Journal of Hydrology 557 (2018) 791–804

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

Monthly paleostreamflow reconstruction from annual tree-ring chronologies

J.H. Stagge [a,*], D.E. Rosenberg [a], R.J. DeRose [b], T.M. Rittenour [c]

[a] Utah State University, Department of Civil and Environmental Engineering, Logan, UT 84321, United States
[b] United States Forest Service, Forest Inventory and Analysis, Rocky Mountain Research Station, Ogden, UT 84401, United States
[c] Utah State University, Department of Geology, Logan, UT 84321, United States

ARTICLE INFO

Article history:
Received 23 March 2017
Received in revised form 27 October 2017
Accepted 21 December 2017
Available online 27 December 2017
This manuscript was handled by A. Bardossy, Editor-in-Chief, with the assistance of Sheng Yue, Associate Editor

Keywords:
Paleoclimate
Streamflow reconstruction
Tree-ring
Global circulation
Dendrochronology

ABSTRACT

Paleoclimate reconstructions are increasingly used to characterize annual climate variability prior to the instrumental record, to improve estimates of climate extremes, and to provide a baseline for climate-change projections. To date, paleoclimate records have seen limited engineering use to estimate hydrologic risks because water systems models and managers usually require streamflow input at the monthly scale. This study explores the hypothesis that monthly streamflows can be adequately modeled by statistically decomposing annual flow reconstructions. To test this hypothesis, a multiple linear regression model for monthly streamflow reconstruction is presented that expands the set of predictors to include annual streamflow reconstructions, reconstructions of global circulation, and potential differences among regional tree-ring chronologies related to tree species and geographic location. This approach is used to reconstruct 600 years of monthly streamflows at two sites on the Bear and Logan rivers in northern Utah. Nash-Sutcliffe Efficiencies remain above zero (0.26–0.60) for all months except April and Pearson's correlation coefficients (R) are 0.94 and 0.88 for the Bear and Logan rivers, respectively, confirming that the model can adequately reproduce monthly flows during the reference period (10/1942 to 9/2015). Incorporating a flexible transition between the previous and concurrent annual reconstructed flows was the most important factor for model skill. Expanding the model to include global climate indices and regional tree-ring chronologies produced smaller, but still significant improvements in model fit. The model presented here is the only approach currently available to reconstruct monthly streamflows directly from tree-ring chronologies and climate reconstructions, rather than using resampling of the observed record. With reasonable estimates of monthly flow that extend back in time many centuries, water managers can challenge systems models with a larger range of natural variability in drought and pluvial events and better evaluate extreme events with recurrence intervals longer than the observed record. Establishing this natural baseline is critical when estimating future hydrologic risks under conditions of a non-stationary climate.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

While grappling with the question of how future climate changes will affect the likelihood and severity of hydrological extremes (floods and droughts), hydrologists, engineers, and water-resources planners have noted the potential for streamflow reconstructions to characterize pre-industrial hydrologic variability over multiple centuries (Bonin and Burn, 2005). By combining reconstructions of the past with climate change projections, it may be possible to place the signal of climate change-induced streamflow trends in the context of long-term natural variability. In addition, streamflow reconstructions can significantly increase the number of scenarios used for drought vulnerability studies or water resources systems optimization. Despite these potential benefits, streamflow reconstructions have not gained widespread use in water systems analysis, in part because flow has typically been reconstructed at an annual resolution, which is generally too coarse for analysis of drought vulnerability and decision-making. This study explores whether monthly streamflows can be adequately predicted from annual tree-ring chronologies and other reconstructed data. To confirm this, we outline and test a novel statistical method to reconstruct monthly flow series.

Existing techniques for annual streamflow reconstruction primarily rely on linear regression to relate carefully chosen tree-ring chronologies to mean annual flow (MAF) during the available

* Corresponding author.
E-mail address: james.stagge@usu.edu (J.H. Stagge).

https://doi.org/10.1016/j.jhydrol.2017.12.057
0022-1694/© 2017 Elsevier B.V. All rights reserved.

instrumental record. This relationship is then applied to the full tree-ring record to reconstruct flows outside the observed period (Loaiciga et al., 1993; Cook and Kairiukstis, 2013). Tree-ring chronologies are carefully developed by selecting sites and species in which the reconstruction variable is the limiting factor for tree growth. Multiple replication samples are often taken for each tree and these measurements across many trees at a given site are combined to generate a master site ring width, through a process called crossdating, whereby common features in the chronology are matched (Fritts, 2012; Douglass, 1941). The resulting chronologies can be linked to climate variables either through simple linear regression (Duvick and Blasing, 1981), multiple linear regression (Meko et al., 1980), or more complex approaches.

More complex reconstruction approaches often rely on Principal Component Analysis (PCA) to extract differences/commonalities across multiple site chronologies and tree species to better capture regional variations (Cook et al., 1994; Hidalgo et al., 2000; Maxwell et al., 2011) or seasonal effects (Brubaker, 1980). Meko et al. (2015) showed that flow reconstruction accuracy can be improved by using PCA and regression to combine chronologies from traditional reconstruction species and less traditional species which capture unique climate signals.

Reconstruction approaches assume that processes relating climate to tree-ring growth during the instrumental record are identical to the reconstructed period (Fritts, 2012). This Principle of Uniformatism has been slightly modified in modern dendrochronology to separate the important climate signal from underlying factors affecting tree growth (Cook, 1987). Uniformitism is particularly important for flow reconstructions, where river reaches should be selected to avoid significant man-made effects. Alternatively, the effects of impoundments or land use change can be removed from the time series to approximate a near-natural flow record. Annual streamflow recontructions have been produced for regions with adequate tree-ring chronologies (Meko et al., 2001; Woodhouse et al., 2006) and have been used for water resources planning, e.g. Woodhouse and Lukas (2006) and Axelson et al. (2009).

Despite the availability of annual streamflow reconstructions, few methodologies currently exist to reconstruct sub-annual flow from annual resolved tree-ring chronologies (Gangopadhyay et al., 2015; Sauchyn and Ilich, 2017). The Sauchyn and Ilich (2017) approach uses stochastic hydrology techniques to generate many feasible sequences of weekly flows that sum to the annual reconstruction while maintaining statistical properties of the observed record. The Gangopadhyay et al. (2015) approach instead resamples annual subsets of temperature and precipitation from the instrumental records and matches them to tree-ring widths in the paleo-record using a K-nearest neighbor approach repeated many times to develop an ensemble of temperature and precipitation timeseries (Gangopadhyay et al., 2009). Temperature and precipitation are then used as inputs for a water balance model (Wolock and McCabe, 1999). This method has been used to generate seasonal streamflow in Nevada (Solander et al., 2010) and monthly streamflow along the Colorado River (Gangopadhyay et al., 2015). While useful for generating runoff in well-studied watersheds, this approach requires a calibrated watershed model that is not always available. Additionally, by resampling from the observed record, the potential monthly time series of temperature and precipitation are limited to re-ordering ≈60–100 observed annual subsets from the instrumental record. While this may be an effective approach for some locations, it is highly desirable to develop methods to reconstruct monthly streamflow directly, without the need for watershed models and the limitations of repeated resampling.

To test the hypothesis that monthly streamflow can be reconstructed from annual tree-ring chronologies, this study introduces a novel approach for reconstructing monthly streamflow that extends fundamental principles of flow reconstruction and then demonstrates these models by reconstructing flow and evaluating goodness of fit for two sites in northern Utah. The candidate models include a simple Monthly Fraction (MF) model, an Annual Percentile (AP) model that directly links annual flow percentile with monthly percentile, and an Annual Percentile with Regression (APR) model that uses multiple annual reconstructions to predict the monthly percentile. Several versions of the APR model are considered, each adding increasing predictors which include: 1) the original annual streamflow reconstructions; 2) climate reconstructions of the El Niño-Southern Oscillation (ENSO); and 3) regionally available tree-ring chronologies. Regional tree-ring chronologies were included as predictors based on the hypothesis that different species, elevations, and site locations might capture different parts of the seasonal hydrologic signal, whereas ENSO was considered because it has shown coherence with streamflow and precipitation in the western U.S. and Utah (Cayan et al., 1999; Schoennagel et al., 2005; Wang et al., 2012; DeFlorio et al., 2013; Zhou et al., 2014). The paper further evaluates and discusses model results in the context of individual predictors, their physical basis, and implications for water management.

2. Models

Three model frameworks are introduced in this study as potential candidates for the reconstruction of mean monthly streamflow. The MF model uses simplistic assumptions to reconstruct monthly streamflows and is included as a "null" model, against which the other models can be compared. The remaining two models constitute the primary approach proposed herein and are presented in order of increasing complexity, each applying the same basic framework, but with increasing numbers of predictors. First, the AP model directly links the reconstructed annual streamflow percentiles to monthly percentiles, using the assumption that the monthly percentile is constant and identical to the reconstructed annual percentile throughout each water year. Second, the APR model estimates monthly percentiles using regression, first considering only lagged annual streamflow percentiles and ultimately incorporating additional predictors such as global climate indices (ENSO) or spatial/species patterns in regional tree-ring chronologies extracted by PCA. The models and fitting procedures are available as an R package, paleoAPR (Stagge, 2017), while the code and data for the specific analyses performed in this paper are available in an online repository athttps://doi.org/10.5281/zenodo.1029739.

2.1. Monthly fraction (MF) model

The MF model assumes that the monthly proportion of total annual streamflow (TAF) is identical across all years. Based on this assumption, the monthly fraction, $\bar{f}_{m,y}$, is determined for each month by dividing the monthly flow volume, $Q_{m,a}$, by annual volume, TAF_y. In this notation, "m" and "a" subscripts correspond to monthly and annual steps, respectively. TAF is equivalent to $12 \times MAF$. Monthly streamflow is then reconstructed by multiplying the appropriate mean monthly fraction, $\bar{f}_m$, for each of the 12 months by reconstructed TAF:

$$\widehat{Q}_m = \bar{f}_m \times TAF_a, \quad \text{where} \quad \bar{f}_m = \frac{\sum_{a=1}^{n} f_{m,y}}{n} \quad \text{and} \quad f_{m,a} = \frac{Q_{m,a}}{TAF_a} \quad (1)$$

where $\widehat{Q}_m$ represents estimated monthly streamflow. Monthly flows reconstructed by the MF model retain the same seasonal shape, but are scaled linearly. For example, if an average of 30% of each year's historical flow volume occurred during June, this proportion is maintained in the reconstruction. Prior to performing

reconstruction, the 12 f_m values were re-scaled to sum to 1, to ensure monthly flows would recreate the original MAF reconstruction. The mean f_m was used to represent the monthly fraction, rather than the median, which made little difference in the case study results.

