Can a paleo-drought record be used to reconstruct streamflow? A case-study for the Missouri River Basin

Michelle Ho[1], Upmanu Lall[1, 2], Edward R. Cook[3]

1. Columbia Water Center, Columbia University, New York, NY, USA

2. Department of Earth and Environmental Engineering, Columbia University, New York, New York, USA

3. Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York 10964, USA

Corresponding author: Michelle Ho. Email: mh3538@columbia.edu, Phone: 212-854-7081

Keywords: paleoclimate, regularized canonical correlation analysis, streamflow reconstruction

Key points

- Streamflow is reconstructed from an existing paleoclimate drought index
- Methodological innovation using rCCA addresses very high dimensional dataset
- Reconstructed streamflow provides insights into extreme, persistent high and low flow

Abstract

Recent advances in paleoclimatology have revealed dramatic long-term hydro-climatic variations that provide a context for limited historical records. A notable dataset derived from a relatively dense network of paleoclimate proxy records in North America is the Living Blended Drought Atlas (LBDA): a gridded tree-ring based reconstruction of summer Palmer Drought Severity Index. This index has been used to assess North American drought frequency, persistence and spatial extent over the past two millennia. Here, we explore whether the LBDA can be used to reconstruct annual streamflow. Relative to streamflow reconstructions that use tree rings within the river basin of interest, the use of a gridded proxy poses a novel challenge. The gridded series have high spatial correlation, since they rely on tree rings over a common radius of influence. A novel algorithm for reconstructing streamflow using regularized canonical regression and inputs of local and global covariates is developed and applied over the Missouri River Basin, as a test case. Effectiveness in reconstruction is demonstrated with reconstructions showing periods where streamflow deficits may have been more severe than during recent droughts (e.g. the Civil War, Dust Bowl and 1950s droughts). The maximum persistence of droughts and floods over the past 500 years far exceed those observed in the instrumental record and periods of multi-decadal variability in the 1500s and 1600s are detected. Challenges for an extension to a national streamflow reconstruction or applications using other gridded paleoclimate datasets such as adequate spatial coverage of streamflow and applicability of annual reconstructions are discussed.

1 Introduction

The concern with anthropogenic climate change and its hydrologic impacts has focused interest on how long term climate variability may impact streamflow (e.g. Nohara et al., 2006; Seager et al., 2007). The consequence of using short records to "over-allocate" the flows of major rivers are often cited as an example of the need for long records that can better inform the possible range of long-term variations of streamflow (Tootle and Piechota, 2006; McGowan et al., 2009; Woodhouse et al., 2010). Continuous records of streamflow in the US span several decades at best. Advances in paleoclimatology in the past few decades have provided opportunities across the world to extend the range of hydroclimatic variability (e.g. Quinn, 1992; Jones and Mann, 2004; Tierney et al., 2010; Gallant and Gergis, 2011; Vance et al., 2012; Cook et al., 2013b; Devineni et al., 2013; Ho et al., 2015). While considerable uncertainty clouds the projections of hydroclimatic states towards the end of the 21st century, in the near to medium term paleoclimate information may be crucial to inform the interannual to decadal variability of regional water availability as indicated by streamflow for reservoir operation, and agricultural and other water use decisions.

Paleoclimate reconstructions have been developed using proxies that typically span the past 1000-2000 years (also known as the Common Era). The North American region has a relatively dense network of high-resolution paleoclimate proxy records, primarily comprised of tree-ring chronologies. Tree-ring-proxy records have been used to assess various components of environmental variations (Fritts, 1976) including drought severity (e.g. Routson et al., 2011; Cook et al., 2015a), pluvials (e.g. Woodhouse et al., 2005; Pederson et al., 2012), streamflow variability (e.g. Woodhouse et al., 2006; Prairie et al., 2008; Allen et al., 2013; Devineni et al., 2013), and precipitation frequency (Woodhouse and Meko, 1997) in addition to enabling comparisons of past climate with projected climate scenarios (e.g. Ault et al., 2014; Cook et al., 2015a; Smerdon et al., 2015).

Studies focused on the reconstruction of tree-ring-based paleo-hydrology (e.g. annual and season streamflow and floods) typically utilize proxies derived from tree-ring networks within or near the catchment region as predictors (e.g. Woodhouse et al., 2006; St. George, 2010; Pederson et al., 2012; Devineni et al., 2013). However, these networks are spatially irregular with record lengths varying across chronologies. An alternative to using spatially and temporally irregular tree-ring chronologies as model predictors is to use an existing derivative of these records, namely the Living Blended Drought Atlas (LBDA) (Cook et al., 2010a). The LBDA is a paleoclimate reconstruction of the summer (June-August) Palmer Drought Severity Index (PDSI) that is gridded across North America on a 0.5° × 0.5° latitude/longitude grid with reconstructions dating back as far as 2000 years. These records are temporally complete over the Conterminous United States (CONUS) from 1473 onward. The LBDA, or its predecessor the North American Drought Atlas (Cook and Krusic, 2004), have been used to assess the frequency and spatial distribution of droughts over the past millennia (e.g. Herweijer et al., 2007; Cook et al., 2013a).

2 A proposal for using PDSI to reconstruct streamflow

The intent of the modeling case study presented here is to develop a suitable framework with which streamflow within the CONUS may be reconstructed using a tree-ring-based reconstruction of the PDSI. In developing the modelling framework we consider: 1) the constraints posed by the LBDA and implications for reconstructing streamflow; 2) possible temporal resolutions (monthly, seasonal or annual) for direct streamflow reconstruction using the LBDA data; 3) how to best use local and far-field LBDA information for local streamflow reconstruction; and 4) provides insights from the 500 year reconstruction of the multi-site Missouri River Basin flows as to the decadal and longer variability of streamflow in the region. Given the existence of the spatially and temporally complete LBDA record over the last 500 years covering the CONUS, we explore whether the LBDA, a reconstruction of PDSI using

tree-ring chronologies, could be a reasonable predictor of streamflow variability. In this case the variable to be reconstructed is annual streamflow with the aim of eventually reconstructing paleoclimate records of streamflow across the CONUS, an undertaking that has not previously been attempted using the LBDA or tree-rings.

The motivation for using the LBDA stems from our understanding that the growth of moisture-limited trees, from which the LBDA is derived, are in part governed by climatic forcings that drive soil moisture availability. That is, given a vector of unspecified climate variables, $\boldsymbol{C}_t$ (where $\boldsymbol{C}$ may be comprised of, but not limited to, climate variables such as temperature, rainfall, wind, soil moisture and radiation) we can define PDSI as $PDSI_t = f_1(\boldsymbol{C}_t)$. Streamflow is also a derivative of a number of climate variables and can similarly be defined as $Q_t = f_2(\boldsymbol{C}_t)$. We seek to determine if it is possible to derive and fit a function f_3 that relates streamflow to PDSI where $Q_t = f_3(PDSI_t)$. Where suitable instrumental records of climate are available, we may derive f_1 and this has been approximated in part using methods such as the Thornthwaite potential evapotranspiration (PET) equation (Thornthwaite, 1948) and the Penman-Monteith PET equation (Monteith, 1965) to varying degrees of success (Lockwood, 1999; Sheffield et al., 2012; Dai, 2013). Consideration of the joint probability distributions $f(PDSI_t, \boldsymbol{C}_t)$ and $f(Q_t, PDSI_t)$ enables the vector of climate variables to be reconstructed by capitalizing on the climate-PDSI relationship and the tree-ring chronology-PDSI relationship given $f(\boldsymbol{C_t}|PDSI_t)f(PDSI_t|\boldsymbol{tree\text{-}ring\ chronology_t})$. Paleoclimate streamflow could similarly be derived by implementing f_2, namely $f(Q_t|\boldsymbol{C}_t)f(\boldsymbol{C}_t|PDSI_t)f(PDSI_t|\boldsymbol{tree\text{-}ring\ chronology_t})$. However, the challenge in this approach is the selection of an appropriate vector of climate variables, many of which may be sparsely observed or unobserved in the instrumental record. Therefore, we consider modelling paleoclimate streamflow through $f(Q_t|PDSI_t)f(PDSI_t|\boldsymbol{tree\text{-}ring\ chronology_t})$. Since the reconstructed $PDSI_t$ in the LBDA is really $E[PDSI_t|\boldsymbol{tree\text{-}ring\ chronology}_t]$, we start by considering $f(\boldsymbol{Q}_t|\boldsymbol{PDSI}_t)$, where $\boldsymbol{Q}_t$ is the streamflow at one or more locations in a river basin, and $\boldsymbol{PDSI}_t$ represents a vector of LBDA values at the gridded locations of the LBDA that can be a potential

predictor of the streamflows. The relationship can be developed using contemporaneous values of $\boldsymbol{PDSI}_t$ and historical $\boldsymbol{Q}_t$. Subsequently, we can apply this relationship to the paleo estimates of $\boldsymbol{PDSI}_t$ recognizing that we could use the expected values of $\boldsymbol{PDSI}_t$ reported in the LBDA, or simulations from the uncertainty distributions of $\boldsymbol{PDSI}_t$ reported in the LBDA, as conditioning variables to derive sequences of $\boldsymbol{Q}_t$.

A key motivation for using the LBDA is that it is spatially complete across CONUS over the past 500 years and a successful streamflow reconstruction would have significant value in assessing national water planning and use strategies and in investigating the different temporal and spatial structures in streamflow, which differ from the LBDA (see Figures S1 and S2 in supporting information). Relative to reconstructions that use tree rings within the river basin of interest, the use of a gridded proxy series to reconstruct watershed processes such as streamflow poses a novel challenge. The gridded LBDA series have high spatial correlation, since they rely on tree rings over a common radius of influence around each grid point, nominally 450 km in this case or roughly the correlation decay *e*-folding distance between grid points of the instrumental PDSI. A novel algorithm for reconstruction that uses local and global covariates for streamflow reconstruction using regularized canonical regression is developed and applied over the Missouri River Basin, as a test case. This provides a proof of concept at a sub-continental scale as to whether the approach is feasible. The Missouri River Basin was chosen as it contains the only major river headwaters in the western US where extensive reconstructions of paleoclimate hydrology have not been undertaken (Driscoll, 2013) and also parallels current efforts to further develop tree-ring chronologies and streamflow reconstruction models using tree rings in the region (Pederson, 2013). A description of the case study region and data are presented in the following section while Section 4 presents initial diagnostics that inform the modeling approach developed in Section 5. Section 5 details how the above proposal of using PDSI to reconstruct streamflow is implemented in the Missouri River Basin. Section 6 provides model verification results and summaries of

the key modes of variability in the 500 year mean streamflow reconstructions of Missouri River Basin streamflow. The final section reviews outstanding questions as to modeling uncertainties, and the challenges for extending the model presented here to a national scale.

3 Case Study Region and Data

3.1 The Missouri River Basin

The Missouri River is the longest river and the second largest river basin in the US, draining an area over 1.3 million km^2 that spans the southern portions of two Canadian provinces and ten states in the US (Missouri River basin boundary and key tributaries shown in Figure 1). The headwaters, which are largely snow-melt fed, are located in the Northern Rocky Mountains. These waters then flow through a largely semi-arid region to its confluence with the Mississippi River near St. Louis, Missouri (Galat et al., 2005). Land use in the Missouri River Basin is dominated by agricultural activities including cropping and grazing which cover 95% of the region (US Army Corps of Engineers, 2006), while the remaining land is used for recreation, transport, urban and industrial use including mining and energy sector activities (Galat et al., 2005). An improved perspective of streamflow variability would therefore be of benefit to the region in terms of managing and balancing the demands for water amongst the various sectors and users.

Precipitation and moisture availability in the Missouri River Basin are characterized by high precipitation in the western mountainous region, which averages over 1000 mm/year to the drier region in the rain shadow east of the Rocky Mountains where average annual precipitation is less than 400 mm/year. Precipitation increases towards the far eastern regions of the Missouri River Basin (Kunkel et al., 2013).

Winter precipitation in the northern Missouri River Basin and in the mountainous regions to the west is related to the El Niño Southern Oscillation (ENSO) signal from the preceding summer and autumn

(Redmond and Koch, 1991). El Niño teleconnections typically manifest as upper level anticyclonic high pressure cells over the northwestern US and result in the northward displacement or splitting of the jet stream and anomalously dry conditions in the Missouri River Basin (Trenberth et al., 1988; Dettinger et al., 1998; Smith et al., 1998). Conversely, La Niña events typically result in wetter conditions in this region. ENSO impacts are modulated by the decadal scale variability in the northern Pacific Ocean with warm decadal phases resulting in a deep Aleutian low and corresponding ridging over the western US thereby enhancing El Niño conditions (Gershunov and Barnett, 1998; Brown and Comrie, 2004). The enhancement of La Niña impacts by a cool phase in the northern Pacific decadal signature is particularly noticeable in the northern Missouri River Basin (Wise, 2010). Both Pacific and Atlantic Ocean influences are seen in the Northern Great Plains with the Great Plains low level jet, originating from the Gulf of Mexico, enhancing summer precipitation in this region (Higgins et al., 1997).

3.2 Streamflow data

Monthly streamflow data for the Missouri River Basin were obtained from the United States Geological Survey (USGS) Surface-Water Daily Data for the Nation (http://nwis.waterdata.usgs.gov/nwis/sw). The streamflow data are from stations included in the USGS's GAGES-II network (U.S. Geological Survey, 2011) within the Missouri River Basin boundary and are reference gauges identified by the USGS as the least-disturbed watersheds with minimal regulation. The selected gauges meet a criterion of data spanning 40 years with less than 5% missing data (Figure 1, Table S1) and results in 55 streamflow gauges, 46 of which are also in the USGS Hydro Climatic Data Network.

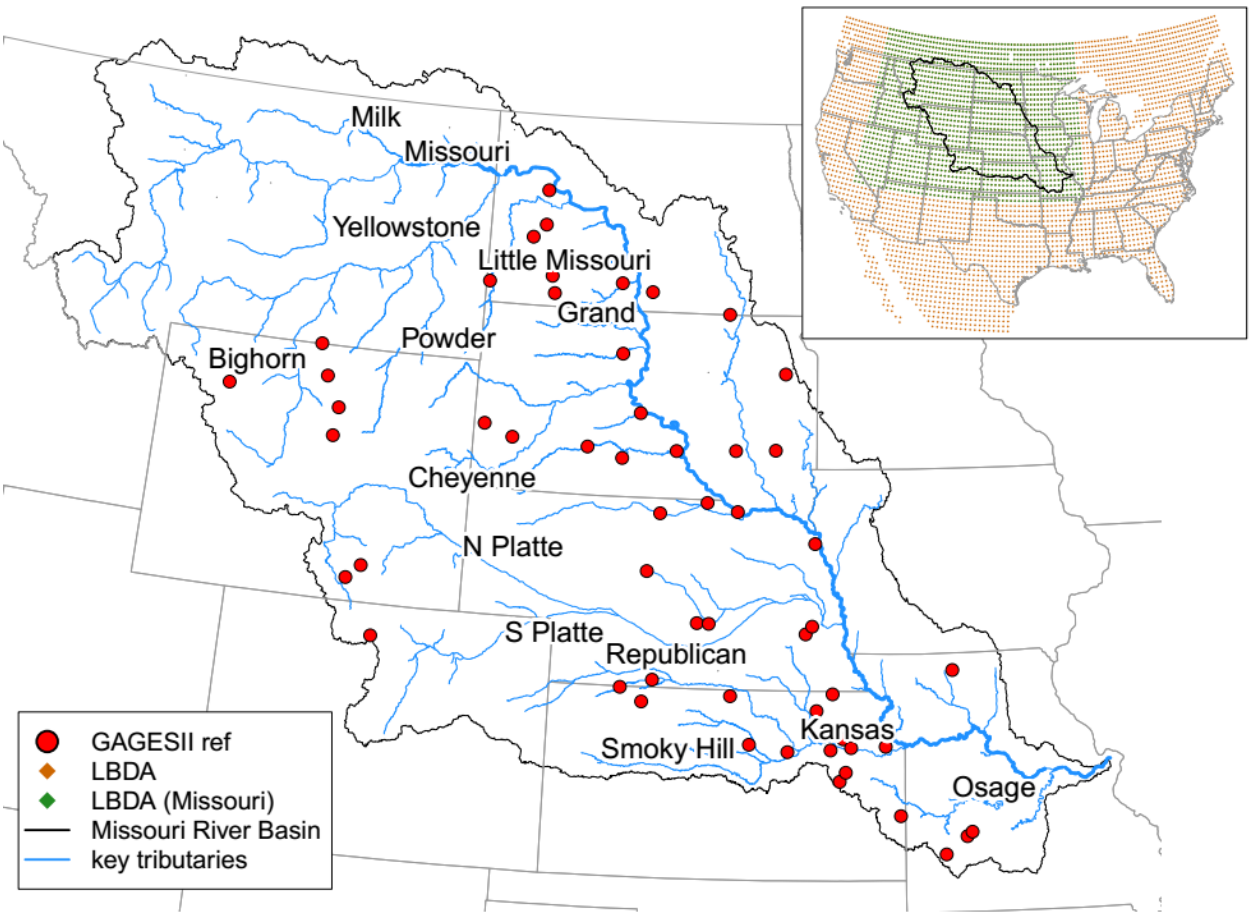


Figure 1. The Missouri River Basin, key tributaries, locations of gauged streamflow used in the analysis, and (inset) the LBDA grids used in the analysis. LBDA grids in the Missouri River Basin region shown in green.

Three of the selected stations had missing monthly values. These values were imputed using multiple imputation by chained equations (MICE) and a method of predictive mean matching (Buuren and Groothuis-Oudshoorn, 2011). Monthly streamflow imputation was conducted using streamflow from the three closest stations and a cosine function to represent a seasonal signal. The number of repetitions in MICE was determined using a rule of thumb method proposed by White et al. (2011) and the multiple imputed monthly values were averaged across the repetitions.

Monthly data was aggregated into annual streamflow data using a calendar year instead of a water year because the average driest month of streamflow at most stations occurred in either December or January. Start and end years with incomplete data were excluded. The resulting annual data spans from

1929 to 2014 with annual record lengths varying between 39 and 85 years after aggregation. Streamflow was logarithmically transformed since that leads to a nearly Gaussian distribution for annual streamflow. Annual streamflow records of zero were replaced with half the minimum annual streamflow prior to the application of the log transform.

3.3 A Paleoclimate Record of North American Drought

The LBDA is an updated version of the seminal North American Drought Atlas (NADA) (Cook et al., 1999; Cook and Krusic, 2004; Cook et al., 2010a), which is a paleoclimate reconstruction of the summer (June to August – JJA) PDSI based on a network of tree-ring chronologies. The LBDA has a spatial resolution of 0.5° × 0.5° latitude/longitude and incorporates information from 1845 tree-ring chronologies, an improvement over previous NADA versions that were informed by fewer tree-ring chronologies and were calculated over coarser grids (Cook et al., 1999; Cook et al., 2004). The LBDA is spatially complete over the CONUS region from 1473-2005 and includes instrumental data from 1979 onwards. A region of the LBDA ranging from 23°N to 52°N and 125°W to 66°W (see green dots in Figure 1) was extracted for the analysis. This broad region extending beyond the CONUS region was selected to capture patterns of LBDA variability relevant to the CONUS. A comparative analysis between an instrumental-based gridded PDSI dataset and the LBDA showed that they are highly correlated, with correlation values significant at the 99% level and similar variance (results not shown here). The instrumental LBDA data post-1978 were therefore also included in the modeling analysis performed here.

4 Initial Diagnostic Analyses

Initial diagnostic analyses were performed using both parametric and non-parametric correlations (Pearson and Kendall correlations respectively) between the LBDA and the log-normal streamflow series. The two correlations measures yielded similar results suggesting a near-Gaussian linear dependency (Pizarro and Lall, 2002). Different levels of temporal aggregation were tested including

monthly, rolling seasonal, bi-seasonal and annual streamflow. Different representations of the LBDA were also tested including using the LBDA grid located nearest to each streamflow gauge, using LBDA grids surrounding each streamflow gauge within a given diameter, and principal components (PCs) (Jolliffe, 2002) and archetype analysis (Cutler and Breiman, 1994) of US-wide and Missouri River Basin region LBDA (see orange and green diamonds respectively in Figure 1). An annual temporal aggregation of streamflow was found to produce the strongest signal (Figure 2 showing correlation results between streamflow gauges in the Missouri River Basin and the nearest LBDA grid and PCs of US-wide LBDA). Diagnostic tests using the nearest LBDA grid were superior and reflects the ability of the point-by-point regression method used for the LBDA to preserve local climate details in the PDSI reconstructions that are also related to streamflow.

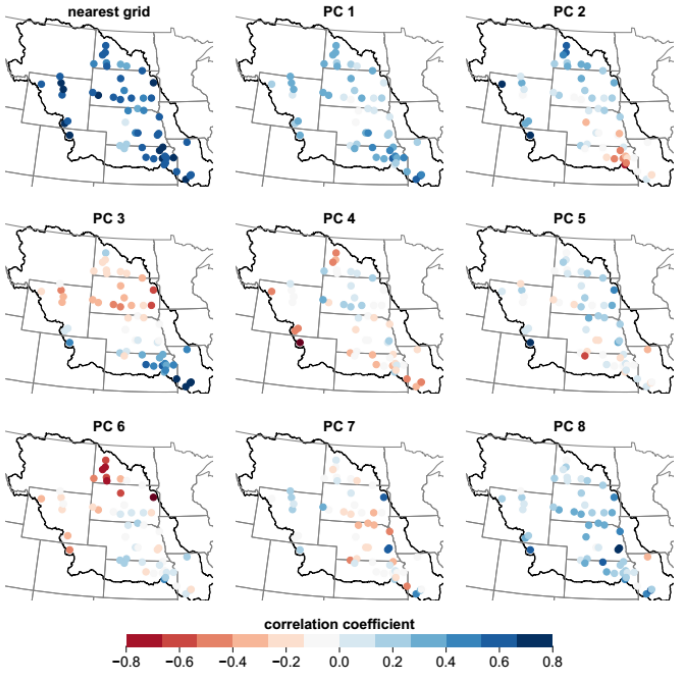


Figure 2. Pearson correlation between annual streamflow and a) closest LBDA grid b)-i) PC1 – PC8 of US-wide LBDA

High correlations were also noted for some streamflow stations with LBDA in the surrounding region (see supporting information) particularly for streamflow records with weaker correlations with the nearest grid (e.g. gauges in south-central North Dakota and on the border of Nebraska and Kansas). We

therefore considered information from LBDA grids within a 450 km radius consistent with the tree chronology search radius used to form the LBDA. Furthermore, given that large-scale climate and weather patterns influence local hydroclimatic conditions (Woodhouse et al., 2002), one needs to also consider the relationships with these larger modes of variability.

Correlations between streamflow and the first eight PCs of LBDA grids across the US (LBDA locations shown in Figure 1, correlations in Figure 2 and PC loading patterns shown in Figure 3) show that the Missouri River Basin streamflow is correlated with PC1, a PC representing overall US LBDA variability. However, these correlations are weak in comparison with correlations with the nearest LBDA grid. The north/south loading pattern of PC 2 is reflected in the change in correlation sign for stations north and south of the Nebraska and Kansas border. Similarly, the east-west difference in the loadings of PC3 leads to strong positive correlations in the downstream reaches of the Basin in Kansas and Missouri and negative correlations in Nebraska and Wyoming. The correlation results between streamflow and PCs of CONUS LBDA suggest that the large-scale modes of variability could provide additional information for modelling streamflow. Further details of the eight PCs of the CONUS LBDA and selection methods are provided in the supporting information.

5 Modeling approach and performance metrics

Here, we present a suite of plausible methods that could be implemented to model streamflow using the LBDA as the covariate(s) (Section 5.1). We justify the selection of a model that incorporates the use of regularized canonical correlation (rCCA), which is further described in Section 5.2. A description of the model performance metrics is provided in Section 5.3.

5.1 Model design and preliminary model assessment

We seek to use log-linear models to quantify the relationship between streamflow at individual gauges and a suite of site-specific LBDA records to facilitate a paleoclimate reconstruction of streamflow in the Missouri River Basin. Keeping in mind the application to streamflow record extension using the reconstructed PDSI values available in the LBDA we consider the following steps:

1. *Instrumental Period* (32-77 years at 55 locations): Use LBDA reconstructed PDSI to estimate the relationship between log transformed annual streamflow and a selection of LBDA records specific to the target streamflow site, namely $f(ln(Q_t)|\boldsymbol{PDSI}_t)$, where $\boldsymbol{PDSI}_t$ is really the expected value of PDSI informed by the tree-ring chronologies, $E[\boldsymbol{PDSI}_t|\boldsymbol{tree\text{-}ring\ chronology}_t]$, from 1929 to 2005 to maximize the overlap between the two sets of variables.

2. *Paleo Period (1473 to 1929):* Use $f(ln(Q_t)|\boldsymbol{PDSI}_t)$ estimated in the previous step with the LBDA reconstructed PDSI, $E[\boldsymbol{PDSI}_t|\boldsymbol{tree\text{-}ring\ chronology}_t]$, to estimate $ln(Q_t)$ prior to 1929 (estimates of $ln(Q_t)$ post 1929 will also be shown for comparison).

Several reconstruction model designs were considered given the initial diagnostic results. These included:

1) Developing one model for each streamflow station with predictors comprised of either:
    a. the LBDA grid located closest to the streamflow gauge;
    b. PCs of the LBDA from a region around the Missouri River Basin (Figure 1, green diamonds) within a multilinear model framework;
    c. canonical correlation analysis with regularization(rCCA described in Section 5.2) using LBDA grids within a 450km radius; or
    d. rCCA using LBDA grids within a 450km radius in addition to the retained PCs of CONUS-wide LBDA.

2) Developing one model for all streamflow stations with predictors comprised of either:
    a. PCs of both streamflow and CONUS-wide LBDA; or
    b. rCCA using streamflow and LBDA information from either the Missouri River Basin region or the CONUS region.

The fine resolution, gridded LBDA data results in a large number of highly correlated, potential predictors in all of the model designs considered with the exception of 1a. Given the 0.5° × 0.5° resolution of the LBDA data, if a reconstruction of streamflow at a single gauge is considered using only a 450 km radius of surrounding LBDA data, one has around 300 potential predictors including the leading eight CONUS LBDA PCs. The length of annual streamflow records in the Missouri River Basin ranges from 39 to 85 years, and thus it is clear that the number of potential predictors is significantly greater than the number of observations available to fit the model. Of course, if one were to consider a model with a simultaneous reconstruction of all the streamflow records, the dimension of the estimation problem becomes even more challenging. Consequently, this is the first issue considered in model development.

A single reconstruction model would be advantageous in its ability to consider all streamflow stations at once rather than fitting 55 individual models. However, the varying streamflow record lengths and differences in record periods would have resulted in large uncertainties as a large degree of annual streamflow imputation would have been required. As a result, model designs under option 2 were not further examined.

