

A Tree-Ring-Based Reconstruction of Delaware River Basin Streamflow Using Hierarchical Bayesian Regression

NARESH DEVINENI

Columbia Water Center, The Earth Institute, Columbia University, New York, New York

UPMANU LALL

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

NEIL PEDERSON AND EDWARD COOK

Tree Ring Laboratory, Lamont-Doherty Earth Observatory, Palisades, New York

(Manuscript received 17 November 2011, in final form 27 December 2012)

ABSTRACT

A hierarchical Bayesian regression model is presented for reconstructing the average summer streamflow at five gauges in the Delaware River basin using eight regional tree-ring chronologies. The model provides estimates of the posterior probability distribution of each reconstructed streamflow series considering parameter uncertainty. The vectors of regression coefficients are modeled as draws from a common multivariate normal distribution with unknown parameters estimated as part of the analysis. This leads to a multilevel structure. The covariance structure of the streamflow residuals across sites is explicitly modeled. The resulting partial pooling of information across multiple stations leads to a reduction in parameter uncertainty. The effect of no pooling and full pooling of station information, as end points of the method, is explored. The no-pooling model considers independent estimation of the regression coefficients for each streamflow gauge with respect to each tree-ring chronology. The full-pooling model considers that the same regression coefficients apply across all streamflow sites for a particular tree-ring chronology. The cross-site correlation of residuals is modeled in all cases. Performance on metrics typically used by tree-ring reconstruction experts, such as reduction of error, coefficient of efficiency, and coverage rates under credible intervals is comparable to, or better, for the partial-pooling model relative to the no-pooling model, and streamflow estimation uncertainty is reduced. Long record simulations from reconstructions are used to develop estimates of the probability of duration and severity of droughts in the region. Analysis of monotonic trends in the reconstructed drought events do not reject the null hypothesis of no trend at the 90% significance over 1754–2000.

1. Introduction

The upper Delaware River Basin System supplies New York City, one of the largest urban water supply systems in the United States. With a cumulative storage capacity of $1.5 \times 10^9 \text{ m}^3$ from five major reservoirs, the Delaware River basin supplies about $3 \times 10^6 \text{ m}^3 \text{ day}^{-1}$ to the city of New York. The Delaware River Basin Commission and the New York City Department of Environmental Protection are primarily responsible for

managing the releases from the major reservoirs to meet the water demand of the city of New York and to maintain downstream ecosystem services (DRBC 2007). The operating rules of this reservoir system (e.g., the minimum water levels to be maintained in the reservoir at any specified time, or the specification of drought return periods) are based on relatively short historical records of data. The typical record length of the naturalized streamflow data for the major reservoirs on the system is 50–60 years. Given that the drought of record in the basin was in the 1960s, that is, about 50 years ago, extended records of hydrologic variability from paleoproxies such as tree rings could be very useful for assessing the likely return period of this drought for regional water supply planning and drought operation.

Corresponding author address: Naresh Devineni, Columbia Water Center, The Earth Institute, Columbia University, New York, NY 10027.

E-mail: nd2339@columbia.edu

DOI: 10.1175/JCLI-D-11-00675.1

© 2013 American Meteorological Society

Recently, impacts on fisheries during the summer low flow period have led to questions concerning the reservoir operating policies that are designed to avert the 1960s drought risk (Kolesar and Serio 2011).

Numerous studies have focused on the use of tree-ring widths for developing proxy climatic and hydrologic series using traditional regression techniques (Stockton and Jacoby 1976; Meko and Graybill 1995). Meko et al. (2007), Woodhouse et al. (2006), and Woodhouse and Lukas (2006a,b) used a stepwise linear regression approach to develop multicentury reconstructed streamflows and investigate the medieval drought in the upper Colorado River basin. Similarly, over the Northeast, Cook and Jacoby (1983) used canonical regression analysis to reconstruct the July–September streamflow for the Potomac River using tree-ring chronologies from nearby sites. Cook and Jacoby (1977) also examined the drought in the Hudson River Valley by reconstructing the Palmer Drought Severity Index (PDSI) using a stepwise regression analysis. Recently, Maxwell et al. (2011) reconstructed the Potomac River streamflow dating back to 950 using a network of tree-ring chronologies from multiple species. Kauffman and Vonck (2011) investigated the frequency and intensity of extreme drought over the lower Delaware River basin, specifically at the mouth of the Delaware River using a reconstructed PDSI.

These paleoreconstruction methods use a regression model fit to the observed streamflow using tree-ring chronologies as predictors. The streamflow data in the preinstrumental (paleo) period are then obtained by applying the estimated regression coefficients to the paleo-period tree-ring indices. The paleoreconstruction often considers multiple proxies and multiple hydroclimatic records to be reconstructed [e.g., gridded PDSI reconstruction as in Cook et al. (1999) or temperature reconstruction over spatial grids using multiple proxies as in Tingley and Huybers (2010a,b)]. The resulting multivariate regression problem can be high dimensional. Given a finite dataset, a practical question is how best to estimate parameter uncertainty. Further, the records often have varying length, and long gaps in data can also pose estimation problems. In this paper we consider only the continuous common record and do not consider expectation–maximization (EM) or related algorithms (Dempster et al. 1977; Schneider 2001) for gap filling. A successful model also needs to preserve the correlation of streamflow across sites to properly constrain stochastic simulations of multireservoir operation (Gangopadhyay et al. 2009).

In this paper, we present a hierarchical Bayesian regression (HBR) model for inferences on the posterior probability distribution of the regression coefficients

and streamflow values at multiple locations of interest using recently developed tree-ring chronologies in the upper Delaware River basin. A multilevel model framework that provides an elegant means of propagating the parameter uncertainties through appropriate conditional distributions is adopted. Further, noting that multiple correlated predictors (regional tree-ring chronologies) from different species may inform streamflow reconstruction in a similar way, the hierarchical model provides for partial pooling of this common information. Partial pooling reduces the equivalent number of independent parameters, resulting in lower uncertainty in parameter estimates, and therefore leads to reduced uncertainty in the reconstructed streamflows. The multilevel or partial-pooling approach improves on estimation on full pooling, which ignores the cross-site variations in response, and on no pooling, which estimates independent regressions across the sites. Those two cases are subsets or end points of the model developed, and are compared as such.

Hierarchical Bayesian models have been used previously in the context of climate field reconstruction over spatial grids (Tingley and Huybers 2010a,b) and reconstructing Northern Hemisphere temperature data using proxy datasets such as tree-ring measurements, pollen indices, and borehole temperatures, among others. (Li et al. 2010). Similarly, dynamic Bayesian space–time models have also been used to develop long lead forecasting for tropical Pacific SSTs (Berliner et al. 2000). Several other hydrologic applications have been developed and demonstrated in Lima and Lall (2009, 2010) and Kwon et al. (2008, 2011). Readers can also refer to Wikle (2003), Raftery (1995), Gelman and Hill (2007), and Gelman et al. (2004) for additional information on hierarchical Bayesian model applications.

The paper is organized as follows. A brief description of the streamflow and tree-ring chronology data used in the study is provided in section 2. Section 3 contains a description of the proposed hierarchical Bayesian regression model and section 4 presents the results and analysis from the model. The drought characterization using the reconstructed streamflow data for the region is presented in section 5. Finally, in section 6, the key results are discussed and summarized.

2. Data description

a. Streamflow data

The location of the five major reservoirs, selected stream gauges, and the tree-ring sites in the upper Delaware River basin (DRB) is shown in Fig. 1. The DRB extends roughly 532 km from its confluence of the East

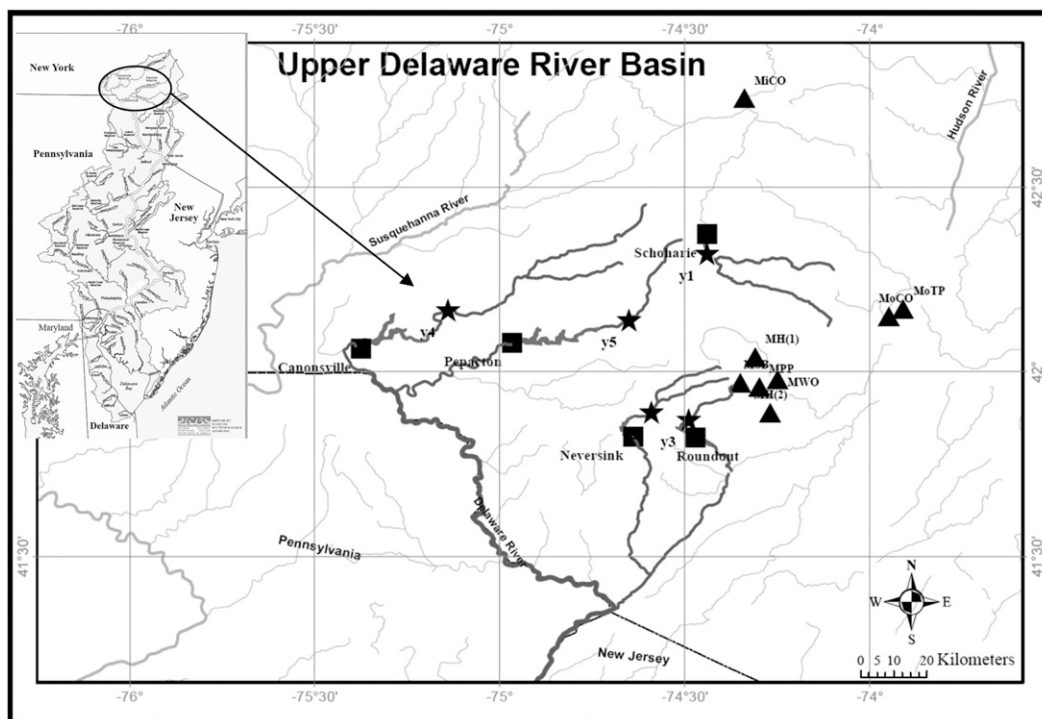


FIG. 1. Location of the tree-ring chronology sites (filled triangle marker), five major reservoirs (filled square marker), and the respective upstream gauges (star marker) in the upper Delaware River basin (DRB). The inset in the figure shows the location of the upper DRB in New York.

and West Branches in New York to the mouth of the Delaware Bay encompassing $\sim 34\,965\text{ km}^2$ and includes the states of New York, New Jersey, Pennsylvania, and Delaware. In this study, we consider five major reservoirs in the upper DRB that serve as the primary water supply systems for the city of New York. Specifics of the U.S. Geological Survey (USGS) stream gauges on the major tributaries corresponding to inflow into each reservoir are provided in Table 1. For the purpose of this study, we selected the stream gauges (from the USGS National Water Information System) on the major creeks feeding into the reservoir such that the inflows are not influenced by any upstream diversions or regulations. The drainage area and record length vary across the stations (Table 1). The Schoharie creek (USGS gauge 1350000) has the longest (98 yr) record. All other stations have data records in the range of 50–64 years.

b. Tree-ring data

Table 2 shows the details of the seven new and one older collection [hemlock(2), *Tsuga canadensis*, in Mohonk, New York] of tree-ring chronologies developed from forests in the upper DRB. Note the multiple chronologies for each site. The chronology site locations are presented in Fig. 1. Of the seven chronologies, one was used in Pederson et al. (2004) (pitch pine in

Mohonk), three were developed and used in Pederson (2005) (*Liriodendron tulipifera* and *Quercus prinus* in Montgomery Place and *Quercus prinus* in Middleburg), and three were used (now updated) in Cook and Jacoby (1977) [hemlock(1), *Tsuga canadensis*, pitch pine, *Pinus rigida*, and *Quercus* subgenus *leucobalanus* in Mohonk]. The hemlock(2) chronology was developed for the North American Drought Atlas (Cook et al. 1999, 2010). The recently updated Mohonk records, except for the *Betula lenta* chronology, are published here for the first time and are available from the International Tree-Ring Databank in their original form.

