

Water Resources Research

RESEARCH ARTICLE

10.1029/2019WR024901

Key Points:

- A novel Bayesian network model for streamflow reconstructions using the spatial dependence induced by the river network and regional trees is presented
- The spatial network structure allows a substantial reduction in the uncertainty associated with paleo-streamflows
- The spatial Markov model improves upon traditional streamflow reconstruction methods

Correspondence to:

A. Ravindranath,
aravind000@citymail.cuny.edu

Citation:

Ravindranath, A., Devineni, N., Lall, U., Cook, E. R., Pederson, G., Martin, J., & Woodhouse, C. (2019). Streamflow reconstruction in the upper Missouri River basin using a novel Bayesian network model. *Water Resources Research*, 55, 7694–7716. <https://doi.org/10.1029/2019WR024901>

Received 31 JAN 2019

Accepted 2 AUG 2019

Accepted article online 8 AUG 2019

Published online 9 SEP 2019

Streamflow Reconstruction in the Upper Missouri River Basin Using a Novel Bayesian Network Model

Arun Ravindranath¹ , Naresh Devineni² , Upmanu Lall³ , Edward R. Cook⁴ , Greg Pederson⁵ , Justin Martin⁵ , and Connie Woodhouse⁶ 

¹Department of Civil Engineering, NOAA-Center for Earth System Sciences and Remote Sensing Technologies, Center for Water Resources and Environmental Research, City University of New York (City College), New York, NY, USA,

²Department of Civil Engineering, City University of New York (City College), New York, NY, USA, ³Department of Earth and Environmental Engineering, Columbia Water Center, The Earth Institute, Columbia University, New York, NY, USA,

⁴Tree Ring Laboratory, Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA, ⁵Northern Rocky Mountain Science Center, US. Geological Survey, Bozeman, MT, USA, ⁶School of Geography and Development, Laboratory of Tree Ring Research, Department of Geoscience, The University of Arizona, Tucson, AZ, USA

Abstract A Bayesian model that uses the spatial dependence induced by the river network topology, and the leading principal components of regional tree ring chronologies for paleo-streamflow reconstruction is presented. In any river basin, a convergent, dendritic network of tributaries come together to form the main stem of a river. Consequently, it is natural to think of a spatial Markov process that recognizes this topological structure to develop a spatially consistent basin-scale streamflow reconstruction model that uses the information in streamflow and tree ring chronology data to inform the reconstructed flows, while maintaining the space-time correlation structure of flows that is critical for water resource assessments and management. Given historical data from multiple streamflow gauges along a river, their tributaries in a watershed, and regional tree ring chronologies, the model is fit and used to simultaneously reconstruct the full network of paleo-streamflow at all gauges in the basin progressing upstream to downstream along the river. Our application to 18 streamflow gauges in the Upper Missouri River Basin shows that the mean adjusted R^2 for the basin is approximately 0.5 with good overall cross-validated skill as measured by five different skill metrics. The spatial network structure produced a substantial reduction in the uncertainty associated with paleo-streamflow as one proceeds downstream in the network aggregating information from upstream gauges and tree ring chronologies. Uncertainty was reduced by more than 50% at six gauges, between 6% and 50% at one gauge, and by less than 5% at the remaining 11 gauges when compared with the traditional principal component regression reconstruction model.

1. Introduction

The operating rules and water release policies of dams and reservoirs are often based on short streamflow records that span a few decades, whereas the factors driving streamflow variability exhibit long periods of systematic variation. Consequently, long climate records or proxies are needed to extend streamflow data to get insights as to how water supply variability is manifest over long periods of time. For instance, streamflow responds to large-scale atmospheric teleconnection patterns of both high and low frequency (Cayan et al., 1999; Hamlet & Lettenmaier, 2000; Hidalgo & Dracup, 2003; Najibi et al., 2017; Nowak et al., 2012; Redmond & Koch, 1991; Wise et al., 2018). Shorter streamflow records may not be representative of longer-term variability in streamflow even with typical stochastic simulation methods that use the recorded data. Paleo-reconstructions that hindcast streamflow records back in time using annual tree ring chronologies have proven to be useful for understanding the statistics of droughts as well as the recurrence characteristics of and regime shifts between wet and dry periods, or periods with high or low interannual variability.

Streamflow reconstructions using the paleo-climatic information from tree ring chronologies have traditionally been performed using multiple linear regression models, nonparametric methods, and hierarchical Bayesian methods, and we review some of the major literature on this in the next section. Building on this literature, we present a novel approach for streamflow reconstruction from tree ring chronologies. The primary motivation is the observation that streamflow processes on the typical convergent dendritic river

network can be best described by a spatial Markov process. Flow at a downstream gauge can be considered to depend on flow at the most immediate upstream gauges, and an exogenous variable that represents processes that determine the local streamflow input in between the upstream gauges and the downstream gauge of interest. In our context, the exogenous variables are appropriately selected tree ring chronologies. The key innovation here is the inclusion of the spatial network and the corresponding induced dependence structure of streamflow, in a Bayesian framework.

$$\begin{aligned}
 f(A, B, C, D, E, F | T_1, \dots, T_{13}) & \quad (1) \\
 &= f(A|B, C, T_1) \times f(B|T_2, T_3) \times f(C|D, E, F, T_5, T_6, T_7) \\
 & \quad \times f(D|T_8, T_9, T_{10}, T_{11}) \times f(E|T_8, T_9, T_{10}, T_{11}) \times f(F|T_4, T_{12}, T_{13})
 \end{aligned}$$

The right-hand side of the equation is the mathematical factorization of the joint conditional density of streamflow at the gages on the network, given the tree ring chronologies, into a product of conditional densities using a version of the *fundamental rule* (Jensen & Nielsen, 2007), consistent with the physical dependence between streamflow gauges, their feeder gauges and tree ring chronologies (see Figure 1).

In the application presented here, we reconstruct paleo-period streamflow records for streamflow gauges in the Upper Missouri River Basin using this model structure, but with the modification that the leading principal components (Wilks, 1995) of appropriately selected tree ring chronologies T_i for each gauge are used as the information from tree ring chronologies.

Section 2 presents a brief review of much of the seminal work on streamflow reconstructions using paleoclimate proxies. Section 3 presents the Bayesian mathematical model used for inference. Section 4 discusses the case study and how this model is applied to reconstructing streamflow in a given watershed. Section 5 reviews the data used and the processing of the data in addition to the predictor selection method employed. Section 6 provides the reconstructions, and their comparison to reconstructions of flows at the same gauges using a traditional principal components regression. Finally, section 7 concludes and summarizes the paper.

2. Previous Streamflow Reconstruction Efforts and the State of the Art

Many studies have considered the use of paleoclimate data in reconstructing historical climatic data for the non instrumental period, thereby creating extended records of climatic data spanning several centuries. Streamflow has been an important variable of interest in this regard. However, the special spatial structure of streamflow networks has not been used to constrain the reconstruction in any of these studies. This provides an opportunity to use information more effectively than in other reconstruction work and is the focus of our paper.

Streamflow reconstructions have traditionally been performed using multiple linear regression models, or modifications thereof, and the predictors used in these regression models are typically tree ring chronologies or their principal components. The regression models used in these studies include stepwise multiple linear regression (Barnett et al., 2010; Earle, 1993; Watson et al., 2009; Woodhouse, 2001; Woodhouse et al., 2006; Woodhouse & Lukas, 2006); standard multiple linear regression or canonical regression, often in the form of principal components regression (PCR) (Barnett et al., 2010; Cook & Jacoby, 1983; Cook et al., 2013; Day & Sandifer, 2015; Gedalof et al., 2004; Ma et al., 2014; Maxwell et al., 2011; Meko & Graybill, 1995; Meko et al., 2001; Stockton & Jacoby, 1976; Timilsena et al., 2007; Watson et al., 2009); or hierarchical Bayesian regression (Bracken et al., 2016; Devineni et al., 2013; Rao et al., 2018). With the exception of Devineni et al. (2013), Rao et al. (2018), and Bracken et al. (2016), the reconstructions in these studies are done using variations of ordinary least squares (OLS) regression for reconstructing flow at each single-site individually (point-by-point regression) rather than multiple sites simultaneously and accounting for spatial correlation across the sites by explicitly estimating the inherent correlation structure either physically or statistically.

Devineni et al. (2013) and Rao et al. (2018) reconstructed flows at multiple streamflow sites or reservoirs within a basin simultaneously, with explicitly modeled spatial dependence of the flows, to produce

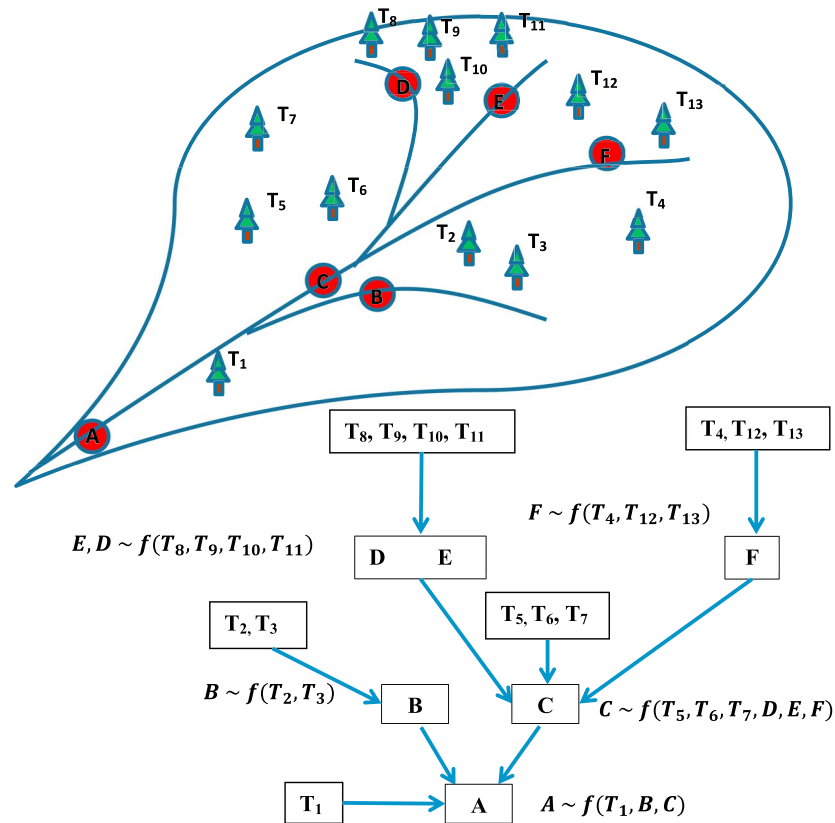


Figure 1. Conceptual sketch of the network Bayesian model. Six streamflow gauges (A–F) and 13 tree ring chronology sites (T_1 – T_{13}) are shown in the sketch in order to illustrate the concept of the graphical network model. Physically informed modeling structure using regional tree ring chronologies and feeder streamflow gauges is explored using factorization into lower dimensional conditional probability distributions as shown in the directed graph. The conditional distributions generated at each stage of the graph serve as statistical interpretations of the modeling structure and lay the groundwork for converting the graphical model into a set of equations for estimating the parameters of the streamflow network’s likelihood function for all gauges (nodes) in the network simultaneously using a Bayesian estimation scheme. Although we allude to the use of regional tree ring chronologies in this figure, in our analysis for the Upper Missouri River Basin, we used the leading principal components of the basin-wide tree predictors.

reconstructions whose outputs were probabilistic in nature; that is, posterior distributions as opposed to conditional means, giving explicit uncertainty estimates in the reconstructions. The hierarchical Bayesian regression used partial pooling to reduce the uncertainty in the reconstruction at each site and across sites. Bracken et al. (2016) followed a similar philosophy for multisite reconstructions, using a Bayesian framework, providing uncertainty information in the distributional reconstructions, using the principal components of tree ring chronologies, and employing a modeling approach that aims to minimize the uncertainty in its estimates. These three papers are the closest in spirit to the current paper.

Although more in line with the methods that produce single-site reconstructions, there are studies that have employed extensions or generalizations of classic regression. Young (1994) used an adaptive, three-way interpolation model that combined the usage of multiple discriminant analysis, multiple linear regression, and normal ratio methods in order to reconstruct streamflow for three gauges in central Arizona over a period of several hundred years using monthly precipitation and annual tree ring chronologies. A study by Meko et al. (2007) used a two-stage linear regression procedure, in which the first stage involved performing a separate regression on each streamflow site, obtaining a single-site reconstruction (SSR) of streamflow for each site and performing a principal component analysis (PCA) on the covariance matrix of the SSRs. Stage two consisted of performing a stepwise regression of the reconstructed streamflow on the scores of the most important principal components (PCs). Patskoski et al. (2015) developed a “hybrid” approach in which sea surface temperature conditions from the tropical Pacific and regional tree ring chronologies from the

watershed itself are used to inform streamflow reconstructions. They used singular spectrum analysis to extract quasi-periodic components from streamflow and Niño 3.4 and non-periodic components from streamflow and tree ring chronologies, so that separate stepwise regressions could be performed for the periodic and non-periodic components of streamflow. Adding the periodic component estimate of streamflow with the non-periodic component estimate of streamflow then gives the reconstruction. This approach was then compared to the more traditional PCR method of streamflow reconstruction. Partial least squares regression has also been considered for streamflow reconstructions (Barnett et al., 2010; Watson et al., 2009).

