STOCHASTIC NONPARAMETRIC FRAMEWORK FOR BASIN WIDE

STREAMFLOW AND SALINITY MODELING: APPLICATION FOR THE

COLORADO RIVER BASIN

By

JAMES ROGER PRAIRIE

B.S., State University of New York

College of Environmental Science and Forestry, 1993

M.S., University of Colorado, 2002

A thesis submitted to the faculty of the Graduate School of the University of

Colorado in partial fulfillment of the requirement for the degree of

Doctor of Philosophy

Department of Civil, Environmental, and Architectural Engineering

2006

This thesis entitled:

Stochastic Nonparametric Framework for Basin Wide Streamflow and Salinity

Modeling: Application for the Colorado River Basin

written by James Roger Prairie

has been approved for the

Department of Civil, Environmental, and Architectural Engineering

_____

Balaji Rajagopalan

_____

Edith Zagona

_____

Kenneth Strzepek

Date _____

The final copy of this thesis has been examined by the signatories, and we find that

both the content and the form meet acceptable presentation standards of scholarly

work in the above mentioned discipline.

Prairie, James Roger (Ph.D., Civil, Environmental, and Architectural Engineering)

Stochastic Nonparametric Framework for Basin Wide Streamflow and Salinity

Modeling: Application for the Colorado River Basin

Thesis directed by Professor Balaji Rajagopalan

The Colorado River system recently experienced the worst five year drought in 100 years of measured streamflows. The vast network of reservoirs that sustain and fuel development were drawn below 50% of system capacity to historic lows. The dry period also impacted water quality, increasing salinity concentration as a result of low flows. Paleo reconstructions of streamflow in the basin indicate that such dry spells are not uncommon. Interestingly, a compact to share water resources among the basin states, developed during wet periods of the early 1920s, is under stress as the states confront economic growth under limited water resources.

Given these factors, the key question for this research is how to plan for effective and sustainable management of water resources in the basin? This requires two key components, (i) a robust framework to generate realistic basin wide streamflow and associated salinity scenarios and (ii) a decision support model to evaluate operating policy alternatives for efficient management and sustainability of water resources in the basin. To achieve this, three inter-related modules were developed including. (i) A stochastic nonparametric model for basin wide streamflow generation based on historic observations conditioned on paleo reconstructed streamflow. This new model is data driven and improves considerably upon traditional approaches, besides being simple, robust and flexible. A nonhomogeneous Markov Chain based approach to

combine hydrologic state information (i.e., wet or dry) from the paleo streamflow and flow magnitude from the historic data is a novel method to combine the strengths of these two different data sets. Together, these constitute a significant methodological contribution from this research. (ii) A basin wide stochastic model for generating salinity scenarios extending the nonparametric flow model for salinity simulation. (iii) A realistic decision model of the basin that evaluates policy alternatives under various flow and salinity scenarios.

This comprehensive framework provides the ability to generate scenarios of basin wide streamflows and salinity that are statistically consistent, realistic and also incorporate paleo information. These scenarios are used in a long-term planning model of the Colorado River Basin to evaluate two alternate reservoir operation policies on various system risk and reliability estimates.

DEDICATION

For my loving wife Annemarie. Her never ending support and encouragement helped me to believe in myself and discover I can complete anything I put my heart in.

ACKNOWLEDGEMENTS

CONTENTS

FIGURES

TABLES

CHAPTER 1

INTRODUCTION

1.1 Motivation

Over the past recent years, the western United Stated has been reeling under a severe dry spell. The Colorado River basin experienced the most severe five-year drought in the observed record from 2000-2004. While this region is generally semiarid and is no stranger to dry periods, the recent dry spell coupled with increasing population, economic growth, and competing water demands is having a decisive impact on the socio economic well being of the region. The vast networks of reservoirs that sustain and fuel development are at alarmingly historic lows not seen since their filling periods (Fulp, 2005), thereby threatening development. The dry period is also having an impact on the water quality; low flows in the streams can lead to increased concentration of salinity, thus degradation of water quality. Furthermore, compacts and decrees to share water resources among the western states developed during the wet periods of the early 1920s and 1930s, are especially under stress, as the states confront growth under limited water resources (Kuhn, 2005). Though increased flows in 2005 brought some relief, it remains to be seen if the relief will persist.