A cursory comparison of the viable models was conducted by comparing the coefficient of determination for each fitted model. The development of individual models for each streamflow gauge using either model 1a (the closest LBDA grid) or 1b (PCs of the LBDA in a region around the Missouri River Basin) resulted in acceptable models of streamflow (mean adjusted R^2 across the two models were

0.37 and 0.33 respectively). However, the spatially complete LBDA presents the opportunity to include a wider variety of local and large-scale information and these models (model designs 1c and 1d) were superior to the simpler models (model designs 1a and 1b). In addition, rCCA using both local and CONUS-wide information (model 1d) resulted in slightly improved model results (mean adjusted $R^2$of 0.74 across all individual station models) over using only local information (model 1c, mean adjusted R^2 of 0.71). We therefore selected model 1d (rCCA using LBDA grids within a 450km radius in addition to the retained PCs of CONUS-wide LBDA) for further analysis. A schematic of the data used to fit this model is shown in Figure 3, while a more detailed description of rCCA is provided in Section 5.2.

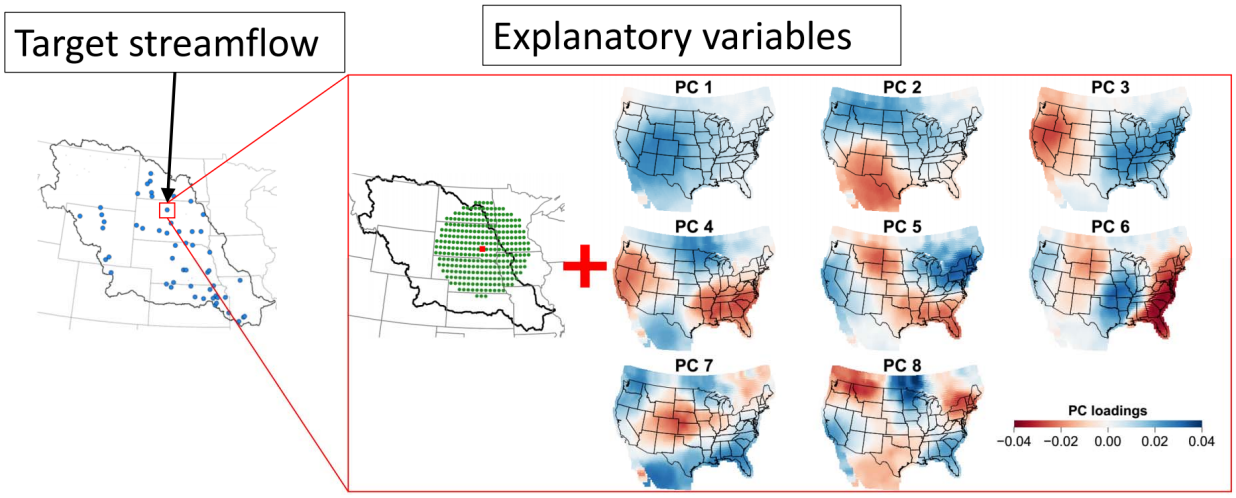


Figure 3. Schematic of LBDA information to be included in the reconstruction model of one streamflow station in the Missouri River Basin. The model inputs are the LBDA within a 450km radius and the first 8 PCs of US-wide LBDA.

5.2 Regularized canonical correlation analysis

Methods such as principal component analysis (PCA, Jolliffe, 2002), archetype analysis (AA, Cutler and Breiman, 1994; Stone and Cutler, 1996; Steinschneider and Lall, 2015) and canonical correlation analysis (CCA, Hotelling, 1936) are typically used for dimension reduction in this setting. Given that the number of potential predictors exceeds the number of observations, their high mutual correlation (>0.95 for many of the adjacent grids), and an interest in exploring a multivariate streamflow response, we explore the use of regularized canonical correlation where the regularization procedure is akin to ridge

302 regression (De Bie and De Moor, 2003). Regularized canonical correlation was selected for use here to
303 capitalize on its ability to tailor the dimension reduction process to maximize the correlation between
304 the explanatory and target variables (in contrast to PCA where the explanatory and target variable
305 relationship is not considered).

306 Canonical correlation was introduced by Hotelling (1936) as a method of linearly transforming two
307 vector variables to canonical form to maximize the correlation between them. Consider two sets of
308 random variables represented by two matrices **X** and **Y**. **X** is a $n \times p$ matrix where $\mathbf{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p]$ containing
309 n observations at p different locations, while **Y** is a $n \times q$ matrix where $\mathbf{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_q]$ containing n
310 observations at q different locations. Both **X** and **Y** have finite variance matrices represented by $\boldsymbol{\Sigma}_{XX}$ and
311 $\boldsymbol{\Sigma}_{YY}$ respectively. The covariance matrix between **X** and **Y** is $\boldsymbol{\Sigma}_{XY}$ while the covariance matrix between **Y**
312 and **X** is $\boldsymbol{\Sigma}_{YX}$. In this application, **X** consists of the LBDA inputs specific to each streamflow gauge and **Y** is
313 the streamflow at one station (i.e. p is large and q = 1). Canonical correlation analysis involves rotating
314 the coordinate axes of both **X** and **Y** to new coordinate systems in order to clearly exhibit correlation
315 between **X** and **Y**. An arbitrary linear combination could be $\mathbf{U}=\mathbf{X}\times\boldsymbol{\alpha}$ and $\mathbf{V}=\mathbf{Y}\times\boldsymbol{\gamma}$ such that the correlation
316 between **U** and **V** is maximized. **U** and **V** yield the first pair of canonical variates, while $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ are the
317 vectors of canonical weights of length p and q respectively. This process may be repeated subject to the
318 constraint that following pairs of canonical variates are orthogonal to previous pairs with a maximum of
319 min(p, q) pairs obtained. In this application, the model of streamflow is applied station by station and
320 therefore only one pair of canonical variates are calculated and the first canonical variate of the LBDA is
321 used to fit the model of streamflow.

322 The correlation of successive pairs of canonical variates can be found using an eigen decomposition of
323 $\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{XY}\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\Sigma}_{YX}$ and $\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{XY}$. The resulting min(p, q) eigenvalues are common to both and the
324 square root of the eigenvalues yield the canonical correlation. The eigenvectors of $\boldsymbol{\Sigma}_{XX}^{-1}\,\boldsymbol{\Sigma}_{XY}\,\boldsymbol{\Sigma}_{YY}^{-1}\,\boldsymbol{\Sigma}_{YX}$ and
325 $\boldsymbol{\Sigma}_{YY}^{-1}\,\boldsymbol{\Sigma}_{YX}\,\boldsymbol{\Sigma}_{XX}^{-1}\,\boldsymbol{\Sigma}_{XY}$ respectively yield the canonical weights $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ that are used to transform **X** and **Y**.

These weights are akin to the beta values in a multiple linear regression. Transforms using the i^{th} eigenvector result in correlations corresponding to the square root of the i^{th} eigenvalue. Here, the weight for **Y** (streamflow) is 1 and the weights for **X** (LBDA) are given by the first eigenvector of $\mathbf{\Sigma}_{YY}^{-1}\mathbf{\Sigma}_{YX}\mathbf{\Sigma}_{XX}^{-1}\mathbf{\Sigma}_{XY}$.

We employ regularization of the CCA process to address the issue of a large number of predictors relative to the number of observations (Vinod, 1976). Regularization is a smoothing process where a "roughness penalty", also known as the regularization parameter (λ), is introduced by converting $\mathbf{\Sigma}_{YY}$ and $\mathbf{\Sigma}_{XX}$ to $\mathbf{\Sigma}_{YY} + \lambda_y\mathbf{I}$ and $\mathbf{\Sigma}_{XX} + \lambda_x\mathbf{I}$ respectively (Leurgans et al., 1993) and is similar to the technique of ridge regression (De Bie and De Moor, 2003). Values of λ range between 0 and 1 with larger λ values indicating a higher degree of smoothing. Regularization also enables the process of matrix inversion to be stabilized. A suitable value of λ is determined using a leave one out cross validation score, where λ_x and λ_y are selected such that the correlation between the transformed datasets are maximized while the degrees of freedom used (as defined by Dijkstra, 2014) are limited to a maximum of *n*-10. If the criteria for the maximum degrees of freedom could not be achieved λ was set to one, otherwise λ was evaluated to two significant figures. No regularization was required for the single variable streamflow, whilst the LBDA was heavily regularized in almost all cases (Figure 4). 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/).

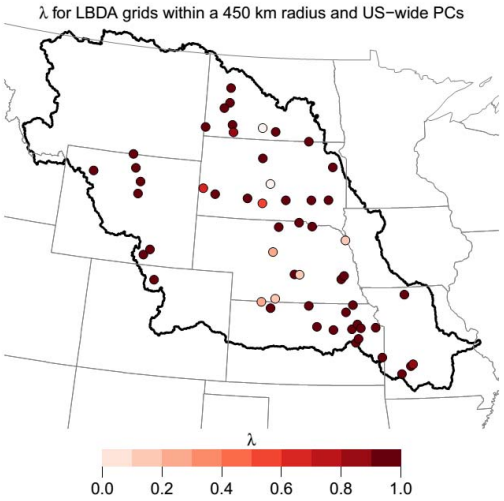


Figure 4. CCA regularization parameter values for the explanatory variables (LBDA grids within a 450km radius and CONUS-wide PCs) for each model of LBDA and streamflow.

5.3 Model performance metrics

The performance of the model selected for further analysis includes verification using a leave-k-out cross-validation procedure. Ten percent of the data (between 3 and 8 years out of a total of 32 and 77 years of streamflow data overlapping the LBDA record) are randomly selected and withheld from model fitting. These values are then predicted from the model fit to the balance of the data. The entire process is repeated 100 times, thus providing a set of 100 k-fold cross-validation samples. A comparison of the model residuals resulting from both calibrated and verified model inputs is made for the 100 cross-validation samples. The coefficient of efficiency (CE) and the reduction of error (RE) (Cook et al., 1994; Wilson et al., 2010) are additional metrics calculated to verify the model. Both CE and RE are similar to the Nash-Sutcliff efficiency test, however, the metrics are normalized using the mean of the verification and calibration data respectively. Namely, CE is defined as

$$CE = 1 - \frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)}{\sum_{i=1}^{n}(x_i - \bar{x}_v)}$$

359 while RE is defined as

$$RE = 1 - \frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)}{\sum_{i=1}^{n}(x_i - \bar{x}_c)}$$

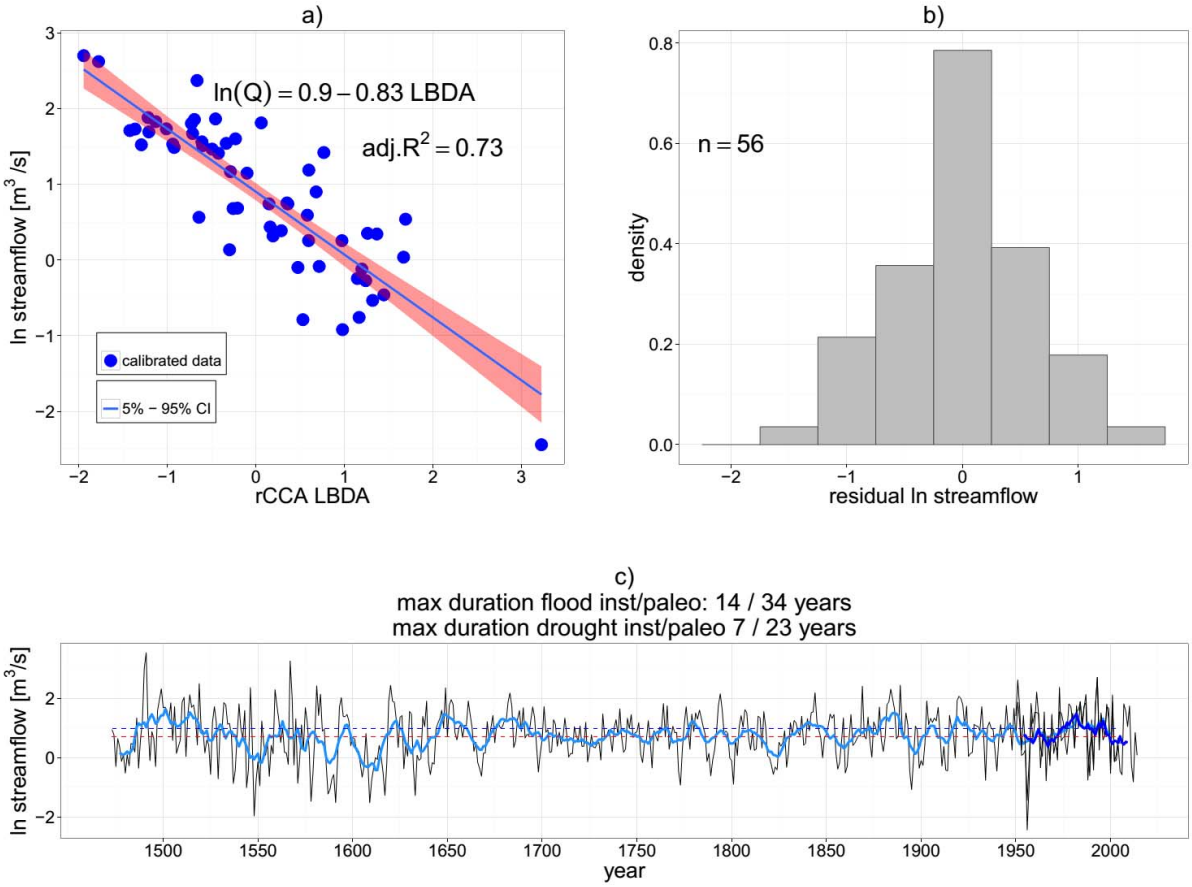
360 Where x_i and $\hat{x}_i$ are the observed and modelled streamflows respectively, while $\bar{x}_v$ and $\bar{x}_c$ are the
361 means of the observed streamflow in the validation and calibration periods respectively.

362 6 Results

363 A log-linear model was individually fitted to each streamflow gauge using logarithmically transformed
364 streamflow and the first canonical variate of the LBDA inputs (schematic of inputs shown in Figure 3)
365 using rCCA. The models were tested using a cross-validation procedure and calibration and verification
366 metrics, as described in Section 5.3 were calculated. Finally, an overall assessment of the dominant
367 modes of temporal variability in reconstructed streamflow variability in the Missouri River Basin was
368 made using a frequency wavelet analysis on the leading PCs of reconstructed streamflow.

369 6.1 Model results

370 Streamflow at each of the 55 stations in the Missouri River Basin was constructed using a least squares
371 regression model of natural-log streamflow and the first regularized canonical variate of the combined
372 LBDA grids within a 450 km radius and first eight PCs of US-wide LBDA as the predictor variable. All
373 results for modeled and reconstructed streamflow are shown in log space, while verification statistics
374 are also calculated for logarithmic streamflow. An example of the input and modeled streamflow is
375 shown for one streamflow station calibrated using data from Turkey Creek near Seneca (Figure 5). A
376 summary of modeled streamflow results and summary statistics of the residuals for both calibrated and
377 verified streamflow across all 100 calibration and verification sets are provided in the supporting
378 information.



379

380 *Figure 5. Model results for Turkey Creek near Seneca (USGS gauge number 06814000) calibrated using all available data and the*
381 *first canonical variate of the LBDA inputs a) input (dots) and modelled (line) natural log streamflow showing 5th – 95th prediction*
382 *interval, b) histogram of model residuals, and c) flood and drought persistence gauged by a threshold of mean±0.5SD of ten year*
383 *running average (blue lines – light blue line is the reconstructed ten year moving average).*

384 The rCCA linear model resulted in 55 models with adjusted R^2 values ranging between 0.56 and 0.90.

385 The model residuals were near normal at almost all stations with the exception of some stations with

386 small catchment regions showing near-uniform residuals. In each streamflow model, the magnitude of

387 rCCA loadings for the eight PCs of CONUS LBDA were similar to those of the ~300 local LBDA grids

388 indicating that information from the eight PCs of CONUS LBDA did not dominate the reconstruction.

389 6.2 Model validation

390 The predictive power of the model was assessed by calculating the coefficient of efficiency (CE) and

391 reduction of error (RE) for all cross validations. The median CE and RE values are shown in Figure 6 a and

b respectively, while distributions of the values are shown in box plots in the supporting information. CE and RE values range from -∞ to 1, with values over 0 indicating that the model predictions are more accurate than the respective climatology used for each statistic.

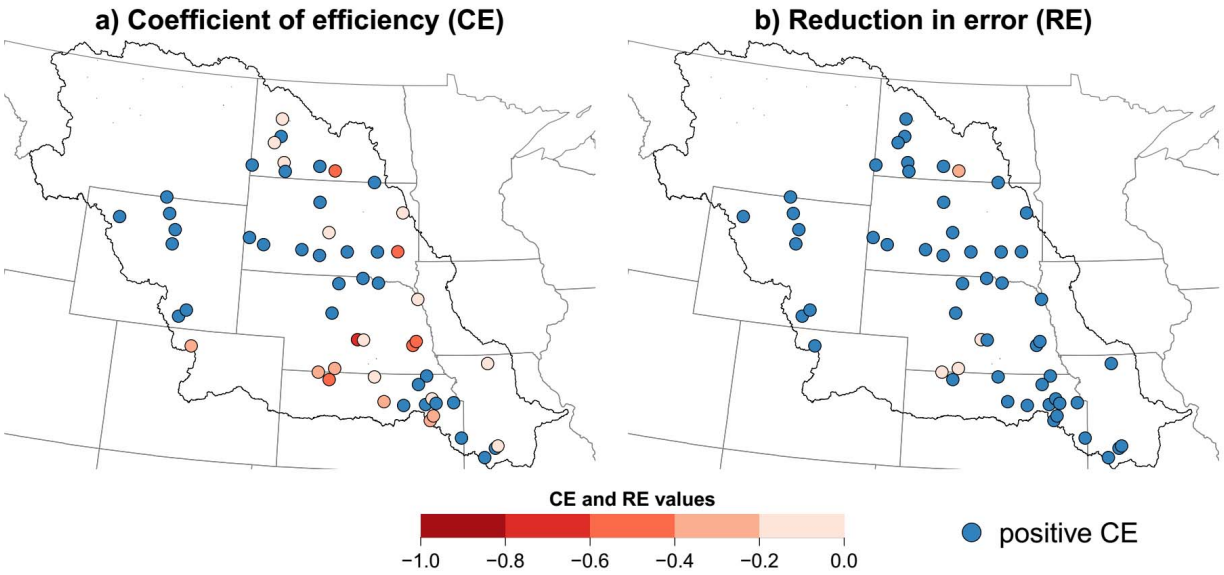
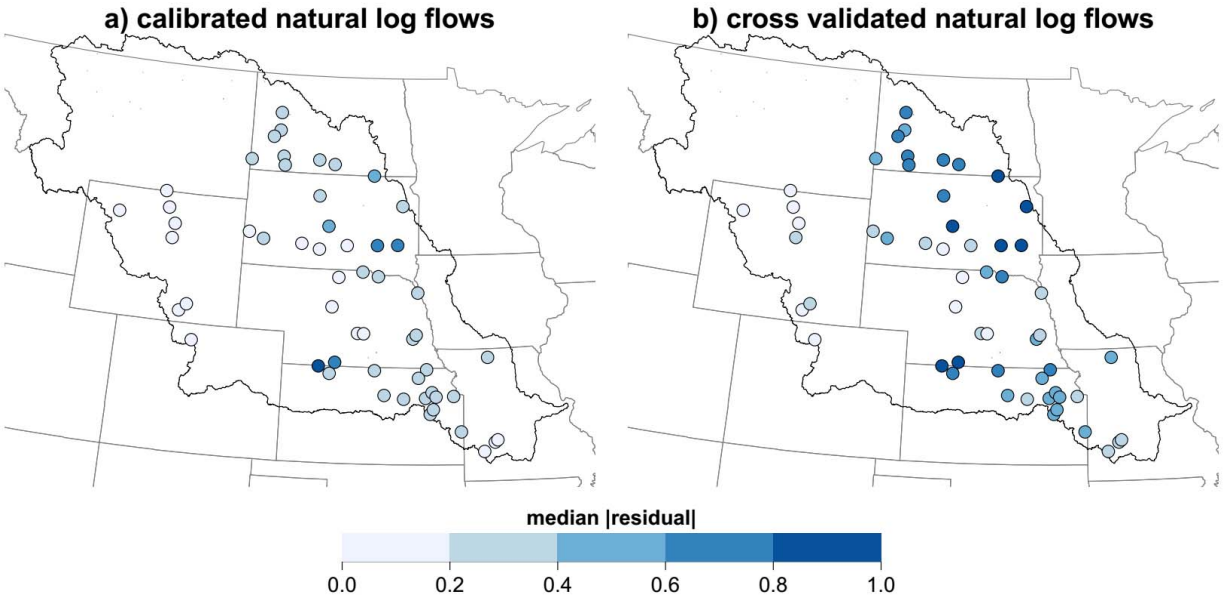


Figure 6. Median values of a) CE and b) RE values for cross validated models of the 55 Missouri River Basin streamflow stations.

The CE values in Figure 6 indicate that many of the streamflow models are able to provide median streamflow predictions with greater accuracy than the climatology of the withheld data. The RE values likewise suggest that the models are able to predict the withheld values with greater accuracy than the climatology of the calibrated values at most stations. The combined CE and RE results suggest there is some skill in reconstructing paleoclimate streamflow using the LBDA.

A comparison of the distribution of residuals resulting from calibrated values vs. validated values was also made using the 100 repetitions of the k-fold cross validation test. The median and the range of the cross-validated absolute residuals are modestly larger than that of the absolute residuals during the fitting process. The differences, accounting for the sample sizes, are not statistically significant in over 90% of 100 cross validated results (using a two-sided t-test and a null hypothesis that the true difference between the means is zero and alpha = 0.05). The median values of the cross validated residuals are

408 shown in Figure 7, while box plots showing the distribution of the cross validated residuals are shown in
409 the supporting information. A summary of the reconstruction is presented in Section 6.3 using PCs of the
410 streamflow reconstructed at the 55 stations.



411

412 *Figure 7. Median cross validated model residuals (absolute value) for a) calibrated and b) verified natural log streamflow inputs*
413 *at the 55 Missouri River Basin streamflow stations. Cross validations were repeated 100 times.*

414 ## 6.3 Reconstructed streamflow analysis

415 Mean annual streamflow was reconstructed for all 55 stations in the Missouri River Basin (data available
416 in Ho et al., 2016). PCA was used to extract the leading modes of variability from the reconstructed
417 Missouri River Basin natural log streamflow at all 55 stations. A combined assessment of the PCs using
418 North's Rule of Thumb and a scree plot showed adjacent degenerate eigenvalues pairs for PCs of order 4
419 and higher and discontinuities at PC4 respectively (shown in supporting information). Furthermore, the
420 spatial pattern of PC4 was difficult to interpret and therefore only the first three PCs are shown in Figure
421 8.

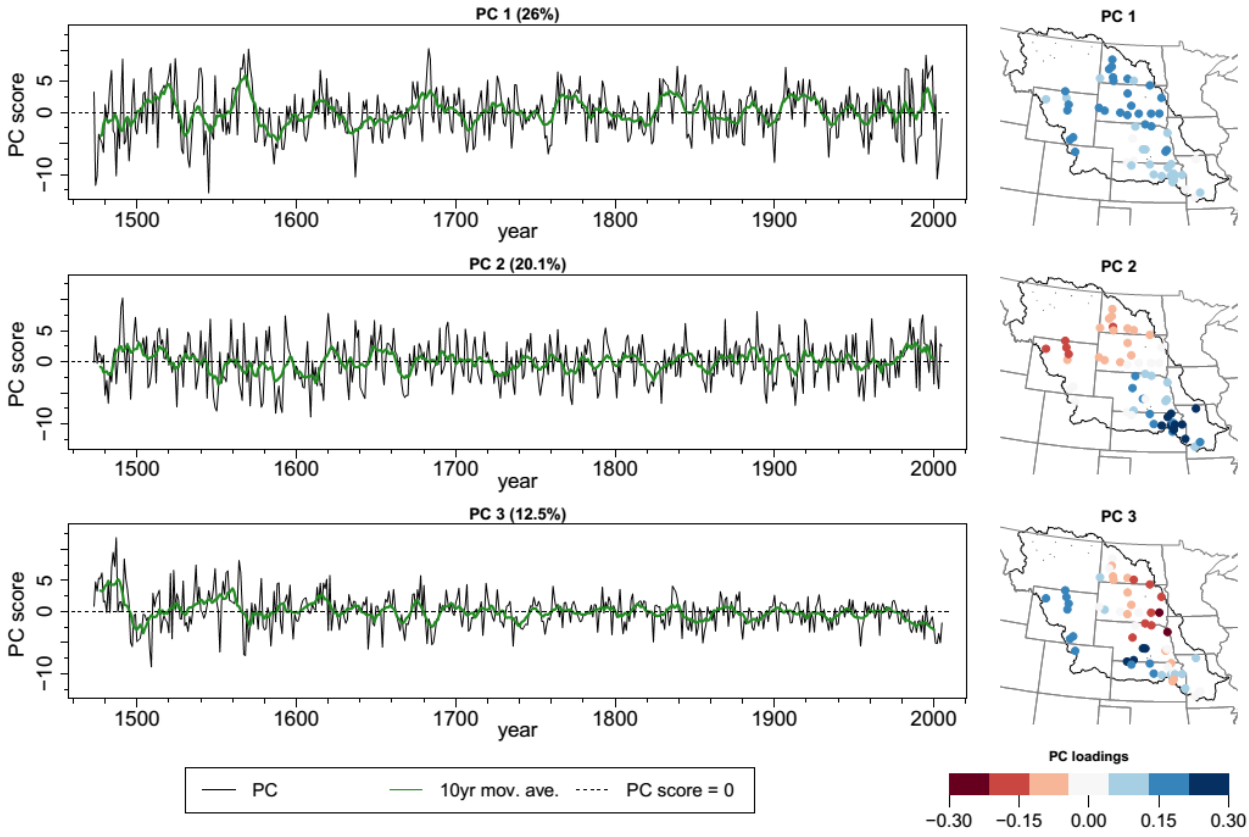


Figure 8. Annual (black) and 10 year moving average (green line) of the first four PCs of reconstructed mean natural log streamflow in the Missouri River Basin (left) with percentage variance explained shown in parenthesis and the corresponding loading patterns (right).

Negative streamflow anomalies coinciding with the 1930s dust bowl drought and 1950s drought are evident in PC1, which has a positive loading pattern across all stations in the Missouri River Basin (Figure 8). The 1950s drought had more severe impacts in the southern half of the Missouri River Basin (Piechota and Dracup, 1996; Cole et al., 2002; Andreadis et al., 2005; Cook et al., 2009) and this is detected in PC2 that is comprised of positive loadings in the southern half of the Basin. The Civil War drought, which largely impacted the Central Plains region from the mid-1850s to mid-1860s (Herweijer et al., 2006), is also evident in PC2.

The severity of these droughts had wide ranging impacts on ecological states, agricultural produce and social activities and are often used as benchmark droughts to which contemporary droughts are compared (Breshears et al., 2005; Hornbeck, 2009). However, the streamflow reconstructions suggest

periods where streamflow deficits may have been more severe than any of these historical droughts. For example, the late 1540s, 1590s, and late 1750s all show negative streamflow anomalies in both PC1 and PC2 (26% and 20.1% variability explained respectively) that are of similar or greater magnitude than the streamflow deficits associated with the Civil War, Dust Bowl or 1950s droughts. Drought durations were also longer when assessed in the context of streamflow variability over the past 500 years. A threshold of above or below 0.5 standard deviation in decadal streamflow at each station was used as an approximation of flood and drought regimes respectively. The maximum duration of continuous periods of either flood or drought regimes were found to be longer in the majority of stations investigated as demonstrated in Figure 5 c that shows the reconstructed annual and decadal streamflow at Turkey Creek and the longest duration drought and flood regime for the station.

