BOYLE | AECOM

In Association with

amec
Leonard Rice Engineers, Inc.
STRATUS CONSULTING
Canyon Water Resources, LLC

Boyle
215 Union Blvd, Suite 500 | Lakewood CO
T 303.987.3443 | F 303.987.3908

Technical Memorandum | Draft

To:	Blaine Dwyer (Boyle AECOM)
Distribution:	
From:	Ben Harding, Subhrendu Gangopadhyay (AMEC Earth & Environmental)
Subject:	CRWAS Phase 1 | Task 6.1 | Literature Review and Method Evaluation | Task 6.2 | Analyses of Tree-Ring Data | Task 6.3 | Recommendation for Extending Historical Hydrology
Date:	February 8, 2009

Introduction

This Technical Memorandum summarizes information developed as part of Task 6 of the Colorado River Water Availability Study (CRWAS or Study).

The objective of Task 6 is to extend historical hydrologic data using currently available tree-ring data and stochastic methods to develop alternate hydrologic traces in formats usable in the CDSS. Sub-tasks 6.1 through 6.3 involve agency coordination, literature review, data analysis and recommendation of approach.

This memorandum provides a review of relevant literature and an evaluation of alternative approaches of developing alternative historical hydrology, which will be implemented in CRWAS Task 6.4 and used in CRWAS Task 6.5. This memorandum also provides the results of a statistical analysis of one tree-ring reconstruction and places those statistics in the context of the statistics of the record of naturalized flows during the historical period of record. Subsequent sections of this technical memorandum discuss: 1) CRWAS Requirements, 2) Concept Basis and Summary of Past Work 3) Candidate Methodologies; 4) Summary Evaluation of Candidate Methodologies 5) Recommended Methodology, 6) Summary Statistics from Prairie et al. (2008) Paper; 7) References, and 8) an Appendix containing a literature review, a detailed description of the recommended methodology, and a glossary.

Based on the evaluation documented below, we recommend the use of an adaptation of a method developed for the Bureau of Reclamation and applied in their recent model studies used in developing guidelines for Lower Basin shortages and coordinated operations for Lake Powell and Lake Mead on the Colorado River (*Lower Colorado River Guidelines*, Reclamation, 2007).

CRWAS Requirements

The State's Request for Proposals for the Study calls for choice and implementation of a method to use information from prehistoric tree-ring records to extend observed records of flows (i.e., to develop an "alternative historical hydrology"). The subsequent contractual scope of work calls for the use of information from paleo records also to be used to extend the data set that represents conditions during the observation period assuming development of climate change (the *climate-perturbed observed flows).* The State's Request for Proposals and the contractual scope of work call for the generation of an ensemble of flow traces containing at least one hundred (50- to 100-year) traces of alternative hydrology. Thus, methods of extending flow traces using information from tree-rings (as detailed herein) must be applicable to both natural flow records and the climate-perturbed record.

The water resources models to be used in the Study are the Bureau of Reclamation's Colorado River Simulation System (CRSS) model and the State of Colorado's StateMod model, part of the Colorado River Decision Support System (CRDSS). Both of these models (as used in the Study) require monthly inflows. The CRSS model requires monthly inflows at 29 inflow points throughout the basin. The CRDSS StateMod models to be used in the Study require monthly flows at hundreds of *baseflow* gage points throughout those portions of the Colorado River Basin within the State of Colorado. Thus, the method that is adopted to extend flow records must be capable of generating traces of monthly flows at two different levels of spatial detail throughout the Colorado River Basin.

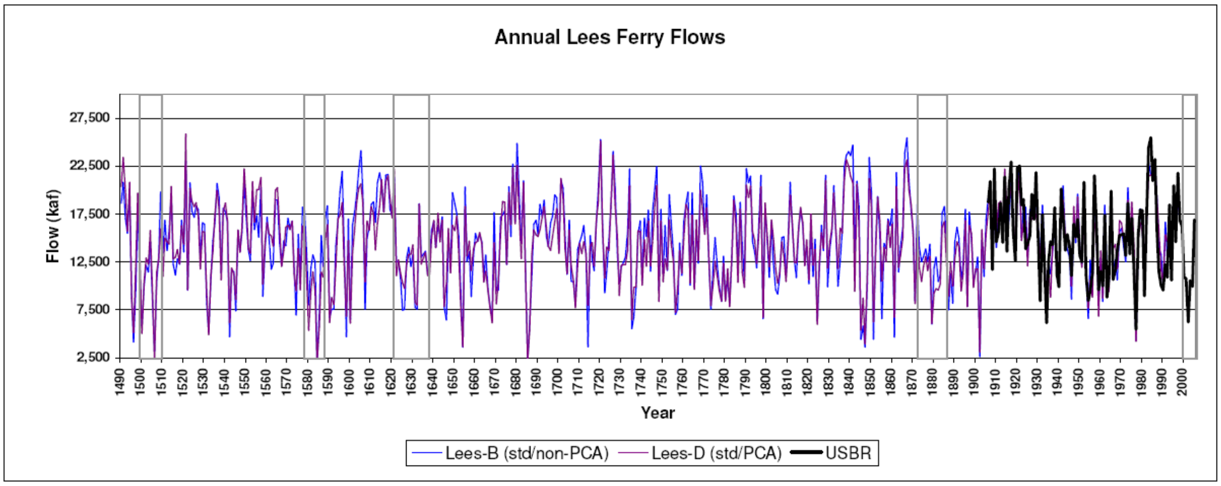
Concept Basis and Summary of Past Work

This section provides a synthesis of concepts that have been used in developing extended hydrologic data from tree-ring chronologies. A more detailed review of relevant publications is included in the Appendix.

Paleohydrologic reconstructions of streamflows are very useful for understanding multi-decadal hydrologic variability and for drought mitigation planning[1]. The overall context for extension of flows using paleohydrology is illustrated using Figure 1, a chart showing paleohydrologic reconstructed annual streamflow for the period 1490-1997 on the Colorado River at Lees Ferry, AZ, (Woodhouse et al., 2006), along with the naturalized observed flows at that site. The *observation period* in Figure 1 extends from 1906 through 2005, while the *pre-observation period* extends from 1490 until 1905. The period over which tree-ring chronologies overlap observed flows extends from 1906 through 1997 (beyond 1997 only a limited number of tree-ring chronologies are available for the Colorado River basin) and is referred to as the *overlap period.* The reconstructions are based on a functional relationship, typically a linear regression, between tree-ring chronologies and the streamflows (e.g., Stockton and Jacoby, 1976; Meko et al., 1995), developed over the overlap period, which is then used to estimate flows during the pre-observation period.

[1] This summary of the methods used in developing paleo-reconstructions is adapted from Gangopadhyay et al. (2008).

Figure 1.
Reconstruction of Colorado River annual flow at Lees Ferry.

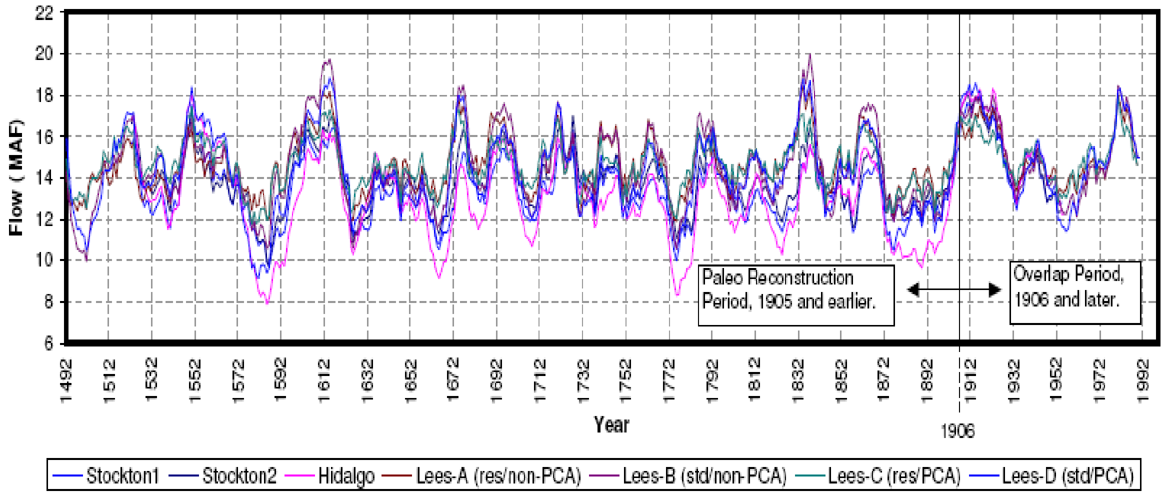


In this approach, a suite of trees are cored to obtain a chronology of tree-ring widths, which are corrected for physiological and other biases to obtain tree-ring growth indices[2]. Tree-ring growth indices for many trees at one site are typically aggregated (usually by averaging) into a *chronology,* which contains a single index value for each year in the chronology. A stepwise regression approach is used to select the best subset of tree-ring chronologies, based on the ability of that subset to predict streamflows at a specified location, and a multiple linear regression (MLR) model is fitted to the observed streamflow. This MLR model is then used to estimate streamflows during the pre-observation period using tree-ring chronologies. Variations of this basic approach have been proposed – for instance, Hidalgo et al., (2000) used the MLR approach on the Principal Components (PC) of the tree-ring indices. The reconstructions in this approach are sensitive to the number of PCs retained, as shown by Hidalgo et al. (2000) in their comparison with traditional MLR-based reconstructions.

