

Updated Streamflow Reconstructions for the Upper Colorado River Basin

Connie A. Woodhouse
NOAA National Climatic Data Center, Boulder CO

Stephen T. Gray
U.S. Geological Survey, Desert Laboratory, Tucson AZ

David M. Meko
University of Arizona, Tucson AZ

Accepted for publication in Water Resource Research.
Copyright 2006 American Geophysical Union.
Further reproduction or electronic distribution is not permitted.

Abstract

Updated proxy reconstructions of water year (October-September) streamflow for four key gages in the upper Colorado River basin were generated using an expanded tree-ring network and longer calibration records than in previous efforts. Reconstructed gages include the Green River at Green River, Utah; Colorado near Cisco, Utah; San Juan near Bluff, Utah; and Colorado at Lees Ferry, Arizona. The reconstructions explain 72-81% of the variance in the gage records, and results are robust across several reconstruction approaches. Time-series plots as well as results of cross-spectral analysis indicate strong spatial coherence in runoff variations across the sub-basins. The Lees Ferry reconstruction suggests a higher long-term mean than previous reconstructions, but strongly supports earlier findings that Colorado River allocations were based on one of the wettest periods in the past five centuries, and that droughts more severe than any 20th-21st century event occurred in the past.

Index terms: 1812 Drought, 1860 Streamflow, 1884 Water supply, 4920 Dendrochronology, 344 Paleoclimatology

1. Introduction

The Colorado River, perhaps the most important regional source of surface water supply in the western United States, was the subject of the first tree-ring based effort aimed at the quantitative reconstruction of streamflow records (Stockton and Jacoby 1976). The reconstruction of annual flows at Lees Ferry, which reflects conditions in the entire Upper Colorado River basin (Figure 1), contained several noteworthy features. The highest sustained flows in the entire record, 1520 to 1961, occurred in the early decades of the 20th century, a period that coincides with the negotiation of the 1922 Colorado River Compact and the resulting allocation of Colorado River flows. In effect, water that was not likely to be in the river on a consistent basis was divided among the basin states. In addition, the most persistent and severe drought occurred in the late 16th century, with flows during this period much lower than for any event in the 20th century.

Two decades later, this landmark reconstruction was the basis for a series of studies that investigated the hydrologic, social, and economic impacts of a severe sustained drought in the Colorado River basin (Young 1995). These studies indicated that under the current Law of the River (the set of legal compacts and regulations that govern the Colorado River), a drought like the 16th century event in Stockton and Jacoby's record would greatly challenge the capacity of the Colorado River to meet water supply needs, and have significant impacts on Compact obligations.

Severe drought conditions in the Colorado River basin, coupled with a large increase in water use over the past two decades, have recently resulted in water demands that have outstripped

natural inflows (Fulp 2005). Moreover, new water projects, additional management concerns such as endangered species, and large increases in population have altered the potential impacts of drought. These conditions have reinvigorated interest in reconstructions of Colorado River flow. Stockton and Jacoby's (1976) original Lees Ferry reconstruction ended in 1961, which has made it difficult to assess recent droughts in a long-term context. In addition, reconstruction methods have evolved greatly in recent decades. Hidalgo et al. (2000) have shown that features of the Stockton and Jacoby reconstruction, including relative drought severity and duration, are sensitive to modeling methodology. Thirty additional years of gage data, new and updated tree-ring collections, and improved methodologies now enable a longer and more robust reconstruction of Colorado River streamflow. The purpose of this paper is to describe and analyze a recently-generated set of updated streamflow reconstructions for Lees Ferry and other key gages in the Upper Colorado River Basin.

2. Data and Methods for Reconstructions

2.1. Streamflow Data

We selected four gages in the Upper Colorado River basin for reconstruction: the Green River at Green River, Utah; Colorado River near Cisco, Utah; San Juan River near Bluff, Utah; and Colorado River at Lees Ferry, Arizona. The selected gages represent flows in the three major sub-basins as well as the total flow of the Upper Colorado Basin (Figure 1). The U.S. Bureau of Reclamation provided estimates of natural flows for these locations that span the years 1906 to 1995 (Jim Prairie, personal communication). These flow values have been adjusted to account

for human impacts through a combination of statistical and expert-system approaches, but the records may still include some anthropogenic signals. Water year (October-September) flow data in millions of cubic meters (MCM) were examined graphically and statistically to assess variability, normality and the degree of persistence in the time series (Table 1). The water year flows are essentially normal, and all display a small amount of persistence at a lag of one year. The San Juan represents a considerably more arid region than the other two basins, as evidenced by the lower mean annual flow and higher coefficient of variation. The San Juan is also the only sub-basin for which the first-order autocorrelation is not significantly greater than zero.

2.2. Tree-Ring Chronology Network

In much of the western U.S., tree-ring widths can provide a proxy for gage records because the same climatic factors, primarily precipitation and evapotranspiration, control both the growth of moisture-limited trees and processes related to streamflow (Meko et al. 1995). Recent collections of new tree-ring data and efforts to update older collections have produced a set of 62 moisture-sensitive tree-ring chronologies in Colorado, southwestern Wyoming, and northeastern Utah that span the common interval from 1600 to 1997 (Figure 1 and Appendix 1, www.ncdc.noaa.gov/paleo/pubs/woodhouse2006/woodhouse2006.html) . Of the 62 chronologies, 17 are from ponderosa pine (*Pinus ponderosa*), 21 from Douglas-fir (*Pseudotsuga menziesii*), 21 from pinyon pine (*Pinus edulis*), and three from limber pine (*Pinus flexilis*). Fifteen or more trees were typically sampled at each site using an increment borer and taking two cores from each tree. In the lab, cores were processed, crossdated, and measured using standard dendrochronological techniques (Stokes and Smiley 1968, Swetnam et al. 1985). All ring-width

series were uniformly processed using the ARSTAN program as follows (Cook 1985). Measured series were standardized using conservative detrending methods (negative exponential/straight line fit or a cubic spline two thirds the length of the series) before using a robust-weighted mean to combine all series into a single site chronology (Cook et al. 1990). Low-order autocorrelation in the chronologies that may, in part, be attributed to biological factors (Fritts 1976) was removed, and the resulting residual chronologies were used in most of the subsequent analyses. However, the low-order autocorrelation in the gage records was closely matched by persistence in the tree-ring data. Consequently, the sensitivity to persistence in the tree-ring data was tested for in the Lees Ferry reconstruction by generating reconstruction models using both the standard (persistence retained) and prewhitened (persistence removed) chronologies. Because the number of series in these chronologies decreases with time, chronologies in the resulting reconstruction models were assessed with regard to subsample signal strength (Wigley et al. 1984).

Statistical analyses support the high quality and suitability of these chronologies for hydroclimatic reconstructions (online Appendix 1). The mean inter-series correlation within each chronology averages 0.79, and mean sensitivity (average relative ring width difference from one ring to the next, Fritts 1976) averages 0.41. These statistics indicate the strong common signal between the trees that make up each chronology and the high degree of variability in ring widths from one year to the next. Both characteristics are consistent with strong tree-ring sensitivity to climatic variability (Cook and Briffa 1990).

2.3. Reconstruction Approaches

Multiple linear regression, with predictors entered forward stepwise (Weisberg 1985), was used to generate the reconstruction models. In an automated process such as stepwise regression, increasing the size of the potential predictor pool also increases the likelihood of a meaningless predictor entering the model by chance alone (Rencher and Pun 1980). To assess the sensitivity of the reconstruction to the size and makeup of the predictor pools, two alternative reconstruction approaches were tested for each gage. First, the “full pool” approach used all chronologies significantly correlated ($p < 0.05$) with the gage record as potential predictors. Correlations were evaluated over the entire gage period (1906-1995) and over both early (1906-1950) and late (1951-1995) sets of years to ensure the stability of the correlation. A second approach, a “watershed-limited” approach, followed the same correlation rules, but the potential predictor set was restricted to chronologies within a 100 kilometer buffer around the watershed upstream from the gage.

Reduction of the predictor pool by a watershed boundary constraint was not feasible for the Lees Ferry gage, as the watershed essentially encompasses all chronologies. The approach taken for that gage was to reduce the predictor pool by principal components analysis (PCA). After first removing chronologies uncorrelated with Lees Ferry streamflow, a PCA was run on the correlation matrix of the chronologies for their full common period of overlap. Mardia et al. (1979, p. 244) suggest that in a regression context, the components having the largest correlations with the predictand, rather than the components with the largest variances, are best

suited for retention. Accordingly, only those components significantly ($p < 0.05$) correlated with streamflow were retained in the pool of potential predictors. The resulting pool has essentially been reduced to concisely express orthogonal modes of common variation in the tree-ring data. Because each component is a linear combination of all tree-ring chronologies correlated with streamflow, the PCA approach is relatively robust to nonclimatic influences (e.g. disturbance, insect outbreaks) at individual sites. For the Lees Ferry reconstruction, model sensitivity to the use of the standard versus the prewhitened chronologies was tested for both the non-PCA and PCA approaches described above. Validation statistics and features of the reconstructed time series were compared to assess sensitivity of results to the alternative model formulations.

The strength of the regression models was summarized by the adjusted R^2 and F -level of the regression equation (Weisberg 1985). Possible multicollinearity of predictors was assessed with the Variance Inflation Factor (VIF, Haan 2002). A forward stepwise approach was used to enter predictors from the predictor pools, with threshold F -values for entry or removal of predictors. Variables were entered in order of their explained residual variance. As a guide, the F -level for a predictor was allowed to have a maximum p -value of 0.05 for entry and 0.10 for retention in the equation. Residuals for all regression models were inspected graphically for non-normality, trend, autocorrelation, and obvious dependence on values of the predictors or predicted flows. Any of these conditions could indicate a need for data transformation. Residuals were tested for normality with the Lilliefors test (Conover 1980).

As a safeguard against model overfitting, the entry of predictors was terminated when it resulted in decreased validation accuracy. The reduction of error (RE; Fritts et al. 1990) and root mean

squared error (RMSE; Weisberg 1985) were generated using two different calibration/validation schemes. In one scheme, a stepwise model was first fit to the full calibration period, recording the order of entry of predictors. The model was then fit to the first half of the data using the same predetermined order of entry for the predictors, and validated on the second half of the data. The calibration and validation halves were then exchanged and the process repeated. In the other validation scheme, leave-one-out cross-validation (Michaelsen 1987) was used to generate a single validation series. In both schemes, the RE and RMSE were calculated for each step and plotted to assess when the validation scores stopped improving. One last method of validation involved using the predictors selected by the stepwise regression process to run a linear neural network (LNN). LNN is an iterative model fitting process based on statistical bootstrapping techniques that was used here to assess bias in the explained variance. If the relationship between tree growth and climate is robust and stable, the results of LNN and stepwise regression should be equivalent (Goodman 1996, Woodhouse 1999).

3. Reconstructions

3.1. Full Pool Stepwise Regression Model Results

Statistics for the initial full pool stepwise regression results using residual chronologies as predictors are listed in Table 2 in the first three lines under full pool models (sub-basins) and the first line under the Lees Ferry models. The regression models all have highly significant F-levels, account for between 72 percent and 81 percent of the variance of flow, and possess significant skill when applied to cross-validation testing. The predictor pools for the models

contain between 24 and 38 chronologies, but the stepwise selection yields four to seven predictor chronologies in the final models.

The residuals analysis indicated that normality of residuals could not be rejected (Lilliefors test, $p < 0.05$) for any of the series. Residuals for one gage, Colorado-Cisco, showed borderline significance of autocorrelation at a one-year lag. For three of the four gages, residuals had a significant ($p < 0.05$) downward trend, suggesting greater tree growth than expected from flow in recent decades. A scatterplot indicated that the variance of residuals increased with the predicted values for the Colorado-Cisco. As neither square-root nor log10 transformation of the flow record offered more than marginal improvement for that gage, the decision was made to use the untransformed flows. The stepwise validation results indicated that strict adherence to the F -enter and F -remove criteria did not result in any obvious overfitting of the models. The numbers of steps and predictors chronologies for the four gages are listed in Table 2. Regression coefficients are listed Appendix 2

(www.ncdc.noaa.gov/paleo/pubs/woodhouse2006/woodhouse2006.html). Linear neural networks using the suite of predictors included in the regression equations yielded explained variance values that were the same as those from the regression approaches. An example of the comparison between a gage record and a reconstruction is shown for Lees Ferry in Figure 2.