PC3 (12.5% variability explained) is a mode of variability largely influenced by differences in streamflow variability between the Rocky Mountains regions and the remainder of the Missouri River Basin. The apparent increase in variability in the 1500s and 1600s in PC3 suggests that differences in streamflow variability between the Rocky Mountains regions and the plains region may have been more pronounced during these two centuries.

PC2 of the reconstructed streamflow has positive loadings in the southern Missouri River Basin. This PC shows an extended period of below average flows, with the exception of a couple of years, in the first two decades of the 1600s and is evident in individual streamflow reconstructions around the border of Nebraska and South Dakota and in Kansas. A similar persistence of decadal-scale low flows are not featured in the instrumental record, suggesting that successive years of low streamflow have persisted for longer periods than what has been observed in historical records.

The PCs of the reconstructed streamflow also show several periods where annual flows are larger than the maximum instrumental streamflow. Periods of anomalously high streamflow include the 1560s for

PC 1 and the early 1490s for PC 2 and 3. These periods coincide with high positive reconstructed PDSI values particularly in the lower Missouri Basin region and along the middle and southern Rocky Mountains in the early 1490s and positive PDSI values across most of the interior plains. Anomalously large streamflow during the 1560s were reconstructed at stream gauges around the interior plains regions from tributaries that join the Missouri River in North and South Dakota (e.g. Cannonball, Grand and White Rivers). However, anomalously high streamflow around the 1560s was not reconstructed for either the Platte or Yellowstone rivers suggesting that the increase in streamflow may have resulted from weather systems delivering moisture in the upper Interior Plains (North and South Dakota), but not in the Rocky Mountains region.

In order to quantitatively assess the temporal modes of variability implied by the PCs, a continuous wavelet transform was applied to identify key periodicities in the main modes of reconstructed streamflow variability. Wavelet transforms for PCs 1-3 along with their global wavelet spectra are shown in Figure 9. The wavelet transforms were implemented using the Morlet wavelet function as the mother wavelet and padded with zeros to limit the edge effects of the wavelet analysis (Torrence and Compo, 1998).

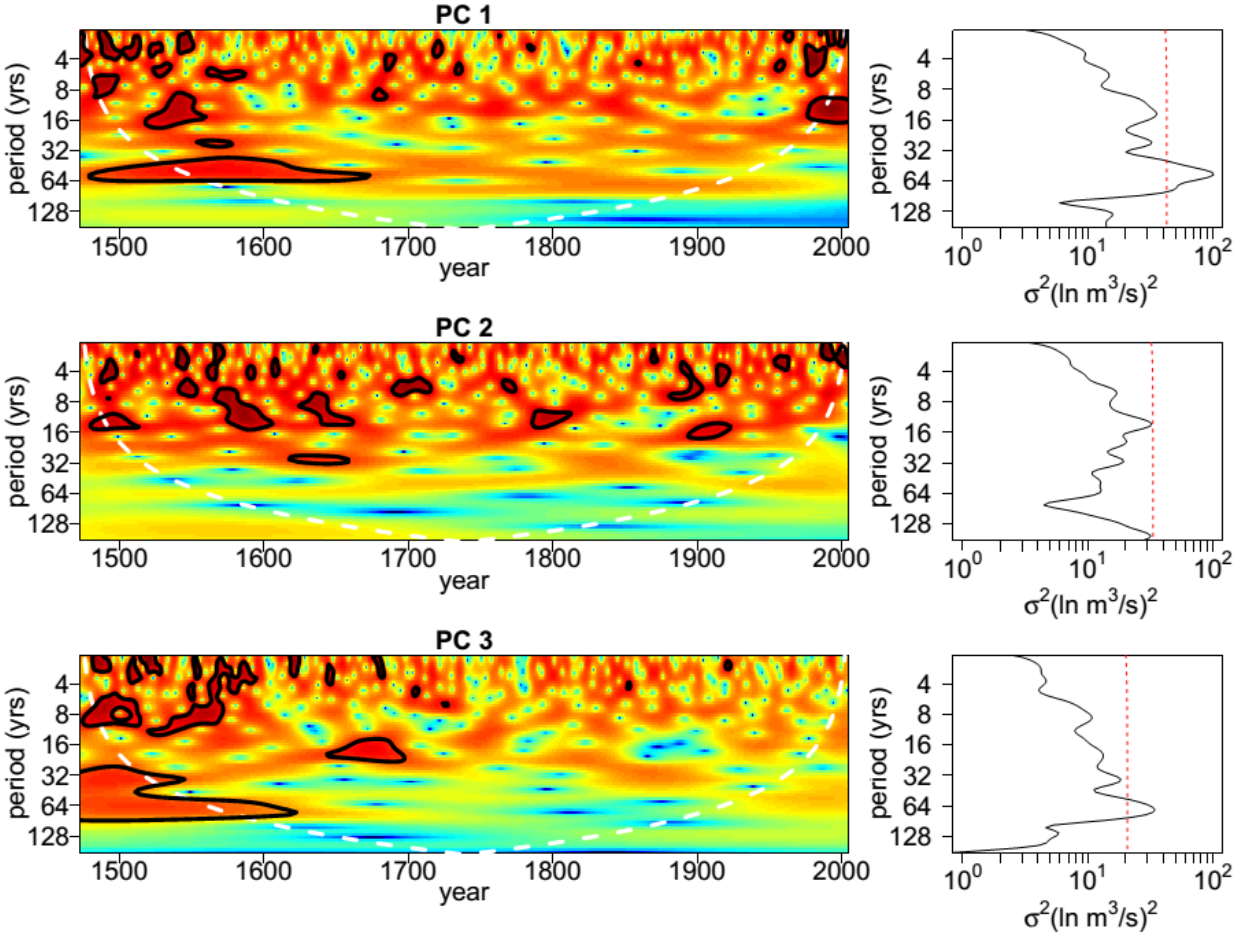


Figure 9. Wavelet transforms of PC 1-3 of reconstructed log streamflow in the Missouri River Basin from 1473-2005 (left), black lines show 95% significance level, dashed white line shows cone of influence under which edge effects may influence the results and the time averaged global wavelet spectra (right) with red dashed line showing 95% confidence level for the wavelet transform using a white noise background spectrum.

All PCs show decadal-scale variability in the 1500s and 1600s, which is significant at the 95% level against white noise for PC 1 and 3. Interestingly, none of the wavelet spectra show a consistent mode of variability that coincide with an El Niño Southern Oscillation frequency (2-7 years Allan, 2000). In addition, previously identified periods of decadal-scale variability in paleoclimate reconstructions of the Pacific Decadal Oscillation (Gedalof and Smith, 2001; MacDonald and Case, 2005) are not reproduced in the streamflow reconstruction. The multi-decadal-scale variability seen in PC1 and PC3 around the 16th century suggest that the persistence of drought and pluvial events was longer around this period compared with historic records and is reflected in previous reconstructions of persistent drought during

this period (Woodhouse and Overpeck, 1998) and in nearby regions (Meko et al., 1995). A similar pattern of multi-decadal-scale variability is also seen in the PCs of LBDA grids within the Missouri River Basin around the 16th century (results not shown here).

7 Discussion and Conclusions

Human activities and social processes are irrefutably intertwined with the hydrological cycle with each influencing and affecting the behavior of the other (Matalas et al., 1982). Consequently, hydrologic investigations need to occur at spatial scales that encompass the social and political regions influencing how water is stored, used, and transported. This requirement necessitates a broader spatial scale to be considered beyond the watershed scale typically addressed in hydrologic assessments (Vogel et al., 2015). Spatially and temporally broad measures of hydroclimatic variability such as the LBDA provide opportunities to conduct such assessments.

We used the LBDA to demonstrate a method of reconstructing mean annual streamflow in the Missouri River Basin from 1473-2005 in order to facilitate a future reconstruction of streamflow across the CONUS. The rCCA approach enabled a large degree of spatially coherent information in the LBDA to be condensed via linear transforms into a canonical variate that explained the majority of information in the target streamflow. The additional regularization step addressed the issue of an ill-posed problem where the number of variables far exceeded the number of observations. Regularization addressed potential issues with overfitting of the model and subsequent poor predictability of uncalibrated values. This site-by-site approach, opposed to a multi-site model, avoided data issues where streamflow records were not concurrently available at all sites. The method also tailored the LBDA information included in the model to each unique site and therefore has the potential to be applied to any individual streamflow site in the United States and, potentially, any streamflow site in North America where there are useful reconstructions in the LBDA.

The analysis of streamflow variability in the Missouri River Basin over the past 500 years was, however, limited by the absence of streamflow gauges in the headwaters of the Missouri River. Based on the selection criteria used here, there are no gauges located upstream of Fort Peck Lake and only one gauge in Montana was included in the analysis. The dearth of gauges in this region therefore omits potentially critical information when drawing conclusions regarding regional streamflow variability as presented in Section 6.3. Future analyses could seek to include additional streamflow gauges, potentially drawing on non-references gauges from the GAGESII database and other sources. While non-reference GAGESII gauges are located within disturbed catchments with potentially regulated streams, it may be plausible that these measurements would still preserve the signatures of interannual streamflow variability which we seek to recover. Encouragingly, the headwater regions of the Missouri River Basin contain a relatively dense network of tree-ring chronologies that would improve the fidelity of both PDSI and subsequent streamflow reconstructions. Furthermore, efforts towards expanding this tree-ring network are currently been made with a view to target and improve Missouri River Basin streamflow reconstructions (Pederson, 2013).

Dendrochronological studies for the purposes of climate reconstruction typically use a targeted sampling method whereby old living trees are selected for sampling to maximize the length of reconstruction and the location of sampling is chosen to ensure that tree growth is largely limited by the climatic factors of interest for reconstruction (Speer, 2010). Over 50 different tree species were used in the formulation of the LBDA, however, species was not a determinant for inclusion in the LBDA. Rather, the tree-ring chronologies specifically selected for use in reconstruction were those that were best correlated with available soil moisture as modeled a priori by PDSI and this was done using a method of point-by-point regression that ensures that the selected chronologies were likely to have a causal association with the gridded instrumental PDSI being reconstructed (Cook et al., 1999).

In developing our approach to reconstructing streamflow from the tree-ring based LBDA, we have capitalized on the fact that both variables 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 remove noise irrelevant to hydrological variability rendering this paleoclimate reconstruction of PDSI particularly suitable for informing streamflow. Furthermore, streamflow and PDSI are inherently different hydrological variables as demonstrated in the temporal and spatial statistics shown in the supporting information. Streamflow is also a derivative of a suite of non-climatic variables such as catchment size, land-use, geology, topography, and groundwater interactions. These catchment-specific attributes are not explicitly considered in our study (with the exception of land use considered through the selection of streamflow gauges from relatively undisturbed catchments) and the inclusion of attributes could provide an improvement to the model (e.g. Thomas and Benson, 1970; Lima and Lall, 2010).

Assessments of hydrological variability on temporal scales broader than traditional hydrological practices are paramount to identifying long-term variations and quantifying the nature of persistent regimes, which are unachievable using comparatively short instrumental record. The understanding of paleoclimate variability facilitates the development of water and land use practices that are appropriate, sustainable, and resilient under previous patterns of hydroclimatic variability and complement efforts towards developing suitable adaption procedures for projected future climate scenarios. Whilst paleoclimate reconstructions of annual streamflow are useful for assessing long-term variability, annual information is often too coarse for water resource applications such as reservoir management and seasonal water allocations. In order to produce relevant data on a monthly or daily time step, a future undertaking could involve temporal disaggregation of the data. One method that could be utilized is a non-parametric k-nearest neighbors approach (Lall and Sharma, 1996; Rajagopalan and Lall, 1999;

Nowak et al., 2010) that would avoid some issues associated with multi-site stochastic generation schemes (Valencia and Schaake, 1973; Segond et al., 2006).The objective of producing paleoclimate reconstructions of streamflow that are relevant for water resource management also suggests that the reconstruction of streamflow that is currently regulated would be valuable. Paleoclimate reconstructions of regulated streamflow could be achieved at an annual scale provided that regulations primarily impact the timing of sub-annual flows or if sufficient data prior to regularization exists. The resulting annual streamflow reconstruction could be disaggregated to a shorter timescale if pre-regulated flow data is available.

This study has capitalized on the availability of the spatially and temporally complete LBDA dataset to attempt a reconstruction of streamflow on a broad spatial scale across the Missouri River Basin. The analysis employed here provides a feasible foundation from which streamflow reconstructions across the CONUS and North America could be implemented using the LBDA. In addition, the methodology presented here could be applied to reconstructing streamflow using other reconstructions of drought including the Monsoon Asia Drought Atlas (Cook et al., 2010b), the Old World Drought Atlas covering Europe, North Africa, and the Middle East (Cook et al., 2015b) and the Australia and New Zealand summer drought atlas (Palmer et al., 2015).

Acknowledgments

We would like to sincerely thank Ben Cook (bc9z@ldeo.columbia.edu) for providing the LBDA data, the US Geological Survey (USGS) for making their monthly streamflow data available (http://waterdata.usgs.gov/nwis/monthly?referred_module=sw), staff at the USGS (Christopher Ryan cmryan@usgs.gov and Thomas Weaver tlweaver@usgs.gov) for clarifying streamflow data availability; Siyan Wang for initial data collection; Scott Steinschneider, Xun Sun and Pierre Gentine at the Columbia Water Center for their invaluable discussions on this work; and Juan A. Ballesteros and one other anonymous reviewer for their constructive comments. Reconstruction results are available on the NOAA paleoclimate data base (https://www.ncdc.noaa.gov/paleo/study/19520). This work is funded by an NSF award 1360446 and NSF award 1401698. Lamont-Doherty contribution number XXXX.

References

Allan, R. J., 2000: ENSO and Climatic Variability in the Past 150 Years. *El Niño and the Southern Oscillation: Multiscale Variability and Global and Regional Impacts*, Henry F. Diaz and Vera Markgraf, Eds., Cambridge University Press, pp.

Allen, E. B., Rittenour, T. M., DeRose, R. J., Bekker, M. F., Kjelgren, R. and Buckley, B. M., 2013: A tree-ring based reconstruction of Logan River streamflow, northern Utah, *Water Resources Research*, **49** (12), 8579-8588, doi: 10.1002/2013wr014273.

Andreadis, K. M., Clark, E. A., Wood, A. W., Hamlet, A. F. and Lettenmaier, D. P., 2005: Twentieth-Century Drought in the Conterminous United States, *Journal of Hydrometeorology*, **6** (6), 985-1001, doi: 10.1175/JHM450.1.

Ault, T. R., Cole, J. E., Overpeck, J. T., Pederson, G. T. and Meko, D. M., 2014: Assessing the risk of persistent drought using climate model simulations and paleoclimate data, *Journal of Climate*, **27** (20), 7529-7549, doi: 10.1175/jcli-d-12-00282.1.

Breshears, D. D., Cobb, N. S., Rich, P. M., Price, K. P., Allen, C. D., Balice, R. G., Romme, W. H., Kastens, J. H., Floyd, M. L., Belnap, J., Anderson, J. J., Myers, O. B. and Meyer, C. W., 2005: Regional vegetation die-off in response to global-change-type drought, *Proceedings of the National Academy of Sciences of the United States of America*, **102** (42), 15144-15148, doi: 10.1073/pnas.0505734102.

Brown, D. P. and Comrie, A. C., 2004: A winter precipitation 'dipole' in the western United States associated with multidecadal ENSO variability, *Geophysical Research Letters*, **31** (9), doi: 10.1029/2003GL018726.

Buuren, S. v. and Groothuis-Oudshoorn, K., 2011: mice: Multivariate Imputation by Chained Equations in R, *Journal of Statistical Software*, **45** (3), 67.

608 Cole, J. E., Overpeck, J. T. and Cook, E. R., 2002: Multiyear La Niña events and persistent drought in the
609 contiguous United States, *Geophysical Research Letters*, **29** (13), 25-1-25-4, doi:
610 10.1029/2001GL013561.
611 Cook, B. I., Miller, R. L. and Seager, R., 2009: Amplification of the North American "Dust Bowl" drought
612 through human-induced land degradation, *Proceedings of the National Academy of Sciences*,
613 **106** (13), 4997-5001, doi: 10.1073/pnas.0810200106.
614 Cook, B. I., Ault, T. R. and Smerdon, J. E., 2015a: Unprecedented 21st century drought risk in the
615 American Southwest and Central Plains, *Science Advances*, **1** (1), doi: 10.1126/sciadv.1400082.
616 Cook, B. I., Smerdon, J. E., Seager, R. and Cook, E. R., 2013a: Pan-Continental Droughts in North America
617 over the Last Millennium, *Journal of Climate*, **27** (1), 383-397, doi: 10.1175/jcli-d-13-00100.1.
618 Cook, E. R. and Krusic, P. J., 2004:The North American Drought Atlas,
619 Cook, E. R., Briffa, K. R. and Jones, P. D., 1994: Spatial regression methods in dendroclimatology: A
620 review and comparison of two techniques, *International Journal of Climatology*, **14** (4), 379-402,
621 doi: 10.1002/joc.3370140404.
622 Cook, E. R., Meko, D. M., Stahle, D. W. and Cleaveland, M. K., 1999: Drought Reconstructions for the
623 Continental United States, *Journal of Climate*, **12** (4), 1145-1162, doi: 10.1175/1520-
624 0442(1999)012<1145:drftcu>2.0.co;2.
625 Cook, E. R., Woodhouse, C. A., Eakin, C. M., Meko, D. M. and Stahle, D. W., 2004: Long-Term Aridity
626 Changes in the Western United States, *Science*, **306** (5698), 1015-1018, doi:
627 10.1126/science.1102586.
628 Cook, E. R., Seager, R., Heim, R. R., Vose, R. S., Herweijer, C. and Woodhouse, C., 2010a: Megadroughts
629 in North America: placing IPCC projections of hydroclimatic change in a long-term palaeoclimate
630 context, *Journal of Quaternary Science*, **25** (1), 48-61, doi: 10.1002/jqs.1303.
631 Cook, E. R., Anchukaitis, K. J., Buckley, B. M., D'Arrigo, R. D., Jacoby, G. C. and Wright, W. E., 2010b:
632 Asian Monsoon Failure and Megadrought During the Last Millennium, *Science*, **328** (5977), 486-
633 489, doi: 10.1126/science.1185188.
634 Cook, E. R., Palmer, J. G., Ahmed, M., Woodhouse, C. A., Fenwick, P., Zafar, M. U., Wahab, M. and Khan,
635 N., 2013b: Five centuries of Upper Indus River flow from tree rings, *Journal of Hydrology*, **486**
636 (0), 365-375, doi: 10.1016/j.jhydrol.2013.02.004.
637 Cook, E. R., Seager, R., Kushnir, Y., Briffa, K. R., Büntgen, U., Frank, D., Krusic, P. J., Tegel, W., van der
638 Schrier, G., Andreu-Hayles, L., Baillie, M., Baittinger, C., Bleicher, N., Bonde, N., Brown, D.,
639 Carrer, M., Cooper, R., Čufar, K., Dittmar, C., Esper, J., Griggs, C., Gunnarson, B., Günther, B.,
640 Gutierrez, E., Haneca, K., Helama, S., Herzig, F., Heussner, K.-U., Hofmann, J., Janda, P., Kontic,
641 R., Köse, N., Kyncl, T., Levanič, T., Linderholm, H., Manning, S., Melvin, T. M., Miles, D., Neuwirth,
642 B., Nicolussi, K., Nola, P., Panayotov, M., Popa, I., Rothe, A., Seftigen, K., Seim, A., Svarva, H.,
643 Svoboda, M., Thun, T., Timonen, M., Touchan, R., Trotsiuk, V., Trouet, V., Walder, F., Ważny, T.,
644 Wilson, R. and Zang, C., 2015b: Old World megadroughts and pluvials during the Common Era,
645 *Science Advances*, **1** (10), doi: 10.1126/sciadv.1500561.
646 Cutler, A. and Breiman, L., 1994: Archetypal Analysis, *Technometrics*, **36** (4), 338-347, doi:
647 10.1080/00401706.1994.10485840.
648 Dai, A., 2013: Increasing drought under global warming in observations and models, *Nature Clim.*
649 *Change*, **3** (1), 52-58, doi: 10.1038/nclimate1633.
650 De Bie, T. and De Moor, B., 2003: On the regularization of canonical correlation analysis, *Int. Sympos. ICA*
651 *and BSS*, 785-790.
652 Dettinger, M. D., Cayan, D. R., Diaz, H. F. and Meko, D. M., 1998: North–South Precipitation Patterns in
653 Western North America on Interannual-to-Decadal Timescales, *Journal of Climate*, **11** (12), 3095-
654 3111, doi: 10.1175/1520-0442(1998)011<3095:NSPPIW>2.0.CO;2.

Devineni, N., Lall, U., Pederson, N. and Cook, E., 2013: A Tree-Ring-Based Reconstruction of Delaware River Basin Streamflow Using Hierarchical Bayesian Regression, *Journal of Climate*, **26** (12), 4357-4374, doi: 10.1175/jcli-d-11-00675.1.

Dijkstra, T., 2014: Ridge regression and its degrees of freedom, *Quality & Quantity*, **48** (6), 3185-3193, doi: 10.1007/s11135-013-9949-7.

Driscoll, D., 2013: *Hydroclimatic information and risk assessment methods for managing hydrologic extremes in the Missouri River Basin - A white paper for future studies*. *Hydroclimatic extremes in the Missouri River Basin*, USGS.

Fritts, H. C., 1976: *Tree Rings and Climate*. Academic Press Inc., 567 pp.

Galat, D. L., Berry Jr, C. R., Peters, E. J. and White, R. G., 2005: Chapter 10 - Missouri River Basin. *Rivers of North America*, Arthur C. Benke and Colbert E. Cushing, Eds., Academic Press, pp. 426-480.

Gallant, A. J. E. and Gergis, J., 2011: An experimental streamflow reconstruction for the River Murray, Australia, 1783-1988, *Water Resources Research*, **47** (12), W00G04, doi: 10.1029/2010wr009832.

Gedalof, Z. e. and Smith, D. J., 2001: Interdecadal climate variability and regime-scale shifts in Pacific North America, *Geophys. Res. Lett.*, **28** (8), 1515-1518, doi: 10.1029/2000gl011779.

Gershunov, A. and Barnett, T. P., 1998: Interdecadal Modulation of ENSO Teleconnections, *Bulletin of the American Meteorological Society*, **79** (12), 2715-2725, doi: 10.1175/1520-0477(1998)079<2715:IMOET>2.0.CO;2.

González, I., Déjean, S., Martin, P. G. P. and Baccini, A., 2008: CCA: An R Package to Extend Canonical Correlation Analysis, *2008*, **23** (12), 14, doi: 10.18637/jss.v023.i12.

Herweijer, C., Seager, R. and Cook, E. R., 2006: North American droughts of the mid to late nineteenth century: a history, simulation and implication for Mediaeval drought, *The Holocene*, **16** (2), 159-171, doi: 10.1191/0959683606hl917rp.

Herweijer, C., Seager, R., Cook, E. R. and Emile-Geay, J., 2007: North American Droughts of the Last Millennium from a Gridded Network of Tree-Ring Data, *Journal of Climate*, **20** (7), 1353-1376, doi: 10.1175/JCLI4042.1.

Higgins, R. W., Yao, Y., Yarosh, E. S., Janowiak, J. E. and Mo, K. C., 1997: Influence of the Great Plains Low-Level Jet on Summertime Precipitation and Moisture Transport over the Central United States, *Journal of Climate*, **10** (3), 481-507, doi: 10.1175/1520-0442(1997)010<0481:IOTGPL>2.0.CO;2.

Ho, M., Kiem, A. S. and Verdon, D. C., 2015: A paleoclimate rainfall reconstruction in the Murray-Darling Basin (MDB), Australia: 2. Assessing hydroclimatic risk using paleoclimate records of wet and dry epochs, *Water Resour. Res.*, **51** (10), doi: 10.1002/2015WR017059.

Ho, M., Lall, U. and Cook, E. R., 2016:Missouri River Basin 533 Year Annual Streamflow Reconstructions,<https://www.ncdc.noaa.gov/paleo/study/19520>.

Hornbeck, R., 2009: The Enduring Impact of the American Dust Bowl: Short and Long-run Adjustments to Environmental Catastrophe, *National Bureau of Economic Research Working Paper Series*, **No. 15605**, doi: 10.3386/w15605.

Hotelling, H., 1936: Relations between two sets of variates, *Biometrika*, **28** (3-4), 321-377, doi: 10.1093/biomet/28.3-4.321.

Jolliffe, I. T., 2002: *Principal Component Analysis*. Second edition ed. Springer-Verlag, 487 pp.

Jones, P. D. and Mann, M. E., 2004: Climate over past millennia, *Rev. Geophys.*, **42** (2), RG2002, doi: 10.1029/2003rg000143.

Kunkel, K. E., Stevens, L. E., Stevens, S. E., Sun, L., Janssen, E., Wuebbles, D., Kruk, M. C., Thomas, D. P., Shulski, M., Umphlett, N., Hubbard, K., Robbins, K., Romolo, L., Akyuz, A., Pathak, T., Bergantino, T. and Dobson, J. G., 2013: Regional Climate Trends and Scenarios for the U.S. National Climate Assessment. Part 4. Climate of the U.S. Great Plains, NOAA Technical Report, NOAA, pp. 82.