Recent research indicates that the larger number of species used for tree-ring-based reconstructions could enhance the final reconstruction (Cook and Pederson 2010; Maxwell et al. 2011; Pederson et al. 2013). All chronologies used here have also been shown to be useful for drought index reconstructions by Cook and Pederson. All time series of tree-ring measurements were processed using standard techniques (Stokes and Smiley 1968; Fritts 1976; Cook 1985; Cook and Kairiukstis 1990). Ring width series with growth distortions, rotten sections, or other gaps, including series from the data bank, were filled using the gap-filling option in auto-regressive standardization (ARSTAN) (see Pederson et al. 2004). All series were transformed using the adaptive power

TABLE 1. Details of the stream gauges on major tributaries and the corresponding reservoir systems in the Delaware River basin. These reservoirs serve as the primary water supply systems for the city of New York. The table shows the number of years of streamflow record and the drainage area corresponding to each stream gauge.

Reservoir	Feeder creek	Stream gauge	Data record	Number of years	Drainage area (km ²)
Schoharie	Schoharie	1350000	1903–2000	98	611
Neversink	Neversink	1435000	1937–2000	64	173
Roundout	Roundout	1365000	1937–2000	64	98
Canonsville	West Branch Delaware River	1423000	1950–2000	51	857
Pepacton	East Branch Delaware River	1413500	1937–2000	64	421

transformation and then standardized to conserve as much long-term variation in ring widths as possible while reducing the influence of nonclimatic forces such as changes in competition (Pederson et al. 2004). The “Friedman Super Smoother” was the primary option used to reduce the influence of disturbance in each series (Friedman 1984; Buckley et al. 2010). The Friedman Super Smoother would occasionally cause ring index inflation or deflation at either end of a series. In these few cases, a cubic smoothing spline two-thirds the length of each series was used (Cook and Peters 1981). The mean value for each year was calculated using a biweight robust function following standardization (Cook 1985).

c. Diagnostic analysis and predictor selection

The tree-ring chronologies represent the annual growth cycle of the trees resulting from less dense (inner portion) early-wood formation during the photosynthetically active growing season (late spring and summer) and the more dense (outer portion) late-wood formation during the fall and winter. These chronologies vary in size each year depending upon the regional climate phenomena. Consequently, the tree rings (measured as the width of early wood plus late wood) are wider during years with adequate moisture availability and narrow during drought years. Hence, analogous to streamflow, the growth index is an integrator of moisture and energy availability in the region. This

commonality between annual growth index and streamflow enables us to develop predictive models that can be utilized to understand the long-term variability of the climate in the region. The summer season average [June–August (JJA)] streamflow for each of the five major gauges in the upper DRB was identified for reconstruction under the hypothesis that growing season of the trees, concurrent to the streamflow, may present the best sensitivity across flows and trees. This relatively dry period is also critical for reservoir operations given the fishing and ecological impacts. Preliminary analyses of the seasonality of inflows (Fig. 2a) show that typically 45% of annual inflows occur during March–May and the flows during JJA contribute 20% of the annual inflows. From a water management perspective, developing reconstructed inflows for the summer season is important to assess the frequency and recurrence of severe droughts and to better quantify the operational rule curves for downstream release purposes (Kolesar and Serio 2011). Summer (JJA) is also the growth season when the trees are photosynthetically active and thus most sensitive to moisture limitations and loss through transpiration.

For a preliminary assessment of this hypothesis, we computed the Pearson correlation coefficient between the tree-ring chronologies and 1) the annual average streamflows and 2) the JJA streamflows for the five stations (Fig. 2). The correlation coefficients statistically

TABLE 2. Details of the tree-ring chronology data used in the study. The information regarding the site, species, number of trees per site and number of cores sampled is given in the table. The species are TSCA (*Tsuga canadensis*), QUSP (*Quercus* subgenus *leucobalanus*), BELE (*Betula lenta*), PIRI (pitch pine), QUPR (*Quercus prinus*), and LITU (*Liriodendron tulipifera*). The Mohonk records are available from the International Tree-Ring Databank (ITRDB) in their original form.

Tree-ring chronology	Site (New York)	Species	Number of trees	Number of cores	Actual data record	Data record used (247 yr)
MH(1)	Mohonk	TSCA	25	43	1626–2002	1754–2000
MWO	Mohonk	QUSP	20	34	1450–2002	1754–2000
MH(2)	Mohonk	TSCA	18	25	1658–2002	1754–2000
MSB	Mohonk	BELE	17	27	1614–2002	1754–2000
MPP	Mohonk	PIRI	23	45	1618–2002	1754–2000
MoCO	Montplace	QUPR	21	34	1727–2002	1754–2000
MoTP	Montplace	LITU	20	32	1754–2002	1754–2000
MiCO	Middlebergh	QUPR	23	42	1507–2000	1754–2000

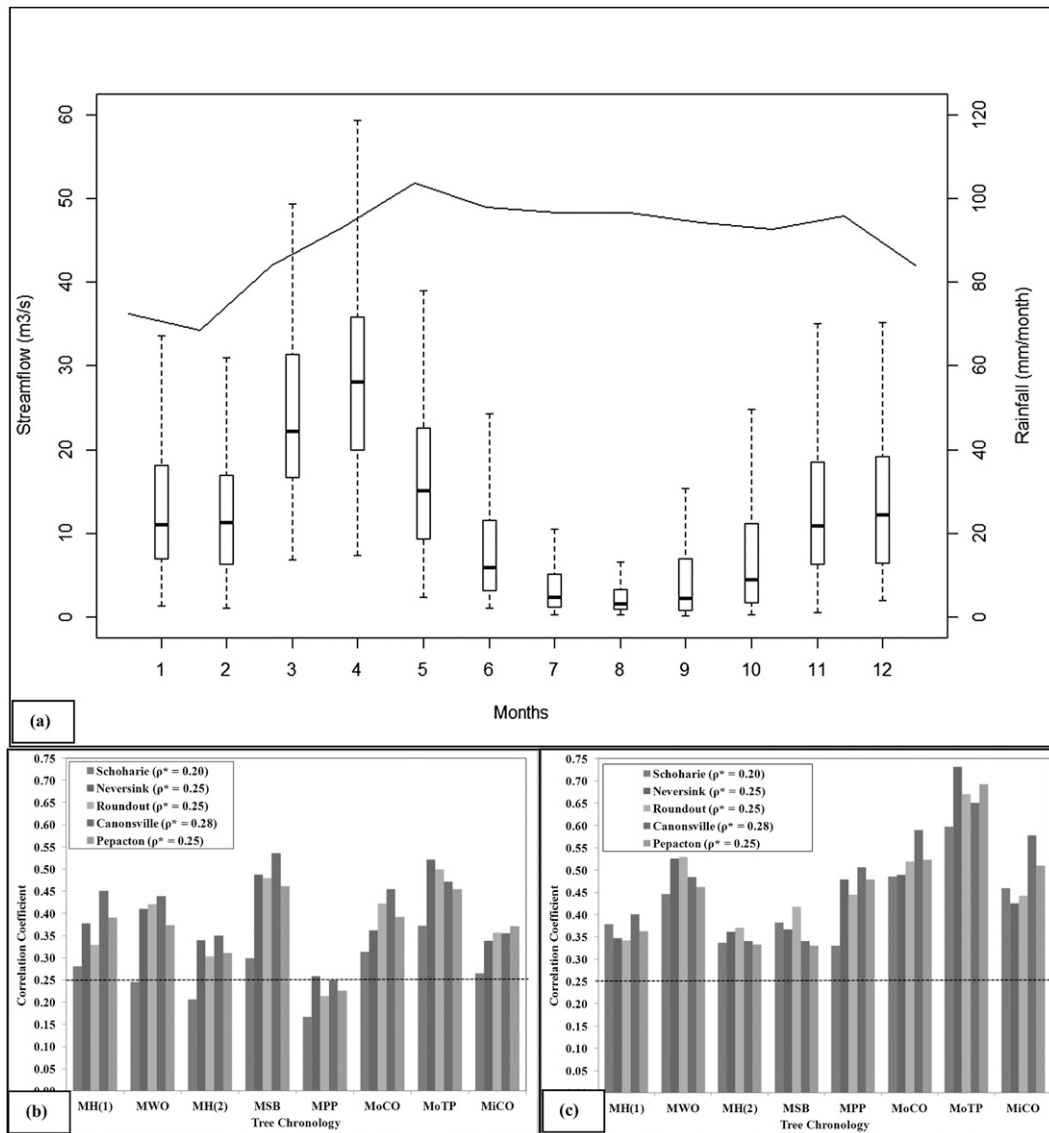


FIG. 2. Diagnostic analysis of streamflows and tree-ring chronologies: (a) boxplot of monthly flows and plot of the mean monthly rainfall for the Schoharie station; (b) correlation coefficient between tree rings and annual average flows; and (c) correlation between tree rings and summer (JJA) season average streamflow at the five selected stations. The horizontal line marks the one-sided 95% significance level for correlation.

significant at the 95% confidence interval for a given sample size are highlighted. One can see that all tree chronologies correlate better with the summer than with annual streamflow. The tree ring–streamflow correlation with winter/spring flows was also lower than for the summer flows. Regional studies also show that the tree-ring response to regional climate is greatest during the active growing season (summer) (Cook and Jacoby 1977, 1983; Maxwell et al. 2011; Pederson et al. 2013).