2.2. Annual percentile (AP) model

The AP model forms the base for all subsequent reconstruction models. It assumes that the streamflow percentile for each month is equivalent to the MAF percentile for the particular water-year. That is, if a given year's MAF is in the 20th percentile, then the mean flow for each month of that water-year was also the 20th percentile, determined by the cumulative probability distribution of flow for each month. Because subsequent models depend on a Gaussian distribution, we transformed the percentile (0–100) to a standard normal value ($-\infty$ to ∞), represented by the variable N:

$$\widehat{N}_m = N_a \tag{2}$$

where $\widehat{N}_m$ is the estimated monthly standard normal and the subscripts m and a again represent monthly and annual values of the normalized streamflow percentile.

For use in the AP model, a univariate probability distribution was fit to the MAF and each of the monthly mean flows to permit transformation to and from the standard normal distribution. Nonparametric alternatives, such as empirical cumulative distributions, were not considered because they limit the ability to extrapolate beyond the original data. Annual and monthly cumulative distributions were fit based on Maximum Likelihood Estimation (MLE) using the fitdistrplus package in R (Delignette-Muller and Dutang, 2015). Candidate distributions were visually compared using skewness-kurtosis plots (Cullen and Frey, 1999) evaluated with 500 nonparametric bootstrap realizations. Final selection of candidate distributions was further validated based on Akaike information criterion (AIC), quantile plots, and results of Kolmogorov-Smirnov and Anderson-Darling tests. Statistical distribution tests were based on 5000 bootstrap simulations, thereby avoiding the issue of applying tests where parameters are estimated from the data (Crutcher, 1975).

2.3. Annual percentile with regression (APR) model

The APR model has the same fundamental structure as the AP model without enforcing a 1:1 relationship between the annual and monthly standard normal values. Instead, a unique regression model for each month was fit, which estimated that month's streamflow percentile based on a number of potential predictors. This regression approaches uses similar principles as other multiple linear regression reconstructions (Meko et al., 1980; Cook et al., 1994; Hidalgo et al., 2000). It is designed to be flexible so that, in future studies, any variable with a demonstrated climate effect could be considered as a predictor.

Three APR models of increasing complexity are considered in this study. The first uses only the standard normal of the lagged percentiles from the annual streamflow reconstruction:

$$\widehat{N}_{m,0} = \beta_{-1}N_{a,-1} + \beta_0 N_{a,0} + \beta_{+1}N_{+1} \tag{3}$$

where $\widehat{N}_{m,0}$ represents monthly standard normal, as in Eq. (2), and $N_{a,-1}$, $N_{a,0}$, and $N_{a,+1}$ are annual normalized percentiles for the previous, concurrent, and future water years, respectively. The associated β values are model coefficients fit using the regression approach described below. Transformation to standard normal values, rather than percentiles, was done to permit regression with standard, normal residuals. The model described in Eq. (3) is referred to as the APR model with lagged predictors. It can be viewed as a temporal disaggregation of the original annual reconstruction, using no additional data. The lagged regression approach alleviates the need to explicitly define the most appropriate water-year bounds, as it is naturally derived based on predictive power.

The second and third variants of the APR model incorporate predictors outside the original annual reconstruction. The second model adds two reconstructions of the ENSO climate index, while the third model adds regional tree-ring principal components. These are referred to as "APR with Climate Indices" and "APR with all predictors", respectively, with the full "APR with all predictors" model taking the form:

$$\widehat{N}_{m,0} = \beta_{-1}N_{a,-1} + \beta_0 N_{a,0} + \beta_{+1}N_{+1} + \beta_{ENSO}\,ENSO + \sum_{j=1}^{8}\beta_j\,PC_j \tag{4}$$

where additional predictors for ENSO and regional tree-ring principal components are included as β_{ENSO} and β_j, respectively.

3. Methods

3.1. Mean annual reconstructions

The candidate monthly reconstruction approaches were applied at two sites with previously published annual streamflow reconstructions in the Bear River watershed of northern Utah. Flows in this region are snowmelt-dominated, with the majority of annual precipitation falling in the form of winter snow. The two sites were located on the Logan River and in the upper Bear River. Both reconstructions (Allen et al., 2013; DeRose et al., 2015) used similar methods to reconstruct the MAF based on the USGS water-year of Oct-Sep.

The Bear River is located in the Intermountain U.S. and is the single largest river in the eastern Great Basin. It drains portions of northeastern Utah, southwestern Wyoming, and Southeastern Idaho, with its headwaters located in the Uinta Mountains. The Bear River is the largest tributary to the Great Salt Lake, providing approximately 60% of its annual inflow. The annual flow reconstruction for the upper Bear River is based on the instrumental flow record from USGS gauge 1011500 (1942 CE-present). This gauge is located near the furthest upstream Utah-Wyoming border and includes flow contribution from a 445 km^2 watershed along the north slope of the Uinta mountains. Flow at this site contributes approximately 8% of the total Bear River flow. Flow at this site is considered near-natural, as it is upstream of major diversions, with only a single, small storage reservoir located upstream. The tree-ring predictor data used to reconstruct MAF was based on a nearby Utah juniper (*Juniperus osteosperma*) chronology. The reconstruction (DeRose et al., 2015) covered the period 800-2010 CE, and explained 67% of annual variation (R^2 = 0.672).

The Logan River is the largest tributary to the Bear River and drains 1389 km^2 of the Bear River Range in northeastern Utah and southeastern Idaho. The MAF reconstruction (Allen et al., 2013) is based on USGS gauge 10109001 (1922 CE – present), which captures the near-natural upper portion (554 km^2) of the Logan River watershed prior to impoundments and diversions near the city of Logan, Utah. The reconstruction covers the period from 1605 to 2010 CE. Two reconstruction models were used, one termed the Local Model that relied on two Rocky Mountain juniper (*Juniperus scopulorum*) chronologies located within the Logan River watershed and another termed the Regional Model that also incorporated a two-needle pinyon (*Pinus edulis*) chronology from north-central Utah and a limber pine (*Pinus flexilis*) chronology from western Wyoming. The Regional Model showed better skill (R^2 = 0.581) than the local model (R^2 = 0.482), but both reconstructions were considered for this study.

3.2. Global circulation index predictors

Global circulation indices are one of the variables which could be included in the APR model framework. For these sites, two reconstructions of the El Niño-Southern Oscillation (ENSO, Trenberth, 1997) were considered as potential predictors. ENSO is the dominant mode of interannual climate variability on Earth and has a documented effect on the climate of North America (DeFlorio et al., 2013; Zhou et al., 2014), and Utah in particular (Cayan et al., 1999; Schoennagel et al., 2005). These reconstructions do not directly measure ENSO, defined by sea surface temperature in the southern Pacific Ocean, but rather capture its signature on tree-ring growth over large regions known to be affected by ENSO. In this way, including reconstructions of ENSO provides a broader sample of tree-ring chronologies to help characterize the spatial variability of interannual water balance.

The first ENSO reconstruction, described in detail by Li et al. (2011), is based on the leading Principal Component (PC) of the North American Drought Atlas, a reconstruction of annually resolved summer (JJA) Palmer Drought Severity Index (Cook et al., 1999). The leading PC of this reconstruction is most correlated with the January-March (JFM) ENSO Nino3 index ($r = 0.51$) and uses this relationship to extend the ENSO record back to 900 CE. Loading for the Li et al. (2011) reconstruction is centered on the American southwest, with northern Utah at the extreme northern range of strong loading values. By comparison, the Li et al. (2013) ENSO reconstruction uses tree-ring chronologies from seven broad climate regions surrounding the Pacific Ocean, including three in North America, two in South America, two in Asia, and one in New Zealand. This reconstruction captures a different portion of the ENSO signal and is best correlated with a region further west in the southern Pacific Ocean (ENSO Nino3.4 index) during the previous November-January (NDJ). The larger archive of tree-ring chronologies used by Li et al. (2013) produces a better estimate during the observed period ($r = 0.80$), but only extends back to 1300 CE. Because their different underlying tree-ring chronologies capture different parts of the ENSO signal (geographic extent and season), these reconstructions are considered complementary to one another. They are included here to demonstrate the flexibility and potential for including large-scale climate index reconstructions, particularly in regions with known teleconnections.

3.3. Tree-ring chronology analysis

The final set of potential reconstruction predictor variables are based on regional tree-ring chronologies (49 chronologies located within a maximum 450 km radius of the reconstruction gauges). All tree-ring sites had limited soil development and occurred on southerly or westerly aspects, occupying open canopy positions. Chronologies included seven different species: Douglas-fir (*Pseudotsuga menziesii*), two-needle pinyon (*Pinus edulis*), singleleaf pinyon (*Pinus monophylla*), limber pine (*Pinus flexilis*), ponderosa pine (*Pinus ponderosa*), Utah juniper (*Juniperus osteosperma*), and Rocky Mountain juniper (*Juniperus scopulorum*). Tree-ring indices (ie, chronologies) were developed using the same approach for each site using the dplR package (Bunn, 2008). Individual tree-ring series were detrended using a cubic-smoothing spline with a frequency cut-off of 50% at a wavelength of $\frac{2}{3}$ of the length of the time-series to remove variability potentially associated with stand dynamics or biological growth trends. Tree-ring series were then averaged using a robust biweight mean. Autocorrelation was retained in the resultant tree-ring chronologies. Preliminary analysis (not presented) showed that correlation between monthly flow percentiles differed across selected tree-ring chronologies, suggesting that different chronologies could be used to highlight unique portions of the seasonal hydroclimatic signal. Such findings have been noted in previous studies, such as Brubaker (1980), which showed that the first two principal components of regional tree-ring chronologies explained different summer and winter precipitation responses and were related to varying species (Douglas-fir and ponderosa pine).

Regional tree-ring chronologies were selected to include a range of species, elevations, and locations in an effort to highlight the most important factors for predicting seasonal flow. Because tree-ring chronologies can be highly correlated, Principal Component Analysis (PCA) was used to isolate unique differences in response, reducing the chronologies to a smaller set of orthogonal Principal Components (PCs) that explain a majority of the original tree-ring variance. The number of retained PCs was based on the Kaiser-Guttman criterion (Kaiser, 1960; Guttman, 1954), defined as PCs with eigenvalues greater than 1, and by identifying a significant change in variance explained through a scree plot. These stopping criteria are thought to slightly overestimate the number of retained components (Jackson, 1993), which is desirable in the context of maximizing seasonal variability across multiple chronologies.