These reconstruction techniques, along with the suite of tree-ring information used, capture very well the variability of the observed flow, but the flow magnitudes generated by these techniques differ in the pre-observation period. This can be seen in seven reconstructions of Lees Ferry flows (Stockton and Jacoby, 1976; Hidalgo et al., 2000, Woodhouse et al., 2006) shown in Figure 2.

[2] Trees actually add a *volume* of new growth each year and that volume varies depending on environmental conditions and other factors such as disease. As the tree diameter increases, a given volume of growth will be contained in a thinner ring. Thus, this geometric effect must be accounted for in the creation of tree-ring indexes. Other effects also require compensation, such as autocorrelation caused by physiological factors such as energy storage.

Figure 2.
Seven reconstructions of Colorado River annual flow at Lees Ferry.



The divergence of streamflows among the various reconstructions during the pre-observation period is due to the use of different reconstruction calibration techniques, different tree-ring data treatment (i.e., standard versus *prewhitened* chronologies, the inclusion of lagged predictors), different tree-ring data, and different gage data (both the years used and the hydrologic time series itself) for the calibration. All of these are potential sources of the differences, and these differences should be expected. The fact that these different reconstructions do vary coherently is a testament to the robustness of the hydroclimatic signal in the trees.

The MLR technique suffers from four main drawbacks:

(i) It assumes the data to be normally distributed,
(ii) It assumes there is no correlation between the predictor variables,
(iii) Outliers can have an undue influence on the fitted MLR model, and
(iv) It can be used to model only annual flows at a single gaging site.

Several statistical methods are available to overcome these shortcomings. For a description of methods to overcome these limitations, the reader is referred to Gangopadhyay et al. (2008) and references contained therein.

Reconstructed annual flows are typically available only for one or a few points in a basin. In the Colorado River, annual reconstructions were first developed at Lees Ferry, because flows at that point represent 90% of the total basin flow, and because Lee Ferry, the point at which the Basin was divided in the Colorado River Compact, is only about one mile downstream.[3] Annual flows at a single point usually cannot be used directly in a typical water resources systems model, which simulate multiple tributary inflows on a monthly or finer time step. This limitation prevented the use of paleo-reconstructed flows in the most

[3] The Paria River joins the Colorado River between the gage at Lees Ferry and the compact point at Lee Ferry. The Paria has an average annual flow of about 20,000 acre-feet. The average annual flow at Lees Ferry during the observation period (1906-2005) is about 15.3 million acre-feet.

important model of the Colorado River system, the CRSS model, which required monthly inflows at 29 inflow points throughout the basin. Thus, a means of disaggregating reconstructed flows, both in time and space, was required before the flows could be used in the Colorado River Basin (and in most water resources models). The first use of paleohydrology to drive a complex water resources model in the Colorado River Basin was in the Severe Sustained Drought Study (Young, 1995, Harding et al., 1995), using disaggregated monthly flows developed by Tarboton (Tarboton, 1995). These disaggregating methods were refined by Tarboton et al. (1998) and by Prairie et al (2007).

The techniques of Prairie et al. (2007) involve first, for any reconstructed annual flow, randomly selecting an *analogue year* from a set of *K neighbor years* taken from the observed record. The K neighbors are selected from the observed record based on their ranked distance (in terms of flow) from the reconstructed annual flow. The K neighbors are weighted so that the closest neighbor has the greatest weight and the farthest has the least weight. The analogue year is randomly selected from the neighbors based on weight. The analogue year is used as a "template" for the spatial and temporal disaggregation. This approach was used to disaggregate reconstructed flows created by Meko et al., (2007) in support of model studies used in developing the Lower Colorado River Guidelines (Reclamation, 2007).

Prairie et al. (2008) developed another method that is well-suited for creating input data sets for complex water resources models. Prairie et al. (2008) used the information in the tree-ring chronologies to construct a stochastic model of annual *sequences* that was in turn used to construct sequences of annual model input. This type of *re-sequencing* approach does not model the individual flow magnitudes, but instead arranges *years* from the observation period in sequences that are statistically consistent with the information about hydrologic conditions (i.e. wet or dry year) contained in the tree-ring chronologies. In the first application of this method, for the Lower Colorado River Guidelines, sequences of annual flows at Lees Ferry were developed and subsequently disaggregated for use in the CRSS model. However, sequences of any type of input data that is associated with a year, including complex, structured data, can be constructed using this approach. This allows ensembles of traces of model input data to be constructed by re-sequencing annual model input data from the observation period.

The State's Request for Proposals and the subsequent contractual scope of work for CRWAS specifically calls for special consideration of the method developed by Prairie et al. (2008). During Task 6.1 evaluations, it became apparent that the statistical model already developed by Prairie for the use in developing the Lower Colorado River Guidelines could be utilized directly to re-sequence CRSS and CRDSS input data to meet the requirements of the Study. This approach is described below as part of three candidate methodologies, from which one methodology is recommended for development of the extended historical hydrology in Tasks 6.4 and 6.5.

Candidate Methodologies

Methods that are applicable to extend hydrologic data for use in CRWAS fall into two principal categories: parametric regression models that produce annual flow sequences and magnitudes at a single point, and non-parametric methods that produce sequences of years from which sequences of flows or structured model input data can be constructed[4]. To date, regression models of annual flows have been developed at Lees Ferry, approximately four other sites in the Upper Colorado River Basin, and at the Little Colorado River in the Lower Colorado River Basin. The water resources modeling that will be conducted as part of CRWAS will require monthly time series of flows at hundreds of locations in the basin. The current inventory of tree-ring chronologies is insufficient to create models at all of the required flow points and, even if sufficient tree-ring data existed, the magnitude of such an effort would probably be infeasible for CRWAS. Thus, a means of generating flows that reflect the information in the tree-ring chronologies at all of the CRSS inflow points and the CRDSS baseflow gages is required. For regression methods that generate annual flows, a disaggregation step will be required. The non-parametric re-sequencing methods include all model input points.

Based on our review of literature and our experience with development and use of reconstructed flows, we have identified for subsequent evaluation three methodologies that we believe are feasible for use in CRWAS to extend historical hydrology. These are:

(1) Parametric reconstruction of annual flows at Lees Ferry followed by disaggregation in space and time using Prairie et al. (2007) with resampling of flows below Lees Ferry.

(2) Parametric reconstruction of annual flows at Lees Ferry and at intra-state sub-basins (e.g., Yampa, Gunnison, etc.) followed by disaggregation in space and time using Prairie et al. (2007), with resampling of flows below Lees Ferry.

(3) Re-sequencing of annual model input data from the observation period record based on an existing model of annual sequences at Lees Ferry using Prairie et al. (2008).

These three methodologies are described the following paragraphs.

(1) Parametric reconstruction of annual flows at Lees Ferry followed by disaggregation using Prairie et al. (2007) with resampling of flows below Lees Ferry.