3.2 Sensitivity of the Reconstruction Models to Predictor Pool

The predictor pool sensitivity tests apply only to the gages on the Colorado-Cisco and San Juan,

as the same predictors were selected from both pools for the Green River gage. Limiting the pool by watershed boundary reduced the number of potential predictors from 38 to 32 chronologies for the Colorado-Cisco and from 24 to 8 chronologies for the San Juan (online Appendix 2). Stepwise regression for the Colorado-Cisco and the San Juan gage yielded two and three predictors, respectively, in the full-pool regression equation that were not in the limited-pool equation. In these cases, as expected, the explained variance is reduced in the watershed-limited models. To address reconstruction sensitivity, reconstructions based on full pool and limited pools of predictors were compared with attention to critical pre-calibration periods, such as the well-known drought on the late 16th century (Stockton and Jacoby 1976; Gray et al. 2004; Stahle et al. 2000). A comparison of reconstructions for the San Juan and Colorado-Cisco from the two different models indicates only slight differences, particularly during periods of drought (Figure 3). On consideration of calibration/validation accuracy, the relative insensitivity of reconstructions to predictor-pool reduction, and ability to reproduce statistical features of the observed record, we decided to adopt the full-pool predictor subsets for the final sub-basin reconstructions and analysis.

3.3. Sensitivity of the Lees Ferry Reconstruction to Modeling Approaches

Reconstructions of Lees Ferry streamflow were tested using four different forms of the predictor tree-ring data: residual chronologies (Lees-A, described in section 3.1), standard chronologies (Lees-B), principal components of residual chronologies (Lees-C), and principal components of standard chronologies (Lees-D). Exploratory analysis suggested 1490 as a reasonable start year for the reconstructions; of the original 62 chronologies, 31 residual chronologies and 30 standard

showed significant correlations with annual streamflow and passed the screening test for time coverage to at least 1490 (Table 2). Stepwise regression on the standard chronologies (Table 2, Lees-B) yielded a reconstruction model with the same number of predictors (7) as for the residual-chronology version, and a slight increase in F-level and variance explained by regression (see Figure 1 for locations of predictor chronologies, and online Appendix 2 for regression coefficients).

The PCA indicated that the residual chronologies have somewhat more spatial structure than the standard chronologies (Appendix 3, www.ncdc.noaa.gov/paleo/pubs/woodhouse2006/woodhouse2006.html). PC 1 is by far the most important component, accounting for 47 percent of the variance of the residual chronologies and 45 percent of the variance of the standard chronologies. For both sets of data, five PCs have eigenvalues exceeding 1.0, and these PCs account for a cumulative 69 percent (residual chronologies) and 68 percent (standard chronologies) of the tree-ring variance.

PC loadings on all chronologies are positive for PC 1 whether the PCA is on residual or standard chronologies. This pattern attests to the strong over-riding common signal in tree growth over the Upper Colorado Basin. There appears to be some species dependence, with highest weights on *Pinus edulis* chronologies. Spatial organization in PCs 2-5 is most obvious for the residual chronologies: maps of loadings (not shown) indicate an east-west contrast in PC 2, a north-south contrast in PC 3, and spatial clustering in PCs 4 and 5.

The predictor pools, based on significant correlation of PCs with streamflow, were PCs 1, 15 and

16 for the residual chronologies, and PCs 1, 17, 28, and 29 for the standard chronologies. Except for PC 1, a high percentage of tree-ring variance accounted for by a PC did not imply strong correlation with streamflow.

In the stepwise procedure for both the residual and standard chronologies, only PC 1 entered as a predictor of flow (Table 2, and online Appendix 2). The final models (Table 2, Lees-C and Lees-D) account for seven to nine percent less variance of flow than the corresponding non-PCA models but, with just one predictor variable, have considerably higher F-levels. Both PCA models verify well as indicated by the high cross-validation RE statistics (Table 2). We repeated the PCA regression exercise with predictor pools made up of the PCs 1-5, rather than PCs screened by correlation with flow, and arrived at the same results -- a final model with just PC 1 as the predictor.

Descriptive statistics for the observed flows and the four alternative Lees Ferry reconstructions for the 1906-95 calibration period are listed in Table 3. For the calibration period, the reconstructed and observed means are forced to be equal by the regression process, and differences in standard deviation simply reflect differences in proportion of variance explained by regression. The skew for all four reconstructions is opposite in sign from that of the observed flows, but given the short sample provided by the calibration period, only the skewness of Lees-C is significantly different from zero at $\alpha=0.05$ (Snedecor and Cochran 1989). On the basis of Lilliefors test (Conover 1980) the assumption of normality could not be rejected for any of the four reconstructions ($\alpha=0.05$). A large contrast is seen in first-order autocorrelation of the two reconstructions based on residual chronologies versus the reconstructions using standard chronologies. The reconstructions by residual chronologies have essentially no first-order

autocorrelation, while the observed flows and the reconstructions by standard chronologies are significantly positively autocorrelated ($p < 0.01$, one-tailed test).

Annual observed flows range from 37% to 166% of the 1906-1995 mean. In general, for any reconstruction we expect departures from the calibration period mean to be underestimated due to compression of variance in regression modeling, but in Table 3 the lowest annual flows in all four reconstructions are lower than the lowest observed flow. This unexpected result might be due to the exaggerated negative skew of the reconstructions. In contrast, no reconstructed flow is as high as the highest observed flow. The five-year running means are as expected, with neither highs nor lows as extreme as in the observed data. As expected when using residual chronologies, the 20-year running means are conservative, and the lows appear to be exaggerated by the reconstructions based on standard chronologies (Table 3).

The four time series of smoothed full-length (1490-1997) Lees Ferry reconstructions track one another closely (Figure 4). All reconstructions indicate a long-term mean flow below the 1906-1995 observed mean. The long-term reconstructed mean ranges from 94.0 percent to 96.5 percent of the observed mean, and so is relatively insensitive to choice of model. If the standard error of an m -year mean of reconstructed values is assumed to be $1/\sqrt{m}$ times the root-mean-square error of the annual reconstructed values (Table 2) and the errors are normally distributed, all four reconstructed means are significantly ($\alpha = 0.05$) different than the observed mean.

Depending on reconstruction model, the long-term standard deviation is greater than (non-PCA models) or less than (PCA models) the standard deviation of observed flows (Table 4). If

climate were equally variable before and during the calibration period, compression of the variance in regression would tend to yield a long-term reconstruction with lower variance than that of the observed flows. The greater standard deviation for the non-PCA models implies more variable climate before the start of the calibration period than after. All four reconstructions are negatively skewed, but the assumption of zero skew can be rejected ($p < 0.05$, $N = 508$) only for the reconstructions from the residual chronologies (Table 4).

Differences in first-order autocorrelation among models were noted for the 1906-1995 calibration period (Table 3), and those differences also apply to the long-term reconstructions (Table 4). A comparison of first-order autocorrelations of reconstructed data for the full reconstruction and the calibration period suggests the autocorrelation in the calibration period is representative of the long-term record. It is also evident, however, that the autocorrelation of the reconstructed flows from residual chronologies is biased low relative to that of the observed flows (Table 3). The impact of the disparity in first-order autocorrelations for model Lees-A was investigated by restoring the persistence to the reconstructed flow with an autoregressive model, and comparing the original reconstruction with the persistence-restored reconstruction. The two series were extremely similar, and although two-year droughts were slightly less common and three-year slightly more common in the persistence-restored reconstruction, there were no distinct differences for longer droughts. The reconstructions from standard chronologies more accurately reflect the first-order autocorrelation of the observed record (Table 3). Model Lees-D is perhaps strongest in this regard because the reconstructed flows are slightly more autocorrelated than the observed flows. This is reasonable because the reconstruction errors are assumed to not be autocorrelated.

Extreme n -year running means are quite similar for the alternative Lees Ferry reconstructions, but somewhat more extreme for the reconstructions using the standard chronologies (Table 4). Regardless of model, the lowest 1-year, 5-year and 20-year means for the full reconstructions are much below those in the observed flows. The lowest reconstructed 20-year means for all models are in the late 1500s (Figure 4; note that this drought is somewhat more severe in the standard chronology PCA model). In the standard chronology models, the highest reconstructed n -year means exceed those in the observed record, with the exception of 5-year means. Smoothed time series of the four reconstructions are in agreement in the exceptional wetness of the early 1900s (Figure 4). The implication is that a period of such sustained wetness had not occurred since the start of the 1600s.

In summary, the above comparison shows that key features of the updated flow reconstructions for Lees Ferry are fairly robust to modeling choices. The models using the standard chronologies appear to more closely match the persistence in the gage record, and the non-PCA version using standard chronologies (Lees-B) has the greatest calibration period accuracy as measured by regression R^2 . On the other hand, the models based on standard chronologies overestimate the severity of multi-decadal droughts (20-year means) in the calibration period, which is worrisome considering that the regression procedure itself tends to compress reconstructed values toward the calibration period mean. Smoothed time series plots (Figure 4) suggest the PCA reconstruction on standard chronologies (Lees-D) is somewhat of an outlier, and gives a worst-case scenario for the severity of extended droughts and wet periods. In view of the fact that the sub-basin gages were reconstructed using the residual chronologies and a non-PCA approach, for

consistency of analysis we used the reconstruction version Lees-A as the baseline record in the sub-basin analysis that follows.

3.4. Spatial Fidelity among Gages and Reconstructions

The relationship between gages within the Upper Colorado River basin, and how those relationships were preserved in the reconstructions, was evaluated by examining the shared variance between the set of gage records and the set of reconstructions. Spatial relationships were then examined with regard to the magnitudes of flow from the sub-basins and their relationship to the total Colorado River flow at Lees Ferry.

In the gage records, all flows are highly correlated ($r > 0.77$) except between the San Juan and the Green ($r = 0.55$), the most widely separated gages (Table 5a). In the reconstructed flow records for the same time period (1906-1995), the same relationships are preserved between the Green, Colorado-Cisco, and Lees Ferry reconstructions. Correlations between the San Juan and the other reconstructions are somewhat inflated, particularly between the Green and San Juan (Table 5a). The relationships for the full reconstructions are quite similar to those for the period of the gage records (Table 5b). The greater shared variance between the San Juan reconstruction and the other reconstructions, compared to the relationships in the gage record, may be due to an absence of tree-ring chronologies located within the San Juan River basin. The result of this may be a weaker representation of local basin variability.

The observed flows from the three sub-basins gages, Green River, Colorado-Cisco and the San

Juan account for nearly all (average of 95.5%) of the total water year flow observed at Lees Ferry from 1906-1995 (Table 6). Over the same years, the average values of contributed flows in the reconstructions are closely matched, as expected due to the regression process. Over the full common reconstruction period, 1569-1997, contributions are also very similar (Table 6). Figure 5 shows the variations in flow at the four gages and the sum of the three sub-basin flows over the full reconstruction period as 5-yr running averages. The match between the three gage sum and the Lees Ferry flow is good ($r = 0.98$, $p < 0.001$), though there are several periods when the sum appears to be somewhat less than Lees Ferry total flow (e.g., the 1630s and the last quarter of the 1600s, both periods of higher flows).

4. Long-term Hydroclimatic Variability in the Upper Colorado River Basin

4.1. Frequency Characteristics of Reconstructed Flows

We used a multi-taper method (MTM) spectral analysis to examine the frequency characteristics of reconstructed flows at Lees Ferry and the three sub-basins gages (Mann and Lees, 1996).

MTM provides a robust means for isolating signal peaks from a time series that may contain both periodic and aperiodic behavior. The MTM spectrum for the Lees Ferry reconstruction (Figure 6a) shows that significant ($p < 0.05$) high-frequency variability in Upper Colorado River flows (2-7 yr) is accompanied by a strong bi-decadal peak centered around ~24 yr. MTM also identifies a significant multidecadal peak around 64 yr. Peaks similar to those in the two to seven year band at Lees Ferry are also present in the spectra for each sub-basin (Figure 6b-d). All of the sub-basin reconstructions show significant bi-decadal peaks, though relative power is

reduced for the Green River gage. The reconstructions for both the Colorado-Cisco and the San Juan show strong multidecadal peaks centered on ~ 64 yr. Cross-spectral MTM reveals significant coherency across the sub-basins at lower frequencies (Figure 7). Coherency and phasing of bi-decadal and multi-decadal peaks is particularly strong.

The wavelet spectra for each of these reconstructed gage records further highlights their coherence in the frequency domain (Figure 8). Wavelet analysis also shows marked non-stationarity in the strength of these signals through time. In particular each of the wavelet spectra are characterized by multidecadal variability (30-70 yr) in the first two centuries followed by a period from the 18th through mid-19th centuries dominated by significant energy in the decadal to bidecadal bands. Beginning in the late 19th century, however, we see a return to significant multidecadal variability. These lower-frequency modes persist until the late 20th century, when the effects of zero padding likely reduce power in the multidecadal bands (Torrence and Compo, 1998).