One can also observe from Fig. 2c that there is a potential opportunity for “grouping” or pooling the relationships across trees and streamflow. For instance,

the correlation coefficient between the annual tree-ring growth index of hemlock [MH(1)] and the summer average streamflow of the five stations is in the range 0.35–0.4, indicating that the hemlock relationship to streamflow is similar across the five stations. Similarly, the correlation of the flows of the five stations to the tulip poplar species (MoTP) ranges from 0.6 to 0.75. This suggests that pooling the regression coefficients across stations with respect to a specific tree-ring chronology may be useful, while pooling regression coefficients for a streamflow station across different tree chronologies may not be as effective.

We also investigated the possibility of lagged correlation (e.g., $t-1$, $t-2$, $t-3$) between tree rings and streamflow that could result from longer use of stored energy in the trees from prior growth (Trumbore et al. 2002; Kagawa et al. 2006). However, we did not find any statistically significant relations (results not shown) at these lags. Hence, only the tree-ring chronologies of the current year were used to predict the summer seasonal streamflows.

The Shapiro–Wilk normality test (Royston 1995) applied to the log transformed summer seasonal (JJA) average flows for each station did not reject the null hypothesis of the transformed values being normally distributed at the 5% level (p values ranged from 0.05 to 0.27). The Box–Cox transform led to the same conclusion. The log transformed streamflows were modeled as the response variables, but all subsequent model validation results are presented in terms of the real space summer flows.

3. Hierarchical Bayesian regression: Methodology

A general multilevel modeling structure that allows pooling of information across stream gauges for regression on available tree-ring chronologies, and considers correlation of model residuals across stations was explored under the hierarchical Bayesian regression framework. We term the general model the “partial pooling model.” Several subsets of this model were explored to develop intuition as to how different end points of this model perform. We consider the following subsets: a no-pooling model, where regression coefficients for each streamflow series are modeled independently, and a full-pooling model, where all streamflow sequences have the same regression coefficient for a specific tree-ring series. In each case, we considered estimation of the full covariance matrix across sites of residuals as well as a diagonal structure that treats them as uncorrelated.

a. The general model (partial pooling)

$$y_{i,t} | \alpha_i, \boldsymbol{\beta}_i = \alpha_i + \mathbf{X}_t \boldsymbol{\beta}_i + \varepsilon_{i,t} \quad i = 1, \dots, 5,$$

$$\varepsilon_{i,t} | \boldsymbol{\Sigma}_e \sim N(0, \boldsymbol{\Sigma}_e), \quad \text{and}$$

$$\boldsymbol{\beta}_i | \boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta \sim \text{MVN}(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta),$$

with priors modeled as

$$\begin{aligned} \alpha_i &\sim N(0, 10^4), \\ \boldsymbol{\mu}_\beta &\sim \text{MVN}(0, 10^4 \mathbf{I}), \\ \boldsymbol{\Sigma}_\beta &\sim \text{Inv-Wishart}_{\nu_0}(\boldsymbol{\Lambda}_0), \quad \text{and} \\ \boldsymbol{\Sigma}_e &\sim \text{Inv-Wishart}_{\nu_1}(\boldsymbol{\Lambda}_1). \end{aligned} \quad (1)$$

Equation (1) represents the regression of $y_{i,t}$ (log of the streamflow) on \mathbf{X}_t , the vector of the eight predictors (tree-ring growth indices) for year t . MVN stands for multivariate normal distribution. The $\boldsymbol{\beta}_i$ is a vector of the eight regression slopes relating these predictors to the streamflow at station i . The α_i are the intercepts of the regression model at each station and can be interpreted as the mean streamflow at site i . Since the drainage area of the river basin varies across sites, the α_i are not expected to have a common mean. They could be modeled as a function of the drainage area, as in Lima and Lall (2010), but given the small number of sites we did not consider that for the current application. The regression errors $\varepsilon_{i,t}$ are assumed to be normally distributed and correlated across the stream gages. Their covariance across streamflow sites is modeled through the 5×5 matrix $\boldsymbol{\Sigma}_e$. The $\boldsymbol{\beta}_i$ can be interpreted as the streamflow sensitivities to the transient climate as recorded in the tree-ring chronologies.

Noting that the correlation of a given tree-ring chronology is very similar across the five stations; we consider a multilevel model that allows for pooling of information across stations for a given tree for estimating the regression slope parameters to reduce the associated uncertainty. The model has a multilevel structure where the model parameters are presumed to be drawn from a common distribution, whose parameters (e.g., the $\boldsymbol{\beta}_i$) are in turn described by a set of hyperparameters (e.g., $\boldsymbol{\mu}_\beta$ and $\boldsymbol{\Sigma}_\beta$). Here $\boldsymbol{\mu}_\beta$ is a vector of length 8 representing the common regression slope for each of the five stations for each of the eight tree-ring chronologies. Correspondingly, $\boldsymbol{\Sigma}_\beta$ is an 8×8 covariance matrix of the $\boldsymbol{\mu}_\beta$. This representation allows partial pooling across the stations by shrinking the estimates of $\boldsymbol{\beta}_i$ toward a common mean $\boldsymbol{\mu}_\beta$ (Gelman and Hill 2007), with dispersion matrix $\boldsymbol{\Sigma}_\beta$, estimated as part of the solution.

The priors for the covariance matrices ($\boldsymbol{\Sigma}_\beta$ and $\boldsymbol{\Sigma}_e$) are taken to be the inverse Wishart distribution with scale matrices ($\boldsymbol{\Lambda}_0$ and $\boldsymbol{\Lambda}_1$) and ν_0 and ν_1 degrees of freedom. In our applications, the scale matrices $\boldsymbol{\Lambda}_0$ and $\boldsymbol{\Lambda}_1$ were specified as an identity matrix \mathbf{I} and the degrees of freedom ν_0 and ν_1 were set to one more than the dimension of the matrix (i.e., the total number of predictors, 8, for $\boldsymbol{\Sigma}_\beta$ and total number stations, 5, for $\boldsymbol{\Sigma}_e$) to induce a uniform prior distribution on the variance (Gelman and Hill 2007). This choice of conjugate priors was made for computational convenience (Gelman 2005). The joint posterior distribution $p(\boldsymbol{\theta} | y)$ for the partial-pooling case of the complete parameter vector $\boldsymbol{\theta}$ [which includes regression intercepts (α_i), slopes ($\boldsymbol{\beta}_i$), and covariance ($\boldsymbol{\Sigma}_e$)] is derived by defining the posterior distribution function $p(\boldsymbol{\theta} | y)$ as follows:

$$\begin{aligned}
 p(\boldsymbol{\theta} | y) \propto & \prod_{i=1}^5 \prod_{t=1}^{n_i} N[y_i(t) | \alpha_i + \boldsymbol{\beta}_i \mathbf{X}(t), \boldsymbol{\Sigma}_e] \text{Inv-Wishart}(\boldsymbol{\Sigma}_e | \nu_1, \boldsymbol{\Lambda}_1) N(\alpha_i | 0, 10^4) \\
 & \times N(\boldsymbol{\beta}_i | \boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta) N(\boldsymbol{\mu}_\beta | 0, 10^4) \text{Inv-Wishart}(\boldsymbol{\Sigma}_\beta | \nu_0, \boldsymbol{\Lambda}_0).
 \end{aligned} \tag{2}$$

Here n_i is the number of observations at station i available for fitting the model.

b. No-pooling model

$$\begin{aligned}
 y_{i,t} | \alpha_i, \boldsymbol{\beta}_i &= \alpha_i + \mathbf{X}_i \boldsymbol{\beta}_i + \varepsilon_{i,t}; \quad i = 1, \dots, 5 \quad \text{and} \\
 \varepsilon_{i,t} &\sim N(0, \boldsymbol{\Sigma}_e),
 \end{aligned}$$

with priors modeled as

$$\begin{aligned}
 \alpha_i &\sim N(0, 10^4), \\
 \boldsymbol{\beta}_i &\sim \text{MVN}(0, 10^4 \mathbf{I}), \quad \text{and} \\
 \boldsymbol{\Sigma}_e &\sim \text{Inv-Wishart}_{\nu_1}(\boldsymbol{\Lambda}_1).
 \end{aligned} \tag{3}$$

In the no-pooling case, the regression coefficients $\boldsymbol{\beta}_i$ are allowed to be estimated independently for each streamflow station. To allow a comparison with the partial-pooling model, we modeled the covariance structure of the errors $\varepsilon_{i,t}$ across the stations. We choose weakly informative prior distributions (Gelman and Hill 2007; Gelman et al. 2004) for the model parameters. The priors for α_i and $\boldsymbol{\beta}_i$ are taken to be normal with a mean of zero and a large variance. Similar to the partial-pooling scheme, the prior distribution for the covariance matrix $\boldsymbol{\Sigma}_e$ is taken to be the inverse Wishart distribution with a scale matrix $\boldsymbol{\Lambda}_1$ and ν_1 degrees of freedom. The joint posterior likelihood $p(\boldsymbol{\theta} | y)$ for the no-pooling case is given as follows:

$$p(\boldsymbol{\theta} | y) \propto \prod_{i=1}^5 \prod_{t=1}^{n_i} N[y_i(t) | \alpha_i + \boldsymbol{\beta}_i \mathbf{X}(t), \boldsymbol{\Sigma}_i] \text{Inv-Wishart}(\boldsymbol{\Sigma}_i | \nu_1, \boldsymbol{\Lambda}_1) N(\alpha_i | 0, 10^4) N(\boldsymbol{\beta}_i | 0, 10^4 \mathbf{I}). \tag{4}$$

c. Full-pooling model

$$\begin{aligned}
 y_{i,t} | \alpha_i, \boldsymbol{\beta} &= \alpha_i + \mathbf{X}_i \boldsymbol{\beta} + \varepsilon_{i,t}; \quad i = 1, \dots, 5 \quad \text{and} \\
 \varepsilon_{i,t} &\sim N(0, \boldsymbol{\Sigma}_e),
 \end{aligned}$$

with priors modeled as

$$\begin{aligned}
 \alpha_i &\sim N(0, 10^4), \\
 \boldsymbol{\beta} &\sim N(0, 10^4), \quad \text{and} \\
 \boldsymbol{\Sigma}_e &\sim \text{Inv-Wishart}_{\nu_1}(\boldsymbol{\Lambda}_1).
 \end{aligned} \tag{5}$$

The full-pooling model has the regression coefficient for all stations for a given tree; that is, $\boldsymbol{\beta}_i = \boldsymbol{\beta}$. The posterior likelihood $p(\boldsymbol{\theta} | y)$ is given as follows:

$$p(\boldsymbol{\theta} | y) \propto \prod_{i=1}^5 \prod_{t=1}^{n_i} N[y_i(t) | \alpha_i + \boldsymbol{\beta} \mathbf{X}(t), \boldsymbol{\Sigma}_e] \text{Inv-Wishart}(\boldsymbol{\Sigma}_e | \nu_1, \boldsymbol{\Lambda}_1) N(\alpha_i | 0, 10^4) N(\boldsymbol{\beta} | 0, 10^4). \tag{6}$$

As before, the error covariance across stations is estimated. This allows for a consistent model structure across the three schemes.