The regional tree-ring chronologies have varying lengths, primarily due to differences in their earliest measurements (800–1763 CE) (Fig. 1). Traditional PCA requires a fully defined data matrix and therefore would require a trade-off, either removing short chronologies to lengthen the time series or producing PCs beginning in 1763 CE based on the common period. To avoid this limitation and to capture potentially valuable information in the shorter chronologies, missing values were imputed using the missMDA package (Josse et al., 2016) in R prior to PCA analysis. The number of components was determined by 10-fold cross-validation, minimizing mean squared error (MSE) across 100 simulations (Josse and Husson, 2012). Missing values were imputed using an iterative algorithm which cycles between estimating missing values and recalculating PCs until the observed and fitted values converge. PC scores were calculated in this way for a 600 year duration and compared to loadings derived from traditional PCA using approximately 400 (1610 CE, 36 chronologies) and 600 year (1390 CE, 19 chronologies) subsets, selected based on inflection points in chronology availability (Fig. 1). A third cut-off point at 900 years was initially considered, but dismissed because the number of available time series (4) was less than the optimal number of PCs to retain. The imputed PCA results did not significantly change results and was therefore used for all subsequent analysis. The Logan and Bear River sites were located close enough to each other that the same PCA scores were used for both locations.

3.4. Model fitting

Each of the model variants were fit based on USGS stream gauge data, calculated by accumulating daily flows to determine mean monthly flow. Models were fit using the paleoAPR package in R (Stagge, 2017). The entire observed period (7/1942 to 9/2016 for Bear River and 10/1921 to 9/2015 for Logan River) was used as the dependent variable for fitting the models. For each site, two reconstruction models were generated, the first using the true observed MAF as training data and the second using the reconstructed MAF. The model using observed MAF provides an upper envelope for model performance, given perfect input data, while the model using reconstructed MAF provides an estimate of model fit that can be expected for the full reconstruction.

The APR models were the only reconstruction models in this study that rely on regression. Regression fitting was performed using elastic nets (Zou and Hastie, 2005), chosen as an alternative to traditional step-wise Ordinary Least Squares regression used in previous reconstructions (Cook et al., 1994; Hidalgo et al., 2000). Elastic nets are designed to solve issues with stepwise regression, reducing overfitting while also handling high dimensional data with multi-collinearity. This is accomplished by blending ridge regression (Hoerl and Kennard, 1970) and LASSO (least absolute shrinkage and selection operator) regression (Tibshirani, 1997; Tibshirani, 1996) using a term, α, that can range from 0 (ridge regression) to 1 (LASSO regression). Blending these models combines the benefits of each: ridge regression is effective at reducing overfitting and handling highly related predictors, while the LASSO approach can perform variable selection by shrinking non-significant coefficients to zero. Ridge and LASSO regression use similar approaches that seek to minimize the sum of mean squared error along with a penalty term. The penalty terms are the primary difference between the approaches, using either the square of model coefficients, $\sum \beta^2$ in ridge regression, or the absolute value of the model coefficients, $\sum |\beta|$ in LASSO regression. The penalty is then multiplied by a shrinkage parameter, λ.

Calibration and validation of the APR models was performed by tuning the blending, α, and shrinkage, λ, terms in the elastic net regression. The best parameter set was chosen from a grid search of a 10×100 $\alpha \times \lambda$ grid using 10-fold cross-validation, repeated 8 times with random selections, with RMSE as the selection criterion. This produces effective validation and variable selection simultaneously because as the α approaches 1 (LASSO regression) and the λ penalty increases, non-significant variables are removed from the regression by setting them to zero.

The APR model assumes that reconstructed flow percentiles should approximate the standard normal distribution, with a mean of 0 and variance of 1, during the reference period. This assumption follows from the univariate probability distributions which were defined using flows during the reference period and are ultimately used to convert monthly flows into the standard normal, and vice versa. To guarantee that reconstructed flow normals during the reference period approximate the standard normal, a post-processing step was added to the APR model to correct the mean and variance for each month, extending this correction across the full reconstruction.

3.5. Model goodness-of-fit

Goodness-of-fit was evaluated using a suite of tests and measures during the common reference period of 10/1942 to 9/2015, to allow for equal comparisons across the two sites. These goodness-of-fit measures included mean error (ME) to estimate systematic model bias, mean absolute error (MAE) and root-mean-squared error (RMSE) to estimate the residual magnitudes, measures of parametric (Pearson, R) and non-parametric (Spearman, $R_{Spearman}$) correlation, and Nash-Sutcliffe Efficiency (NSE) to estimate overall model predictive power. In addition, residual and time-series plots were generated to allow for visual inspection of residual patterns or timing issues that might not be captured by these metrics (McCuen et al., 2006; Criss and Winston, 2008). Finally, model fit was evaluated for each month separately using the same goodness-of-fit metrics and residual plots. This allowed for a more detailed review of model performance, particularly in months with low flow, which could otherwise be obscured by high flow periods in the full time series.

4. Results

4.1. Monthly fraction (MF) model fitting

Estimation of the MF model required only the calculation of the mean monthly proportion of flow. The Logan and Bear River sites showed similar seasonal patterns, with strong seasonality and flow peaks in the summer caused by delayed melting of the winter snowpack (Fig. 2). Fig. 2 presents all available years, with the mean monthly fraction used for reconstruction shown as a dark line. Logan River (Fig. 2a) had a higher proportion as baseflow and more gradual rising and falling limbs around a peak which occurred in May or June. The Bear River (Fig. 2b) had a more drastic seasonal peak flow, that predominantly occurred in June and occasionally May.

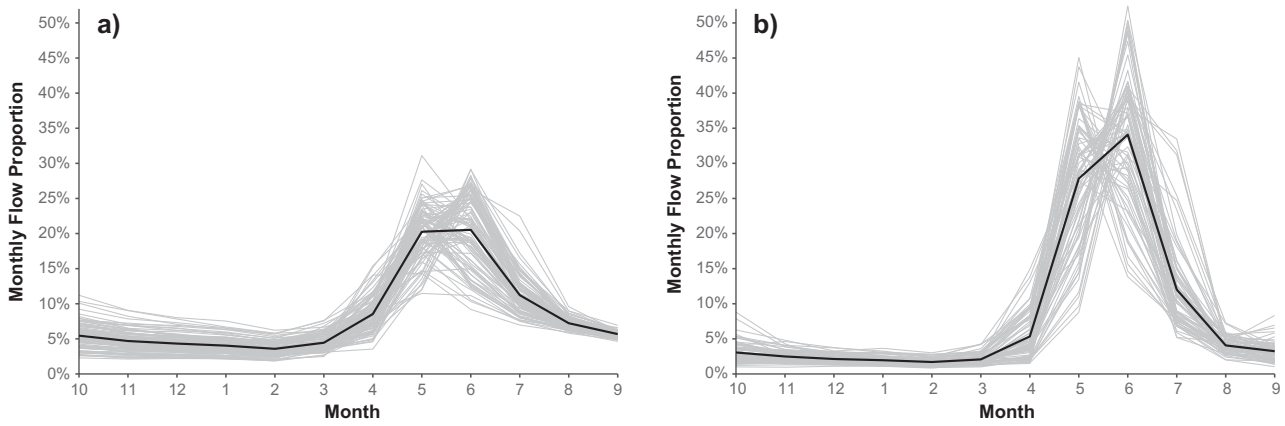

Fig. 2. Monthly flow proportion relative to mean annual flow (%) for Logan (a) and Bear River headwaters (b).

4.2. Annual Percentile (AP) model fitting

The primary model fitting task for the AP model was to determine the best cumulative probability distribution for the MAF and each of the monthly flows. For both sites, skewness-kurtosis plots recommended the use of a gamma distribution for monthly flows and a logistic distribution for MAF. Choice of these distributions was further supported by low AIC values and non-significant tests for violations of the candidate empirical cumulative distributions. The logistic distribution used for annual flows was nearly Gaussian, but with slightly thicker tails, while the gamma distribution used for monthly flows accounted for the positively skewed nature of monthly flows combined with a lower bound at zero.

4.3. APR model with lagged annual reconstructions

The APR model with lagged annual flows reconstruction was equivalent to a temporal disaggregation of the original reconstruction and required model coefficients for the effect of the lagged (−1, 0, +1) annual standard normal values on each of the monthly normalized percentiles. The resulting fit (Fig. 3) showed that the importance of the reconstructed annual flow, quantified by the coefficient, transitions between the reconstructed flow from the previous year (−1) to the concurrent year (0) for both locations. For the Logan River (Fig. 3a), this transition was gradual. The crossover point occurred in March, while the highest model coefficients occurred during the peak and falling limb of the seasonal hydrograph (June through September), before being carried over into the next water-year. The Bear River had a much more abrupt transition (Fig. 3b) between the previous water-year, which explained October through March, and the concurrent water year, which explained monthly flow from June through the end of the water year. As expected, the future (+1) water-year reconstruction had little to no predictive power for monthly flows.

4.4. Tree-ring chronology analysis

From the imputed regional tree-ring dataset, which included all 49 chronologies (Fig. 1), 8 PCs were retained based on minimizing MSE from repeated 10-fold cross-validation. This result was similar to the result from a traditional PCA and a smaller (36 chronologies) dataset, which recommended retaining 8 PCs based on the Kaiser-Guttman criterion. As expected, the smaller 600-year, 19-chronology dataset had a smaller range of species and sites, which resulted in the retention of only 5 PCs. Retaining 8 PCs with imputation was therefore deemed reasonable.

In all tests, the first PC (PC1) explained a large proportion of the total variance, from 38.4% in the imputed dataset to 33.9% and 35.7% in the 400- and 600-year datasets, respectively. The proportion of variance explained by subsequent PCs fell after PC1, from 7.8% to 2.7% between PCs 2 and 8, and ultimately explained 73.1% of the total variance. Loadings for the imputed dataset were nearly identical to loadings from the traditional PCA based on spatial and species patterns, which further validated the imputation scheme.

While a total of eight PCs were considered as predictors, focus was given to the first five, as these represented a majority of the explained variance. PC1 was the only PC where all chronologies loaded in the same direction (Fig. 4). Because of this common pattern and the relatively large variance explained, PC1 appeared to characterize the general dry or wet state of the entire region. Loading for PC1 was dominated by Douglas-fir from low-elevation sites (Fig. 4). However, this does not mean that Douglas-fir was necessarily best at capturing the region's hydroclimatology. For example, both annual reconstructions (Allen et al., 2013; DeRose et al., 2015) originally considered Douglas-fir chronologies, but instead relied on other species that better explained MAF. Instead, Douglas-fir likely dominated PC1 because they have good predictive skill, and were the best represented species (43%) among the available chronologies, covering the largest spatial extent.

While PC1 explained regional hydroclimatology and a major portion of the variability among all chronologies, subsequent PCs captured smaller spatial and species-specific patterns and deviations from the larger climate signal (Fig. 4). PC loadings are shown spatially and by species in Fig. 4 and were summarized as:

- PC1: General hydroclimatology, as captured by Douglas-fir at lower elevations
- PC2: Dipole between Bear River Douglas-fir and Southern Utah pinyon/ponderosa pine
- PC3: High-elevation sites, dominated by two limber pine and two Douglas-fir chronologies
- PC4: East/west dipole between Utah juniper species and Wyoming Douglas-fir/limber pines
- PC5: Dipole between juniper sites near the Logan River and a singleleaf pinyon site in the Great Basin.

