

1 **Tree-Ring Reconstruction of Streamflow by Scatterplot Smoothing**

2

3 David M. Meko¹ and Katherine K. Hirschboeck¹

4 ¹Laboratory of Tree-Ring Research, University of Arizona

5

6 A new modeling approach for reconstruction of streamflow from tree-ring widths is
7 described and illustrated for the Salt-Verde River basin, Arizona. The approach,
8 proposed to deal with relationships that are nonlinear and possibly weaker at higher flows
9 than at low flows, includes a combination of parametric and nonparametric statistical
10 techniques. Main steps are 1) filtering and scaling of tree-ring chronologies to adjust for
11 lags and signal-strength differences in climate response, 2) weighting over sites to
12 emphasize common signal for flow, and 3) interpolation of reconstructed flows from a
13 locally weighted polynomial (loess curve) fit to the scatterplot of observed flows on the
14 weighted tree-ring variable. Error bars for the reconstruction are derived by weighted
15 bootstrapping of cross-validation residuals of loess-estimated flows. A comparison of
16 reconstructions by the new method and by a more traditional approach shows important
17 differences, especially in the inferred severity of low flows and high flows.

18

19 **1. Introduction**

20

21 Reconstruction of streamflow from tree rings is based on a statistically calibrated model
22 relating a time series of observed flows to indices of tree-ring width at one or more sites.
23 The tree-ring indices, or site chronologies (Cook et al. 1990) are viewed as proxies for
24 moisture variations in runoff-producing parts of the watershed, and in this sense are
25 surrogate precipitation series. The statistical model is usually multiple linear regression
26 (MLR) and the predictors, or the chronologies, are sometimes reduced by averaging or
27 principal components analysis, and sometimes lagged to adjust for non-climatic
28 persistence in the tree ring series and to allow for the possibility imperfectly
29 synchronized relationships of chronologies and flow (e.g., Stockton and Jacoby 1976;
30 Smith and Stockton 1981; Meko and Graybill 1995; Cleaveland 2000; Meko et al. 2001;
31 Woodhouse et al. 2006). Depending on the basin, flow transformation has been applied
32 in such studies to mitigate problems with violations of assumptions on regression
33 residuals. For example, log-transformation was used in reconstructions of annual flow
34 for the Salt River (Smith and Stockton 1981), and Sacramento River (Meko et al. 2001)
35 to deal with non-normality of residuals and non-constancy of regression error-variance,
36 also called heteroskedasticity. As the tree-ring index, defined as the ratio of measured
37 ring width to a fitted “age-trend”, is itself an arbitrary transformation of ring width, there
38 is no reason to expect the index to be ideally suited as a flow predictor. Flow
39 reconstruction has indeed been reported to benefit from quadratic transformation of the
40 tree-ring chronologies used as predictors (Cleaveland 2001). If a tree-ring index is
41 viewed as a precipitation proxy it is reasonable to expect some sort of nonlinear
42 relationship of the index with streamflow, as the runoff/rainfall ratio for semi-arid basins
43 generally is not linear (e.g., Sellers 1960). Moreover, under very wet conditions the
44 incremental increase in tree growth with increasing moisture would logically decrease, as
45 soil moisture would no longer be the main limiting factor to growth.

46 Although some combination of data transformation of predictand and predictors
47 might effectively deal with the problems mentioned above, it is important to note that
48 transformation of the predictand is not without its own drawbacks. One is that the
49 regression is optimized on transformed units (e.g., minimizing the sum of squares of log-

50 transformed flow). Another is that the reconstruction must generally be back-
51 transformed to original units if it is to be practically useful to most users (e.g., water
52 managers), and in doing so some convenient properties of the regression estimates are
53 lost. For example, while regression guarantees that the calibration-period means of
54 observed and reconstructed predictand are equal, the equality no longer holds for the
55 predictand back-transformed to original units. A third is that the error bars for
56 reconstructed flow, while constant for the transformed flow, vary from year-to-year for
57 back-transformed flow. A fourth drawback is that the ideal form of transform may not
58 be obvious -- whether for straightening the flow-tree ring relationship or for reducing
59 heteroskedasticity. Log-transformation, for example, is just one of a family of power
60 transformations (Hoaglin et al. 1983), and its predominant use in dendrohydrology is
61 largely due to convenience and familiarity.

62 The purpose of this paper is to describe and illustrate a new approach to
63 streamflow reconstruction that circumvents flow transformation but still addresses the
64 issues of nonlinearity and nonconstant error variance that plague reconstruction efforts,
65 especially in smaller semi-arid basins. The approach adapts both nonparametric and
66 robust parametric modeling to deal with some of the challenges posed by tree-ring and
67 flow data in such basins. Tree-ring chronologies are first individually filtered and
68 scaled to emphasize the flow signal. The resulting series are then weighted to reduce
69 site-specific noise. Finally, reconstructed flow is interpolated from a smoothed
70 scatterplot of observed flow on the weighted derived tree-ring variable. A confidence
71 interval for reconstructed flows is estimated by weighted bootstrapping that reflects any
72 increased scatter in the relationship of tree-ring index to flow as conditions become
73 wetter.

74 The trial basin selected to illustrate the method is the combined Salt-Tonto-Verde
75 basin, in south-central Arizona. We generate a reconstruction of water-year flow, 1451-
76 1982, by the new method. That reconstruction is compared with a reconstruction by
77 linear regression with log-transformed flow to assess the assess sensitivity of inferred
78 hydrologic drought history to reconstruction methodology.
79

80 **2. Methods**

81

82 The method described here is based on an idealized relationship between flow
83 and tree-ring index (Figure 1). The sketch might apply to a gaged or natural-flow series
84 of annual flows, and a moisture-limited tree-ring index at a site in or near the watershed.
85 The first relevant aspect of this idealized relationship is curvature. It is reasonable to
86 expect that the incremental change in tree-ring index with increasing flow might decrease
87 as conditions become so moist that soil moisture is no longer limiting to growth. The
88 second aspect is smoothness, reflecting the likely limitation of sampled data to capture
89 only the gross features of any relationship of tree-growth to moisture variation. The third
90 aspect is the amplified scatter of the relationship during wet years. Amplified scatter
91 could arise from several sources. Gaged streamflow, as derived from stage-discharge
92 relationships, is less accurate for high flows than for low flows; this characteristic is
93 reflected in the convention for rating stream-gage accuracy as a percentage of measured
94 flow (e.g., Mosely and McKerchar 1993). Moreover, high flow might also result from
95 localized heavy rains that might or might not be commensurate with precipitation
96 received at any particular tree-ring site. Finally, high flows may leave dramatically
97 different imprints in the soil moisture zone of the trees depending on the intensity and
98 duration of the contributing storms.

99 The conceptual idea behind the method proposed here is that a suitably smoothed
100 scatterplot such as that in Figure 1 can be directly applied as a reconstruction model.
101 Inference from the smoothed line can yield flow estimates in original flow units, without
102 the need for data transformation. The procedure must of course be objective, such that
103 the smooth curve can be regenerated by different researchers from the same basic data.
104 The procedure must also allow for the incorporation of flow signals from multiple tree-
105 ring sites, when the signal may differ in strength, linearity, and lag properties across sites.
106 Finally, estimates of uncertainty of reconstructed flows must reflect the increasing scatter
107 of the relationship at high flows.

108 This paper a multi-stage reconstruction modeling approach proposed for
109 precipitation reconstruction (Meko 1997) and later adapted to flow reconstruction (Meko
110 et al. 2001). The main extensions are incorporation of 1) quadratic regression and

111 nonparametric smoothing to allow flexibility to handle curvilinear relationships, and 2)
112 weighted bootstrapping to generate confidence intervals consistent with an observed
113 amplification of reconstruction error in wetter years. The main steps are summarized in
114 Figure 2 and are described in more detail below.

115

116 2.1 Filtering and scaling. The analysis begins with a time series of flow and a set of
117 residual tree-ring chronologies (Cook et al. 1990). Each tree-ring chronology is first
118 individually filtered and scaled to emphasize its signal for flow. Essentially, this step
119 consists of multiple separate single-site reconstructions of flow, as each chronology is
120 converted into an estimated flow time series. The statistical model for filtering and
121 scaling is quadratic regression of flow on current and lagged tree-ring index. A
122 contemporaneous (no lags) model is specified, followed by the possible addition of lagged
123 terms. The contemporaneous model takes one of three possible alternative forms:

$$\begin{aligned} & 1) y_t = b_o + b_1 w_t + b_2 w_t^2 + e_t \\ 124 \quad & 2) y_t = b_o + b_1 w_t^2 + e_t \\ & 3) y_t = b_o + b_1 w_t + e_t \end{aligned} \tag{0.1}$$

125

126 where y_t is flow in year t , w_t is the tree-ring index in year t , e_t is a noise term, and
127 $\{b_o, b_1, b_2\}$ are parameters estimated by robust regression using the Huber weighting
128 function. Robust regression is preferred here because of the likelihood of outliers in
129 relating tree-growth at a particular point (single chronology) to a flow series that
130 responds to spatially distributed runoff. While ordinary least squares minimizes the sum
131 of squares of residuals, robust regression minimizes a weighted sum of squares of
132 residuals, such that outliers have diminished influence on the fit (Myers 1990). Huber's
133 weighting function assigns equal and maximum weight to observations whose residuals
134 are within some specified threshold distance from zero, and truncates the weights for
135 other residuals according to an influence function (Huber 1973). More details on the
136 particular implementation of robust regression used are included in Appendix B.

137 The selection of model-form from the candidates in (0.1) is guided by the signs
138 of parameter estimates from trial-and-error fitting of all three forms. Priority is given to

139 model (1) if its parameter estimates \hat{b}_1 and \hat{b}_2 are both positive. If either estimated
 140 parameter is negative, model (1) is rejected, and whichever of (2) or (3) has the higher
 141 regression R^2 is accepted. These rules ensure that the curve is monotonically increasing
 142 and concave upward.

143 The filtering-and-scaling model is then expanded by adding lagged terms that
 144 may contribute significantly to the prediction accuracy. A stepwise procedure, again
 145 using robust regression with the Huber weighting function, is used to incorporate the
 146 lagged terms. If y_t is more highly correlated with the tree-ring index w_t than with the
 147 squared index w_t^2 , the pool of potential lagged predictors for stepwise is
 148 $\{w_{t-2}, w_{t-1}, w_{t+1}, w_{t+2}\}$; otherwise the pool is $\{w_{t-2}^2, w_{t-1}^2, w_{t+1}^2, w_{t+2}^2\}$. Stepwise entry of
 149 lagged terms is governed by the size of the partial correlation (Mardia et al. 1979)
 150 between residuals of the model at the current step and potential predictors not yet in the
 151 model. After a lagged predictor is selected, the model is re-fit and the process is
 152 continued until all lags are entered. To avoid over-fitting, the final model selection is
 153 guided by cross-validation (leave-9-out), such that entry is truncated with the step
 154 immediately preceding the first rise in median cross-validation error. The particular
 155 choice of 9 observations for deletion in cross-validation is arbitrary, but is sufficiently
 156 large so that no cross-validation prediction relies on a tree-ring value also used to
 157 calibrate the model when the regression includes predictors lagged up to ± 4 years from
 158 the year of flow.

159 The final filtering-and-scaling model arrived at by the procedure described above
 160 can hypothetically range in complexity from a simple linear regression of flow on the
 161 tree-ring chronology

$$y_t = b_0 + b_1 w_t + e_t \quad (0.2)$$

162
 163
 164 to a quadratic model with four lagged terms

$$y_t = b_0 + b_1 w_t + b_2 w_t^2 + b_3 w_{t-1}^2 + b_4 w_{t-2}^2 + b_5 w_{t+1}^2 + b_6 w_{t+2}^2 + e_t \quad (0.3)$$

166 As each chronology is modeled separately, a different form of model may be arrived at
 167 for each tree-ring chronology in the predictor network.