For each model, the parameters $\boldsymbol{\theta}$ are estimated using WinBUGS (the Windows-based version of Bayesian inference using Gibbs sampling, Lunn et al. 2000; Spiegelhalter et al. 1996). WinBUGS employs the Gibbs sampler, a Markov chain Monte Carlo (MCMC) method, for simulating the posterior probability distribution of the parameters conditional on the current choice of parameters and the data. The Gibbs sampler sequentially samples one parameter from the conditional distribution of that parameter relative to the others and provides an effective sampling-based numerical solution

for parameter estimation (Gilks and Roberts 1995). We simulated five chains starting from random initial values for the parameters to verify the convergence of the posterior distribution based on the shrink factor suggested by Gelman and Rubin (1992). The shrink factor compares the variance in the sampled parameters within the chains and across the chains to describe the improvement in the estimates for an increasing number of iterations. Gelman and Rubin suggest running the chains until the estimated shrink factors are less than 1.1 for all the parameters. For this application, each chain was run for a 500-cycle burn in to discard the initial state followed by 15 000 simulations of model parameters until the shrink factor was close to 1. The R2WinBUGS and WinBUGS

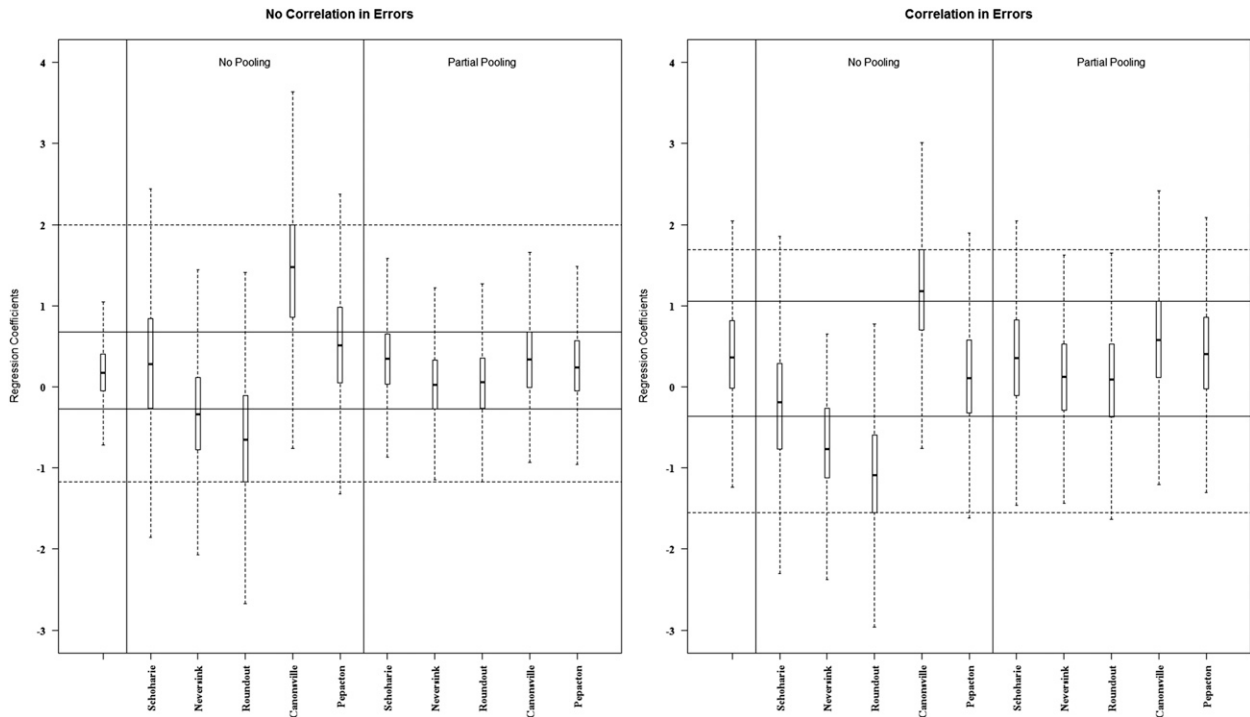


FIG. 3. Boxplots of the posterior distribution of the regression coefficients (i.e., β_j) from the full-pooling, no-pooling, and partial-pooling models for MH(1) species (*Tsuga canadensis*) for all stations: (left) diagonal covariance matrix for residuals and (right) full covariance matrix for residuals. The interquartile range across the stations for each model scheme is shown as a horizontal line (solid for partial pooling, and dashed for no pooling).

codes with detailed instructions and the relevant data to implement the above described simulation can be found at <http://water.columbia.edu/education/software/>.

4. Results and analysis

The fit of the three models (no pooling, full pooling, and partial pooling) was compared initially using the Bayesian deviance statistics, that is, the measure of lack of fit \bar{D} , the effective number of parameters pD , and the deviance information criteria (DIC) computed using the posterior distributions of the model parameters (Gelman et al. 2004). The \bar{D} statistic defined as $\bar{D} = -2 \log p(y | \theta)$ is a measure of the model's lack of fit, expressed as a negative factor multiplied by the log likelihood averaged over the posterior samples of the data given the model and its parameters. Lower \bar{D} indicates better model fit. For the no-pooling, full-pooling, and the partial-pooling models, the \bar{D} values are 107, 113, and 94, respectively. Similarly, the pD (defined as $\bar{D} - \hat{D}$), a measure of the effective number of parameters in the model for the no-pooling, full-pooling, and partial-pooling models are 60, 28, and 45, respectively. For the no-pooling and full-pooling cases the number of model parameters can easily be estimated from the

model fit. For example, in the no-pooling case, the model has 5 regression intercepts, 40 (5 stations \times 8 chronologies) regression coefficients or slopes, and 15 variance-covariance terms. Note that \hat{D} is similar to \bar{D} but is a point estimate computed on the parameter means. The reduction in pD for the partial-pooling model is seen as a result of pooling of information in the multilevel structure. Finally, DIC ($=\bar{D} + pD$) provides an assessment of the model predictive skill. Smaller values of the DIC indicate better estimated predictive capacity of the model. The DIC of the no-pooling, full-pooling, and partial-pooling models are 169, 141, and 139, indicating that the partial-pooling HBR model has better predictive capacity compared to the no-pooling or full-pooling regression models. A detailed discussion on the Bayesian deviance statistics can be found in Gelman et al. (2004).

a. Comparison of regression parameters and the reconstructed flows

The posterior probability distributions of the regression slope parameter vector (i.e., β_j) for the five streamflow stations from the no-pooling, full-pooling, and partial-pooling models for the *Tsuga canadensis* species are illustrated in Fig. 3. The posterior distributions of the

regression coefficients under two model schemes (with correlated errors and without correlated errors) are shown in separate plots. In each subset, the first column corresponds to the regression coefficient for the full-pooling case, where all stations have the same coefficient. This is followed by the boxplots of the regression coefficients for the five stations under no pooling and the five stations under partial pooling. The results for uncorrelated residuals are presented in Fig. 3a and for the case where the covariance matrix is modeled are presented in Fig. 3b. The interquartile range (IQR) from the full-pooling, no-pooling, and partial-pooling models for the estimated β_i is shown in Table 3. Note that the interquartile range across the β_i for a particular tree shrinks substantially across the sites for the partial-pooling case relative to no pooling. Also, the IQR increases for the full-pooling and partial-pooling models as the correlation of the residuals is estimated, as expected given the additional parameters to be estimated.

The reduction in variance of the regression coefficients in turn leads to a reduction in the uncertainty in the streamflow estimates. An examination of the correlation structure resulting from the posterior distribution from the no-pooling, full-pooling, and partial-pooling models of the resulting $y_i(t)$ indicated the need for modeling the full 5×5 covariance matrix Σ_e . The uncertainty in estimating the coefficients is larger for the partial-pooling case where the covariance is modeled explicitly, but still less than for the no-pooling case. Subsequent results are presented only for the case where the full covariance matrix is modeled.

A boxplot of the posterior distribution of the mean of the vector of regression coefficients across sites (i.e., μ_β) for each tree species from the joint normal distribution is shown in Fig. 4. Most of the chronologies [except the tree rings from the MH(2) (*Tsuga canadensis*) chronology] have positive coefficients. Given the high correlation in the chronologies of the tree species [0.8 for MH(1) and MH(2)], the coefficients are expected to be negatively correlated. The off-diagonal elements in the estimated Σ_β are typically small, suggesting that even though the tree-ring chronologies are correlated with each other, the correlation of the μ_β is weak.

The posterior probability distributions of the reconstructed flows from the no-pooling approach and the partial-pooling approach for the Roundout and Canonsville stations during the period 1754–2000 are compared in Fig. 5. For the sake of brevity, all subsequent results are presented for only the no-pooling and partial-pooling models, with results from the full-pooling model discussed only where appropriate. Predicting or extrapolating data back in time using the tree-ring indices may be associated with a larger uncertainty due to potential

extrapolation of the fitted data. From a comparison of the uncertainty bands (5th and 95th percentiles) in Fig. 5, we can see that the partial-pooling HBR approach results in a modest reduction in uncertainty in estimating the posterior distribution of the flows. This can be also be seen from Fig. 5c, which shows the boxplot of the width (i.e., the difference of the 95th percentile and the 5th percentile) for the 247 years from the no-pooling and partial-pooling models. The reduction in uncertainty is similar for other stations.