A k-nearest neighbors (k-*nn*) nonparametric method was used by Gangopadhyay et al. (2009) in order to reconstruct naturalized annual streamflow ensembles from tree ring chronologies in the Upper Colorado River basin. The use of hydrologic/physical models has also been explored (Lutz et al., 2012; Saito et al., 2008). Saito et al. (2008) used a mechanistic watershed model of their own design to reconstruct and project water year streamflow. This was a concerted effort to combine information from tree ring chronology records with watershed modeling in order to produce estimates of streamflow.

Nguyen and Galelli (2018) used a linear dynamic systems modeling approach to streamflow reconstruction in northern Thailand. Ho et al. (2016) demonstrated another modeling approach that was applied to reconstructing streamflow in the Missouri River Basin using a gridded paleo-proxy called the Living Blended Drought Atlas (LBDA). Since the LBDA series have a high degree of spatial correlation, regularized canonical correlation analysis was applied to LBDA and the result used as input to a log-linear reconstruction model.

The approach to streamflow reconstruction developed in this paper is distinct from the approaches reviewed above, which did not consider the network structure of flows in a river basin. Although the approaches of Devineni et al. (2013), Bracken et al. (2016), and Rao et al. (2018) considered spatial correlation structure, the resulting spatial correlation matrix could be as large as $N_s * N_s$, where N_s is the number of sites, and estimating this covariance matrix reliably can be a challenge with finite data as N_s increases. Our method dramatically reduces the number of correlations across sites that need to be modeled in a physically meaningful way through the network-based spatial Markov process.

3. The Model: A Bayesian Spatial Markov Model

The spatial Markov network model structure is specified as follows. We proceed sequentially from the *terminal*, or most downstream, gauge on the main stem of the river to identify its immediate upstream gauges, and repeat this process to identify the Spatial Markov Network consistent with the representation in Figure 1. The resulting joint probability distribution model (see an example of its factorization in equation (1)) is then fit using a Bayesian approach. Subsequently, paleo-streamflow along the network can be simulated by using the conditional distributions from the upstream to the downstream nodes for each year. The ensemble of such draws from the posterior conditional distributions then preserves the spatial dependence of streamflow conditional to the state of the tree ring chronologies for each of the years. This is an advance over independently regressing each station on a set of tree ring chronologies, or trying to fit the full spatial correlation matrix, or using PCA to reduce dimensionality.

For our application, we found that the log of the streamflow at each site was normally distributed and that linear models suffice for describing the conditional relations with upstream flows and with tree ring chronologies over the network. The coefficients of each of the linear models are the *hyperparameters* of the likelihood function for the network.

The Bayesian approach used for parameter estimation considers the unknown model parameters as random variables. In our application, noninformative *prior distributions* are specified for each unknown parameter (Gelman & Hill, 2007). The distribution placed on the parameters is then updated using available data as a means of “training” the model and obtaining the updated *posterior distribution* estimates of the parameters. These posterior estimates of the hyperparameters are then used to calculate the posterior estimates of the likelihood function, as well as the conditional distributions of the streamflow along the network. One can then simulate from these distributions or report the mean values and other statistics. The *Bayesian Spatial Markov*, or BSM model can then be summarized as below in equation (2).

$$\begin{aligned}
 \ln(y_t^{(i)}) &\sim N(\mu_t, \sigma_t^2) \\
 \ln(y_t^{(i)}) &= \begin{cases} \alpha_i + \mathbf{x}_t^{(i)} \boldsymbol{\beta}_i + \sum_{j=1}^m \ln(y_t^{(j)}) \beta_i^{(j)} + \varepsilon_t^{(i)} & \text{if feeder sites apply} \\ \alpha_i + \mathbf{x}_t^{(i)} \boldsymbol{\beta}_i + \varepsilon_t^{(i)} & \text{if feeder sites do not apply} \end{cases} \\
 \varepsilon_t^{(i)} &\sim N(0, \tau_i); \quad \tau_i = \frac{1}{\sigma_t^2}; \quad \sigma_t \sim U(0, 100) \\
 \alpha_i &\sim N(0, \tau_i = 0.001) \\
 \boldsymbol{\beta}_i &\sim MVN(0, 100I) \\
 \beta_i^{(j)} &\sim N(0, 100) \\
 p(\boldsymbol{\theta} | \text{data}) &\propto \prod N \left(\ln(\mathbf{y}_S) \mid \begin{cases} \alpha + \mathbf{x} \boldsymbol{\beta}_T + \boldsymbol{\beta}_S \ln(\mathbf{y}_S) + \varepsilon & \text{if feeder sites apply} \\ \alpha + \mathbf{x} \boldsymbol{\beta}_T + \varepsilon & \text{if feeder sites do not apply} \end{cases} \right) \\
 &* N(\alpha | 0, 1000) * N(\boldsymbol{\beta}_T | 0, 100) * N(\boldsymbol{\beta}_S | 0, 100) * U(\boldsymbol{\sigma}_S | 0, 100)
 \end{aligned} \tag{2}$$

In equation (2), bold quantities represent vectors, $N()$ represents the normal (Gaussian) distribution, $MVN()$ represents the multivariate normal distribution, $U(0, 100)$ represents the Uniform distribution on the interval (0, 100), τ is the reciprocal of the variance and is known as the *precision* of the distribution, I is the identity matrix of appropriate dimension, the subscript t is a time-varying index on the annual resolution (water year), and the superscript (i) indexes the gauge being modeled. The superscript i_j represents the subset of gauges that are immediately upstream and contributing to i , where j runs from 1, ..., m in the situation that m is the number of sites upstream of site i that feed site i . The streamflow data $y_t^{(j)}$ that act as feeder sites for gauge i are weighted by regression coefficients $\beta_i^{(j)}$, where the subscript on these coefficients represent the site being modeled, and the superscript represents the specific upstream feeder gauge. Both cases in equation (2) include an intercept term α_i , the model error term $\varepsilon_t^{(i)}$, a vector of the leading PCs specific to each gauge $\mathbf{x}_t^{(i)}$ as deterministic predictors, and the regression slopes $\boldsymbol{\beta}_i$. While we used the leading principal components of the tree ring chronologies as the predictors for our study region, one can appropriately select local or regional tree ring chronologies depending on the application. The joint posterior likelihood $p(\boldsymbol{\theta} | \text{data})$ of the full parameter vector $\boldsymbol{\theta}$ is given at the end.

4. Case Study: The UMRB

We now provide an application of the model to streamflow records from 18 gauges in the Upper Missouri River Basin (UMRB). Figure 2 shows the portion of the MRB that is under study, along with a demarcation of all 18 streamflow gauges used. The terminal gauge in the network reconstruction scheme is Landusky. The Missouri River Basin provides a case study for which the reconstruction method developed in this paper is applied, evaluated, and cross-validated. The period over which the reconstruction was done is 1800–1989. It was desired to hindcast further into the past, but 1800 was chosen because all tree ring chronologies used began on or before that year, thus restricting our analysis to this common period of record of the tree ring data.

The Missouri River Basin is the second largest drainage basin in the United States, draining about one sixth of the conterminous United States and roughly 9,691 square miles of Canada (Galat et al., 2005). The basin has a watershed area spanning over 500,000 square miles, and the Missouri River, which is the longest river in the United States, produces annual yields of 40 million acre-feet (United States Department of the Interior, Bureau of Reclamation, 2016). The river headwaters are in the Rocky Mountains, where snowmelt is the main source of water. The river flows east across the Great Plains to its confluence with the Mississippi River. The United States Bureau of Reclamation has constructed over 40 dams on the river's tributaries that have positively impacted agricultural development, and the various facilities in the basin provide other benefits as diverse as flood control, navigation, irrigation, power generation, water supply, recreation, fish and wildlife support, ecological and biodiversity support, and water quality (United States Department of the Interior, Bureau of Reclamation, 2016). This basin is particularly interesting, due to its size, importance,

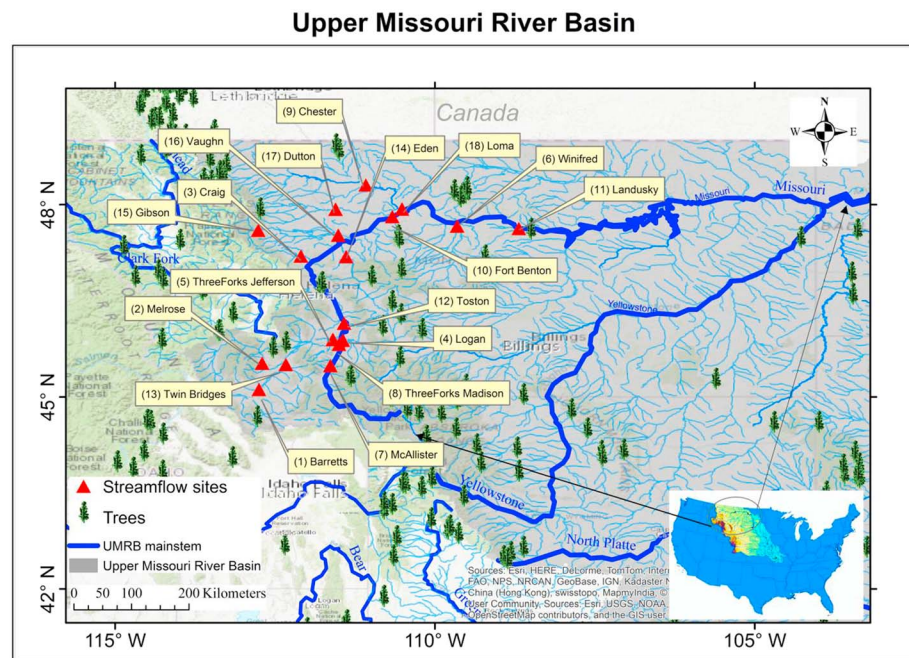


Figure 2. A map of the Upper Missouri River Basin (UMRB) with all 18 streamflow gauges. The red triangles are the streamflow gauges under consideration, which naturally form a network. The original 375 tree ring chronology sites are depicted as trees in the map. The bold blue curve is the main stem of the Missouri River, while the thinner, lighter blue lines are tributaries and smaller order streams. The small inset shows a map of the United States along with the location of the Missouri River Basin, with the UMRB (upper portion of the basin) circled.

and geographic, topographic and climatic complexities. For instance, Wise et al. (2018) found that different seasonal controls affect the upper and lower portions of the basin, and that streamflow and temperature trends were, and in the future will be, quite different between these two portions. Since 1898, when record keeping of streamflow in the UMRB officially began, nine out of 10 of the biggest flood events in the UMRB have occurred after 1970 (Livneh et al., 2016). The investigators found that it was generally wetter regional and seasonal conditions, with respect to the 1895–1974 climatology, coupled with land surface and antecedent soil moisture conditions that often contributed to these flood events (Livneh et al., 2016; Najibi et al., 2017). It is possible that events such as these have occurred in the distant past, and thus, we must extend the historical record in order to see this. It is also necessary to understand how patterns of streamflow variability over time are connected with low-frequency climate modes, which change over long periods of time. This also requires an extended streamflow record. Given the importance of the UMRB, it is a sound application of our reconstruction model. To this end, the following steps were taken in model planning and design:

- Step 1. A map of the complete Missouri River watershed (in this case, MRB) was generated and a portion of the basin was selected for the focus of the analysis. The region of focus here is the upper portion of the watershed (the UMRB). The map is made with the river's main stem, tributaries, streamflow gauges, and tree ring chronologies demarcated on the map as a way of understanding and mapping out the physical dendro-riverine network in full. See Figure 2.
- Step 2. Based on the map generated in step 1 (Figure 2), we consider the streamflow gauges as nodes and direction of streamflow as arrows in order to translate the information presented on the map into a directed graph (digraph) of the streamflow network in the chosen subbasin (UMRB). See Figure 3 for an illustration of the digraph corresponding to the UMRB reconstruction model. This digraph becomes the basis for the network model that will represent the river network structure of the subbasin.
- Step 3. The graphical network model (Figure 3) is translated into a spatial Markov network model, and from there, into a dependence model, as propounded in equation (1). The regression equations that specify the functional relationships between the components of the dependence model are written