168 The filtering-and-scaling model for a given chronology is calibrated on the full
 169 available overlap of tree-ring index and flow data, and is then applied to the longer tree-
 170 ring record to generate the long-term filtered-and-scaled tree-ring index, which extends to
 171 the start of that particular chronology. This long series can alternatively be considered a
 172 single-site reconstruction of flow, as models like (0.2) and(0.3) are reconstruction
 173 models. For brevity, the individual predicted time series \hat{y}_t generated by substitution of
 174 tree-ring indices into a filtering-and-scaling regression model is referred to from here on
 175 as a SSR (single site reconstruction).

176
 177 2.2 Weighting. The SSRs for the various tree-ring sites are next weighted into a single
 178 time series through principal components analysis (PCA). The PCA is run on the
 179 covariance matrix instead of the correlation matrix of the SSRs because the differing
 180 variances of the SSRs contain useful information: the variances are proportional to the
 181 percent variance of flow that could be explained by the single-site regression models.
 182 The weighted tree-ring variable can be expressed as

$$183 \quad x_t = \sum_{i=1}^{i=n_c} a_i (\hat{y}_{t,i} - \bar{\hat{y}}_i) \quad (0.4)$$

184 where $\hat{y}_{t,i}$ is the i^{th} SSR in year t , $\bar{\hat{y}}_i$ is the sample mean of the i^{th} SSR, a_i is the loading of
 185 principal component #1 (PC1) on the i^{th} SSR, and n_c is the number of chronologies in the
 186 network. The sample means in (0.4) are computed on the period in common period to the
 187 various SSRs. Because the PCA is run on this common period, the weighted series x_t can
 188 extend only over this period. Note that x_t is also the time series of scores of PC1 (Mardia
 189 et al. 1979), the single linear combination explaining the greatest percentage of variance
 190 of the SSRs. This time series, relying on tree-growth variations at multiple sites, is likely
 191 a more robust indicator of flow than any of the individual SSRs.

192
 193 2.3 Loess-curve estimation. A scatter plot of observed flows y_t on site-weighted SSRs x_t
 194 is smoothed with locally weighted polynomial regression, or “loess” (Cleveland and
 195 Devlin 1988) such that the smoothed line is a nonparametric description of the
 196 relationship between y_t and x_t . The loess model assumes generation of y by

197
$$y_i = g(x_i) + \varepsilon_i \tag{0.5}$$

198 Where ε_i are independent, normal variables with mean zero and variance σ^2 , and $g(x_i)$ is
 199 some smooth function locally fit to restricted ranges of x . Loess begins with
 200 specification of some subset of points along the x-axis at which separate polynomial
 201 regressions of y_i on x_i are to be run. Let any one of these points be designated x_0 . A
 202 regression is run on the set of observations $\{x, y\}$ with x_i nearest x_0 , and in this sense the
 203 regression is “local”. The user specifies the size of the desired neighborhood around x_0 as
 204 a decimal fraction of the total number of observations in the scatterplot through a
 205 smoothing parameter α . For example, $\alpha = 0.5$ specifies that the neighborhood of points
 206 for the local regression will be comprised of the 50 percent of observations with x_i nearest
 207 x_0 . The larger the smoothing parameter α , the less localized the fit, and the smoother the
 208 curve. The fit is “weighted” in the sense that within the neighborhood more importance
 209 is given to observations nearer x_0 in estimating the regression parameters. We use a
 210 locally linear (polynomial degree 1) model for the loess estimation. The estimation
 211 procedure minimizes a weighted sum-of-squares of errors

212

213
$$S = \sum_{i=1}^k g_i(x_0)(y_i - \beta_0 - \beta_1 x_i)^2 \tag{0.6}$$

214 where β_0 and β_1 are regression parameters, $g_i(x_0)$ is a weighting function, and the
 215 summation is over the k observations in the neighborhood of x_0 . The local regression
 216 procedure is repeated for a set of points x_0 distributed along the x-axis, and a predicted
 217 flow \hat{y}_0 is generated at each x_0 . Straight-line segments joining the estimates \hat{y}_0 at each x_0
 218 constitute the loess curve.

219 In our implementation of the loess, the target points x_0 are specified as the
 220 minimum x_i , maximum x_i and $0.05j$ quantiles of x_i , where $j = 1, \dots, 19$. The smoothing
 221 parameter, α , is arrived at by a trial-and-error process starting with trial values $\{\alpha=0.3,$
 222 $0.4, 0.5, 0.6, 0.7, 0.8\}$. The lowest trial α giving a monotonically increasing curve is
 223 selected as the final α . This constraint, consistent with our conceptual model, requires

224 that increasing tree-growth to be associated with increasing flow. Following Martinez
225 and Martinez (2005), the weights g in (0.6) were computed with the tri-cube weight
226 function (Appendix C).

227

228

229 2.4 Interpolation of reconstructed flows. Reconstructed flow for any given year is
230 interpolated from the loess curve using by linear interpolation. Interpolation is necessary
231 because the historical values of x_t will not in general coincide with an x_0 at which loess-
232 curve estimates were made. Furthermore, extension of the loess curve beyond the
233 extremes of x_t in the calibration period may be necessary if flow is to be inferred for years
234 outside the calibration period. For such years, we extrapolate the loess curve linearly: a
235 straight line between the loess curve estimates at the 0.05 quantile and minimum
236 calibration x_t is extended to lower x_t , and a straight line between the loess curve estimates
237 at the 0.95 quantile and maximum calibration x_t is extended to higher x_t .

238

239 2.5 Error bars for reconstructed flows. Reconstructed flows must be accompanied by
240 some estimate of uncertainty. The conceptual model represented by the smoothed
241 scatterplot in Figure 1 implies increased uncertainty of reconstruction with increased
242 wetness. If the loess residuals $\hat{e}_t = y_t - \hat{y}_t$, where y_t is observed flow and \hat{y}_t is flow
243 estimated from the smoothed scatterplot, happen to be approximately normally
244 distributed with constant variance, error bars could be estimated directly from the normal
245 distribution. With nonconstant error variance, as in Figure 1, a theoretical approach to
246 error bars is less tenable. We propose a heuristic approach that also does not rely on any
247 distributional assumption on the residuals and is consistent with the observed pattern of
248 scatter: a confidence interval for any given reconstructed flow is estimated from the
249 cumulative distribution function of weighted-bootstrap cross-validation residuals of
250 reconstructed flows. A time series of cross-validation residuals is first generated by the
251 following steps: 1) the loess curve for the selected value of smoothing parameter α is
252 repeatedly re-fit, each time leaving out a different sequence of 9 consecutive
253 observations, 2) at each re-fitting, the loess curve is applied to predict the flow for the

254 central of the 9 omitted observations, and 3) the residuals (observed minus predicted)
255 from each step are assembled into a single series of residuals¹.

256 For any reconstructed flow, the cross-validation residuals are resampled with
257 replacement (bootstrap) 1000 times to generate a distribution of errors tailored to that
258 particular level of reconstructed flow. A “neighborhood” of reconstructed flows is
259 defined based on the k reconstructed flows in the calibration period nearest the target
260 reconstructed flow. The number k is specified as decimal fraction (e.g., 0.6) of the total
261 number of calibration-period observations. A weighted bootstrap is used so that residuals
262 for predicted flows closer to the target reconstructed flow are more highly represented
263 than residuals for predicted flows further from the target. The bi-square function is used
264 to generate the weights. The bi-square function is computed (Appendix C) and any
265 weights less than 1/100 the maximum weight are dropped.

266 The 1000 noise values are then added to the reconstructed flow to generate 1000
267 noise-added reconstructed flow values for the target year. The $\alpha / 2$ and $1 - \alpha / 2$
268 probability points of the empirical cumulative distribution function of the 1000 values is
269 the $100(1 - \alpha)$ percent confidence interval for the reconstructed flows. For example, the
270 0.10 and 0.90 probability points give the 80 percent confidence interval.

271 The 1000 noise-added reconstructions are also the basic data for estimating
272 uncertainty of statistics derived from the annual reconstructions. For example, a
273 confidence interval for 5-year running mean reconstructed flow can be directly tallied by
274 smoothing each of the 1000 annual series with that running mean and computing the
275 probability points of the empirical cumulative distribution function of the smoothed
276 series in each year.

277

¹ Note that (2) cannot be applied to the first 4 and last 4 observations, because 9 observations are not available there for extracting central values. For those starting and ending observations, fewer than 9 observations are omitted, but a buffer of 4 observations is retained between the cross-validated observation and the nearest observation of the estimation set. For example, at the front end, observations 1-5 are omitted to get the cross-validation residual for observation 1, observations 1-6 are omitted to get the cross-validation residual for observation 2, etc.

278 **3. Sample Application**

279

280 The sample basin is the combined watersheds of the Salt, Verde and Tonto Rivers,
281 Arizona (Figure 3). This semi-arid basin, a major source of water supply for south-
282 central Arizona, has been the subject of previous dendrohydrologic studies (Smith and
283 Stockton 1981; Graybill et al. 2006). Elevation in the 11,142 mi² (28,858 km²) basin
284 ranges from 637 m to 3,846 m (Hawkins 2006). The distribution of annual precipitation
285 is bimodal, with a winter peak associated with disturbances in the westerlies, and a
286 summer peak associated with summer convective storms (Sellers and Hill 1974).
287 Snowmelt accounts for about 39 percent of the annual precipitation (Serreze et al. 1999).
288 Runoff depends strongly on snowmelt (Molotch et al. 2002), and usually peaks in spring
289 (Anderson and White 1986). Total annual precipitation is quite variable over the basin,
290 and ranges from less than 15 inches (380 mm) in the lower elevations to more than 25
291 inches (635 mm) in the high mountains (Anderson and White 1986).

292

293 3.2 Tree-ring data. Tree-ring data for the example consists of autoregressive-residual
294 chronologies of total ring-width for 10 sites in Arizona, western New Mexico, and
295 southeastern Utah (Figure 3, Table 1). Field collections were made in fall of 2005 to
296 update sites sites 1 and 2. The rest of the sites were previously collected by other
297 researchers, and their data were obtained either from the International Tree-Ring Data
298 Bank (ITRDB) or the Laboratory of Tree-Ring Research at the University of Arizona.
299 The starting point for processing was measured ring widths of individual cores or cross-
300 sections. Correlation analysis was used to identify any ring-width series with obvious
301 dating or measuring errors, and these were eliminated from the data set. Each ring-width
302 series was detrended with a cubic smoothing spline with amplitude of frequency response
303 equal to 0.95 at a wavelength twice the series length (Cook and Peters 1981). Core
304 indices were computed by the ratio method, converted to residual indices by
305 autoregressive modeling, and then averaged over cores at the site to produce site
306 chronologies (e.g., see Cook et al. 1990). The common period of coverage by all 10
307 chronologies was 1451-1983. By suggested guidelines (Wigley et al. 1984), all

308 chronologies were sufficiently well-replicated, with subsample signal strength exceeding
309 0.85 for this interval. Variance-stabilization was applied in computing each site
310 chronology to adjust for the expected statistical dependence of the chronology variance
311 on time-varying sample size, or number of cores (Osborn et al. 1997). Step-by-step
312 details of the tree-ring chronology development can be found elsewhere (Appendix 3,
313 SRP Final Report).

314 3.1 Flow data. The observed streamflow series, referred to as “flow” in the remainder of
315 this paper, is the total volume of flow for the water year (October-September) summed
316 over three gages: the Salt River near Roosevelt, the Verde River below Bartlett Dam,
317 and the Tonto River near Roosevelt² (Figure 3). The flow series over 1914-2007 is
318 highly positively skewed and has negligible autocorrelation (Table 2, Figure 4) . A
319 characteristic of the annual flows is the occasional very high flow: flow exceeds 300
320 percent of the median in six years. Some periods (e.g., 1950s) are notable for a long gap
321 between wet years.