The correlation of flow across streamflow stations is also estimated from the posterior distribution of the flows. The observed correlation between stations ranged between 0.92 (for Neversink and Roundout) and 0.74 (for Neversink and Canonsville). The median correlation between Neversink and Roundout estimated from the 1000 posterior draws is 0.89 with the interquartile range between 0.88 and 0.93. Similarly, the median of the correlation between Neversink and Canonsville estimated from the 1000 posterior draws is 0.72 with the interquartile range between 0.65 and 0.79. Results for other stations are similar. In the next section, we present the results from cross-validation over varying calibration periods using performance statistics common to the tree-ring reconstruction literature.

b. Validation tests under varying calibration periods

We used two performance metrics, reduction of error (RE) and coefficient of efficiency (CE), as measures of model performance to compare the reconstructed posterior mean of the streamflow estimates with the actual streamflow data. These metrics were estimated using the leave- m -out cross-validation method. The procedure is carried out by leaving out m randomly selected data points from the observational dataset for validation, and the model is developed using the remaining $(n - m)$ observations (n is the total number of observational data points). This process is repeated several times to obtain an ensemble of validation metrics resulting from each randomly selected model. We used the 50-yr (1950–99) common data period across all streamflow stations for this purpose. The cross-validation approach taken was to draw a sample of size 40 from the 50-yr common record without replacement and fit the Bayesian model on this dataset: then predictions were made on the 10 observations that were left out. This procedure was repeated 50 times to compute the validation statistics. Note that analysts who fit Bayesian models are typically interested in the coverage rates (discussed later), uncertainty levels, and model checking using the posterior draws. However, we include comparisons across the cross-validated statistics here so that traditional tree-ring analysts who use such a procedure are given

TABLE 3. Interquartile ranges for the regression coefficients (i.e., β_i) from both model schemes.

Reservoir	No correlation in errors			Correlation in errors		
	Full pool	No pool	Partial pool	Full pool	No pool	Partial pool
MH(1)						
Schoharie	0.45	1.10	0.62	0.83	1.05	0.93
Neversink		0.88	0.60		0.86	0.82
Roundout		1.06	0.62		0.96	0.89
Canonsville		1.13	0.69		0.99	0.94
Pepacton		0.94	0.62		0.90	0.88
MWO						
Schoharie	0.43	0.99	0.65	0.76	1.04	0.87
Neversink		0.96	0.60		0.85	0.74
Roundout		1.11	0.62		0.99	0.86
Canonsville		1.04	0.64		0.97	0.83
Pepacton		0.96	0.61		0.95	0.78
MH(2)						
Schoharie	0.54	1.16	0.66	0.90	0.91	0.93
Neversink		1.01	0.66		0.75	0.88
Roundout		1.24	0.66		0.95	0.94
Canonsville		1.28	0.73		0.95	0.99
Pepacton		1.18	0.69		0.89	0.93
MSB						
Schoharie	0.20	0.43	0.34	0.31	0.43	0.41
Neversink		0.39	0.31		0.35	0.34
Roundout		0.46	0.35		0.44	0.38
Canonsville		0.45	0.38		0.42	0.40
Pepacton		0.42	0.35		0.35	0.39
MPP						
Schoharie	0.18	0.38	0.30	0.30	0.36	0.35
Neversink		0.34	0.30		0.32	0.32
Roundout		0.41	0.32		0.39	0.38
Canonsville		0.47	0.36		0.40	0.37
Pepacton		0.41	0.31		0.34	0.31
MoCO						
Schoharie	0.39	0.95	0.64	0.67	1.04	0.77
Neversink		0.86	0.57		0.69	0.72
Roundout		0.94	0.57		0.83	0.81
Canonsville		0.90	0.63		0.90	0.76
Pepacton		0.78	0.55		0.81	0.73
MoTP						
Schoharie	0.27	0.64	0.46	0.49	0.66	0.63
Neversink		0.60	0.44		0.54	0.54
Roundout		0.75	0.46		0.70	0.59
Canonsville		0.79	0.51		0.65	0.57
Pepacton		0.67	0.45		0.58	0.56

TABLE 3. (Continued)

Reservoir	No correlation in errors			Correlation in errors		
	MiCO					
	Full pool	No pool	Partial pool	Full pool	No pool	Partial pool
Soharie	0.31	0.74	0.55	0.62	0.83	0.66
Neversink		0.66	0.49		0.65	0.56
Roundout		0.81	0.54		0.73	0.63
Canonsville		0.88	0.57		0.75	0.65
Pepacton		0.78	0.52		0.65	0.61

a benchmark consistent with their approach. The entire Bayesian fitting process is repeated for each sample of size 40.

The reduction of error is defined as

$$RE \equiv 1 - \left[\frac{\sum(x_i - \hat{x}_i)^2}{\sum(x_i - \bar{x}_c)^2} \right]. \tag{7}$$

In Eq. (7) x_i and \hat{x}_i are the observed and the predicted posterior mean of the streamflows (transformed back to real space by taking antilogs) in year i of the validation period and \bar{x}_c is the mean of the observational data in the calibration period (Lorenz 1956; Fritts 1976). RE ranges from $-\infty$ to $+1$ and is similar to the R^2 statistic. $RE > 0$ indicates that the reconstructed streamflow contains useful information not contained in the calibration period. Similarly, $RE < 0$ indicates that the reconstructions are poorer than climatology; that is, the reconstructions are not better than the mean flows in the calibration period. The coefficient of efficiency is defined as

$$CE \equiv 1 - \left[\frac{\sum(x_i - \hat{x}_i)^2}{\sum(x_i - \bar{x}_v)^2} \right]. \tag{8}$$

In Eq. (8) x_i and \hat{x}_i are the observed and the predicted streamflows in year i of the validation period, and \bar{x}_v is the mean of the observational data in the validation period. $CE < 0$ indicates that the reconstructions are poorer than validation climatology; that is, the reconstructions are not better than the mean flows in the validation period. CE is similar to RE, but used as a measure to evaluate the model under the validation period, as it is a more rigorous metric. For more details on the CE and RE, see Cook et al. (1999) and Fritts (1976).

The results for RE and CE performance under cross-validation for both no pooling and partial pooling for each station are shown in Fig. 6. We observe that under most calibration periods, both the no-pooling and

partial-pooling methods show RE and CE greater than zero, indicating that the reconstructed streamflows (from both methods) contain useful information not contained in the calibration period. Further, we also observe that, on average, the RE and CE across all the validation periods from the partial-pooling HBR method is comparable to or better than the no-pooling method for the stations. The improved average metrics for the HBR method reflect the reduction in the uncertainties in estimating the model parameter and the resulting flows. A comparison of the average bias and variance of the estimates from both the no-pooling and partial-pooling methods showed that the reduction in average error for the partial-pooling method is primarily due to a lower parameter variance.

In addition to computing the cross-validated RE and CE, the performance of the posterior probability distribution is assessed by examining the model’s ability

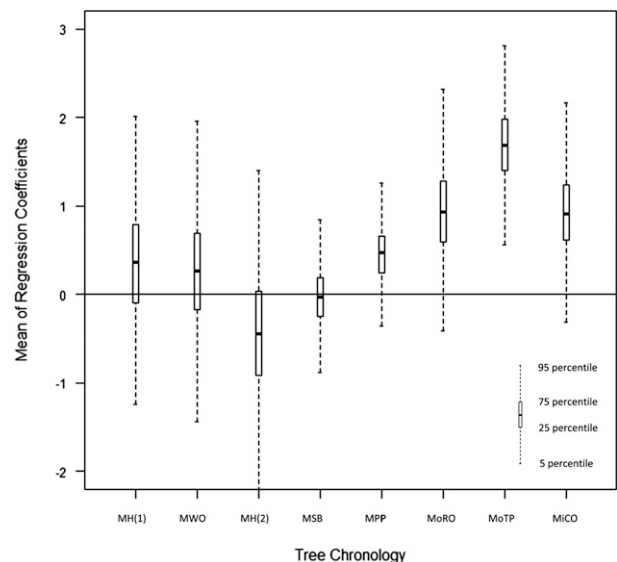


FIG. 4. Boxplots of the μ_β coefficients from the partial pooling hierarchical Bayesian regression model. MH(1) and MH(2) are strongly correlated series.

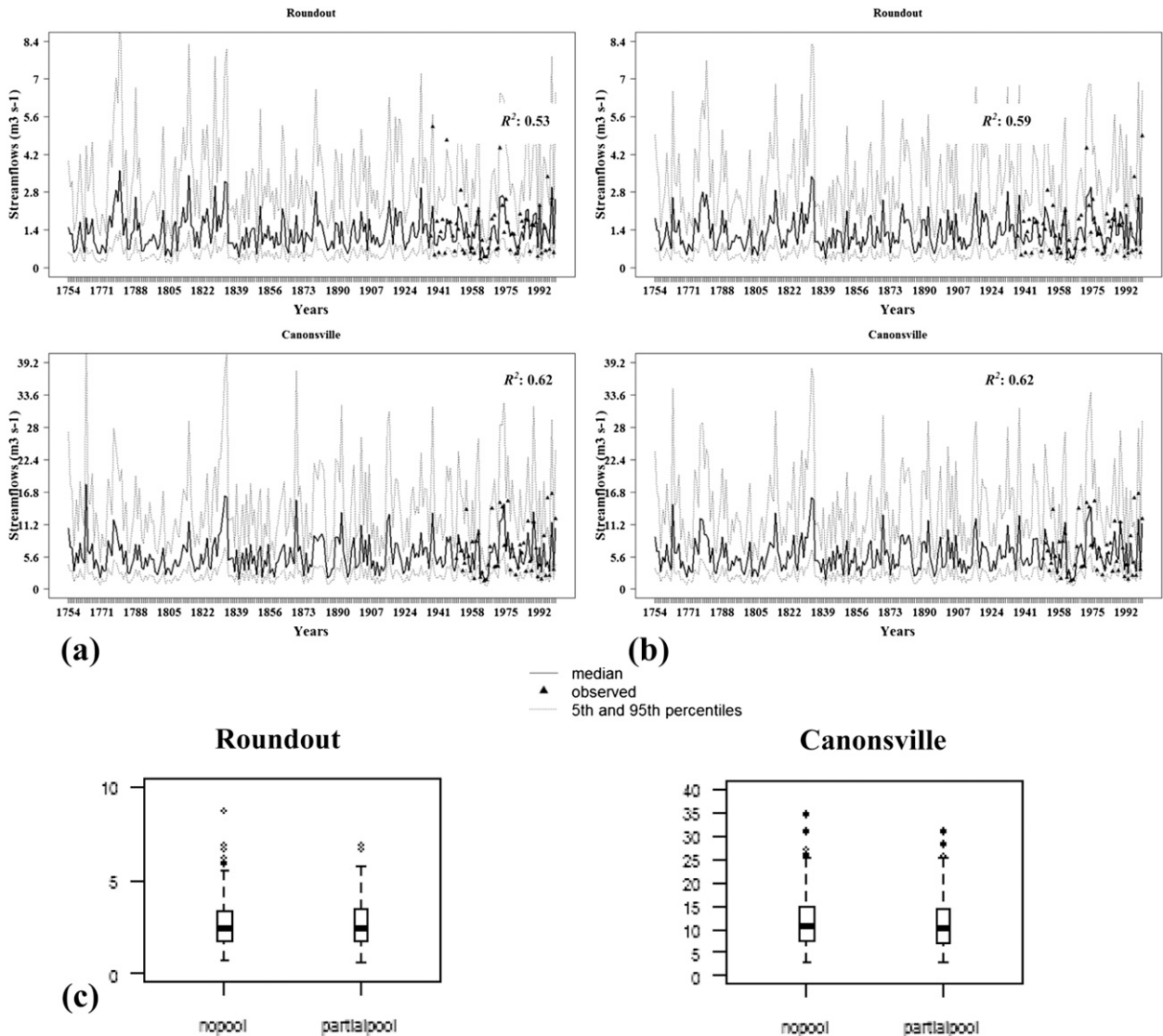


FIG. 5. JJA reconstructed seasonal average streamflow for the Roundout and Canonsville stations from (a) no-pooling traditional regression and (b) partial-pooling hierarchical Bayesian regression, along with the uncertainty bands that represent the 5th and 95th percentile flows. (c) The boxplots of the in the width of the uncertainty bands (95th percentile – 5th percentile) for the two models.