There is increasing evidence that the year to year variations in the western United States hydro-climate are driven by large scale climate features (e.g., ENSO, PDO, AMO, etc.). Hoerling and Kumar (2003), from climate model simulations, show tantalizing evidence that the dry spell could be a result of cooler than normal tropical Eastern Pacific and warmer than normal tropical Western Pacific (a La Nina pattern)

and Indian oceans. This long break from the El Nino active period of 1980s and 1990s could be a strong factor for the dry spell. Adding to these woes are recent indications that the annual cycle in precipitation is shifting earlier in the year for the western United States, perhaps caused by global climate change. In particular, the spring warmth has been occurring earlier, coupled with decreased winter snow pack. This precipitates a reduced inflow into the rivers, as a majority of them are driven by snowmelt (Regonda et al., 2005; Mote, 2003; Cayan et al., 2001).

Further studies have examined trends in soil moisture, runoff, and drought characteristics over the United States and found increasing trends for moisture and runoff and decreasing drought trends across much of the country (Andreadis and Lettenmaier, 2006). In this study, the exception that stood out was in the interior West and Southwest where trends indicated a significant increase in drought duration and severity.

Given the high stakes, the Bureau of Reclamation (Reclamation), the federal agency charged with managing water resources in the Colorado River system, is confronted with three immediate questions for efficient water resources planning and management:

1. How unusual is the current dry spell?

2. How can we simulate stream flow scenarios that are consistent with the current dry spell and other realistic future conditions?

3. What is the impact of these conditions on salinity?

Answering the first question will place the current dry period in context; the second seeks to provide a robust framework that can incorporate such information in

generating realistic streamflow scenarios, which will drive planning and management models; the third addresses the sensitivity of salinity levels to drought conditions in the basin, a key variable in maintaining the water quality standards in the Colorado River system.

1.2   State of present modeling methods

With these pressures in the Colorado River basin, water managers are dependant on basin scale models to understand the impact of natural flow variability on key decision variables relating to both streamflow and salinity concentration. Reclamation uses a long-term planning model termed the Colorado River Simulation System (CRSS) (DOI, 1987). Streamflow scenarios, typically, obtained from stochastic models, drive the CRSS allowing basin managers to understand the impact of operational policies on various aspects of the system, both in terms of salinity and water quantity. Alternative policies can also be compared and risk and reliability estimates of the system can be evaluated, all of which is useful in long-term planning and management. Clearly, the stochastic model needs to generate realistic and statistically consistent scenarios for effective management. Realistic scenarios include flows within or just beyond the range of flow magnitudes in the observed record. Where, statistically consistent scenarios share similar distributional properties with the observed record.

Currently, Reclamation depends on the index sequential method (ISM) (Ouarda et al., 1997) to generate stochastic hydrologic scenarios or traces. This stochastic method entails a sequential block bootstrap of the observed data where the block size is determined by the simulation or planning horizon. Bootstrapping is one of several data resampling techniques described by Efron (1982). Essentially, the ISM loops

through the historic data sequentially picking a block of flows from the observed record so no new sequences (i.e., wet and dry periods) are generated and thus, limited variability arise in the simulations. This is a severe limitation of the ISM especially when using it to answer the research questions posed earlier. Scenarios of salinity are also modeled using the ISM and suffer from similar drawbacks. Clearly, an improved, robust, and flexible method is required.