Lall, U. and Sharma, A., 1996: A Nearest Neighbor Bootstrap For Resampling Hydrologic Time Series, *Water Resources Research*, **32** (3), 679-693, doi: 10.1029/95WR02966.
Leurgans, S. E., Moyeed, R. A. and Silverman, B. W., 1993: Canonical Correlation Analysis when the Data are Curves, *Journal of the Royal Statistical Society. Series B (Methodological)*, **55** (3), 725-740.
Lima, C. H. R. and Lall, U., 2010: Spatial scaling in a changing climate: A hierarchical bayesian model for non-stationary multi-site annual maximum and monthly streamflow, *Journal of Hydrology*, **383** (3–4), 307-318, doi: 10.1016/j.jhydrol.2009.12.045.
Lockwood, J. G., 1999: Is Potential Evapotranspiration and Its Relationship with Actual Evapotranspiration Sensitive to Elevated Atmospheric CO2 Levels?, *Climatic Change*, **41** (2), 193-212, doi: 10.1023/A:1005469416067.
MacDonald, G. M. and Case, R. A., 2005: Variations in the Pacific Decadal Oscillation over the past millennium, *Geophys. Res. Lett.*, **32** (8), L08703, doi: 10.1029/2005gl022478.
Matalas, N. C., Landwehr, J. M. and Wolman, M. G., 1982: Prediction in Water Management. *Scientific basis of water management*, National Academy, pp. 118-122.
McGowan, H. A., Marx, S. K., Denholm, J., Soderholm, J. and Kamber, B. S., 2009: Reconstructing annual inflows to the headwater catchments of the Murray River, Australia, using the Pacific Decadal Oscillation, *Geophys. Res. Lett.*, **36** (6), L06707, doi: 10.1029/2008gl037049.
Meko, D., Stockton, C. W. and Boggess, W. R., 1995: The tree-ring record of sever sustained drought, *JAWRA Journal of the American Water Resources Association*, **31** (5), 789-801, doi: 10.1111/j.1752-1688.1995.tb03401.x.
Monteith, J., 1965: Evaporation and environment, *Symp. Soc. Exp. Biol*, 4.
Nohara, D., Kitoh, A., Hosaka, M. and Oki, T., 2006: Impact of Climate Change on River Discharge Projected by Multimodel Ensemble, *Journal of Hydrometeorology*, **7** (5), 1076-1089, doi: 10.1175/JHM531.1.
Nowak, K., Prairie, J., Rajagopalan, B. and Lall, U., 2010: A nonparametric stochastic approach for multisite disaggregation of annual to daily streamflow, *Water Resources Research*, **46** (8), W08529, doi: 10.1029/2009WR008530.
Palmer, J. G., Cook, E. R., Turney, C. S. M., Allen, K., Fenwick, P., Cook, B. I., O'Donnell, A., Lough, J., Grierson, P. and Baker, P., 2015: Drought variability in the eastern Australia and New Zealand summer drought atlas (ANZDA, CE 1500–2012) modulated by the Interdecadal Pacific Oscillation, *Environmental Research Letters*, **10** (12), 124002, doi: 10.1088/1748-9326/10/12/124002.
Pederson, G. T., 2013, Multi-century perspectives on current and future streamflow in the Missouri River Basin, viewed 2015/09/08, <
Pederson, N., Bell, A. R., Cook, E. R., Lall, U., Devineni, N., Seager, R., Eggleston, K. and Vranes, K. P., 2012: Is an Epic Pluvial Masking the Water Insecurity of the Greater New York City Region?, *Journal of Climate*, **26** (4), 1339-1354, doi: 10.1175/jcli-d-11-00723.1.
Piechota, T. C. and Dracup, J. A., 1996: Drought and Regional Hydrologic Variation in the United States: Associations with the El Niño-Southern Oscillation, *Water Resources Research*, **32** (5), 1359-1373, doi: 10.1029/96WR00353.
Pizarro, G. and Lall, U., 2002: El Niño-induced Flooding in the U.S. West: What Can We Expect, *Eos Trans. AGU*, **82** (32), 349-352, doi: 10.1029/2002EO000255.
Prairie, J., Nowak, K., Rajagopalan, B., Lall, U. and Fulp, T., 2008: A stochastic nonparametric approach for streamflow generation combining observational and paleoreconstructed data, *Water Resources Research*, **44** (6), W06423, doi: 10.1029/2007wr006684.
Quinn, W. H., 1992: A study of the Southern Oscillation-related climatic activity for A.D 622-1990 incorporating Nile River flood data. *El Niño: Historical and Paleoclimatic Aspects of the Southern Ocean*, Henry F. Diaz and Vera Markgraf, Eds., Cambridge University Press, pp. 119-149.

751 Rajagopalan, B. and Lall, U., 1999: A k-nearest-neighbor simulator for daily precipitation and other
752 weather variables, *Water Resources Research*, **35** (10), 3089-3101, doi:
753 10.1029/1999WR900028.
754 Redmond, K. T. and Koch, R. W., 1991: Surface Climate and Streamflow Variability in the Western United
755 States and Their Relationship to Large-Scale Circulation Indices, *Water Resources Research*, **27**
756 (9), 2381-2399, doi: 10.1029/91WR00690.
757 Routson, C. C., Woodhouse, C. A. and Overpeck, J. T., 2011: Second century megadrought in the Rio
758 Grande headwaters, Colorado: How unusual was medieval drought?, *Geophys. Res. Lett.*, **38**
759 (22), L22703, doi: 10.1029/2011gl050015.
760 Seager, R., Ting, M., Held, I., Kushnir, Y., Lu, J., Vecchi, G., Huang, H.-P., Harnik, N., Leetmaa, A., Lau, N.-
761 C., Li, C., Velez, J. and Naik, N., 2007: Model Projections of an Imminent Transition to a More
762 Arid Climate in Southwestern North America, *Science*, **316** (5828), 1181-1184, doi:
763 10.1126/science.1139601.
764 Segond, M. L., Onof, C. and Wheater, H. S., 2006: Spatial–temporal disaggregation of daily rainfall from a
765 generalized linear model, *Journal of Hydrology*, **331** (3–4), 674-689, doi:
766 10.1016/j.jhydrol.2006.06.019.
767 Sheffield, J., Wood, E. F. and Roderick, M. L., 2012: Little change in global drought over the past 60
768 years, *Nature*, **491** (7424), 435-438, doi: 10.1038/nature11575.
769 Smerdon, J. E., Cook, B. I., Cook, E. R. and Seager, R., 2015: Bridging Past and Future Climate across
770 Paleoclimatic Reconstructions, Observations, and Models: A Hydroclimate Case Study, *Journal of*
771 *Climate*, **28** (8), 3212-3231, doi: 10.1175/JCLI-D-14-00417.1.
772 Smith, S. R., Green, P. M., Leonardi, A. P. and O'Brien, J. J., 1998: Role of Multiple-Level Tropospheric
773 Circulations in Forcing ENSO Winter Precipitation Anomalies, *Monthly Weather Review*, **126**
774 (12), 3102-3116, doi: 10.1175/1520-0493(1998)126<3102:ROMLTC>2.0.CO;2.
775 Speer, J. H., 2010: *Fundamentals of tree-ring research*. University of Arizona Press, 333 pp.
776 St. George, S., 2010: Tree Rings as Paleoflood and Paleostage Indicators. *Tree Rings and Natural*
777 *Hazards: A State-of-Art*, Markus Stoffel, Michelle Bollschweiler, R. David Butler, and H. Brian
778 Luckman, Eds., Springer Netherlands, pp. 233-239.
779 Steinschneider, S. and Lall, U., 2015: Daily Precipitation and Tropical Moisture Exports across the Eastern
780 United States: An Application of Archetypal Analysis to Identify Spatiotemporal Structure,
781 *Journal of Climate*, **28** (21), 8585-8602, doi: 10.1175/JCLI-D-15-0340.1.
782 Stone, E. and Cutler, A., 1996: Introduction to archetypal analysis of spatio-temporal dynamics, *Physica*
783 *D: Nonlinear Phenomena*, **96** (1–4), 110-131, doi: 10.1016/0167-2789(96)00016-4.
784 Thomas, D. M. and Benson, M. A., 1970: Generalization of streamflow characteristics from drainage-
785 basin characteristics, Water Supply Paper, 1975, pp.
786 Thornthwaite, C. W., 1948: An Approach toward a Rational Classification of Climate, *Geographical*
787 *Review*, **38** (1), 55-94, doi: 10.2307/210739.
788 Tierney, J. E., Oppo, D. W., Rosenthal, Y., Russell, J. M. and Linsley, B. K., 2010: Coordinated hydrological
789 regimes in the Indo - Pacific region during the past two millennia, *Paleoceanography*, **25**,
790 PA1102, doi: 10.1029/2009PA001871.
791 Tootle, G. A. and Piechota, T. C., 2006: Relationships between Pacific and Atlantic ocean sea surface
792 temperatures and U.S. streamflow variability, *Water Resources Research*, **42** (7), n/a-n/a, doi:
793 10.1029/2005WR004184.
794 Torrence, C. and Compo, G. P., 1998: A Practical Guide to Wavelet Analysis, *Bulletin of the American*
795 *Meteorological Society*, **79** (1), 61-78, doi: 10.1175/1520-
796 0477(1998)079<0061:APGTWA>2.0.CO;2.
797 Trenberth, K. E., Branstator, G. W. and Arkin, P. A., 1988: Origins of the 1988 North American Drought,
798 *Science*, **242** (4886), 1640-1645, doi: 10.1126/science.242.4886.1640.

U.S. Geological Survey, 2011, GAGES-II: Geospatial Attributes of Gages for Evaluating Streamflow, viewed 06/19/2015, <http://water.usgs.gov/GIS/metadata/usgswrd/XML/gagesII_Sept2011.xml>.

US Army Corps of Engineers, 2006: Missouri River Mainstem Reservoir System - Master Water Control Manual - Missouri River Basin, pp. 431.

Valencia, D. and Schaake, J. C., 1973: Disaggregation processes in stochastic hydrology, *Water Resour. Res*, **9** (3), 580-585, doi: 10.1029/WR009i003p00580.

Vance, T. R., van Ommen, T. D., Curran, M. A. J., Plummer, C. T. and Moy, A. D., 2012: A Millennial Proxy Record of ENSO and Eastern Australian Rainfall from the Law Dome Ice Core, East Antarctica, *Journal of Climate*, **26** (3), 710-725, doi: 10.1175/jcli-d-12-00003.1.

Vinod, H. D., 1976: Canonical ridge and econometrics of joint production, *Journal of Econometrics*, **4** (2), 147-166, doi: 10.1016/0304-4076(76)90010-5.

Vogel, R. M., Lall, U., Cai, X., Rajagopalan, B., Weiskel, P., Hooper, R. P. and Matalas, N. C., 2015: Hydrology: The interdisciplinary science of water, *Water Resources Research*, **51** (6), 4409-4430, doi: 10.1002/2015WR017049.

White, I. R., Royston, P. and Wood, A. M., 2011: Multiple imputation using chained equations: Issues and guidance for practice, *Statistics in Medicine*, **30** (4), 377-399, doi: 10.1002/sim.4067.

Wilson, R., Cook, E., D'Arrigo, R., Riedwyl, N., Evans, M. N., Tudhope, A. and Allan, R., 2010: Reconstructing ENSO: the influence of method, proxy data, climate forcing and teleconnections, *Journal of Quaternary Science*, **25** (1), 62-78, doi: 10.1002/jqs.1297.

Wise, E. K., 2010: Spatiotemporal variability of the precipitation dipole transition zone in the western United States, *Geophysical Research Letters*, **37** (7), doi: 10.1029/2009GL042193.

Woodhouse, C. A. and Meko, D., 1997: Number of Winter Precipitation Days Reconstructed from Southwestern Tree Rings, *Journal of Climate*, **10** (10), 2663-2669, doi: 10.1175/1520-0442(1997)010<2663:NOWPDR>2.0.CO;2.

Woodhouse, C. A. and Overpeck, J. T., 1998: 2000 years of drought variability in the central United States, *Bulletin of the American Meteorological Society*, **79** (12), 2693, doi.

Woodhouse, C. A., Gray, S. T. and Meko, D. M., 2006: Updated streamflow reconstructions for the Upper Colorado River Basin, *Water Resources Research*, **42** (5), W05415, doi: 10.1029/2005wr004455.

Woodhouse, C. A., Kunkel, K. E., Easterling, D. R. and Cook, E. R., 2005: The twentieth-century pluvial in the western United States, *Geophysical Research Letters*, **32** (7), doi: 10.1029/2005GL022413.

Woodhouse, C. A., Meko, D. M., MacDonald, G. M., Stahle, D. W. and Cook, E. R., 2010: A 1,200-year perspective of 21st century drought in southwestern North America, *Proceedings of the National Academy of Sciences*, **107** (50), 21283-21288, doi: 10.1073/pnas.0911197107.

Woodhouse, C. A., Lukas, J., Brice, B., Hirschboeck, K., Hartmann, H., Kostuk, M., Lay, E., Martinex, D., McMahan, B. and Shah, A., 2002, Tree rings and streamflow, viewed 2015/06/01, <http://treeflow.info/content/tree-rings-and-streamflow>.

Can a paleo-drought record be used to reconstruct streamflow? A case-study for the Missouri River Basin

Michelle Ho[1], Upmanu Lall[1, 2], Edward R. Cook[3]

1. Columbia Water Center, Columbia University, New York, NY, USA

2. Department of Earth and Environmental Engineering, Columbia University, New York, New York, USA

3. Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York 10964, USA

Corresponding author: Michelle Ho. Email: mh3538@columbia.edu, Phone: 212-854-7081

Keywords: paleoclimate, regularized canonical correlation analysis, streamflow reconstruction

Key points

- Streamflow is reconstructed from an existing paleoclimate drought index
- Methodological innovation using rCCA addresses very high dimensional dataset
- Reconstructed streamflow provides insights into extreme, persistent high and low flow

Abstract

Recent advances in paleoclimatology have revealed dramatic long-term hydro-climatic variations that provide a context for limited historical records. A notable dataset derived from a relatively dense network of paleoclimate proxy records in North America is the Living Blended Drought Atlas (LBDA): a gridded tree-ring based reconstruction of summer Palmer Drought Severity Index. This index has been used to assess North American drought frequency, persistence and spatial extent over the past two millennia. Here, we explore whether the LBDA can be used to reconstruct annual streamflow. Relative to streamflow reconstructions that use tree rings within the river basin of interest, the use of a gridded proxy poses a novel challenge. The gridded series have high spatial correlation, since they rely on tree rings over a common radius of influence. A novel algorithm for reconstructing streamflow using regularized canonical regression and inputs of local and global covariates is developed and applied over the Missouri River Basin, as a test case. Effectiveness in reconstruction is demonstrated with reconstructions showing periods where streamflow deficits may have been more severe than during recent droughts (e.g. the Civil War, Dust Bowl and 1950s droughts). The maximum persistence of droughts and floods over the past 500 years far exceed those observed in the instrumental record and periods of multi-decadal variability in the 1500s and 1600s are detected. Challenges for an extension to a national streamflow reconstruction or applications using other gridded paleoclimate datasets such as adequate spatial coverage of streamflow and applicability of annual reconstructions are discussed.

Michelle Ho 4/6/2016 8:25 AM
Deleted: A

Michelle Ho 4/6/2016 8:25 AM
Deleted: in the past few decades

Michelle Ho 4/6/2016 8:14 AM
Deleted:

Michelle Ho 4/6/2016 8:25 AM
Deleted: the

Michelle Ho 4/6/2016 8:19 AM
Deleted: annually-resolved

Michelle Ho 4/6/2016 8:15 AM
Deleted: the

Michelle Ho 4/6/2016 8:25 AM
Deleted: (June-August)

Michelle Ho 4/6/2016 8:21 AM
Deleted: (PDSI)

Michelle Ho 4/6/2016 8:15 AM
Deleted: the

Michelle Ho 4/6/2016 8:15 AM
Deleted: of North American droughts

Michelle Ho 4/6/2016 8:27 AM
Deleted: series to reconstruct watershed processes such as streamflow

Michelle Ho 4/6/2016 8:18 AM
Deleted: o

Michelle Ho 4/6/2016 8:19 AM
Deleted: for streamflow reconstruction

Michelle Ho 4/6/2016 8:28 AM
Deleted: streamflow

Michelle Ho 4/6/2016 8:17 AM
Deleted: in recent history

Michelle Ho 4/6/2016 8:13 AM
Deleted: An analysis of drought and flood epochs using a

Michelle Ho 4/6/2016 8:13 AM
Deleted: simple threshold analysis shows that t

Michelle Ho 4/6/2016 8:14 AM
Deleted: . A wavelet analysis reveals

Michelle Ho 4/6/2016 8:24 AM
Deleted: streamflow

1 Introduction

The concern with anthropogenic climate change and its hydrologic impacts has focused interest on how long term climate variability may impact streamflow (e.g. Nohara et al., 2006; Seager et al., 2007). The consequence of using short records to "over-allocate" the flows of major rivers are often cited as an example of the need for long records that can better inform the possible range of long-term variations of streamflow (Tootle and Piechota, 2006; McGowan et al., 2009; Woodhouse et al., 2010). Continuous records of streamflow in the US span several decades at best. Advances in paleoclimatology in the past few decades have provided opportunities across the world to extend the range of hydroclimatic variability (e.g. Quinn, 1992; Jones and Mann, 2004; Tierney et al., 2010; Gallant and Gergis, 2011; Vance et al., 2012; Cook et al., 2013b; Devineni et al., 2013; Ho et al., 2015). While considerable uncertainty clouds the projections of hydroclimatic states towards the end of the 21st century, in the near to medium term paleoclimate information may be crucial to inform the interannual to decadal variability of regional water availability as indicated by streamflow for reservoir operation, and agricultural and other water use decisions.

Paleoclimate reconstructions have been developed using proxies that typically span the past 1000-2000 years (also known as the Common Era). The North American region has a relatively dense network of high-resolution paleoclimate proxy records, primarily comprised of tree-ring chronologies. Tree-ring-proxy records have been used to assess various components of environmental variations (Fritts, 1976) including drought severity (e.g. Routson et al., 2011; Cook et al., 2015a), pluvials (e.g. Woodhouse et al., 2005; Pederson et al., 2012), streamflow variability (e.g. Woodhouse et al., 2006; Prairie et al., 2008; Allen et al., 2013; Devineni et al., 2013), and precipitation frequency (Woodhouse and Meko, 1997) in addition to enabling comparisons of past climate with projected climate scenarios (e.g. Ault et al., 2014; Cook et al., 2015a; Smerdon et al., 2015).

Michelle Ho 3/6/2016 6:01 PM
Deleted: the Colorado River is

Studies focused on the reconstruction of tree-ring-based paleo-hydrology (e.g. annual and season streamflow and floods) typically utilize proxies derived from tree-ring networks within or near the catchment region as predictors (e.g. Woodhouse et al., 2006; St. George, 2010; Pederson et al., 2012; Devineni et al., 2013). However, these networks are spatially irregular with record lengths varying across chronologies. An alternative to using spatially and temporally irregular tree-ring chronologies as model predictors is to use an existing derivative of these records, namely the Living Blended Drought Atlas (LBDA) (Cook et al., 2010a). The LBDA is a paleoclimate reconstruction of the summer (June-August) Palmer Drought Severity Index (PDSI) that is gridded across North America on a 0.5° × 0.5° latitude/longitude grid with reconstructions dating back as far as 2000 years. These records are temporally complete over the Conterminous United States (CONUS) from 1473 onward. The LBDA, or its predecessor the North American Drought Atlas (Cook and Krusic, 2004), have been used to assess the frequency and spatial distribution of droughts over the past millennia (e.g. Herweijer et al., 2007; Cook et al., 2013a).

2 A proposal for using PDSI to reconstruct streamflow

The intent of the modeling case study presented here is to develop a suitable framework with which streamflow within the CONUS may be reconstructed using a tree-ring-based reconstruction of the PDSI. In developing the modelling framework we consider; 1) the constraints posed by the LBDA and implications for reconstructing streamflow; 2) possible temporal resolutions (monthly, seasonal or annual) for direct streamflow reconstruction using the LBDA data; 3) how to best use local and far-field LBDA information for local streamflow reconstruction; and 4) provides insights from the 500 year reconstruction of the multi-site Missouri River Basin flows as to the decadal and longer variability of streamflow in the region. Given the existence of the spatially and temporally complete LBDA record over the last 500 years covering the CONUS, we explore whether the LBDA, a reconstruction of PDSI using

Michelle Ho 3/6/2016 6:13 PM
Deleted: s

tree-ring chronologies, could be a reasonable predictor of streamflow variability. In this case the variable to be reconstructed is annual streamflow with the aim of eventually reconstructing paleoclimate records of streamflow across the CONUS, an undertaking that has not previously been attempted using the LBDA or tree-rings.

Michelle Ho 3/6/2016 5:23 PM
Deleted: ,

Michelle Ho 3/6/2016 5:23 PM
Deleted: i

Michelle Ho 3/6/2016 5:23 PM
Deleted: ,

The motivation for using the LBDA stems from our understanding that the growth of moisture-limited trees, from which the LBDA is derived, are in part governed by climatic forcings that drive soil moisture availability. That is, given a vector of unspecified climate variables, $\boldsymbol{C}_t$ (where $\boldsymbol{C}$ may be comprised of, but not limited to, climate variables such as temperature, rainfall, wind, soil moisture and radiation) we can define PDSI as $PDSI_t = f_1(\boldsymbol{C}_t)$. Streamflow is also a derivative of a number of climate variables and can similarly be defined as $Q_t = f_2(\boldsymbol{C}_t)$. We seek to determine if it is possible to derive and fit a function f_3 that relates streamflow to PDSI where $Q_t = f_3(PDSI_t)$. Where suitable instrumental records of climate are available, we may derive f_1 and this has been approximated in part using methods such as the Thornthwaite potential evapotranspiration (PET) equation (Thornthwaite, 1948) and the Penman-Monteith PET equation (Monteith, 1965) to varying degrees of success (Lockwood, 1999; Sheffield et al., 2012; Dai, 2013). Consideration of the joint probability distributions $f(PDSI_t, \boldsymbol{C}_t)$ and $f(Q_t, PDSI_t)$ enables the vector of climate variables to be reconstructed by capitalizing on the climate-PDSI relationship and the tree-ring chronology-PDSI relationship given $f(\boldsymbol{C_t}|PDSI_t)f(PDSI_t|\boldsymbol{tree\text{-}ring\ chronology_t})$. Paleoclimate streamflow could similarly be derived by implementing f_2, namely $f(Q_t|\boldsymbol{C}_t)f(\boldsymbol{C}_t|PDSI_t)f(PDSI_t|\boldsymbol{tree\text{-}ring\ chronology_t})$. However, the challenge in this approach is the selection of an appropriate vector of climate variables, many of which may be sparsely observed or unobserved in the instrumental record. Therefore, we consider modelling paleoclimate streamflow through $f(Q_t|PDSI_t)f(PDSI_t|\boldsymbol{tree\text{-}ring\ chronology_t})$. Since the reconstructed $PDSI_t$ in the LBDA is really $E[PDSI_t|\boldsymbol{tree\text{-}ring\ chronology}_t]$, we start by considering $f(\boldsymbol{Q}_t|\boldsymbol{PDSI}_t)$, where $\boldsymbol{Q}_t$ is the streamflow at one or more locations in a river basin, and $\boldsymbol{PDSI}_t$ represents a vector of LBDA values at the gridded locations of the LBDA that can be a potential

predictor of the streamflows. The relationship can be developed using contemporaneous values of $\boldsymbol{PDSI}_t$ and historical $\boldsymbol{Q}_t$. Subsequently, we can apply this relationship to the paleo estimates of $\boldsymbol{PDSI}_t$ recognizing that we could use the expected values of $\boldsymbol{PDSI}_t$ reported in the LBDA, or simulations from the uncertainty distributions of $\boldsymbol{PDSI}_t$ reported in the LBDA, as conditioning variables to derive sequences of $\boldsymbol{Q}_t$.

A key motivation for using the LBDA is that it is spatially complete across CONUS over the past 500 years and a successful streamflow reconstruction would have significant value in assessing national water planning and use strategies and in investigating the different temporal and spatial structures in streamflow, which differ from the LBDA (see Figures S1 and S2 in supporting information). Relative to reconstructions that use tree rings within the river basin of interest, the use of a gridded proxy series to reconstruct watershed processes such as streamflow poses a novel challenge. The gridded LBDA series have high spatial correlation, since they rely on tree rings over a common radius of influence around each grid point, nominally 450 km in this case or roughly the correlation decay *e*-folding distance between grid points of the instrumental PDSI. A novel algorithm for reconstruction that uses local and global covariates for streamflow reconstruction using regularized canonical regression is developed and applied over the Missouri River Basin, as a test case. This provides a proof of concept at a sub-continental scale as to whether the approach is feasible. The Missouri River Basin was chosen as it contains the only major river headwaters in the western US where extensive reconstructions of paleoclimate hydrology have not been undertaken (Driscoll, 2013) and also parallels current efforts to further develop tree-ring chronologies and streamflow reconstruction models using tree rings in the region (Pederson, 2013). A description of the case study region and data are presented in the following section while Section 4 presents initial diagnostics that inform the modeling approach developed in Section 5. Section 5 details how the above proposal of using PDSI to reconstruct streamflow is implemented in the Missouri River Basin. Section 6 provides model verification results and summaries of

the key modes of variability in the 500 year mean streamflow reconstructions of Missouri River Basin streamflow. The final section reviews outstanding questions as to modeling uncertainties, and the challenges for extending the model presented here to a national scale.

3 Case Study Region and Data

3.1 The Missouri River Basin

The Missouri River is the longest river and the second largest river basin in the US, draining an area over 1.3 million km^2 that spans the southern portions of two Canadian provinces and ten states in the US (Missouri River basin boundary and key tributaries shown in Figure 1). The headwaters, which are largely snow-melt fed, are located in the Northern Rocky Mountains. These waters then flow through a largely semi-arid region to its confluence with the Mississippi River near St. Louis, Missouri (Galat et al., 2005). Land use in the Missouri River Basin is dominated by agricultural activities including cropping and grazing which cover 95% of the region (US Army Corps of Engineers, 2006), while the remaining land is used for recreation, transport, urban and industrial use including mining and energy sector activities (Galat et al., 2005). An improved perspective of streamflow variability would therefore be of benefit to the region in terms of managing and balancing the demands for water amongst the various sectors and users.

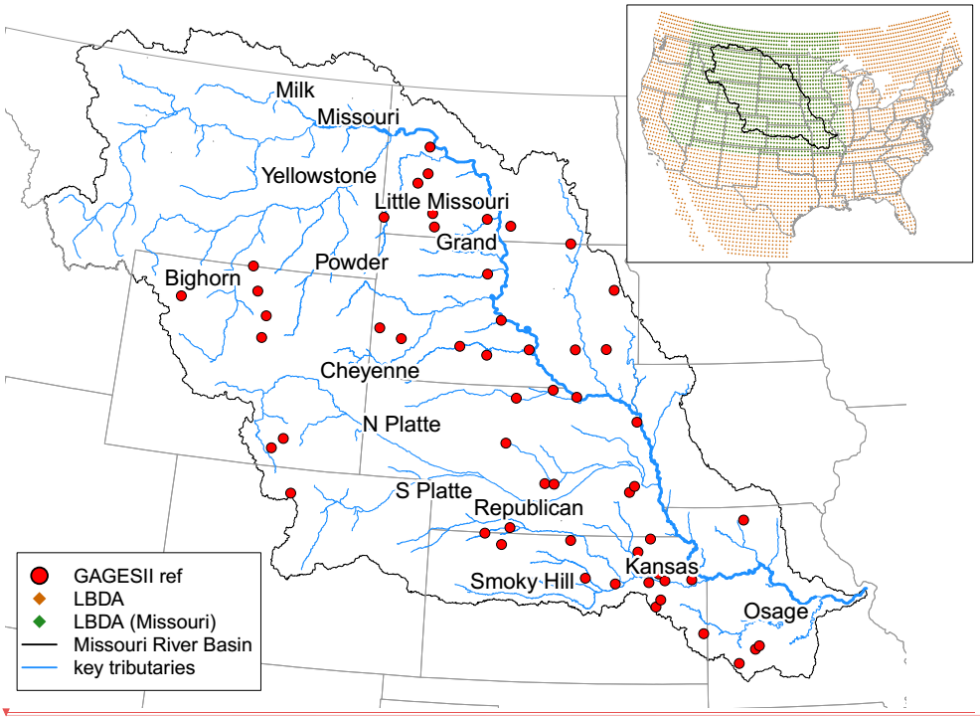
Precipitation and moisture availability in the Missouri River Basin are characterized by high precipitation in the western mountainous region, which averages over 1000 mm/year to the drier region in the rain shadow east of the Rocky Mountains where average annual precipitation is less than 400 mm/year. Precipitation increases towards the far eastern regions of the Missouri River Basin (Kunkel et al., 2013).