Subsequent PCs (Supplemental figures) become more site-specific and explained less variance.

4.5. APR model with additional reconstructions

More complex APR models were fitted, first adding ENSO indices to the lagged annual flow reconstructions, and then adding regional tree-ring PCs, termed "all predictors". Model coefficients for the lagged annual flow reconstruction had nearly identical

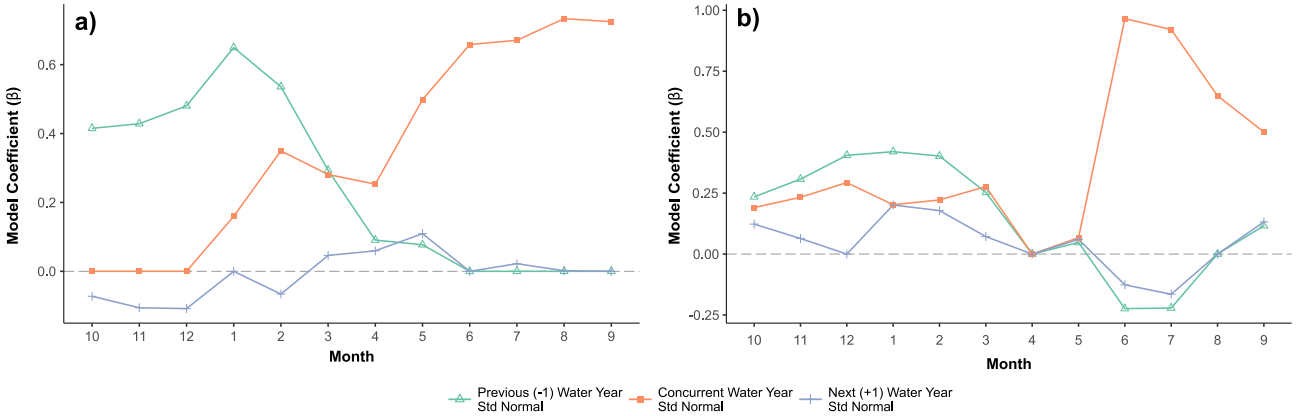

Fig. 3. APR model coefficients for concurrent and lagged (−1, +1) years at the (a) Logan and the (b) Bear rivers.

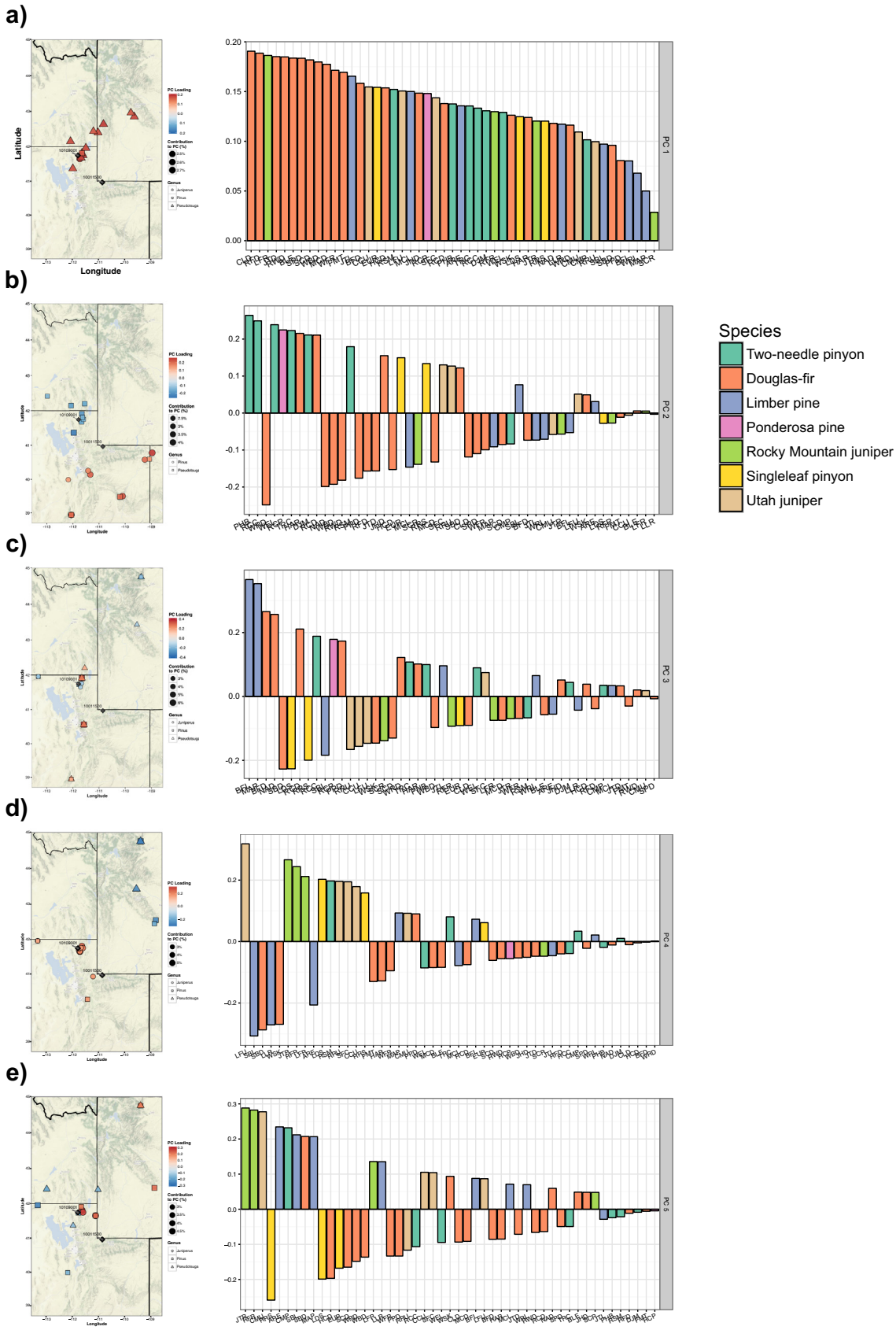

Fig. 4. PC loadings from imputed values presented spatially (left) and by species (right). Loadings are shown for PCs 1–5 (a–e). Spatial maps only present sites with loadings greater than the mean to prevent over-plotting.

shapes as those presented in Section 4.3 and Fig. 3, with a transition from the prior year's reconstruction to the current year, crossing in spring (March-May). This transition was more gradual for the Logan River and more dramatic for Bear River.

The Logan and Bear River sites had broadly similar responses to the two ENSO reconstructions. At both locations, ENSO 3.4 reconstructed using Pacific Ocean proxies (Li et al., 2013) produced a negative response, beginning in January (Fig. 5). This effect continued throughout the winter, spring, and summer for the Logan River (Fig. 5a), while its effect was absent during the spring for the Bear River reconstruction (Fig. 5c). The negative sign of this coefficient suggests that estimated flows at both sites should be decreased when November through January ENSO 3.4 reconstructed by Pacific proxies is strongly positive, and vice versa. ENSO3 as quantified by tree-rings from the North American Drought Atlas (NADA) (Li et al., 2011) amplified the effect of the Pacific Proxies during March and April, but contributed an opposite effect during the summer (June through August) (Fig. 5a and c). See Section 5.2 for a more detailed discussion of these model predictors.

For both sites, PC1 had an insignificant effect as a predictor within the APR model. The only significant regional tree-ring predictor during the fall (October-December) was PC3 (Fig. 5b and d). During the winter, PCs 6 and 7 provided additional information for both sites. PC4 was the most important component during the winter for the Bear River site (Fig. 5b), but does not appear in the Logan model (Fig. 5d). Spring showed little consistency among PC coefficients at both sites. Several regional tree-ring signals were important during the summer, with components 2, 5, and 7 providing the largest and most consistent effects at both sites (Fig. 5b and d). PC8 displayed a unique effect during the summer, with a strongly negative coefficient for the Logan River site and a weakly positive coefficient for the Bear River site. Lagged PC effects were initially considered, using lags of −2 to +2 years, but were ultimately removed due to weak explanatory skill and concerns about overfitting.

4.6. Goodness-of-fit

The APR model with all predictors produced the best fit across all goodness-of-fit measures, which included analyses of the full time-series, monthly fit, residual patterns, and visual inspection. Considering the full time-series, goodness-of-fit improved for nearly all measures (MAE, RMSE, NSE, R) with increasing numbers of parameters, except for mean error (Table 1). Correlations (R) for the APR model with all predictors and the observed MAF was 0.97 for the Logan River and 0.96 for the Bear River. This represents an assumed best possible fit, given that observed annual flows were exactly known. In all cases, using annual flows reconstructed from tree-rings lowered the goodness-of-fit, decreasing R to 0.93 and 0.94, for the Logan and Bear Rivers, respectively. For many fit measures, as additional parameters were added to the model, not only did the overall fit improve, but the gap between observed annual flows and reconstructed annual flows decreased. The MF model is unique in that it excelled at minimizing ME, i.e. systematic bias, but tended to perform only moderate well to poorly for other measures.