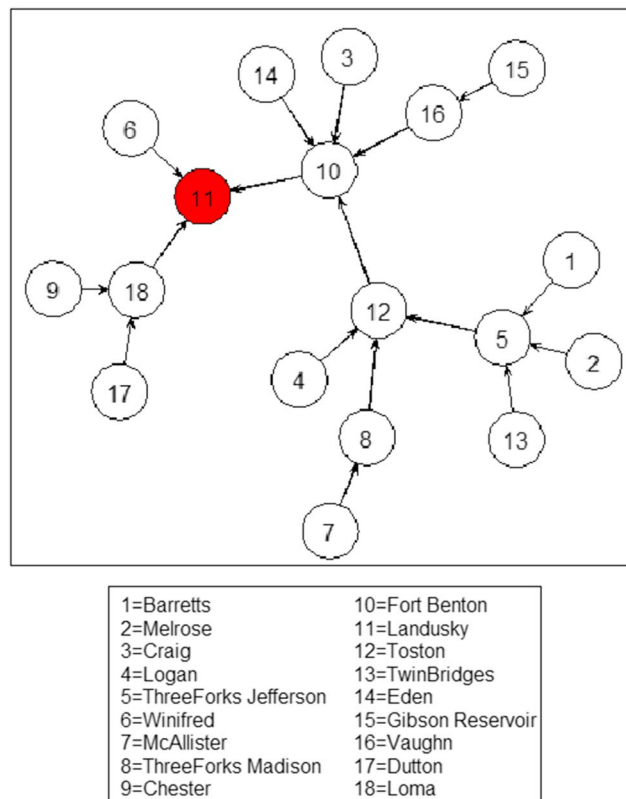


Figure 3. The skeletal structure of the streamflow network, illustrating how the individual streamflow gauges feed one another in the physical riverine network of the Upper Missouri River Basin. The gauges are represented by circles with numbers inscribed. The legend below the network graph gives the gauge names corresponding to the numbers. The terminal gauge (Landusky) is colored in red. The arrows indicate the direction of flow from each gauge, and the gauges that serve as feeders to other gauges, giving a cohesive network structure, can easily be seen from the orientation of these arrows.

out, and a Bayesian estimation scheme is adopted for estimating the model parameters. Predictor selection for these regressions is done as a means of finding the best tree ring chronologies from a host of candidates to inform the flows at each streamflow gauge. The incorporation of feeder gauges as additional predictors is decided based on the digraph from step 2 (Figure 3). The dependence model, which is a joint likelihood of the streamflow gauges and tree ring chronologies in the network under consideration, is then estimated under a Bayesian framework. This amounts to estimating the parameters in the regression equations simultaneously across all of the streamflow gauges in the network. The reconstructed streamflow data are generated from the joint likelihood simultaneously (using a simulation approach) for all gauges in the network with appropriate spatial dependence and using the posterior estimates of the mean and variance parameters, which in turn were estimated based on the training, or observed, streamflow data. In the simulation approach, we first estimate the posterior distribution of the streamflow of the gauges without feeders using the model parameters and the regression equations, and then use the median of these posterior estimates as predictors for the gauges with feeders in order to estimate their posterior streamflow. In doing this, we follow the natural stream order within the river network. Refer to equation (2) in section 2 for the full specification of the equations.

In equation (2), $i = 1, 2, \dots, 18$ for the 18 gauges in the UMRB being modeled in this study. The model parameters are estimated simultaneously across all 18 gauges using the likelihood function. Cross-site correlations as a means of modeling the spatial dependence structure along the river network and the physical relationship across streamflow gauges is implicitly modeled in this framework.

A sample of the regression equations in the model are provided for part of the network terminating at the Landusky gauge. If one refers to Figure 3, and the portion of the network presented therein leading from Dutton (gauge 17) to Loma (gauge 18) to Landusky (gauge 11), the set of equations describing this portion of the network for the purposes of illustrating the streamflow reconstruction are as follows:

$$\begin{aligned}
 & \ln(y_i^{(11)}) \sim N(\mu_i^{(11)}, \tau_{11}) \\
 & \tau_{11} = \frac{1}{\sigma_{11}^2}; \sigma_{11} \sim U(0, 100) \\
 & \mu_i^{(11)} = \alpha_{11} + \ln(y_i^{(6)})\beta_{11}^{(6)} + \ln(y_i^{(10)})\beta_{11}^{(10)} + \ln(y_i^{(18)})\beta_{11}^{(18)} + \mathbf{x}_t^{(11)}\beta_{11} \\
 & \alpha_{11} \sim N(0, 0.001) \\
 & \beta_{11} \sim MVN(0, 100\mathbf{I}) \\
 & \beta_{11}^{(j)} \sim N(0, 100), \text{ for } j = 6, 10, 18 \\
 (3b) \quad & \ln(y_i^{(18)}) \sim N(\mu_{18}, \tau_{18}) \\
 & \tau_{18} = \frac{1}{\sigma_{18}^2}; \sigma_{18} \sim U(0, 100) \\
 & \mu_i^{(18)} = \alpha_{18} + \mathbf{x}_t^{(18)}\beta_{18} + \ln(y_i^{(17)})\beta_{18}^{(17)} \\
 & \alpha_{18} \sim N(0, 0.001) \\
 & \beta_{18} \sim MVN(0, 100\mathbf{I}) \\
 & \beta_{18}^{(17)} \sim N(0, 100) \\
 & \ln(y_i^{(17)}) \sim N(\mu_t^{(17)}, \tau_{17}) \\
 & \tau_{17} = \frac{1}{\sigma_{17}^2}; \sigma_{17} \sim U(0, 100) \\
 & \mu_t^{(17)} = \alpha_{17} + \mathbf{x}_t^{(17)}\beta_{17} \\
 & \alpha_{17} \sim N(0, 0.001) \\
 & \beta_{17} \sim MVN(0, 100\mathbf{I})
 \end{aligned} \tag{3}$$

In equation (3), $\mathbf{x}_t^{(17)}$, denotes the relevant leading principal components of the tree ring chronologies selected for gauge 17 (Dutton). Posterior distributions of the parameters are estimated using the Gibbs sampling algorithm for the Markov Chain Monte Carlo method (Gelman & Hill, 2007; Robert & Casella, 2011) on the basis of the prior distributions assigned. Note that we have omitted the analogous equations for gauge 9, which acts as a feeder to gauge 18, for the sake of brevity.

5. Data, Model Selection and Cross-Validation Procedure

Given the model structure and modeling procedure as outlined in sections 2 and 3, the analysis was carried out using naturalized streamflow data from 18 gauges in the UMRB and tree ring chronology data from the UMRB. In this section, we describe in full the data used, the sources of that data, the predictor/model selection procedure and summarize the results of this selection procedure.

5.1. Streamflow Data

Monthly naturalized streamflow data corresponding to 31 streamflow gauges in the MRB were compiled using estimates of natural streamflow developed by Cary and Parrett (1996), Brekke et al. (2009), and Larry Dolan (MTDNRC, personal communication, 2009). Each streamflow gauge had chunks of missing data in disparate patches. Eighteen stations in the UMRB that make up the longest continuous river network are chosen for the analysis (Figure 2). The monthly data were then aggregated to water year totals; that is, the sum of the monthly streamflow data from 1 October through 30 September of the following calendar year was calculated. Taking the standardized version of the natural logarithms of these water year totals constitutes the predictand of the model. Table 1 displays detailed information on the selected stream gauges.

Table 1
The Basic Data Corresponding to the Streamflow Gauges in the UMRB Streamflow Network Considered in This Paper

Gauge	USGS gauge number	River	Drainage area (km ²)	Years of record	Record length (Years)
Barretts	06016000	Beaverhead	7,071	1928–1989	62
Melrose	06025500	Big Hole	6,402	1930–2002	73
Craig	06073500	Dearborn	834	1947–2015	69
Logan	06052500	Gallatin	4,633	1929–2002	74
Three Forks	06036650	Jefferson	24,755	1929–1989	61
Winifred	06114700	Judith	7,115	1928–1989	62
McAllister	06041000	Madison	5,570	1929–1989	62
Three Forks	06042500	Madison	6,353	1929–2002	74
Chester	06101500	Marias	12,805	1928–1989	62
Fort Benton	06090800	Missouri	62,929	1929–1989	61
Landusky	06115200	Missouri	105,280	1928–1989	62
Toston	06054500	Missouri	37,920	1929–2002	74
Twin Bridges	06023000	Ruby	2,512	1939–2002	64
Eden	06077500	Smith	4,113	1928–1989	62
Gibson Reservoir	06079500	Sun	1,489	1928–1989	62
Vaughn	06089000	Sun	4,595	1928–1989	62
Dutton	06108000	Teton	3,206	1928–1989	62
Loma	06108800	Teton	5,250	1928–1989	62

Note. The columns, from left to right, contain the following information: gauge name, USGS gauge number, the name of the river in which the streamflow gauge is placed, the drainage area of the gauge, the years during which we have the measured/observed water year total streamflow values, and the length of record (i.e. number of years for which the observations are complete).

5.2. Tree Ring Chronology Data

The tree ring chronology data used here come from an overall network of 374 tree ring chronology sites located throughout the MRB (Figure 2) and developed specifically for this project. The tree ring chronology network serves as the suite of candidate predictor variables to be considered in regression for reconstructing MRB streamflow, and the predictor selection procedure (described in the next subsection) narrowed the number of prospective tree ring chronologies considerably for each streamflow gauge reconstructed here.

5.3. Predictor Variable/Tree Ring Chronology Selection

The selection of tree ring chronologies for reconstructing the 18 streamflow gauges follows the basic model outlined in Cook et al. (1999). First, a 1,000-km (621-mile) search radius from each streamflow gauge location was used to find the tree ring chronologies that were plausibly correlated with water year streamflow. This search radius is much larger than the 450-km (280-mile) one used by Cook et al. (1999) for optimal reconstruction of PDSI. This is because the topology of watersheds is such that the main runoff-producing regions near where the trees are growing can be quite far from the downstream gauge locations, thus making the search radius between the gauge and tree ring chronology series necessarily greater than a simple correlation decay distance model for precipitation and PDSI. The 1,000-km (621-mile) search radius covers a sizable fraction of the UMRB. Consequently, most of the 375 tree ring chronologies were located within 1,000 km (621 miles) of each streamflow gauge. Thus, the number of candidate tree ring predictors found per gauge ranged from 237 to 310.

Once the pool of candidate tree ring predictors was found for a given streamflow gauge, a time period common to all data was chosen for correlating each tree ring chronology with its target streamflow record. In this case it was 1947–1979, which is the longest period in common between the 18 streamflow records and the 374 candidate tree ring chronologies. Each water year streamflow record was then correlated against its pool of candidate tree ring chronologies and those chronologies that correlated at the two-tailed 95% significance level were retained as the actual predictors of streamflow. The number of tree ring chronologies that passed this 95% screening ranged from 8 to 94 at the 18 streamflow gauges, with a mean of 52 and a median of 58, which is a dramatic decrease from the pool of candidate predictors.

Next, we performed a PCA on each suite of screened tree ring predictors for each of the 18 streamflow gauges to further reduce the dimensionality of the predictors. The PCA was performed using the `prcomp` function in R software/programming language, version 3.2.2 (R Core Team, 2015). The `prcomp` function is found in the “stats” package and Singular Value Decomposition was used to carry out the PCA computation using the covariance matrix. We selected the top five PCs as the final suite of tree ring chronology predictors for each of these 18 gauges. Note that the PCs are different for different gauges; however, in all the cases, the top five PCs were accounting for at least 50% of the variance in the original set of screened predictors. One can increase or decrease the number of PCs that finally go into the BSM model depending on the application. In our case, for the purpose of demonstrating the application and the utility of the BSM model, the top five PCs offered a reasonable amount of variance in the original set of chosen predictors while ensuring low dimensionality.

5.4. The Cross-Validation Procedure

When presenting a modeling methodology, it is both traditional and necessary to use a cross-validation procedure to validate the goodness of the model. We chose to implement such a cross-validation procedure using a leave-five-years-out cross validation. Given the entire observed water year totals for each of the 18 gauges (see Table 1 for record lengths per station), five consecutive years from the common record (1947–1983) were chosen at random as validation years and the BSM model was built using the remaining observations, also known as the calibration years, for each gauge and the data removed from those five years were hence interpolated by the model. Five validation metrics, namely, coverage rates of 95% credible intervals ($CR_{0.95}$), the reduction of error statistic (RE), the coefficient of efficiency statistic (CE), a normalized root-mean-square error (RMSE), and ranked probability skill score (RPSS), were then calculated based on comparing the actual data from the five validation years with the model-interpolated results as a way of seeing how well the model performs. See Cook et al. (1999) for descriptions of RE and CE. CE is also equivalent to the Nash-Sutcliffe Efficiency statistic commonly used in hydrology. This cross-validation procedure was repeated 50 times and boxplots for each metric were created based on the distribution of the 50 values obtained.

We include a comparison of the cross-validation results of the BSM model with those of the standard Principal Components Regression (PCR) model. By “standard,” we mean that we did not apply the simulation method that is described later on in detail in section 6.1, as this is not how PCR is traditionally carried out for reconstructions. For the PCR model, we simply applied classical OLS regression using the same five PCs identified in section 5.3 for each gauge as the predictors, and log-seasonal total streamflow as the predictand, in order to generate point estimates of the regression parameters and reconstructions. We applied the exact same cross-validation procedure described above with these point reconstructions with the OLS framework instead of the Bayesian framework. No network structure was implemented in the PCR; hence, feeder streamflow data were never considered as a predictor for any of the gauges in the PCR method. We were only able to do this for the RE, CE, and RMSE statistics, as RPSS is only appropriate for models with probabilistic output and coverage rates of credible intervals is strictly a Bayesian evaluation metric (Li et al., 2010). We then compared the resulting cross-validation distributions with those generated by the BSM model. We will quickly explain each cross-validation metric in the following paragraphs.