Recent paleo reconstructions of streamflow for the pre-observed period, from tree-ring chronologies (Woodhouse et al., 2006) show droughts of greater magnitude and duration occurring periodically, indicating that the recent drought is not unusual in the basin. Clearly, the rich information provided by paleo reconstructions should be incorporated in any stochastic streamflow models to enable the generation of a realistic variety of plausible flow scenarios for robust planning and management of water resources in the basin. However, the magnitudes of reconstructed streamflows vary widely due to the reconstruction methodology (Woodhouse et al., 2006). This apparent weakness of the paleo reconstructed flow data has made their use in a water resources planning context contentious, despite their availability for many decades. Despite these apparent weaknesses, few argue about the duration and frequency of dry and wet (i.e., the hydrologic state) periods from the reconstructions. Due to these short comings of the paleo reconstructions, typically this rich information is not used in planning and management for the Colorado River. A creative methodology is required to combine the observed and paleo data to take full advantage of the insights from the paleo data.

1.3 Outline of this study

Given the research questions and, the limitations and inadequacies of the existing efforts, a new integrated framework is needed. This thesis develops such a framework (shown in Figure 1-1) with four key components. Each of these components are described in Chapters 2 through 5, with a summary of conclusions and recommendations for future extensions of this research outlined in Chapter 6. A brief outline of the various chapters is presented below.
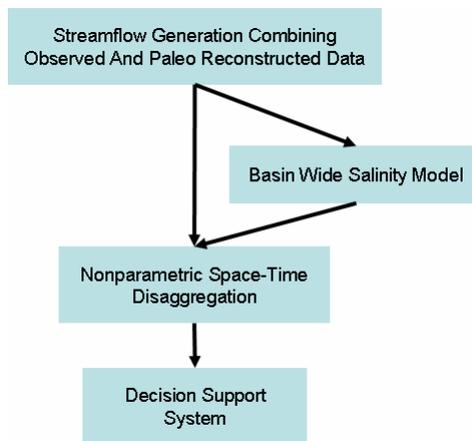
Figure 1-1 Flowchart of study.

In Chapter 2 a stochastic nonparametric space-time disaggregation method is developed. This generates statistically consistent monthly streamflow scenarios across all the locations in the Colorado River Basin from annual streamflows at a single aggregate site. This nonparametric method is simple, flexible and improves upon the traditional alternatives; furthermore, the framework provides the basin stakeholders with a method they can easily understand, a critical requirement for its use in practice. The method also generates streamflow magnitudes and sequences not seen in the past and can be used in short and long-term planning. This is presented in a peer reviewed journal article accepted for publication in Water Resources Research (Prairie et al., 2006b).

Chapter 3 extends the space-time disaggregation framework from the previous chapter to develop a basin wide salinity scenario generation model. Salinity management is important from the water quality standpoint but also with respect to treaty agreements with Mexico (Colorado River Basin Salinity Control Forum, 2005). Salinity is strongly linked with the streamflow magnitude, i.e., increased streamflow generally leads to increased salt mass but less concentration due to dilution; the opposite occurs during low streamflows. This relationship was used by Prairie et al. (2005) in a regression framework to estimate salinity at a single site based on the streamflow. Combining this single-site salinity estimation model with the space-time streamflow disaggregation from Chapter 2, a new basin wide salinity generation framework is developed. This generates monthly salinity scenarios at all the locations in the Colorado streamflow network simultaneously, preserving the spatial-temporal statistical properties of the observed salinity and also the streamflow-salinity relationship at individual locations.