Goodness-of-fit at the monthly scale provided more detailed information about model performance. Using the NSE as a goodness of fit measure, it was clear that the MF and AP models have little predictive power (NSE < 0) for much of the early water year, from October until late winter or spring (Fig. 6). The MF model performed particularly poorly during this period. There was a rapid increase in model skill during May and June, when the majority of annual flow volume occurred (Fig. 2). The MF and AP models performed similarly during this period, with the MF model

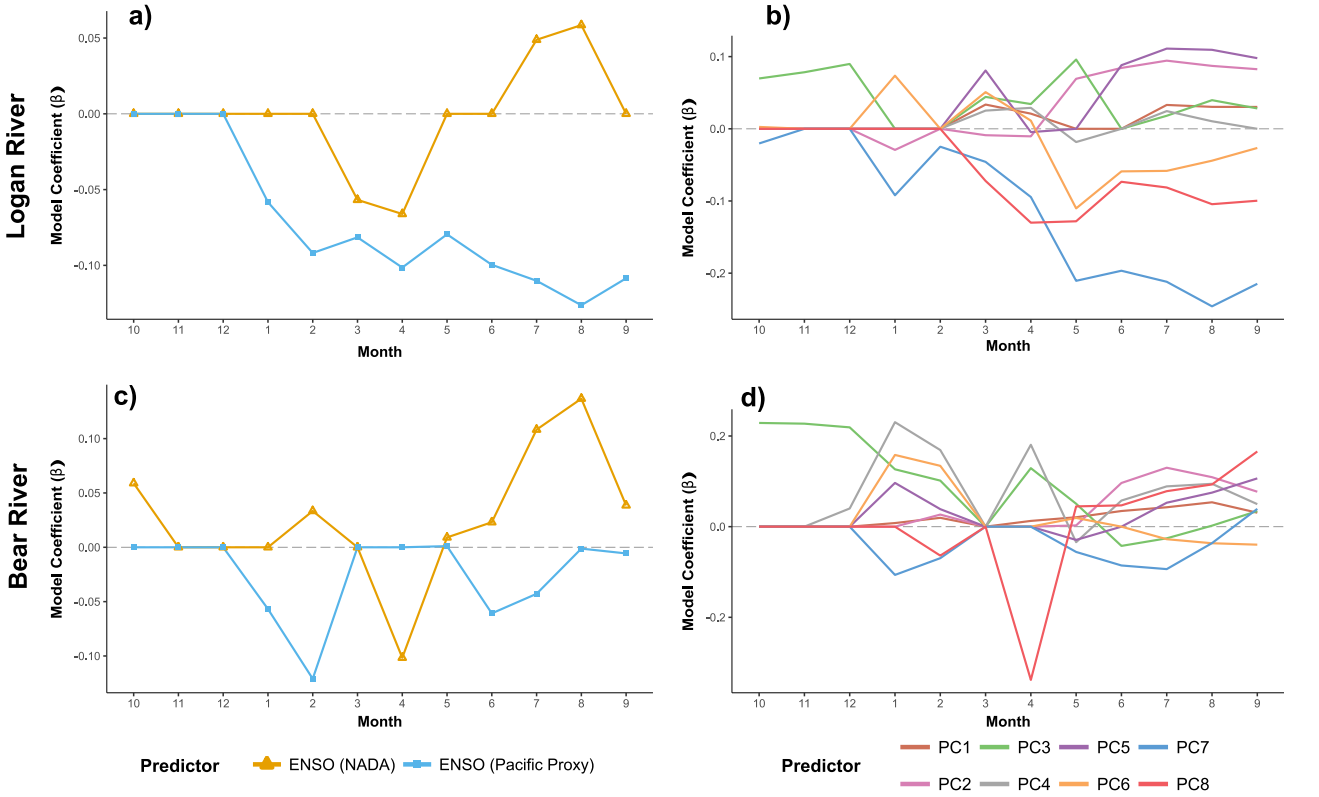

Fig. 5. Model coefficients for the Logan River (a and b) and the Bear River (c and d). Global circulation indices are shown on the left (a and c), while tree-ring PCs are shown on the right (b and d).

Table 1
Annual goodness of fit for Logan River. ME represents mean error, MAE represents mean absolute error, RMSE represents root mean squared error, NSE represents Nash-Sutcliffe Efficiency, R represents Pearson correlation, and $R_{Spearman}$ represents Spearman's rank correlation.

Model	MAF	ME (m^3/s)	MAE (m^3/s)	RMSE (m^3/s)	NSE	R	$R_{Spearman}$
MF	Observed	0.008	1.30	2.00	0.896	0.950	0.870
AP	Observed	0.082	1.16	1.87	0.910	0.957	0.880
APR + Lags	Observed	−0.004	1.02	1.70	0.924	0.963	0.913
APR + Clim Ind	Observed	−0.032	1.00	1.67	0.928	0.964	0.913
APR + All Predictors	Observed	−0.026	0.94	1.62	0.932	0.966	0.925
MF	Reconstructed	0.007	1.70	2.83	0.794	0.893	0.857
AP	Reconstructed	−0.088	1.66	2.93	0.779	0.888	0.846
APR + Lags	Reconstructed	−0.033	1.59	2.91	0.781	0.890	0.885
APR + Clim Ind	Reconstructed	−0.035	1.52	2.73	0.808	0.903	0.890
APR + All Predictors	Reconstructed	−0.008	1.35	2.35	0.858	0.928	0.902

performing slightly better for reconstructed annual flows (Fig. 6b and d). NSE remained high for the remainder of the water year.

Removing the pre-defined water year in the APR model with lagged annual reconstruction predictors represented the largest structural improvement among the models (Fig. 7). Improvement by the APR model occurred primarily during the early part of the water year, from October until March or April (Fig. 7). April was unique in its consistently poor reconstruction skill across all models.

Expanding the APR model to include global climate indices and regional tree-rings produced smaller, but still significant improvements in model fit. The largest improvements occurred in the Bear River, where inclusion of these additional predictors improved fits for all months, though still showing relatively poor fits during the spring season (Fig. 6d). Fit for the Logan River was quite good without these additional predictors, with the most noticeable improvement occurring during the difficult spring months and the summer (Fig. 6b). Fit during October through January on the Logan River site remained relatively low in quality, though still much improved over the AP or MF models. When the model predictors were added successively (Fig. 7), the ENSO reconstructions provide a smaller incremental improvement than the regional tree-rings.

A sample of the historical reconstruction for both sites is presented in Fig. 8, highlighting a period of severe regional drought during the 1730s CE (Woodhouse and Brown, 2001). The entire reconstructed timeseries is available online athttp://www.

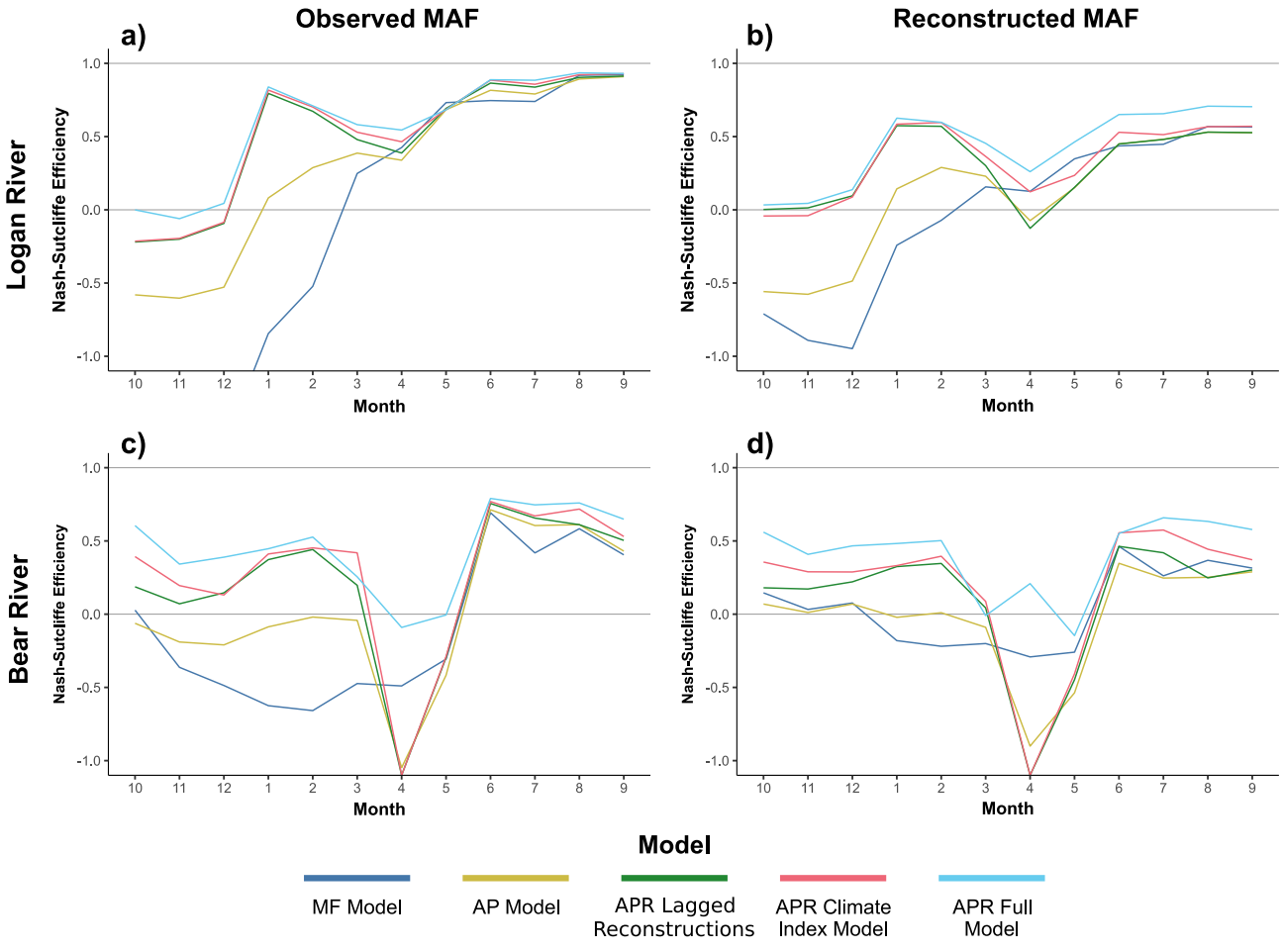

Fig. 6. Nash-Sutcliffe Efficiency (NSE) by month.

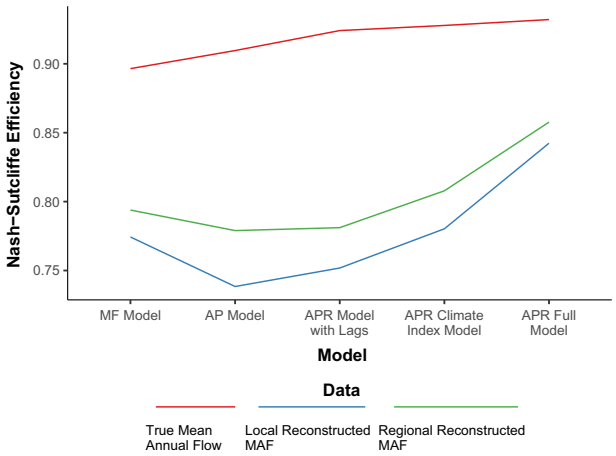

Fig. 7. Nash-Sutcliffe Efficiency (NSE) for the Logan River. Models of increasing complexity are presented along the horizontal axis, while colors represent the input reconstruction dataset.

paleoflow.org. By instead focusing on the observed period and visually comparing reconstructed to observed flows (Fig. 8b), it was possible to confirm the good fit of the APR model and to identify structural errors in the MF and AP models. In the MF and AP models, the transition between water years was not smooth and sometimes produced drastic jumps in flow, particularly between years with very different annual reconstructed flow, such as between 1961 and 1962 CE on the Logan River (Fig. 8b). This shift from a dry to wet year produced a sudden increase in flow at the water year transition, prior to the high flow period recorded in the observations. This type of error was not present in the APR models due to the smooth transition between annual reconstructions. Including all predictors in the APR model smoothed annual transitions and improved the fit and shape of the rising and falling hydrograph limbs (Fig. 8b and d), despite only producing a modest improvement in monthly and annual fit metrics. This finding was further supported by reduced residuals for medium and low flows in the non-peak season when using regional tree-rings and global climate indices.