322

323 3.3 Filtering and scaling. A comparison of basic descriptive statistics of flow and tree-
324 ring chronologies suggests some obstacles to inferring flows from the tree-ring data
325 (Table 2). Differences in mean or variance between flow and chronologies are
326 unimportant to reconstruction quality, as regression methods are indifferent to linear
327 rescaling of the variables. Skew and autocorrelation, however, deserve further attention.
328 First-order autocorrelation is small for all series – slightly negative for the chronologies
329 and slightly positive for the flow. Autocorrelation reaches weak significance for only one
330 site. The negative autocorrelation for chronologies might seem surprising as these are
331 autoregressive-residual chronologies. However it should be noted that the autoregressive
332 (AR) models for whitening chronologies are based on the full lengths of tree-ring series,
333 while the statistics in Table 2 are for 1914-1982, a common period later used as the
334 reconstruction calibration period. The differing signs of autocorrelation in flow and
335 chronologies might indicate the tree-ring data has been slightly “over-whitened” for the

² Some splicing of records from different gages was necessary for the Verde and Tonto series. The specific gages used are identified in Appendix 1, SRP Final Report

336 objective of flow reconstruction, but the small size of autocorrelations makes this a minor
337 inconsistency.

338 Flow is significantly positively skewed, while eight of the ten chronologies are
339 negatively skewed. This negative skew of chronologies is slight, and reaches
340 significance (0.05 α -level) for only one chronology. In contrast, flow is highly positively
341 skewed. The contrast is readily apparent in quantile-quantile plots, as illustrated for four
342 chronologies in Figure 5. As a linear model would essentially transfer the tree-ring
343 distribution shape onto that of the reconstructed flows, likely problems with the linear
344 model become apparent. First is underestimation of both high and low flows. In other
345 words, the severity of high flows would be understated, and the severity of low flows
346 overstated. A scatter plot of flow on tree-ring index for one of the chronologies
347 illustrates the problem: the least-squares straight line is too low at both the high-flow
348 and low-flow ends (Figure 6A). These characteristics reflect the inability of a straight
349 line to follow the concave curvature in the scatterplot. Another regression complication,
350 heteroskedasticity, is indicated when the residuals from the straight-line fit are plotted
351 against predicted flows: the classic fan-shaped pattern indicates greater variance of errors
352 when predicted flow is high (Figure 6B).

353 Scatter plots of flow on other chronologies (not shown) were also more-or-less
354 curvilinear. The robust regression modeling of flow on lagged chronologies accordingly
355 identified quadratic models as appropriate for the scaling and filtering for each of the 10
356 sites (Table 3). Plots of predicted flow on tree-ring index for four of the sites for which
357 non-lagged models were selected are shown in Figure 7. The quadratic fits are superior
358 to straight lines in capturing the curvature of the relationships, but are notably deficient in
359 parts of the plots, especially toward the high-flow side. For example, a more extreme
360 curvature in the high-index side of the plot for site 4 appears necessary to track the point
361 cluster. Signal strength as measured by the accuracy of the least squares quadratic fit
362 varies greatly over sites – variance explained ranges from 17 to 62 percent (Table 3). All
363 models have positive skill as measured by the reduction-of-error statistic applied in cross-
364 validation and split-sample validation. For three of the sites, the selected filtering-and-
365 scaling model included lags. Plots of estimated flow on the tree-ring index for those sites

366 is of course not a smooth line because the flow estimate for a given year does not depend
367 exclusively on the tree-ring index or its squared value in that year alone. While justified
368 statistically in the stepwise procedure, the lags that entered are not particularly important
369 in influencing the flow estimates. At site 1, for example, the jagged departures in the
370 fitted line from what would be a smooth concave curve— departures imposed by the
371 lagged predictor -- are relatively minor compared with the overall y-axis range of the
372 fitted line (Figure 8).

373

374 3.4 Weighting. The PCA on the SSRs indicates that the first three components account
375 for 79 percent of the variance of SSRs for their 1451-1982 common period (Table 4).
376 PC1, which we adopt as a natural weighting function expressing common variance
377 among SSRs, itself accounts for 63 percent of the variance. The loadings for PC1 reflect
378 the differential signal strength in chronologies, as reflected in the variance-explained
379 statistics for the filtering-and-scaling models listed in Table 3. For example, the site with
380 the highest PC1 loading has the highest individual flow-variance explained, and the three
381 sites with the lowest loadings explain the lowest percentages of flow variance. Use of
382 PC1 as a weighting function to combine the information on flow from the various
383 chronologies therefore most highly weights those chronologies with the strongest
384 individual flow signal, and vice versa.

385

386 3.5 Loess-curve estimation and cross-validation. A scatterplot of observed flows on
387 scores of PC1 of the SSRs is the framework for the loess plot. The scatterplot is repeated
388 with loess curves corresponding to four different trial-and-error choices of smoothing
389 parameter α in Figure 9. The loess curves for $\alpha = 0.3$ and $\alpha = 0.4$ were rejected because
390 they were not monotonically increasing; those curves would imply decreasing flow with
391 increasing tree-ring index for at least part of the range of tree ring index. Loess curves
392 for $\alpha \geq 0.5$ increased monotonically, and would have been acceptable by that criterion.
393 The curves for $\alpha = 0.5$ (not shown) was rejected for containing irregularities (not smooth

394 enough), while the curve for $\alpha = 0.8$ was so smooth that tracking suffered at very high
395 flows and very low flows.

396 The final selected loess fit ($\alpha=0.6$) is plotted again in Figure 10, with a linear
397 extension to accommodate tree-ring data outside its range encountered in the 1914-82
398 calibration period. This curve constitutes the nonparametric reconstruction model. Any
399 PC1 score x_i from the tree-ring record is a pointer to the x-axis of the loess plot, and the
400 corresponding flow is linearly interpolated from a lookup table corresponding to the loess
401 plot. The dashed arrows on this figure illustrate the use of a the curve to infer a flow of
402 3.6 maf (4.4 bcm) from a PC1 score of 110.
403

404 3.6 Error bars. Weighted-bootstrap pseudo-populations of residuals and shapes of
405 weighting functions for generating error bars for reconstructed flows are illustrated in
406 Figure 11 for four consecutive years in the 1500s. The plotted points for each of the four
407 frames are identical, as these are just the residuals (observed flow minus predicted flow)
408 1914-82. The subset of those residuals bootstrapped for each reconstructed year varies,
409 depending on the reconstructed flow. For the extremely dry year 1506, the pseudo-
410 population comes from the left (dry) side of the plot, and consists of residuals tightly
411 clustered around zero. In contrast, for the wet year 1509 the residuals have a much wider
412 range.

413 The weighted-bootstrap sample for generating error bars differs for each year of
414 the reconstructed as the reconstructed flow differs from year to year. The widths of the
415 error bars consequently differ, as shown in Figure 12 for a snapshot of ten years near the
416 start of the 16th century. Note that this snapshot includes the four years whose pseudo-
417 populations of residuals were depicted in Figure 11. The confidence interval is narrow in
418 1506, reflecting the small residuals making up that pseudo-population. The interval is
419 wider for 1507-1509, reflecting the larger residuals bootstrapped. Small differences in
420 width of confidence interval for these years come from slight differences in the
421 bootstrapped population and different weighting functions within those populations.
422 The confidence interval on annual reconstructed flows accordingly varies from year to
423 year over the full-length reconstruction (Figure 13A), and the variation carries over to
424 any smoothed version of the reconstruction, such as the 5-year running-mean (Figure
425 13B). The 80% confidence interval is considerably tighter for wet periods than for dry
426 periods (e.g., early 1600s vs 1660s.). For this particular example, the estimated record
427 low is 0.57 maf (.70 bcm), for the period 1666-1670; the confidence interval indicates a
428 10 percent probability the 5-year mean flow then was as low as 0.42 maf (0.52 bcm).

429

430 3.6 Comparative reconstruction. The reconstruction generated by loess was compared
431 with a reconstruction by linear regression. To simplify the comparison, both methods
432 used the same ten tree-ring chronologies and calibration period. The comparison
433 reconstruction was done using the log-10 transformed flows as a predictand, and the time

434 series of scores of PC1 of the residual chronologies as the predictor. No lags were used
435 in the comparison model, as these were shown to be of minor importance to the loess
436 reconstruction.

437 The regression R^2 indicated that the comparison reconstruction accounted for 69%
438 of the variance of flow in the 1914-82 calibration period. This accuracy is misleading,
439 however, because it applies to log-transformed flows. After back-transformation to
440 original flow units, the variance explained as computed by

$$441 \quad V = 1 - \text{SSE}/\text{SST} \quad (0.7)$$

442 where SSE is the sum of squares of errors (observed minus predicted flows) and SST is
443 the sum of squares of departures of observed flows from the 1914-82 mean, drops to
444 $V = 0.59$. The corresponding value for the loess reconstruction is $V = 0.69$. The loess
445 reconstruction by this measure is therefore a closer fit than the linear-regression
446 reconstruction to the observed flows.

447 The mean, standard deviation, and skew and first-order autocorrelation of
448 observed flows are closer to those of the loess reconstruction than to those of the linear-
449 regression reconstruction (Table 5). The higher standard deviation for the loess
450 reconstruction follows directly from its explaining more variance of the observed flows.
451 Both reconstructions somewhat underestimate the very small positive first-order
452 autocorrelation of the observed flows.

453 Perhaps the more interesting comparisons for alternative reconstruction methods
454 are in the key features of the long-term reconstruction. Mean, standard deviation, skew
455 and first-order autocorrelation are all higher for the 1451-1982 loess reconstruction than
456 for the linear-regression reconstruction. A scatterplot of reconstructed flows by the two
457 methods has a curvature, indicating primarily that the loess approach give greater
458 extremes on the high-flow end and lesser extremes on the low-flow end (Figure 14). A
459 time series plot for a snapshot of reconstructed flows in the 1500s shows that the
460 differences in individual years are appreciable (Figure 15). For example, in wet years
461 1509 and 1549 the loess reconstruction is on the order of 0.5 maf (0.6 bcm) higher than
462 the regression reconstruction. Differences in dry years are much smaller – on the order
463 of 0.1 maf (0.1 bcm).

464 Cumulative distribution functions (cdfs) for the two reconstructions similarly
465 point to the tendency for the loess reconstruction to have both higher highs and higher
466 lows than the linear-regression reconstruction (Figure 16). For the long-term record and
467 the 1914-82 calibration period, the cdf of the loess reconstruction is shifted to the right of
468 the cdf of the linear regression reconstruction. The shift is most apparent in the tails.

469 Differences in individual years are expected to carry over to differences in multi-
470 year flow anomalies. In Figure 17, the extremes in running means of length 1-20 years
471 are compared. The highest running means (wettest periods) are plotted in Figure 17A,
472 and the lowest running means (driest periods) in Figure 17B. Large differences in the
473 two reconstructions are apparent even for running means longer than 10 years. For
474 example, the wettest 20-year period is about 25 percent higher in flow for the loess
475 reconstruction, and the driest 20-year period about 5 percent higher in flow for the loess
476 reconstruction.

477

478 **4. Discussion and Conclusions**

479 Scatterplot smoothing is a useful and intuitively appealing approach to reconstruction of
480 streamflow from tree rings when nonlinearity and heteroskedasticity of errors make linear
481 regression problematic. An ideal setting for this approach would be a simple bivariate
482 problem in which flow was to be inferred from a single tree-ring series – which might be
483 an average over site chronologies. In practice, the problem is more complicated because
484 individual tree-ring chronologies might have greatly varying strength of signal for flow,
485 the relationships between chronologies and flow might include lags, and sites might be
486 clustered spatially so that averaging over chronologies would unfairly emphasize some
487 parts of the basin over others that may be of more importance to runoff. The approach
488 presented here relies on a combination of parametric and non-parametric methods to
489 address these complications.

490 The intermediate step of converting individual tree-ring chronologies to estimates
491 of flow (“scaling and filtering”) before weighting them into the predictor for the loess
492 scatterplot runs the risk of overfitting. We attempted to minimize this risk by restricting

493 the filtering-and-scaling models to have a fairly simple quadratic form and by using
494 cross-validation to guard against unwarranted entry of lagged predictors.