As mentioned in the previous section, paleo reconstructed streamflows on the Colorado River, based on tree-rings, indicate periodic occurrence of longer and severe droughts, suggesting that the recent unprecedented drought in the Colorado basin is not unusual (Woodhouse et al. 2006). The magnitudes of paleo reconstructed streamflows are less reliable, especially during extreme years inhibiting their direct use in water management and planning (Woodhouse and Brown, 2001). However, the information on the hydrologic state (wet or dry) is very reliable. Thus, the question then is how to combine the long paleo reconstructed streamflow information of lesser reliability with the shorter but reliable observed data to develop a framework for

streamflow simulation? This is addressed in Chapter 4, where a new and innovative methodology is developed. In this methodology, the paleo reconstructed streamflows are used in a nonhomogeneous Markov Chain model to generate a hydrologic state consistent with the paleo reconstruction and a nonparametric conditional bootstrap method is used to generate the streamflow magnitude from the observed record conditioned on the generated hydrologic state. This methodology combines the strengths of both of the data sets in an effective manner, producing robust and realistic hydrologic sequences.

The stochastic flow and salt scenarios simulated from the methodologies developed in the previous three chapters are used in Chapter 5 as inputs in Reclamation's long-term policy model, CRSS. The impact of two different reservoir operating policies on the various components of the Colorado River's water resources system are evaluated for different hydrologic scenarios. Reliability estimates of various system components are also estimated; thus, providing a robust basis for sustainable water resources planning and management in the basin.

Chapter 6 summarizes the key findings and contributions from the four components of the proposed research framework and recommends potential future extensions of this research.

CHAPTER 2

A STOCHASTIC NONPARAMETRIC TECHNIQUE FOR SPACE-TIME DISAGGREGATION OF STREAMFLOWS

2.1   Introduction

Synthetic simulation of streamflow sequences is used in a variety of applications including reservoir operation and for evaluating water supply reliability. Multiple reservoirs and stream sections are often considered in a system's operation plan. For this purpose, streamflows generated at different sites need to be consistent. This implies, that the flow at a downstream gauge is the sum of tributary flows; the annual flow is the sum of monthly flows; the monthly fraction of flows in wet/dry years are representative; and the dependencies of flows between the sites have to be reproduced. To this end, the disaggregation problem can be thought of as simulation from the conditional probability density function (PDF) $f(X|Z)$, where X is a vector of disaggregated (e.g., monthly) flows and Z is the aggregate (e.g., annual) flows and other terms (e.g., the first months correlation with the last month of the previous year), subject to the condition that the disaggregated flows add up to the aggregate flows, which is the additivity property. Often a simpler approach has been used consisting of fitting a model of the form:

$$X = AZ + B\varepsilon \tag{1}$$

Where Z is usually taken to be just the annual flow and A and B are matrices of the model parameters that are estimated to ensure the additivity property and ε is the stochastic term. Notice that the above form is that of a linear regression which has a rich developmental history; consequently, the main assumption is that the stochastic

term and hence the data (X and Z) are assumed to be normally distributed. To achieve this, the data are typically transformed to a normal distribution by appropriate transforms before the model is fit. The simulation proceeds as follows: (i) An aggregate streamflow is generated from an appropriate linear or nonlinear model or equivalent dataset. (ii) The simulated aggregate flow is then disaggregated using the above model. The simulated flows are back transformed to the original space. This linear stochastic framework for streamflow disaggregation was first developed by Valencia and Schaake (1973) and subsequently modified and improved by several others (Mejia and Rousselle, 1976; Lane, 1979; Salas et al., 1980; Stedinger and Vogel, 1984; Stedinger et al., 1985; Salas, 1985; Santos and Salas, 1992).

Since these models are fit in the transformed space the additivity of the disaggregated flows to the aggregate flows in the original space after back transformation is not guaranteed. Hence, several adjustments have to be made (e.g., Lane, 1982; Stedinger and Vogel, 1984; Grygier and Stedinger, 1988). Furthermore, the model is designed to reproduce the statistics in the transformed space but reproduction is not guaranteed in the original space.