Winter precipitation in the northern Missouri River Basin and in the mountainous regions to the west is related to the El Niño Southern Oscillation (ENSO) signal from the preceding summer and autumn

(Redmond and Koch, 1991). El Niño teleconnections typically manifest as upper level anticyclonic high pressure cells over the northwestern US and result in the northward displacement or splitting of the jet stream and anomalously dry conditions in the Missouri River Basin (Trenberth et al., 1988; Dettinger et al., 1998; Smith et al., 1998). Conversely, La Niña events typically result in wetter conditions in this region. ENSO impacts are modulated by the decadal scale variability in the northern Pacific Ocean with warm decadal phases resulting in a deep Aleutian low and corresponding ridging over the western US thereby enhancing El Niño conditions (Gershunov and Barnett, 1998; Brown and Comrie, 2004). The enhancement of La Niña impacts by a cool phase in the northern Pacific decadal signature is particularly noticeable in the northern Missouri River Basin (Wise, 2010). Both Pacific and Atlantic Ocean influences are seen in the Northern Great Plains with the Great Plains low level jet, originating from the Gulf of Mexico, enhancing summer precipitation in this region (Higgins et al., 1997).

3.2 Streamflow data

Monthly streamflow data for the Missouri River Basin were obtained from the United States Geological Survey (USGS) Surface-Water Daily Data for the Nation (http://nwis.waterdata.usgs.gov/nwis/sw). The streamflow data are from stations included in the USGS's GAGES-II network (U.S. Geological Survey, 2011) within the Missouri River Basin boundary and are reference gauges identified by the USGS as the least-disturbed watersheds with minimal regulation. The selected gauges meet a criterion of data spanning 40 years with less than 5% missing data (Figure 1, Table S1) and results in 55 streamflow gauges, 46 of which are also in the USGS Hydro Climatic Data Network.



Michelle Ho 3/6/2016 3:53 PM

Deleted:

Figure 1. The Missouri River Basin, key tributaries, locations of gauged streamflow used in the analysis, and (inset) the LBDA grids used in the analysis. LBDA grids in the Missouri River Basin region shown in green.

Three of the selected stations had missing monthly values. These values were imputed using multiple imputation by chained equations (MICE) and a method of predictive mean matching (Buuren and Groothuis-Oudshoorn, 2011). Monthly streamflow imputation was conducted using streamflow from the three closest stations and a cosine function to represent a seasonal signal. The number of repetitions in MICE was determined using a rule of thumb method proposed by White et al. (2011) and the multiple imputed monthly values were averaged across the repetitions.

Monthly data was aggregated into annual streamflow data using a calendar year instead of a water year because the average driest month of streamflow at most stations occurred in either December or January. Start and end years with incomplete data were excluded. The resulting annual data spans from

1929 to 2014 with annual record lengths varying between 39 and 85 years after aggregation. Streamflow was logarithmically transformed since that leads to a nearly Gaussian distribution for annual streamflow. Annual streamflow records of zero were replaced with half the minimum annual streamflow prior to the application of the log transform.

3.3 A Paleoclimate Record of North American Drought

The LBDA is an updated version of the seminal North American Drought Atlas (NADA) (Cook et al., 1999; Cook and Krusic, 2004; Cook et al., 2010a), which is a paleoclimate reconstruction of the summer (June to August – JJA) PDSI based on a network of tree-ring chronologies. The LBDA has a spatial resolution of 0.5° × 0.5° latitude/longitude and incorporates information from 1845 tree-ring chronologies, an improvement over previous NADA versions that were informed by fewer tree-ring chronologies and were calculated over coarser grids (Cook et al., 1999; Cook et al., 2004). The LBDA is spatially complete over the CONUS region from 1473-2005 and includes instrumental data from 1979 onwards. A region of the LBDA ranging from 23°N to 52°N and 125°W to 66°W (see green dots in Figure 1) was extracted for the analysis. This broad region extending beyond the CONUS region was selected to capture patterns of LBDA variability relevant to the CONUS. A comparative analysis between an instrumental-based gridded PDSI dataset and the LBDA showed that they are highly correlated, with correlation values significant at the 99% level and similar variance (results not shown here). The instrumental LBDA data post-1978 were therefore also included in the modeling analysis performed here.

4 Initial Diagnostic Analyses

Initial diagnostic analyses were performed using both parametric and non-parametric correlations (Pearson and Kendall correlations respectively) between the LBDA and the log-normal streamflow series. The two correlations measures yielded similar results suggesting a near-Gaussian linear dependency (Pizarro and Lall, 2002). Different levels of temporal aggregation were tested including

monthly, rolling seasonal, bi-seasonal and annual streamflow. Different representations of the LBDA were also tested including using the LBDA grid located nearest to each streamflow gauge, using LBDA grids surrounding each streamflow gauge within a given diameter, and principal components (PCs) (Jolliffe, 2002) and archetype analysis (Cutler and Breiman, 1994) of US-wide and Missouri River Basin region LBDA (see orange and green diamonds respectively in Figure 1). An annual temporal aggregation of streamflow was found to produce the strongest signal (Figure 2 showing correlation results between streamflow gauges in the Missouri River Basin and the nearest LBDA grid and PCs of US-wide LBDA). Diagnostic tests using the nearest LBDA grid were superior and reflects the ability of the point-by-point regression method used for the LBDA to preserve local climate details in the PDSI reconstructions that are also related to streamflow.

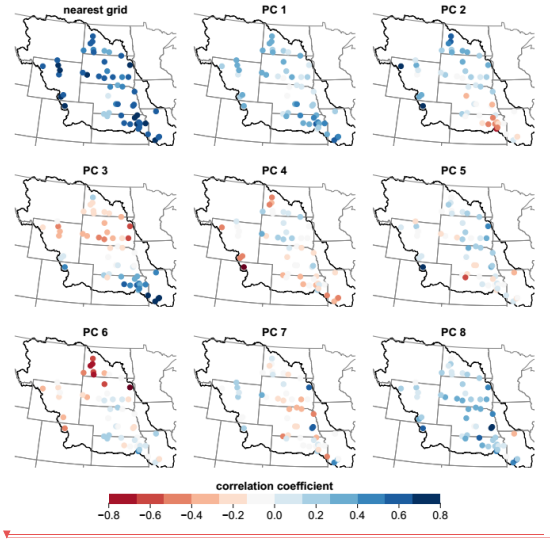


Figure 2. Pearson correlation between annual streamflow and a) closest LBDA grid b)-i) PC1 – PC8 of US-wide LBDA

High correlations were also noted for some streamflow stations with LBDA in the surrounding region (see supporting information) particularly for streamflow records with weaker correlations with the nearest grid (e.g. gauges in south-central North Dakota and on the border of Nebraska and Kansas). We

Michelle Ho 3/6/2016 4:01 PM

Deleted:

247 therefore considered information from LBDA grids within a 450 km radius consistent with the tree
248 chronology search radius used to form the LBDA. Furthermore, given that large-scale climate and
249 weather patterns influence local hydroclimatic conditions (Woodhouse et al., 2002), one needs to also
250 consider the relationships with these larger modes of variability.

251 Correlations between streamflow and the first eight PCs of LBDA grids across the US (LBDA locations
252 shown in Figure 1, correlations in Figure 2 and PC loading patterns shown in Figure 3) show that the
253 Missouri River Basin streamflow is correlated with PC1, a PC representing overall US LBDA variability.
254 However, these correlations are weak in comparison with correlations with the nearest LBDA grid. The
255 north/south loading pattern of PC 2 is reflected in the change in correlation sign for stations north and
256 south of the Nebraska and Kansas border. Similarly, the east-west difference in the loadings of PC3 leads
257 to strong positive correlations in the downstream reaches of the Basin in Kansas and Missouri and
258 negative correlations in Nebraska and Wyoming. The correlation results between streamflow and PCs of
259 CONUS LBDA suggest that the large-scale modes of variability could provide additional information for
260 modelling streamflow. Further details of the eight PCs of the CONUS LBDA and selection methods are
261 provided in the supporting information.

262 5 Modeling approach and performance metrics

263 Here, we present a suite of plausible methods that could be implemented to model streamflow using
264 the LBDA as the covariate(s) (Section 5.1). We justify the selection of a model that incorporates the use
265 of regularized canonical correlation (rCCA), which is further described in Section 5.2. A description of the
266 model performance metrics is provided in Section 5.3.

5.1 Model design and preliminary model assessment

We seek to use log-linear models to quantify the relationship between streamflow at individual gauges and a suite of site-specific LBDA records to facilitate a paleoclimate reconstruction of streamflow in the Missouri River Basin. Keeping in mind the application to streamflow record extension using the reconstructed PDSI values available in the LBDA we consider the following steps:

1. *Instrumental Period* (32-77 years at 55 locations): Use LBDA reconstructed PDSI to estimate the relationship between log transformed annual streamflow and a selection of LBDA records specific to the target streamflow site, namely $f(ln(Q_t)|\boldsymbol{PDSI}_t)$, where $\boldsymbol{PDSI_t}$ is really the expected value of PDSI informed by the tree-ring chronologies, $E[\boldsymbol{PDSI}_t|\boldsymbol{tree\text{-}ring\ chronology}_t]$, from 1929 to 2005 to maximize the overlap between the two sets of variables.

2. *Paleo Period (1473 to 1929):* Use $f(ln(Q_t)|\boldsymbol{PDSI}_t)$ estimated in the previous step with the LBDA reconstructed PDSI, $E[\boldsymbol{PDSI}_t|\boldsymbol{tree\text{-}ring\ chronology}_t]$, to estimate $ln(Q_t)$ prior to 1929 (estimates of $ln(Q_t)$ post 1929 will also be shown for comparison).

Several reconstruction model designs were considered given the initial diagnostic results. These included:

1) Developing one model for each streamflow station with predictors comprised of either:
    a. the LBDA grid located closest to the streamflow gauge;
    b. PCs of the LBDA from a region around the Missouri River Basin (Figure 1, green diamonds) within a multilinear model framework;
    c. canonical correlation analysis with regularization(rCCA described in Section 5.2) using LBDA grids within a 450km radius; or
    d. rCCA using LBDA grids within a 450km radius in addition to the retained PCs of CONUS-wide LBDA.

2) Developing one model for all streamflow stations with predictors comprised of either:
    a. PCs of both streamflow and CONUS-wide LBDA; or
    b. rCCA using streamflow and LBDA information from either the Missouri River Basin region or the CONUS region.

The fine resolution, gridded LBDA data results in a large number of highly correlated, potential predictors in all of the model designs considered with the exception of 1a. Given the 0.5° × 0.5° resolution of the LBDA data, if a reconstruction of streamflow at a single gauge is considered using only a 450 km radius of surrounding LBDA data, one has around 300 potential predictors including the leading eight CONUS LBDA PCs. The length of annual streamflow records in the Missouri River Basin ranges from 39 to 85 years, and thus it is clear that the number of potential predictors is significantly greater than the number of observations available to fit the model. Of course, if one were to consider a model with a simultaneous reconstruction of all the streamflow records, the dimension of the estimation problem becomes even more challenging. Consequently, this is the first issue considered in model development.

A single reconstruction model would be advantageous in its ability to consider all streamflow stations at once rather than fitting 55 individual models. However, the varying streamflow record lengths and differences in record periods would have resulted in large uncertainties as a large degree of annual streamflow imputation would have been required. As a result, model designs under option 2 were not further examined.

A cursory comparison of the viable models was conducted by comparing the coefficient of determination for each fitted model. The development of individual models for each streamflow gauge using either model 1a (the closest LBDA grid) or 1b (PCs of the LBDA in a region around the Missouri River Basin) resulted in acceptable models of streamflow (mean adjusted R^2 across the two models were

0.37 and 0.33 respectively). However, the spatially complete LBDA presents the opportunity to include a wider variety of local and large-scale information and these models (model designs 1c and 1d) were superior to the simpler models (model designs 1a and 1b). In addition, rCCA using both local and CONUS-wide information (model 1d) resulted in slightly improved model results (mean adjusted $R^2$of 0.74 across all individual station models) over using only local information (model 1c, mean adjusted R^2 of 0.71). We therefore selected model 1d (rCCA using LBDA grids within a 450km radius in addition to the retained PCs of CONUS-wide LBDA) for further analysis. A schematic of the data used to fit this model is shown in Figure 3, while a more detailed description of rCCA is provided in Section 5.2.

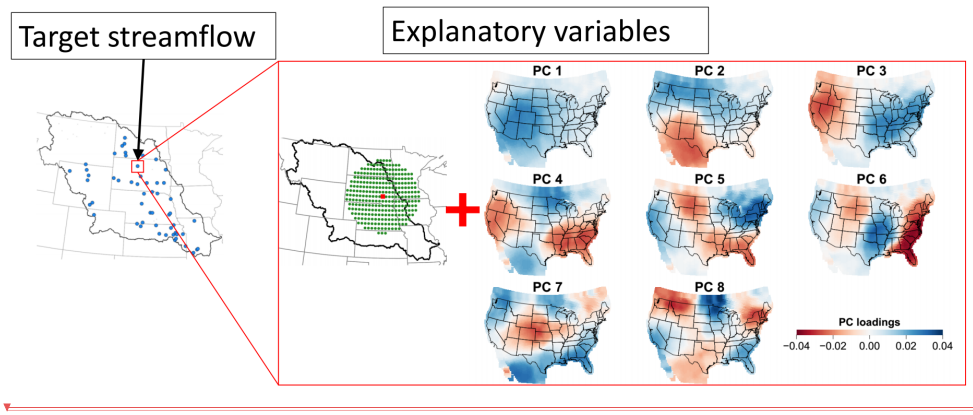
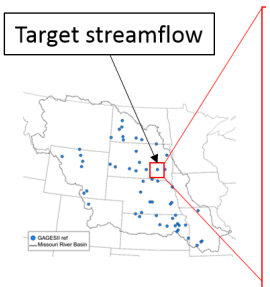


Figure 3. Schematic of LBDA information to be included in the reconstruction model of one streamflow station in the Missouri River Basin. The model inputs are the LBDA within a 450km radius and the first 8 PCs of US-wide LBDA.

5.2 Regularized canonical correlation analysis

Methods such as principal component analysis (PCA, Jolliffe, 2002), archetype analysis (AA, Cutler and Breiman, 1994; Stone and Cutler, 1996; Steinschneider and Lall, 2015) and canonical correlation analysis (CCA, Hotelling, 1936) are typically used for dimension reduction in this setting. Given that the number of potential predictors exceeds the number of observations, their high mutual correlation (>0.95 for many of the adjacent grids), and an interest in exploring a multivariate streamflow response, we explore the use of regularized canonical correlation where the regularization procedure is akin to ridge

Michelle Ho 3/6/2016 4:31 PM



Deleted:

regression (De Bie and De Moor, 2003). Regularized canonical correlation was selected for use here to capitalize on its ability to tailor the dimension reduction process to maximize the correlation between the explanatory and target variables (in contrast to PCA where the explanatory and target variable relationship is not considered).

Canonical correlation was introduced by Hotelling (1936) as a method of linearly transforming two vector variables to canonical form to maximize the correlation between them. Consider two sets of random variables represented by two matrices **X** and **Y**. **X** is a $n \times p$ matrix where $\mathbf{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p]$ containing n observations at p different locations, while **Y** is a $n \times q$ matrix where $\mathbf{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_q]$ containing n observations at q different locations. Both **X** and **Y** have finite variance matrices represented by $\boldsymbol{\Sigma}_{XX}$ and $\boldsymbol{\Sigma}_{YY}$ respectively. The covariance matrix between **X** and **Y** is $\boldsymbol{\Sigma}_{XY}$ while the covariance matrix between **Y** and **X** is $\boldsymbol{\Sigma}_{YX}$. In this application, **X** consists of the LBDA inputs specific to each streamflow gauge and **Y** is the streamflow at one station (i.e. p is large and q = 1). Canonical correlation analysis involves rotating the coordinate axes of both **X** and **Y** to new coordinate systems in order to clearly exhibit correlation between **X** and **Y**. An arbitrary linear combination could be $\mathbf{U}=\mathbf{X}\times\boldsymbol{\alpha}$ and $\mathbf{V}=\mathbf{Y}\times\boldsymbol{\gamma}$ such that the correlation between **U** and **V** is maximized. **U** and **V** yield the first pair of canonical variates, while $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ are the vectors of canonical weights of length p and q respectively. This process may be repeated subject to the constraint that following pairs of canonical variates are orthogonal to previous pairs with a maximum of min(p, q) pairs obtained. In this application, the model of streamflow is applied station by station and therefore only one pair of canonical variates are calculated and the first canonical variate of the LBDA is used to fit the model of streamflow.

The correlation of successive pairs of canonical variates can be found using an eigen decomposition of $\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{XY}\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\Sigma}_{YX}$ and $\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{XY}$. The resulting min(p, q) eigenvalues are common to both and the square root of the eigenvalues yield the canonical correlation. The eigenvectors of $\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{XY}\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\Sigma}_{YX}$ and $\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{XY}$ respectively yield the canonical weights $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ that are used to transform **X** and **Y**.

These weights are akin to the beta values in a multiple linear regression. Transforms using the i^{th} eigenvector result in correlations corresponding to the square root of the i^{th} eigenvalue. Here, the weight for **Y** (streamflow) is 1 and the weights for **X** (LBDA) are given by the first eigenvector of $\mathbf{\Sigma}_{YY}^{-1}\mathbf{\Sigma}_{YX}\mathbf{\Sigma}_{XX}^{-1}\mathbf{\Sigma}_{XY}$.

We employ regularization of the CCA process to address the issue of a large number of predictors relative to the number of observations (Vinod, 1976). Regularization is a smoothing process where a "roughness penalty", also known as the regularization parameter (λ), is introduced by converting $\mathbf{\Sigma}_{YY}$ and $\mathbf{\Sigma}_{XX}$ to $\mathbf{\Sigma}_{YY} + \lambda_y\mathbf{I}$ and $\mathbf{\Sigma}_{XX} + \lambda_x\mathbf{I}$ respectively (Leurgans et al., 1993) and is similar to the technique of ridge regression (De Bie and De Moor, 2003). Values of λ range between 0 and 1 with larger λ values indicating a higher degree of smoothing. Regularization also enables the process of matrix inversion to be stabilized. A suitable value of λ is determined using a leave one out cross validation score, where λ_x and λ_y are selected such that the correlation between the transformed datasets are maximized while the degrees of freedom used (as defined by Dijkstra, 2014) are limited to a maximum of n-10. If the criteria for the maximum degrees of freedom could not be achieved λ was set to one, otherwise λ was evaluated to two significant figures. No regularization was required for the single variable streamflow, whilst the LBDA was heavily regularized in almost all cases (Figure 4). 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/).

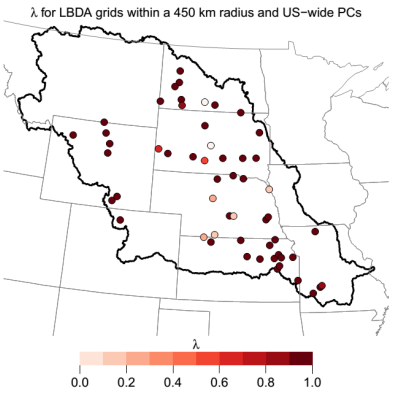


Figure 4. CCA regularization parameter values for the explanatory variables (LBDA grids within a 450km radius and CONUS-wide PCs) for each model of LBDA and streamflow.

5.3 Model performance metrics

The performance of the model selected for further analysis includes verification using a leave-k-out cross-validation procedure. Ten percent of the data (between 3 and 8 years out of a total of 32 and 77 years of streamflow data overlapping the LBDA record) are randomly selected and withheld from model fitting. These values are then predicted from the model fit to the balance of the data. The entire process is repeated 100 times, thus providing a set of 100 k-fold cross-validation samples. A comparison of the model residuals resulting from both calibrated and verified model inputs is made for the 100 cross-validation samples. The coefficient of efficiency (CE) and the reduction of error (RE) (Cook et al., 1994; Wilson et al., 2010) are additional metrics calculated to verify the model. Both CE and RE are similar to the Nash-Sutcliff efficiency test, however, the metrics are normalized using the mean of the verification and calibration data respectively. Namely, CE is defined as

$$CE = 1 - \frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)}{\sum_{i=1}^{n}(x_i - \bar{x}_v)}$$

while RE is defined as

$$RE = 1 - \frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)}{\sum_{i=1}^{n}(x_i - \bar{x}_c)}$$

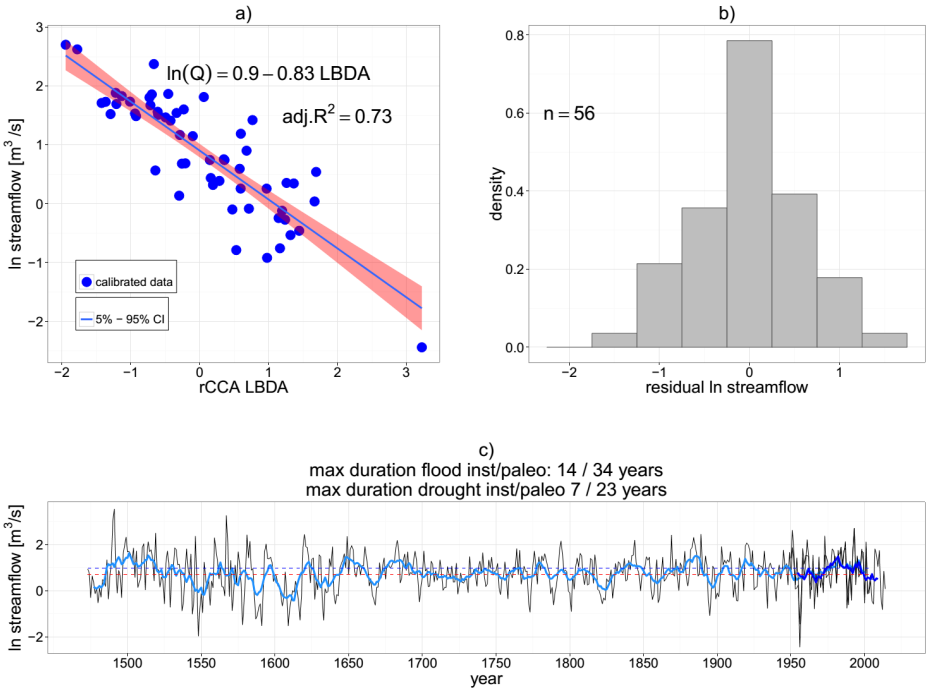
Where x_i and $\hat{x}_i$ are the observed and modelled streamflows respectively, while $\bar{x}_v$ and $\bar{x}_c$ are the means of the observed streamflow in the validation and calibration periods respectively.

6 Results

A log-linear model was individually fitted to each streamflow gauge using logarithmically transformed streamflow and the first canonical variate of the LBDA inputs (schematic of inputs shown in Figure 3) using rCCA. The models were tested using a cross-validation procedure and calibration and verification metrics, as described in Section 5.3 were calculated. Finally, an overall assessment of the dominant modes of temporal variability in reconstructed streamflow variability in the Missouri River Basin was made using a frequency wavelet analysis on the leading PCs of reconstructed streamflow.

6.1 Model results

Streamflow at each of the 55 stations in the Missouri River Basin was constructed using a least squares regression model of natural-log streamflow and the first regularized canonical variate of the combined LBDA grids within a 450 km radius and first eight PCs of US-wide LBDA as the predictor variable. All results for modeled and reconstructed streamflow are shown in log space, while verification statistics are also calculated for logarithmic streamflow. An example of the input and modeled streamflow is shown for one streamflow station calibrated using data from Turkey Creek near Seneca (Figure 5). A summary of modeled streamflow results and summary statistics of the residuals for both calibrated and verified streamflow across all 100 calibration and verification sets are provided in the supporting information.



409

410 *Figure 5. Model results for Turkey Creek near Seneca (USGS gauge number 06814000) calibrated using all available data and the*
411 *first canonical variate of the LBDA inputs a) input (dots) and modelled (line) natural log streamflow showing 5^{th} – 95^{th} prediction*
412 *interval, b) histogram of model residuals, and c) flood and drought persistence gauged by a threshold of mean±0.5SD of ten year*
413 *running average (blue lines – light blue line is the reconstructed ten year moving average).*

414 The rCCA linear model resulted in 55 models with adjusted R^2 values ranging between 0.56 and 0.90.

415 The model residuals were near normal at almost all stations with the exception of some stations with

416 small catchment regions showing near-uniform residuals. In each streamflow model, the magnitude of

417 rCCA loadings for the eight PCs of CONUS LBDA were similar to those of the ~300 local LBDA grids

418 indicating that information from the eight PCs of CONUS LBDA did not dominate the reconstruction.

419 6.2 Model validation

420 The predictive power of the model was assessed by calculating the coefficient of efficiency (CE) and

421 reduction of error (RE) for all cross validations. The median CE and RE values are shown in Figure 6 a and

b respectively, while distributions of the values are shown in box plots in the supporting information. CE and RE values range from -∞ to 1, with values over 0 indicating that the model predictions are more accurate than the respective climatology used for each statistic.

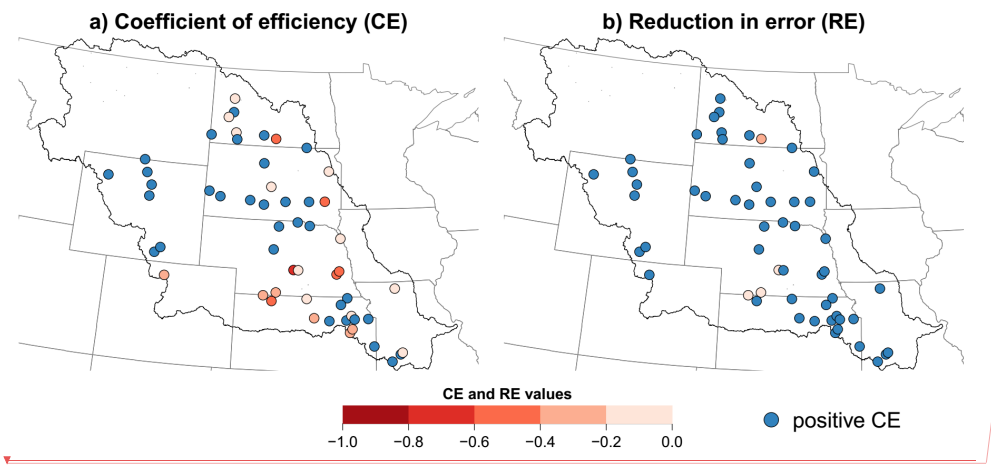
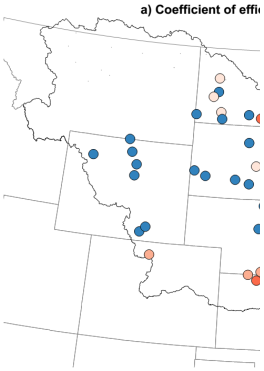


Figure 6. Median values of a) CE and b) RE values for cross validated models of the 55 Missouri River Basin streamflow stations.