4.2. Basin-Scale Flow Variability

The Lees Ferry and sub-basin streamflow reconstructions enable an examination of the spatial characteristics of long-term drought variability in the upper Colorado River basin. We first compared 5-year, 10-year, and 20-year averages of streamflow in the Lees Ferry reconstruction with averaged flows in the three sub-basins to determine the degree of drought variability across the upper Colorado River basin. In general, there is a strong tendency for extreme low flows at Lees Ferry to be matched by extreme low flows in all three of the sub-basins. Of the driest 5-

year periods (lowest 15% of flows) in the Lees Ferry reconstruction, none ranked above the driest tercile in the Colorado-Cisco reconstruction, which accounts for the greatest proportion of Lees Ferry flow. Two of driest Lees Ferry flow periods ranked in the middle tercile in the Green River reconstruction record (1728-1732, 1628-1632). There were nine periods in the San Juan reconstruction that fell within the middle tercile that were dry periods in the Lees Ferry record. Four of these periods occurred in the 1580s and 1590s, which is known regionally as an extreme drought throughout the western U.S. (e.g., Stahle et al. 2000). While there were some extremely dry years in the San Juan reconstruction over this period (e.g., 1590), this period was also marked by several wet years (e.g., 1589, 1595, 1599).

Important regional variations do exist within extreme dry periods (Table 7). Rankings of 5-year averages show that the driest 5-year period in the Lees Ferry record, 1844-1848, was extremely dry in the Green and Colorado-Cisco records (driest and third driest, respectively), but was somewhat less extreme in the San Juan (17th driest). The second most extreme 5-year low-flow period in the Lees record, 1622-26, was similarly dry in the Colorado-Cisco and San Juan records (second driest and driest, respectively), but to a much lesser extent in the Green (63rd driest). Regional variability in extreme low flows is also evident over longer time scales. The period 1622-1631 was the driest ten-year period in the Lees Ferry reconstruction. As in the five-year periods, low flows in 1620s are less extreme in the Green River record, but are markedly low in both the San Juan and Colorado-Cisco records (ranks 71st, third, and sixth, respectively). In contrast, the Green River appears to be most strongly impacted by decadal-scale droughts in the 1870s and 1880s. As suggested above, the San Juan appears to be less sensitive to the low flows in the 1580s and 1590s, and this is evident at both 10-yr and 20-year time scales. The 20-

year period ending in 1592 is the driest such period in the Lees Ferry and Colorado-Cisco reconstructions, and the sixth driest in the Green reconstruction, but it was the 48th driest period in the San Juan reconstruction.

Regional drought variability was also examined in the context of its impacts on Lees Ferry flows. Rankings for ten-year moving averages of flow in the three sub-basins were divided into terciles. Periods when the value for one basin fell in the dry tercile while flow in another basin fell into the wet tercile, were tabulated (Table 8). Again, droughts tend to be widespread, affecting, to some degree, all three sub-basins simultaneously. However, in 15 of these ten-year periods, contrasting conditions exist between two basins. Most commonly (eight periods), high flows in the San Juan reconstruction coincide with low flows in the Green reconstruction. Dry conditions in the San Juan and wet in the Green are far less common (three periods). In two periods, the Green is dry while the Colorado-Cisco is wet, and there is one case each when the San Juan is wet and Colorado-Cisco dry and vice versa. The contrasting conditions in the pairs of sub-basins appear to balance each other with respect to Lees Ferry flow for the most part, with Lees Ferry flow for these periods most often falling in the middle tercile. However, in four of the eight periods when Green River is low and the San Juan river is high, flow at Lees Ferry is in the driest tercile, and except for one of the eight periods, Lees Ferry flow is lower than the median. This suggests that low flow conditions in the Green River can over-ride wet conditions in the San Juan, and to some extent, moderate conditions in the Colorado-Cisco, to influence Lees Ferry flows. Greater sensitivity of Lees Ferry flow to the Green than to the San Juan is of course expected, given the much larger percentage of flow contribution from the Green (Table 1 and Table 6).

As shown in the spectral analysis (section 4.1), streamflow at Lees Ferry and the three sub-basins also varies significantly over multidecadal timescales. To highlight this lower-frequency variability, each of the reconstructions was smoothed with a 50-year cubic spline (Figure 9). The smoothed time series display a pattern of high magnitude variations in the 16th and 17th centuries and the 19th and 20th centuries, with dampened variability centered on the 18th century. The driest multi-decadal period in the Lees Ferry reconstruction occurs in the late-16th century. The low-flow period at the end of the 19th century shares a similar magnitude. In this multidecadal context the 1950s drought is also notable as the 4th lowest flow period at Lees Ferry. Generally high-flow regimes occurred across the basin in the early 17th and early 20th centuries. The most recent decades of the reconstruction were also quite wet. As in the case of the multi-year and decadal flow-regimes discussed above, the magnitude of departures for these multidecadal flow regimes varies somewhat across the basin. This is particularly true for the early 1700s through the mid 1800s, which is the period when the wavelet analyses show a significant loss of multidecadal power in the basin. However, the timing and duration of multidecadal flow regimes is markedly coherent across the Upper Colorado River Basin.

4.3. Comparison with Previous Lees Ferry Reconstructions

Because of the central importance of the Lees Ferry record to the allocation of Colorado River water supply, it is important that the reconstruction be as accurate as possible, and that the uncertainty be appreciated. The discussion in section 3.3 dealt with uncertainty due to modeling choices: the use of standard versus residual chronologies and the decision to use individual

chronologies or chronologies reduced by PCA in the regressions. Previous reconstruction efforts (Stockton and Jacoby 1976; Hidalgo et al. 2000) not only used different modeling procedures from ours, but also a different tree-ring network and a much shorter calibration period. In this section we compare our updated reconstructions, versions Lees-A and Lees-D, with reconstructions by Stockton and Jacoby (1976) and Hidalgo et al. (2000). We refer to these two previous reconstructions as SJ1976 and HDP2000. The comparison focuses on two statistics: the long-term mean annual flow, and the most severe sustained drought as measured by the lowest reconstructed 20-year moving average of flow. Lees-A is our model using regression of flow on residual chronologies. Lees-D is our model using regression of flow on PCs of standard chronologies. Those two versions were selected for the comparison because they represent the most conservative (wettest) and least conservative (driest) of the alternative reconstructions from the updated chronologies (see section 3.3).

Time series plots of smoothed reconstructions (Figure 10) generally agree in timing of highs and lows, but disagree considerably on the magnitude of some flow anomalies. The plots for the updated reconstruction generally show wetter conditions than the previous reconstructions. HDP2000 represents the driest scenario, with greatly amplified low-flow features in the late 1500s, late 1700s and near 1900. Much less disagreement among the four reconstructions is evident in the calibration period than in the pre-calibration period.

Selected calibration and reconstruction statistics for the four models are listed in Table 9. Flow statistics are given in units of both billion cubic meters (BCM) and million acre-feet (MAF) to facilitate comparison with previous published studies. Note that the reconstructions differ

considerably in calibration period as well as in the number of tree-ring chronologies on which the final reconstructions depend. Agreement of the reconstructions in the calibration period (Figure 10) is not surprising as all four models have high R^2 values (Table 9). Perhaps the most striking disagreement in the models is the magnitude of the late-1500s drought (the period of the lowest 20-year mean), which is estimated at 11.2 BCM (9.1 MAF) by HDP2000 and 15.6 BCM (12.6 MAF) by Lees-A. The updated reconstructions suggest the long-term mean annual flow is not a low as previously estimated. Our driest updated reconstruction model (Lees-D) gives a long-term mean of 17.6 BCM (14.3 MAF), which is some 0.9 BCM (0.8 MAF) higher than the original estimate by Stockton and Jacoby (1976).

Differences in the reconstructions are undoubtedly related to differences in the basic data and the statistical models used for reconstruction. The most obvious data difference between this and past efforts would be that different chronologies were used as predictors. Two of Stockton and Jacoby's (1976) original sites were recollected, but it is not evident that any of the same trees were sampled. Gage data were different as well, and both the tree-ring data and gage data in previous efforts resulted in a calibration period nearly half the length of the calibration period used in this study. Differences could also result from data processing and decisions in detrending the raw ring widths. SJ1976 and HDP2000 used standard chronologies, and models with PCs of lagged chronologies as predictors. Over the common period 1906-1961, the SJ1976 reconstruction showed a lag-1 autocorrelation of 0.36 and HDP2000 0.41, which are similar to the 0.22 and 0.31 for the models from this study that were based on standard chronologies. All of these lag-1 values are also consistent with values for the gage record (0.25). The inclusion of lagged predictors may have the effect of enhancing the persistence in the extreme low flow

years. The reliance on just seven tree-ring chronologies to sample the runoff variations over the entire Upper Colorado River Basin – as with updated reconstruction Lees-A -- might also present a case of potential undersampling of the watershed. However, we note that the Lees-A is closely tracked by Lees-D, a PC-based reconstruction with weights on 31 chronologies distributed over the watershed (Figure 4).

We can rule out the choice of calibration period as a major source of differences among the reconstructions; re-calibrating our model Lees-A on 1914-61 (following SJ1976) instead of 1906-95 did not appreciably affect the inferred magnitudes of past droughts. The accuracy of the naturalized flow values is clearly important to the estimated severity of reconstructed droughts: SJ1976 reported important differences in drought severity and in long-term mean reconstructed flow depending on the version of the natural flow record (several existed at that time) used to calibrate their reconstruction model.

5. Recent drought (2000-2004) in a multi-century perspective

To assess the long-term standing of the most recent drought on the Colorado River, the observed natural flows at Lees Ferry averaged over the heart of the recent drought (water years 2000-2004) can be compared with five-year running means of the Lees Ferry reconstruction. Because of the unexplained variance in the regression however, we must allow for the possibility that the true five-year mean for any reconstructed period may have been lower than the reconstructed five-year mean. For this assessment, error bars were placed on the reconstructed five-year

running means. The standard error of a five-year mean was estimated as $s_m = \text{RMSE}_{\text{cv}} / \sqrt{5}$, where RMSE_{cv} is the cross-validation root-mean-square error of the annual reconstructed values. The computed standard error and the assumption that the errors are normally distributed yield confidence intervals and threshold levels of reconstructed five-year mean flow with specific empirical non-exceedance probabilities.

The reconstructed (Lees-A) five-year means for Lees Ferry along with the threshold levels of flow with 0.25 and 0.10 non-exceedance probability are plotted in Figure 11 with the observed 1999-2004 mean as a baseline (“Lowest Observed”) for comparison. The five-year mean for 1999-2004 was 12,187 MCM, or 64.9% of the 1906-95 mean natural flow. The time series plots indicate that only one five-year period, 1844-48, was drier than 1999-2004 (Figure 11a). Annual reconstructed flow during this period averaged 63% of normal.

A probabilistic interpretation of the reconstruction indicates, however, that several other periods have a reasonably large probability of being drier than 1999-2004. Two additional periods, in the early 1500s and early 1600s, have a 25% or greater chance of being as dry as 1999-2004 (Figure 11b). Six periods in addition to the 1840s have a 10% or greater chance of being drier than 1999-2004 (Figure 11c). During the signature drought of 1844-48, the probability is 10% that the true 5-year mean flow was as low as 54.8% of normal (10,290 MCM or 8.3422 MAF). It should be emphasized that Lees-A is the most conservative (wettest) of our Lees Ferry reconstructions, and that other versions give even more frequent past occurrences of flow lower than in 1999-2004.

6. Discussion and Conclusions

6.1. The Updated Reconstructions

An updated and expanded set of tree-ring chronologies has enabled the generation of high-quality water-year streamflow reconstructions for four key gages in the Upper Colorado River basin; the Green River at Green River, UT; Colorado River near Cisco; San Juan near Bluff; and the full Upper Colorado River at Lees Ferry (available online at www.ncdc.noaa.gov/paleo/pubs/woodhouse2006/woodhouse2006.html). These reconstructions span the common years 1569 to 1997, and account for more than 70% of the variance in the gage records. Based on the extensive sensitivity analyses, differences in predictor pools and data reduction methods had little significant impact on important features (e.g. long-term mean, runs of drought years, etc.) of the reconstructions. The use of standard versus prewhitened chronologies does have some impact on the magnitude of reconstructed high and low flows, and the standard chronology models retain a degree of low-order autocorrelation similar to that in the gage record.