Alternate approaches to disaggregation (Tao and Delleur, 1976; Todini, 1980; Koutsoyiannis, 1992; Koutsoyiannis and Manetas, 1996; Koutsoyiannis, 2001) allow representation of non-Gaussian data directly in the disaggregation scheme to avoid the need for data transformation. These techniques can incorporate the skewness from the historic data into the stochastic term (Tao and Delleur, 1976; Todini, 1980, Koutsoyiannis, 1999). Koutsoyiannis (2001) provides a stepwise disaggregation scheme that incorporates an adjustment procedure that preserves the additivity

property and certain higher order statistics. These methods are iterative in nature and thus, computationally intensive besides requiring assumptions of linearity.

Recent advances in nonparametric methods (see Lall (1995) for an overview of nonparametric methods and their applications to hydroclimatic data) provide an attractive alternative to linear parametric methods. Unlike the linear approach where a single linear model is fit to the entire data, the nonparametric methods involve local functional fitting. The function is fit to a small number of neighbors at each point. This approach has the ability to capture any arbitrary features (nonlinearities, non-normal, etc.) exhibited by the data. Nonparametric methods have been applied to a variety of hydroclimate modeling questions including stochastic daily weather generation (Rajagopalan and Lall, 1999; Yates et al., 2003), streamflow simulation (Sharma et al., 1997; Lall and Sharma, 1996; Prairie et al., 2006a), streamflow forecasting (Grantz et al., 2005, Singhrattna et al., 2005) and for flood frequency estimation (Moon and Lall, 1994) to mention a few.

Kernel estimator based nonparametric streamflow simulation at a single site was developed by Sharma et al. (1997) where they also demonstrate its advantage over traditional linear models. Sharma and O'Neil (2002) improved on this to capture the interannual dependence. However, kernel methods can be inefficient in higher dimensions (e.g., space-time disaggregation) as noted by Sharma and O'Neil (2002) and as such, difficult to implement in multivariate problems such as space-time disaggregation in a network. Lall and Sharma (1996) developed a K-nearest neighbor (K-NN) bootstrap approach to time series modeling and applied it to streamflow simulation. Being a bootstrap method values not observed in the historic data will not

be generated in the simulations. To address this, a modified version of the K-NN bootstrap was developed by Prairie et al. (2005, 2006a) and this was further used in streamflow forecasting (Grantz et al., 2005; Singhrattna et al., 2005). Semi-parametric approaches that combine the traditional linear modeling and bootstrap methods for streamflow simulation have also been developed (Souza Filho and Lall, 2003; Srinivas and Srinvasan, 2001).

Tarboton et al. (1998) developed a kernel based approach (an extension of their single site methodology in Sharma et al. (1997)) for temporal (i.e., annual to monthly) streamflow disaggregation. Kumar et al. (2000) adopted K-NN bootstrap techniques in conjunction with an optimization scheme for spatial and temporal disaggregation of monthly streamflows to daily flows. They indicate that disaggregating monthly flow to daily involves a higher dimensional problem that cannot always be well represented by traditional parametric disaggregation techniques. Additionally, daily flows typically display nonlinear flow dynamics that are not adequately modeled with traditional techniques. The optimization framework allows for increased flexibility in specifying the functional relationships the disaggregation scheme needs to preserve but at a great computational cost. Srinivas and Srinivasan (2005) developed a semi-parametric disaggregation method for a multi-site model they termed as Hybrid Moving block bootstrap Multi-site model (HMM). In this approach a parametric model (such as a linear auto regressive model) is fit to the data and the residuals from this model are resampled by block bootstrapping (the nonparametric component). This method is able to incorporate the strengths of both parametric and nonparametric models but still requires multiple steps.

In practical terms, there is a need for a robust, simple and parsimonious approach for space-time streamflow disaggregation that can capture the features exhibited by the data. To this end, here we develop a K-NN based disaggregation framework. The proposed framework and the algorithm are first described followed by its application to four streamflow sites on the Upper Colorado River basin, concluding with a summary and discussion of applications and the future direction for this research.

