiting trends generally resulted in improved model fit suggesting that similar trends were captured in the LBDA and that the trends are likely climate related. In contrast, nonreference streamflow gauges with trends resulted in smaller adjusted  $R^2$  values and may reflect human impacts on streamflow. However, the location of both reference and nonreference gauges exhibiting both positive and negative trends appear to cluster regionally (see Figure 1ii) making the separation of climate induced and human induced trends difficult to identify. For some nonreference stations, detrending the streamflow data may improve the model fit. In addition, the use of naturalized flows or nonreference streamflow data prior to catchment and flow alterations is a potential avenue for further validations of the models and testing of climate related impacts. However, amassing such a data set requires detailed information on catchment development and streamflow extraction timelines, which are not readily available across the CONUS to the level of detail required, and is therefore beyond the current scope of this work.

## 6. Conclusions

The development of CONUS-wide streamflow reconstructions that span 555 years has provided an opportunity to characterize annual streamflow variability on a spatial and temporal scale that has not previously been attempted. While some details of streamflow variability at small spatial scales are omitted using regional-scale analysis methods, key periods of widespread drought and regional wet spells since 1451 could still be identified. The goal of these reconstructions was to elucidate continental-scale variability and regional patterns of variability and we have therefore made the data available for the first PC of streamflow variability in each cluster (available through the NOAA paleoclimate data base). The U.S. lacks a National Water Planning and investment analysis effort. However, as significant investments in water infrastructure are considered, national-scale analyses and implications of drought will become important. Our effort is intended to inform such analyses. For more detailed streamflow reconstructions at a gauge level, we refer readers to previous studies that were focused on specific gauges or catchments (a summary of tree-ring-informed streamflow reconstructions in North America is available on the TreeFlow website) [Woodhouse *et al.*, 2002].

The CONUS-wide streamflow reconstructions provide water resource managers currently lacking catchment-specific paleoclimate reconstructions with the opportunity to conduct preliminary analyses of long-term water resource vulnerabilities using our regional reconstructions. It is intended that the provision of regional results would encourage water resource managers to recognize the value of paleoclimate streamflow reconstructions and seek more detailed, catchment and river-specific paleoclimate reconstructions in the future.

The multicentennial reconstructions provide an opportunity to quantify recent streamflow variability in the context of long-term climate variations. In addition, the streamflow reconstructions show periods of persistent streamflow extremes, which in many cases have not been observed in the instrumental record. Importantly, the development of this CONUS-wide reconstruction of streamflow will be useful in informing the development of a national water plan, which is yet to be established in the U.S., and helping to build a roadmap for the future sustainable use of water.

## References

- Adams, K. D., R. M. Negrini, E. R. Cook, and S. Rajagopal (2015), Annually resolved late Holocene paleohydrology of the southern Sierra Nevada and Tulare Lake, California, *Water Resour. Res.*, 51, 9708–9724, doi:10.1002/2015WR017850.
- Barlow, M., S. Nigam, and E. H. Berbery (2001), ENSO, Pacific decadal variability, and U.S. summertime precipitation, drought, and stream flow, *J. Clim.*, 14(9), 2105–2128, doi:10.1175/1520-0442(2001)014 <2105:EPDVAU>2.0.CO;2.
- Bekker, M. F., R. Justin DeRose, B. M. Buckley, R. K. Kjelgren, and N. S. Gill (2014), A 576-year Weber River streamflow reconstruction from tree-rings for water resource risk assessment in the Wasatch Front, Utah, *J. Am. Water Resour. Assoc.*, 50(5), 1338–1348, doi:10.1111/jawr.12191.

## Acknowledgments

Our thanks go to Benjamin I. Cook for providing us with a copy of the updated LBDA data set; Indrani Pal and Tara Troy for their helpful suggestions and informative discussions; and Pierre Gentine and Dali Plavsic for assistance in running the cross validation computations. We also thank the Associate Editor and three anonymous reviewers for their constructive comments that enabled us to clarify and improve the paper. Reconstruction results are available on the NOAA paleoclimate data base (URL: <https://www.ncdc.noaa.gov/paleo/study/20435>). Streamflow data are from U.S. Geological Survey [2011] available from <https://waterdata.usgs.gov/nwis/sw>. This work is funded by an NSF award 1360446. Lamont-Doherty contribution number 8101.