The Lees Ferry reconstructions presented here differ from the efforts of Stockton and Jacoby (1976) and Hidalgo et al. (2000) in suggesting a higher long-term mean for Upper Colorado River flows, and to some degree, less extreme multi-year droughts. While the choice of predictor pools and calibration datasets may factor into these differences, statistical

reconstruction methodology, particularly the treatment of autocorrelation, also contributes to reduced drought magnitude and an increased long-term mean.

Spatially, the relationships between reconstructed sub-basin flows are similar to those in the gage records, except for the San Juan reconstruction, which is somewhat more highly correlated with the other gage reconstructions over the instrumental period. This enhanced similarity is lessened over the full reconstruction period. It is possible that the higher correlation between the San Juan and other basins is due to the lack of tree-ring chronologies actually located in the San Juan River basin. However, exploratory analyses using several recently generated tree-ring chronologies in the San Juan basin did not change these results (Woodhouse, unpublished). The reconstructions also capture the contribution of sub-basin flows to total Colorado River flow at Lees Ferry. The sub-basin flows together account for about 96% of upper Colorado River flow and contributions from the three basins are relatively stable over the 431-year common period.

As seen in the comparisons of the Lees Ferry and sub-basin reconstructions, over the past four centuries severe multi-year and decadal-scale droughts in the upper Colorado River basin have tended to be widespread events. The most severe 5-, 10- and 20-year droughts recorded at Lees Ferry are always reflected in the sub-basin gages, although there are sub-regional differences in the magnitude of droughts. When the influence of sub-basin conditions on Lees Ferry flow is examined, most periods of low flow in one sub-basin coincide with low flows in the other sub-basins. There are some exceptions, in particular when flow in the Green River is low and the San Juan flow is high. In most of these periods of contrasting drought conditions, Lees Ferry flows are average, but a few cases (e.g. the 1930s) suggest that drought in the Green River can

have an overriding influence on flows at Lees Ferry, even when high flows prevail on the San Juan. Likewise multidecadal flow regimes tend to be strongly coherent across the basin. Again the magnitude of these persistent high and low-flow events varies across the basin, but the timing and duration of these regimes is consistent among the reconstructions.

6.2. Upper Colorado River Droughts and Possible Climatic Drivers

The coherency of many single and multi-year droughts across the reconstructions points to common drivers for high-frequency variations in regional hydroclimate. Spectral analysis of the Lees Ferry reconstruction (Figure 6) shows significant variability in a three to seven year band associated with the El Niño Southern Oscillation (ENSO; Cayan et al. 1999). Similar high-frequency peaks exist in the sub-basin reconstructions. Examination of gaged values and ENSO indicates a good correspondence between La Niña events and low flows on the San Juan, but the relationship is less clear in the other gages. This agrees with Cayan and Webb (1992) who found that streamflow in the southwestern part of Colorado typically shares the strong southwestern U.S. response to ENSO (i.e., increased winter precipitation during El Niño events), while the response is much weaker at gages north of this region, and Hidalgo and Dracup (2003) who reported the ENSO response is much weaker in the Colorado Headwaters and Upper Green River areas.

Coherency between flows at multidecadal and longer time scales also suggests that remote forcing or region-wide circulation features influence lower-frequency variations in the Upper

Colorado River. Although statistical associations have been demonstrated between North American drought and North Atlantic (Enfield et al. 2001), North Pacific (Cayan et al. 1998, McCabe et al. 2004) and Indian Ocean (Hoerling and Kumar 2004) variability, more research is needed to understand how slow changes in sea surface temperatures are tied to Upper Colorado River flow regimes.

Overall, intra-basin variations in reconstructed drought magnitude, combined with spectral analyses suggesting variability over a broad range of timescales (interannual to multidecadal), indicate complex and possibly non-stationary linkages between the Upper Colorado River and regional to remote forcings. Independent proxy data for ocean variability (i.e. not from western North American tree-rings) and modeling studies are needed to better examine the long-term relationships between Colorado River flows and potential climatic drivers.

6.3. Implications for Management

The recent drought has been a wake-up call for many water management agencies throughout the Colorado River basin. This drought (2000-2004), as measured by 5-year running means of water-year total flow at Lees Ferry, is a markedly severe event in the context of the tree-ring reconstruction extending to 1490, and the probability is low ($p < 0.10$) that any 5-year period since 1850 has been as dry. But the current drought is not without precedence in the tree-ring record. Average reconstructed annual flow for the period 1844-48 was lower than the observed flow for 1999-2004. In view of reconstruction error, it is helpful to evaluate tree-ring

reconstructions probabilistically, and such an evaluation suggests that eight periods between 1536 and 1850 had at least a 10% probability of being as dry as 1999-2004. In addition, longer duration droughts have occurred in the past. The Lees Ferry reconstruction contains one sequence each of six, eight, and eleven consecutive years with flows below the 1906-1995 average (1663-1668, 1776-1783, and 1873-1883). Overall, these analyses demonstrate that severe, sustained droughts are a defining feature of Upper Colorado River hydroclimate. Flows in the Upper Colorado are also shown to be non-stationary over decadal and longer timescales, making short-term records inappropriate for most planning and forecast applications.

Although our results differ in some respects from those of Stockton and Jacoby (1976), the underlying messages are the same. The long-term perspective provided by tree-ring reconstructions points to looming conflict between water demand and supply in the upper Colorado River basin. This suggestion has even greater relevance today. Demands on the Colorado River over the past decades have risen to meet or exceed average water-availability. Any variations or shifts in climate can have a significant impact on the system (Harding et al. 1995, Christensen et al. 2004,). The sensitivity of the Colorado River system became abundantly clear with the onset of the recent drought. Though the southern portion of the Upper Colorado, as well as many areas in the Lower Basin, gained a measure of drought relief in the winter of 2004-2005, major reservoirs on the Colorado River remained far below capacity in 2005. In the future, predicted climatic changes, including a shift in the ratio of snowfall to rainfall and earlier snowmelt and runoff (Cayan et al. 2001, Stewart et al. 2004), will likely compound the strain on water resources throughout the entire Colorado River Basin.

Many such climatic changes may have already begun in the western U.S. (Mote 2005), and rising temperatures will also increase demands for irrigation and hydropower generation. Proxy reconstructions can aid in planning for these scenarios by providing insights into the range of natural variability and a means to explore extreme climatic events and persistent climatic changes that are poorly captured in observational records. Reconstructions of annual streamflow for large rivers are particularly useful in that they integrate climatic variability over large regions, provide essential data for water managers, and complement existing reconstructions of seasonal climate variability (e.g. Cook et al. 2004). In concert with information on projected future changes, information on long-term variability must guide planning for drought management and economic development in the basin if we are to adequately face the social, legal and environmental challenges that coming decades will undoubtedly present.

Acknowledgements

Gray was funded by the U.S. Geological Survey and Wyoming Water Development Commission. Meko was funded by a grant from the Arizona Board of Regents Technology and Research Initiative Fund. Woodhouse received funding from the NOAA Office of Global Programs Climate Change Data and Detection program (grant GC02-046). We greatly appreciate the comments of Edward Cook and two anonymous reviewers. We also thank Jeff Lukas, Mark Losleben, Margot Kaye, Gary Bolton, Kurt Chowanski, Stephen Jackson, Julio Betancourt, and R.G. Eddy for field and laboratory assistance in tree-ring chronology data collections and chronology development, and James Prairie (USBR) for providing the estimates of natural flow for Colorado River basin gages used in the calibrations.

References

Cayan, D.R. and R. H. Webb (1992), El Niño/Southern Oscillation and streamflow in the western United States. in Diaz, H. F. and Markgraf, V. (eds.), *El Niño: Historical and Paleoclimatic Aspects of the Southern Oscillation*, Cambridge University Press, pp. 29-68.

Cayan, D. R., M. D. Dettinger, H. F. Diaz, and N. E. Graham (1998), Decadal variability of precipitation over western North America, *J. Climate*, 11, 3148-3166.

Cayan, D.R., K.T. Redmond, and L.G. Riddle (1999), ENSO and hydrologic extremes in the western United States, *J. Climate*, 12, 2881-2893.

Cayan, D. R., S.A. Kammerdiener, M.D. Dettinger, J.M. Caprio, D.H. Peterson (2001), Changes in the onset of Spring in the Western United States, *Bull. Amer. Meteorol. Soc.*, 82, 399-415.

Christensen, N.S., A.W. Wood, N. Voisin, D.P. Lettenmaier, and R.N. Palmer (2004), Effects of climate change on the hydrology and water resources of the Colorado River Basin, *Clim. Change*, 62, 337-363.

Conover W. (1980), *Practical Nonparametric Statistics, 2nd Edition*, John Wiley & Sons, New York.

Cook E. R. (1985), A time series approach to tree-ring standardization, Ph.D. Dissertation, University of Arizona, Tucson.

Cook E. R. and K. Briffa (1990), A comparison of some tree-ring standardization methods, in *Methods of Dendrochronology, Applications in the Environmental Sciences*, edited by E. R. Cook and L. A. Kairiukstis, pp. 153-162. Kluwer Academic Publishers, Dordrecht.

Cook E. R., K. Briffa, S. Shiyatov, and V. Mazepa (1990), Tree-ring standardization and growth-trend estimation, in *Methods of Dendrochronology, Applications in the Environmental Sciences*, edited by E. R. Cook and L. A. Kairiukstis, pp. 104-123. Kluwer Academic Publishers, Dordrecht.

Cook, E.R., C.A. Woodhouse, C.M. Eakin, D.M. Meko, and D.W. Stahle (2004), Long-term aridity changes in the western United States. *Science*, 306, 1015-1018.

Enfield, D. B., A. M. Mestas-Nuñez and P. J. Trimble (2001), The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental U.S., *Geophys. Res. Lett.*, 28, 2077-2080.

Fritts H. C. (1976), *Tree Rings and Climate*, Academic Press, London.

Fritts H. C., J. Guiot, and G.A. Gordon (1990), Verification, in *Methods of Dendrochronology, Applications in the Environmental Sciences*, edited by E. R. Cook and L. A. Kairiukstis, pp. 178-185, Kluwer Academic Publishers, Dordrecht.

Fulp, T. (2005), How low can it go? *Southwest Hydrology*, 4, 16-17, (March/April).

Gray, S. T., S.T. Jackson, and J.L. Betancourt (2004), Tree-ring based reconstructions of interannual to decadal-scale precipitation variability for northeastern Utah, *J. Amer. Water Resour. Assoc.*, 40, 947-960.

Goodman, P.H. (1996), NevProp software, version 3, Users' Manual, University of Nevada, Reno, (<ftp://ftp.scs.unr.edu/pub/cbmr/nevpropdir>).

Haan C. T. (2002), *Statistical methods in Hydrology, Second Edition*, Iowa State University Press, Ames, Iowa.

Harding, B.L., T.B. Sangoyomi, and E.A. Payton (1995), Impacts of severe sustained drought in Colorado River water resources, *Water Resour. Bull.*, 316, 815-824.

Hidalgo, H.G. and J.A. Dracup (2003), ENSO and PDO effects on hydroclimatic variability in the Upper Colorado River Basin, *J. Hydrometeorology*, 4, 5-23.

Hidalgo H. G., T.C. Piechota, and J.A. Dracup (2000), Alternative principal components regression procedures for dendrohydrologic reconstructions, *Water Resour. Res.*, 36, 3241-3249.

Hoerling, M. and A. Kumar (2003), The perfect ocean for drought, *Science*, 299, 691-694.

Mann, M.E., and J. Lees (1996), Robust estimation of background noise and signal detection in climatic time series, *Clim. Chang.*, 33, 409-445.

Mardia K., Kent J. and Bibby J. (1979) *Multivariate Analysis*. Academic Press.

McCabe, G. J., M.A.Palecki, and J.L. Betancourt (2004), Pacific and Atlantic Ocean influences on multidecadal drought frequency in the United States, *Proceedings of the National Academy of Sciences*, 101, 4136-4141

Michaelsen J. (1987), Cross-validation in statistical climate forecast models, *J. Clim. and Appl. Meteorol.*, 26, 1589-1600.

Meko D. M., C.W. Stockton, and W.R. Boggess (1995), The tree-ring record of severe sustained drought, *Water Resour. Bull.*, 31, 789-801.

Mote, P.W., A.F. Hamlet, M.R. Clark, and D.P. Lettenmaier (2005), Declining mountain snowpack in western North America, *Bull. Amer. Meteorol. Soc.*, 86, 39-49.

Rencher A. C., and F.C. Pun (1980), Inflation of R^2 in best subset regression, *Technometrics*, 22, 49-53.