to cover the observed flows within a specified credible interval. Here, we estimated the coverage rates (Li et al. 2010) for the 90% credible intervals for the validation periods for both the models. For each validation period, we count the number of failures or the number of observations (during the validation period) that are outside the 5th and 95th percentile of the posterior distribution resulting from the model developed using the remaining years as calibration for each station. Henceforth by computing the total number of failures from all the randomly selected models, we estimated the coverage rate as the percentage of failures in the total

of 500 (50×10 years) samples. The average coverage rate across the stations is approximately 92% for the no-pooling model and 91.5% for the partial-pooling model, indicating the robustness of the fitted Bayesian models.

From the above results, we see that the performance of the partial-pooling HBR method for streamflow reconstructions of the Delaware River is comparable to or better than the no-pooling traditional regression method. In the next section, we use simulations of the reconstructed streamflow for regional drought characterization.

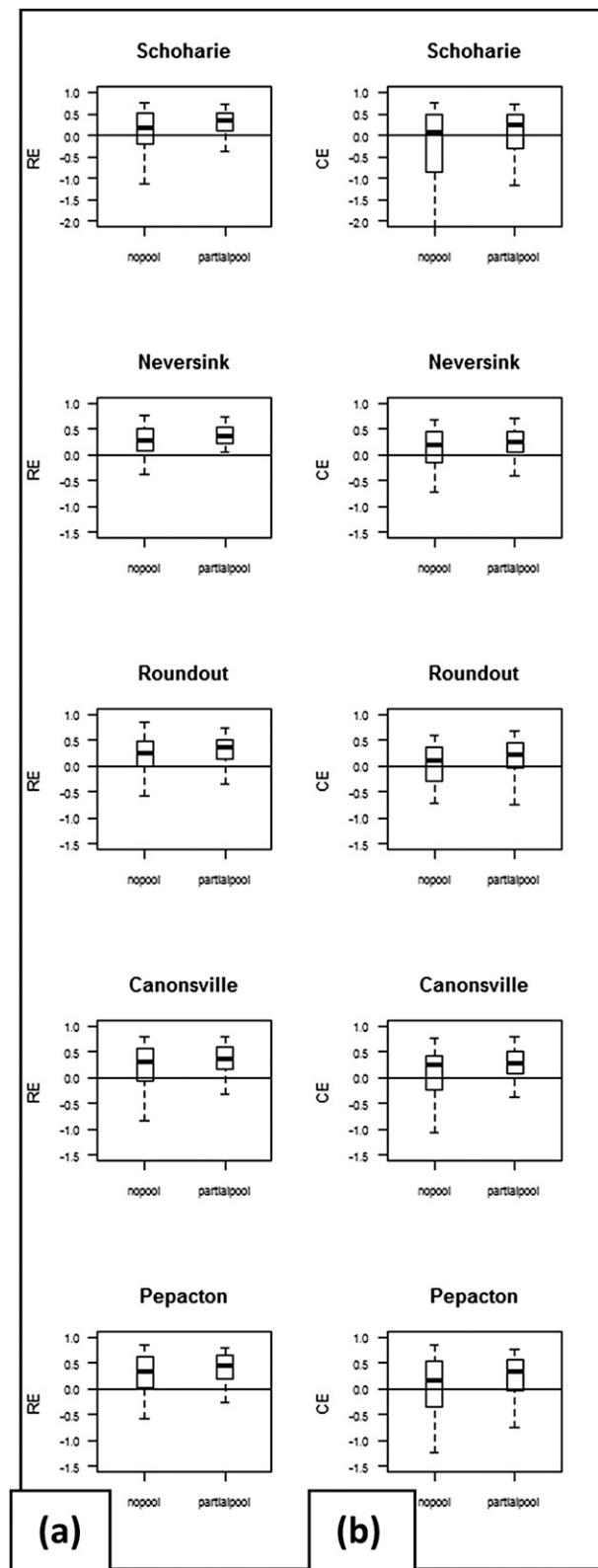


FIG. 6. Boxplots of (a) reduction of error (RE) and (b) coefficient of efficiency (CE) for the 50 randomly selected cross-validation cases.

5. Drought characterization based on reconstructed streamflows

The drought of record in the region is the one from the early to mid 1960s (Namias 1966, 1967). A similar drought could cause severe stress on regional water resources, given the increased population and services today. Subsequent moderate droughts have led to water restrictions in the region due to reduced reservoir storages (NYCDEP 2011). In this section, we attempt to 1) characterize the duration and severity of the 1960s drought along with its return periods and 2) investigate for any changes/trends in the extreme drought events using the reconstructed streamflow from the general partial pooling model.

a. Quantifying the duration and severity of droughts

We define a drought as an event during which the streamflow is continuously below a certain level. A schematic representation (based on historical data for the Canonsville station) of the drought statistics is shown in Fig. 7. For a selected threshold (90% of mean observed summer flows here), a drought event is defined as the sequence of years that are under the threshold with event duration defined as the numbers of years the flow is continuously below the threshold. The magnitude or the severity is the cumulative deficit over the drought duration estimated as the area under the curve below the threshold. The number of historical drought events and their severity for the Canonsville station is shown in Fig. 7b. The 1960s drought is seen to be the most severe in terms of duration (5 yr) and severity (a cumulative deficit of $132 \times 10^6 \text{ m}^3$). Hence, purely based on the historical record, the return period of this drought event is approximately 1 in 54.

b. Quantifying the duration and severity of droughts based on reconstructed flows

We used the posterior probability distribution of the 247-yr-long reconstructed streamflow records conditioned on tree-ring data to quantify the duration and severity of droughts greater than or equal to the 1960s drought in the historical record for each station. For example, for the Canonsville station the 1960s drought with a severity of $132 \times 10^6 \text{ m}^3$ —we identify all events that are more severe in terms of duration and severity in each simulation of length 247 years from the posterior probability distribution of the model. The return period of drought severity can then be estimated from the number of such events in a 247-yr simulation. One thousand realizations, each of length 247 yr, were generated from the posterior distribution and the number of events that exceed the duration and severity of the

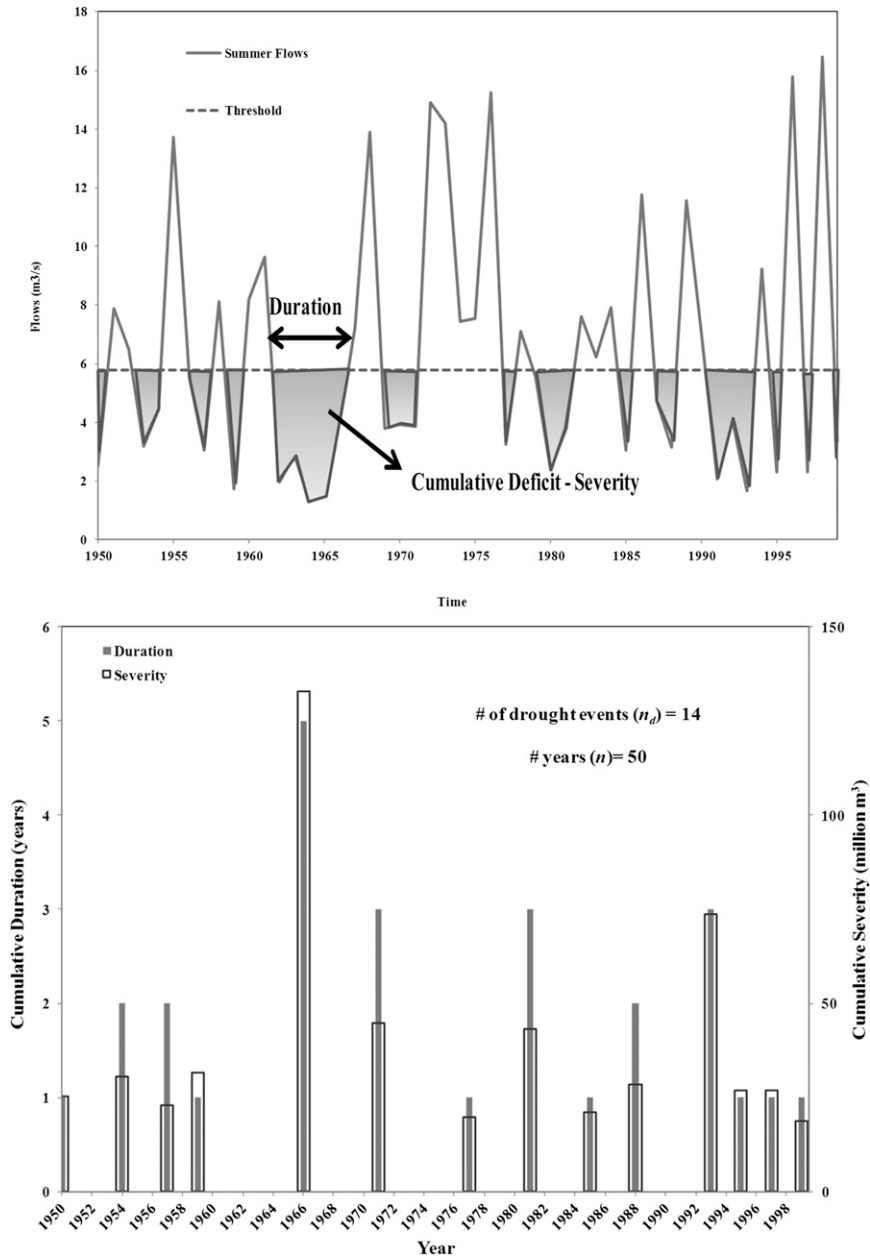


FIG. 7. (top) Schematic representation of the duration and severity of drought and (bottom) drought events based on a selected threshold of 90% of the average streamflow in the historical streamflow data record of 50 years for the Canonsville station.