This approach was used by the Bureau of Reclamation in developing the Lower Colorado River Guidelines (Reclamation, 2007, Appendix N) where it is referred to as the *Direct Paleo* approach. The approach involves three steps:

(1) *Reconstruct Annual Flows at Lees Ferry.* Use a regression model based on tree-ring chronologies in the region, to reconstruct a time series of annual flows at Lees Ferry (Meko et al., 2007).
(2) *Disaggregate Upper Basin Flows.* For each annual flow in the time series resulting from Step 1, disaggregate the annual flow at Lees Ferry to twelve monthly flows at each point of interest in the Upper Colorado River basin using the method of Prairie et al. (2007). The disaggregation would be done at two scales, one for modeling of the Big River using CRSS, which requires 20 inflow points in the Upper Basin, and

[4] Non-parametric re-sequencing approaches often rely on regression-based reconstructions as the basis for establishing state transition probabilities. The approach by Gangopadhyay et al. (2008) can develop state sequences directly from observed flows and tree-ring chronologies.

one for modeling intra-state tributaries using CRDSS, which will require hundreds of inflow points.

(3) *Select Lower Basin Flows.* The CRSS model requires 9 inflow points in the Lower Basin below Lees Ferry. Because these flows do not contribute to Lees Ferry they cannot be generated by disaggregating the Lees Ferry flow. Instead, for each annual flow in the reconstructed time series developed in Step 1, the lower basin flows are set to those in the analogue year used in the disaggregation.

Advantages:

(1) Can generate Lees Ferry flows that have not occurred in the observed record. In any natural record, the range of flows (bounded by the maximum and minimum flows) can only increase as the period of record is extended by the passage of time and the operation of natural processes. Extension of the observed record by the use of paleohydrology would be expected to reveal new maximum and minimum flows, and parametric regression methods can be expected to generate flows outside the range of observed data, which is consistent with the behavior of natural time series of flows.

(2) A Parametric reconstruction of annual flows at Lees Ferry is available from Meko et al. (2007).

Disadvantages:

(1) The regression model used to reconstruct annual flows is based on the relationship between observed flows and tree-ring growth, and will not be valid under a changed climate. Thus, the reconstructed annual time series itself must somehow be adjusted to reflect projected climate change. The only information available to make this adjustment would be the observed streamflows, the climate-perturbed flows, and the reconstructed annual flow time series itself. It may be that the only feasible way to adjust the reconstructed flows would be to develop a regression model that would relate the magnitude of the climate-perturbed monthly flow at some point to the magnitude of the observed annual flow at that point for the same year (and perhaps to some antecedent flows). It is not clear that this approach would succeed and, even if it did, the additional statistical model would introduce additional uncertainty to estimates and would increase the effort of the work.

(2) Parametric regression methods can produce high or low flows that are physically unrealistic. It may be difficult to eliminate exceedingly high or low flows without changing variance in the flows as a whole.

(3) Parametric regression models require much effort and judgment to construct without over- or under-fitting.

(4) There is some uncertainty regarding the ability of the method to reconstruct extreme (high and low) flow magnitudes.

(5) Reconstructions from different regression models (utilizing different overlap periods or different tree-ring chronologies) will have different mean flows and different variability. There are several reconstructions available for Lees Ferry, so the choice of which reconstruction to use will to some extent impact results.

(2) Parametric reconstruction of annual flows at sub-basins (e.g., Yampa, Gunnison, etc.) followed by disaggregation using Prairie et al. (2007)

This approach is applicable to the tributary basins within Colorado that will be simulated using CRDSS; a separate reconstruction (using Method 1 or 3) would be required for simulation of the entire Colorado River Basin using CRSS. To develop reconstructions of annual flows on tributary basins in Colorado this approach would extend the work of Woodhouse et al. (2006) who reconstructed annual flows for the San Juan River at Bluff, Utah, and the Colorado River at Cisco, Utah. Additional reconstructions would be required for the Yampa River, White River, Gunnison River and Dolores River. The approach would require two steps:

(1) Develop reconstructions of annual flows at appropriate points on the Yampa, White, Gunnison and Dolores rivers. Use the approach of Woodhouse et al.,

(2) Disaggregate annual point flows to monthly flows at the flow points required by CRDSS. Use the approach of Prairie et al. (2007).

This Method 2 differs from Method 1 by relying on multiple reconstructions at points in the sub-basins in Colorado instead of relying on a single reconstruction at Lees Ferry. Thus, the required disaggregation within the sub-basins to the CRDSS baseflow gage points would be based on independent reconstructions rather than on the Lees Ferry reconstruction.

<u>Advantages</u>:

(1) As with Method 1, can generate Lees Ferry flows that have not occurred in the observed record. In any natural record, the range of flows (bounded by the maximum and minimum flows) can only increase as the period of record lengthens. Extension of the observed record by the use of paleohydrology would be expected to reveal new maximum and minimum flows, and parametric regression methods can be expected to generate flows outside the range of observed data, which is consistent with the behavior of natural time series of flows.

(2) May reveal changes during prehistoric periods in the degree to which variability of flow (and wet and dry spells) is correlated among the sub-basins. Changes in inter-basin correlation might exercise water supply systems in ways that exceed historical experience.

<u>Disadvantages</u>:

(1) This approach will suffer the same disadvantages as for Method 1.

(2) This approach will require development of new reconstructions of annual flows for some tributaries (Yampa, White, Gunnison, and Dolores rivers).

(3) There is no guarantee that there will be sufficient chronologies and useful correlations from which to develop the new annual reconstructions for tributaries.

(4) Preserving the space-time co-variability statistics among the different sub-basins will prove to be challenging because the disaggregation algorithm would have to be implemented in multiple stages and additional steps would be required to ensure that system-wide mass balance is preserved.

(3) Re-sequencing of years from observed record based on an existing re-construction at Lees Ferry using Prairie et al. (2008).

This approach provides a method to combine transient state information (wet/dry) from paleohydrologic reconstructions of hydrologic markers (e.g., streamflow, Palmer Drought Severity Index – PDSI) with observed streamflow records to generate ensembles of traces of streamflow. As described in Prairie et al. (2008) and applied by Reclamation (2007) in developing the Lower Colorado River Guidelines, this method was used to generate annual sequences of annual flow at Lees Ferry. In this application, it would be used to develop sequences of annual model input for either CRSS or CRDSS. It is described more fully below and in Prairie et al. (2008). The method proceeds in four steps:

(1) Working from an existing reconstruction at Lees Ferry, e.g. Meko (2007), calculate the probabilities of the transition from one hydrologic state (wet or dry) in one year to the another hydrologic state (or the same hydrologic state) in the subsequent year. With these transition probabilities, construct a Markov chain model of state transition (detailed in Appendix Section B).

(2) Run the Markov chain model developed in Step 1 to generate an ensemble of sequences of annual hydrologic states (e.g. a wet or dry year).

(3) Using a nearest neighbor re-sampling approach (detailed in Appendix Section B) conditioned on state and observed flow, generate an ensemble of sequences of years from the observed record. This sequence is simply a list of calendar years from the observed record, e.g. 1955, 1986, 1985, …, etc. that preserves the paleo-state transition statistics.

(4) Construct input data sets for CRSS and CRDSS from the sequences of years by traversing a sequence resulting from Step 3 and, for each year on the list, appending the entire annual input data set for that year onto an input file.

<u>Advantages</u>:

(1) The method is very efficient in its requirements of labor for model development. It is largely data-driven and requires only one parameter (the number of neighbors, i.e., similar years), and this parameter is usually estimated based on *heuristics* (e.g., Yates et al., 2003). Because of this efficiency, automated validation and skill assessment techniques can be used to evaluate the quality of a model. It is also much less time-consuming to develop new models based on projected streamflows that reflect climate change.

(2) The sequences used to extend observed flows can also be used to re-sequence the climate-perturbed flows if the assumption is adopted that climate change will not significantly change the *sequence* of wet/dry years (states). Although there is an expectation that a change in sequencing will likely occur, we are not aware of any definitive scientific evidence about the nature of those changes. Absent that information, it is our judgment that the best information available for CRWAS is the variety of sequencing found in the paleohydrology record. Adopting this assumption allows consistent sequencing to be used to extend the historical and climate-perturbed record. Maintaining a consistency of sequencing simplifies the analysis of the sensitivity of system impacts across impact analyses that use different assumptions about future climate.

(3) If the assumption described in (2) above is adopted, the same re-sequencing can be applied to both the “Big River” and CRDSS datasets and to all climate change conditions without construction of a new model.

(4) Even if new states must be assigned to years based on adjusted streamflows, the effort of doing so and the effort of re-running the re-sampling process to generate ensembles of state sequences and year sequences are not excessive.