495 Variations on the proposed approach might be preferable in some circumstances.
496 In the absence of heteroskedasticity of errors, the weighted bootstrapping could be omitted
497 and reconstruction error bars generated from the error variance and some suitably fitted
498 error distribution. A simpler approach to reduction of the individual tree-ring
499 chronologies to a single predictor might also be possible if all chronologies showed a
500 similarly strong signal for flow and if lagged response could be ignored. In the extreme,
501 this might amount to the loess-curve “predictor” being merely the arithmetic average of a
502 number of tree-ring chronologies in the basin. A less extreme departure would be to use
503 some other form of regression model for the single-site reconstructions. We used robust
504 regression partly based on exploratory analysis that showed some severe outliers in
505 scatterplots of flow on individual tree-ring chronologies. Without this complication,
506 ordinary regression, with quadratic or some other power-transformation of the tree-ring
507 indices, might be preferred. Another possibility would be generate the single-site
508 reconstructions by loess curves (scatterplot smoothing) instead of quadratic regression.
509 We chose not to do that in the interest of automating the reconstruction process: for
510 reconstruction problems with many tree-ring chronologies, the scatterplot smoothing was
511 judged to be too tedious in requiring a large number of subjective decisions (e.g., on
512 smoothing parameters for individual scatterplots).

513 The method described here can be readily extended to accommodate time-nested
514 tree-ring reconstruction models (e.g., Meko et al. 2001) . For brevity, we have restricted
515 the sample application to reconstruction for a specific time period uniformly covered by
516 all the tree ring chronologies. In practice, some subset of chronologies will extend much
517 further back in time, and some may come nearly up to present. For those situations, the
518 modeling procedure would simply be repeated such that multiple reconstructions were
519 generated. If these happened to overlap, as is usual, priority could be given in assigning
520 the final reconstructed flow for a given year to the model deemed superior in some way
521 (e.g., lowest validation error variance).

522 The proposed approach is perhaps most blatantly subjective in selection of the
523 smoothing parameter α for the loess curve. Our conceptual model dictated constraints on
524 this choice. The conceptual model is that trees grow faster as conditions become wetter,
525 but may respond less to increased moisture under very wet conditions: we accordingly
526 assume the curve fit to a scatterplot of flow on tree-ring index should increase
527 monotonically, and perhaps increase in steepness toward high flows. These constraints
528 would perhaps not apply in some studies. For example, situations could be envisioned in
529 which very moist conditions become detrimental to growth. Besides the smoothing
530 parameter, the other parameter typically varied in loess smoothing is the degree- of-
531 polynomial, λ (Martinez and Martinez 2005). We choose a locally linear model,
532 corresponding to $\lambda = 1$. The locally-linear model has been recommended for having well-
533 behaved end-effects (Martinez and Martinez 2005). The loess curve by the local-linear
534 model could also be readily extended to allow estimation of flow for those years with
535 tree-ring data outside its calibration-period range.

536 The confidence intervals from weighted-boostrapping of cross-validation
537 residuals are of course not exactly reproducible, as a bootstrap sampling will generate a
538 different sample when repeated. The size of the bootstrap sample (e.g., 1000 vs 5000)
539 will affect the repeatability of the confidence interval; an optimal size for particular
540 problems could be fine-tuned with Monte Carlo sensitivity studies. Another limitation
541 that should be mentioned is the violation of the regression assumption of constant error
542 variance. This violation makes can affect properties (e.g., bias, variance) of estimated
543 regression predictors coefficients, and have important implications for hypothesis testing
544 using those coefficients (e.g., Kennedy 2003). Inferences based on significance of
545 coefficients is not an issue in flow reconstruction. The degree to which the violation
546 negatively impacts the predictions in a flow reconstruction model is unknown. The most
547 important points in the reconstruction application are that 1) the smooth curve gives and
548 acceptable and reasonably robust fit to the scatterplot, and 2) the error bars reflect any
549 systematic dependence of reconstruction accuracy on level of predicted flow.
550 Application of the method presented here should include cross-validation and graphical
551 quality control to ensure that these points are addressed. Alternatively, data
552 transformation could be investigated in combination with the loess approach to

553 circumvent the problem of heteroskedasticity of errors – with the tradeoff of likely need
554 for backtransformation to make reconstructed flows usable.

555 The reconstruction approach proposed here is not claimed to be superior in
556 general to any existing statistical reconstruction method, but is proposed as a useful
557 alternative for reconstruction scenarios plagued by nonlinearity of relationships and
558 heteroskedasticity of errors. The scenario is most likely in small, semi-arid basins, where
559 flashy flow regimes may produce highly-skewed flows and where high flows are unlikely
560 to leave a footprint in tree-growth commensurate with the flow anomaly. Comparative
561 exercises can point out sensitivity of major reconstructed time series features to choice of
562 reconstruction method. Our limited comparison with a more convention linear-regression
563 reconstruction model indicates the loess procedure as implemented here does give
564 practically significant differences in reconstructed features. The more extreme highs and
565 less extreme lows by the loess reconstruction are a direct consequence of the effort to
566 deal with curvature in the relationship between tree-ring index and flow. Other types of
567 statistical models, not investigated here, may be applicable to flow-reconstruction
568 problems in which the data are not particularly well-suited for regression. Response
569 surfaces (Graumlich 1993), neural networks (Zhang et al. 1999; Woodhouse 1999; Ni et
570 al. 2002), and classification trees (Meko and Baisan 2001) are some of the other
571 techniques that have been used to deal with nonlinearity in tree-ring reconstruction
572 models.

573

574 **Acknowledgements.** This work was supported by the Salt River Project.

575

576

577 **References**

578

579 Anderson, T., and White, Natalie D., 1986, Arizona Surface-Water Resources, National
580 Water Summary 1985 -- Hydrologic events and surface-water resources, United
581 States Geological Survey Water-Supply Paper 2300, 145-150.

582 Box, G.E.P., and Jenkins, G.M., 1976, Time series analysis: forecasting and control: San
583 Francisco, Holden Day, p. 575 pp.

584 Cleaveland, M.K., 2000, A 963-year reconstruction of summer (JJA) streamflow in the
585 White River, Arkansas, USA, from tree rings: *The Holocene*, v. 10, no. 1, p. 33-
586 41.

587 Cleveland, W.S., and Devlin, S.J., 1988, Locally weighted regression: An approach to
588 regression analysis by local fitting: *Journal of the American Statistical*
589 *Association*, v. 83, no. 403, p. 596-610.

590 Cook, E.R., Briffa, K., Shiyatov, S., and Mazepa, V., 1990, Tree-ring standardization and
591 growth-trend estimation, *in* Cook, E.R., and Kairiukstis, L.A., eds., *Methods of*
592 *Dendrochronology, Applications in the Environmental Sciences*: Kluwer
593 Academic Publishers, p. 104-123.

594 -----, and Peters, K., 1981, The smoothing spline: A new approach to standardizing forest
595 interior tree-ring width series for dendroclimatic studies, *Tree-Ring Bulletin* 41,
596 45-53.

597 Graumlich, L., 1993, A 1000-year record of temperature and precipitation in the Sierra
598 Nevada, *Quaternary Research* 39, 249-255.

599 Graybill, D.A., Gregory, D.A., Funkhouser, G.S., and Nials, F.L., 2006, Long-term
600 streamflow reconstructions, river channel morphology, and aboriginal irrigation
601 systems along the Salt and Gila Rivers, *In: Environmental Change and Human*
602 *Adaptation in the Ancient Southwest*, edited by Jeffrey S. Dean and David E.
603 Doyel. University of Utah Press.

604 Hawkins, T.W., 2006a, Parameterization of the snowmelt runoff model for the Salt-Verde
605 system, Arizona, during drought conditions: *Journal of the Arizona-Nevada*
606 *Academy of Science*, v. 380, no. 2, p. 63-73.

607 Hoaglin, D.C., Mosteller, F., and Tukey, J.W., 1983, *Understanding Robust and*
608 *Exploratory Data Analysis*: John Wiley & Sons, Inc., New York, p. 447.

609 Huber, P.J., 1973, Robust regression: Asymptotics, conjectures, and Monte Carlo: *Annals*
610 *of Statistics*, v. 1, p. 799-821.

611 Kennedy, P., 2003, *A guide to econometrics*: Cambridge, Massachusetts, The MIT Press.

612 Mardia, K., Kent, J., and Bibby, J., 1979, *Multivariate Analysis*: Academic Press, 518.

613 Martinez, W.L., and Martinez, A.R., 2005, *Exploratory data analysis with MATLAB*:
614 New York, Chapman & Hall/CRC.

615 Meko, D.M., 1997, Dendroclimatic reconstruction with time varying subsets of tree
616 indices: *Journal of Climate*, v. 10, p. 687-696.

617 -----, and Baisan, C.H., 2001, Pilot study of latewood-width of conifers as an indicator of
618 variability of summer rainfall in the north American Monsoon region:
619 *International J. of Climatology*, v. 21, p. 697-708.

620 -----, and Graybill, D.A., 1995, Tree-ring reconstruction of Upper Gila River discharge:
621 *Water Resources Bulletin*, v. 31, no. 4, p. 605-616.

622 -----, Therrell, M.D., Baisan, C.H., and Hughes, M.K., 2001, Sacramento River flow
623 reconstructed to A.D. 869 from tree rings: *J. of the American Water Resources*
624 *Association*, v. 37, no. 4, p. 1029-1040.

625 Michaelsen, J., 1987, Cross-validation in statistical climate forecast models, *J. of Climate*
626 *and Applied Meteorology* 26, 1589-1600.

627 Molotch, N.P., Fassnacht, S.R., Helfrich, S.R., and Bales, R., C., 2002, Estimating the
628 distribution of snow water equivalent and snow extent beneath the cloud-cover in
629 the Salt-Verde river basin, Arizona: 59th Eastern Snow conference, Stowe,
630 Vermont USA.

631 Mosely, M.P., and McKerchar, A.I., 1993, Streamflow, *in* David R. Maidment, ed.,
632 *Handbook of Hydrology*: McGraw-Hill Book Company.

633 Myers, R.H., 1990, Classical and modern regression with applications, second edition:
634 Pacific Grove, California, Duxbury.

635 Ni, F., Cavazos, T., Hughes, M.K., Comrie, A.C., and Funkhouser, G., 2002, Cool-season
636 precipitation in the southwestern USA since A.D. 1000: Comparison of linear and
637 nonlinear techniques for reconstruction: *International Journal of Climatology*, v.
638 22, p. 1645-1662.

639 Osborn, T.J., Briffa, K.R., and Jones, P.D., 1997, Adjusting variance for sample-size in
640 tree-ring chronologies and other regional mean timeseries: *Dendrochronologia*, v.
641 15, p. 89-99.

642 Sellers, W.D., and Hill, R.H., 197, *Arizona Climate 1931-1972*, The University of
643 Arizona Press, Tucson, Arizona, 616 pp.

644 Serreze, M., Clark, Armstrong, R.L., McGuiness, D.A., and Pulwarty, R.S., 1999,
645 Characteristics of the western United States snowpack from snowpack telemetry
646 (SNOTEL) data: *Water Resources Research*, v. 35, no. 7, p. 2145-2160.

647 Smith, L.P., and Stockton, C.W., 1981, Reconstructed streamflow for the Salt and Verde
648 Rivers from tree-ring data: *Water Resources Bulletin*, v. 17, no. 6, p. 939-947.

649 Stockton, C.W., and Jacoby, G.C., 1976, Long-term surface-water supply and streamflow
650 trends in the Upper Colorado River Basin, *Lake Powell Research Project Bulletin*
651 No. 18: National Science Foundation, 70 pp.

652 Wigley, T.M.L., Briffa, K.R., and Jones, P.D., 1984, On the average value of correlated
653 time series, with applications in dendroclimatology and hydrometeorology:
654 *Journal of Climate and Applied Meteorology*, v. 23, p. 201-213.

655 Woodhouse, C.A., 1999, Artificial neural networks and dendroclimatic reconstructions:
656 An example from the Front Range, Colorado, USA: *The Holocene*, v. 9, no. 5, p.
657 521-529.

658 Zhang, Q.-B., Hebda, R.J., Zhang, Q.-J., and Alfaro, R.I., 1999, Modeling tree-ring
659 growth responses to climatic variables using artificial neural networks: *Forest*
660 *Science*, v. 46, no. 2, p. 229-238.

1 **Appendix A. Statistical software**

2 All computations were done in MATLAB©, using a combination of functions written by
3 us and available from MATLAB© and the Computational Statistics Toolbox (Martinez
4 and Martinez 2002). Numerous user-written functions and scripts were also used in the
5 analysis.