The CE values in Figure 6 indicate that many of the streamflow models are able to provide median streamflow predictions with greater accuracy than the climatology of the withheld data. The RE values likewise suggest that the models are able to predict the withheld values with greater accuracy than the climatology of the calibrated values at most stations. The combined CE and RE results suggest there is some skill in reconstructing paleoclimate streamflow using the LBDA.

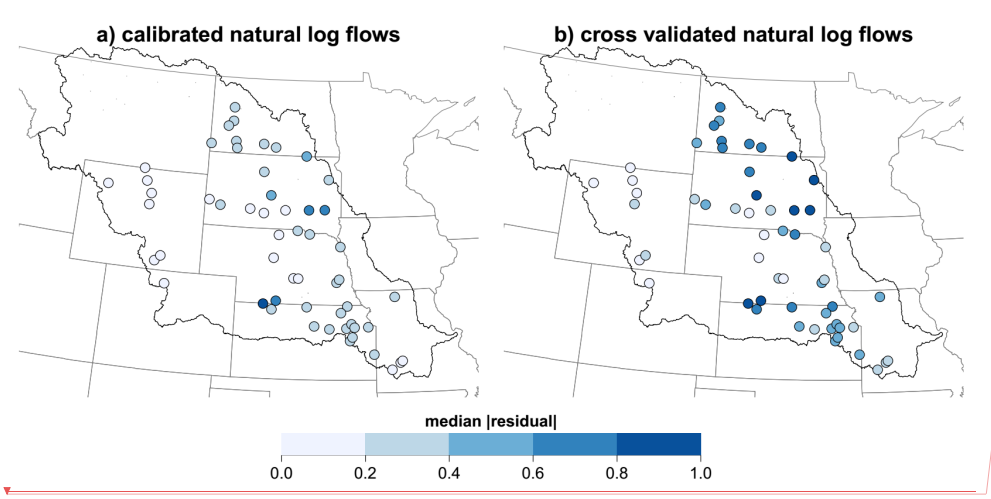
A comparison of the distribution of residuals resulting from calibrated values vs. validated values was also made using the 100 repetitions of the k-fold cross validation test. The median and the range of the cross-validated absolute residuals are modestly larger than that of the absolute residuals during the fitting process. The differences, accounting for the sample sizes, are not statistically significant in over 90% of 100 cross validated results (using a two-sided t-test and a null hypothesis that the true difference between the means is zero and alpha = 0.05). The median values of the cross validated residuals are

Michelle Ho 3/6/2016 4:45 PM



Deleted:

shown in Figure 7, while box plots showing the distribution of the cross validated residuals are shown in the supporting information. A summary of the reconstruction is presented in Section 6.3 using PCs of the streamflow reconstructed at the 55 stations.

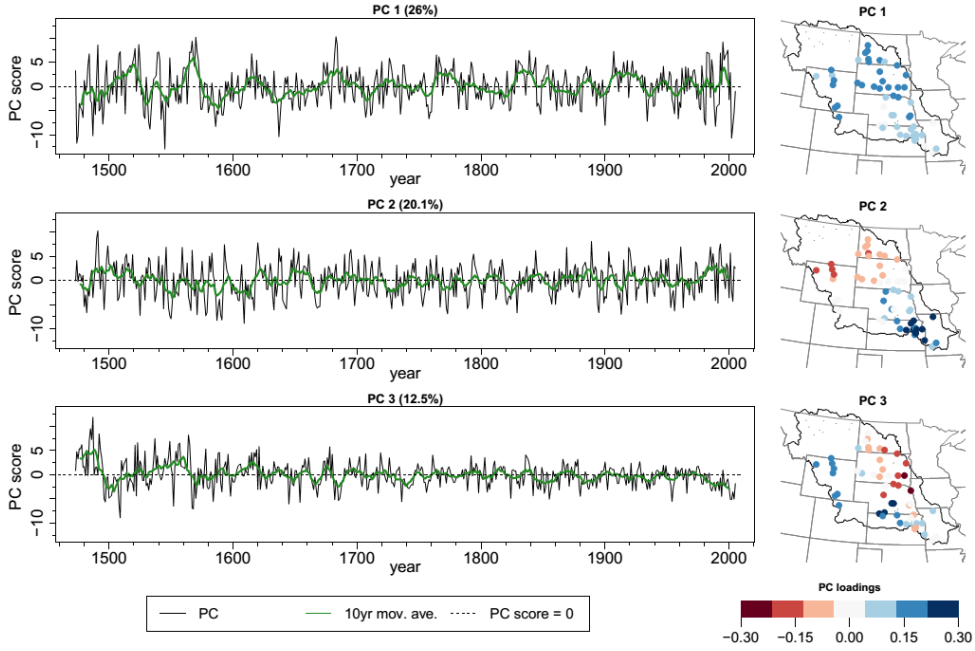


Michelle Ho 3/6/2016 4:46 PM
a) calibrated natura
Deleted:

Figure 7. Median cross validated model residuals (absolute value) for a) calibrated and b) verified natural log streamflow inputs at the 55 Missouri River Basin streamflow stations. Cross validations were repeated 100 times.

6.3 Reconstructed streamflow analysis

Mean annual streamflow was reconstructed for all 55 stations in the Missouri River Basin (data available in Ho et al., 2016). PCA was used to extract the leading modes of variability from the reconstructed Missouri River Basin natural log streamflow at all 55 stations. A combined assessment of the PCs using North's Rule of Thumb and a scree plot showed adjacent degenerate eigenvalues pairs for PCs of order 4 and higher and discontinuities at PC4 respectively (shown in supporting information). Furthermore, the spatial pattern of PC4 was difficult to interpret and therefore only the first three PCs are shown in Figure 8.



454

455 *Figure 8. Annual (black) and 10 year moving average (green line) of the first four PCs of reconstructed mean natural log*
456 *streamflow in the Missouri River Basin (left) with percentage variance explained shown in parenthesis and the corresponding*
457 *loading patterns (right).*

458 Negative streamflow anomalies coinciding with the 1930s dust bowl drought and 1950s drought are
459 evident in PC1, which has a positive loading pattern across all stations in the Missouri River Basin (Figure
460 8). The 1950s drought had more severe impacts in the southern half of the Missouri River Basin
461 (Piechota and Dracup, 1996; Cole et al., 2002; Andreadis et al., 2005; Cook et al., 2009) and this is
462 detected in PC2 that is comprised of positive loadings in the southern half of the Basin. The Civil War
463 drought, which largely impacted the Central Plains region from the mid-1850s to mid-1860s (Herweijer
464 et al., 2006), is also evident in PC2.

465 The severity of these droughts had wide ranging impacts on ecological states, agricultural produce and
466 social activities and are often used as benchmark droughts to which contemporary droughts are
467 compared (Breshears et al., 2005; Hornbeck, 2009). However, the streamflow reconstructions suggest

468 periods where streamflow deficits may have been more severe than any of these historical droughts. For
469 example, the late 1540s, 1590s, and late 1750s all show negative streamflow anomalies in both PC1 and
470 PC2 (26% and 20.1% variability explained respectively) that are of similar or greater magnitude than the
471 streamflow deficits associated with the Civil War, Dust Bowl or 1950s droughts. Drought durations were
472 also longer when assessed in the context of streamflow variability over the past 500 years. A threshold
473 of above or below 0.5 standard deviation in decadal streamflow at each station was used as an
474 approximation of flood and drought regimes respectively. The maximum duration of continuous periods
475 of either flood or drought regimes were found to be longer in the majority of stations investigated as
476 demonstrated in Figure 5 c that shows the reconstructed annual and decadal streamflow at Turkey
477 Creek and the longest duration drought and flood regime for the station.

478 PC3 (12.5% variability explained) is a mode of variability largely influenced by differences in streamflow
479 variability between the Rocky Mountains regions and the remainder of the Missouri River Basin. The
480 apparent increase in variability in the 1500s and 1600s in PC3 suggests that differences in streamflow
481 variability between the Rocky Mountains regions and the plains region may have been more
482 pronounced during these two centuries.

483 PC2 of the reconstructed streamflow has positive loadings in the southern Missouri River Basin. This PC
484 shows an extended period of below average flows, with the exception of a couple of years, in the first
485 two decades of the 1600s and is evident in individual streamflow reconstructions around the border of
486 Nebraska and South Dakota and in Kansas. A similar persistence of decadal-scale low flows are not
487 featured in the instrumental record, suggesting that successive years of low streamflow have persisted
488 for longer periods than what has been observed in historical records.

489 The PCs of the reconstructed streamflow also show several periods where annual flows are larger than
490 the maximum instrumental streamflow. Periods of anomalously high streamflow include the 1560s for

PC 1 and the early 1490s for PC 2 and 3. These periods coincide with high positive reconstructed PDSI values particularly in the lower Missouri Basin region and along the middle and southern Rocky Mountains in the early 1490s and positive PDSI values across most of the interior plains. Anomalously large streamflow during the 1560s were reconstructed at stream gauges around the interior plains regions from tributaries that join the Missouri River in North and South Dakota (e.g. Cannonball, Grand and White Rivers). However, anomalously high streamflow around the 1560s was not reconstructed for either the Platte or Yellowstone rivers suggesting that the increase in streamflow may have resulted from weather systems delivering moisture in the upper Interior Plains (North and South Dakota), but not in the Rocky Mountains region.

In order to quantitatively assess the temporal modes of variability implied by the PCs, a continuous wavelet transform was applied to identify key periodicities in the main modes of reconstructed streamflow variability. Wavelet transforms for PCs 1-3 along with their global wavelet spectra are shown in Figure 9. The wavelet transforms were implemented using the Morlet wavelet function as the mother wavelet and padded with zeros to limit the edge effects of the wavelet analysis (Torrence and Compo, 1998).

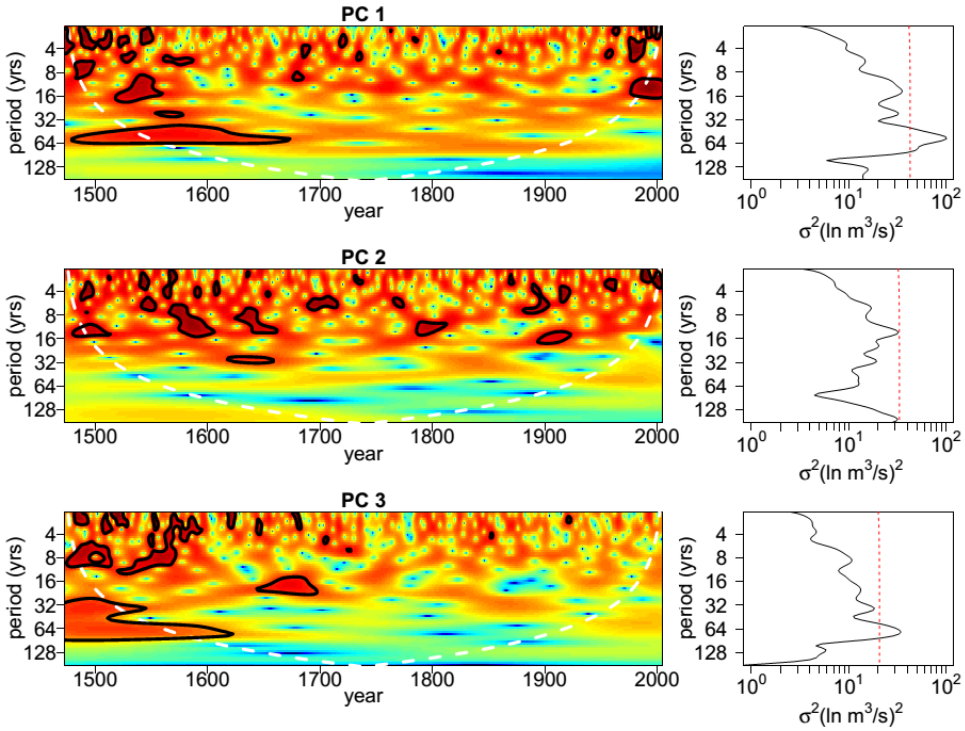


Figure 9. Wavelet transforms of PC 1-3 of reconstructed log streamflow in the Missouri River Basin from 1473-2005 (left), black lines show 95% significance level, dashed white line shows cone of influence under which edge effects may influence the results and the time averaged global wavelet spectra (right) with red dashed line showing 95% confidence level for the wavelet transform using a white noise background spectrum.

All PCs show decadal-scale variability in the 1500s and 1600s, which is significant at the 95% level against white noise for PC 1 and 3. Interestingly, none of the wavelet spectra show a consistent mode of variability that coincide with an El Niño Southern Oscillation frequency (2-7 years Allan, 2000). In addition, previously identified periods of decadal-scale variability in paleoclimate reconstructions of the Pacific Decadal Oscillation (Gedalof and Smith, 2001; MacDonald and Case, 2005) are not reproduced in the streamflow reconstruction. The multi-decadal-scale variability seen in PC1 and PC3 around the 16th century suggest that the persistence of drought and pluvial events was longer around this period compared with historic records and is reflected in previous reconstructions of persistent drought during

this period (Woodhouse and Overpeck, 1998) and in nearby regions (Meko et al., 1995). A similar pattern of multi-decadal-scale variability is also seen in the PCs of LBDA grids within the Missouri River Basin around the 16th century (results not shown here).

7 Discussion and Conclusions

Human activities and social processes are irrefutably intertwined with the hydrological cycle with each influencing and affecting the behavior of the other (Matalas et al., 1982). Consequently, hydrologic investigations need to occur at spatial scales that encompass the social and political regions influencing how water is stored, used, and transported. This requirement necessitates a broader spatial scale to be considered beyond the watershed scale typically addressed in hydrologic assessments (Vogel et al., 2015). Spatially and temporally broad measures of hydroclimatic variability such as the LBDA provide opportunities to conduct such assessments.

We used the LBDA to demonstrate a method of reconstructing mean annual streamflow in the Missouri River Basin from 1473-2005 in order to facilitate a future reconstruction of streamflow across the CONUS. The rCCA approach enabled a large degree of spatially coherent information in the LBDA to be condensed via linear transforms into a canonical variate that explained the majority of information in the target streamflow. The additional regularization step addressed the issue of an ill-posed problem where the number of variables far exceeded the number of observations. Regularization addressed potential issues with overfitting of the model and subsequent poor predictability of uncalibrated values. This site-by-site approach, opposed to a multi-site model, avoided data issues where streamflow records were not concurrently available at all sites. The method also tailored the LBDA information included in the model to each unique site and therefore has the potential to be applied to any individual streamflow site in the United States and, potentially, any streamflow site in North America where there are useful reconstructions in the LBDA.

The analysis of streamflow variability in the Missouri River Basin over the past 500 years was, however, limited by the absence of streamflow gauges in the headwaters of the Missouri River. Based on the selection criteria used here, there are no gauges located upstream of Fort Peck Lake and only one gauge in Montana was included in the analysis. The dearth of gauges in this region therefore omits potentially critical information when drawing conclusions regarding regional streamflow variability as presented in Section 6.3. Future analyses could seek to include additional streamflow gauges, potentially drawing on non-references gauges from the GAGESII database and other sources. While non-reference GAGESII gauges are located within disturbed catchments with potentially regulated streams, it may be plausible that these measurements would still preserve the signatures of interannual streamflow variability which we seek to recover. Encouragingly, the headwater regions of the Missouri River Basin contain a relatively dense network of tree-ring chronologies that would improve the fidelity of both PDSI and subsequent streamflow reconstructions. Furthermore, efforts towards expanding this tree-ring network are currently been made with a view to target and improve Missouri River Basin streamflow reconstructions (Pederson, 2013).

Dendrochronological studies for the purposes of climate reconstruction typically use a targeted sampling method whereby old living trees are selected for sampling to maximize the length of reconstruction and the location of sampling is chosen to ensure that tree growth is largely limited by the climatic factors of interest for reconstruction (Speer, 2010). Over 50 different tree species were used in the formulation of the LBDA, however, species was not a determinant for inclusion in the LBDA. Rather, the tree-ring chronologies specifically selected for use in reconstruction were those that were best correlated with available soil moisture as modeled a priori by PDSI and this was done using a method of point-by-point regression that ensures that the selected chronologies were likely to have a causal association with the gridded instrumental PDSI being reconstructed (Cook et al., 1999).

In developing our approach to reconstructing streamflow from the tree-ring based LBDA, we have capitalized on the fact that both variables 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 remove noise irrelevant to hydrological variability rendering this paleoclimate reconstruction of PDSI particularly suitable for informing streamflow. Furthermore, streamflow and PDSI are inherently different hydrological variables as demonstrated in the temporal and spatial statistics shown in the supporting information. Streamflow is also a derivative of a suite of non-climatic variables such as catchment size, land-use, geology, topography, and groundwater interactions. These catchment-specific attributes are not explicitly considered in our study (with the exception of land use considered through the selection of streamflow gauges from relatively undisturbed catchments) and the inclusion of attributes could provide an improvement to the model (e.g. Thomas and Benson, 1970; Lima and Lall, 2010).

Assessments of hydrological variability on temporal scales broader than traditional hydrological practices are paramount to identifying long-term variations and quantifying the nature of persistent regimes, which are unachievable using comparatively short instrumental record. The understanding of paleoclimate variability facilitates the development of water and land use practices that are appropriate, sustainable, and resilient under previous patterns of hydroclimatic variability and complement efforts towards developing suitable adaption procedures for projected future climate scenarios. Whilst paleoclimate reconstructions of annual streamflow are useful for assessing long-term variability, annual information is often too coarse for water resource applications such as reservoir management and seasonal water allocations. In order to produce relevant data on a monthly or daily time step, a future undertaking could involve temporal disaggregation of the data. One method that could be utilized is a non-parametric k-nearest neighbors approach (Lall and Sharma, 1996; Rajagopalan and Lall, 1999;

590 Nowak et al., 2010) that would avoid some issues associated with multi-site stochastic generation
591 schemes (Valencia and Schaake, 1973; Segond et al., 2006).The objective of producing paleoclimate
592 reconstructions of streamflow that are relevant for water resource management also suggests that the
593 reconstruction of streamflow that is currently regulated would be valuable. Paleoclimate
594 reconstructions of regulated streamflow could be achieved at an annual scale provided that regulations
595 primarily impact the timing of sub-annual flows or if sufficient data prior to regularization exists. The
596 resulting annual streamflow reconstruction could be disaggregated to a shorter timescale if pre-
597 regulated flow data is available.

598 This study has capitalized on the availability of the spatially and temporally complete LBDA dataset to
599 attempt a reconstruction of streamflow on a broad spatial scale across the Missouri River Basin. The
600 analysis employed here provides a feasible foundation from which streamflow reconstructions across
601 the CONUS and North America could be implemented using the LBDA. In addition, the methodology
602 presented here could be applied to reconstructing streamflow using other reconstructions of drought
603 including the Monsoon Asia Drought Atlas (Cook et al., 2010b), the Old World Drought Atlas covering
604 Europe, North Africa, and the Middle East (Cook et al., 2015b) and the Australia and New Zealand
605 summer drought atlas (Palmer et al., 2015).
606

Michelle Ho 3/6/2016 6:27 PM
Deleted: similar

Michelle Ho 3/6/2016 6:28 PM
Deleted: reconstructions could be attempted

Michelle Ho 3/6/2016 6:28 PM
Deleted: continental-scale

Michelle Ho 3/6/2016 6:28 PM
Deleted: s

Michelle Ho 3/6/2016 6:29 PM
Deleted: and

Michelle Ho 3/6/2016 6:29 PM
Deleted: recently completed

Acknowledgments

We would like to sincerely thank Ben Cook (bc9z@ldeo.columbia.edu) for providing the LBDA data, the US Geological Survey (USGS) for making their monthly streamflow data available (http://waterdata.usgs.gov/nwis/monthly?referred_module=sw), staff at the USGS (Christopher Ryan cmryan@usgs.gov and Thomas Weaver tlweaver@usgs.gov) for clarifying streamflow data availability; Siyan Wang for initial data collection; Scott Steinschneider, Xun Sun and Pierre Gentine at the Columbia Water Center for their invaluable discussions on this work; and Juan A. Ballesteros and one other anonymous reviewer for their constructive comments. Reconstruction results are available on the NOAA paleoclimate data base (https://www.ncdc.noaa.gov/paleo/study/19520). This work is funded by an NSF award 1360446 and NSF award 1401698. Lamont-Doherty contribution number XXXX.

Michelle Ho 4/6/2016 7:48 AM
Deleted: and

References

Allan, R. J., 2000: ENSO and Climatic Variability in the Past 150 Years. *El Niño and the Southern Oscillation: Multiscale Variability and Global and Regional Impacts*, Henry F. Diaz and Vera Markgraf, Eds., Cambridge University Press, pp.

Allen, E. B., Rittenour, T. M., DeRose, R. J., Bekker, M. F., Kjelgren, R. and Buckley, B. M., 2013: A tree-ring based reconstruction of Logan River streamflow, northern Utah, *Water Resources Research*, **49** (12), 8579-8588, doi: 10.1002/2013wr014273.

Andreadis, K. M., Clark, E. A., Wood, A. W., Hamlet, A. F. and Lettenmaier, D. P., 2005: Twentieth-Century Drought in the Conterminous United States, *Journal of Hydrometeorology*, **6** (6), 985-1001, doi: 10.1175/JHM450.1.

Ault, T. R., Cole, J. E., Overpeck, J. T., Pederson, G. T. and Meko, D. M., 2014: Assessing the risk of persistent drought using climate model simulations and paleoclimate data, *Journal of Climate*, **27** (20), 7529-7549, doi: 10.1175/jcli-d-12-00282.1.

Breshears, D. D., Cobb, N. S., Rich, P. M., Price, K. P., Allen, C. D., Balice, R. G., Romme, W. H., Kastens, J. H., Floyd, M. L., Belnap, J., Anderson, J. J., Myers, O. B. and Meyer, C. W., 2005: Regional vegetation die-off in response to global-change-type drought, *Proceedings of the National Academy of Sciences of the United States of America*, **102** (42), 15144-15148, doi: 10.1073/pnas.0505734102.

Brown, D. P. and Comrie, A. C., 2004: A winter precipitation 'dipole' in the western United States associated with multidecadal ENSO variability, *Geophysical Research Letters*, **31** (9), doi: 10.1029/2003GL018726.

Buuren, S. v. and Groothuis-Oudshoorn, K., 2011: mice: Multivariate Imputation by Chained Equations in R, *Journal of Statistical Software*, **45** (3), 67.

Cole, J. E., Overpeck, J. T. and Cook, E. R., 2002: Multiyear La Niña events and persistent drought in the contiguous United States, *Geophysical Research Letters*, **29** (13), 25-1-25-4, doi: 10.1029/2001GL013561.

Cook, B. I., Miller, R. L. and Seager, R., 2009: Amplification of the North American "Dust Bowl" drought through human-induced land degradation, *Proceedings of the National Academy of Sciences*, **106** (13), 4997-5001, doi: 10.1073/pnas.0810200106.

Cook, B. I., Ault, T. R. and Smerdon, J. E., 2015a: Unprecedented 21st century drought risk in the American Southwest and Central Plains, *Science Advances*, **1** (1), doi: 10.1126/sciadv.1400082.

Cook, B. I., Smerdon, J. E., Seager, R. and Cook, E. R., 2013a: Pan-Continental Droughts in North America over the Last Millennium, *Journal of Climate*, **27** (1), 383-397, doi: 10.1175/jcli-d-13-00100.1.

Cook, E. R. and Krusic, P. J., 2004:The North American Drought Atlas,

Cook, E. R., Briffa, K. R. and Jones, P. D., 1994: Spatial regression methods in dendroclimatology: A review and comparison of two techniques, *International Journal of Climatology*, **14** (4), 379-402, doi: 10.1002/joc.3370140404.

Cook, E. R., Meko, D. M., Stahle, D. W. and Cleaveland, M. K., 1999: Drought Reconstructions for the Continental United States, *Journal of Climate*, **12** (4), 1145-1162, doi: 10.1175/1520-0442(1999)012<1145:drftcu>2.0.co;2.

Cook, E. R., Woodhouse, C. A., Eakin, C. M., Meko, D. M. and Stahle, D. W., 2004: Long-Term Aridity Changes in the Western United States, *Science*, **306** (5698), 1015-1018, doi: 10.1126/science.1102586.

Cook, E. R., Seager, R., Heim, R. R., Vose, R. S., Herweijer, C. and Woodhouse, C., 2010a: Megadroughts in North America: placing IPCC projections of hydroclimatic change in a long-term palaeoclimate context, *Journal of Quaternary Science*, **25** (1), 48-61, doi: 10.1002/jqs.1303.

Cook, E. R., Anchukaitis, K. J., Buckley, B. M., D'Arrigo, R. D., Jacoby, G. C. and Wright, W. E., 2010b: Asian Monsoon Failure and Megadrought During the Last Millennium, *Science*, **328** (5977), 486-489, doi: 10.1126/science.1185188.

Cook, E. R., Palmer, J. G., Ahmed, M., Woodhouse, C. A., Fenwick, P., Zafar, M. U., Wahab, M. and Khan, N., 2013b: Five centuries of Upper Indus River flow from tree rings, *Journal of Hydrology*, **486** (0), 365-375, doi: 10.1016/j.jhydrol.2013.02.004.

Cook, E. R., Seager, R., Kushnir, Y., Briffa, K. R., Büntgen, U., Frank, D., Krusic, P. J., Tegel, W., van der Schrier, G., Andreu-Hayles, L., Baillie, M., Baittinger, C., Bleicher, N., Bonde, N., Brown, D., Carrer, M., Cooper, R., Čufar, K., Dittmar, C., Esper, J., Griggs, C., Gunnarson, B., Günther, B., Gutierrez, E., Haneca, K., Helama, S., Herzig, F., Heussner, K.-U., Hofmann, J., Janda, P., Kontic, R., Köse, N., Kyncl, T., Levanič, T., Linderholm, H., Manning, S., Melvin, T. M., Miles, D., Neuwirth, B., Nicolussi, K., Nola, P., Panayotov, M., Popa, I., Rothe, A., Seftigen, K., Seim, A., Svarva, H., Svoboda, M., Thun, T., Timonen, M., Touchan, R., Trotsiuk, V., Trouet, V., Walder, F., Ważny, T., Wilson, R. and Zang, C., 2015b: Old World megadroughts and pluvials during the Common Era, *Science Advances*, **1** (10), doi: 10.1126/sciadv.1500561.

Cutler, A. and Breiman, L., 1994: Archetypal Analysis, *Technometrics*, **36** (4), 338-347, doi: 10.1080/00401706.1994.10485840.

Dai, A., 2013: Increasing drought under global warming in observations and models, *Nature Clim. Change*, **3** (1), 52-58, doi: 10.1038/nclimate1633.

De Bie, T. and De Moor, B., 2003: On the regularization of canonical correlation analysis, *Int. Sympos. ICA and BSS*, 785-790.

Dettinger, M. D., Cayan, D. R., Diaz, H. F. and Meko, D. M., 1998: North–South Precipitation Patterns in Western North America on Interannual-to-Decadal Timescales, *Journal of Climate*, **11** (12), 3095-3111, doi: 10.1175/1520-0442(1998)011<3095:NSPPIW>2.0.CO;2.