- Brock, G., V. Pihur, S. Datta, and S. Datta (2008), clValid: An R package for cluster validation, *J. Stat. Software*, 25(4), 22, doi:10.18637/jss.v025.i04.
- Busby, M. W. (1964), *Yearly Variations in Runoff for the Conterminous United States, 1931–1960*, p. 49, United States Government Printing Office, Washington, D. C.
- Carrier, C., A. Kalra, and S. Ahmad (2013), Using paleo reconstructions to improve streamflow forecast lead time in the Western United States, *J. Am. Water Resour. Assoc.*, 49(6), 1351–1366, doi:10.1111/jawr.12088.
- Cleaveland, M. K. (2000), A 963-year reconstruction of summer (JJA) stream flow in the White River, Arkansas, USA, from tree-rings, *Holocene*, 10(1), 33–41, doi:10.1191/095968300666157027.
- Cook, E. R., D. M. Meko, D. W. Stahle, and M. K. Cleaveland (1999), Drought reconstructions for the Continental United States, *J. Clim.*, 12(4), 1145–1162, doi:10.1175/1520-0442(1999)012<1145:DRFTCU>2.0.CO;2.
- Cook, E. R., C. A. Woodhouse, C. M. Eakin, D. M. Meko, and D. W. Stahle (2004), Long-term aridity changes in the Western United States, *Science*, 306(5698), 1015–1018, doi:10.1126/science.1102586.
- Cook, E. R., R. Seager, 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. Quat. Sci.*, 25(1), 48–61, doi:10.1002/jqs.1303.
- Coulthard, B., D. J. Smith, and D. M. Meko (2016), Is worst-case scenario streamflow drought underestimated in British Columbia? A multi-century perspective for the south coast, derived from tree-rings, *J. Hydrol.*, 534, 205–218, doi:10.1016/j.jhydrol.2015.12.030.
- 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(6), 1117–1130, doi:10.1175/JHM-386.1.
- De Bie, T., and B. De Moor (2003), On the regularization of canonical correlation analysis, in *International Symposium on ICA and BSS*, pp. 785–790, ICA2003, Nara, Japan.
- DeRose, R. J., M. F. Bekker, S. Y. Wang, B. M. Buckley, R. K. Kjelgren, T. Bardsley, T. M. Rittenour, and E. B. Allen (2015), A millennium-length reconstruction of Bear River stream flow, Utah, *J. Hydrol.*, 529(Part 2), 524–534, doi:10.1016/j.jhydrol.2015.01.014.
- Dettinger, M. D., and H. F. Diaz (2000), Global characteristics of stream flow seasonality and variability, *J. Hydrometeorol.*, 1(4), 289–310, doi:10.1175/1525-7541(2000)001<0289:GCOSFS>2.0.CO;2.
- Devineni, N., U. Lall, N. Pederson, and E. Cook (2013), A tree-ring-based reconstruction of Delaware River Basin streamflow using hierarchical Bayesian regression, *J. Clim.*, 26(12), 4357–4374, doi:10.1175/JCLI-D-11-00675.1.
- Diehl, T. H., and M. A. Harris (2014), Withdrawal and consumption of water by thermoelectric power plants in the United States, 2010, *Sci. Invest. Rep.* 2014-5184, p. 38, U.S. Geological Survey, Reston, Va.
- Dijkstra, T. (2014), Ridge regression and its degrees of freedom, *Qual. Quant.*, 48(6), 3185–3193, doi:10.1007/s11135-013-9949-7.
- DuMars, C. T., and J. D. Minier (2004), The evolution of groundwater rights and groundwater management in New Mexico and the western United States, *Hydrogeol. J.*, 12(1), 40–51, doi:10.1007/s10040-003-0303-3.
- Enfield, D. B., A. M. Mestas-Núñez, and P. J. Trimble (2001), The Atlantic Multidecadal Oscillation and its relation to rainfall and river flows in the continental U.S., *Geophys. Res. Lett.*, 28(10), 2077–2080, doi:10.1029/2000GL012745.
- Falcone, J. A., D. M. Carlisle, D. M. Wolock, and M. R. Meador (2010), GAGES: A stream gage database for evaluating natural and altered flow conditions in the conterminous United States, *Ecology*, 91(2), 621, doi:10.1890/09-0889.1.