The coverage rate of 95% credible intervals is a statistic defined in the following way (Devineni et al., 2013; Li et al., 2010):

$$CR_{0.95} = \frac{\sum_{t \in \text{val}} \mathbf{1}_{I_t}(\mathbf{y}_t^{(i)})}{|\text{val}|} \quad (4)$$

where $\text{val} =$ set of five randomly chosen validation years, $I_t = \left[q_{0.025}(\widehat{\mathbf{y}}_t^{(i)}), q_{0.975}(\widehat{\mathbf{y}}_t^{(i)}) \right]$ is the interval whose lower bound is the 2.5th percentile of the model-generated posterior distribution of flows corresponding to validation year t in gauge i and whose upper bound is the 97.5th percentile of the same, $\mathbf{y}_t^{(i)}$ is the observed water year total flow for year t , gauge i , and $\mathbf{1}$ is the indicator function on I_t . Hence, $\mathbf{1}$ is equal to 1 if $\mathbf{y}_t^{(i)}$ is in the credible interval I_t and is 0 otherwise. In simple English, equation (4) describes the relative proportion of validation years during which the actual total streamflow datum is inside the corresponding credible interval. It is desirable for this value to be as close to 1 as possible.

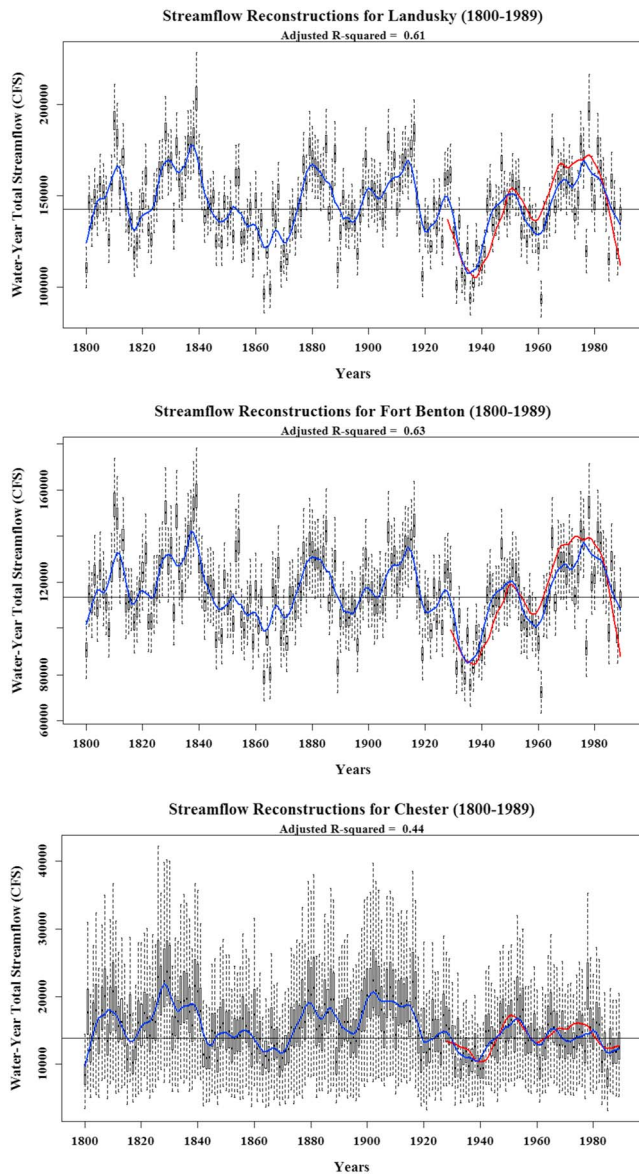


Figure 4. The reconstructions for (top) Landusky, (middle) Fort Benton, and (bottom) Chester presented as boxplot time series. The boxplots represent the annual posterior distribution estimates of the streamflow. The medians of these boxplots/distributions are considered to be the actual reconstruction values when computing R^2 , and the 10-year lowess-smoothed red time series that has been superimposed on the boxplots represents the trends in observed flows. The blue curves are 10-year lowess-smoothed time series based on the boxplot medians. The adjusted R^2 of the reconstruction is included beneath the chart title.

The reduction of error (RE; Fritts, 1976) and coefficient of efficiency (CE; Briffa et al., 1988) statistics are defined as

$$RE = 1 - \frac{\sum_{t \in \text{eval}} \left(y_t^{(i)} - \hat{y}_t^{(i)} \right)^2}{\sum_{t \in \text{eval}} \left(y_t^{(i)} - \bar{y}_c^{(i)} \right)^2} \quad (5)$$

$$CE = 1 - \frac{\sum_{t \in \text{eval}} \left(y_t^{(i)} - \hat{y}_t^{(i)} \right)^2}{\sum_{t \in \text{eval}} \left(y_t^{(i)} - \bar{y}_v^{(i)} \right)^2} \quad (6)$$

where $\hat{y}_t^{(i)}$ in both RE and CE represents the median of the BSM-generated posterior distribution for year t and gauge i , $\bar{y}_c^{(i)}$ is the average value of the observational data in the calibration period, and $\bar{y}_v^{(i)}$ is the average value of the observational data in the validation period. RE ranges from $-\infty$ to 1 and is a comparative measure of how accurate and informative the model forecast is with respect to the mean of the calibration period. $RE < 0$ indicates that the model contains less information than the mean value of the calibration data, whereas $RE > 0$ indicates that the model contains more information than the mean value of the calibration data (Cook et al., 1999; Devineni et al., 2013). CE is a similar measure that compares the model performance with the information contained in the validation period mean instead of the calibration data. Hence, $CE < 0$ implies that the validation data contain more information than the model forecast, whereas $CE > 0$ implies that the model forecast contains more information.

The fourth validation metric is the RMSE statistic normalized by the median of the entire observed record. The normalization is done in order to understand the average model forecast/estimation error with respect to the “typical” level of streamflow at a given gauge. RMSE values across gauges cannot be fairly compared directly, or even understood in isolation, without understanding the typical magnitude of the flows at each gauge. With the accumulation of flow as one moves further downstream, the average flow at downstream gauges will be much larger than the average flow at upstream gauges. This metric is therefore defined as

$$RMSE = \sqrt{\frac{1}{|\text{val}|} \sum_{t \in \text{eval}} \left(y_t^{(i)} - \hat{y}_t^{(i)} \right)^2} \quad (7a)$$

$$\text{Median-Normalized RMSE} = 100 \times \frac{RMSE}{\text{median}(\mathbf{y}^{(i)})} \quad (7b)$$

where $\mathbf{y}^{(i)}$ represents the entire streamflow record for gauge i . It is clear from the definition of median-normalized RMSE that we have chosen to express this statistic as a percentage for ease of interpretation. This metric expresses the average degree of discrepancy between the observations and the model interpolations over the entire validation period.

The final cross-validation metric chosen is the ranked probability skill score, or RPSS. The RPSS is expressed as a ratio of two RPS, or ranked probability score, values. The RPS (Epstein, 1969; Murphy, 1969, 1971) is a validation metric associated with categorical probabilistic forecasts and measures the cumulative squared error between the status of an observation existing in a particular category and the probability under the forecast model of being in that category. The RPSS compares the RPS of the forecast model with the RPS of a reference forecast system. Hence, the RPS metric is defined as (Wilks, 1995)

$$RPS = \sum_{m=1}^J (Y_m - O_m)^2 \quad (8a)$$

where J = total number of categories, $Y_m = \sum_{i=1}^m P(\text{category } i)$, $O_m = \sum_{i=1}^m o_i$, $m=1, \dots, J$, $P(\text{category } i)$ is the probability of an observation being in category i , and o_i is an indicator variable that equals 1 if the observation is in category i and 0 otherwise. The RPSS is then defined accordingly:

$$RPSS = 1 - \frac{\overline{RPS_{\text{forecast}}}}{\overline{RPS_{\text{reference}}}} \quad (8b)$$

where $\overline{RPS_{\text{forecast}}}$ is the RPS of the forecast model (in our case, the network Bayes model) averaged over the entire 5-year validation period, and $\overline{RPS_{\text{reference}}}$ is the RPS of the reference forecast averaged over the validation period. The RPSS ranges from $-\infty$ to 1. $RPSS < 0$ indicates that the reference forecast has higher skill than the forecast model, $RPSS = 0$ implies equal skill and $RPSS > 0$ implies that the forecast model has higher skill, with $RPSS = 1$ being a perfect score.

6. Results and Discussion

6.1. Reconstruction Results

The streamflow reconstructions for Landusky, Fort Benton, and Chester, respectively, are shown in Figure 4 as a sample of the reconstructions, specifically exemplifying two major gauges with multiple feeder gauges (Landusky and Fort Benton) and one gauge without any feeders and only trees ring chronologies instead (Chester). The record of observed streamflow data varies slightly from gauge to gauge, as seen in column five of Table 1. The reconstructions create a streamflow time series that spans 190 years, from 1800–1989. The record period common to all of the selected trees determined the time span of the reconstructions. The reconstructions themselves are presented as time series composed of boxplots instead of single points, as the reconstructions for each year are simulated from the likelihood using estimates of the posterior distribution of the mean and variance parameters for those years. The boxplots graphically depict those posterior distributions. The lowess-smoothed (Loader, 1999) time series of medians is shown as a blue curve passing through the boxplots, and the average of the observed data is shown as a black horizontal line. In each of these plots, we also include the lowess-smoothed observed time series in red. The adjusted R^2 statistic is calculated using the median of the reconstruction (posterior) distributions as the fitted values. The adjusted R^2 value for the Landusky gauge is 0.61 (Table 2).

The second panel of Figure 4 shows the results for Fort Benton, which is a feeder gauge to Landusky and a major junction in the network in its own right. The adjusted R^2 value for this reconstruction is 0.63 (Table 2). Finally, the first panel of Figure 4 shows the reconstruction time series and the adjusted R^2 value from our Bayes model for Chester. Note that Chester has no feeder streams. The adjusted R^2 value based only on tree rings is therefore a more modest 0.44.

The adjusted R^2 values, with only two slight exceptions, tend to increase between the feeder gauges and the gauges they feed. This is perfectly intuitive, as the streamflow gauges that are connected to upstream feeder gauges are receiving additional information from their feeders, which in turn contain large-scale variability signals received from tree ring chronology data. Aside from that, adjusted R^2 values are generally between 0.40 and 0.60, with only four gauges below 0.40. These adjusted R^2 values are in line with other paleo streamflow reconstruction efforts. For instance, Watson et al. (2009) were able to explain between 40% and 64% of observed variance in their streamflow reconstructions for the headwaters of the Wind River in Wyoming. Barnett et al. (2010) were able to explain 44% to 65% variance in their streamflow reconstructions for the Upper Green River Basin. Ho et al. (2016) were able to explain an even greater share of the variance in the Missouri River Basin, where the adjusted R^2 values of their reconstructions ranged from 0.56 to 0.90. However, there is a significant overlap in the range of R^2 values in this study with those in our study. Other studies are similar for other basins. The adjusted R^2 values for each gauge corresponding to the BSM model reconstructions are shown in column 2 of Table 2.

Table 2
The Results of the BSM Reconstructions

Streamflow gauge	R^2_{adj} (Bayes)	Average rCV over observed period	Average rCV over reconstructed record	Significant predictors as determined by the model
Barretts	0.35	0.257	0.266	PC1 and PC5
Twin Bridges	0.45	0.311	0.337	PC1, PC2, PC4, PC5
Melrose	0.39	0.306	0.312	PC1, PC2, PC3, PC4, PC5
Three Forks Jefferson	0.50	0.109	0.113	PC1, PC2, PC4, PC5, and Twin Bridges, Melrose and Barretts streamflow
Logan	0.56	0.199	0.204	PC1, PC2, PC4, and PC5
McAllister	0.62	0.153	0.159	PC1, PC2, PC4, PC5
Three Forks Madison	0.53	0.077	0.080	PC3, PC4, PC5, and McAllister streamflow
Toston	0.59	0.071	0.076	PC1, PC4, PC5, and Logan, Three Forks Jefferson and Three Forks Madison Streamflow
Craig	0.45	0.466	0.553	PC1 and PC4
Eden	0.59	0.356	0.372	PC1, PC2, and PC3
Gibson Reservoir	0.29	0.396	0.468	PC1, PC2, PC3, PC4, PC5
Vaughn	0.39	0.115	0.136	PC1, PC2, and Gibson Reservoir streamflow
Fort Benton	0.63	0.063	0.066	PC4, PC5, and Eden, Toston and Vaughn streamflow
Winifred	0.47	0.444	0.476	PC1, PC2, PC3, PC4, and PC5
Chester	0.44	0.317	0.362	PC1, PC2, PC4, and PC5
Dutton	0.42	0.268	0.290	PC1, PC2, and PC4
Loma	0.40	0.005	0.005	PC3, PC4, PC5, and Dutton streamflow
Landusky	0.61	0.050	0.052	PC2, PC4, PC5, and Loma, Fort Benton and Winifred streamflow

Note. Data are presented (from left to right columns in the table, starting with column 2) in terms of the adjusted R^2 of the linear model fit, the relative uncertainty (measured as robust coefficient of variation, or rCV) of the reconstruction calculated over the observed record, the relative uncertainty of the reconstruction calculated over the paleo record, and the predictors that were deemed important by analyzing the posterior distribution of the model parameters (see Figure 5). BSM = Bayesian Spatial Markov model.