695 Devineni, N., Lall, U., Pederson, N. and Cook, E., 2013: A Tree-Ring-Based Reconstruction of Delaware
696 River Basin Streamflow Using Hierarchical Bayesian Regression, *Journal of Climate*, **26** (12),
697 4357-4374, doi: 10.1175/jcli-d-11-00675.1.
698 Dijkstra, T., 2014: Ridge regression and its degrees of freedom, *Quality & Quantity*, **48** (6), 3185-3193,
699 doi: 10.1007/s11135-013-9949-7.
700 Driscoll, D., 2013: *Hydroclimatic information and risk assessment methods for managing hydrologic*
701 *extremes in the Missouri River Basin - A white paper for future studies*. *Hydroclimatic extremes in*
702 *the Missouri River Basin*, USGS.
703 Fritts, H. C., 1976: *Tree Rings and Climate*. Academic Press Inc., 567 pp.
704 Galat, D. L., Berry Jr, C. R., Peters, E. J. and White, R. G., 2005: Chapter 10 - Missouri River Basin. *Rivers of*
705 *North America*, Arthur C. Benke and Colbert E. Cushing, Eds., Academic Press, pp. 426-480.
706 Gallant, A. J. E. and Gergis, J., 2011: An experimental streamflow reconstruction for the River Murray,
707 Australia, 1783-1988, *Water Resources Research*, **47** (12), W00G04, doi:
708 10.1029/2010wr009832.
709 Gedalof, Z. e. and Smith, D. J., 2001: Interdecadal climate variability and regime-scale shifts in Pacific
710 North America, *Geophys. Res. Lett.*, **28** (8), 1515-1518, doi: 10.1029/2000gl011779.
711 Gershunov, A. and Barnett, T. P., 1998: Interdecadal Modulation of ENSO Teleconnections, *Bulletin of*
712 *the American Meteorological Society*, **79** (12), 2715-2725, doi: 10.1175/1520-
713 0477(1998)079<2715:IMOET>2.0.CO;2.
714 González, I., Déjean, S., Martin, P. G. P. and Baccini, A., 2008: CCA: An R Package to Extend Canonical
715 Correlation Analysis, *2008*, **23** (12), 14, doi: 10.18637/jss.v023.i12.
716 Herweijer, C., Seager, R. and Cook, E. R., 2006: North American droughts of the mid to late nineteenth
717 century: a history, simulation and implication for Mediaeval drought, *The Holocene*, **16** (2), 159-
718 171, doi: 10.1191/0959683606hl917rp.
719 Herweijer, C., Seager, R., Cook, E. R. and Emile-Geay, J., 2007: North American Droughts of the Last
720 Millennium from a Gridded Network of Tree-Ring Data, *Journal of Climate*, **20** (7), 1353-1376,
721 doi: 10.1175/JCLI4042.1.
722 Higgins, R. W., Yao, Y., Yarosh, E. S., Janowiak, J. E. and Mo, K. C., 1997: Influence of the Great Plains
723 Low-Level Jet on Summertime Precipitation and Moisture Transport over the Central United
724 States, *Journal of Climate*, **10** (3), 481-507, doi: 10.1175/1520-
725 0442(1997)010<0481:IOTGPL>2.0.CO;2.
726 Ho, M., Kiem, A. S. and Verdon, D. C., 2015: A paleoclimate rainfall reconstruction in the Murray-Darling
727 Basin (MDB), Australia: 2. Assessing hydroclimatic risk using paleoclimate records of wet and dry
728 epochs, *Water Resour. Res.*, **51** (10), doi: 10.1002/2015WR017059.
729 Ho, M., Lall, U. and Cook, E. R., 2016:Missouri River Basin 533 Year Annual Streamflow
730 Reconstructions,<https://www.ncdc.noaa.gov/paleo/study/19520>.
731 Hornbeck, R., 2009: The Enduring Impact of the American Dust Bowl: Short and Long-run Adjustments to
732 Environmental Catastrophe, *National Bureau of Economic Research Working Paper Series*, **No.**
733 **15605**, doi: 10.3386/w15605.
734 Hotelling, H., 1936: Relations between two sets of variates, *Biometrika*, **28** (3-4), 321-377, doi:
735 10.1093/biomet/28.3-4.321.
736 Jolliffe, I. T., 2002: *Principal Component Analysis*. Second edition ed. Springer-Verlag, 487 pp.
737 Jones, P. D. and Mann, M. E., 2004: Climate over past millennia, *Rev. Geophys.*, **42** (2), RG2002, doi:
738 10.1029/2003rg000143.
739 Kunkel, K. E., Stevens, L. E., Stevens, S. E., Sun, L., Janssen, E., Wuebbles, D., Kruk, M. C., Thomas, D. P.,
740 Shulski, M., Umphlett, N., Hubbard, K., Robbins, K., Romolo, L., Akyuz, A., Pathak, T., Bergantino,
741 T. and Dobson, J. G., 2013: Regional Climate Trends and Scenarios for the U.S. National Climate
742 Assessment. Part 4. Climate of the U.S. Great Plains, NOAA Technical Report, NOAA, pp. 82.

743 Lall, U. and Sharma, A., 1996: A Nearest Neighbor Bootstrap For Resampling Hydrologic Time Series,
744 *Water Resources Research*, **32** (3), 679-693, doi: 10.1029/95WR02966.
745 Leurgans, S. E., Moyeed, R. A. and Silverman, B. W., 1993: Canonical Correlation Analysis when the Data
746 are Curves, *Journal of the Royal Statistical Society. Series B (Methodological)*, **55** (3), 725-740.
747 Lima, C. H. R. and Lall, U., 2010: Spatial scaling in a changing climate: A hierarchical bayesian model for
748 non-stationary multi-site annual maximum and monthly streamflow, *Journal of Hydrology*, **383**
749 (3–4), 307-318, doi: 10.1016/j.jhydrol.2009.12.045.
750 Lockwood, J. G., 1999: Is Potential Evapotranspiration and Its Relationship with Actual
751 Evapotranspiration Sensitive to Elevated Atmospheric CO2 Levels?, *Climatic Change*, **41** (2), 193-
752 212, doi: 10.1023/A:1005469416067.
753 MacDonald, G. M. and Case, R. A., 2005: Variations in the Pacific Decadal Oscillation over the past
754 millennium, *Geophys. Res. Lett.*, **32** (8), L08703, doi: 10.1029/2005gl022478.
755 Matalas, N. C., Landwehr, J. M. and Wolman, M. G., 1982: Prediction in Water Management. *Scientific*
756 *basis of water management*, National Academy, pp. 118-122.
757 McGowan, H. A., Marx, S. K., Denholm, J., Soderholm, J. and Kamber, B. S., 2009: Reconstructing annual
758 inflows to the headwater catchments of the Murray River, Australia, using the Pacific Decadal
759 Oscillation, *Geophys. Res. Lett.*, **36** (6), L06707, doi: 10.1029/2008gl037049.
760 Meko, D., Stockton, C. W. and Boggess, W. R., 1995: The tree-ring record of sever sustained drought,
761 *JAWRA Journal of the American Water Resources Association*, **31** (5), 789-801, doi:
762 10.1111/j.1752-1688.1995.tb03401.x.
763 Monteith, J., 1965: Evaporation and environment, *Symp. Soc. Exp. Biol*, 4.
764 Nohara, D., Kitoh, A., Hosaka, M. and Oki, T., 2006: Impact of Climate Change on River Discharge
765 Projected by Multimodel Ensemble, *Journal of Hydrometeorology*, **7** (5), 1076-1089, doi:
766 10.1175/JHM531.1.
767 Nowak, K., Prairie, J., Rajagopalan, B. and Lall, U., 2010: A nonparametric stochastic approach for
768 multisite disaggregation of annual to daily streamflow, *Water Resources Research*, **46** (8),
769 W08529, doi: 10.1029/2009WR008530.
770 Palmer, J. G., Cook, E. R., Turney, C. S. M., Allen, K., Fenwick, P., Cook, B. I., O'Donnell, A., Lough, J.,
771 Grierson, P. and Baker, P., 2015: Drought variability in the eastern Australia and New Zealand
772 summer drought atlas (ANZDA, CE 1500–2012) modulated by the Interdecadal Pacific
773 Oscillation, *Environmental Research Letters*, **10** (12), 124002, doi: 10.1088/1748-
774 9326/10/12/124002.
775 Pederson, G. T., 2013, Multi-century perspectives on current and future streamflow in the Missouri River
776 Basin, viewed 2015/09/08, <
777 Pederson, N., Bell, A. R., Cook, E. R., Lall, U., Devineni, N., Seager, R., Eggleston, K. and Vranes, K. P.,
778 2012: Is an Epic Pluvial Masking the Water Insecurity of the Greater New York City Region?,
779 *Journal of Climate*, **26** (4), 1339-1354, doi: 10.1175/jcli-d-11-00723.1.
780 Piechota, T. C. and Dracup, J. A., 1996: Drought and Regional Hydrologic Variation in the United States:
781 Associations with the El Niño-Southern Oscillation, *Water Resources Research*, **32** (5), 1359-
782 1373, doi: 10.1029/96WR00353.
783 Pizarro, G. and Lall, U., 2002: El Niño-induced Flooding in the U.S. West: What Can We Expect, *Eos Trans.*
784 *AGU*, **82** (32), 349-352, doi: 10.1029/2002EO000255.
785 Prairie, J., Nowak, K., Rajagopalan, B., Lall, U. and Fulp, T., 2008: A stochastic nonparametric approach
786 for streamflow generation combining observational and paleoreconstructed data, *Water*
787 *Resources Research*, **44** (6), W06423, doi: 10.1029/2007wr006684.
788 Quinn, W. H., 1992: A study of the Southern Oscillation-related climatic activity for A.D 622-1990
789 incorporating Nile River flood data. *El Niño: Historical and Paleoclimatic Aspects of the Southern*
790 *Ocean*, Henry F. Diaz and Vera Markgraf, Eds., Cambridge University Press, pp. 119-149.

Rajagopalan, B. and Lall, U., 1999: A k-nearest-neighbor simulator for daily precipitation and other weather variables, *Water Resources Research*, **35** (10), 3089-3101, doi: 10.1029/1999WR900028.

Redmond, K. T. and Koch, R. W., 1991: Surface Climate and Streamflow Variability in the Western United States and Their Relationship to Large-Scale Circulation Indices, *Water Resources Research*, **27** (9), 2381-2399, doi: 10.1029/91WR00690.

Routson, C. C., Woodhouse, C. A. and Overpeck, J. T., 2011: Second century megadrought in the Rio Grande headwaters, Colorado: How unusual was medieval drought?, *Geophys. Res. Lett.*, **38** (22), L22703, doi: 10.1029/2011gl050015.

Seager, R., Ting, M., Held, I., Kushnir, Y., Lu, J., Vecchi, G., Huang, H.-P., Harnik, N., Leetmaa, A., Lau, N.-C., Li, C., Velez, J. and Naik, N., 2007: Model Projections of an Imminent Transition to a More Arid Climate in Southwestern North America, *Science*, **316** (5828), 1181-1184, doi: 10.1126/science.1139601.

Segond, M. L., Onof, C. and Wheater, H. S., 2006: Spatial–temporal disaggregation of daily rainfall from a generalized linear model, *Journal of Hydrology*, **331** (3–4), 674-689, doi: 10.1016/j.jhydrol.2006.06.019.

Sheffield, J., Wood, E. F. and Roderick, M. L., 2012: Little change in global drought over the past 60 years, *Nature*, **491** (7424), 435-438, doi: 10.1038/nature11575.

Smerdon, J. E., Cook, B. I., Cook, E. R. and Seager, R., 2015: Bridging Past and Future Climate across Paleoclimatic Reconstructions, Observations, and Models: A Hydroclimate Case Study, *Journal of Climate*, **28** (8), 3212-3231, doi: 10.1175/JCLI-D-14-00417.1.

Smith, S. R., Green, P. M., Leonardi, A. P. and O'Brien, J. J., 1998: Role of Multiple-Level Tropospheric Circulations in Forcing ENSO Winter Precipitation Anomalies, *Monthly Weather Review*, **126** (12), 3102-3116, doi: 10.1175/1520-0493(1998)126<3102:ROMLTC>2.0.CO;2.

Speer, J. H., 2010: *Fundamentals of tree-ring research*. University of Arizona Press, 333 pp.

St. George, S., 2010: Tree Rings as Paleoflood and Paleostage Indicators. *Tree Rings and Natural Hazards: A State-of-Art*, Markus Stoffel, Michelle Bollschweiler, R. David Butler, and H. Brian Luckman, Eds., Springer Netherlands, pp. 233-239.

Steinschneider, S. and Lall, U., 2015: Daily Precipitation and Tropical Moisture Exports across the Eastern United States: An Application of Archetypal Analysis to Identify Spatiotemporal Structure, *Journal of Climate*, **28** (21), 8585-8602, doi: 10.1175/JCLI-D-15-0340.1.

Stone, E. and Cutler, A., 1996: Introduction to archetypal analysis of spatio-temporal dynamics, *Physica D: Nonlinear Phenomena*, **96** (1–4), 110-131, doi: 10.1016/0167-2789(96)00016-4.

Thomas, D. M. and Benson, M. A., 1970: Generalization of streamflow characteristics from drainage-basin characteristics, Water Supply Paper, 1975, pp.

Thornthwaite, C. W., 1948: An Approach toward a Rational Classification of Climate, *Geographical Review*, **38** (1), 55-94, doi: 10.2307/210739.

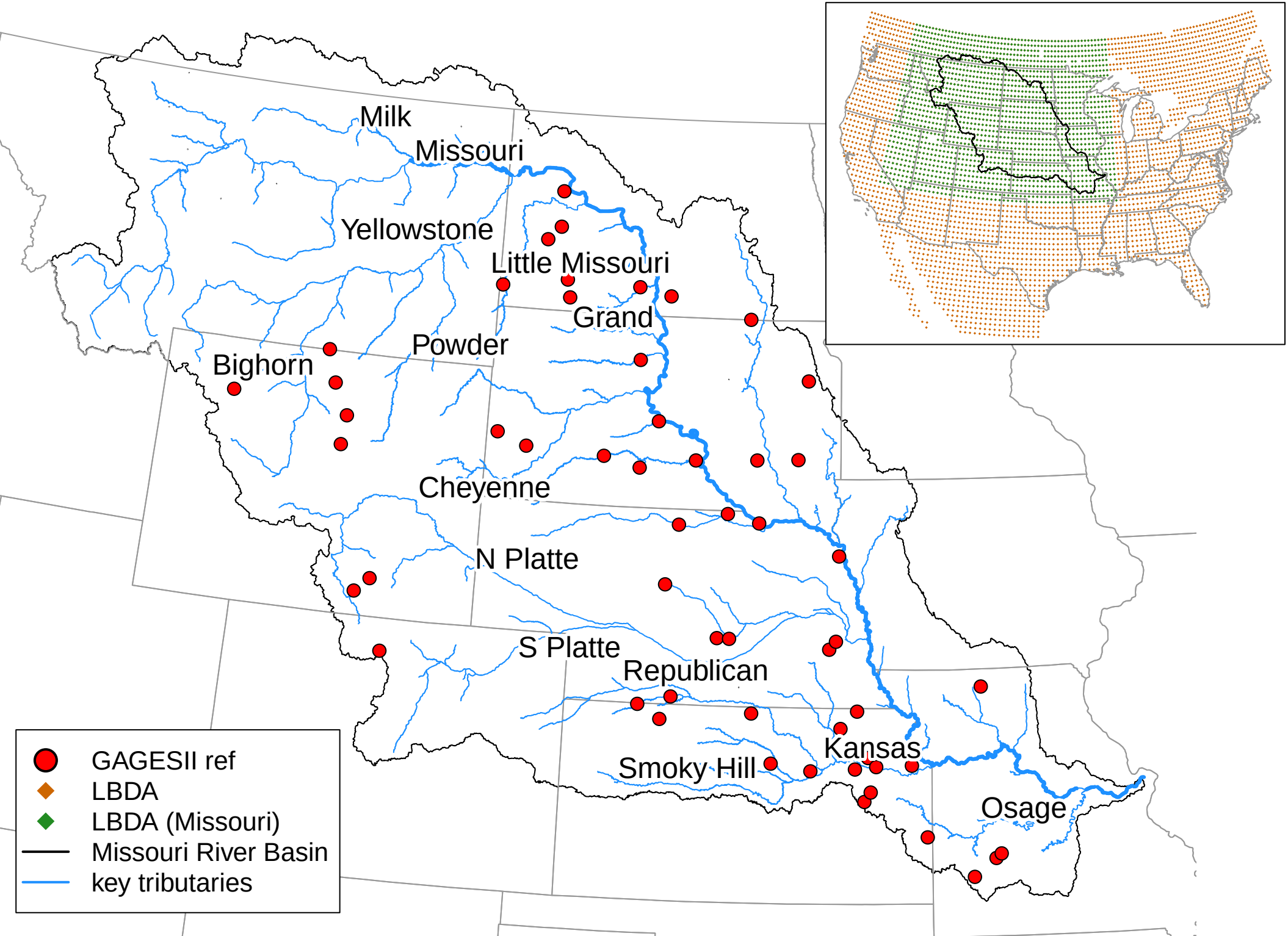
Tierney, J. E., Oppo, D. W., Rosenthal, Y., Russell, J. M. and Linsley, B. K., 2010: Coordinated hydrological regimes in the Indo - Pacific region during the past two millennia, *Paleoceanography*, **25**, PA1102, doi: 10.1029/2009PA001871.

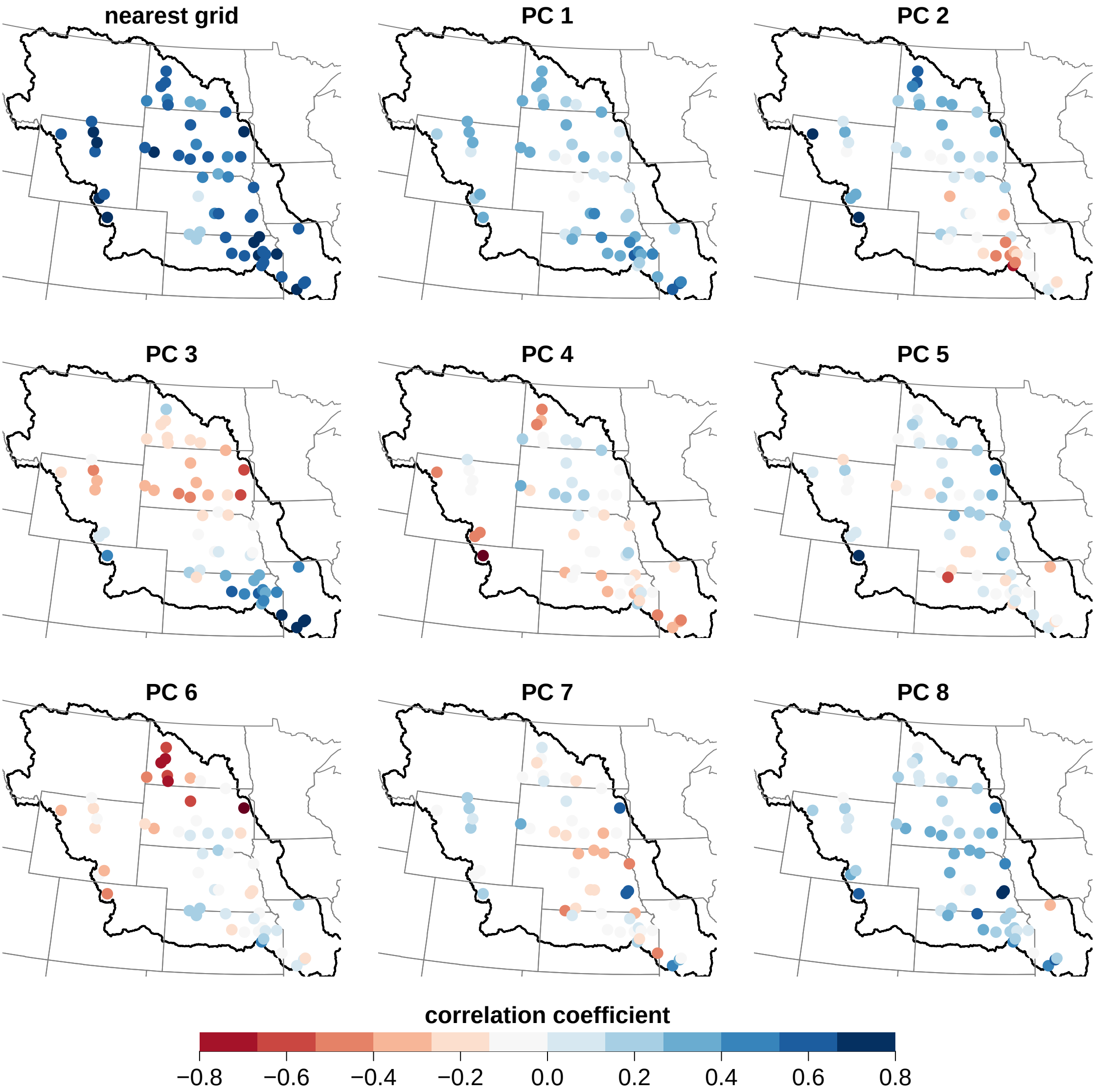
Tootle, G. A. and Piechota, T. C., 2006: Relationships between Pacific and Atlantic ocean sea surface temperatures and U.S. streamflow variability, *Water Resources Research*, **42** (7), n/a-n/a, doi: 10.1029/2005WR004184.

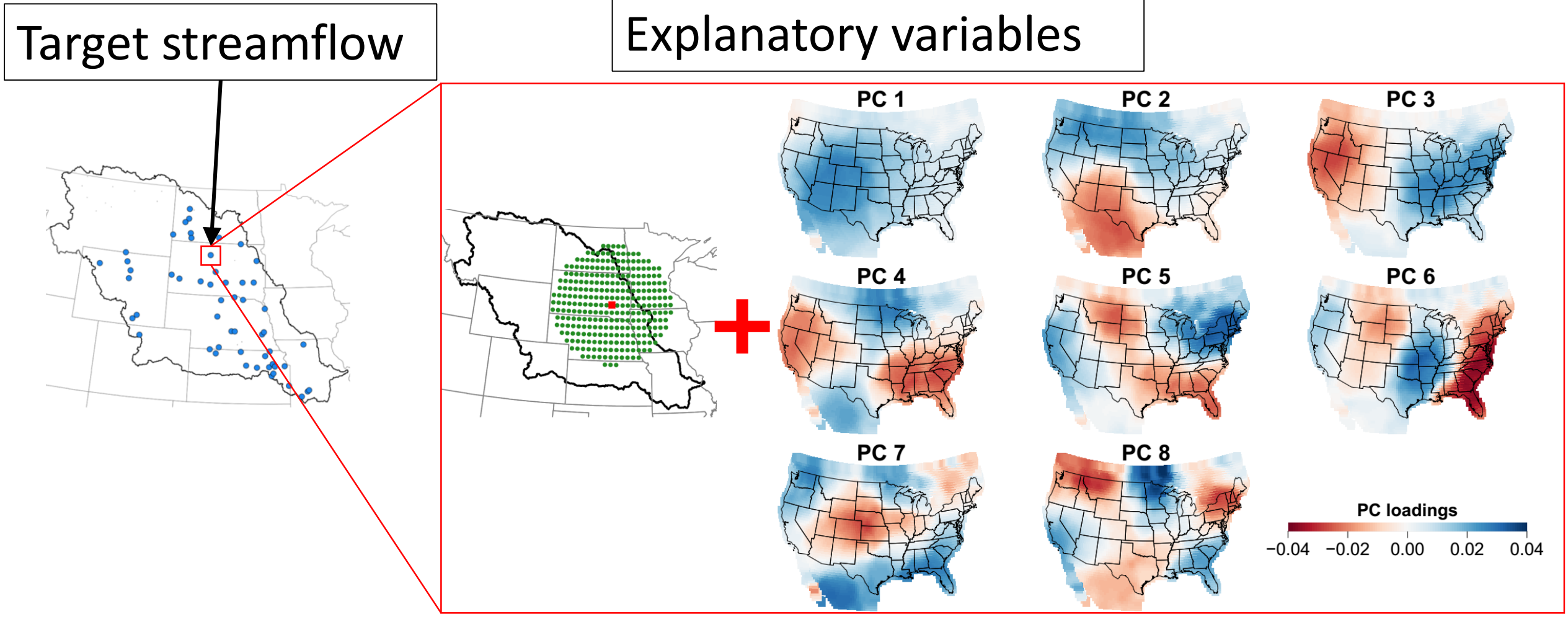
Torrence, C. and Compo, G. P., 1998: A Practical Guide to Wavelet Analysis, *Bulletin of the American Meteorological Society*, **79** (1), 61-78, doi: 10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2.

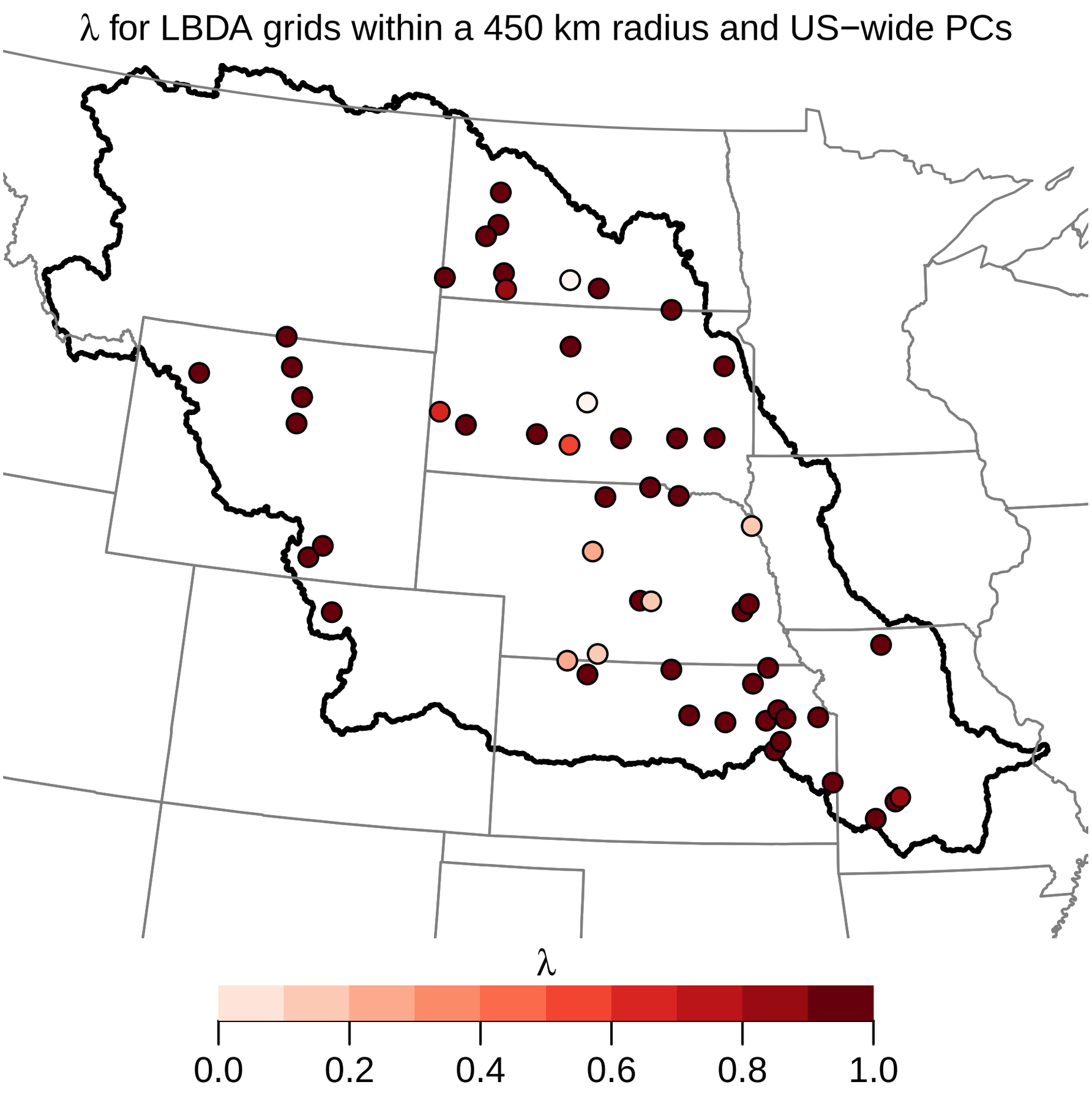
Trenberth, K. E., Branstator, G. W. and Arkin, P. A., 1988: Origins of the 1988 North American Drought, *Science*, **242** (4886), 1640-1645, doi: 10.1126/science.242.4886.1640.