1960s drought at each streamflow station was counted for each realization. The results from the partial-pooling model for the exceedance attributes of the severity and duration of the 1960s drought are shown in Fig. 8. From Fig. 8a we see that the median return period of the 1960s drought is around 80 yr for the region with an interquartile range between 50 and 125 yr. However, we also see that the return periods for the Neversink and the Roundout stations, which have small drainage

areas, are lower than the regional median. To understand the correlated nature of these droughts, in Fig. 8b we show the histograms of the number of stations that are simultaneously under drought from all the simulations. We observe that most commonly all five stations are under drought. However, Neversink experiences more frequent droughts of the 1960s severity and corresponds to the case where there is a lone station under drought. Further, the number of simulations that

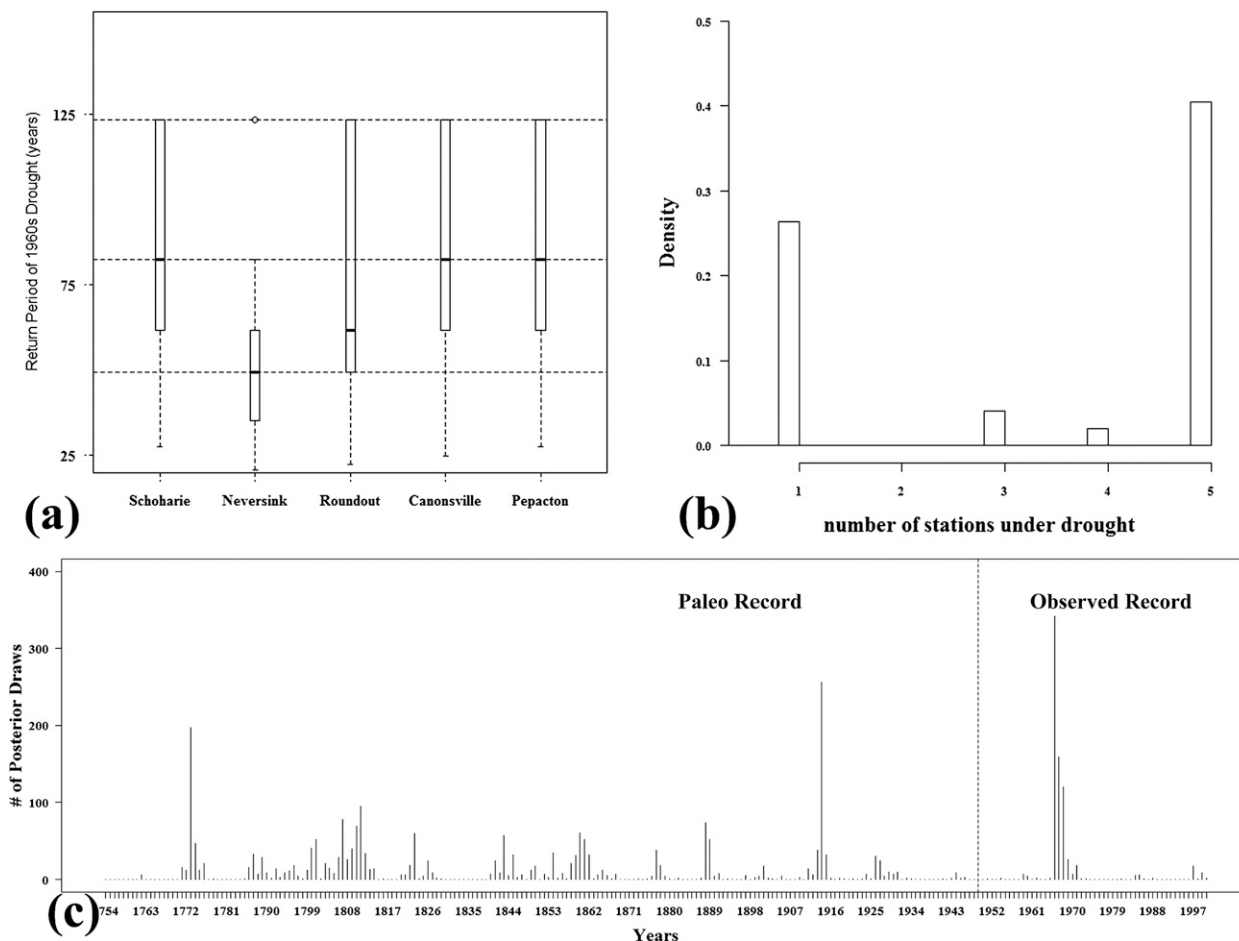


FIG. 8. (a) Boxplots of the return period of the 1960s drought identified for the five stations from the 1000 simulations of the partial-pooling model; (b) histogram of the number of stations that are simultaneously under drought; and (c) time series marking Canonsville droughts with duration and severity greater than 1960s drought from posterior simulations, marked at the end of each drought.

indicate an exceedance of the 1960s drought for Canonsville over the paleoreconstructed record is illustrated in Fig. 8c. Note that each bar (showing the total number of posterior draws per year that indicate the drought) corresponds to the year in which the simulated drought ends. A cluster of these years, as in 1965–70, indicates a high probability of such a drought. The 1960s period is striking in this regard. However, it is within the fitting sample and, hence, is expected to be more prominent. The years 1912–14, 1850s–60s, 1790–1810, and the 1770s appear to be other periods of interest. Similar observations were also made by Pederson et al. in the Northeast region. The simulations provide the ability to also analyze reservoir fill and drain probabilities as a function of drought intermittence and recurrence. Bayesian regime and changepoint analysis models integrated with the reconstruction model could also be employed to inform reservoir and drought management policies.

c. Trend analysis to detect changes in drought events

The analysis presented thus far attempted to analyze the probability of occurrence of the 1960s drought using the long run simulations of reconstructed streamflow for each station. In this section, we assess monotonic trends in the joint drought events from the 1000 posterior draws that exceed the 1960s drought threshold level using the Mann–Kendall nonparametric trend tests (Mann 1945; Helsel and Hirsch 1992). The Mann–Kendall test is a rank-based test that is typically used for detecting trends in extremes with no assumption of the underlying distribution of the data (Helsel and Hirsch 1992). For each posterior draw of streamflow, we identify the drought events in the 247-yr simulation with a duration and severity greater than the target duration and severity, and apply the Mann–Kendall trend test for monotonic increase or decrease in the incidence of these historical drought events. The incidence

is recorded as a binary variable (1 for exceedance, 0 for nonexceedance). Results show that the slope (τ) from the test as percentiles across the 1000 simulations ranges from -0.2 to 0.2 , suggesting that there is little evidence for a monotonic trend in the incidence of droughts for the target threshold.

6. Discussion and summary

A restricted goal of our hierarchical Bayesian model for streamflow reconstruction was to consider 1) the use of processed tree-ring chronologies as the primary predictors; 2) all chronologies of the common length; 3) no expectation-maximization (Dempster et al. 1977; Schneider 2001) or similar algorithm for imputation of predictors or response variables, that is, full record lengths; 4) no exogenous predictors for streamflow such as drainage area; and 5) no consideration of the spatial structure on the river drainage network.

We were interested in seeing whether partial pooling through hierarchical Bayesian regression (HBR) offered 1) reductions in uncertainty over the relatively short lengths of time series available in the Delaware basin for a small number of streamflow stations for relatively small drainage area basins, 2) insights on how best to pool (or share information about the regression coefficients) across streamflow stations or trees, and 3) an ability to model multivariate correlations across sites and trees. These questions were explored with the Delaware application, with the idea that we would eventually build a HBR model that considers the entire reconstruction process more generally, relaxing the restrictions imposed here. For example, as in Lima and Lall (2010), one can easily extend the model to consider exogenous predictors for the slopes and intercepts and to link different levels together. Intuitively, such an approach is advantageous compared to some traditional applications where, for instance, only the conditional mean of one modeling step (e.g., tree-ring chronology processing) is used as “data” for the next modeling step (e.g., flow reconstruction). Extensions of this model to explore whether there are cyclical patterns (e.g., related to the North Atlantic Oscillation or other low-frequency climate modes) or hidden states may allow further investigation of the drought onset and withdrawal as part of system operation. So far, very few such models have been pursued for this application. Integration of models such as the wavelet autoregressive models presented in Kwon et al. (2007) would be of interest in this regard.

A relatively simple HBR model structure was consequently used and justified in the application. We found partial pooling across stations for a particular tree

species, and modeling the correlation of residuals across streamflow stations to be best for this dataset. This is also consistent with the biological and climate intuition of tree-ring specialists and hydrologists. Comparisons with traditional models in our preliminary work (not reported here) showed that the HBR is competitive or superior in terms of the validation statistics typically used by paleoclimate modelers. We are able to see the effectiveness of the HBR under partial pooling to deliver reduced uncertainty in reconstruction and improved cross-validation performance statistics, as well as provide ways to assess the joint probability distribution of drought severity and duration and its uncertainty. Such information is important both in terms of adding value from long records and in terms of developing precision in the return period estimates. Finally, nonstationarity in drought incidence was explored for the longer record using simulations from the posterior density of the reconstructed flows and the conclusion was that the evidence for monotonic trend was not statistically significant. The implications for water resource managers are that, at least for now, 1) drought planning and management could use the information from the historical and paleoreconstructions for the dry period in each year, 2) the return periods of droughts of different severity and duration can be estimated, including its uncertainty, and 3) synthetic streamflow sequences for summer period inflows into the reservoir system can be developed to explore the risk implications for different drought sequences.

Acknowledgments. This research was supported through the Consortium on Climate Risk in the Urban Northeast (CCRUN), part of the NOAA RISA program and through the NSF Grant 0934516 titled “Reconstructing Climate from Tree Ring Data.” The authors thank Dr. Martin Tingley for his critical reviews that were very helpful in improving the analysis and presentation in the manuscript. The authors would also like to thank the other two anonymous reviewers whose valuable comments led to significant improvements in the manuscript.