Snedecor G. W., and W.G. Cochran (1989), *Statistical Methods, 8th rev. ed.*, Revised by Iowa State University Statistics Department Staff, Iowa State University Press.

Stahle D. W., Cook E. R., Cleaveland M. K., Therrell M. D., Meko D. M., Grissino-Mayer H. D., Watson E. and Luckman B. H. (2000), Tree-ring data document 16th century megadrought over North America, *Eos: Trans. Amer. Geophys. Union*, 81, 121-125.

Stewart, I.T., D.R. Cayan, and M.D. Dettinger (2004), Changes in snowmelt runoff timing in western North America under a 'Business as Usual' climate change scenario, *Clim. Change*, 62, 217-232.

Stockton C. W. and G.C. Jacoby (1976), Long-term surface-water supply and streamflow trends in the Upper Colorado River Basin, Lake Powell Research Project Bulletin No. 18. National Science Foundation, Arlington, VA.

Stokes M. A. and T.L. Smiley (1968), *An Introduction to Tree-ring Dating*, The University of Arizona Press, Tucson.

Swetnam, T. W., M. A. Thompson, and E. K. Sutherland (1985), Using dendrochronology

to measure radial growth of defoliated trees, Agriculture Handbook 639, USDA Forest Service, Washington, DC.

Torrence, C. and G.P. Compo (1998), A practical guide to wavelet analysis, *Bull. Amer. Meteor. Soc.*, 79, 61-78.

Weisberg S. (1985), *Applied Linear Regression, 2nd ed*, John Wiley, New York.

Wigley, T., K. Briffa, and P.D. Jones (1984), On the average value of correlated time series, with applications in dendroclimatology and hydrometeorology, *J. Climate and Appl. Meteor.*, 23, 201-213.

Woodhouse, C.A. (1999), Artificial neural networks and dendroclimatic reconstructions: an example from the Front Range, Colorado, USA, *Holocene*, 9, 521-529.

Young R. A. (1995), Coping with a severe sustained drought on the Colorado River: introduction and overview, *Water Resour. Bull.*, 316, 779-788.

Figure Captions

Figure 1. Location of gages at (A) Green River at Green River, UT, (B) Colorado River near Cisco, UT, (C) San Juan River near Bluff, UT, and (D) Lees Ferry, Arizona (dots) and tree-ring chronologies (triangles). The upper Colorado River basin is outlined in a solid line and the sub-basins discussed are outlined by the dotted and solid lines (Green, Colorado with Yampa and Gunnison, and the San Juan basins). Tree-ring chronologies used in Lees Ferry stepwise regression are circled; a heavy black line indicates chronologies used in regression equations calibrated with both standard and residual chronologies, a gray line indicates chronologies used in the standard chronology calibration, and a thin black indicates chronologies used in the residual chronology calibration.

Figure 2. Comparison of observed and reconstructed streamflow, Lees Ferry gage (black line) and Lees-A reconstruction (gray line), 1906-1997 (gage to 1995).

Figure 3. San Juan near Bluff, Utah reconstructed streamflow, 1569-1997, from two models, full predictor pool (gray lines) and watershed-limited predictor pool (black lines), annual values, top, and 10-year running average, bottom.

Figure 4. Reconstructed 20-yr running means of Colorado River at Lees Ferry, Arizona, by alternative statistical models. Horizontal lines are the observed mean of the unsmoothed flows for the 1906-95 calibration period (dashed) and the reconstructed mean of unsmoothed flows for

the entire (1490-1997) Lees-A reconstruction (solid). See text and Table 3 for definitions of the models.

Figure 5. Five-year running averages of reconstructed annual streamflow, 1571-1995 for (a, dark line) Lees Ferry, Arizona, (a, light line) the sum of the flow for the three reconstructions, and the three sub-basins; (b) Colorado near Cisco, Utah; (c) Green River at Green River, Utah; and (d) San Juan River near Bluff, Utah.

Figure 6. Multi-taper method spectral analyses (Mann and Lees, 1996) of reconstructed flows for (a) the Colorado at Lees Ferry, Arizona; (b) the Green River at Green River, Utah; (c) the Colorado near Cisco, Utah; (d) the San Juan near Bluff, Utah. All spectra cover the common period from 1569-1997 AD. Peaks are shown versus the 95% confidence level (dotted black line). These analyses were performed using a 3×2 pi taper under red noise assumptions.

Figure 7. Multi-taper method cross-spectral analysis (Mann and Lees, 1996) of reconstructed flows at the major sub-basin gages on the Green River at Green River, Utah; Colorado River near Cisco, Utah; and San Juan River near Bluff, Utah. The coherency spectra (top) are plotted against the 95% confidence interval (dotted line). Phasing of spectral peaks is shown in the lower panel.

Figure 8. Wavelet power spectra of reconstructed flows for (a) the Colorado River at Lee's Ferry, Utah; (b) Green River at Green River, Utah; (c) Colorado River near Cisco, Utah; (d) San Juan River near Bluff, Utah. A derivative to the Gaussian wavelet ("Mexican Hat"; see Torrence

and Compo, 1998) was used under red noise assumptions, and each reconstruction was padded with zeros to avoid wrap-around effects. Black contours in the power spectra represent the 95% confidence level compared to red noise. The cone-shaped net shows portions of the spectrum where power may be reduced through the effects of zero padding (Torrence and Compo, 1998). All spectra cover the common period from 1569-1997 AD.

Figure 9. Reconstructed Upper Colorado River flows, smoothed with a 50-year spline.

Figure 10. Twenty-year running means of four alternative reconstructions of the annual flow of the Colorado River at Lees Ferry for common period 1520-1961. Lees-A is our updated reconstruction from residual chronologies. Lees-D is our updated reconstruction from PCs of standard chronologies (see text). SJ1976 is the mean of two reconstructions generated by equations 2 and 3 on page 24 of that reference. HPD2000 is the PC-based reconstruction of Hidalgo et al. (2000). The horizontal lines are the 1906-2004 observed mean (solid) and the lowest observed 20-year running mean of the 1906-2004 period (dash-dot).

Figure 11. Current drought in long-term context from reconstructed five-year running means of natural flow at Lees Ferry, Arizona. (a) observed and reconstructed flow. (b) observed flow and flow with 0.25 non-exceedance probability derived from reconstruction and its estimated error variance. (c) observed flow and reconstructed flow with 0.10 non-exceedance probability. Flow plotted as percentage of 1906-95 mean of observed mean annual flow, or 18,788 million cubic meters (15.232 MAF). Reconstruction series from model Lees-A (see text).

Table 1. Metadata and descriptive statistics of annual flows.

Gage Name ¹	USGS ID	Basin Area (10 ⁶ ha)	Flow Statistics ²			
			Mean (10 ⁶ m ³)	cv	skew	r ₁
A Green R. at Green River, UT	9315000	11.6161	6704	0.30	0.38	0.26**
B Colorado R. nr Cisco, UT	9180500	6.2419	8505	0.28	0.22	0.25**
C San Juan R. nr Bluff, UT	9379500	5.9570	2711	0.40	0.32	0.12
D Colorado R. at Lees Ferry, AZ	9380000	28.9562	18778	0.28	0.15	0.25**

¹Gage locations coded by letter (A,B,C,D) shown on map in Figure 1

²Mean, coefficient of variation, skewness coefficient and first-order autocorrelation computed from 1906-95 annual (water-year total) flows. Asterisks mark significance of first-order autocorrelation based on one-tailed test (* for $\alpha=0.05$, ** for $\alpha = 0.01$). None of the skewness coefficients are significantly different from zero at $\alpha=0.05$.

Table 2. Regression statistics for reconstruction models. Top: sub-basin models based on full-pool and watershed-limited pool of potential predictors. Bottom: statistics for four alternative modeling choices for Lees Ferry record. Validation statistics RE and RMSE are based on cross-validation.

gage	predictors in # steps/ pool		R ²	F-level	RE	RMSE
	pool	# predictors				
<i>full pool</i>						
GR/UT	28	9/7	0.72	30.0	0.66	1149.2
CO/Cisco	38	5/5	0.77	54.0	0.73	1248.7
SJ/Bluff	24	4/4	0.73	56.7	0.70	589.7
<i>limited pool</i>						
GR/UT	18	9/7	0.72	30.0	0.66	1149.2
CO/Cisco	32	4/4	0.73	57.7	0.69	1330.4
SJ/Bluff	8	3/3	0.67	58.8	0.64	640.6
<i>Lees Ferry¹</i>						
Lees-A (res)	31	7/7	0.81	48.7	0.76	2579.1
Lees-B (std)	30	7/7	0.84	61.2	0.81	2337.1
Lees-C (res,PCA)	3	1/1	0.72	226.9	0.71	2861.3
Lees-D (std,PCA)	4	1/1	0.77	294.7	0.76	2599.5

¹See Table 3 for definitions of models

Table 3. Statistics of observed and reconstructed flow of Colorado River at Lees Ferry for 1906-95 calibration period.

Series ¹	Statistics ²				Running Means as Percentage of Normal ³					
					Lowest			Highest		
	mean	stdev	skew	r(1)	1-yr	5-yr	20-yr	1-yr	5-yr	20-yr
Lees-A	18778	4787	-0.14	0.04	31	79	89	157	139	111
Lees-B	18778	4885	-0.26	0.22	28	76	83	151	142	115
Lees-C	18778	4526	-0.52	-0.05	25	81	90	141	130	108
Lees-D	18778	4679	-0.47	0.31	28	73	83	148	139	115
Obs.	18778	5332	0.15	0.25	37	72	85	166	145	116

¹Lees-A is reconstruction from residual chronologies; Lees-B is from standard chronologies; Lees-C is from PCs of residual chronologies Lees-D is from PCs of standard chronologies; Obs is the observed natural flow record (see text).

²Statistics: mean and standard deviation in MCM, skewness, lag-1 autocorrelation

³Normal defined as mean of observed flows for calibration period 1906-95

Table 4. Statistics¹ of reconstructed flow of Colorado River at Lees Ferry, 1490-1997, and observed flow, 1906-95. Series and columns defined as in Table 3.

Series	Statistics				Running Means as Percentage of Normal					
					Lowest			Highest		
	mean	stdev	skew	r(1)	1-yr	5-yr	20-yr	1-yr	5-yr	20-yr
Lees-A	18097	5555	-0.30	0.04	11	63	83	167	142	115
Lees-B	17957	5616	-0.21	0.29	11	56	77	167	156	124
Lees-C	18124	4990	-0.32	-0.00	13	66	81	172	130	112
Lees-D	17656	5193	-0.19	0.34	14	50	68	170	143	119
Obs.	18778	5332	0.15	0.25	37	72	85	166	145	116

¹See Table 3 for definitions of columns

Table 5. Inter-basin correlations of observed and reconstructed flows; (a) correlation matrices of observed and reconstruction flow for the period 1906-1995, and (b) correlation matrix of reconstructed flows for period 1569-1997. GRUT = Green River at Green River, COCI = Colorado River at Cisco, SJBL = San Juan near Bluff, Lees = Colorado at Lees Ferry.

a.	GRUT	COCI	SJBL
Observed			
COCI	0.85		
SJBL	0.55	0.77	
Lees	0.92	0.98	0.79
Reconstructed			
COCI	0.87		
SJBL	0.71	0.84	
Lees	0.93	0.95	0.83

b.	GRUT	COCI	SJBL
COCI	0.87		
SJBL	0.69	0.83	
Lees	0.93	0.96	0.82

Table 6. Percentage of observed and reconstructed annual flow at Lees Ferry contributed by sub-basins.

	3 gage total	GRUT	COCI	SJBL
Observed 1906-95	95.5	35.8	45.4	14.2
Reconstructed 1906-95	95.6	35.9	45.5	14.3
Reconstructed 1569-1997	96.3	35.7	46.7	13.9

Table 7. Ranked sub-basin flows during lowest 5-year, 10-year, and 20-year moving averages of reconstructed flow at Lees Ferry. The ending years of the ten lowest-flow periods at Lees Ferry are listed under “5-year”, “10-year” and “20-year”. Sub-basin gages are: Green River at Green River (GR), Colorado at Cisco (CC) and San Juan at Bluff (SJ).