One of the cornerstones of the BSM model is the usage of feeder gauge linkages to not only mimic the physical network structure of the river basin but also to reduce uncertainty in the reconstructions. In order to test this hypothesis, we compare a model for reconstructing the flows at the aforementioned 18 gauges in the UMRB without considering the network structure of the watershed. In other words, we do not consider feeder gauge linkages in this reconstruction model, which is created purely for the purposes of testing the hypothesis of uncertainty reduction. This test model is constructed as follows. Using the predictors (PCs) selected in section 5.3 for the BSM model, we used standard OLS with the observed streamflow values to estimate the regression coefficients. Using these point estimates of the regression coefficients, we estimate the mean and variance parameters by calculating the fitted values and model standard error for each of the 190 years and then simulate the reconstruction distributions from the Gaussian likelihood function to obtain distributional paleo reconstructions as we did with the BSM model. Note that the feeder stream linkages were omitted from this procedure, as this is a unique feature of our BSM model, but the idea of simulating streamflow reconstructions as probabilistic, distributional hindcasts was adapted to the classical PCR framework for the sake of direct comparisons of model skill metrics. The adjusted R^2 statistics were also calculated for this model, and these values are found in column 2 of Table 3. As can be seen from this Table, the values are roughly the same between both the PCR and BSM models, confirming the earlier assertion that the adjusted R^2 values from the BSM model are similar to other paleo reconstruction efforts. For gauges without feeders, the BSM model ought to generate comparable reconstructions to those of a classical regression-based framework, as the only difference between the two methods are the estimation schemes; one being Bayesian (BSM) and the other being classical OLS (PCR).

Columns three and four of Tables 2 and 3 show, for each streamflow gauge and each of BSM and PCR methods, respectively, the *robust coefficient of variation* (rCV) values averaged over the portion of the reconstructed records corresponding to the observations and the hindcasted portion of the reconstructed record, respectively. The rCV is defined as the interquartile range divided by the median, and one can easily see how this, in robust statistics, is analogous to the coefficient of variation in classical statistics. For each gauge, the rCV was computed for each of the annual reconstruction distributions and then averaged to one statistic.

Table 3
The Results of the PCR Reconstructions

Streamflow gauge	R^2_{adj} (PCR)	Average rCV over observed period	Average rCV over reconstructed record
Barretts	0.34	0.253	0.262
Twin Bridges	0.45	0.307	0.334
Melrose	0.38	0.302	0.307
Three Forks Jefferson	0.53	0.217	0.222
Logan	0.55	0.194	0.199
McAllister	0.62	0.151	0.156
Three Forks Madison	0.55	0.172	0.179
Toston	0.59	0.185	0.194
Craig	0.44	0.444	0.529
Eden	0.58	0.350	0.368
Gibson Reservoir	0.29	0.388	0.460
Vaughn	0.38	0.336	0.396
Fort Benton	0.62	0.190	0.195
Winifred	0.47	0.435	0.468
Chester	0.42	0.311	0.359
Dutton	0.40	0.263	0.285
Loma	0.39	0.269	0.293
Landusky	0.57	0.201	0.206

Note. Data are presented (from left to right columns in the table, starting with column 2) in terms of the adjusted R^2 of the ordinary least squares model fit, the relative uncertainty (measured as robust coefficient of variation, or rCV) of the reconstruction calculated over the observed record, and the relative uncertainty of the reconstruction calculated over the paleo record. PCR = principal component regression.

The rCV values can give us an idea of the uncertainty reduction within gauges and across gauges. It can be compared across gauges, since the uncertainty reported in this statistic is measured relative to the average volumetric flow at each gauge.

The average rCV for the observed portion of the reconstructions is always less than or equal to the average rCV for the reconstructed portion, indicating greater uncertainty in the unobserved portion of the reconstructions, going back in time. This is generally true for both the BSM and PCR models. This trend is not surprising, as we expect to see more uncertainty as we extrapolate further into the past with no observations to guide us directly, using only the estimates derived from observations at a later point in the record when the behavior of physical streamflow drivers may have been quite different. However, the difference in average rCV value is not very significant between the observed and reconstructed portions for either model. This indicates that the level of uncertainty in the reconstructions during the paleo period is not much worse than that of the reconstructions during the observed period, which is a comforting sign.

The level of uncertainty in the BSM model reconstructions is typically at least an order of magnitude lower for gauges that have feeders than for those that do not have feeders. If we compare the rCV values for the gauges without feeders between the BSM and PCR models, the values are virtually the same. However, for the gauges with feeders, BSM has a significant uncertainty reduction in the reconstructions for both the observed and paleo periods when compared with the classical PCR model. Hence, our suspicion that the spatial Markov structure would lead to a significant uncertainty reduction was correct in the case of the gauges with feeders.

To illustrate these results better, we present the spatial distribution of the percentage reduction in the uncertainty for each gauge in Figure 5. For each gauge, we computed $\left(\frac{rCV_{PCR} - rCV_{BSM}}{rCV_{PCR}}\right) \times 100$ for the reconstruction period as a measure of uncertainty reduction. A positive (negative) measure indicates that the rCV in the BSM (PCR) model is lower than the rCV of the PCR (BSM) model. As mentioned above, we can see that there is a significant uncertainty reduction for gauges that have feeder gauges as predictors in addition to the tree ring chronologies. We observed more than 50% reduction in uncertainty for almost all such gauges. Three Forks Jefferson presents the only exception, for which the uncertainty reduction according to Figure 5 was between 5% and 50%. The gauges that serve as the feeders (i.e., those having only tree ring chronologies, and not tree ring-based reconstructed flows, as predictors) have uncertainty reduction ranging between

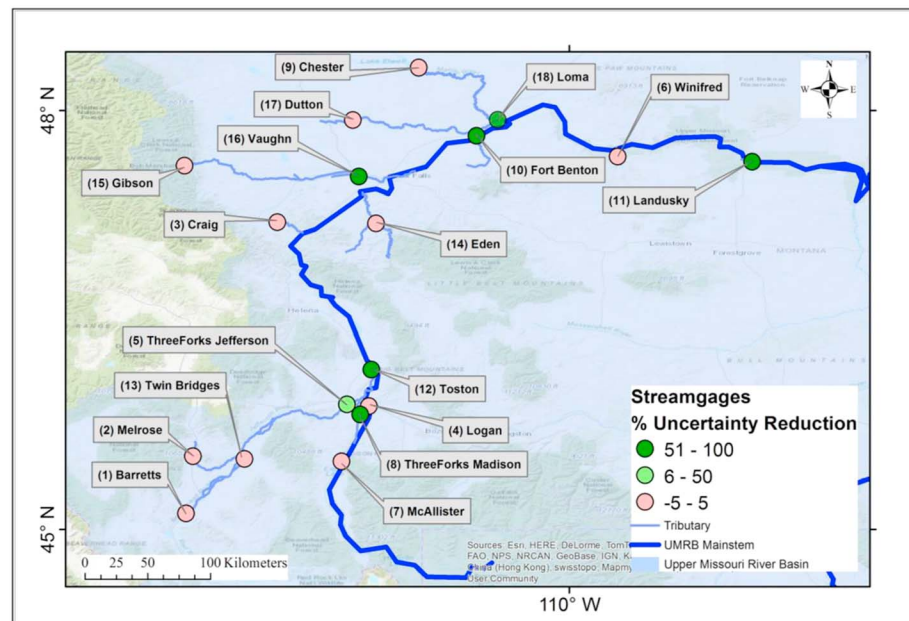


Figure 5. A map of the percentage reduction in uncertainty for each of the 18 gauges as measured by the rCV metric over the reconstruction period. Uncertainty reduction is measured as the relative difference between rCV of the PCR model and rCV of the BSM model.

−5% and +5%. PCR model is marginally better in some of these latter gauges; however, since the margin is within 5%, it can be argued that both PCR and BSM model outputs are equally valid for these gauges.

Figure 6 shows the posterior distributions of the regression coefficients corresponding to each explanatory variable in the model for the same three gauges as depicted in Figure 4: Chester, Fort Benton, and Landusky. Included on these boxplots are the zero line as a horizontal black line that crosses the entire graph and the 95% credible intervals as red dots. Strictly speaking, it would seem proper to accept a particular predictor variable (PC) as significant in explaining streamflow variability over time if the line through zero was completely outside of the confidence bounds (superimposed dots). However, we feel it is appropriate to grant a little more leeway and consider a predictor variable to be deemed insignificant by the model only if the zero line passes through the interquartile range, that is, if the zero line is within the boundaries of the box. By this definition, Figure 6 shows us that PC2, PC4, and PC5, along with all three feeders (Fort Benton, Winifred, and Loma) were found significant for Landusky streamflow reconstruction by the BSM model. However, in terms of explained variance in the final model, the partial- R^2 s of the PCs must be considered tiny compared to those of the feeder streams. This is because, as already formed estimates of upstream streamflow based to some degree on the same tree ring data used in the PCs for reconstructing the downstream gauges, the feeder streams are more highly correlated with the lower streamflow record than any of the PCs used as additional predictors. For Fort Benton, PC4 and PC5 were significant paleo-predictors and Eden, Toston, and Vaughn streamflow were significant feeders for the reconstructions at this gauge. Finally, PC1, PC2, PC4, and PC5 were deemed to be significant predictors for streamflow reconstruction at Chester. The fifth column of Table 2 lists the predictor variables among both tree ring PCs and feeder gauges that were determined to be important by our reconstruction model.

6.2. Cross-Validation Results

The cross-validation results, using the procedure described in section 5.4, are depicted in Figures 7 through 11. The comparisons with the standard PCR model are included in Figures 8, 9 and 10, with the PCR cross-validation output presented as gray boxes overlaying the white boxes belonging to the BSM cross-validation.

The results for the first cross-validation metric, the coverage rate of 95% credible intervals, or $CR_{0.95}$, for our model, are found in Figure 7. Eleven of the gauges cover the true streamflow value for all five of the randomly selected validation years for every one of the 50 iterations of the cross-validation procedure. For Three Forks Jefferson, while the average coverage rate of the cross-validation empirical distribution is 1.0,

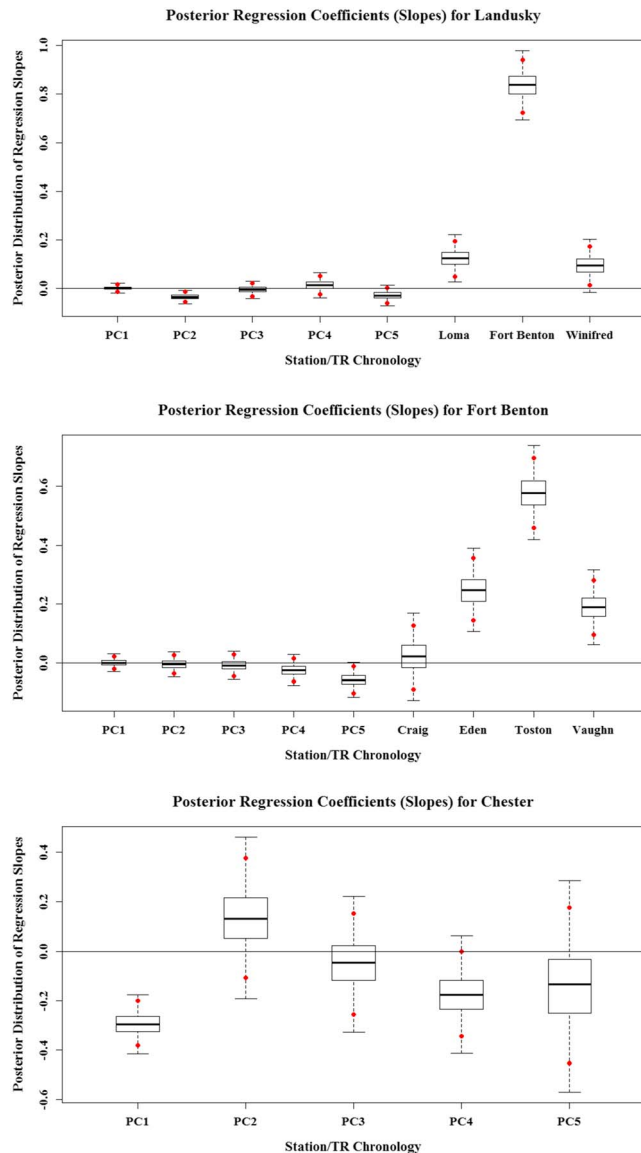


Figure 6. The boxplots representing the estimated posterior distributions of the regression parameters (slopes) for the same gauges depicted in Figure 4 (top: Landusky; middle: Fort Benton; and bottom: Chester). The lower and upper red dots superimposed on the boxplots are the 2.5th and 97.5th percentile values, respectively, of the distribution.

there is some skew toward coverage rates that are lower. The probability mass remains above a coverage rate of 0.8, indicating that for this gauge, the model is generating plausible values in the output distribution with good probability. However, Landusky, Loma, Fort Benton, Toston, Vaughn, and Three Forks Madison are streamflow gauges for which the median value of this distribution of 50 cross-validated $CR_{0.95}$ values is less than one. While Landusky, Fort Benton, Toston, Vaughn, and Three Forks Madison have median cross-validation distributions at or above 0.50, indicating good coverage, Loma is the only gauge that shows an anomalously deficient result. The cross-validation distribution for coverage rates for Fort Benton and Three Forks Madison show quite wide variability, but maintain strong median values. All of this indicates that the model consistently produces plausible estimated streamflow totals with high probability in each of its posterior estimates across 17 of the 18 gauges, indicating a reasonably high precision in the predictive model.