(5) The adaptation we suggest here to the re-sequencing approach in Prairie et al. (2008) would build a sequence of complete annual model input data for individual years. By doing so, the method will preserve all spatial correlations and seasonal patterns (including shifts in the seasonal patterns of flows due to projected climate change).

(6) The method will not produce an annual condition that is not contained in the observed (or climate-perturbed) data set, so it cannot produce a physically unrealistic flow.

(7) The method naturally produces ensembles of flow traces, which are appropriate for risk and uncertainty analysis.

Disadvantages:

(1) The method adopts the assumption that spatial correlations are unchanged in pre-history and in the future. If spatial correlations actually vary substantially over time, the method may not exercise water resources systems as thoroughly as would be the case if variability in spatial correlations could be simulated.

(2) The method will not produce an annual condition that is not contained in the observed data set. If annual maximum and minimum flows actually were substantially different in pre-history, the method may not exercise water resources systems as thoroughly as would be the case if variability in flow extremes could be simulated.

Summary Evaluation of Candidate Methodologies

Both regression-based methods and non-parametric methods (and combinations of the two methods) have been used successfully to develop reconstructions of hydrology based on tree-ring records. Of the three methods we have identified as candidates for use in CRWAS, Method 1 and a variant of Method 3 have been applied in the Colorado River Basin. Method 2 has not been applied in the Colorado River Basin. Method 2 has a substantial technical risk due to the uncertainty about whether sufficient tree-ring chronologies, having adequate correlation with the tributary flows, exist to develop regression models for those tributaries for which reconstructed flows do not exist. It is not feasible within the resources of the CRWAS scope to collect new tree-ring chronologies if the data available in existing archives is insufficient. Method 2 also has other technical and level-of-effort disadvantages. The offsetting advantage of Method 2, that it might capture additional spatial variability from the paleo record, does not appear to compensate for the substantial technical risk and the higher level of effort that would result from adopting it.

Method 1 has been applied successfully. The vast majority of reconstructions of annual flows at a single point have been completed using regression techniques. These techniques are generally accepted among scientists, engineers and water managers.

Disaggregation of reconstructed annual flows is less common, and the application of the new technique of Prairie et al. (2007) has only been accomplished recently. However, there is little question about the technical feasibility of Method 1 for extending observed flows. It is less certain how Method 1 could be used to meet the requirement of CRWAS, that paleohydrology be used to blend information about drought (and wet spells) from the paleo record with projections of climate change. Difficulty with Method 1 arises because regression techniques depend on developing a mathematical model of the causal physical relationships that affect both tree growth and streamflow. Under a changed climate, these causal relationships would change, but there is little existing insight into how to quantify that change. Perhaps the most feasible way to apply Method 1 to adjust flows to reflect the impact of climate change would be to develop a regression equation, at a given point and for a given month of the year, to estimate the change in flow. It is not clear that this approach would succeed, and it would certainly add effort and complexity to the work.

Method 3 has been applied to develop reconstructions by re-sequencing observed flows (and model input data sets representing observed conditions), but has not, to our knowledge, been applied to develop model input or flow sequences based on flows altered to reflect the impact of climate change. However, doing so presents no mechanical challenges, and our assessment of the approach indicates that it will blend information available from the paleo record with projections of climate change without significant risk of introducing substantial bias.

Recommended Methodology

Based on the advantages and disadvantages of the candidate methodologies set out in our evaluation above, we suggest the use of Method 3 above, a re-sequencing of years from the observed record based on an existing re-construction at Lees Ferry adapted from Prairie et al. (2008). This method has the required level of accuracy, cost-effectiveness, good documentation, and compatibility with CRWAS goals. Method 3 has two disadvantages:

- It assumes that observed spatial correlations are the same in prehistory and in the future; and
- It cannot produce an annual condition that is not contained in the observed record.

We do not find that these disadvantages would impair the results of CRWAS. While we can safely assume that spatial correlations are not constant, there is little scientific evidence on which to quantify *past* changes in those correlations, and the hydrologic modeling approach we will suggest will represent future climate-driven changes in spatial relationships among streamflows in the study area. We also judge that the observed record contains a sufficient range of annual flows to represent single-year events, and the hydrologic modeling approach we will suggest will represent climate-driven changes in the range of annual flows.

Method 3 has three compelling advantages:

- It is efficient in terms of the level of effort required and in terms of computational requirements;
- If one assumption is adopted, the same sequences can be applied to observed flows and flows altered to reflect the impact of climate change, and to input data

from both the CRSS and CRDSS models. This will make it easier to understand the degree to which other assumptions regarding climate change and hydrology influence the results of the Study; and

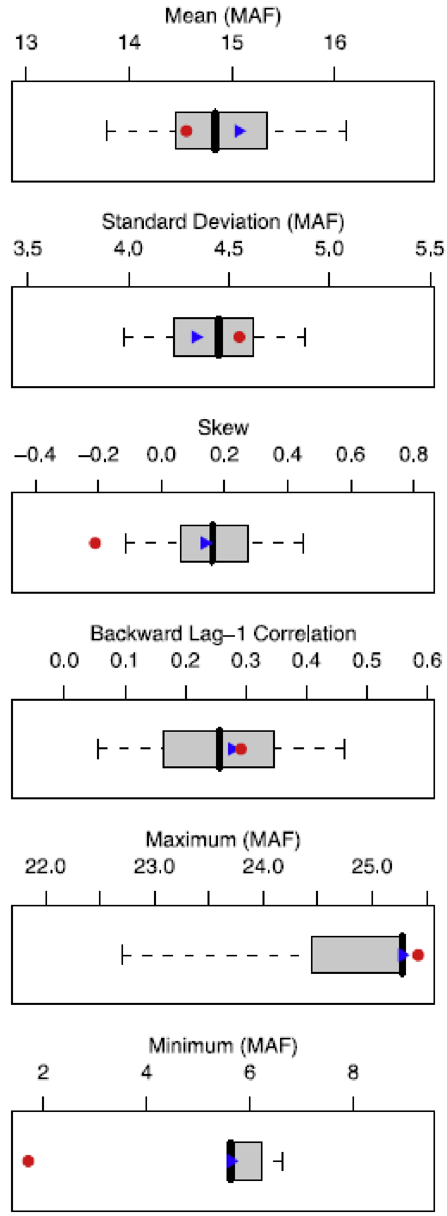
- The data, methods, and codes required to apply the approach exist and are familiar to the Study team.

A more complete description of the Prairie et al. (2008) method is provided in the Appendix.

Summary Statistics from Prairie et al. (2008) Paper

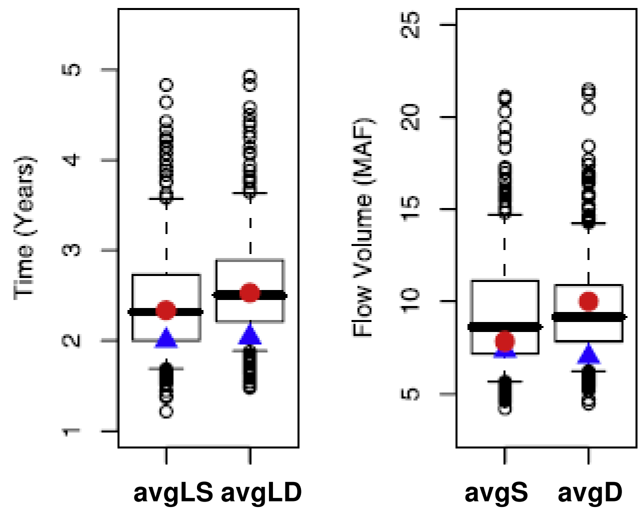
A suite of basic distributional statistics for annual flows at Lees Ferry were presented by Prairie et al, (2008), including (Figure 3) the annual (1) mean, (2) standard deviation, (3) coefficient of skew, (4) maximum, (5) minimum, and (6) lag-1 autocorrelation. Surplus and drought statistics, including the average length of surplus (avgLS), average length of drought (avgLD), average volume of surplus (avgS), and average volume of deficit (avgD) are shown in Figure 4. Surplus (drought) is defined as values above (below) a threshold, here the median of the observed record (1906-2005). The proposed framework by Prairie et al. (2008) is referred to as NPC in the following plots, and 500 simulations, each 100 years long, were used in generating the box plots. On each of the following plots the corresponding statistic for the observation period is shown as a blue triangle. The corresponding statistic for the entire set of reconstructed flows (by Woodhouse, et al., 2006) is shown as a red dot.