Lees rank	5 year	GR rank	CC rank	SJ rank	10 year	GR rank	CC rank	SJ rank	20 year	GR rank	CC rank	SJ rank
1	1848	1	3	17	1631	71	6	3	1592	6	1	48
2	1626	63	2	1	1782	9	8	17	1593	7	2	82
3	1847	8	8	9	1592	5	3	57	1641	57	14	13
4	1846	3	12	54	1632	40	1	2	1889	2	9	5
5	1759	24	16	29	1781	11	9	12	1598	3	3	97
6	1686	11	6	22	1593	1	7	92	1890	1	15	20
7	1584	16	1	8	1883	3	25	4	1671	13	16	45
8	1883	7	36	2	1737	43	10	16	1638	52	55	28
9	1668	2	13	44	1879	4	26	18	1639	66	57	25
10	1902	10	5	3	1880	2	18	24	1640	77	64	21

Table 8. Terciles and percentiles of 10-year moving average reconstructed flow at Lees Ferry during periods of contrasting flow anomalies in sub-basins. Year listed is last of ten.

Contrast	Lees Flow Tercile	Lees Flow Percentile
Green dry/ San Juan wet		
1597	middle	0.546
1600	dry	0.171
1674	dry	0.283
1743	middle	0.406
1856	middle	0.475
1941	dry	0.276
1942	dry	0.320
1943	middle	0.363
Green wet/ San Juan dry		
1735	middle	0.463
1766	middle	0.658
1774	middle	0.518
Green dry/ Colorado wet		
1818	middle	0.615
1823	middle	0.489
San Juan wet/ Colorado dry		
1859	middle	0.499
San Juan dry/ Colorado wet		
1820	middle	0.610

Table 9. Comparative statistics of Lees Ferry reconstructions

Model ¹	Calibration ²			Reconstruction ³	
	Period	N _c	R ²	Lowest 20-yr Mean	Long-Term Mean
Lees-A	1906-1995	7	0.81	15.6 ± 1.2 BCM (12.6 ± 0.9 MAF)	18.1 ± 0.2 BCM (14.7 ± 0.2 MAF)
Lees-D	1906-1995	30	0.77	12.8 ± 1.1 BCM (10.4 ± 0.9 MAF)	17.6 ± 0.2 BCM (14.3 ± 0.2 MAF)
SJ1976	1914-1961	17	0.78*	13.5 BCM (10.9 MAF)	16.7 BCM (13.5 MAF)
HPD2000	1914-1961	6	0.82	11.2 ± 1.0 BCM (9.1 ± 0.8 MAF)	16.3 ± 0.2 BCM (13.2 ± 0.2 MAF)

¹Reconstruction model (see text)

²Calibration period, number of contributing chronologies, and proportion of variance explained by regression model

³Statistics of long-term reconstruction, expressed in units of billion cubic meters and million acre-feet along with 95% confidence interval estimated from the cross-validation root-mean-square error (see text); statistics for common period 1520-1961

*SJ1976 is a mean of two reconstructions with R² values of 0.78 and 0.87 (Stockton and Jacoby 1976); no cross-validation was performed for these models

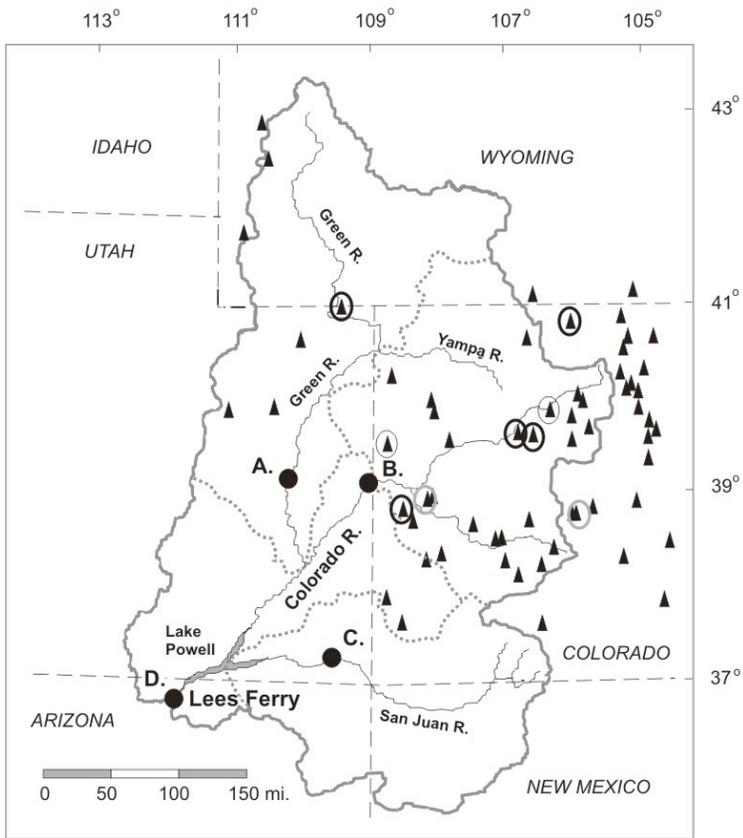


Figure 1. Location of gages at (A) Green River at Green River, UT, (B) Colorado River near Cisco, UT, (C) San Juan River near Bluff, UT, and (D) Lees Ferry (dots) and tree-ring chronologies (triangles). The upper Colorado River basin is outlined in a solid line and the sub-basins discussed are outlined by the dotted and solid lines (Green, Colorado with Yampa and Gunnison, and the San Juan basins). Tree-ring chronologies used in Lees Ferry stepwise regression are circled; a heavy black line indicates chronologies used in regression equations calibrated on both standard and residual chronologies, a gray line indicates chronologies used in standard chronology calibration, and thin black line indicates chronologies used in residual chronology calibration.

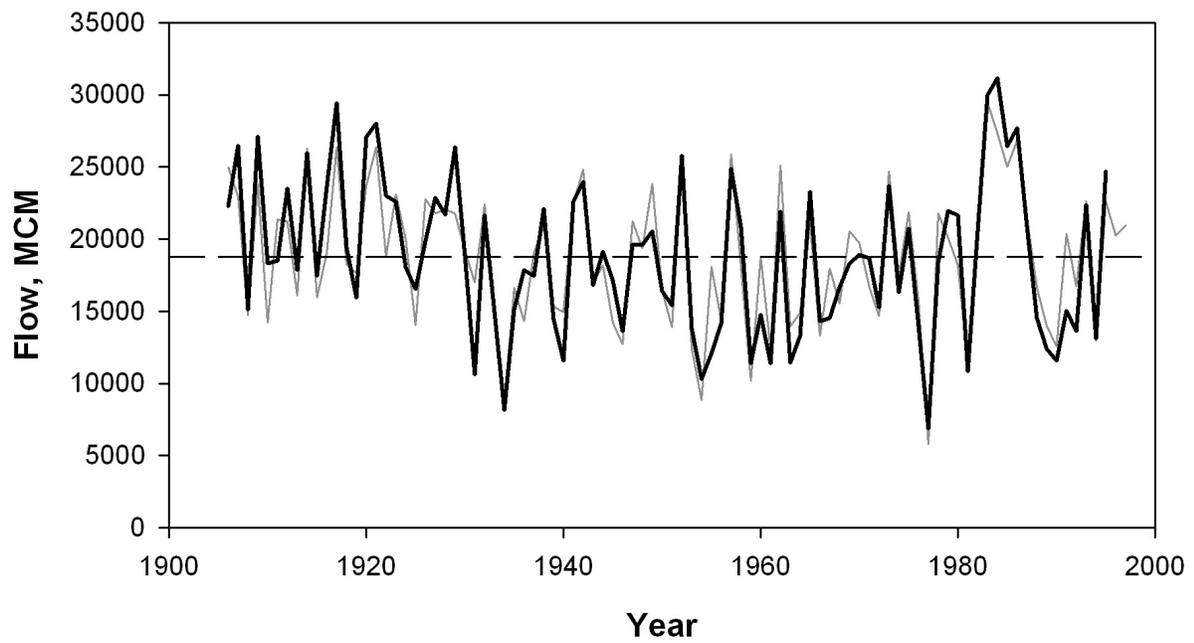


Figure 2. Comparison of observed and reconstructed streamflow, Lees Ferry gage (black line) and Lees-A reconstruction (gray line), 1906-1997 (gage to 1995).

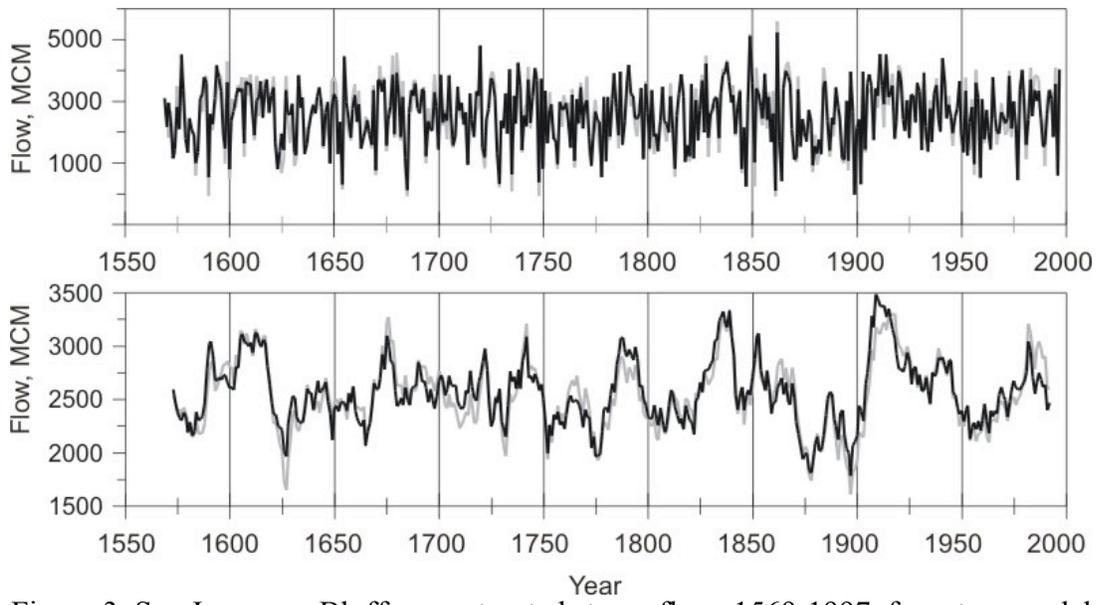


Figure 3. San Juan near Bluff reconstructed streamflow, 1569-1997, from two models, full predictor pool (gray lines) and watershed-limited predictor pool (black line) for annual values, top, and 10-year running average, bottom.

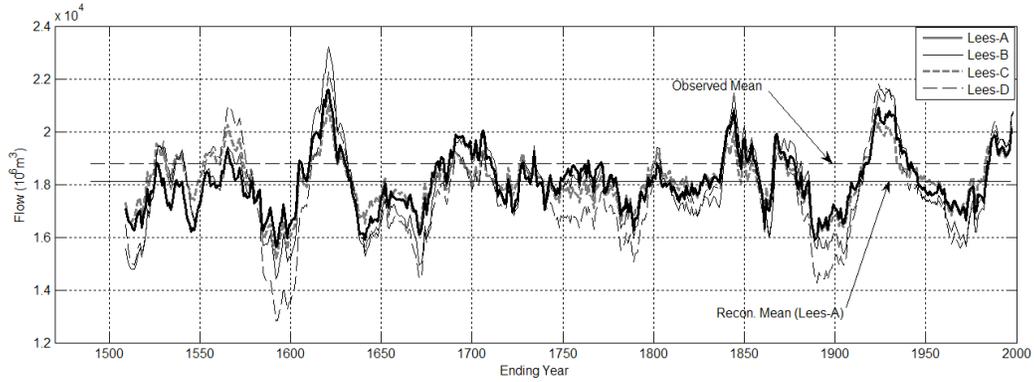


Figure 4. Reconstructed 20-yr running means of Colorado River at Lees Ferry, Arizona, by alternative statistical models. Horizontal lines are the observed mean of the unsmoothed flows for the 1906-95 calibration period (dashed) and the reconstructed mean of unsmoothed flows for the entire (1490-1997) Lees-A reconstruction (solid). See text and Table 3 caption for definitions of the models.