Figure 8 shows the cross-validation results for the RE statistic. This figure contains the cross-validation results for the network Bayes model and the PCR model as discussed earlier. The results for the network Bayes model (the white boxes) suggest that the information conveyed by the model is considerably greater than the information contained on average in the calibration period for all of the streamflow gauges without exception. All the boxes (with the exception of Gibson, which depict the cross-validation distribution of the RE statistic are above the zero line, and the medians of these distributions are above zero. This indicates a strong result. The performance of the PCR model (gray boxes), given the same data, is generally less impressive, although not bad in its own right. The cross-validation distributions for PCR have medians that are consistently lower than those of the network Bayes model for all 18 gauges. In addition to this, the level of uncertainty is considerably larger in the PCR RE cross-validation distributions for a good majority of the gauges; in other words, the BSM model has much less uncertainty in its cross-validation distributions of the RE statistic than the PCR model does for most gauges. Hence, we conclude that the BSM is a potential improvement over PCR in cross validation with the RE statistic, lending credence to the notion that BSM is better at reducing uncertainties and at generating greater accuracy in its reconstructions. We will now cite a few examples of the RE statistic's use in cross validation in some related studies. While there are many more examples that can be cited, we cite just a few prominent ones for the sake of brevity. The RE statistic was also used by Cook and Jacoby (1983), Cook et al. (2013), Devineni et al. (2013), Bracken et al. (2016), and Ho et al.

(2016) for validation. Cook and Jacoby (1983), for streamflow reconstructions in the Potomac River, found RE values not too far from the median values seen in Figure 8; typically between 0.2 and 0.5, with three notable exceptions. Cook et al. (2013) found similar results when reconstructing streamflow in the Indus River Basin. Devineni et al. (2013), when reconstructing streamflow at five reservoirs in the Upper Delaware River Basin using a hierarchical Bayesian, partial-pooling regression model, found average RE values to be between 0.45 and 0.50, with low variability around this median value. Bracken et al. (2016) compared their RE validation results with those of classical, single-site regression models, and found generally higher values from their spatial dependence model. Ho et al. (2016), a study that also reconstructed streamflow in the Missouri River Basin, found that vast majority of the RE values in cross-validation mode were significantly positive. Rao et al. (2018) found similar results in their reconstructions for the Indus River Basin. Hence, our findings indicate that our model is at least as successful in cross-validation mode as other major studies in streamflow reconstruction.

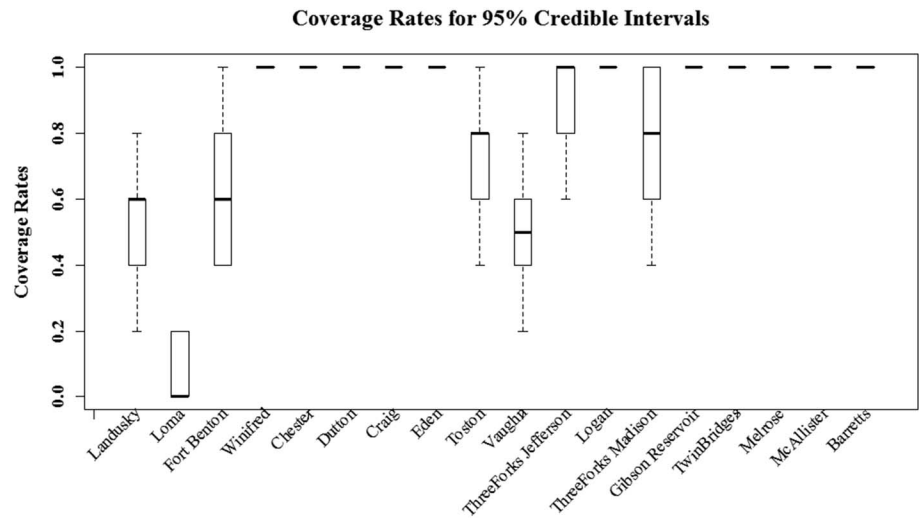


Figure 7. The cross-validation distributions for the coverage rates of 95% credible intervals (denoted $CR_{0.95}$). The boxplots represent the distributions for each of the 18 stations. The distributions are composed of 50 runs of a leave-five-out cross validation, in which five randomly chosen consecutive years during the common record of 1947–1983 were cross-validated against the observations for those chosen years using this statistic.

The results for CE are depicted in Figure 9. The cross-validation distributions for the CE statistic for the BSM model for all gauges except Landusky have significant crossing below the zero line. The median values for these distributions are above zero for all but two stations: Twin Bridges and McAllister. This indicates that the average streamflow observations over the validation period hold less information than the median of the model-generated posterior distributions. The results are somewhat less impressive than those of the RE statistic, as CE is a more stringent measure. We also applied the same cross-validation procedure to the PCR model, and found that the network Bayes model outperformed PCR in this cross validation for all 18 stations in terms of accuracy; in other words, there was a greater tendency toward more positive values in the BSM cross validation for CE. The uncertainty of the PCR boxes is considerably larger than for the Bayes model for

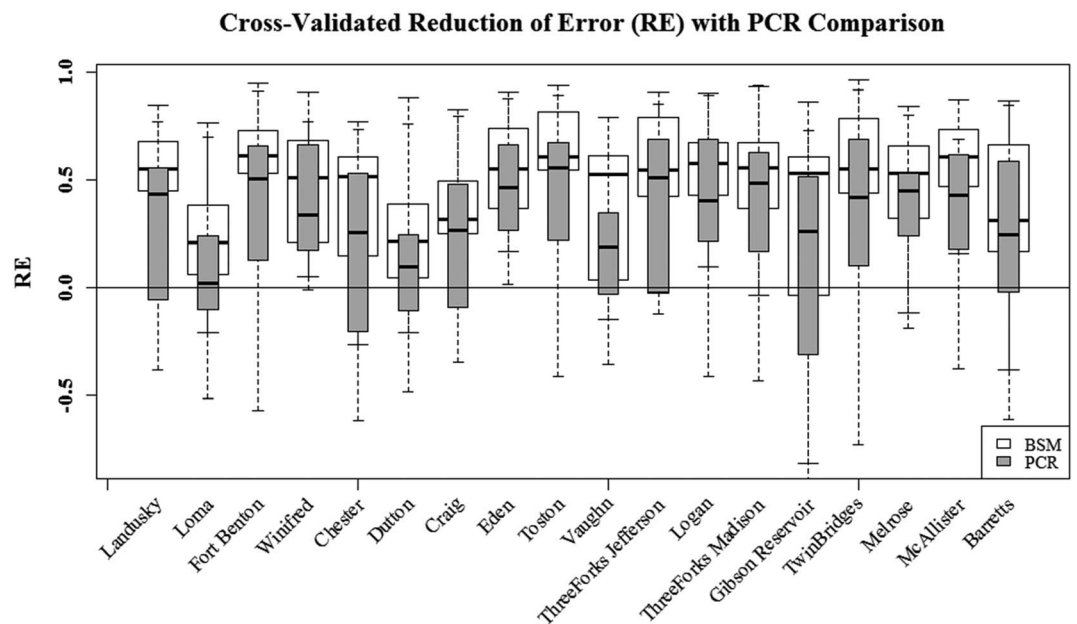


Figure 8. The cross-validation distributions for the reduction of error (RE) statistic with the principal component regression (PCR) cross validations included. Refer to the caption of Figure 7 for an understanding of the nature of these distributions. BSM = Bayesian Spatial Markov model.

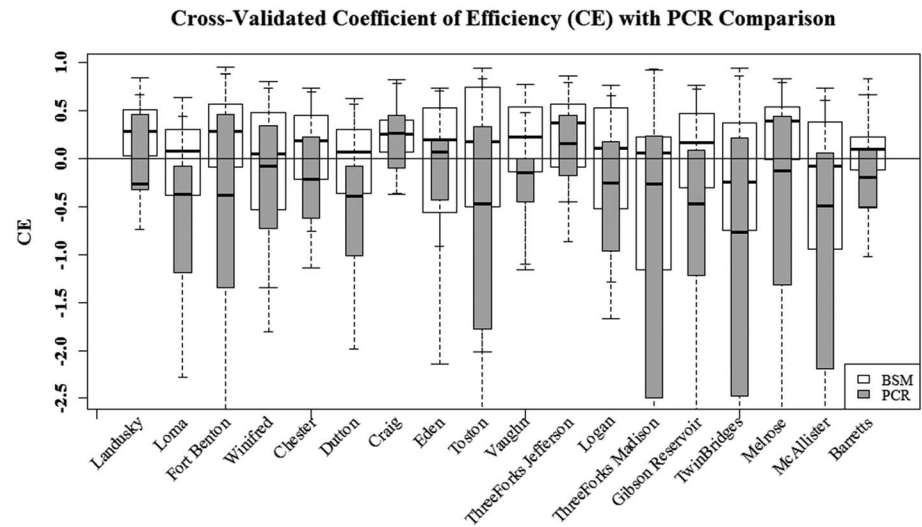


Figure 9. The cross-validation distributions for the coefficient of efficiency (CE) statistic with the principal component regression (PCR) cross validations included. Refer to the caption of Figure 7 for an understanding of the nature of these distributions. BSM = Bayesian Spatial Markov model.

almost all of the gauges. For both RE and CE, this is due to the fact that the BSM model is more effective in reducing uncertainty in its reconstructions. We conclude that the BSM model improves upon the PCR model in cross validation with the CE statistic. Cook et al. (2013), Devineni et al. (2013), Ho et al. (2016) and Rao et al. (2018) all found results that were similar in their reconstructions; in particular, it was a bit more difficult to achieve positive values with the CE statistic than with the RE statistic, and hence a slightly greater number of negative values were encountered in the validation with this statistic.

Figure 10 shows the median-normalized RMSE cross-validation distributions for all streamflow gauges for both the BSM model (white boxes) and the PCR model (gray boxes). Without exception, there was a general tendency for the RMSE of the BSM model to be less than 30% of the median observed streamflow water year total, meaning that the average error incurred by the model-estimated total flow, as measured by the median

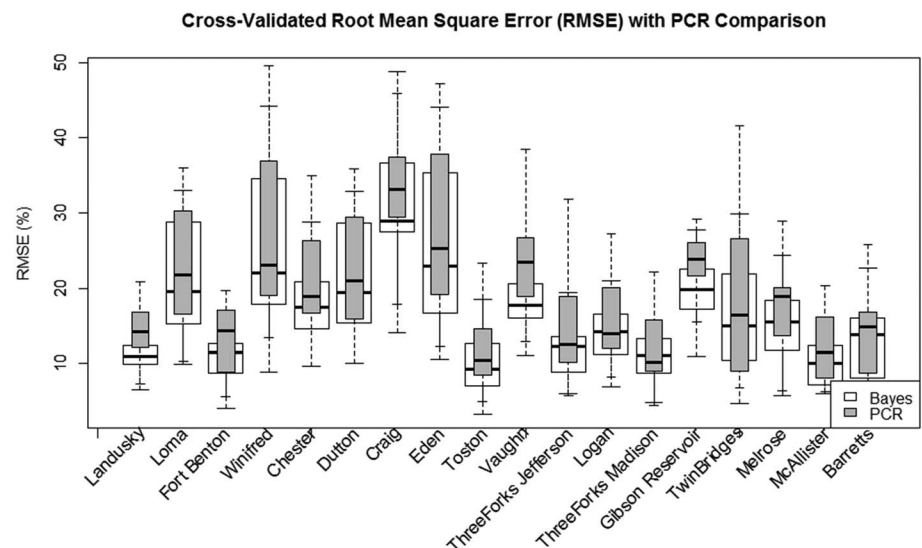


Figure 10. The cross-validation distributions for the median-normalized root-mean-square error (RMSE) statistic with the principal component regression (PCR) cross validations included. Refer to the caption of Figure 7 for an understanding of the nature of these distributions. The median-normalized RMSE is the RMSE calculated over the validation period and corresponding to a particular streamflow gauge, divided by the median value of the entire observed streamflow data for that streamflow gauge.