- Fye, F. K., D. W. Stahle, and E. R. Cook (2003), Paleoclimatic analogs to twentieth-century moisture regimes across the United States, *Bull. Am. Meteorol. Soc.*, 84(7), 901–909, doi:10.1175/BAMS-84-7-901.
- Ghil, M., et al. (2002), Advanced spectral methods for climatic time series, *Rev. Geophys.*, 40(1), 3–1–3–41, doi:10.1029/2000RG000092.
- Glantz, M. H. (1982), Consequences and responsibilities in drought forecasting: The case of Yakima, 1977, *Water Resour. Res.*, 18(1), 3–13, doi:10.1029/WR018001p00003.
- Gleick, P. H. (2016), Water strategies for the next administration, *Science*, 354(6312), 555–556, doi:10.1126/science.aj2221.
- González, I., S. Déjean, P. G. P. Martin, and A. Baccini (2008), CCA: An R package to extend canonical correlation analysis, 23(12), 1–14, doi:10.18637/jss.v023.i12.
- Grafton, R. Q., C. Landry, G. D. Libecap, and R. J. O'Brien (2010), *Water Markets: Australia's Murray-Darling Basin and the US Southwest*, p. 35, National Bureau of Economic Research, Cambridge, Mass.
- Graham, N. E., and M. K. Hughes (2007), Reconstructing the Mediaeval low stands of Mono Lake, Sierra Nevada, California, USA, *Holocene*, 17(8), 1197–1210, doi:10.1177/0959683607085126.
- Gray, S. T., J. L. Betancourt, C. L. Fastie, and S. T. Jackson (2003), Patterns and sources of multidecadal oscillations in drought-sensitive tree-ring records from the central and southern Rocky Mountains, *Geophys. Res. Lett.*, 30(6), 1316, doi:10.1029/2002GL016154.
- Gregory, J. N. (1991), *American Exodus: The Dust Bowl Migration and Okie Culture in California*, Oxford Univ. Press, New York.
- Hammett, K. M. (1990), Land use, water use, streamflow characteristics, and water-quality characteristics of the Charlotte Harbor inflow area, Florida, *U.S. Geol. Surv. Water Supply Pap.*, 2359A, 64 pp.
- Handl, J., J. Knowles, and D. B. Kell (2005), Computational cluster validation in post-genomic data analysis, *Bioinformatics*, 21(15), 3201–3212, doi:10.1093/bioinformatics/bti517.
- Heddinghaus, T. R., and P. Sabol (1991), A review of the Palmer Drought Severity Index and where do we go from here?, in *7th Conference on Applied Climatology*, pp. 242–246, Am. Meteorol. Soc., Salt Lake City, Utah.
- Heim, R., R. Vose, J. Lawrimore, and E. Cook (2007), Putting current North America drought conditions into a multi-century perspective. Part 2: Using the blended product in operational drought monitoring, AGU Fall Meeting Abstracts PP31F-02.
- Hirschboeck, K. K., and D. M. Meko (2005), *A Tree Ring Based Assessment of Synchronous Extreme Streamflow Episodes in the Upper Colorado and Salt-Verde-Tonto River Basins*, 31 pp., Univ. of Ariz. and Salt River Proj., Tucson.
- Ho, M., U. Lall, and E. R. Cook (2016), Can a paleo-drought record be used to reconstruct streamflow? A case-study for the Missouri River Basin, *Water Resour. Res.*, 52, 5195–5212, doi:10.1002/2015WR018444.
- Hotelling, H. (1936), Relations between two sets of variates, *Biometrika*, 28(3–4), 321–377, doi:10.1093/biomet/28.3-4.321.
- Iglesias, E., A. Garrido, and A. Gómez-Ramos (2003), Evaluation of drought management in irrigated areas, *Agric. Econ.*, 29(2), 211–229, doi:10.1111/j.1574-0862.2003.tb00158.x.
- Karl, T. R., and R. W. Knight (1998), Secular trends of precipitation amount, frequency, and intensity in the United States, *Bull. Am. Meteorol. Soc.*, 79(2), 231–241, doi:10.1175/1520-0477(1998)079<0231:STOPAF>2.0.CO;2.
- Karl, T. R., and A. J. Koscielny (1982), Drought in the United States: 1895–1981, *J. Climatol.*, 2(4), 313–329, doi:10.1002/joc.3370020402.
- Leurgans, S. E., R. A. Moyeed, and B. W. Silverman (1993), Canonical correlation analysis when the data are curves, *J. R. Stat. Soc. Ser. B*, 55(3), 725–740.