Without post-processing, the regression equations often reduced the variance of the estimated monthly flow normalized anomalies. The resulting variance ranged from 0.24 to 0.91 across all months, with median values of 0.71 and 0.66 for the Logan and Bear River sites, respectively. This resulted in underestimation of extreme high flows and overestimation of extreme low flows. Increasing variance to 1 through post-processing greatly improved the fit for extreme flows and the resulting flow distribution, while slightly decreasing overall goodness of fit.

5. Discussion

5.1. Reconstruction model comparison

Of the models proposed in this study, the APR model provides the most favorable measures of goodness-of-fit. This model produces NSE greater than zero (0.26–0.60) for all months except April and Pearson's correlation coefficients (R) are 0.94 and 0.88 for the Bear and Logan Rivers, respectively, confirming that the model can adequately reproduce monthly reference period flows. Goodness-of-fit decreases slightly when using reconstructed MAF, rather than observed MAF, but this decrease is minor and reasonable given the challenge and potential benefit of reconstructing centuries of streamflow at a monthly scale.

The MF model is the simplest approach, but it makes a strong assumption about seasonal flows, namely that the hydrograph shape is constant and thus independent of annual flow. Because of this assumption, the MF model accurately captures flows during the peak, but tends to produce errors in timing and magnitude for

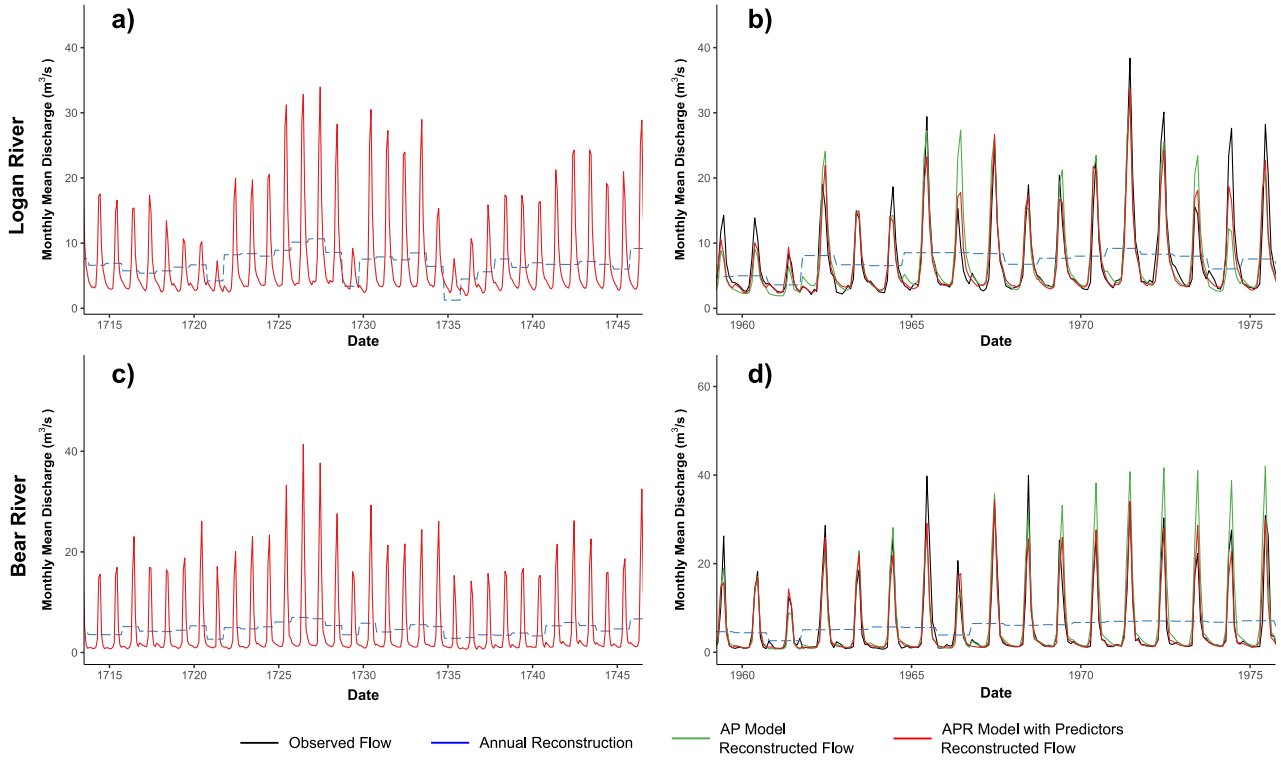

Fig. 8. Reconstructed flows at the Logan River site for subsets of the (a) historical and (b) observed periods, and the same for (c) historical and (d) observed periods at the Bear River site.

all other parts of the year. This is particularly an issue for low flows, which are scaled linearly with annual flow. This assumption is not reasonable because groundwater sources typically provide a buffer to extreme low flows. Additionally, the MF model produces artifacts at the transitions between water years. The only metric for which the MF model excels is mean error, a measure of consistent model bias. This is because the MF model relies on the seasonal mean proportion, which in turn ensures that reconstructed flows are evenly distributed around mean flows, even if accuracy is poor.

The AP model addresses the issue of hydrograph shape, allowing the shape to change with annual flow percentile, but it retains the limitation of applying the same percentile across the entire water-year. This also produces calculation artifacts at the water-year transitions and scaling errors that propagate backwards in time from the peak season. This results in little or no predictive power prior to the peak flow season. Another potential limitation of the AP approach is the potential for generating negative flows if the univariate distribution used to transform flows does not have a lower bound. Here, we have prevented this by selecting a distribution that is strictly positive.

The introduction of regression in the APR model provides the most significant improvement in reconstruction quality. With this improvement, the gap in NSE decreased between models using reconstructed MAF and models using observed MAF. Models using observed MAF have perfect flow estimates at the annual scale and were therefore considered the likely maximum goodness-of-fit. For the sites examined here, using a flexible regression approach allowed fall and winter low flows to retain information from the prior flow peak, rather than depending on the peak flow yet to occur. Adding additional predictors, such as global climate indices and regional tree-ring signals provides further model improvement, particularly in difficult to reconstruct months. Regional tree-rings were the second most important set of predictors, followed by global climate indices, defined here by two ENSO reconstructions. When considering the use of these models for other sites, the modeler must consider the potential trade-off between improved model fit and reconstruction length because the reconstruction will be limited by the length of the shortest reconstruction. The approach provided here attempted to minimize this limitation for the set of regional tree-rings through a process of PCA imputation.

It should be noted that correlation (R) values shown here for the full time series are not directly comparable to R values from annual reconstructions because some of the variance explained is due to capturing the seasonal pattern of flows. Still, the R values are exceptionally good and further supported by excellent model skill measured for individual months.

The APR model had difficulty modeling flows at both sites during the months of April and May. During these months, all coefficients for external predictors were either set to zero or the coefficients became large and opposed. These are the two opposed behaviors of LASSO and ridge regression, respectively, within the elastic net framework as one or the other attempts to handle poor predictive skill. For the Bear River in particular, none of the reconstructed climate indices or regional tree-ring chronologies showed any predictive capabilities in April, including the lagged annual reconstruction. This is likely because April and May occur at the beginning of the snowmelt season when minor changes in temperature near 0 °C and solar radiation can produce large changes in snowmelt, runoff, and flow. Rapid changes in temperature that likely drive snowmelt timing and amount are unlikely to be directly captured in growing season ring-width indices such as those used in this study. For future implementations of the APR model, it will be important to consider additional climatic predictors to help improve the explanation of spring snowmelt timing.

5.2. Interpretation of model predictors

In addition to providing the best model fit, the inclusion of external predictors in the APR model provides an opportunity to explore the effect of global teleconnections and regional hydroclimatology on local streamflow as a form of validation. However, it is important to note this study was not designed to test these links or to isolate their effects. All predictors were considered simultaneously in the APR model and therefore, if the effect of a given driver is accounted for by another measure, it would not appear to be significant. For instance, if the effect of ENSO is already accounted for in the annual MAF reconstruction or modeled better by the regional patterns of tree-rings, it would not appear in the final model.

The canonical understanding of ENSO effects on western U.S. climate is that positive ENSO index values are associated with a warmer, drier winter in northern states and a wetter, cooler winter in the south (Cayan et al., 1999; Redmond and Koch, 1991). The Logan and Bear River sites are situated near the geographic transition between the drier north and wetter south impacts of positive ENSO (DeFlorio et al., 2013; Zhou et al., 2014). However this traditional understanding of the teleconnection does not always hold because it can be modified by other, more local effects. Additionally, recent studies have questioned ENSO's effect across large portions of the western U.S. (Malevich and Woodhouse, 2017). For all of these reasons, ENSO was screened as a potential predictor alongside regional and local tree-rings using multiple regression.

The APR model results associated with the Li et al. (2013) reconstruction using Pacific Ocean proxies were consistent with prior studies (Allen et al., 2013; Cayan et al., 1999; Redmond and Koch, 1991; Kurtzman and Scanlon, 2007). The Logan River exhibited a negative relationship to ENSO due to its northern location which began in the winter and continued throughout the year. The Bear River site's more southern location exhibited only a minor winter response. While the lack of strong ENSO relationship in the Bear River reconstruction can partially be attributed to its latitude, its unique orientation in the watershed may also play a role. The Bear River site is located along the north slope of the Uinta Mountains, the most prominent east–west-oriented mountain in the United States, and which exceeds elevations of 4,000 m. This large mountain range receives substantial winter snow due to orographic effects on Pacific westerlies and can block or redirect southern storm tracks.

Both sites showed a consistent and opposite response during summer to the ENSO3 reconstruction based on the NADA summer Palmer Drought Severity Index (PDSI) reconstruction (Li et al., 2011). Based on the summer effect, which coincides with the underlying NADA summer (JJA) reconstruction, we hypothesize that the contribution of the Li et al. (2011) reconstruction is more strongly related to capturing some of the NADA tree-ring signal than truly measuring an ENSO3 teleconnection on summer flows. It should be noted that ENSO winter effects for both reconstructions are slightly delayed from their ENSO definition. Li et al. (2013) reconstructs ENSO3.4 during November through January and this reconstruction enters the flow model in January, while Li et al. (2011) reconstructs ENSO3 for January through March and enters the flow model in March and April. This provides greater confidence that there is a true influence rather than modeling noise.