1 **Appendix B. Robust Regression and Huber Weighting Function**

2 In linear regression the predictand for any given observation or year i is modeled as

3
$$y_i = \mathbf{x}_i' \mathbf{b} + e_i, \quad i = 1, 2, \dots, n \quad (1.1)$$

4 where y_i is the predictand, \mathbf{x}_i is a row vector of predictors, \mathbf{b} is a vector of coefficients,

5 and e_i is the noise. Predictions can be generated by $\hat{y}_i = \mathbf{x}_i' \hat{\mathbf{b}}$, where the “hat” denotes

6 estimated values, such that the residuals \hat{e}_i are defined as

7
$$\hat{e}_i = y_i - \hat{y}_i \quad (1.2)$$

8 While ordinary least squares regression seeks to minimize $\sum_{i=1}^n \hat{e}_i^2$, robust regression seeks

9 to minimize $\sum_{i=1}^n w_i \hat{e}_i^2$ where w_i are weights that can be used discount the importance of

10 outliers in the fit (Myers 1990).

11 Robust regression with the Huber function was implemented with MATLAB
12 function `robustfit`. An ordinary least squares fit (non-robust) is first run, and the
13 diagonals of the “hat” matrix, $h_{i,j}$, $j = i$ stored. (The “hat” has been dropped from the
14 estimated quantities in the following to avoid clutter.) Scaled residuals are then
15 computed as

16
$$r_i = \frac{e_i}{a\hat{\sigma}\sqrt{1-h_{i,i}}} \quad (1.3)$$

17 where $a = 1.345$ is a specified Huber “tuning factor”, and

18
$$\hat{\sigma} = \frac{\text{MED}(|e_i|)}{0.6745} \quad (1.4)$$

19 is a scale factor reflecting the variability of the residuals. The Huber weights are given
20 by

$$21 \quad w_i = \begin{cases} 1; & |r| \leq 1 \\ 1/|r|; & |r| > 1 \end{cases} \quad (1.5)$$

22

23 The robust estimates of regression parameters are derived iteratively:

- 24 1) Run an ordinary least squares to obtain initial, non-robust, parameter estimates
25 \mathbf{b}_0 , initial residuals $e_{i,0}$, and diagonal elements of the hat matrix, $h_{i,i}$
- 26 2) Using those residuals, compute the initial scaled residuals $r_{i,0}$ from (1.3) and initial
27 weights $w_{i,0}$ from (1.4) and (1.5)
- 28 3) Use weighted least squares (e.g., Weisberg 1985) to get new robust parameter
29 estimates $\mathbf{b}_{R0} = (\mathbf{X}'\mathbf{W}_0\mathbf{X})^{-1} \mathbf{X}'\mathbf{W}_0\mathbf{y}$, where \mathbf{W}_0 is a diagonal matrix of weights with
30 i^{th} diagonal element $w_{i,0}$
- 31 4) Let the parameter estimates from step 3 take the role of starting parameters, and
32 apply (1.3), (1.4), and (1.5) to get new residuals, a new value of $\hat{\sigma}$, and new
33 weights
- 34 5) Go back to step 3
- 35 6) Repeat steps 3 and 4 until convergence is reached

36

37 In function `robustfit`, the constant 0.6745 makes the estimate unbiased for the
38 normal distribution. If there are p columns in \mathbf{X} , the smallest p absolute deviations are
39 excluded when computing the median absolute residual, $\text{MED}(|e_i|)$. Convergence is

40 assumed when the maximum change in any of the parameters from the previous step
41 is very small -- does not exceed some specified small number (tied to the machine
42 precision). The iteratively reweighted least squares is otherwise continued to a
43 maximum possible 50 iterations,

1 **Appendix C. Bisquare and tri-cube weighting functions**

2 The bisquare and tri-cube weighting functions are described in the context of robust
3 statistics by Martinez and Martinez (2005). Given a set of observations $\{x_i, y_i\}$ and
4 some arbitrary point x_0 on the x-axis, define the k nearest neighbors as the k points with x_i
5 closest to x_0 . Define the relative distance of any observation from x_0 as

6
$$u = \frac{|x_0 - x_i|}{\Delta_k(x_0)} \tag{1.1}$$

7 where $\Delta_k(x_0)$ is the largest $|x_0 - x_i|$ for any of the k nearest neighbors. The bisquare
8 weights are

9
$$w_i(x_0) = \begin{cases} (1-u^2)^2; & 0 \leq u < 1 \\ 0; & \text{otherwise} \end{cases} \tag{1.2}$$

10 and the tri-cube weights are

11
$$w_i^*(x_0) = \begin{cases} (1-u^3)^3; & 0 \leq u < 1 \\ 0; & \text{otherwise} \end{cases} \tag{1.3}$$

12

13 Equations (1.2) and (1.3) along with the definition of u specify that the weight
14 decreases from a maximum for the observation nearest x_0 to zero for the farthest
15 observation in the neighborhood, and remains zero for any observation outside the
16 neighborhood. Weights are cannot be less than zero or greater than 1, and a weight 1
17 occurs only if some observation happens to have an x -value exactly at x_0 . The bisquare
18 and tri-cube weighting functions both decline with a sigmoid shape, but steepness of
19 decline differs such that the tri-cube weights are higher than the bisquare weights for
20 points near x_0 and lower than the bisquare weights for points far from x_0 (Figure ??).

21

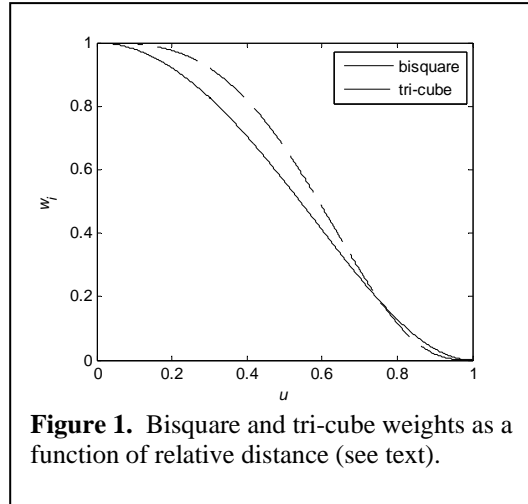


Table 1. Site information on tree-ring chronologies.

N ^a	Name ^b	Species ^c	Location ^d			N _{tree} ^e	Period ^f		s ^g
			Lat	Lon	El(ft)				
1	Wahl Knoll	PSME	34.0	-109.4	9367	23(3)	1435	2005	a
2	Black Mtn	PSME	33.4	-108.2	8721	41(8)	1327	2005	a
3	Navajo Mtn	PIED	37.0	-110.8	7384	13(6)	1330	1989	b
4	Kane Springs	PIED	37.5	-109.9	6350	26(5)	1361	1988	b
5	Betatakin Cyn	PSME	36.7	-110.5	6599	26(9)	1306	1989	b
6	Walnut Cyn	PIPO	35.2	-111.5	6696	17(2)	1451	1987	b
7	Spider Rock	MIX	36.1	-109.3	6105	25(6)	1399	1989	b
8	Satan Pass	PSME	35.6	-108.1	7384	17(2)	1410	1990	b
9	Dinnebito	PIED	36.2	-110.5	6202	28(6)	1410	1983	b
10	El Malpais	PSME	35.0	-108.1	7826	21(16)	1100	1990	c

^aSite as numbered on map

^bSite name

^cSpecies: PSME=Pseudotsuga menziesii; PIED= Pinus edulis
PIPO=Pinus ponderosa; Mix = mix of PSME and PIPO

^dLocation: latitude and longitude in decimal degrees, elevation
in feet above sea level

^eMaximum number of trees in any year (number in 1541)

^fStart and end year of site chronology

^gSource of data: a=site updated by D. Meko in July 2005;

b=Southwest Archaeology Project (J. S. Dean);

c=Site collected by H. Grissino-Mayer (ringwidths are subset of
file nm572.rwl downloaded from International Tree-Ring Data Bank

Table 2. Statistics of annual flows and tree-ring chronologies.

Statistics ^b					
Series ^a	Mean	Std Dev	Skew	Normal	r(1)
Flow	1.201	0.8521	1.50**	F**	0.12
Site 1	1.001	0.1515	-0.22	P	-0.17
Site 2	1.009	0.2372	-0.25	P	-0.15
Site 3	0.984	0.2371	-0.02	F*	-0.06
Site 4	0.984	0.1987	-0.38	P	-0.18
Site 5	1.007	0.1869	-0.36	P	-0.11
Site 6	1.027	0.2239	-0.53*	P	-0.06
Site 7	1.002	0.2057	-0.09	P	-0.18
Site 8	0.999	0.2794	0.30	P	-0.15
Site 9	1.018	0.2454	0.06	P	-0.07
Site 10	1.002	0.2951	-0.04	P	-0.19*

^aSeries: first series is annual flows, others are residual tree-ring chronologies as numbered on map in Figure 2

^bStatistics are the mean, standard deviation, skewness, Lilliefors test for normality, and first-order autocorrelation. Significance at 0.05 and 0.01 alpha-levels are flagged by "*" and "**". Flow statistics (mean and standard deviation) are in millions of cubic feet, and index statistics are dimensionless. Analysis period is 1914-82.

Table 3. Summary statistics of filtering-and-scaling models for tree-ring chronologies.

N ^a	Model ^b	VE ^c	Cross-validation ^d			Split-sample ^e		
			RMSE (maf)	MAE (maf)	MedAE (maf)	RE	REA	REB
1	2:2000	0.26	25.2	15.6	8.4	0.17	0.23	0.23
2	2:0000	0.17	25.9	16.5	10.0	0.12	0.10	0.26
3	2:0000	0.46	20.8	13.9	9.5	0.38	0.49	0.41
4	2:0000	0.36	22.4	14.8	8.6	0.28	0.42	0.29
5	2:0000	0.34	22.3	14.9	10.1	0.29	0.43	0.22
6	2:0002	0.31	23.9	16.0	9.5	0.22	0.20	0.40
7	2:0000	0.41	21.7	13.8	8.1	0.32	0.49	0.29
8	2:0000	0.40	22.4	15.1	9.9	0.28	0.40	0.40
9	2:0020	0.62	19.0	13.7	9.8	0.50	0.57	0.65
10	12:0000	0.23	24.8	17.0	11.1	0.12	0.25	0.13

^aN: sites number as used on map

^bModel: code defining regression model of flow on current and lagged tree-ring index; digits before colon indicate power(s) of current-year index (e.g., 2 denotes 'squared'); columns after colon indicate powers on index at lags t-1, t-2, t+1 and t+2 years relative to the year of flow

^cVE: variance-explained statistic computed from observed and predicted flows on the loess plot. The statistic is defined as $VE = 1 - SSE/SST$, where SSE is the sum of squares of differences of observed flows and predicted flows and SST is the sum of squares of departures of observed flows from their mean.

^dCross-validation: statistics from leave-9-out cross-validation of the loess curve. RMSE=root mean square error; MAE=mean absolute error; MedAE=median absolute error; RE=reduction-of-error statistic (see text)

^eSplit-sample: Reduction-of-error statistics from split-sample calibration-validation with calibration on first half of data and validation on second half (REA) and for calibration on the second half and validation on the first half (REB)

Table 4. Summary of principal components analysis on filtered and scaled tree-ring chronologies, 1451-1982. Table truncated to include only first three components accounting for cumulative 79 percent of variance.

Loadings			

Site ^a	PC#1	PC#2	PC#3

1	0.178	0.407	-0.416
2	0.190	0.366	-0.090
3	0.394	-0.315	0.381
4	0.358	-0.192	0.163
5	0.295	0.018	0.238
6	0.308	-0.144	-0.638
7	0.333	0.234	-0.059
8	0.289	0.414	0.357
9	0.478	-0.398	-0.201
10	0.207	0.398	0.108
Var	62.7	10.8	5.4

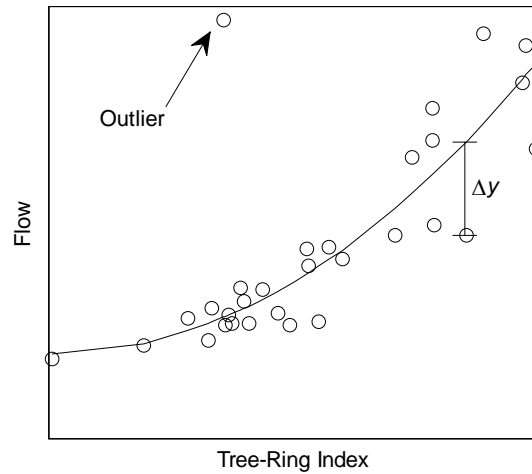
^aSite: First 10 rows correspond to tree-ring sites as numbered on map in Figure 2 and listed in Table 1. Last row is percentage of variance accounted for by the PC

Table 5. Summary statistics observed flows and flows reconstructed by two alternative methods.