- Lins, H. F. (1985), Streamflow variability in the United States: 1931–78, *J. Clim. Appl. Meteorol.*, 24(5), 463–471, doi:10.1175/1520-0450(1985)024<0463:SVITUS>2.0.CO;2.
- Lins, H. F., and J. R. Slack (1999), Streamflow trends in the United States, *Geophys. Res. Lett.*, 26(2), 227–230, doi:10.1029/1998GL900291.
- Lins, H. F., and J. R. Slack (2005), Seasonal and regional characteristics of U.S. streamflow trends in the United States from 1940 to 1999, *Phys. Geogr.*, 26(6), 489–501, doi:10.2747/0272-3646.26.6.489.
- Maechler, M., P. Rousseeuw, A. Struyf, M. Hubert, and K. Hornik (2015), *Cluster: Cluster Analysis Basics and Extensions. R package version 2.0.3.*
- Margolis, E. Q., D. M. Meko, and R. Touchan (2011), A tree-ring reconstruction of streamflow in the Santa Fe River, New Mexico, *J. Hydrol.*, 397(1–2), 118–127, doi:10.1016/j.jhydrol.2010.11.042.
- 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(13), 1399–1410, doi:10.1002/(SICI)1097-0088(19991115)19:13<1399::AID-JOC457>3.0.CO;2-A.
- McCabe, G. J., and D. M. Wolock (2002), A step increase in streamflow in the conterminous United States, *Geophys. Res. Lett.*, 29(24), 38-1–38-4, doi:10.1029/2002GL015999.
- Meko, D. M., and C. A. Woodhouse (2005), Tree-ring footprint of joint hydrologic drought in Sacramento and Upper Colorado River Basins, Western USA, *J. Hydrol.*, 308(1–4), 196–213, doi:10.1016/j.jhydrol.2004.11.003.
- Meko, D. M., C. A. Woodhouse, and K. Morino (2012), Dendrochronology and links to streamflow, *J. Hydrol.*, 412–413, 200–209, doi:10.1016/j.jhydrol.2010.11.041.
- Mo, K. C., J.-K. E. Schemm, and S.-H. Yoo (2009), Influence of ENSO and the Atlantic Multidecadal Oscillation on drought over the United States, *J. Clim.*, 22(22), 5962–5982, doi:10.1175/2009JCLI2966.1.
- Murtagh, F., and P. Legendre (2014), Ward's hierarchical agglomerative clustering method: Which algorithms implement Ward's criterion?, *J. Classif.*, 31(3), 274–295, doi:10.1007/s00357-014-9161-z.
- Nace, R. L., and E. J. Pluhowski (1965), Drought of the 1950's with special reference to the midcontinent, *U.S. Geol. Surv. Water Supply Pap.*, 1804, 93 pp.
- Namias, J. (1966), Nature and possible causes of the Northeastern United States drought during 1962–65, *Mon. Weather Rev.*, 94(9), 543–554, doi:10.1175/1520-0493(1966)094<0543:NAPCOT>2.3.CO;2.
- 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, doi:10.1029/2011GL048650.
- Palmer, W. C. (1965), *Meteorological Drought*, 58 pp., US Weather Bureau, Washington, D. C.
- Pederson, N., A. R. Bell, E. R. Cook, U. Lall, N. Devineni, R. Seager, K. Eggleston, and K. P. Vrane (2012), Is an epic pluvial masking the water insecurities of the greater New York City region?, *J. Clim.*, 26(4), 1339–1354, doi:10.1175/JCLI-D-11-00723.1.
- Phillips, D. H., Y. Reinink, T. E. Skarupa, C. E. Ester, and J. A. Skindlov (2009), Water resources planning and management at the Salt River Project, Arizona, USA, *Irrig. Drain. Syst.*, 23(2), 109, doi:10.1007/s10795-009-9063-0.
- Piechota, T. C., and J. A. Dracup (1996), Drought and regional hydrologic variation in the United States: Associations with the El Niño–Southern Oscillation, *Water Resour. Res.*, 32(5), 1359–1373, doi:10.1029/96WR00353.
- Poff, N. L., J. D. Olden, D. M. Pepin, and B. P. Bledsoe (2006), Placing global stream flow variability in geographic and geomorphic contexts, *River Res. Appl.*, 22(2), 149–166, doi:10.1002/rra.902.
- Probst, J. L., and Y. Tardy (1987), Long range streamflow and world continental runoff fluctuations since the beginning of this century, *J. Hydrol.*, 94(3), 289–311, doi:10.1016/0022-1694(87)90057-6.