2.2 K-Nearest Neighbor Based Disaggregation Framework

The framework follows the work of Tarboton et al. (1998) except that the kernel based density estimation is replaced with a K-nearest neighbor approach. We describe the framework and the implementation algorithm for the case of a temporal (annual to monthly) disaggregation and the same steps follow for spatial disaggregation. As an example, consider X to be a $d=12$ dimensional monthly flow vector where Z is the annual flow. As mentioned earlier, the disaggregation problem amounts to simulation from the conditional PDF $f(X|Z)$ with the constraint that the disaggregated flows sum up to the aggregate flow. The conditional PDF can be written as

$$f(X|Z) = f(X,Z)/\int f(X,Z)dX \tag{2}$$

The numerator in the above equation requires the estimation of a $d+1$ dimensional joint density function $f(X,Z)$. However, due to the additivity requirement all the mass of the this joint PDF is situated on the $d-$dimensional hyperplane defined by

$$X_1 + X_2 + \cdots + X_d = Z \tag{3}$$

Thus, for a particular value of Z (the aggregate annual flow) the conditional PDF can be visualized geometrically as the probability density on a $d-1$ dimensional hyperplane slice through the $d-$ dimensional density $f(X)$. The conditional PDF can be specified through a rotation of the vector X into a new vector Y whose last coordinate is aligned perpendicular to the hyperplane defined by (3). Tarboton et al. (1998) describe this in detail and illustrate this point very well in their Figure 1. The conditional PDF is constructed in the rotated space ($f(Y|Z)$) and the simulation is also done in this rotated space before back rotation. In the Tarboton et al. (1998) framework kernel density estimators are used to construct this conditional PDF and subsequently for simulation. As mentioned earlier, the kernel methods are known to be inefficient and cumbersome to implement in higher dimensions. This limits their ability to extend the approach to space and time disaggregation.

We depart from the Tarboton et al. (1998) framework here and instead develop a K-NN based bootstrap approach to construct and simulate from the conditional PDF ($f(Y|Z)$). The methodology is described in the algorithm below.

2.3   The Algorithm

The steps involved in the algorithm are as follows:

1. The historic data of monthly flows are oriented in $\mathbf{X}$ such that seasons are across rows and years are across columns. $\mathbf{X}$ is rotated into $\mathbf{Y}$ by the rotation matrix $\mathbf{R}$ where,

$$\mathbf{Y} = \mathbf{R}\mathbf{X} \qquad (4)$$

The procedure for obtaining the rotation matrix is described in detail in the appendix of Tarboton et al. (1998), here we summarize from their description.

The rotation matrix is developed from a standard basis (basis vectors aligned with the coordinate axes) which is orthonormal but does not have a basis vector perpendicular to the conditioning plane defined by (3). One of the standard basis vectors is replaced by a vector perpendicular to the conditioning plane. Operationally this entails starting with an identity matrix and replacing the last row with $1/\sqrt{d}$. The basis set is then no longer orthonormal. The Gram Schmidt orthonormalization procedure is applied to the remaining $d-1$ standard basis vectors to obtain an orthonormal basis that now includes a vector perpendicular to the conditioning plane. The resulting $\mathbf{R}$ matrix is orthonormal and has the property $\mathbf{R}^{\mathrm{T}} = \mathbf{R}^{-1}$. Further, note $\mathbf{R}$ is only a function of the dimension d.

The last row of the matrix $\mathbf{Y}$ is $Y_d = Z/\sqrt{d} = Z'$. The first $d-1$ components of the vector Y can be denoted as U and the last component is Z', i.e., $Y = (U, Z')$. Hence, the simulation involves re-sampling from the conditional PDF ($f(U|Z')$)

2. An aggregate flow (i.e., annual flow) z^* is generated from an appropriate model fitted to the annual flow data. This could be a traditional auto regressive model (Salas, 1985) or a K-NN bootstrap approach (Lall and Sharma, 1996) or a kernel density estimator based method (Sharma et al., 1997) or a modified K-NN bootstrap (Prairie et al., 2005, 2006a) or a block bootstrap resampling (Vogel and Shallcross, 1996). Here we used the modified K-NN (Prairie et al., 2006a).