Statistics ^a				
Period	Mean	Stdev	Skew	r(1)

1914-1982				
Loess	1.229	0.698	0.92	0.01
Regr	1.112	0.556	0.66	-0.06
Obs	1.201	0.852	1.50	0.12
1541-1982				
Loess	1.229	0.699	1.08	0.05
Regr	1.102	0.553	0.73	0.01

^aStatistics: mean (maf), standard deviation and first-order autocorrelation



Sketch of idealized relationship between tree-ring index and annual streamflow for a tree-ring chronology in a semi-arid watershed. Growth-response to additional moisture flattens at high moisture levels, leading to concave curvilinear scatter. Relationship weakens as moisture increases, leading to larger “errors” (Δy), or departures from fitted smooth line, for high flows. Smoothed line is quadratic fit to data with the outlier omitted.

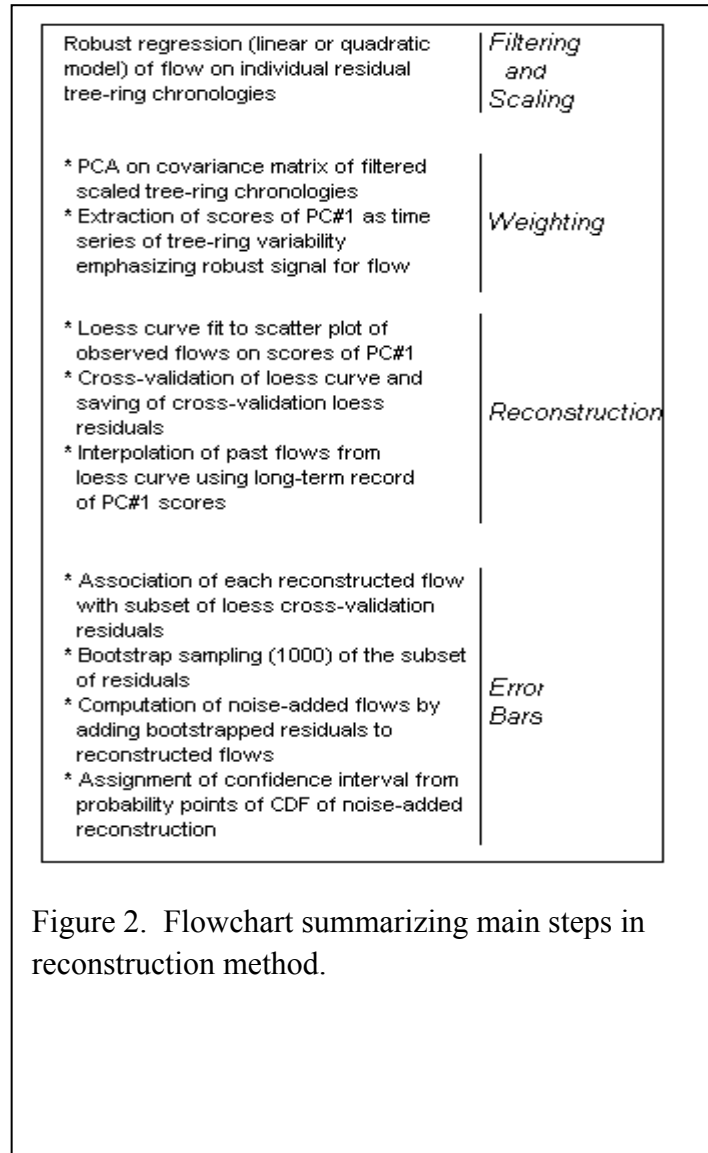


Figure 2. Flowchart summarizing main steps in reconstruction method.

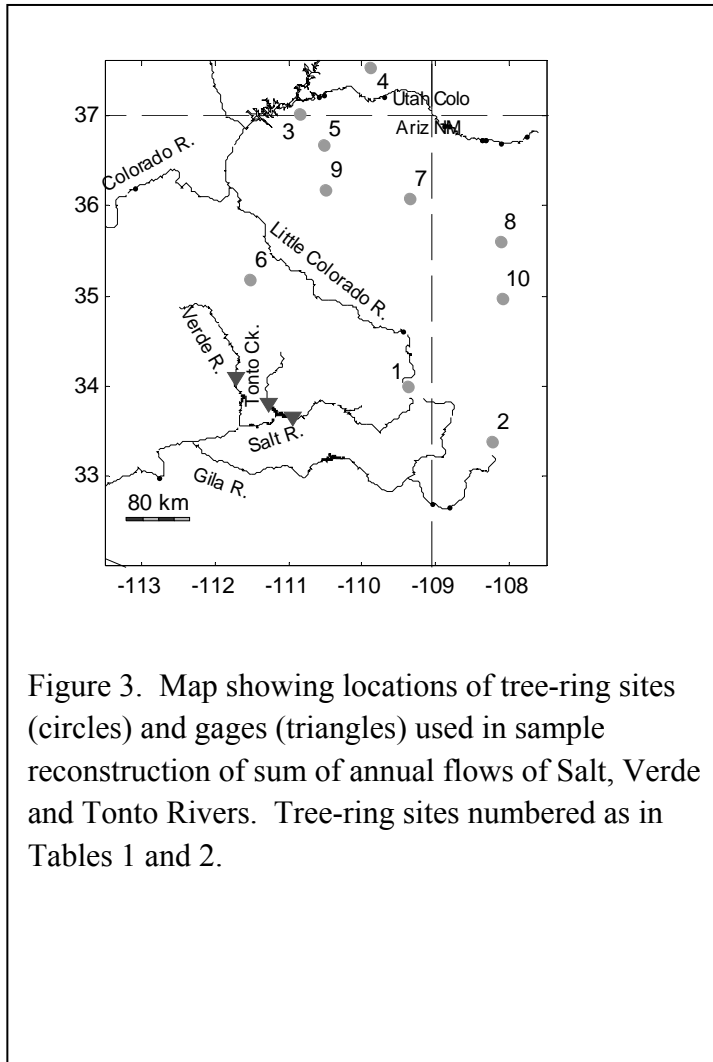


Figure 3. Map showing locations of tree-ring sites (circles) and gages (triangles) used in sample reconstruction of sum of annual flows of Salt, Verde and Tonto Rivers. Tree-ring sites numbered as in Tables 1 and 2.

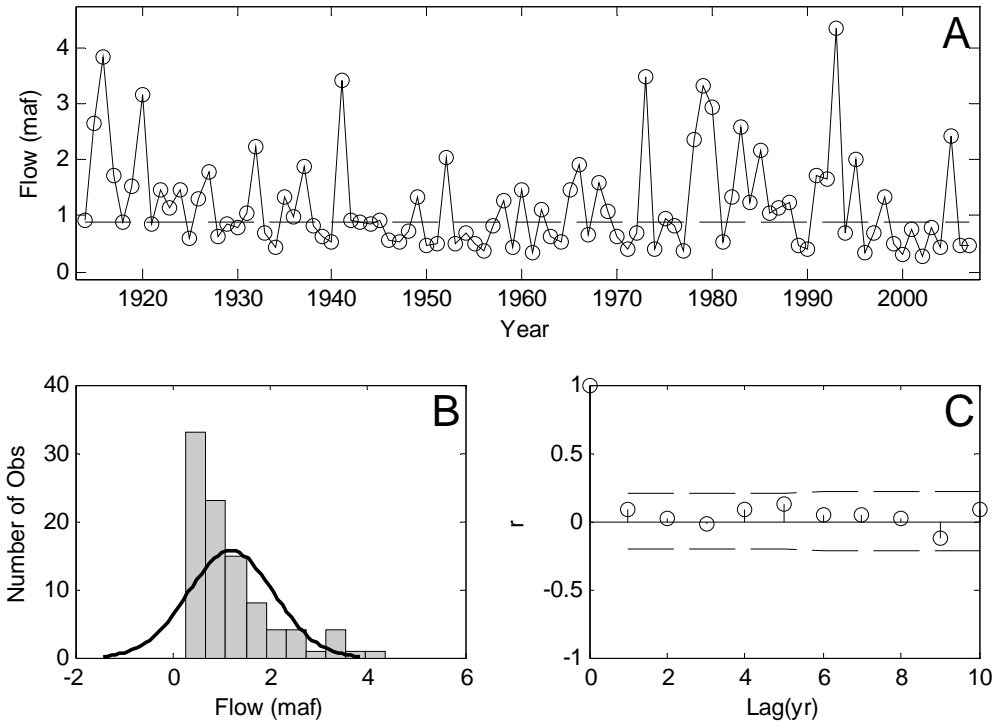


Figure 4. Statistical characteristics of observed water-year total flows, 1914-2007. (A) Time series plot, with dashed horizontal line at 1914-2006 median. (B) Histogram with superposed pdf of theoretical normal distribution. (C) Autocorrelation function with 95% confidence interval defines as two time the large-lag standard error (Box and Jenkins 1976).

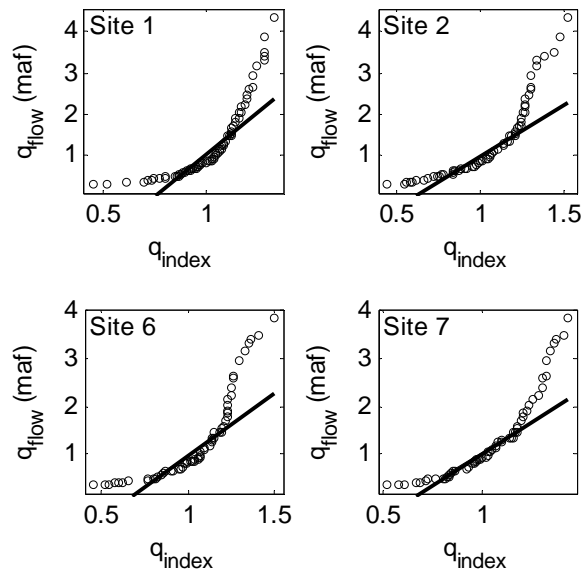


Figure 5. Quantile-quantile plots illustrating differences in shapes of distributions of annual flows and tree-ring indices. Flow series is SVT water-year total. Tree-ring series are residual chronologies (Cook et al. 1990) for four sites in study are (numbered as in Figure 3 and Table 1).

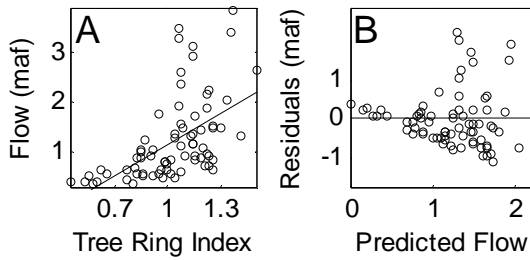


Figure 6. Scatter plots illustrating effects of nonlinearity and non-variance of errors in reconstruction by simple linear regression of flow on untransformed tree-ring index. (A) observed flows against tree-ring index. (B) regression residuals against flows generated by simple linear regression model. Data are SVT flows and chronology #6 (Table 1, Figure 3) for common period 1914-87.