- Rajagopalan, B., K. Nowak, J. Prairie, M. Hoerling, B. Harding, J. Barsugli, A. Ray, and B. Udall (2009), Water supply risk on the Colorado River: Can management mitigate?, *Water Resour. Res.*, 45, W08201, doi:10.1029/2008WR007652.
- 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(9), 2381–2399, doi:10.1029/91WR00690.
- Scanlon, B. R., R. C. Reedy, D. A. Stonestrom, D. E. Pradic, and K. F. Dennehy (2005), Impact of land use and land cover change on groundwater recharge and quality in the southwestern US, *Global Change Biol.*, 11(10), 1577–1593, doi:10.1111/j.1365-2486.2005.01026.x.
- Schubert, S. D., M. J. Suarez, P. J. Pegion, R. D. Koster, and J. T. Bacmeister (2004), On the cause of the 1930s Dust Bowl, *Science*, 303(5665), 1855–1859, doi:10.1126/science.1095048.
- Smith, L. P., and C. W. Stockton (1981), Reconstructed stream flow for the salt and Verde Rivers from tree-ring data, *J. Am. Water Resour. Assoc.*, 17(6), 939–947, doi:10.1111/j.1752-1688.1981.tb01925.x.
- Stahle, D. F., F. Fye, E. Cook, and R. D. Griffin (2007), Tree-ring reconstructed megadroughts over North America since A.D. 1300, *Clim. Change*, 83(1–2), 133–149, doi:10.1007/s10584-006-9171-x.
- Stockton, C. W., and G. C. Jacoby (1976), Long-term surface-water supply and streamflow trends in the Upper Colorado River Basin, in *Lake Powell Research Project Bulletin*, 70 pp., National Science Foundation, Arlington, Va.
- Stromberg, J. C., K. J. Bagstad, J. M. Leenhouts, S. J. Lite, and E. Makings (2005), Effects of stream flow intermittency on riparian vegetation of a semiarid region river (San Pedro River, Arizona), *River Res. Appl.*, 21(8), 925–938, doi:10.1002/rra.858.
- Theil, H. (1992), A rank-invariant method of linear and polynomial regression analysis, in *Henri Theil's Contributions to Economics and Econometrics: Econometric Theory and Methodology*, edited by B. Raj and J. Koerts, pp. 345–381, Springer, Dordrecht, Netherlands.
- Tootle, G. A., T. C. Piechota, and A. Singh (2005), Coupled oceanic-atmospheric variability and U.S. streamflow, *Water Resour. Res.*, 41, W12408, doi:10.1029/2005WR004381.
- Torrence, C., and G. P. Compo (1995), *Wavelet Software*, Boulder, Colo. [Available at <http://paos.colorado.edu/research/wavelets/software.html>.]
- U.S. Geological Survey (2011), *GAGES-II: Geospatial Attributes of Gages for Evaluating Streamflow*, Reston, Va. [Available at [http://water.usgs.gov/GIS/metadata/usgsrd/XML/gagesII\\_Sept2011.xml](http://water.usgs.gov/GIS/metadata/usgsrd/XML/gagesII_Sept2011.xml).]
- U.S. Geological Survey (2012), *USGS Hydro-Climatic Data Network 2009 (HCDN-2009)*, Reston, Va.
- Vinod, H. D. (1976), Canonical ridge and econometrics of joint production, *J. Econometrics*, 4(2), 147–166, doi:10.1016/0304-4076(76)90010-5.
- Walsh, J. E., and A. Mostek (1980), A quantitative analysis of meteorological anomaly patterns over the United States, 1900–1977, *Mon. Weather Rev.*, 108(5), 615–630, doi:10.1175/1520-0493(1980)108<615:AQAMOA>2.0.CO;2.
- Ward, J. H. (1963), Hierarchical grouping to optimize an objective function, *J. Am. Stat. Assoc.*, 58(301), 236–244, doi:10.1080/01621459.1963.10500845.
- Watson, T. A., F. Anthony Barnett, S. T. Gray, and G. A. Tootle (2009), Reconstructed streamflows for the headwaters of the Wind River, Wyoming, United States, *J. Am. Water Resour. Assoc.*, 45(1), 224–236, doi:10.1111/j.1752-1688.2008.00274.x.
- 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, *Weather Clim. Extremes*, 11, 80–94, doi:10.1016/j.wace.2015.11.007.