REFERENCES

- Berliner, L. M., C. K. Wikle, and N. Cressie, 2000: Long-lead prediction of Pacific SSTs via Bayesian dynamic modeling. *J. Climate*, **13**, 3953–3968.
- Buckley, B. M., and Coauthors, 2010: Climate as a contributing factor in the demise of Angkor, Cambodia. *Proc. Natl. Acad. Sci. USA*, **107**, 6748–6752.
- Cook, E. R., 1985: A time series analysis approach to tree-ring standardization. Ph.D. dissertation, University of Arizona, 342 pp.

- , and G. Jacoby, 1977: Tree-ring drought relationships in the Hudson Valley, New York. *Science*, **198**, 399–401.
- , and K. Peters, 1981: The smoothing spline: A new approach to standardizing forest interior ring-width series for dendroclimatic studies. *Tree-Ring Bull.*, **41**, 45–53.
- , and G. Jacoby, 1983: Potomac River streamflow since 1730 as reconstructed by tree rings. *J. Climate Appl. Meteor.*, **22**, 1659–1672.
- , and L. A. Kairiukstis, 1990: *Methods of Dendrochronology: Applications in the Environmental Sciences*. Kluwer Academic, 304 pp.
- , and N. Pederson, 2010: Uncertainty, emergence, and statistics in dendrochronology. *Dendroclimatology: Progress and Prospects*, M. K. Hughes, T. W. Swetnam, and H. F. Diaz, Eds., Vol. 11, *Developments in Paleoenvironmental Research*, Springer Verlag, 77–112.
- , D. Meko, D. Stahle, and M. Cleaveland, 1999: Drought reconstructions for the continental United States. *J. Climate*, **12**, 1145–1162.
- , R. Seager, R. R. Heim Jr., R. S. Vose, C. Herweijer, and C. Woodhouse, 2010: Megadroughts in North America: Placing IPCC projections of hydroclimatic change in a long-term palaeoclimate context. *J. Quat. Sci.*, **25**, 48–61.
- Dempster, A. P., N. M. Laird, and D. B. Rubin, 1977: Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc.*, **39**, 1–38.
- DRBC, cited 2007: Flexible flow management program. [Available online at <http://water.usgs.gov/osw/odrm/>.]
- Friedman, J. H., 1984: A variable span smoother. Stanford University SLAC PUB-3477 STAN-LCS 005, 30 pp. [Available online at <http://www.slac.stanford.edu/cgi-wrap/getdoc/slac-pub-3477.pdf>.]
- Fritts, H. C., 1976: *Tree Rings and Climate*. Academic Press, 567 pp.
- Gangopadhyay, S., B. L. Harding, B. Rajagopalan, J. J. Lukas, and T. J. Fulp, 2009: A nonparametric approach for paleohydrologic reconstruction of annual streamflow ensembles. *Water Resour. Res.*, **45**, W06417, doi:10.1029/2008WR007201.
- Gelman, A., 2005: Prior distribution for variance parameters in hierarchical models. *Bayesian Anal.*, **1**, 1–19.
- , and D. B. Rubin, 1992: Inference from iterative simulation using multiple sequences. *Stat. Sci.*, **7**, 457–511.
- , and J. Hill, 2007: *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, 648 pp.
- , J. B. Carlin, H. S. Stern, and D. B. Rubin, 2004: *Bayesian Data Analysis*. Chapman & Hall, 668 pp.
- Gilks, W. R., and G. O. Roberts, 1995: Strategies for improving MCMC. *Markov Chain Monte Carlo in Practice: Interdisciplinary Statistics*, W. R. Gilks, S. Richardson, and D. Spiegelhalter, Eds., Chapman & Hall, 89–114.
- Helsel, D. R., and R. M. Hirsch, 1992: *Statistical Methods in Water Resources*. Studies in Environmental Science Series, Vol. 49, Elsevier Science, 522 pp.
- Kagawa, A., A. Sugimoto, and T. C. Maximov, 2006: ^{13}C pulse-labelling of photoassimilates reveals carbon allocation within and between tree rings. *Plant Cell Environ.*, **29**, 1571–1584.
- Kauffman, G. J., and K. J. Vonck, 2011: Frequency and intensity of extreme drought in the Delaware basin, 1600–2002. *Water Resour. Res.*, **47**, W05521, doi:10.1029/2009WR008821.
- Kolesar, P., and J. Serio, 2011: Breaking the deadlock: Improving water-release policies on the Delaware River through operations research. *INFORMS Interfaces*, **41**, 18–34.
- Kwon, H.-H., U. Lall, and A. F. Khalil, 2007: Stochastic simulation model for nonstationary time series using an autoregressive wavelet decomposition: Applications to rainfall and temperature. *Water Resour. Res.*, **43**, W05407, doi:10.1029/2006WR005258.
- , C. Brown, and U. Lall, 2008: Climate informed flood frequency analysis and prediction in Montana using hierarchical Bayesian modeling. *Geophys. Res. Lett.*, **35**, L05404, doi:10.1029/2007GL032220.
- , U. Lall, and V. Engel, 2011: Predicting foraging wading bird populations in Everglades National Park from seasonal hydrologic statistics under different management scenarios. *Water Resour. Res.*, **47**, W09510, doi:10.1029/2010WR009552.
- Li, B., D. W. Nychka, and C. M. Ammann, 2010: The value of multiproxy reconstruction of past climate. *J. Amer. Stat. Assoc.*, **105**, 883–911.
- Lima, C. H. R., and U. Lall, 2009: Hierarchical Bayesian modeling of multisite daily rainfall occurrence: Rainy season onset, peak and end. *Water Resour. Res.*, **45**, W07422, doi:10.1029/2008WR007485.
- , and —, 2010: Spatial scaling in a changing climate: A hierarchical Bayesian model for non-stationary multi-site annual maximum and monthly streamflow. *J. Hydrol.*, **383**, 307–318, doi:10.1016/j.jhydrol.2009.12.045.
- Lorenz, E. N., 1956: Empirical orthogonal functions and statistical weather prediction. MIT Department of Meteorology Statistical Forecasting Scientific Rep. 1, 57 pp.
- Lunn, D. J., A. Thomas, N. Best, and D. Spiegelhalter, 2000: WinBUGS—A Bayesian modelling framework: Concepts, structure, and extensibility. *Stat. Comput.*, **10**, 325–337.
- Mann, H. B., 1945: Nonparametric tests against trend. *Econometrica*, **13**, 245–259.
- Maxwell, R. S., A. E. Hessel, E. R. Cook, and N. Pederson, 2011: A multispecies tree ring reconstruction of Potomac River streamflow (950–2001). *Water Resour. Res.*, **47**, W05512, doi:10.1029/2010WR010019.
- Meko, D. M., and D. A. Graybill, 1995: Tree-ring reconstruction of Upper Gila River discharge. *Water Resour. Bull.*, **31**, 605–616.
- , C. A. Woodhouse, C. H. Baisan, T. Knight, J. J. Lukas, M. K. Hughes, and M. W. Salzer, 2007: Medieval drought in the Upper Colorado River basin. *Geophys. Res. Lett.*, **34**, L10705, doi:10.1029/2007GL029988.
- Namias, J., 1966: Nature and possible causes of the northeastern United States drought during 1962–1965. *Mon. Wea. Rev.*, **94**, 543–557.
- , 1967: Further studies of drought over northeastern United States. *Mon. Wea. Rev.*, **95**, 497–508.
- NYCDEP, cited 2011: History of drought and water consumption. [Available online at http://www.nyc.gov/html/dep/html/drinking_water/droughthist.shtml.]
- Pederson, N., 2005: Climatic sensitivity and growth of southern temperate trees in the Eastern US: Implications for the carbon cycle. Ph.D. dissertation, Columbia University, 186 pp.
- , E. R. Cook, G. C. Jacoby, D. M. Peteet, and K. L. Griffin, 2004: The influence of winter temperatures on the annual radial growth of six northern-range-margin tree species. *Dendrochronologia*, **22**, 7–29.
- , A. R. Bell, E. R. Cook, U. Lall, N. Devineni, R. Seager, K. Eggleston, and K. J. Vranes, 2013: Is an epic pluvial masking the water insecurity of the Greater New York City region? *J. Climate*, **26**, 1339–1354.
- Raftery, A., 1995: Bayesian model selection in social research. *Sociol. Methodol.*, **25**, 111–163.

- Royston, P., 1995: Remark AS R94: A remark on algorithm AS 181: The W test for normality. *Appl. Stat.*, **44**, 547–551.
- Schneider, T., 2001: Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values. *J. Climate*, **14**, 853–871.
- Spiegelhalter, D., A. Thomas, N. Best, and W. Gilks, 1996: BUGS 0.5: Bayesian inference using Gibbs sampling manual (version ii). Medical Research Council Biostatistics Unit Manual, 59 pp.
- Stockton, C. W., and G. C. Jacoby, 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.
- Stokes, M. A., and T. L. Smiley, 1968: *An Introduction to Tree-Ring Dating*. University of Arizona Press, 73 pp.
- Tingley, M. P., and P. Huybers, 2010a: A Bayesian algorithm for reconstructing climate anomalies in space and time. Part I: Development and applications to paleoclimate reconstruction problems. *J. Climate*, **23**, 2759–2781.
- , and —, 2010b: A Bayesian algorithm for reconstructing climate anomalies in space and time. Part II: Comparison with the regularized expectation–maximization algorithm. *J. Climate*, **23**, 2782–2800.
- Trumbore, S., J. Gaudinski, P. Hanson, and J. Southon, 2002: Quantifying ecosystem–atmosphere carbon exchange with a ¹⁴C label. *Eos, Trans. Amer. Geophys. Union*, **83**, 265–268.
- Wikle, C. K., 2003: Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology*, **84**, 1382–1394.
- Woodhouse, C. A., and J. Lukas, 2006a: Drought, tree rings and water resource management in Colorado. *Can. Water Resour. J.*, **31**, 1–14, doi:10.4296/cwrj3104297.
- , and —, 2006b: Multi-century tree-ring reconstructions of Colorado streamflow for water resource planning. *Climatic Change*, **78**, 293–315, doi:10.1007/s10584-006-9055-0.
- , S. T. Gray, and D. M. Meko, 2006: Updated streamflow reconstructions for the Upper Colorado River basin. *Water Resour. Res.*, **42**, W05415, doi:10.1029/2005WR004455.