If a simple K-NN based approach is applied the annual flows will be resampled from the historic data only generating values seen in the historic record. To generate annual values not seen before either the kernel density estimator, the modified K-NN, or a traditional parametric model can be implemented.

3. K nearest neighbors (corresponding to K historic years) of the generated

$z'_{sim} = z^* / \sqrt{d}$ are identified. The nearest neighbors are obtained by computing

the distance between the generated z'_{sim} and the historic Z'. The neighbors are

assigned weights based on the function

$$W(k) = \frac{1}{k \sum_{i=1}^{K} \frac{1}{i}} \qquad \text{where } k = 1,2,..K \qquad (5)$$

This weight function gives more weight to the nearest neighbors and less to the

farthest neighbors. For further discussion on the choice of the weight function

readers are referred to Lall and Sharma (1996).

The number of nearest neighbors, K is based on the heuristic scheme $K = \sqrt{N}$

where N equals the sample size (Lall and Sharma, 1996), following the

asymptotic arguments of Fukunaga (1990). Objective criteria such as Generalized

Cross Validation (GCV) can also be used. The above heuristic scheme has

performed well in a variety of applications (Lall and Sharma, 1996; Rajagopalan

and Lall 1999; Yates et al., 2003).

Using these weights as a probability metric, one of the neighbors is resampled.

Suppose the selected neighbor corresponds to year j of the historic record.

4. The corresponding vector Y^* is created as

$$Y^* = (U_j, z'_{sim}) \qquad (6)$$

5. The final step is the back rotation to the original space

$$x^* = R^T Y^* \qquad (7)$$

x^* is the vector of disaggregated (i.e., monthly) flows that will sum to z^*.

Steps 2 though 5 are repeated to generate ensembles of monthly streamflows. The same steps can be used for spatial disaggregation, in which case the matrix **X** represents the spatial streamflows and Z represents the spatial aggregate flow.

Even though we resample historic data, steps 4 and 5 enable the simulation of monthly values not seen in the historic record and can also generate negative values. However, in our application here the negative values simulated were extremely small in number - less than 0.4% of the simulated values for all gauges were negative.

2.3.1 Numerical Example

To further explain the algorithm described above a simple numerical example is presented. In this example we assume two variables (say two seasons) and they sum to the aggregate flows. The **X** matrix (seasonal flows) and the vector Z (aggregate flows) are given as

$$\mathbf{X} = \begin{bmatrix} 222 & 585 \\ 232 & 1206 \end{bmatrix}, Z = \begin{bmatrix} 454 \\ 1791 \end{bmatrix} \tag{8}$$

The rotation matrix **R** is obtained as described in step 1 of algorithm resulting in

$$\mathbf{R} = \begin{bmatrix} 0.7071068 & -0.7071068 \\ 0.7071068 & 0.7071068 \end{bmatrix} \tag{9}$$

The rotated matrix **Y** is computed as

$$\mathbf{Y} = \mathbf{RX} = \begin{bmatrix} -7 & -439 \\ 321 & 1266 \end{bmatrix} \tag{10}$$

Note that the last row of **Y** is equal to $Z/\sqrt{d}$ (here $d = 2$). Suppose the simulated aggregate flow is $z_{sim} = 736$, then $z'_{sim} = \dfrac{736}{\sqrt{2}} = 520$. Based on the

resampling method described in step 3 of the algorithm above, suppose that we chose

the second year then the vector

$$Y^* = (u_{2,1}, z'_{sim}) = \begin{bmatrix} -439 & 520 \end{bmatrix} \qquad (11)$$

The disaggregated vector x_{sim} is obtained as

$$x_{sim} = R^T Y^* = \begin{bmatrix} 0.7071068 & 0.7071068 \\ -0.7071068 & 0.7071068 \end{bmatrix}^T \begin{bmatrix} -439 \\ 520 \end{bmatrix} = \begin{bmatrix} 57 \\ 679 \end{bmatrix} \qquad (12)$$

Note the additivity property $\sum x_{sim} = z_{sim}$ is satisfied.