The Pacific Decadal Oscillation (PDO, Mantua and Hare, 2002) was initially considered as a third predictor because it has a demonstrated link to northern Utah climate over the recent observed climate record, particularly when considered in conjunction with ENSO (Wang et al., 2009b; Wang et al., 2009a). Inclusion of a PDO reconstruction predictor (Biondi et al., 2001) increased model fit significantly during the fall; however this teleconnection was ultimately excluded from this study because other PDO

reconstructions show little agreement – and thus high uncertainty – outside the observed period (Newman et al., 2016). Further, MacDonald and Case (2005) concluded that while ENSO reconstructions showed a persistent effect on northern Utah climate, the multidecadal variability of the PDO for the past two hundred years has not been stable. Therefore, while future studies could examine the link between Logan and Bear River flow during fall and tree-rings along the west coast of the United States and Mexico, it was deemed too uncertain for consideration here.

Regional tree-ring predictors were also useful for validating model results. PC1 was likely not important for explaining regional wet/dry variability because this measure of overall wetness is already captured by the annual flow reconstructions in the model. PC2 was an important predictor for monthly summer flow at both sites. The spatial pattern of PC2 mirrors the ENSO dipole, where northerly sites (predominantly Douglas-fir) responded in an opposite manner to more southerly species, which were primarily two-needle pinyon, but also ponderosa pine and Douglas-fir. This counterintuitive result is best explained by considering that the annual streamflow reconstruction already explains much of the variability associated with tree-ring sites located near the gauges, leaving the north–south chronology dichotomy to mediate baseflow recession, and prevent flows from decreasing excessively during dry years. PC3 directly captured variability in ring width at high-elevation sites and should therefore reflect variation in monthly streamflow not predicted by the more prevalent low-elevation sites. PC3 was most important in late fall and winter (October through January) and again during April and May. This likely represents a rough measure of temperature, signifying the transition to snow in the fall and melt in the spring. PC4, though only a factor in the Bear River winter, indicated an east/west dipole identified in previous studies (Wang et al., 2009b). The other important winter predictors, PCs 6 and 7, relied on a few important tree-rings along the Wyoming border, which appear to be related to snow or winter flow. Summer predictors are more varied spatially, but it is important to note that PC8, which is significant for the Logan reconstruction, relies on several tree-ring chronologies near the Logan site that behave slightly differently from the surrounding area. This can be seen as a fine-tuning predictor.

5.3. Sources of uncertainty

As with all climate reconstructions, the underlying climate proxies represent a source of uncertainty. In this case, uncertainty regarding the tree-ring chronologies was minimized by carefully selecting chronology sites, using multiple samples per tree, and ensuring sufficient replication (sample depth) for each chronology. Other sources of uncertainty include the univariate probability distributions used to normalize flows. It is assumed that reference period flows used to fit these distributions are representative. We have carefully chosen the Logan and Bear River sites to minimize upstream watershed changes that produce discontinuities or non-climate related trends in the observed flow record.

5.4. Implications for water management

The model presented here is the first approach to reconstruct monthly streamflows directly from tree-ring chronologies and climate reconstructions, without routing climate reconstructions through a hydrologic model. With reasonable estimates of monthly flow, extending many centuries back in time, water managers can challenge systems models with a larger range of natural variability in the duration and severity of drought and pluvial events. This could significantly increase the potential for the use of streamflow reconstructions in water management applications, permitting better drought vulnerability simulations and more robust optimization of reservoir operations. For example, engineers are often asked to design systems to address drought events with 100-year return periods in watersheds where the observed flow record is shorter than 100 years. Extrapolation using extreme value analysis can estimate the magnitude of a hypothetical event, but multi-century flow reconstructions likely include more useful and realistic scenarios. Multi-year or decadal periods of drought and successive small drought events can challenge water management operations and reservoir recovery in ways that a single, severe drought event does not. Establishing a long catalog of near-natural flow is also critical when estimating future hydrologic risks under conditions of a non-stationary climate.

Other available methodologies used to reconstruct sub-annual streamflows (Gangopadhyay et al., 2015; Sauchyn and Ilich, 2017) do not use tree-rings to directly estimate monthly flows. Instead, they generate many feasible sub-annual scenarios, either by stochastic hydrology (Sauchyn and Ilich, 2017) or resampling temperature and precipitation as inputs to a water balance model (Gangopadhyay et al., 2015). These approaches have benefits, producing many ensembles to be used for water systems model simulations and ensuring water balance within large watersheds (Gangopadhyay et al., 2015). However, they also have several limitations not present in the APR model. First, by resampling complete water years, the number of potential temperature and precipitation time series are limited to the years in the observed record (Gangopadhyay et al., 2015). If this method were applied to the Logan and Bear Rivers, there would be 88 and 68 potential annual segments, respectively. Using $k = 10$ nearest neighbor resampling, the same years would be repeated across many realizations. This ensures realistic years, but severely limits variation within the range of observed conditions and prevents reasonable extrapolation slightly outside this range. Second, by requiring a fully-developed water balance model, this approach limits its applicability to regions that have adequate gauge data and modeling.

The APR model proposed here has both greater flexibility and lower data requirements, avoiding the need for developing a full hydrologic model, while also allowing a freer range of flows, still constrained by the historical flow distribution. The APR model also provides a direct statistical link between tree-ring proxies and global circulation drivers that can be used for model validation. We assert that the approach outlined here provides several important improvements over the method of Gangopadhyay et al. (2015). At the same time, we note that these two methods are fundamentally different and designed to address different research questions. The Gangopadhyay et al. (2015) method focuses on capturing uncertainty and ranges of potential flow, while the APR method proposed here focuses on a single best estimate.

Given the value of monthly streamflow reconstructions for water managers and the potential growth in this new field of research, it is important for future studies to test the global applicability of both methods by evaluating their use outside the western U.S. This study, as well as the prior efforts to reconstruct monthly flows, has focused on a relatively narrow geographical area of the western U.S. Given the potential value for monthly reconstructions in water resources and the availability of climate-sensitive tree-ring chronologies elsewhere in the world, there is an important opportunity to test whether these approaches can be applied globally.

6. Conclusions

A new framework for generating monthly streamflows directly from annual reconstructions was introduced and demonstrated using two sites in northern Utah. Several potential models were

evaluated, beginning with a simple monthly flow ratio (MF model), introducing a link between annual and monthly percentiles (AP model), allowing lagged annual flows to predict monthly percentiles (APR model), and ultimately including regional and global indices as regressors (APR model with predictors). Of the different models tested, the APR model was shown to successfully reproduce flows across nearly all months and across the entire range of flows. This model can be applied using only data from annual flow reconstructions or can incorporate regional tree-ring chronologies and global climate index reconstructions. Adding these additional predictors was shown to improve predictive skill, while also providing insight with regard to streamflow drivers. More simple model variants were presented as a comparison and to highlight potential modeling challenges, such as sudden flow jumps at water year transitions.

While annual paleostreamflow reconstructions have great potential for use in water resources planning (Tingstad et al., 2014; Rice et al., 2009; Woodhouse and Lukas, 2006; Axelson et al., 2009), their annual resolution is a potential limitation. Monthly reconstructions, rather than annual, are therefore extremely valuable for drought vulnerability planning. This study shows that monthly flow reconstructions are feasible and presents a new model that makes monthly paleostreamflow reconstruction more tractable. This will hopefully improve the utility and accessibility of these reconstructions for water managers, systems analysts, and decision makers.

Acknowledgements

This work was funded by Utah Mineral Lease funds.

This article was prepared in part by an employee of the USDA Forest Service as part of official duties and is therefore in the public domain.

We would like to acknowledge two anonymous reviewers whose comments helped to improve this paper. We would also like to acknowledge the many dendrochronologists who provided tree-ring data to the International Tree-Ring Data Bank: Matt Bekker, Stephen Gray, Erika Wise, Connie Woodhouse, and Jeff Lukas.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, athttps://doi.org/10.1016/j.jhydrol.2017.12.057.

References

Allen, E.B., Rittenour, T.M., DeRose, R.J., Bekker, M.F., Kjelgren, R., Buckley, B.M., 2013. A tree-ring based reconstruction of Logan River streamflow, northern Utah. Water Resour. Res. 49, 8579–8588. https://doi.org/10.1002/2013WR014273.

Axelson, J.N., Sauchyn, D.J., Barichivich, J., 2009. New reconstructions of streamflow variability in the South Saskatchewan River Basin from a network of tree ring chronologies, Alberta, Canada. Water Resour. Res. 45. https://doi.org/10.1029/2008WR007639.

Biondi, F., Gershunov, A., Cayan, D.R., 2001. North pacific decadal climate variability since 1661. J. Clim. 14, 5–10. https://doi.org/10.1175/1520-0442(2001)0140005:NPDCVS2.0.CO;2.

Bonin, D.V., Burn, D.H., 2005. Use of tree ring reconstructed streamflows to assess drought. Can. J. Civ. Eng. 32, 1114–1123. https://doi.org/10.1139/l05-069.

Brubaker, L.B., 1980. Spatial patterns of tree growth Anomalies in the pacific northwest. Ecology 61, 798–807. https://doi.org/10.2307/1936750.

Bunn, A.G., 2008. A dendrochronology program library in R (dplR). Dendrochronologia 26, 115–124. https://doi.org/10.1016/j.dendro.2008.01.002.

Cayan, D.R., Redmond, K.T., Riddle, L.G., 1999. ENSO and hydrologic extremes in the Western United States. J. Clim. 12, 2881–2893. https://doi.org/10.1175/1520-0442(1999) 0122881:EAHEIT2.0.CO;2.

Cook, E.R., 1987. The decomposition of tree-ring series for environmental studies. Tree-Ring Bull.

Cook, E.R., Briffa, K.R., Jones, P.D., 1994. Spatial regression methods in dendroclimatology: a review and comparison of two techniques. Int. J. Climatol. 14, 379–402. https://doi.org/10.1002/joc.3370140404.

Cook, E.R., Kairiukstis, L.A., 2013. Methods of dendrochronology: applications in the environmental sciences. Springer Sci. Business Media.

Cook, E.R., Meko, D.M., Stahle, D.W., Cleaveland, M.K., 1999. Drought reconstructions for the continental United States. J. Clim. 12, 1145–1162. https://doi.org/10.1175/1520-0442(1999)0121145:DRFTCU2.0.CO;2.

Criss, R.E., Winston, W.E., 2008. Do Nash values have value? Discussion and alternate proposals. Hydrol. Process. 22, 2723–2725. https://doi.org/10.1002/hyp.7072.

Crutcher, H.L., 1975. A note on the possible misuse of the Kolmogorov-Smirnov test. J. Appl. Meteorol. 14, 1600–1603. https://doi.org/10.1175/1520-0450(1975)0141600:ANOTPM2.0.CO;2.

Cullen, A.C., Frey, H.C., 1999. Probabilistic techniques in exposure assessment: a handbook for dealing with variability and uncertainty in models and inputs. Springer Sci. Business Media.

DeFlorio, M.J., Pierce, D.W., Cayan, D.R., Miller, A.J., 2013. Western U.S. extreme precipitation events and their relation to ENSO and PDO in CCSM4. J. Clim. 26, 4231–4243. https://doi.org/10.1175/JCLI-D-12-00257.1.