Figure 3.
Box plots of basic statistics from NPC simulations.
Source: Prairie et al. (2008), Figure 7



The box plots show the distribution of statistics generated using the techniques of Prairie et al., (2008). The reconstruction of paleo-flows by Woodhouse et al. (shown by the red dot) uses a parametric regression technique (methodology 1, above) which models flow magnitudes using a linear regression model. This approach can generate flows that are larger or smaller than those in the observed record. This is demonstrated in Figure 1, where the maximum in the Woodhouse et al. (2006) reconstruction is slightly higher than the maximum from the observed record. An even more dramatic demonstration of this effect is seen in the minimum flow in the Woodhouse et al. (2006) reconstruction, which is considerably smaller than the observed minimum flow (less than 2 MAF versus approximately 5.5 MAF). The tendency of the parametric regression technique to generate very low flows also explains the bias in the coefficient of skew of annual flows generated by that technique. The maximum and minimum flows found across all ensembles of the reconstructions made using the NPC are equal to the corresponding statistic from the observed record.

Figure 4.
Box plots of drought and surplus statistics from NPC simulations.
Source: Prairie et al. (2008), Figure 8



Statistics of the observation period are shown as blue triangles, and those of the Woodhouse et al. (2006) reconstruction are shown as red circles. Note that the median values of avgLS and avgLD from the NPC method and the corresponding values from the Woodhouse et al. (2006) reconstruction agree, which demonstrates that the NPC technique preserves the statistics related to flow sequences from the Woodhouse et al. (2006) reconstruction. Both are longer than for the observed record, which demonstrates that the paleo-period contained longer droughts and wet spells than does the observation period. Both methods indicate that droughts during the paleo-period were more severe than during the observation period, as the median value of avgD for the NPC method and the corresponding value for the Woodhouse et al. (2006) reconstruction are notably higher than the value for the observation period.

References

Gangopadhyay, S., B. Harding, B. Rajagopalan, J. Lukas, T. Fulp. “A non-parametric approach for paleohydrologic reconstruction of annual streamflow ensembles.” Water Resources Research, 2008, in review.

Haan, C.T. Statistical Methods in Hydrology, Iowa State University Press, 1977.

Harding, B.L., Sangoyomi, T.B. and Payton, E. A. Impacts of a Severe Sustained Drought on Colorado River Water Resources, Water Resources Bulletin (1995) 31(7).

Hidalgo, H., T. Piechota, J. Dracup. “Alternative principal components regression procedures for dendrohyrologic reconstructions,” Water Resources Research. 2000, 36(11).

Jain, S., and J. K. Eischeid. “What a difference a century makes: understanding the changing hydrologic regime and storage requirements in the Upper Colorado River Basin. Geophysical Research Letters, 2008. 35, L16401, doi:10.1029/2008GL034715.

Meko, D. M., and D. A. Graybill. “Tree-ring reconstruction of Upper Gila River discharge,” Water Resources Bulletin, 1995. 31(4).

Meko, D. M., M. D. Therrell, C.H. Baisan, and M. K. Hughes. “Sacramento River flow reconstructed to A.D. 869 from tree-rings.” Journal of the American Water Resources Association 2001 Volume 37(4).

Meko, D.M., C.A. Woodhouse, C.A. Baisan, T. Knight, J.J. Lukas, M.K. Hughs, and M.W. Salzer. “Medieval drought in the Upper Colorado River Basin.” Geophysical Research Letters 2007 34(5), L10705, doi:10.1029/2007GL029988.

Peters, K., G. Jacoby, and E. Cook. “Principal Components Analysis of Tree-Ring Sites.” Tree-Ring Bulletin 41. 1981.

Piechota, T., J. Timilsena, G. Tootle, and J. Hidalgo. “The Western U.S. Drought: How Bad is it?” EOS 2004 85(32).

Prairie J., B. Rajagopalan, U. Lall, T. Fulp. “A stochastic nonparametric technique for space-time disaggregation of streamflows.” Water Resources Research 2007 Volume 43, W03432, doi:10.1029/2005WR004721.

Prairie, J., K. Nowak, B. Rajagopalan, U. Lall, and T. Fulp. “A stochastic nonparametric approach for streamflow generation combining observational and paleoreconstructed data.” Water Resources Research 2008 Volume 44, W06423, doi:10.1029/2007WR006684.

Reclamation (U.S. Department of the Interior, Bureau of Reclamation). 2007. Final Environmental Impact Statement, Colorado River Interim Guidelines for Lower Basin Shortages and Coordinated Operations for Lake Powell and Lake Mead.

Shamir, E., D. Meko, N. Graham, and K. Georgakakos. “Hydrologic model framework for water resources planning in the Santa Cruz River, Southeastern Arizona.” Journal of the American Water Resources Association 2007 Volume 43(5).

Tarboton, D. G. Hydrologic Scenarios for Severe Sustained Drought in the Southwestern United States, Water Resources Bulletin (1995) 31(7).

Woodhouse, C. A., and J. T. Overpeck. “2000 years of drought variability in the central United States.” Bulletin of the American Meteorological Society 1998 79(12).

Woodhouse, C. A., S. Gray, and D. Meko. “Updated streamflow reconstructions for the upper Colorado River Basin.” Water Resources Research 2006 42(5), W05415, doi:10.1029/2005WR004455.

Yates, D., S. Gangopadhyay, B. Rajagopalan, and K. Strzepek. A technique for generating regional climate scenarios using a nearest-neighbor algorithm, Water Resources Research (2003) 39(7), 1199, doi:10.1029/2002WR001769

Young, R. A. Coping With a Severe Sustained Drought on the Colorado River: Introduction and Overview, Water Resources Bulletin (1995) 31(7).

Appendix

A. Literature Review
   i) Principal Publications
   ii) Other Relevant Publications
B. Description of Recommended Methodology
   i) Homogeneous Markov Chain Simulation with Resampling
   ii) Non-Homogeneous Markov Chain Simulation with Resampling
C. Glossary

A. Literature Review

i) Principal Publications

This section provides brief reviews of the most relevant publications. A complete list of publications that were considered is included in the next section, *Other Relevant Sources*.

Hidalgo, H., T. Piechota, J. Dracup. "Alternative principal components regression procedures for dendrohyrologic reconstructions," Water Resources Research*. 2000, 36(11).*

Assessing different procedures in Principal Component Analysis (PCA)-based regression and an alternative cross validation method for the selection of principal components. PCA is used to extract independent variables from tree-ring data, the significant Principal Components (PCs) are prescreened by an objective criterion before they are included in the regression model, stepwise regression, t-testing, or rotation can be used to determine the PCs for the final model. Cross validation is a way of independently testing models and selecting models with better predictive skill. When compared to traditional streamflow reconstructions in the Upper Colorado River Basin, results of this alternative PCA-based regression model show more intense drought periods.

Jain, S., and J. K. Eischeid. "What a difference a century makes: understanding the changing hydrologic regime and storage requirements in the Upper Colorado River Basin. Geophysical Research Letters*, 2008. 35, L16401, doi:10.1029/2008GL034715.*

Examining long-term variability and change in the Upper Colorado annual runoff volume-quantified as shifts in the mean, interannual variability, and persistence in a recent tree-ring based reconstruction extending back to 762AD. Used a simple model for reservoir storage requirement to show sensitivity to the changing hydrologic regime, with episodes of abrupt shifts toward significantly higher storage requirements, often not readily evident in runoff statistics.

Meko, D. M., and D. A. Graybill. "Tree-ring reconstruction of Upper Gila River discharge," Water Resources Bulletin*, 1995. 31(4).*

Paleohydrologic reconstructions of Gila River flows in Southern Arizona since 1663. Used multiple linear regression for reconstruction with ARMA pre-whitening for filtering and Akaike Information criterion for order selection. Found that the 20th century is unusual for its clustering of high-discharge years, severity of multiyear drought and amplification of low-frequency discharge variations. Also noted connection to low flow time periods in the upper Colorado River Basin.

Meko, D. M., M. D. Therrell, C.H. Baisan, and M. K. Hughes. "Sacramento River flow reconstructed to A.D. 869 from tree-rings." Journal of the American Water Resources Association 2001 Volume 37(4).