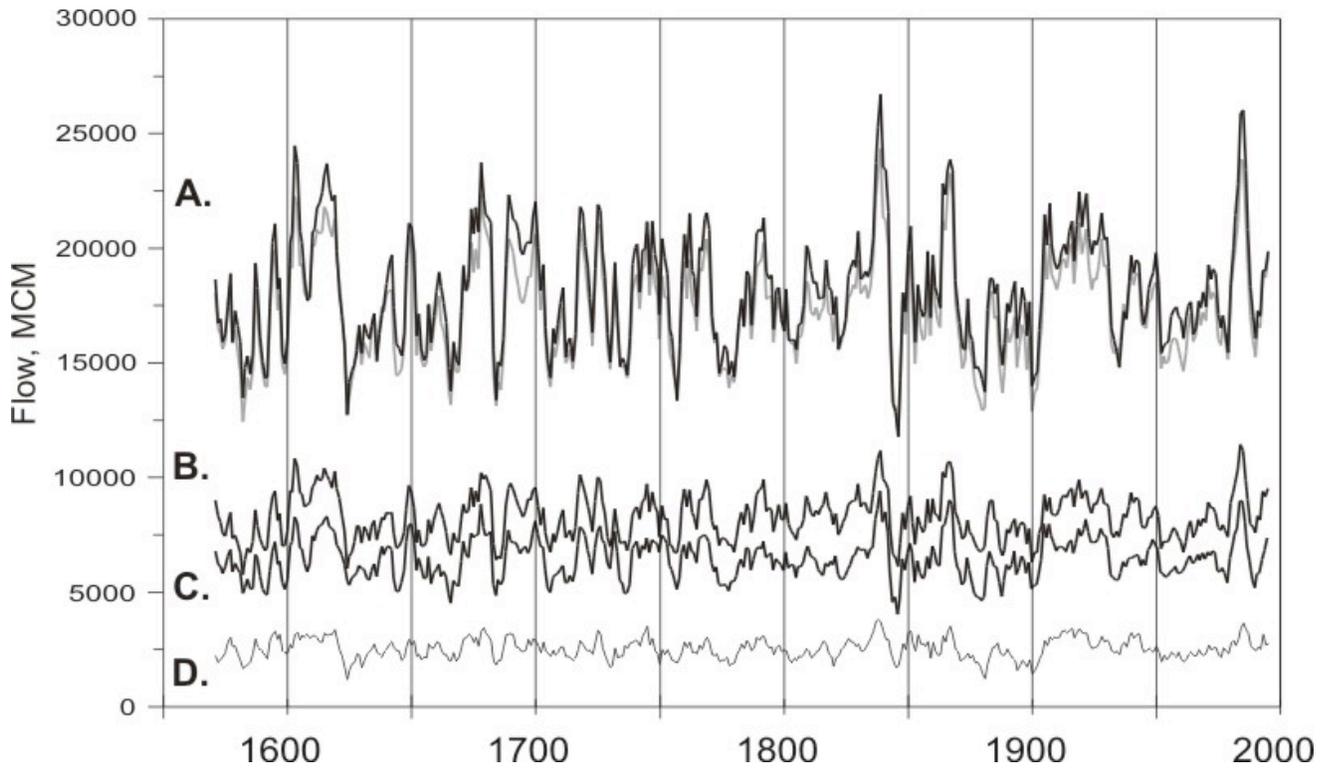


Figure 5. Five-year running averages of reconstructed annual streamflow, 1571-1995 for (A, dark line) Lees Ferry, (A, light line) the sum of the flow for the three reconstructions, and the three sub-basins, (B) Colorado near Cisco, (C) Green River, and (D) San Juan River.

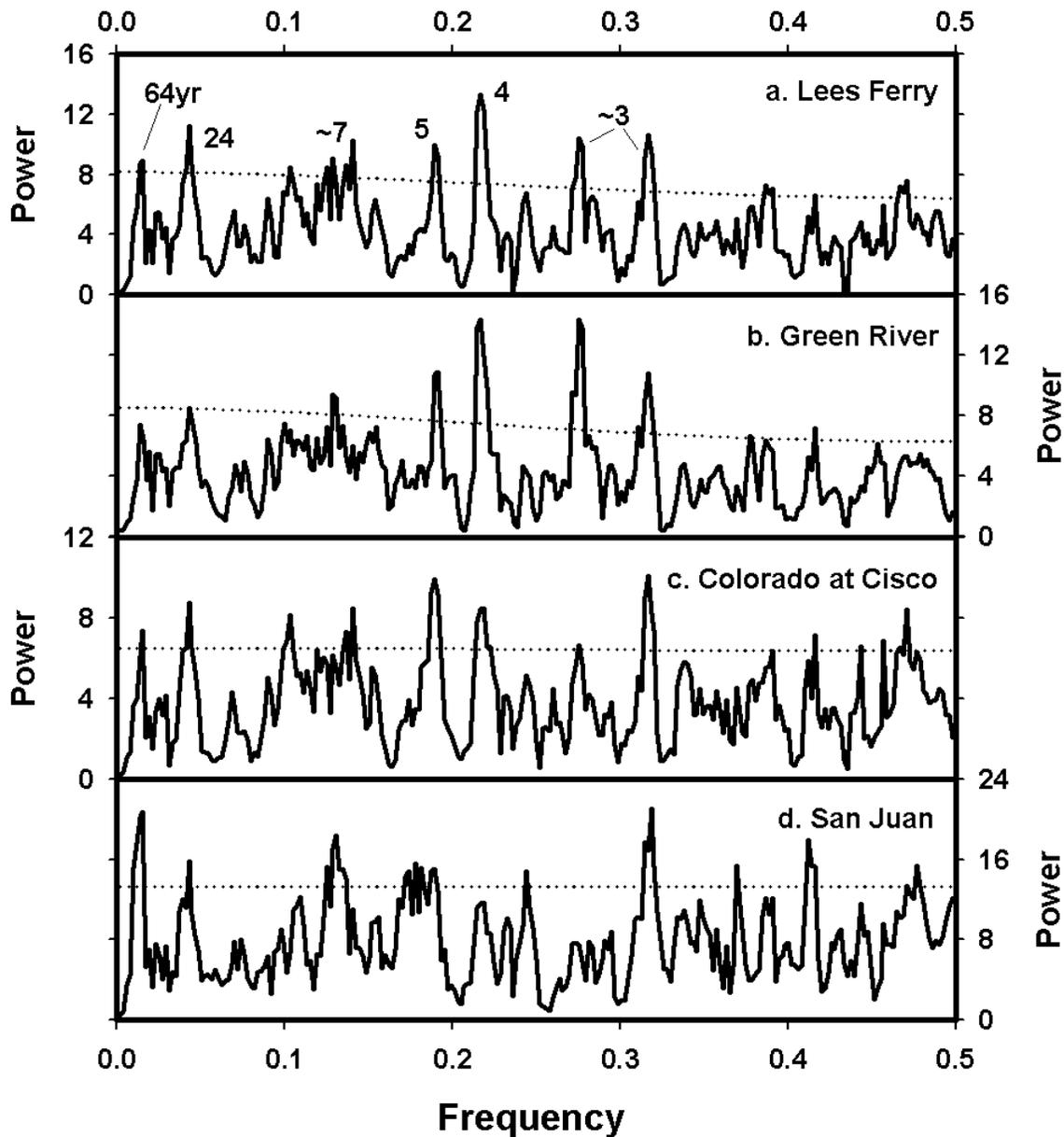


Figure 6. Multi-taper method spectral analyses (Mann and Lees, 1996) of reconstructed flows for (a) the Colorado at Lees Ferry, Arizona; (b) the Green River at Green River, Utah; (c) the Colorado at Cisco, Utah; (d) the San Juan near Bluff, Utah. All spectra cover the common period from 1569-1997 AD. Peaks are shown versus the 95% confidence level (dotted black line). These analyses were performed using a $3 \times 2 \pi$ taper under red noise assumptions.

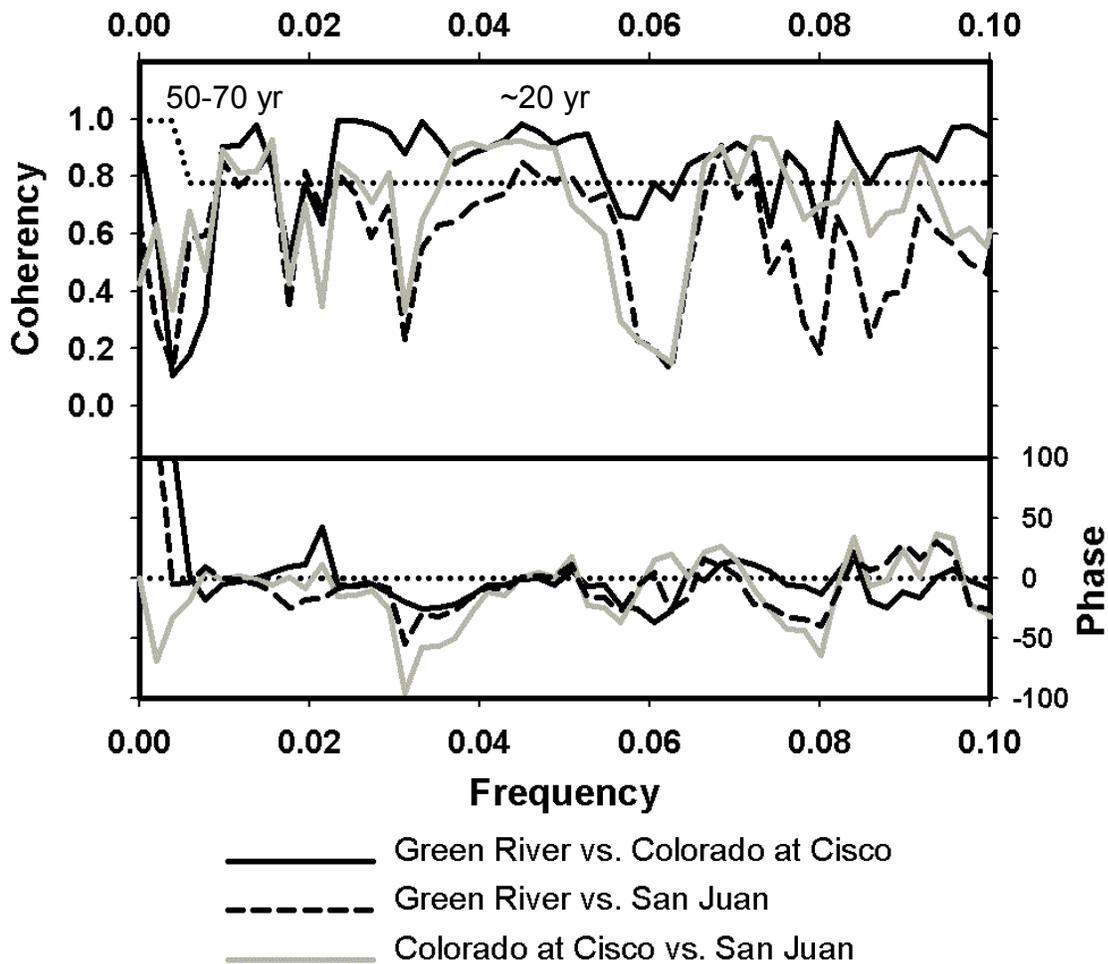


Figure 7. Multi-taper method cross-spectral analysis (Mann and Lees, 1996) of reconstructed flows at the major sub-basin gages on the Green River at Green River, Utah; Colorado River near Cisco, Utah; and San Juan River near Bluff, Utah. The coherency spectra (top) are plotted against the 95% confidence interval (dotted line). Phasing of spectral peaks is shown in the lower panel.