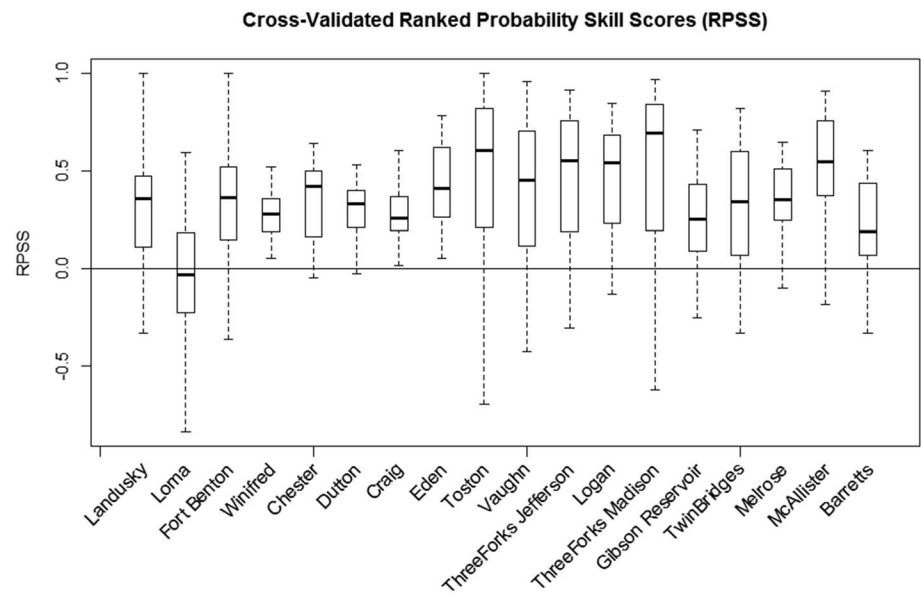


Figure 11. The cross-validation distributions for the ranked probability skill score (RPSS) statistic. Refer to the caption of Figure 7 for an understanding of the nature of these distributions.

of the model distributions, was fairly normal, and in some cases small, given the typical order of magnitude of the observed flows. The level of uncertainty in the cross-validation distributions for the Bayes model, as displayed in Figure 10, was quite low for all but four gauges (Loma, Winifred, Dutton, and Eden), where excessive skewness is also a problem. There is therefore a convergence, or clustering, around the median of the cross-validation distributions, which indicates a consistency in the model estimations of flow. The uncertainty levels in the PCR cross-validation distributions for RMSE were smaller in some cases, such as Craig and Gibson. However, the median of the cross-validation distributions for RMSE were higher for PCR than BSM for every station except Three Forks Madison. The differences in performance here are quite clear, but not dramatic. In general, the BSM model shows a lower rate of error between the model interpolations and observations over the validation period than the PCR model, indicating a tendency to higher accuracy for the BSM model.

Finally, Figure 11 shows the results of the 5-fold cross-validation runs for each of the 18 streamflow gauges. In our case, we created three categories by dividing the observed streamflow totals into terciles for each streamflow gauge, so that the first category was defined as streamflow data less than the 33rd quantile of the entire observed record and the third category was defined as streamflow data less than the 67th quantile of the entire observed record. The second category was naturally defined as anything between the two tercile values. The reference forecast assigned a probability of 33% to the first category, 34% to the second category, and 33% to the third category. The cross-validation distributions for all of the gauges with the exception of Loma remain entirely above zero. This indicates a high level of skill in the BSM model reconstructions.

6.3. Verification of Spatial Dependence

Figure 12 shows (in the order of top left, top right, middle left, middle right, and bottom left, bottom right); the correlation matrix plots for the BSM reconstructions, the PCR model reconstructions, and the actual streamflow data, the legend showing the gauge names associated to each of the numbers (1–18) in the left and top margins of the aforementioned spatial correlation matrices, the scatter plot, with lowess smoother, of the correlation values depicted in the correlation matrix plots plotted against the distances (in kilometers) between each streamflow gauge, and a scatter plot of the correlation values found in the top left correlation matrix against the correlation values found in the top right correlation matrix. This figure depicts the fidelity of the spatial variability of the flows over the common period of record (1947–1989), as this is the period of time over which all observations exist across all gauges. This same time frame was used for the reconstructions as well to maintain consistency. The reconstructed streamflow data for BSM used in this correlation

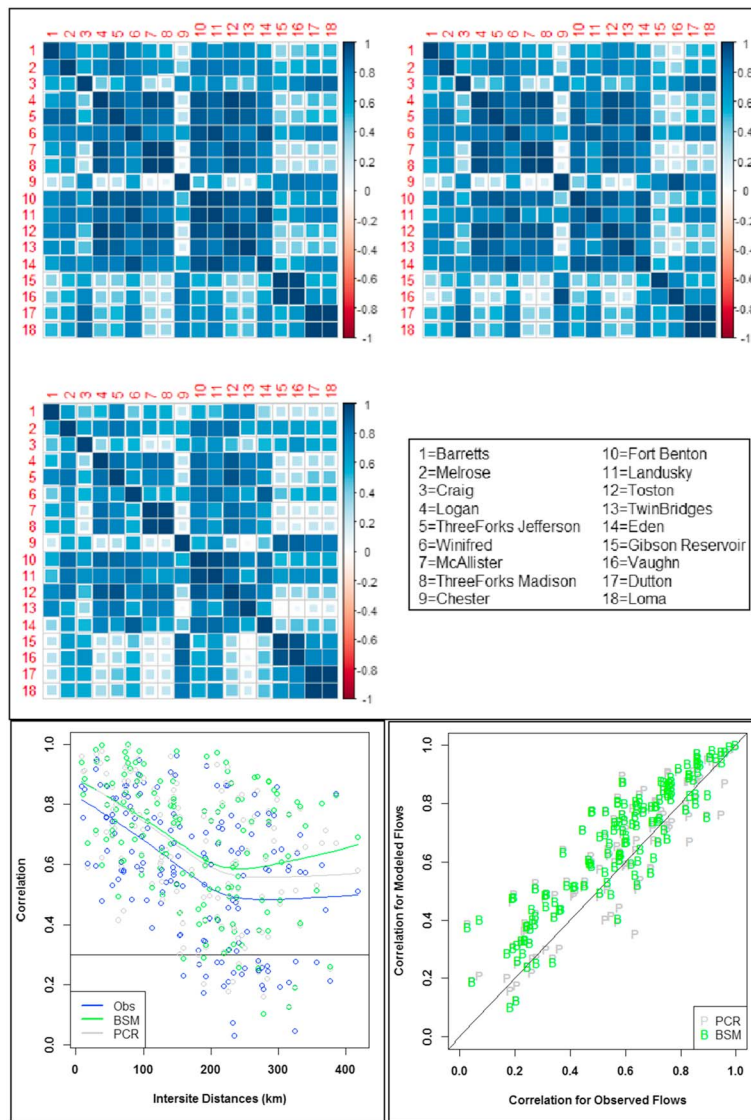


Figure 12. A correlation matrix plot for the BSM median reconstructed flows (top left), the PCR reconstructed flows (top right), the actual streamflow water year totals (middle row, left side), a scatter plot of these correlation values against the distance values between the stations (bottom left), and the correlation values from the correlation matrix plot for BSM median reconstructed flows plotted against the same for the PCR reconstructed flows (bottom right). The correlations were made between the calculated flows at any two of the 18 gauges, and the values are plotted as colors depending on their magnitude and sign, with a scale provided for the color/correlation value ranges. The gauges are labeled by numbers between 1 and 18, and the gauge name corresponding to each number is provided in the accompanying legend (middle row, right side). The scatterplot contains the same correlation values plotted against the distance values between each of the 18 stations in kilometer (x axis). A Lowess smoother for the scatterplot is included for each of observations, BSM-reconstructed streamflow and PCR-reconstructed streamflow correlations. PCR = principal component regression; BSM = Bayesian Spatial Markov model.

matrix plot is derived from the model estimates of the posterior distribution of flow by using the median of the reconstructions. We can see from the figure that the correlation patterns from the BSM reconstruction model are very similar to those of the PCR reconstructions and that both approximate the actual spatial correlation structure fairly well.

The bottom right panel, or the last panel depicting the scatter plot of the correlation values of BSM against PCR, shows that the correlation values are nearly identical, with BSM having slightly greater concentration of values above the $y=x$ line; that is, BSM tends to have slightly stronger correlation values. Bracken et al.

(2016) depicts a very similar scatter plot, comparing their model intersite correlations with the intersite correlations from “published regression-based estimates”. Devineni et al. (2013) capture intersite correlations using a covariance matrix to model the spatial dependence between the streamflow gauges, and Bracken et al. (2016) capture intersite correlations using a Gaussian elliptical copula.

The correlation matrices demonstrate that the spatial correlation structure evident in the reconstructions is actually stronger than that of the observations. This is likely because the selected tree ring chronologies for each gauge were chosen from such a wide radius, leading to trees that were common to different gauges or trees among the ones chosen that, even if they differed from gauge to gauge, were sufficiently close to other gauges that they contained spatial information from that other gauge and carried that information into the streamflow reconstruction of the local gauge. The spatial correlation structure appears to be a bit stronger in the BSM model than the PCR model, but this is difficult to decipher from the correlation matrix plots by themselves. The scatter plot of these same correlation values against the distances between each of the gauges (bottom left panel of Figure 12) shows the expected result of the correlation between gauges diminishing as the distance between the gauges grows. This is easily seen from the lowess curve placed on the scatterplot. We can also see from this graph that the trends are virtually identical between the correlated observations, BSM model reconstructions and PCR reconstructions over the 1947–1989 period. Finally, we can decisively confirm that the correlation magnitudes are roughly equal for the closest gauges for both the PCR and BSM models. However, for the more distant gauges, the BSM and PCR lowess curves begin to diverge, and show that spatial correlation is ultimately stronger for the BSM model, as the correlation magnitudes decay faster for the PCR model (and a bit closer to those of the gauge records at distances >200 km) as the intersite distances increase. This is not surprising, as the BSM model uses the same trees to inform its reconstructions, but includes the streamflow linkages as additional predictors for certain gauges, hence increasing the spatial correlation values. Overall, we observe that the reconstructed flows by BSM and PCR retain a pattern of variability that is reasonably close to the actual. This is important in ensemble streamflow simulations, which in turn is important for reservoir operations and water release decisions.

7. Summary and Conclusions

A novel model is introduced and demonstrated for reconstructing streamflow at multiple gauges in a watershed accounting for the topology and spatial dependence across the river network. The model becomes a compact, mathematical representation of the stochastic process induced by the river network and by the tree ring chronologies on to the streamflow process. Once the parameters are estimated using the training data set, the reconstructions are generated as simulations from estimated joint probability distribution, with appropriate spatial dependence across the sites, as informed by both the trees and the streamflow network. The reconstructions for the streamflow network in the Upper Missouri River Basin consisting of 18 gauges demonstrated the utility of the approach in improving the model performance as one moved through the network, and in reproducing the spatial dependence structure of the flows, which is vital for basin-level reservoir reliability analysis.

It was found that the adjusted R^2 values of the reconstructions were within the 0.40–0.60 range, which is typical of most reconstruction efforts. It was also found that the uncertainty in the reconstructions was reduced by more than 50% for streamflow gauges that had feeders when compared to the PCR reconstruction model. For gauges that did not have feeders, the uncertainty was either increased or decreased by less than 5% when compared to the PCR model, giving a reconstruction that was as good as those given by the traditional PCR model. Based on the improvement in the relative uncertainties for streamflow gauges that had feeder streams versus those that only relied on local trees for reconstruction (Figure 5) and the improved performance of BSM in cross-validation against PCR, we have demonstrated that factorizing the spatial dependence structure using the spatial Markov model aligned with the drainage network reduces uncertainties and increases accuracy. The network Bayesian approach taken here provides uncertainty estimates through the posterior distribution estimates that constitute its output, so that uncertainty can be measured and better understood. However, we also note that the enhanced performance is seen in gauges that are informed by the upstream gauges in the network and the tree ring chronologies. Where there are only tree ring chronologies to inform the model, BSM and PCR offer similar performance.

Reconstructions for streamflow gauges that do not have feeders are subject to the same disadvantages of an ordinary Gaussian regression which does not pool spatial information at all across gauges based on critical commonalities. This is a consideration for the future that has the potential to result in a model that will improve even further on the results presented here, particularly in reducing uncertainties even further and hence improving accuracy. Bracken et al. (2016) presents a multisite streamflow reconstruction framework that also takes intersite dependencies and the spatial correlation structure between gauges into account by using a Gaussian elliptical copula to capture the multisite joint probability distribution and intersite relationships. It is possible that incorporating this by adding a joint spatial distribution layer in the network or injecting a partial-pooling (Devineni et al., 2013) within the Bayesian network model would improve the reconstructions for the sites that do not have feeders.

Acknowledgments

This research is supported by the National Science Foundation, Paleo Perspective on Climate Change (P2C2) program—Awards 1401698 and 1404188. (1) National Science Foundation, Water Sustainability and Climate (WSC) program—Award: 1360446. (2) U.S. Department of Energy Early CAREER Award DE-SC0018124 for the second author N. D. (3) U.S. Bureau of Reclamation WaterSMART Program (Sustain and Manage America's Resources for Tomorrow). (4) The state of Montana Department of Natural Resources and Conservation. (5) The U.S. Geological Survey Land Resources Mission Area and the North Central Climate Adaptation Science Center. (6) Lamont-Doherty Earth Observatory contribution number 8349. The statements contained within the manuscript/research article are not the opinions of the funding agency or the U.S. Government but reflect the authors' opinions. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. NOAA's National Center for Environmental Information (NOAA NCEI) Paleoclimatology Data (<https://www.ncdc.noaa.gov/paleo-search/study/27334>). The authors can be contacted for any details on the data files. Data used in this research are from (a) naturalized streamflow data, (b) tree ring chronology data, (c) river network: National Water Information System (NWIS) of U.S. Geological Survey (USGS) (<http://waterdata.usgs.gov/nwis/>). The authors would like to thank Nasser Najibi of the City College of New York for his assistance in preparing the ArcGIS maps shown in Figures 2 and 5.