Paleohydrological reconstruction for the Sacramento River derived by regressing log10 flow on principal components of tree-ring indices. Monte Carlo analysis of reconstructed n-year running means shows that the gauged record contains drought extremes for periods from 6 to 10 years in length but not for periods outside this range. Also noted persistently high or low flows over 50-year periods for some parts of the long term history

Meko, D.M., C.A. Woodhouse, C.A. Baisan, T. Knight, J.J. Lukas, M.K. Hughs, and M.W. Salzer. "Medieval drought in the Upper Colorado River Basin." Geophysical Research Letters 2007 34(5), L10705, doi:10.1029/2007GL029988.

Extending the Colorado River at Lee Ferry chronology into the Medieval Climate Anomaly. The most extreme low frequency feature of this time period (A.D. 762-2005) is a drought in the mid 1100s characterized by a decrease of more than 15% in mean annual flow averaged over 25 years and the absence of high annual flows over a period of six decades. This drought is consistent with dry conditions in the Great Basin and Colorado plateau at the same time but there are notable regional differences in intensity.

Peters, K., G. Jacoby, and E. Cook. "Principal Components Analysis of Tree-Ring Sites." Tree-Ring Bulletin 41. 1981.

Using a principal components model to analyze tree-ring data for site heterogeneity and to develop chronologies that correlate better with local climate data. Used data from Forest-tundra ecotone of the Yukon Territory and found that more climatic information can be extracted using this technique.

Piechota, T., J. Timilsena, G. Tootle, and J. Hidalgo. "The Western U.S. Drought: How Bad is it?" EOS 2004 85(32).

Used Principal Component Analysis and regression to reconstruct stream flows for 500 years using tree-ring data. Found that the period from 1999 through 2004 is the 7th worst drought in the 500 year tree-ring history. The worst drought occurred in the late 1500s and lasted for 20 years.

Prairie J., B. Rajagopalan, U. Lall, T. Fulp. "A stochastic nonparametric technique for space-time disaggregation of streamflows." Water Resources Research 2007 Volume 43, W03432, doi:10.1029/2005WR004721.

A nonparametric method for space-time disaggregation of streamflow is presented in this paper. The disaggregation is a two-step process where temporal disaggregation (e.g., annual to monthly) is followed by spatial disaggregation (e.g., from an index gage to sub-basin gages). Streamflow ensembles for each gage is developed using k-nearest neighbor bootstrap. This method preserves all space-time co-variability statistics of streamflow, and was demonstrated with an application to the Upper Colorado River Basin. In addition, this method was particularly successful in preserving multi-modal probability distribution of flow which has been observed in historical streamflow time series.

Prairie, J., K. Nowak, B. Rajagopalan, U. Lall, and T. Fulp. "A stochastic nonparametric approach for streamflow generation combining observational and paleoreconstructed data." Water Resources Research 2008 Volume 44, W06423, doi:10.1029/2007WR006684.

A method for generating streamflow scenarios by combining paleohydrologic reconstructions and historical streamflow. A nonhomogeneous Markov chain model developed with the paleo data is used to generate the system state (wet or dry) and a nonparametric k-nearest neighbor time series bootstrap of the historical stream flow data conditioned on the system state is used to generate stream flow scenarios.

Shamir, E., D. Meko, N. Graham, and K. Georgakakos. "Hydrologic model framework for water resources planning in the Santa Cruz River, Southeastern Arizona." Journal of the American Water Resources Association 2007 Volume 43(5).

Stream flow ensembles for the upper Santa Cruz River were used to evaluate risk of ground-water depletion under different water utilization scenarios. The model includes, stochastic generation of hourly precipitation scenarios, transformation of generated precipitation into daily streamflow, surface-water to ground-water interactions and inter-annual variability based on either tree-ring reconstructed annual precipitation or historic streamflow gauge data. Regression modeling was used for paleo-reconstruction and explained 57% of the variance for the winter; regressions for other seasons were not as successful.

Woodhouse, C. A., and J. T. Overpeck. "2000 years of drought variability in the central United States." Bulletin of the American Meteorological Society 1998 79(12).

A comprehensive review of paleoclimatic literature for the past 2000 years reveals that the most severe droughts in the 100 year instrumentally recorded history are surpassed in duration and spatial extent by earlier droughts recorded by historical documents, tree-rings, archaeological remains, lake sediment and geomorphic data. The study contains a list of references with proxy data variables and locations that were used for the study. Based on these findings and GCM predictions it is likely that this area will experience future droughts greater than any on record in the past hundred years. As such water resource management plans should take into consideration the full variability of the system not just the past 100 years.

Woodhouse, C. A., S. Gray, and D. Meko. "Updated streamflow reconstructions for the upper Colorado River Basin." Water Resources Research 2006 42(5), W05415, doi:10.1029/2005WR004455.

Paleohydrologic reconstructions of streamflow data for four gauges in the Upper Colorado River Basin based on expanded tree-ring data. Reconstructions were done using multiple linear regression and results account for 72-81% of the variance in the gauge records. Findings confirm that Colorado River allocations are based on one of the wettest periods in the past 500 years.

World Data Center for Paleoclimatology Website,
http://www.ncdc.noaa.gov/paleo/treering.html

National Climate Data Center (NCDC) website which contains links for tree-ring chronologies all over the US and the world.

ii) Other Relevant Publications

The following publications have been reviewed but are not summarized here.

Briffa, K.R., P.D. Jones, and F.H. Schweingruber. Tree-Ring Density Reconstructions of Summer Temperature Patterns across Western North America since 1600, Journal of Climate. 1992 5(7):735-754.

Cayan, D. R. and R. H. Webb. “El Nino / Southern Oscillation and Streamflow in the Western U.S.” Historical and Paleoclimatic Aspects of the Southern Oscillation 1992. 29-68.

Cook, E.R. and L.A. Kairiukstis. “Methods of Dendrochronology.” Kluwer Academic. 1990. 394 pp.

Graumlich and Brubaker. “Reconstruction of annual temperature (1590-1979) for Longmire, Washington, derived from tree-rings.” Quaternary Research. 1986. 25(2): 223-234.

Gangopadhyay, S., B. Harding, B. Rajagopalan, J. Lukas, T. Fulp. “A non-parametric approach for paleohydrologic reconstruction of annual streamflow ensembles.” Water Resources Research. 2008. In review.

Gray, S.T., J.L. Betancourt, C.L. Fastie, S.T. Jackson. “Patterns and sources of multidecadal oscillations in drought-sensitive tree-ring records from the central and southern Rocky Mountains.” Geophysical Research Letters 2003. 30(6).

Hidalgo, H.G., J.A. Dracup, G.M. MacDonald, and J.A. King. “Comparison of tree species sensitivity to high and low extreme hydroclimatic events.” Physical Geography. 2001. 22, 115-134.

Loaciga, H.A., L. Haston, J. Michaelson. “Dendrohydrology and long-term hydrologic phenomena.” Reviews of Geophysics. 1993. 31(2): 151-172.

Meko D. M., C.W. Stockton and W.R. Boggess. “The tree-ring record of severe sustained drought.” Water Resources Bulletin. 1995. 31(5): 789-801.

Meko, D.M. and C.A. Woodhouse. “Application of streamflow reconstruction to water resources management.” Accepted for publication in: Tree-rings and Climate: Sharpening the Focus. eds. H.F. Diaz and M.K. Hughes. 2007. Kluwer/Springer.

Michaelsen J., H.A. Loaiciga, L. Haston and S. Garver. “Estimating drought probabilities in California using tree-rings.” California Department of Water Resources Report B- 57105. Department of Geography, University of California, Santa Barbara. 1990.

Schulman E. “Tree-ring hydrology of the Colorado River basin.” University of Arizona Bulletin XVI(4). University of Arizona, Tucson. 1945. 51 p.

Schulman, E. “Tree-ring indices of rainfall, temperature, and river flow.” In Compendium of Meteorology, American Meteorological Society. Boston. 1951. pp 1024-1029.

Schulman, E. “Dendroclimatic Changes in Semiarid America.” University of Arizona Press. Tucson, AZ. 1956. 142 p.

Stockton, C.W. “Long-term streamflow records reconstructed from tree-rings.” Papers of the Laboratory of Tree-Ring Research 5. University of Arizona Press. Tucson, AZ. 1975. pp. 1-111.

Stokes and Smiley. “An Introduction to Tree-ring Dating.” University of Arizona Press. 1968. 73 pp.

Tarboton, D. G., A. Sharma, and U. Lall (1998), Disaggregation procedures for stochastic hydrology based on nonparametric density estimation, Water Resour. Res., 34(1), 107–119.