Delignette-Muller, M.L., Dutang, C., 2015. fitdistrplus: an R package for fitting distributions. J. Stat. Softw. 64, 1–34. URLhttp://www.jstatsoft.org/v64/i04/.

DeRose, R.J., Bekker, M.F., Wang, S.Y., Buckley, B.M., Kjelgren, R.K., Bardsley, T., Rittenour, T.M., Allen, E.B., 2015. A millennium-length reconstruction of Bear River stream flow, Utah. J. Hydrol. 529 (Part 2), 524–534. https://doi.org/10.1016/j.jhydrol.2015.01.014.

Douglass, A.E., 1941. Crossdating in dendrochronology. J. Forest. 39, 825–831.

Duvick, D.N., Blasing, T.J., 1981. A dendroclimatic reconstruction of annual precipitation amounts in Iowa since 1680. Water Resour. Res. 17, 1183–1189. https://doi.org/10.1029/WR017i004p01183.

Fritts, H., 2012. Tree Rings and Climate. Elsevier.

Gangopadhyay, S., Harding, B.L., Rajagopalan, B., Lukas, J.J., Fulp, T.J., 2009. A nonparametric approach for paleohydrologic reconstruction of annual streamflow ensembles. Water Resour. Res. 45. https://doi.org/10.1029/2008WR007201.

Gangopadhyay, S., McCabe, G.J., Woodhouse, C.A., 2015. Beyond annual streamflow reconstructions for the Upper Colorado River Basin: a paleo-water-balance approach. Water Resour. Res. 51, 9763–9774. https://doi.org/10.1002/2015WR017283.

Guttman, L., 1954. Some necessary conditions for common-factor analysis. Psychometrika 19, 149–161. https://doi.org/10.1007/BF02289162.

Hidalgo, H.G., Piechota, T.C., Dracup, J.A., 2000. Alternative principal components regression procedures for dendrohydrologic reconstructions. Water Resour. Res. 36, 3241–3249. https://doi.org/10.1029/2000WR900097.

Hoerl, A.E., Kennard, R.W., 1970. Ridge regression: biased estimation for nonorthogonal problems. Technometrics 12, 55–67. https://doi.org/10.1080/00401706.1970.10488634.

Jackson, D.A., 1993. Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. Ecology 74, 2204–2214. https://doi.org/10.2307/1939574.

Josse, J., Husson, F., 2012. Selecting the number of components in principal component analysis using cross-validation approximations. Comput. Stat. Data Anal. 56, 1869–1879. https://doi.org/10.1016/j.csda.2011.11.012.

Julie, Josse, Franois, Husson, 2016. missMDA: a package for handling missing values in multivariate data analysis. J. Stat. Softw. 70. https://doi.org/10.18637/jss.v070.i01.

Kaiser, H.F., 1960. The application of electronic computers to factor analysis. Educ. Psychol. Measur. 20, 141–151. https://doi.org/10.1177/001316446002000116.

Kurtzman, D., Scanlon, B.R., 2007. El NioSouthern Oscillation and Pacific Decadal Oscillation impacts on precipitation in the southern and central United States: evaluation of spatial distribution and predictions. Water Resour. Res. 43, W10427. https://doi.org/10.1029/2007WR005863.

Li, J., Xie, S.P., Cook, E.R., Morales, M.S., Christie, D.A., Johnson, N.C., Chen, F., D'Arrigo, R., Fowler, A.M., Gou, X., Fang, K., 2013. El Nino modulations over the past seven centuries. Nat. Clim. Change 3, 822–826. https://doi.org/10.1038/nclimate1936.

Li, J.B., Xie, S.P., Cook, E.R., Huang, G., D'Arrigo, R., Liu, F., Ma, J., Zheng, X.T., 2011. Interdecadal modulation of El Nino amplitude during the past millennium. Nat. Clim. Change 1, 114–118. https://doi.org/10.1038/Nclimate1086.

Loaiciga, H.A., Haston, L., Michaelsen, J., 1993. Dendrohydrology and long-term hydrologic phenomena. Rev. Geophys. 31, 151–171. https://doi.org/10.1029/93RG00056.

MacDonald, G.M., Case, R.A., 2005. Variations in the Pacific Decadal Oscillation over the past millennium. Geophys. Res. Lett. 32, L08703. https://doi.org/10.1029/2005GL022478.

Malevich, S.B., Woodhouse, C.A., 2017. Pacific SSTs, mid-latitude atmospheric circulation, and widespread interannual anomalies in Western US streamflow. Geophys. Res. Lett. https://doi.org/10.1002/2017GL073536.

Mantua, N.J., Hare, S.R., 2002. The pacific decadal oscillation. J. Oceanogr. 58, 35–44. https://doi.org/10.1023/A:1015820616384.

Maxwell, R.S., Hessl, A.E., Cook, E.R., Pederson, N., 2011. A multispecies tree ring reconstruction of Potomac River streamflow (9502001). Water Resour. Res. 47.

McCuen, R.H., Knight, Z., Cutter, A.G., 2006. Evaluation of the Nash Sutcliffe efficiency index. J. Hydrol. Eng. 11, 597–602. https://doi.org/10.1061/(ASCE)1084-0699(2006) 11:6(597).

Meko, D.M., Friedman, J.M., Touchan, R., Edmondson, J.R., Griffin, E.R., Scott, J.A., 2015. Alternative standardization approaches to improving streamflow reconstructions with ring-width indices of riparian trees. Holocene 25, 1093–1101. https://doi.org/10.1177/0959683615580181.

Meko, D.M., Stockton, C.W., Boggess, W.R., 1980. A tree-ring reconstruction of drought in Southern California. JAWRA J. Am. Water Resour. Assoc. 16, 594–600. https://doi.org/10.1111/j.1752-1688.1980.tb02436.x.

Meko, D.M., Therrell, M.D., Baisan, C.H., Hughes, M.K., 2001. Sacramento river flow reconstructed to A.D. 869 from tree rings. JAWRA J. Am. Water Resour. Assoc. 37, 1029–1039. https://doi.org/10.1111/j.1752-1688.2001.tb05530.x.

Newman, M., Alexander, M.A., Ault, T.R., Cobb, K.M., Deser, C., Di Lorenzo, E., Mantua, N.J., Miller, A.J., Minobe, S., Nakamura, H., Schneider, N., Vimont, D.J., Phillips, A.S., Scott, J.D., Smith, C.A., 2016. The pacific decadal oscillation, Revisited. J. Clim. 29, 4399–4427. https://doi.org/10.1175/JCLI-D-15-0508.1.

Redmond, K.T., Koch, R.W., 1991. Surface climate and streamflow variability in the western united states and their relationship to large-scale circulation indices. Water Resour. Res. 27, 2381–2399. https://doi.org/10.1029/91WR00690.

Rice, J.L., Woodhouse, C.A., Lukas, J.J., 2009. Science and decision making: water management and tree-ring data in the Western United States. JAWRA J. Am. Water Resour. Assoc. 45, 1248–1259. https://doi.org/10.1111/j.1752-1688.2009.00358.x.

Sauchyn, D., Ilich, N., 2017. Nine hundred years of weekly streamflows: stochastic downscaling of ensemble tree-ring reconstructions. Water Resour. Res. https://doi.org/10.1002/2017WR021585.

Schoennagel, T., Veblen, T.T., Romme, W.H., Sibold, J.S., Cook, E.R., 2005. Enso and Pdo variability affect drought-induced fire occurrence in rocky mountain subalpine forests. Ecol. Appl. 15, 2000–2014. https://doi.org/10.1890/04-1579.

Solander, K., Saito, L., Biondi, F., 2010. Streamflow simulation using a water-balance model with annually-resolved inputs. J. Hydrol. 387, 46–53. https://doi.org/10.1016/j.jhydrol.2010.03.028.

Stagge, J., 2017. paleoAPR: R package for monthly paleo streamflow reconstruction. R Package Version 1. https://doi.org/10.5281/zenodo.889893.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B (Methodol.) 58, 267–288. URL:http://www.jstor.org/stable/2346178.

Tibshirani, R., 1997. The lasso method for variable selection in the cox model. Stat. Med. 16, 385–395. https://doi.org/10.1002/(SICI)1097-0258(19970228)16:4385::AID-SIM3803.0.CO;2-3.

Tingstad, A.H., Groves, D.G., Lempert, R.J., 2014. Paleoclimate scenarios to inform decision making in water resource management: example from southern california's inland empire. J. Water Resour. Plann. Manage. 140, 04014025. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000403.

Trenberth, K.E., 1997. The definition of El Nio. Bull. Am. Meteorol. Soc. 78, 2771–2777. https://doi.org/10.1175/1520-0477(1997)0782771:TDOENO2.0.CO;2.

Wang, S.Y., Gillies, R.R., Jin, J., Hipps, L.E., 2009a. Coherence between the great salt lake level and the pacific quasi-decadal oscillation. J. Clim. 23, 2161–2177. https://doi.org/10.1175/2009JCLI2979.1.

Wang, S.Y., Gillies, R.R., Jin, J., Hipps, L.E., 2009b. Recent rainfall cycle in the intermountain region as a quadrature amplitude modulation from the Pacific decadal oscillation. Geophys. Res. Lett. 36, L02705. https://doi.org/10.1029/2008GL036329.

Wang, S.Y., Gillies, R.R., Reichler, T., 2012. Multidecadal drought cycles in the great basin recorded by the great salt lake: modulation from a transition-phase teleconnection. J. Clim. 25, 1711–1721. https://doi.org/10.1175/2011JCLI4225.1.

Wolock, D., McCabe, G., 1999. Effects of potential climatic change on annual runoff in the conterminous United States. J. Am. Water Resour. Assoc. 35, e1350.

Woodhouse, C.A., Brown, P.M., 2001. Tree-ring evidence for great plains drought. Tree-Ring Res. 57, 89–103. URL:http://arizona.openrepository.com/arizona/handle/10150/262536.

Woodhouse, C.A., Gray, S.T., Meko, D.M., 2006. Updated streamflow reconstructions for the Upper Colorado River Basin. Water Resour. Res. 42.

Woodhouse, C.A., Lukas, J.J., 2006. Multi-century tree-ring reconstructions of colorado streamflow for water resource planning. Clim. Change 78, 293–315. https://doi.org/10.1007/s10584-006-9055-0.

Zhou, Z.Q., Xie, S.P., Zheng, X.T., Liu, Q., Wang, H., 2014. Global warming induced changes in El Nio teleconnections over the North Pacific and North America. J. Clim. 27, 9050–9064. https://doi.org/10.1175/JCLI-D-14-00254.1.

Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. J. R. Stat. Soc.: Ser. B (Stat. Methodol.) 67, 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x.