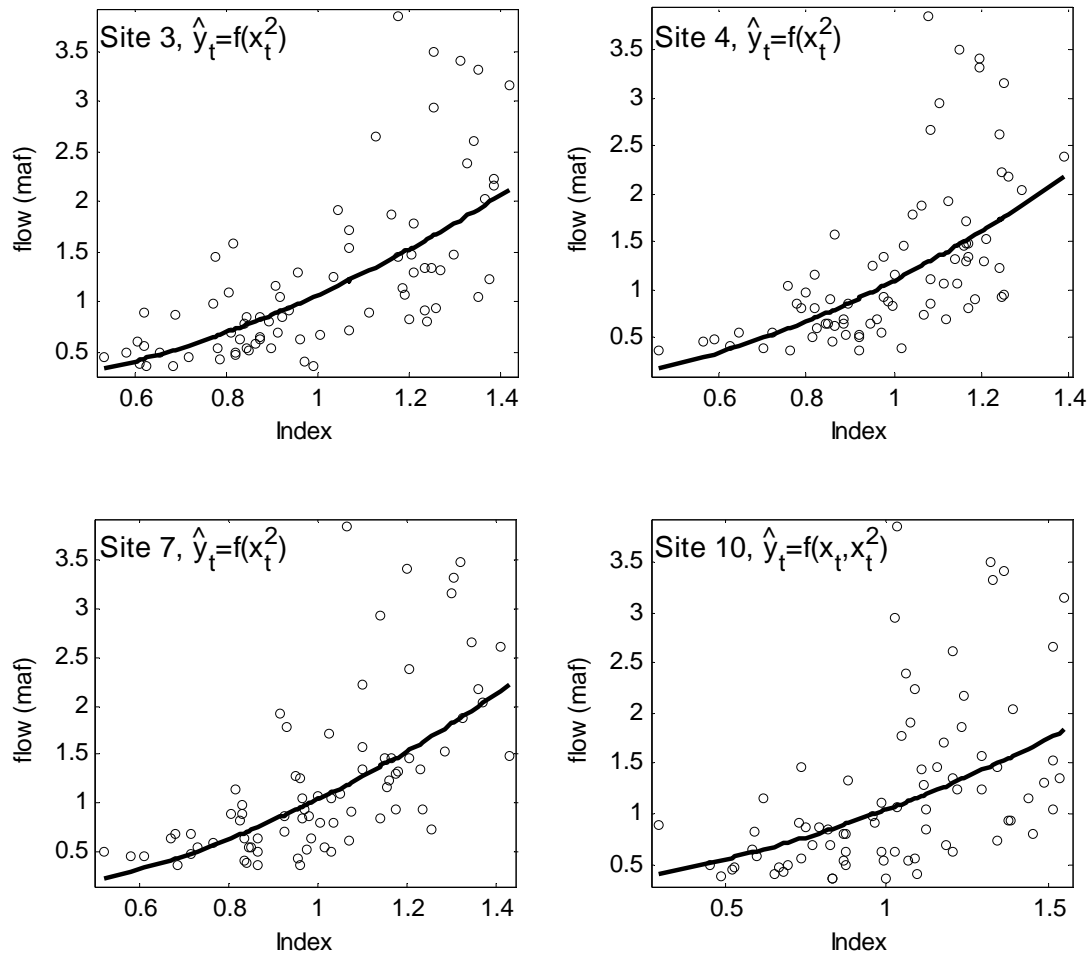


Figure 7. Sample scatterplots and predictions from quadratic robust regression models scaling and filtering tree-ring chronologies into estimates of flow. Sites numbered as in Figure 3 and Table 1. Models for all ten sites listed in Table 3.

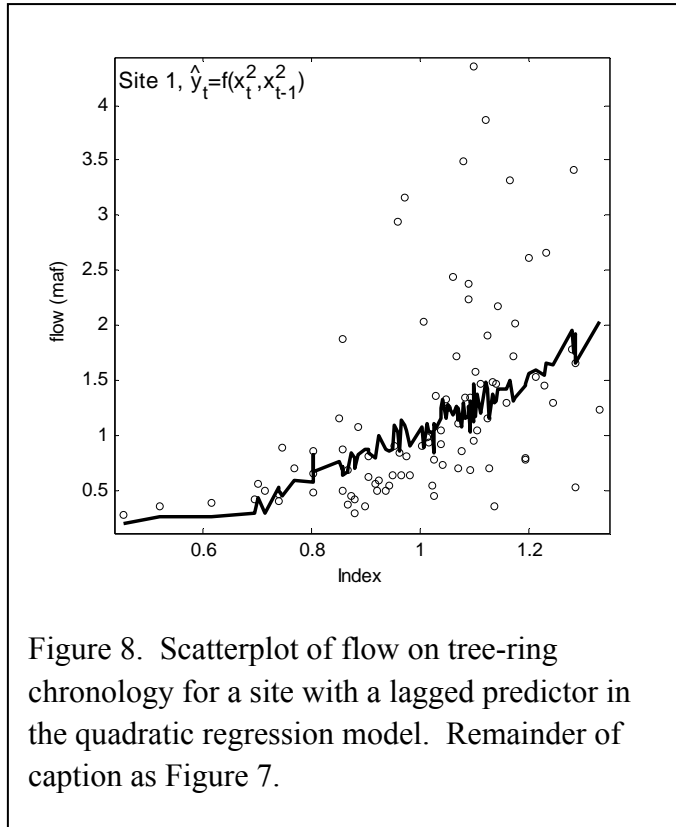


Figure 8. Scatterplot of flow on tree-ring chronology for a site with a lagged predictor in the quadratic regression model. Remainder of caption as Figure 7.

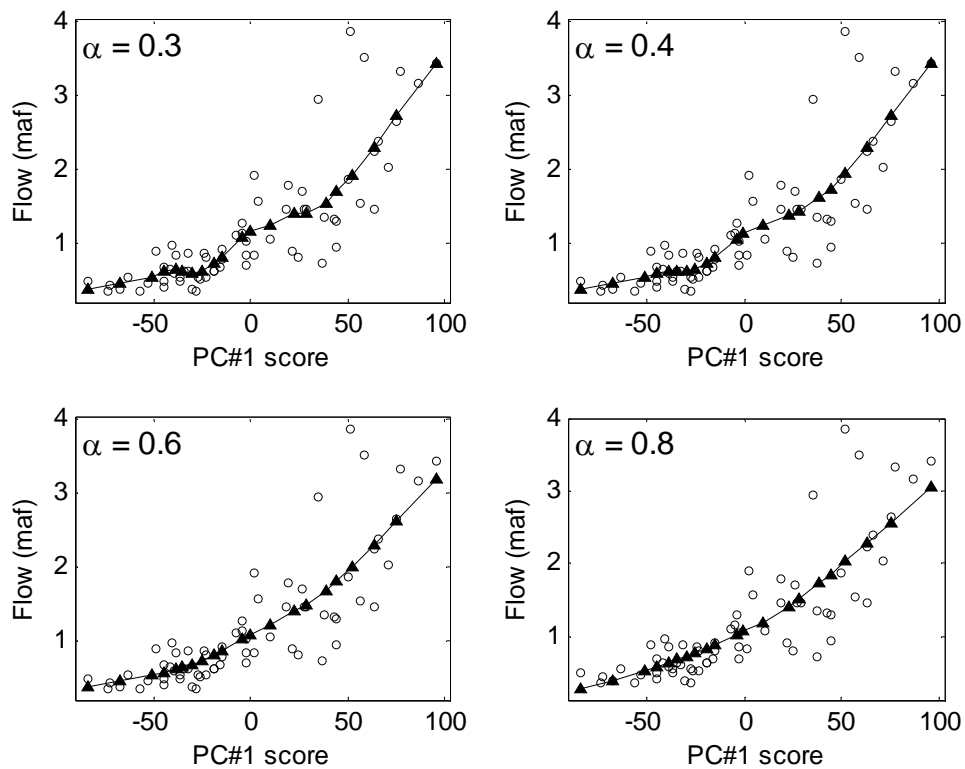


Figure 9. Plots illustrating change in smoothness of loess plot with change in smoothing parameter, α . Open circles are observed data. Loess curve is composed of straight line segments connecting triangles located at minimum, maximum, and 0.05, 0.10, ..., 0.95 quantiles of PC#1 scores. Analysis period 1914-82.

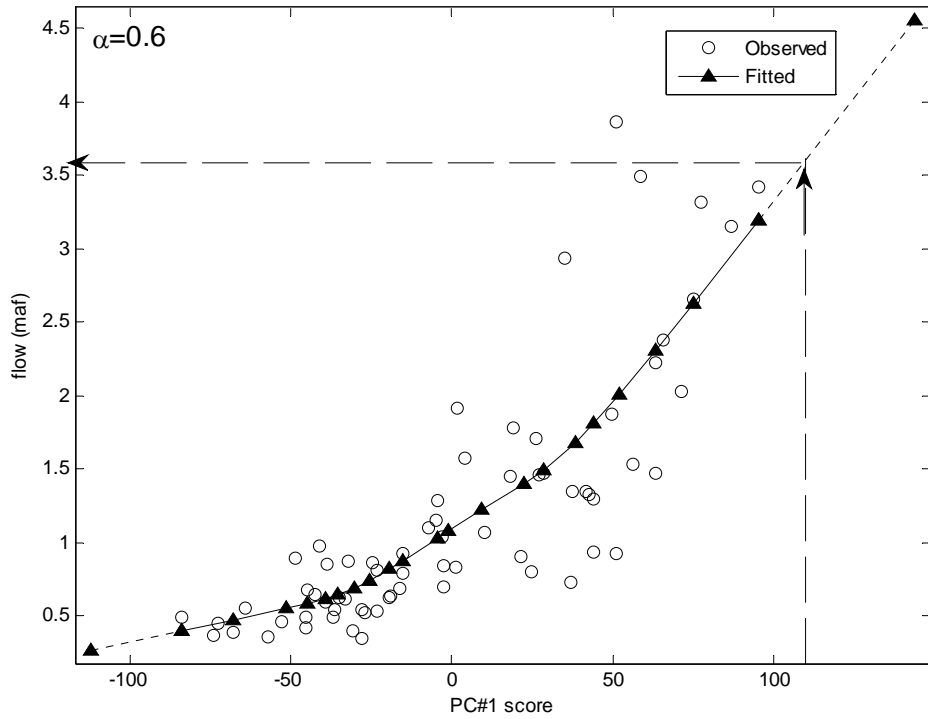


Figure 10. Loess curve used for interpolating final reconstruction. Dotted line is linear extension to allow interpolation outside the range of data in the 1914-1982 loess calibration period. Arrows illustrate using the loess curve to infer a flow of 3.6 maf from a PC#1 score of 110.

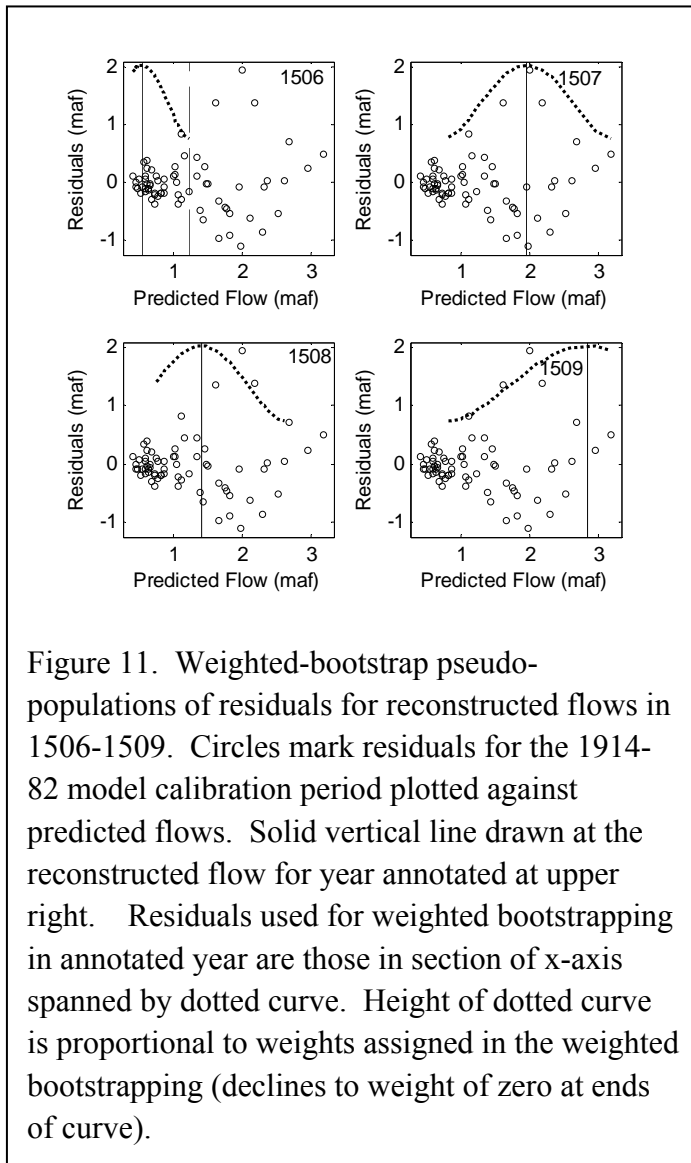
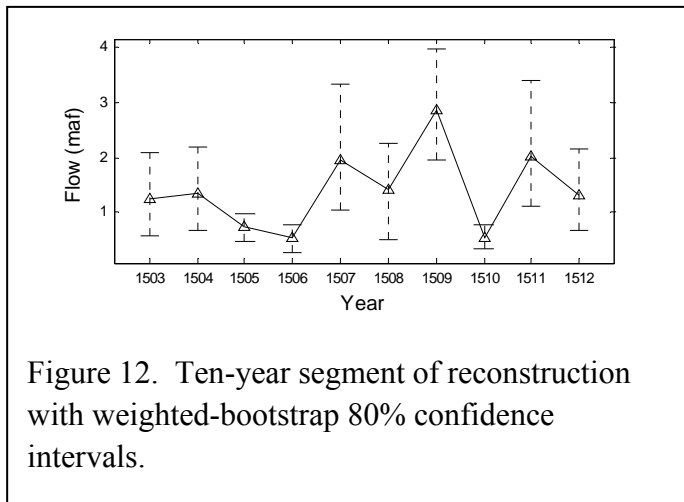