B. Description of Recommended Methodology

The proposed approach is based on theories of stochastic hydrology and paleohydrology covered in the reviewed and referenced literature. Two approaches are available to use Markov chain models to develop alternative hydrologic[5] sequences from historical observed and reconstructed time series for the CRWAS. These are re-sequencing of observed hydrology based on a homogeneous (static) transition probability derived from observed flow states and re-sequencing of observed hydrology based on non-homogeneous (varying) transition probability derived from observed flow states. The first approach is referred to as the *Homogeneous Markov Chain (HMC) Simulation with Resampling*, and the second approach is referred to as the *Non-homogeneous Markov Chain (NHMC) Simulation with Resampling*.

The HMC approach was not used by Prairie et al. (2008) and is not recommended for use in CRWAS. However, it is presented here to provide background that will facilitate a better understanding of the NHMC approach. The NHMC approach to re-sequencing flows was the approach used by Prairie et al., (2008) and is recommended for use in CRWAS.

i) Homogeneous Markov Chain Simulation with Resampling

Markov chain simulation using a single (hence the notion of homogeneity) transition probability matrix is given in Haan (1977). In application to the Colorado River the reconstructed flows at the Lees Ferry gage are classified into two states – dry and wet, based on a flow threshold such as the mean flow at the gage. Years where the annual flow is below the threshold flow are classified as dry years and years where the annual flow is equal to or above the threshold flow are classified as wet years. Using the sequence of dry and wet years, four state transition probabilities are developed. These are, (i) dry-dry, (ii) dry-wet, (iii) wet-dry, and (iv) wet-wet. In this application, the transition probabilities are calculated based on the entire record of reconstructed and observed flows. The historical Lees Ferry naturalized flow time series is also classified into two states, and the years are placed in two bins according to their state.

These transition probabilities are used in the Markov chain algorithm (e.g., Haan, 1977) to simulate dry/wet sequences and generate ensembles of time series of dry/wet states (state traces). Each state trace is constructed as follows: An initial state is selected randomly. The next trace is generated based on the probability of transition from the first state. For example, if the first state is dry, the relevant transition probabilities would be dry-dry and dry-wet. This process is repeated until a trace of the desired length is obtained. Additional traces are constructed until an ensemble of traces of the desired size is obtained. Flow traces are generated based on each state trace to create an ensemble of flow traces. Each state trace is traversed from start to end, and for each state in the sequence an historical year is sampled with replacement from the dry or wet bin, depending on the state in the trace. So, each of the flow traces will be a trace consisting of a set of observed years in a sequence determined by the Markov chain model as captured in the corresponding state trace. Flow data corresponding to the resequenced years can be used to force water allocation or other impact models or for statistical analysis of droughts or other time series statistics.

---

[5] Re-sequencing techniques can be applied to streamflows, water use, or meteorological data, hence the general term “hydrology” is used to demonstrate that point.

The reader is referred to Haan (1977) for a complete description of the HMC approach.

ii) Non-homogeneous Markov Chain Simulation with Resampling

Application of the non-homogeneous Markov Chain simulation with resampling with an application to the Lees Ferry gage ensemble streamflow simulation is described in the paper by Prairie et al. (2008). Conceptually, this approach provides a method to combine transient state information (wet/dry) from paleohydrologic reconstructions of hydrologic markers (e.g., streamflow, Palmer Drought Severity Index – PDSI) with observed streamflow records to generate streamflow ensembles.

The basic concept of the NHMC approach as developed by Prairie, et al. is very similar to the conventional homogeneous Markov Chain algorithm (e.g., Haan, 1977) for simulating hydrologic states except that transition probabilities are calculated using a moving window framework. In application to the CRWAS, we suggest that the NHMC model be based on the paleoreconstructed Lees Ferry streamflows for the period 1490-1905 (Woodhouse et al., 2006) as was done by Prairie, et al. (2008). The NHMC approach in Prairie et al. (2008) is a two-step methodology. The first step is the modeling hydrologic state and the second step is modeling flow magnitudes. A brief description of these steps is given below. The reader is referred to the Prairie et al. (2008) paper for details of the algorithm.

Modeling Hydrologic State

Similar to the homogeneous Markov chain approach, reconstructed streamflow at the Lees Ferry gage (e.g. those by Woodhouse et al. [2006][6]) will be used to derive a set of two-state non-homogeneous transition probabilities following the Prairie et al. (2008) algorithm. The first step in the process is to assign a dry or wet state to each of the years of the reconstruction period 1490 through 1905. Based on a flow threshold, for example, the mean or median of observed flow (1906-2005) for the Lees Ferry gage, one would categorize each year as dry or wet based on the annual flow for that year in the paleoreconstructed flows. If the reconstructed flow in any year is below the threshold flow, then that year will be assigned state dry, else the year will be assigned to state wet. Thus the paleoreconstructed flow time series will be transformed into a binary time series (dry = 0; wet = 1).

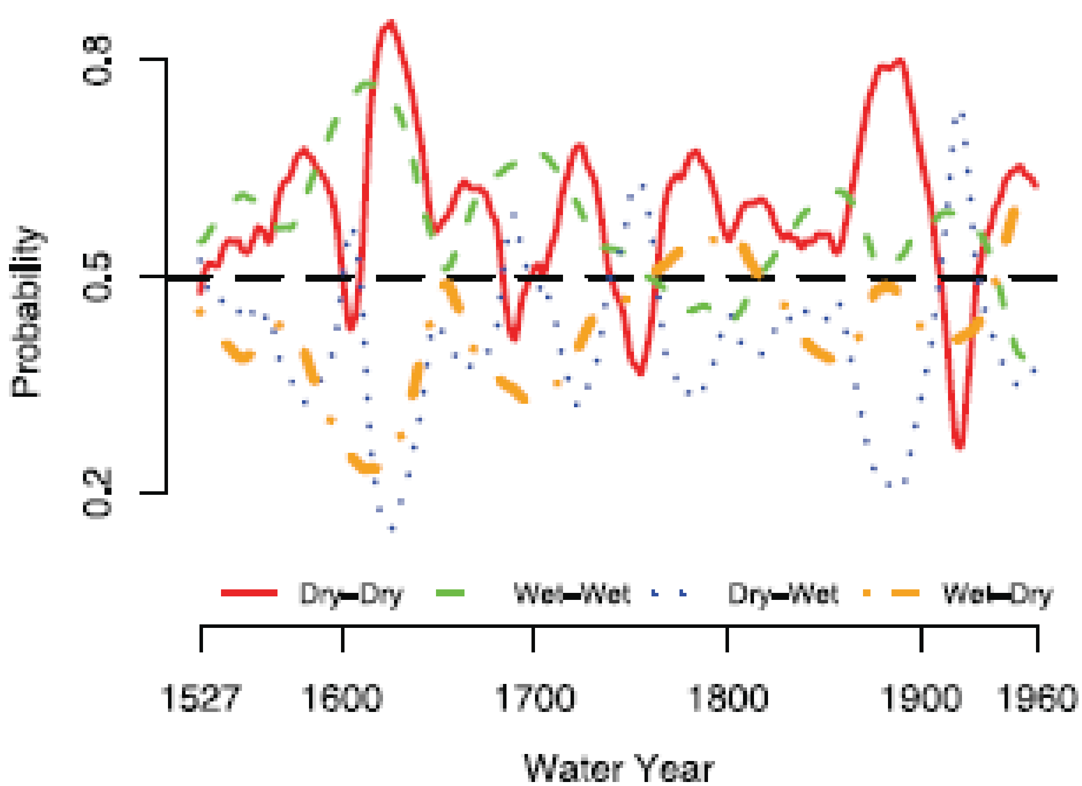
This binary time series will be used to develop the transient transition probability matrix corresponding to four transitions, (i) dry-dry; (ii) dry-wet, (iii) wet-dry, and (iv) wet-wet. In the NHMC, the transition probability matrix is not static but changes with time over the length of the reconstructed record. A moving window of selected width centered on each of the years (e.g. 1490 through 1905 for the Woodhouse reconstruction, and accounting for end effects) is used to develop the four transition probabilities: dry to dry, dry to wet, wet to dry, and wet to wet. The window width is derived using objective least squares cross-validation criteria (LSCV). The derivation of the transition probabilities is completed using the non-parametric kernel smoothing algorithm and the relevant equations in the Prairie et al. (2008) paper are equations (1) through (3). The LSCV criterion is given in equation (4)

[6] Another paleohydrologic reconstruction of Lees Ferry flows has been developed by Meko et al. (2007) for the period 762-1905. A non-parametric paleohydrologic reconstruction of Lees Ferry flow for the period 1400-1905 has been developed by Gangopadhyay et al. (2008). This paper is currently in review following revision.

of the Prairie et al. (2008) paper. The LSCV criterion is used in the technique adopted in Prairie et al. (2008) to determine the optimal window widths to calculate the transition probabilities. It should also be noted that in estimating the transition probabilities, a quadratic smoothing function is also applied.