References

- Barnett, F. A., Gray, S. T., & Tootle, G. A. (2010). Upper Green River Basin (United States) streamflow reconstructions. *Journal of Hydrologic Engineering*, 15(7), 567–579. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000213](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000213)
- Bracken, C., Rajagopalan, B., & Woodhouse, C. (2016). A Bayesian hierarchical nonhomogeneous hidden Markov model for multisite streamflow reconstructions. *Water Resources Research*, 52, 7837–7850. <https://doi.org/10.1002/2016WR018887>
- Brekke, L. D., Kiang, J. E., Olsen, J. R., Pulwarty, R. S., Raff, D. A., Turnipseed, D. P., et al. (2009). *Climate change and water resources management—A federal perspective*, Circular (Vol. 1331, p. 65 p. (Also available online at). Reston, VA: U.S. Geological Survey. <https://pubs.usgs.gov/circ/1331/>
- Briffa, K. R., Jones, P. D., Pilcher, J. R., & Hughes, M. K. (1988). Reconstructing summer temperatures in northern Fennoscandia back to A.D. 1700 using tree-ring data from Scots pine. *Arctic and Alpine Research*, 20(4), 385–394. <https://doi.org/10.2307/1551336>
- Cary, L., & Parrett, C. (1996). Synthesis of natural flows at selected sites in the Upper Missouri River basin, Montana, 1928–89. *U.S. Geological Survey Water-Resources Investigation Report*, 95-4261.
- Cayan, D. R., Redmond, K. T., & Riddle, L. G. (1999). ENSO and hydrologic extremes in the western United States. *Journal of Climate*, 12(9), 2881–2893. [https://doi.org/10.1175/1520-0442\(1999\)012<2881:EAHEIT>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<2881:EAHEIT>2.0.CO;2)
- Cook, E. R., & Jacoby, G. C. (1983). Potomac River streamflow since 1730 as reconstructed by tree rings. *Journal of Climate and Applied Meteorology*, 22(10), 1659–1672. [https://doi.org/10.1175/1520-0450\(1983\)022<1659:PRSSAR>2.0.CO;2](https://doi.org/10.1175/1520-0450(1983)022<1659:PRSSAR>2.0.CO;2)
- Cook, E. R., Meko, D. M., Stahle, D. W., & Cleaveland, M. K. (1999). Drought reconstructions for the continental United States. *Journal of Climate*, 12(4), 1145–1162. [https://doi.org/10.1175/1520-0442\(1999\)012<1145:DRFTCU>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<1145:DRFTCU>2.0.CO;2)
- Cook, E. R., Palmer, J. G., Ahmed, M., Woodhouse, C. A., Fenwick, P., Zafar, M. U., et al. (2013). Five centuries of upper Indus River flow from tree rings. *Journal of Hydrology*, 486, 365–375. <https://doi.org/10.1016/j.jhydrol.2013.02.004>
- Day, C. A., & Sandifer, J. (2015). An annual streamflow reconstruction of the Red River, Kentucky using a white pine (*Pinus strobus*) chronology. *Journal of Geography and Earth Sciences*, 3(1), 1–14. <https://doi.org/10.15640/jges.v3n1a1>
- Devineni, N., Lall, U., Cook, E. R., & Pederson, N. (2013). A tree-ring based reconstruction of Delaware River basin streamflow using hierarchical Bayesian regression. *Journal of Climate*, 26(12), 4357–4374. <https://doi.org/10.1175/JCLI-D-11-00675.1>
- Earle, C. J. (1993). Asynchronous droughts in California streamflow as reconstructed from tree rings. *Quaternary Research*, 39(3), 290–299. <https://doi.org/10.1006/qres.1993.1036>
- Epstein, E. S. (1969). A scoring system for probability forecasts of ranked categories. *Journal of Applied Meteorology*, 8(6), 985–987. [https://doi.org/10.1175/1520-0450\(1969\)008<3C0985:ASSFPF%3E2.0.CO;3B2](https://doi.org/10.1175/1520-0450(1969)008<3C0985:ASSFPF%3E2.0.CO;3B2)
- Fritts, H. C. (1976). *Tree rings and climate*, (p. 567). San Diego, CA: Academic Press.
- Galat, D.L., Berry Jr., C.R., Peters, E.J., & White, R.G. (2005). *Missouri River basin*, In Arthur Benke, Colbert Cushing, Rivers of North America, pp. 427–480, Elsevier Academic Press.
- 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 Resources Research*, 45, W06417. <https://doi.org/10.1029/2008WR007201>
- Gedalof, Z., Peterson, D. L., & Mantua, N. J. (2004). Columbia River flow and drought since 1750. *Journal of the American Water Resources Association*, 40(6), 1579–1592. <https://doi.org/10.1111/j.1752-1688.2004.tb01607.x>
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.
- Hamlet, A. F., & Lettenmaier, D. P. (2000). Long-range climate forecasting and its use for water management in the Pacific Northwest region of North America. *Journal of Hydroinformatics*, 2(3), 163–182. <http://jh.iwaponline.com/content/2/3/163>, <https://doi.org/10.2166/hydro.2000.0015>
- Hidalgo, H. G., & Dracup, J. A. (2003). ENSO and PDO effects on hydroclimatic variations of the upper Colorado River basin. *Journal of Hydrometeorology*, 4(1), 5–23. [https://doi.org/10.1175/1525-7541\(2003\)004<0005:EAPEOH>2.0.CO;2](https://doi.org/10.1175/1525-7541(2003)004<0005:EAPEOH>2.0.CO;2)
- Ho, M., Lall, U., & Cook, E. R. (2016). Can a paleodrought record be used to reconstruct streamflow?: A case study for the Missouri River basin. *Water Resources Research*, 52, 5195–5212. <https://doi.org/10.1002/2015WR018444>
- Jensen, F. V., & Nielsen, T. D. (2007). *Bayesian networks and decision graphs*. New York, NY: Springer.
- Li, B., Nychka, D. W., & Ammann, C. M. (2010). The value of multiproxy reconstruction of past climate. *Journal of the American Statistical Association*, 105(491), 883–895. <https://doi.org/10.1198/jasa.2010.ap09379>
- Livneh, B., Hoerling, M.P., Badger, A., & Eischeid, J. (2016). Causes for hydrologic extremes in the Upper Missouri River Basin. NOAA Climate Assessment Report to USACE (39 pp.). Retrieved from <https://www.drought.gov/drought/sites/drought.gov.drought/files/2016-mrb-climate-assessment-report-final.pdf>
- Loader, C. (1999). *Local regression and likelihood*. New York, NY: Springer.
- Lutz, E. R., Hamlet, A. F., & Littell, J. S. (2012). Paleoreconstruction of cool season precipitation and warm season streamflow in the Pacific Northwest with applications to climate change assessments. *Water Resources Research*, 48, W01525. <https://doi.org/10.1029/2011WR010687>
- Ma, L., Liu, T., Wu, Q., Ji, L., & Gao, R. (2014). Tree ring-based reconstruction of October to November runoffs in the Jialai River since 1826. *Journal of Hydrologic Engineering*, 19(1), 140–147. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000763](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000763)

- Maxwell, R. S., Hessel, A. E., Cook, E. R., & Pederson, N. (2011). A multispecies tree ring reconstruction of Potomac River streamflow (950–2001). *Water Resources Research*, 47, W05 512. <https://doi.org/10.1029/2010WR010019>
- Meko, D. M., & Graybill, D. A. (1995). Tree-ring reconstruction of Upper Gila River discharge. *Journal of the American Water Resources Association*, 31(4), 605–616. <https://doi.org/10.1111/j.1752-1688.1995.tb03388.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. *Journal of the American Water Resources Association*, 37(4), 1029–1039. <https://doi.org/10.1111/j.1752-1688.2001.tb05530.x>
- Meko, D. M., Woodhouse, C. A., Baisan, C. A., Knight, T., Lukas, J. J., Hughes, M. K., & Salzer, M. W. (2007). Medieval drought in the Upper Colorado River Basin. *Geophysical Research Letters*, 34, L10705. <https://doi.org/10.1029/2007GL029988>
- Murphy, A. H. (1969). On the “ranked probability score”. *Journal of Applied Meteorology*, 8(6), 988–989. [https://doi.org/10.1175/1520-0450\(1969\)008<0988:OTPS>2.0.CO;3B2](https://doi.org/10.1175/1520-0450(1969)008<0988:OTPS>2.0.CO;3B2)
- Murphy, A. H. (1971). A note on the ranked probability score. *Journal of Applied Meteorology*, 10(1), 155–156. [https://doi.org/10.1175/1520-0450\(1971\)010<0155:ANOTRP>2.0.CO;3B2](https://doi.org/10.1175/1520-0450(1971)010<0155:ANOTRP>2.0.CO;3B2)
- Najibi, N., Devineni, N., & Lu, M. (2017). Hydroclimate drivers and atmospheric teleconnections of long duration floods: An application to large reservoirs in the Missouri River Basin. *Advances in Water Resources*, 100, 153–167. <https://doi.org/10.1016/j.advwatres.2016.12.004>
- Nguyen, H. T. T., & Galelli, S. (2018). A linear dynamical systems approach to streamflow reconstruction reveals history of regime shifts in northern Thailand. *Water Resources Research*, 54(3), 2057–2077. <https://doi.org/10.1002/2017WR022114>
- Nowak, K., Hoerling, M., Rajagopalan, B., & Zagona, E. (2012). Colorado River basin hydroclimatic variability. *Journal of Climate*, 25(12), 4389–4403. <https://doi.org/10.1175/JCLI-D-11-00406.1>
- Patskoski, J., Sankarasubramanian, A., & Wang, H. (2015). Reconstructed streamflow using SST and tree-ring chronologies over the Southeastern United States. *Journal of Hydrology*, 527, 761–775. <https://doi.org/10.1016/j.jhydrol.2015.05.041>
- R Core Team (2015). R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Rao, M. P., Cook, E. R., Cook, B. I., Palmer, J., Uriarte, M., Devineni, N., et al. (2018). Six centuries of upper Indus Basin streamflow variability and its climatic drivers. *Resources Research*, 54, 5687–5701. <https://doi.org/10.1029/2018WR023080>
- 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 Resources Research*, 27(9), 2381–2399. <https://doi.org/10.1029/91WR00690>
- Robert, C., & Casella, G. (2011). A short history of Markov Chain Monte Carlo: Subjective recollections from incomplete data. *Statistical Science*, 26(1), 102–115. <https://doi.org/10.1214/10-STS351>
- Saito, L., Biondi, F., Salas, J. D., Panorska, A. K., & Kozubowski, T. J. (2008). A Watershed modeling approach to streamflow reconstruction from tree-ring records. *Environmental Research Letters*, 3, 024006. <https://doi.org/10.1088/1748-9326/3/2/024006>
- Stockton, C. W. & Jacoby Jr., G. C. (1976). Long-term surface-water supply and streamflow trends in the Upper Colorado River Basin based on tree-ring analyses. National Science Foundation Lake Powell Research Project Bull. 18, 70 pp.
- Timilsena, J., Piechota, T. C., Hidalgo, H., & Tootle, G. (2007). Five hundred years of hydrological drought in the Upper Colorado River Basin. *Journal of the American Water Resources Association*, 43(3), 798–812. <https://doi.org/10.1111/j.1752-1688.2007.00064.x>
- US Department of the Interior, Bureau of Reclamation (2016). MRB fact sheet <https://www.usbr.gov/climate/secure/docs/2016secure/factsheet/MissouriRiverBasinFactSheet.pdf> (accessed: 10/20/2017)
- Watson, T. A., Barnett, F. A., Gray, S. T., & Tootle, G. A. (2009). Reconstructed streamflows for the headwaters of the wind river, Wyoming, United States. *Journal of the American Water Resources Association*, 45(1), 224–236. <https://doi.org/10.1111/j.1752-1688.2008.00274.x>
- Wilks, D. S. (1995). *Statistical methods in the atmospheric sciences* (2nd ed.). Burlington, MA: Elsevier/Academic Press.
- Wise, E., Woodhouse, C. A., McCabe, G., Pederson, G. T., & St-Jacques, J. M. (2018). Hydroclimatology of the Missouri River Basin. *Journal of Hydrometeorology*, 19, 161–182. <https://doi.org/10.1175/JHM-D-17-0155.1>
- Woodhouse, C. A. (2001). A tree-ring reconstruction of streamflow for the Colorado front range. *Journal of the American Water Resources Association*, 37(3), 561–569. <https://doi.org/10.1111/j.1752-1688.2001.tb05493.x>
- Woodhouse, C. A., Gray, S. T., & Meko, D. M. (2006). Updated streamflow reconstructions for the Upper Colorado River Basin. *Water Resources Research*, 42, W05415. <https://doi.org/10.1029/2005WR004455>
- Woodhouse, C. A., & Lukas, J. J. (2006). Multi-century tree ring reconstructions of Colorado streamflow for water resource planning. *Climate Change*, 78(2–4), 293–315. <https://doi.org/10.1007/s10584-006-9055-0>
- Young, K. C. (1994). Reconstructing streamflow time series in Central Arizona using monthly precipitation and tree ring records. *Journal of Climate*, 7(3), 361–374. [https://doi.org/10.1175/1520-0442\(1994\)007<0361:RSTSI>2.0.CO;2](https://doi.org/10.1175/1520-0442(1994)007<0361:RSTSI>2.0.CO;2)