Figure 11. Weighted-bootstrap pseudo-populations of residuals for reconstructed flows in 1506-1509. Circles mark residuals for the 1914-82 model calibration period plotted against predicted flows. Solid vertical line drawn at the reconstructed flow for year annotated at upper right. Residuals used for weighted bootstrapping in annotated year are those in section of x-axis spanned by dotted curve. Height of dotted curve is proportional to weights assigned in the weighted bootstrapping (declines to weight of zero at ends of curve).



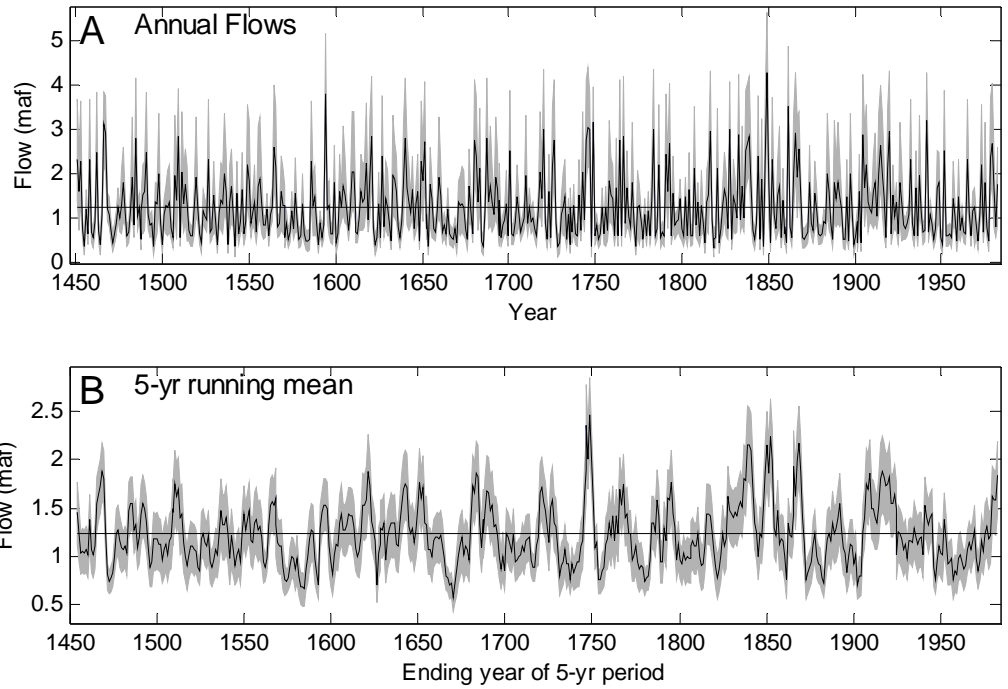


Figure 13. Time series plots of annual reconstructed flows with 80 percent confidence interval. (Top) Annual flows, 1541-1982. (Bottom) Five-year running mean of annual flows. Horizontal line in both plots is the 1541-1982 mean of annual reconstruction.

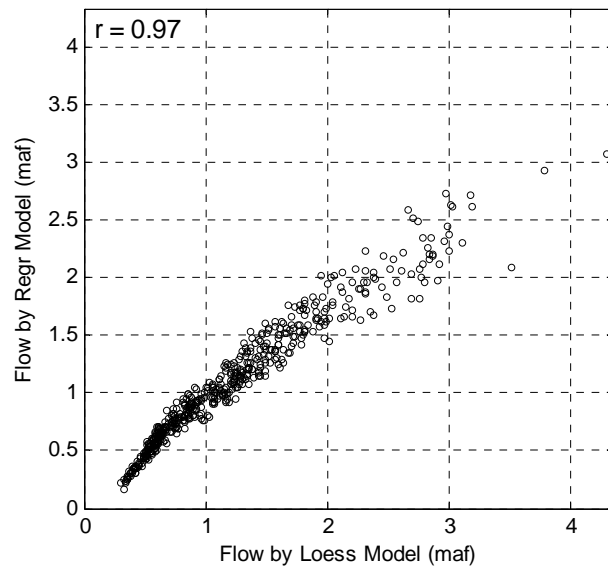


Figure 14. Scatterplot of reconstructed flows by linear regression model on reconstructed flows by loess model, 1451-1982.

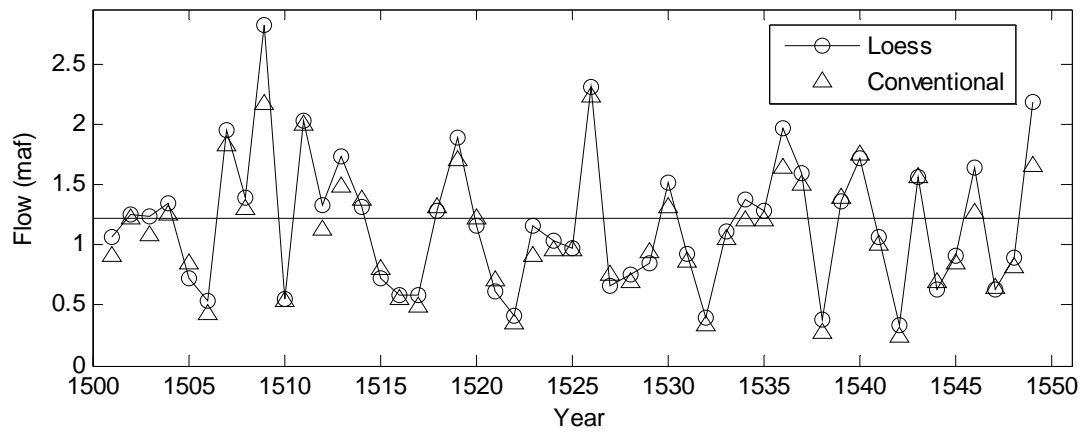


Figure 15. Fifty-year time series segment comparing annual reconstructed flows by loess model with those by conventional model. Conventional model is linear regression of flow on scores of first principal component of the 10 residual tree-ring chronologies.

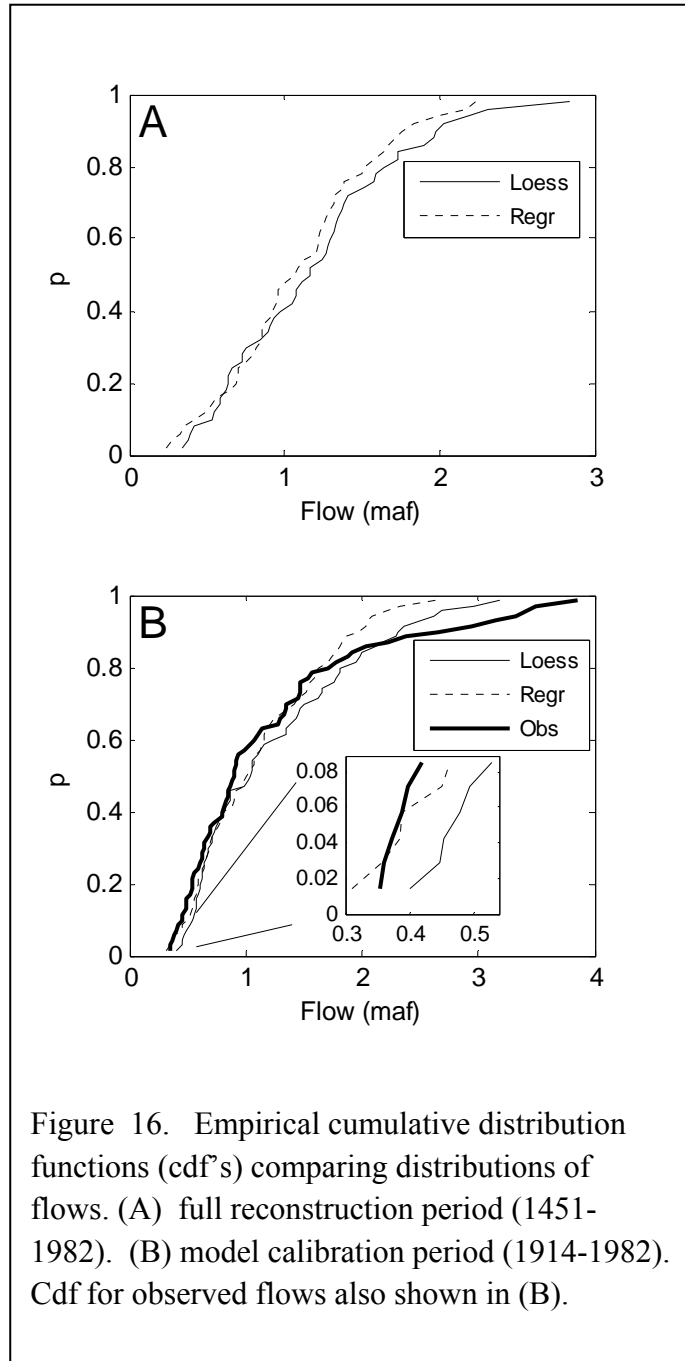


Figure 16. Empirical cumulative distribution functions (cdf's) comparing distributions of flows. (A) full reconstruction period (1451-1982). (B) model calibration period (1914-1982). Cdf for observed flows also shown in (B).

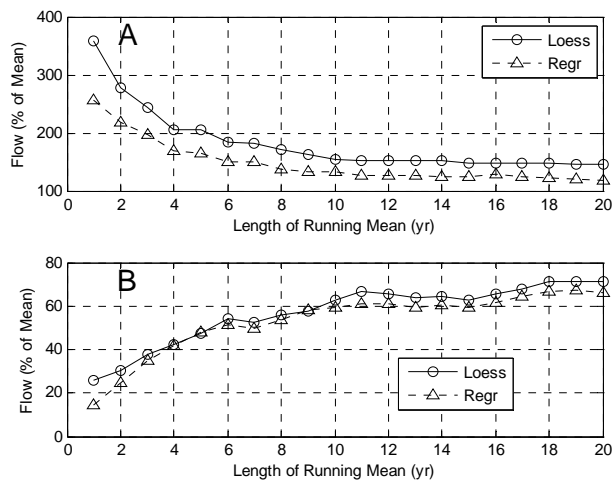


Figure 17. Extreme running means of flow reconstructed by loess curve and regression. (A) Highest running means of length 1-20 years. (B) Lowest running means of length 1-20 years. Analysis period 1541-1982.