The non-homogeneous transition probabilities provide an opportunity to capture the fact that certain extended periods have a higher (lower) probability of transitioning from dry (wet) to wet (dry) states. This is shown in Figure B-1. Note that during the late 19th Century the probability of a dry-dry transition was well above 50%, while the probability of a dry-wet transition (the complementary transition) was well below 50%. The relative magnitude of these transition probabilities reversed in the early part of the 20th Century. Similar changes are apparent throughout the period shown.

Figure B-1.
Plots of Transition Probabilities.
Source: Prairie et al. (2008), Figure 6



Modeling Flow Magnitudes

The streamflow magnitudes in a dataset reconstructed using the approach from Prairie et al. (2008) are modeled based on the observed data, and conditioned upon the hydrologic state simulated from the paleoreconstructed streamflow data (refer to section above). This model of annual flow magnitudes is given in equation (5) of the Prairie et al. (2008) paper as a conditional probability density function (PDF). Flow magnitudes are simulated based on a set of neighbors. The neighbors are a set of years which are constrained in three respects with the simulated year – (i) flow magnitude, (ii) hydrologic state, and (iii) hydrologic state transition. The conditional PDF of flow is generated using this resampling procedure and is referred to as K-NN (K nearest neighbor) bootstrap approach.

The NHMC algorithm randomly selects an epoch of given length (100 years in this case), randomly selects the initial state (dry or wet), and marches through the transitions (Haan, 1977) to get the simulated 100-year state time series. The difference from the HMC is that the transition probability matrix for NHMC changes as the algorithm moves from year to year in the selected epoch. The historical Lees Ferry naturalized flow time series for NHMC is also classified into two states as in the HMC. The NHMC algorithm as developed by Prairie, et al. (2008) employs a k-nearest neighbor algorithm to select similar flow years once the state transition has been simulated. In this case, CRSS natural flow data will be available for the period 1906-2006. Complete CRDSS flow data is available for the period 1908-2007. Gridded observed climate data are available from 1950 through 1999.[7] The intersection of these two datasets covers the period 1950-1999, so only years from this period will be used to model streamflow magnitudes. Therefore, each of the generated traces will consist of a 100-year re-sequence of the years 1950-1999. Flow data corresponding to the re-sequenced years (from either the CRDSS or CRSS datasets) will be used to populate the water allocation model.

The NHMC algorithm with resampling was successfully applied in the recent USBR EIS study (see Reclamation, 2007, Appendix N), and provides an elegant approach to combine observed hydrology and paleohydrologic state information in a consistent stochastic hydrology simulation framework. The implementation of the NHMC algorithm is given in section 3.3 of the Prairie et al. (2008) paper.

[7] The 1950-1999 data set was used for bias correction of downscaled GCM projections and is preferred over similar data sets with longer records.

C. Glossary

Autocorrelation— Measure of repeated pattern in a given dataset. For example, lag-one autocorrelation in a time-series can be estimated by calculating the correlation coefficient by lagging or leading the time-series by one time unit.

Baseflow—As used in the Colorado Decision Support System. A flow volume or flow rate that represents streamflows absent man's influence, including diversions, return flows, reservoir operations and pumping. Synonymous with "naturalized flow". If all of anthropogenic influence is removed, baseflows are often called virgin flows or natural flows.

Chronology— The time-series of annual ring-width indices derived from measuring actual tree-ring widths and filtering for age and geometry related growth trends for one site is called a chronology. Also see, dendrochronology.

Dendrochronology— The analysis of the annual growth rings of trees, leading to the calculation of significant indices of climate and general chronology of the past. The width of a tree-ring is determined by the temperature and/or moisture that prevailed during the year of its formation. Since stress from temperature and/or moisture variations reduces the width of the seasonal growth of a tree ring, dendrochronology has important application in the study of long-term climatic variations.

Ensemble— The concept of an ensemble comes from the fact that the results from repeating an experiment or a model simulation on a stochastic system may be expected to give different outcomes with each repetition. An ensemble is a collection of a number of outcomes of repeated experiments or simulations, each of which represents a possible state that the real system might be in. Variation of the results across the ensemble members gives an estimate of uncertainty. In climate science, ensembles of climate projections made with the same GCM but different initial conditions only characterize the uncertainty associated with internal climate variability, whereas multi-model ensembles including simulations by several models also include the impact of model differences.

Heuristic—A commonsense rule or rules adopted to solve a problem, e.g. "a rule of thumb."

Naturalized flow— Refer to baseflow.

Nearest-neighbor resampling—As used herein, nearest neighbor resampling is a form of regression based on pattern matching. Given a predictor object and a distinct population of values or objects, with both the predictor object and the population described or associated with the same set of attributes, a value from the population can be associated with the predictor object based on the similarity of the attributes between the predictor object and the members of the population. This similarity can be measured by the distance, in the space defined by the attributes, from the predictor object to each member of the population. In a case with two attributes, the distance between two points *i* and *j* is calculated as the conventional Euclidean distance, $d = ((x_i - x_j)^2 + (y_i - y_j)^2)^{1/2}$, but distance can be calculated in other ways. Rather than pick the member of the population that is nearest the predictor object, a member can be randomly selected from a subset of the population that is made up of the *k* members, called *neighbors*, that are nearest the predictor, a method referred to as the *k-nearest neighbor* or *k-NN* method. Selection from the neighbors can be unweighted, or can use a weight based on distance.

Non-parametric statistics—Non-parametric statistical techniques do not rely on the assumption that the data are drawn from a given probability distribution, and therefore the methods do not depend on a population fitting any parameterized distributions (e.g. a normal distribution). A histogram is an example of a non-parametric model. Non-parametric models are not truly free of parameters—a histogram has at least one parameter, which is the width of the bins—instead the term is meant to distinguish the method from conventional statistics where a distribution and the structure of the model must be assumed *a priori*. The structure of a non-parametric model is determined from data

Pre-whitening— Tree-ring chronologies typically contain significant low-order (short-term) autocorrelation, which is either retained in the "standard" chronologies, or removed using autoregressive modeling in a process called pre-whitening to produce "residual" chronologies.

Principal Component Analysis (PCA)—Principal Component Analysis is a transformation technique that is often used to reduce the dimensionality of data. That is, it can be used to extract and select a few *principal components* from a set of data explained by many variables. It can be thought of as a tool to reveal the internal structure of complex data that best explains the variance in the data. In the Colorado River Basin there may be many tens of tree-ring chronologies. Instead of using a statistical model that involves all of the relevant tree-ring chronologies, PCA can be used to generate the principal components from the entire data set, and then a model (e.g. a regression equation, or a non-parametric model) can be based on a few of the principal components. The number of PCs used in a model is usually the number that explains approximately 90% of the total variance in the data.

Resampling— Repeatedly selecting a sub-set of values from a set of values typically with replacement following a weight distribution.
Residual Chronology— Tree-ring chronologies for which low-order autocorrelation is removed typically using autoregressive modeling.

Sampling with replacement— Sampling with replacement" means that the sampled item. The term comes from physical sampling, e.g. numbered balls in an urn, where the sampled ball would be physically replaced in the urn to preserve the population statistics. In this case an historical year is allowed to remain in the population so it can be sampled again—it is "replaced". "Resampling" is used as a short-hand term for sampling with replacement.

Sequence—As used here, a particular ordering of annual flows. The observed record is one sequence. Re-ordering the observed annual flows creates a new sequence. Used instead of "trace" to emphasize that the annual flow magnitudes have not been changed.

Standard Chronology— Tree-ring chronologies for which low-order autocorrelation is retained.

Stochastic Process— A stochastic process is one that involves a random component (e.g. rainfall) with the result that there is more than one possible future outcome from the process as it evolves over time.

Trace— A single set of values from resampling, also referred to as a realization.