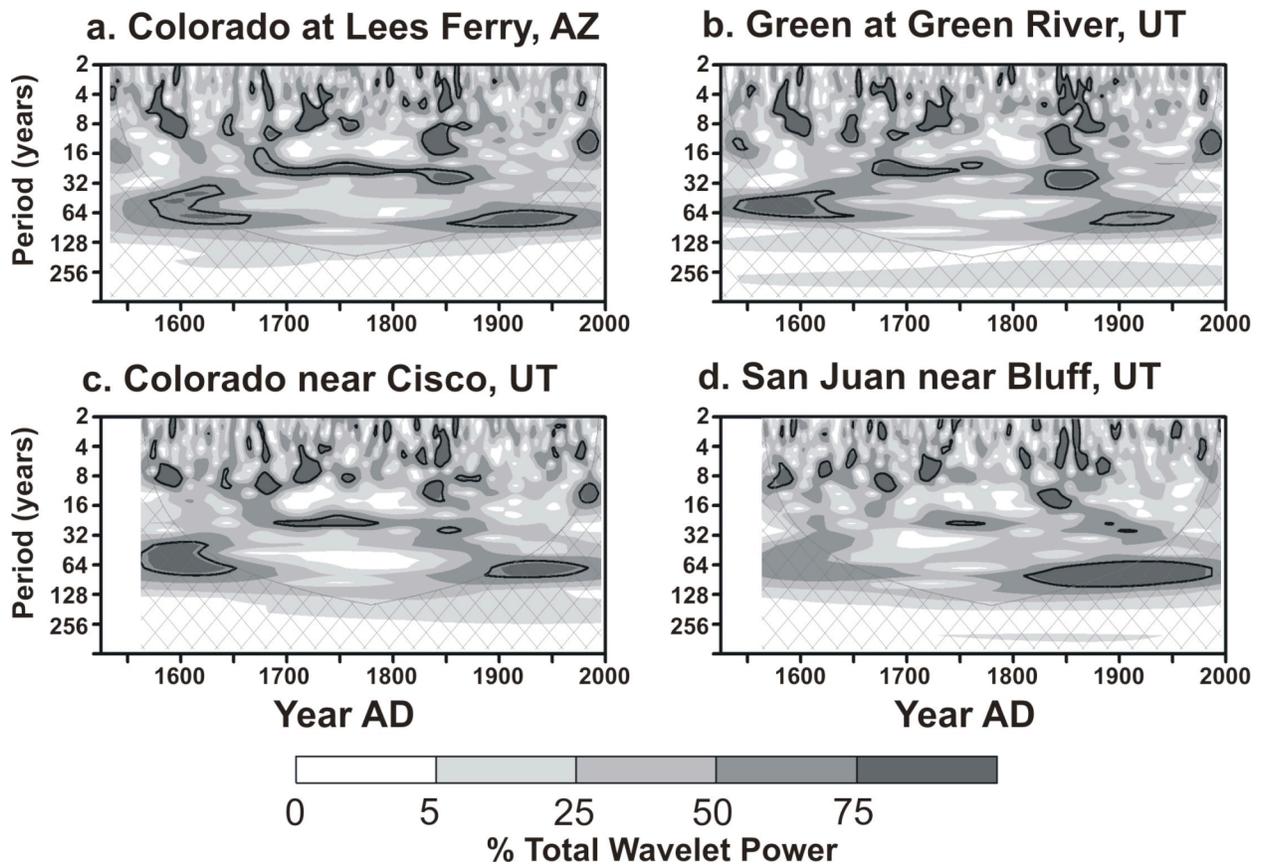


Figure 8. Wavelet power spectra of reconstructed flows for (a) the Colorado River at Lee’s Ferry, Utah; (b) Green River near Green River, Utah; (c) Colorado River near Cisco, Utah; (d) San Juan River near Bluff, Utah. A derivative to the Gaussian wavelet (“Mexican Hat”; see Torrence and Compo, 1998) was used under red noise assumptions, and each reconstruction was padded with zeros to avoid wrap-around effects. Black contours in the power spectra represent the 95% confidence level compared to red noise. The cone-shaped net shows portions of the spectrum where power may be reduced through the effects of zero padding (Torrence and Compo, 1998). All spectra cover the common period from 1569-1997 AD.

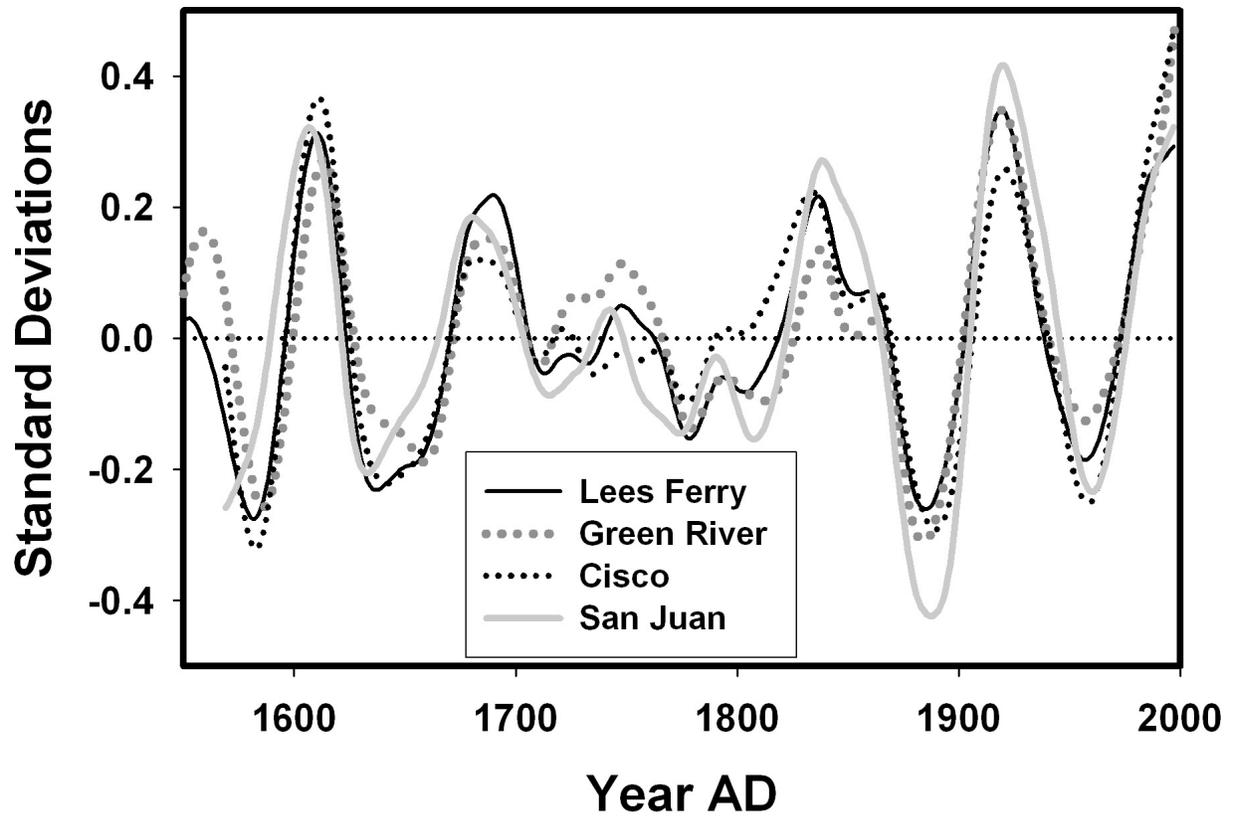


Figure 9. Reconstructed Upper Colorado River flows, smoothed with a 50-year spline.

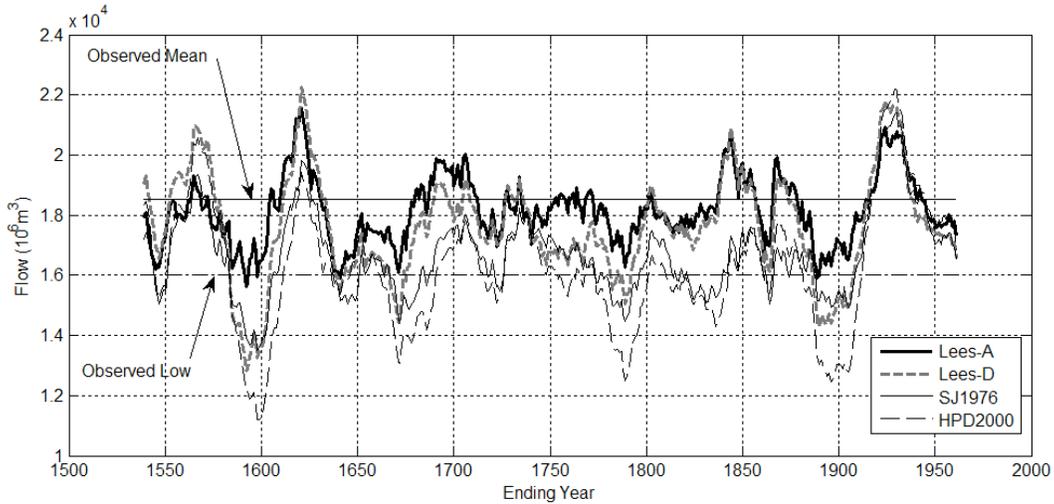


Figure 10. Twenty-year running means of four alternative reconstructions of the annual flow of the Colorado River at Lees Ferry for common period 1520-1961. Lees-A is our updated reconstruction from residual chronologies. Lees-D is our updated reconstruction from PCs of standard chronologies (see text). SJ1976 is the mean of two reconstructions generated by equations 2 and 3 on page 24 of that reference. HPD2000 is the PC-based reconstruction of Hidalgo et al. (2000). The horizontal lines are the 1906-2004 observed mean (solid) and the lowest observed 20-year running mean of the 1906-2004 period (dash-dot).

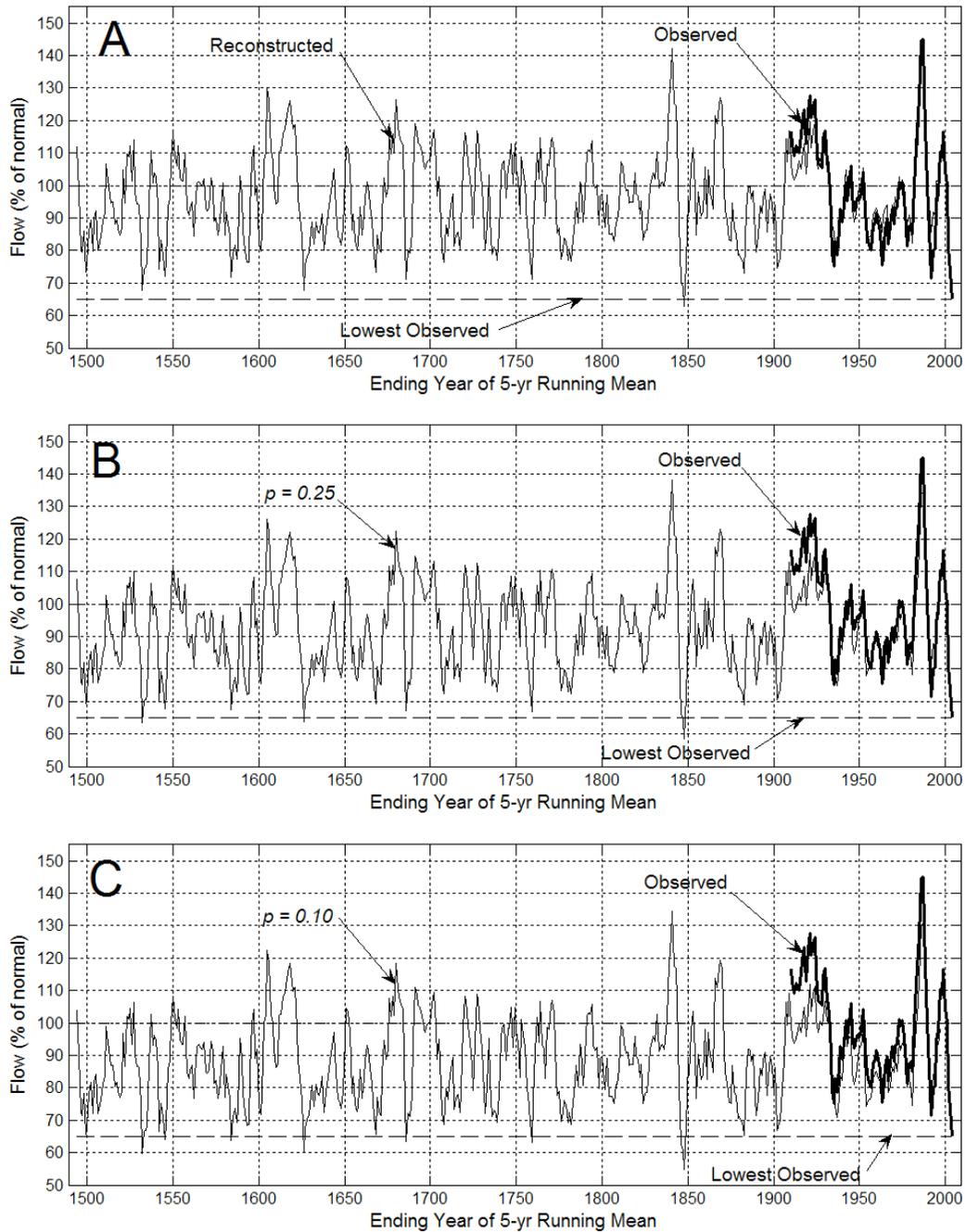


Figure 11. Current drought in long-term context from reconstructed five-year running means of natural flow at Lees Ferry, Arizona. (A) observed and reconstructed flow. (B) observed flow and flow with 0.25 non-exceedance probability derived from reconstruction and its estimated error variance. (C) observed flow and reconstructed flow with 0.10 non-exceedance probability. Flow plotted as percentage of 1906-95 mean of observed mean annual flow, or 18,788 million cubic meters (15.232 MAF). Reconstruction series from model Lees-A (see text).