- Wilcox, B. P., Y. U. N. Huang, and J. W. Walker (2008), Long-term trends in streamflow from semiarid rangelands: Uncovering drivers of change, *Global Change Biol.*, 14(7), 1676–1689, doi:10.1111/j.1365-2486.2008.01578.x.
- Wilhite, D. A. (1983), Government response to drought in the United States: With particular reference to the Great Plains, *J. Clim. Appl. Meteorol.*, 22(1), 40–50, doi:10.1175/1520-0450(1983)022<0040:GRTDIT>2.0.CO;2.
- 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(1), 62–78, doi:10.1002/jqs.1297.
- Woodhouse, C., J. Lukas, K. Morino, D. Meko, and K. Hirschboeck (2016), Using the past to plan for the future? The value of paleoclimate reconstructions for water resource planning, in *Water Policy and Planning in a Variable and Changing Climate*, pp. 161–182, CRC Press, Boca Raton, Fla.
- Woodhouse, C. A., and J. J. Lukas (2006a), Drought, tree-rings and water resource management in Colorado, *Can. Water Resour. J.*, 31(4), 297–310, doi:10.4296/cwrj3104297.
- Woodhouse, C. A., and J. J. Lukas (2006b), Multi-century tree-ring reconstructions of Colorado streamflow for water resource planning, *Clim. Change*, 78(2), 293–315, doi:10.1007/s10584-006-9055-0.
- Woodhouse, C. A., J. Lukas, B. Brice, K. Hirschboeck, H. Hartmann, M. Kostuk, E. Lay, D. Martinex, B. McMahan, and A. Shah (2002), *Tree Rings and Streamflow*, University of Arizona, Tucson, Ariz. [Available at <http://treeflow.info/content/tree-rings-and-streamflow>.]
- Woodhouse, C. A., S. T. Gray, and D. M. Meko (2006a), Updated streamflow reconstructions for the Upper Colorado River Basin, *Water Resour. Res.*, 42, W05415, doi:10.1029/2005WR004455.
- Woodhouse, C. A., J. Lukas, B. Brice, K. Hirschboeck, H. Hartmann, M. Kostuk, E. Lay, D. Martinex, B. McMahan, and A. Shah (2006b), *Data Access by Basin*, University of Arizona, Tucson, Ariz. [Available at <http://treeflow.info/content/tree-rings-and-streamflow>.]
- Woodhouse, C. A., D. W. Stahle, and J. Villanueva Díaz (2012), Rio Grande and Rio Conchos water supply variability over the past 500 years, *Clim. Res.*, 51(2), 147–158, doi:10.3354/cr01059.