U.S. Geological Survey, 2011, GAGES-II: Geospatial Attributes of Gages for Evaluating Streamflow, viewed 06/19/2015, <http://water.usgs.gov/GIS/metadata/usgswrd/XML/gagesII_Sept2011.xml>.
US Army Corps of Engineers, 2006: Missouri River Mainstem Reservoir System - Master Water Control Manual - Missouri River Basin, pp. 431.
Valencia, D. and Schaake, J. C., 1973: Disaggregation processes in stochastic hydrology, *Water Resour. Res*, **9** (3), 580-585, doi: 10.1029/WR009i003p00580.
Vance, T. R., van Ommen, T. D., Curran, M. A. J., Plummer, C. T. and Moy, A. D., 2012: A Millennial Proxy Record of ENSO and Eastern Australian Rainfall from the Law Dome Ice Core, East Antarctica, *Journal of Climate*, **26** (3), 710-725, doi: 10.1175/jcli-d-12-00003.1.
Vinod, H. D., 1976: Canonical ridge and econometrics of joint production, *Journal of Econometrics*, **4** (2), 147-166, doi: 10.1016/0304-4076(76)90010-5.
Vogel, R. M., Lall, U., Cai, X., Rajagopalan, B., Weiskel, P., Hooper, R. P. and Matalas, N. C., 2015: Hydrology: The interdisciplinary science of water, *Water Resources Research*, **51** (6), 4409-4430, doi: 10.1002/2015WR017049.
White, I. R., Royston, P. and Wood, A. M., 2011: Multiple imputation using chained equations: Issues and guidance for practice, *Statistics in Medicine*, **30** (4), 377-399, doi: 10.1002/sim.4067.
Wilson, R., Cook, E., D'Arrigo, R., Riedwyl, N., Evans, M. N., Tudhope, A. and Allan, R., 2010: Reconstructing ENSO: the influence of method, proxy data, climate forcing and teleconnections, *Journal of Quaternary Science*, **25** (1), 62-78, doi: 10.1002/jqs.1297.
Wise, E. K., 2010: Spatiotemporal variability of the precipitation dipole transition zone in the western United States, *Geophysical Research Letters*, **37** (7), doi: 10.1029/2009GL042193.
Woodhouse, C. A. and Meko, D., 1997: Number of Winter Precipitation Days Reconstructed from Southwestern Tree Rings, *Journal of Climate*, **10** (10), 2663-2669, doi: 10.1175/1520-0442(1997)010<2663:NOWPDR>2.0.CO;2.
Woodhouse, C. A. and Overpeck, J. T., 1998: 2000 years of drought variability in the central United States, *Bulletin of the American Meteorological Society*, **79** (12), 2693, doi.
Woodhouse, C. A., Gray, S. T. and Meko, D. M., 2006: Updated streamflow reconstructions for the Upper Colorado River Basin, *Water Resources Research*, **42** (5), W05415, doi: 10.1029/2005wr004455.
Woodhouse, C. A., Kunkel, K. E., Easterling, D. R. and Cook, E. R., 2005: The twentieth-century pluvial in the western United States, *Geophysical Research Letters*, **32** (7), doi: 10.1029/2005GL022413.
Woodhouse, C. A., Meko, D. M., MacDonald, G. M., Stahle, D. W. and Cook, E. R., 2010: A 1,200-year perspective of 21st century drought in southwestern North America, *Proceedings of the National Academy of Sciences*, **107** (50), 21283-21288, doi: 10.1073/pnas.0911197107.
Woodhouse, C. A., Lukas, J., Brice, B., Hirschboeck, K., Hartmann, H., Kostuk, M., Lay, E., Martinex, D., McMahan, B. and Shah, A., 2002, Tree rings and streamflow, viewed 2015/06/01, <http://treeflow.info/content/tree-rings-and-streamflow>.

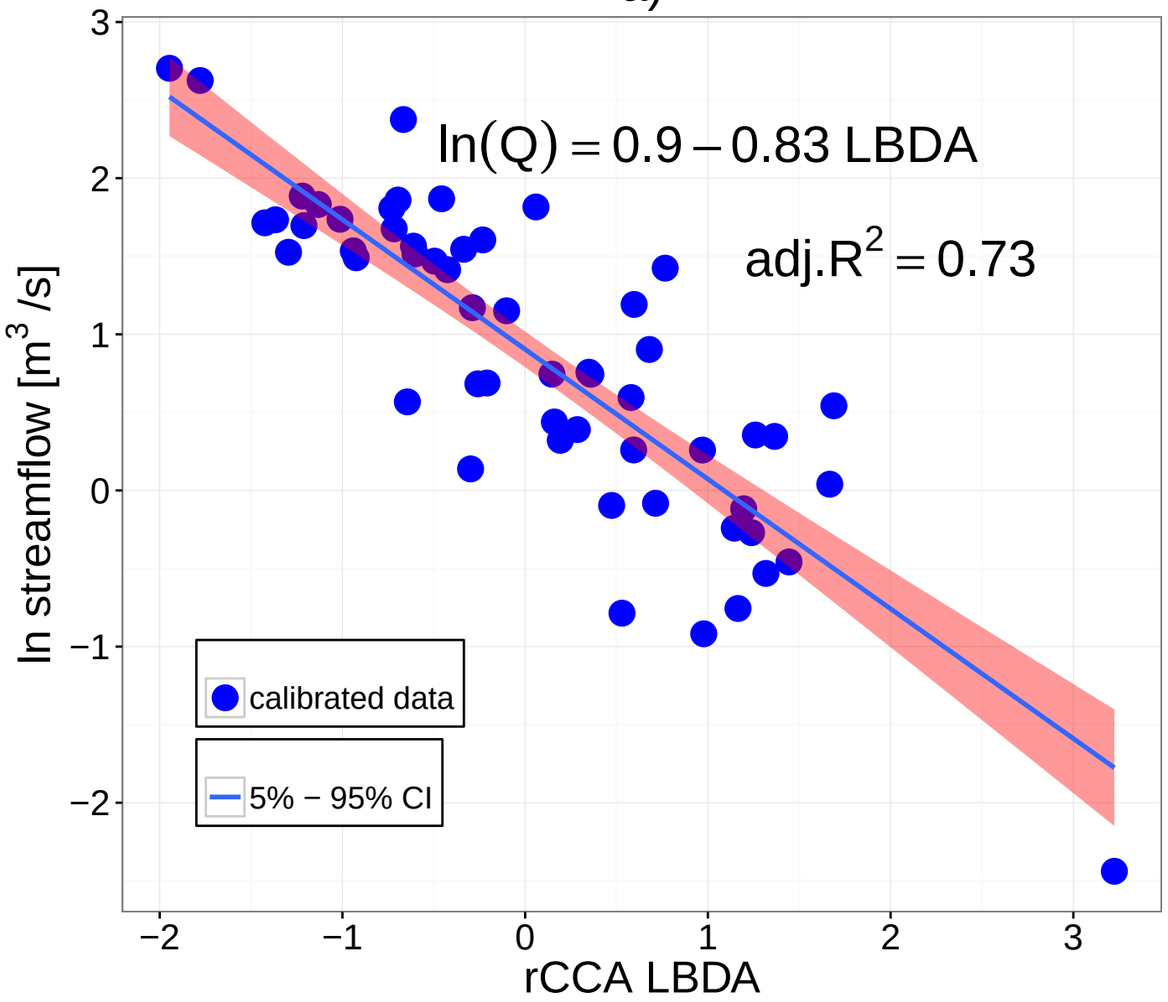

Milk
Missouri
Yellowstone
Little Missouri
Grand
Powder
Bighorn
Cheyenne
N Platte
S Platte
Republican
Smoky Hill
Kansas
Osage
GAGESII ref
LBDA
LBDA (Missouri)
Missouri River Basin
key tributaries


nearest grid
PC 1
PC 2
PC 3
PC 4
PC 5
PC 6
PC 7
PC 8
correlation coefficient
−0.8
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6
0.8

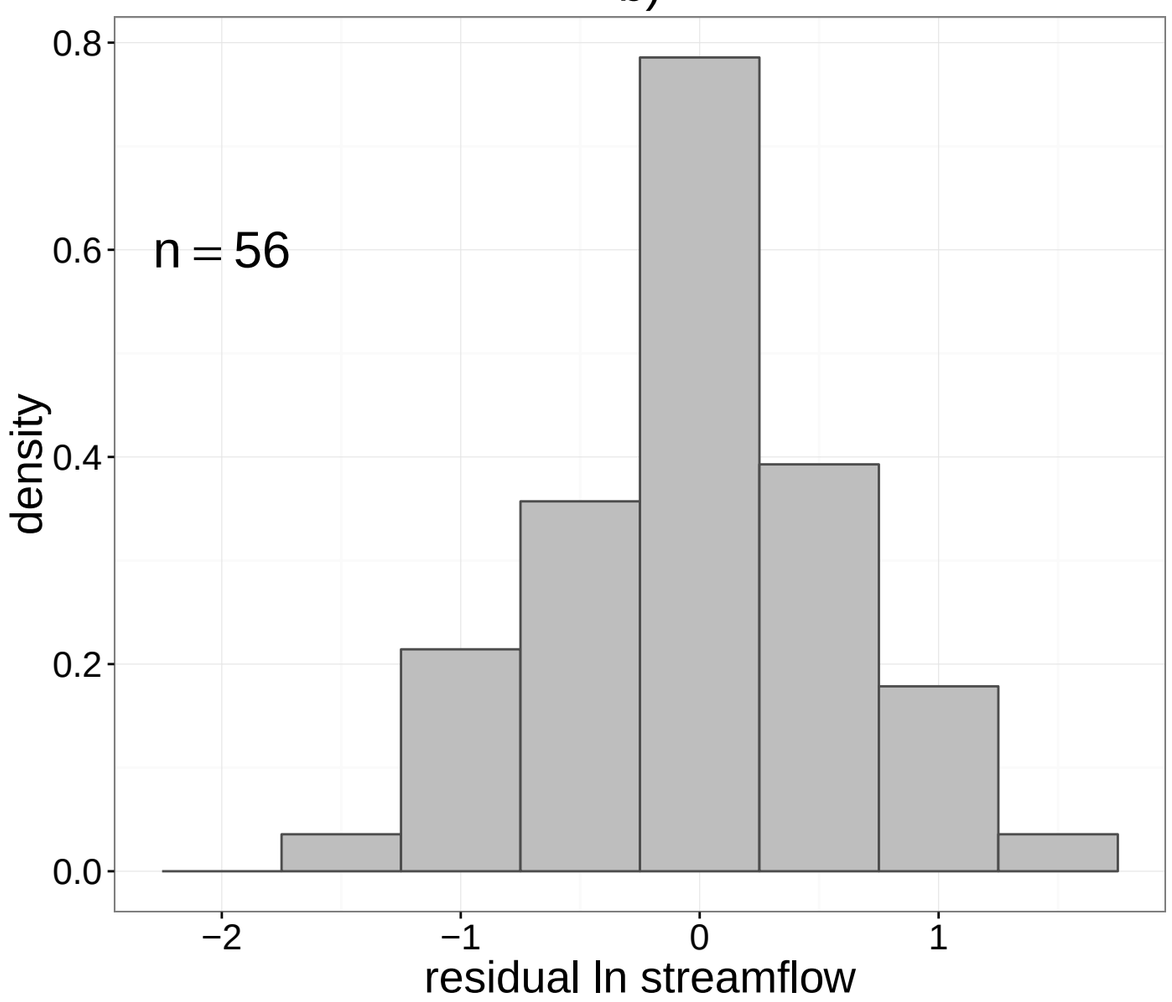

Target streamflow
Explanatory variables
PC 1
PC 2
PC 3
PC 4
PC 5
PC 6
PC 7
PC 8
PC loadings
−0.04
−0.02
0.00
0.02
0.04


λ for LBDA grids within a 450 km radius and US−wide PCs
λ
0.0
0.2
0.4
0.6
0.8
1.0

a)



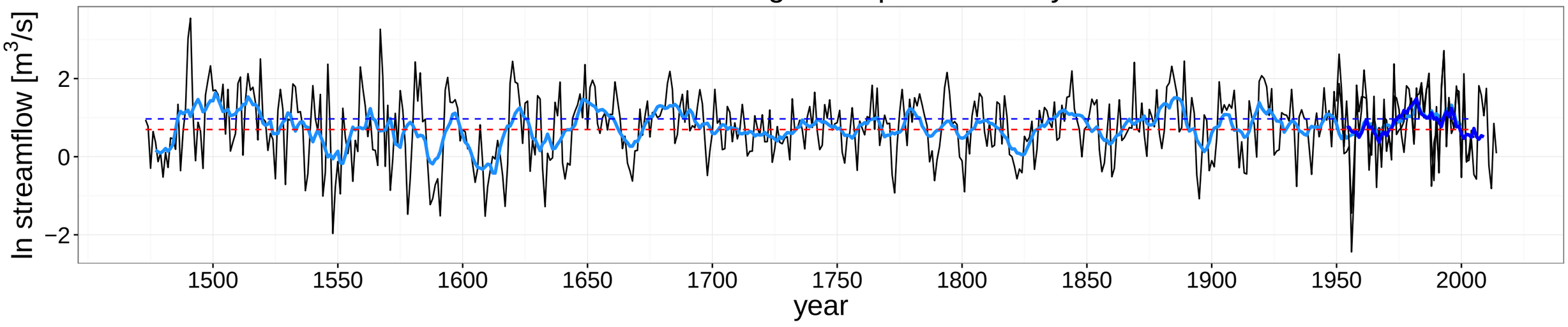
b)

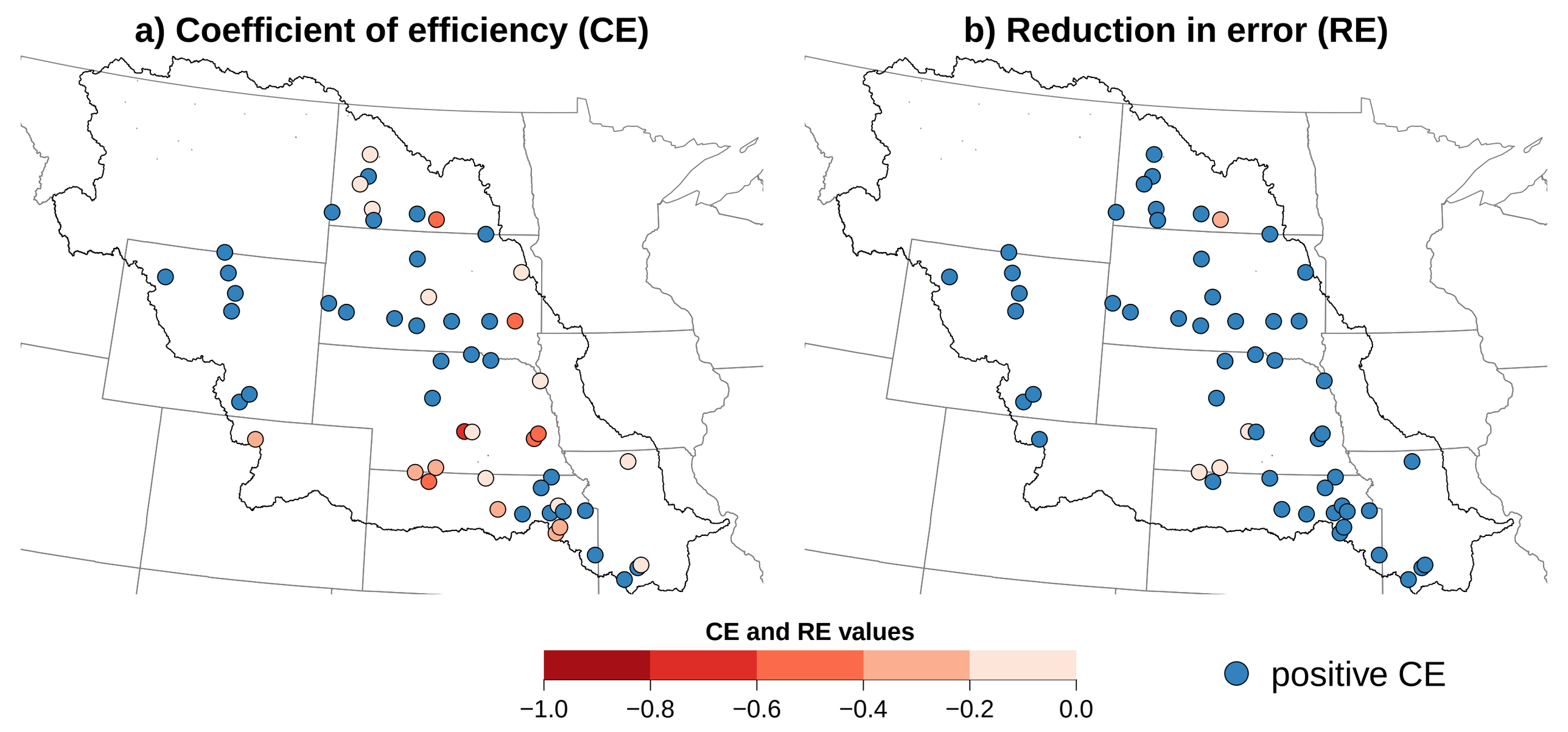


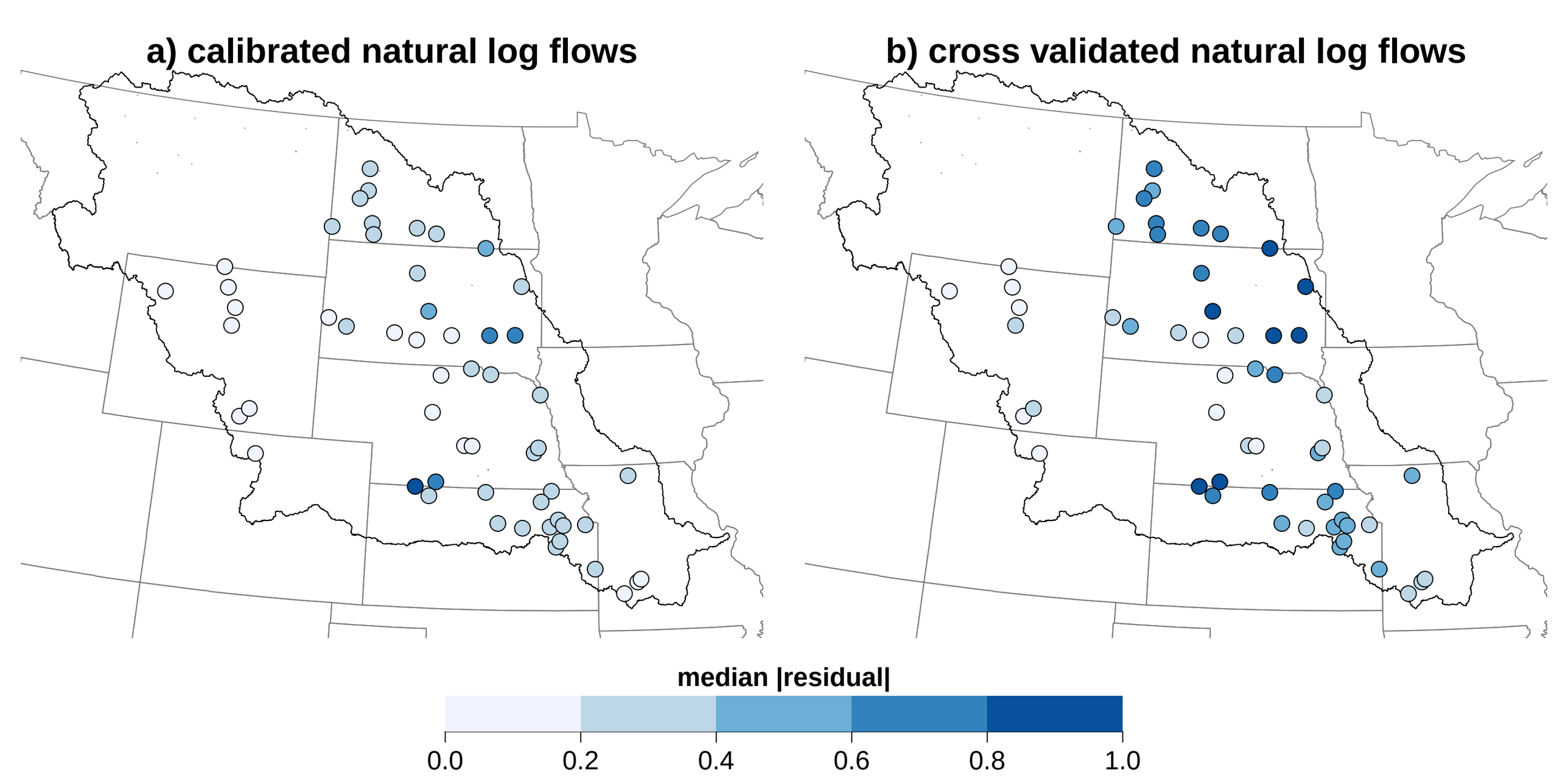
c)

max duration flood inst/paleo: 14 / 34 years

max duration drought inst/paleo 7 / 23 years




a) Coefficient of efficiency (CE)
b) Reduction in error (RE)
CE and RE values
−1.0
−0.8
−0.6
−0.4
−0.2
0.0
positive CE


a) calibrated natural log flows
b) cross validated natural log flows
median |residual|
0.0
0.2
0.4
0.6
0.8
1.0

PC 1 (26%)

PC score

year

PC 1

PC 2 (20.1%)

PC score

year

PC 2

PC 3 (12.5%)

PC score

year

PC 3

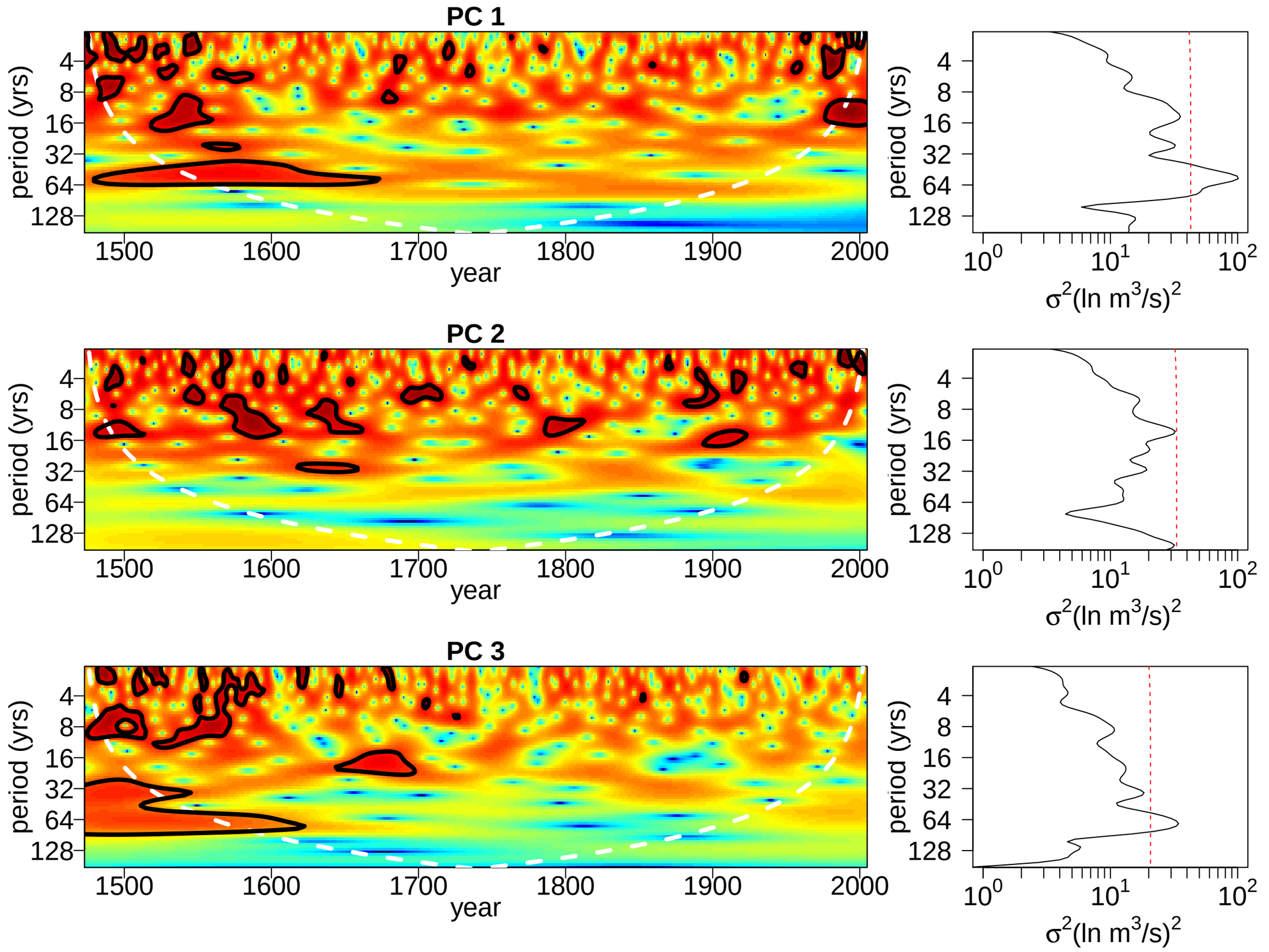
PC

10yr mov. ave.

PC score = 0

PC loadings

−0.30 −0.15 0.00 0.15 0.30


PC 1
PC 2
PC 3
period (yrs)
4
8
16
32
64
128
1500
1600
1700
1800
1900
2000
year
10^0
10^1
10^2
σ^2(ln $m^3/s)^2$