2.3.2 Model Evaluation

The performance of the K-NN space-time disaggregation approach is evaluated

by applying it to four streamflow locations on the Upper Colorado River basin shown

in Figure 2-1. These gauges are Colorado River near Cisco, Utah (site 1); Green River

at Green River, Utah (site 2); San Juan River near Bluff, Utah (site 3); and Colorado

River at Lees Ferry, Arizona (site 4).
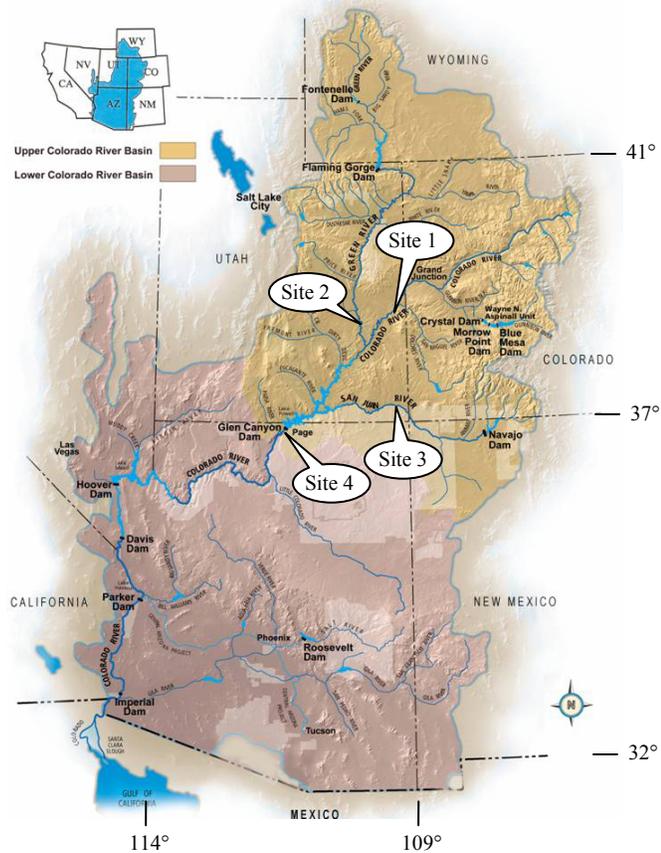
Figure 2-1 Streamflow locations within the Upper Colorado River basin.

Monthly natural streamflows at these locations are available for the 98 year period spanning 1906 – 2003. Naturalized streamflows are computed by removing anthropogenic impacts (i.e., reservoir regulation, consumptive water use, etc.) from the recorded historic flows[1].

The disaggregation schematic is shown in Figure 2-2. In this, we begin with an annual streamflow at an index gauge which is temporally disaggregated to twelve monthly flows. The monthly flows are then disaggregated to flows at the spatial locations. Thus, the disaggregation algorithm is applied twice, first for the temporal and second for the spatial disaggregation. The index gauge is an imaginary gauge

---

[1] The natural flow data and additional reports describing these data are available at
http://www.usbr.gov/lc/region/g4000/NaturalFlow/index.html.

whose monthly flows are created as the sum of the monthly flows at all the four locations. The annual flow at the index gauge was generated from the modified K-NN lag-1 approach (Prairie et al., 2005, 2006b). Using the space-time disaggregation approach we made 500 simulations each of 98 years length. The following statistics are calculated from the simulations and compared with those from the historic data to evaluate the performance of the proposed approach.
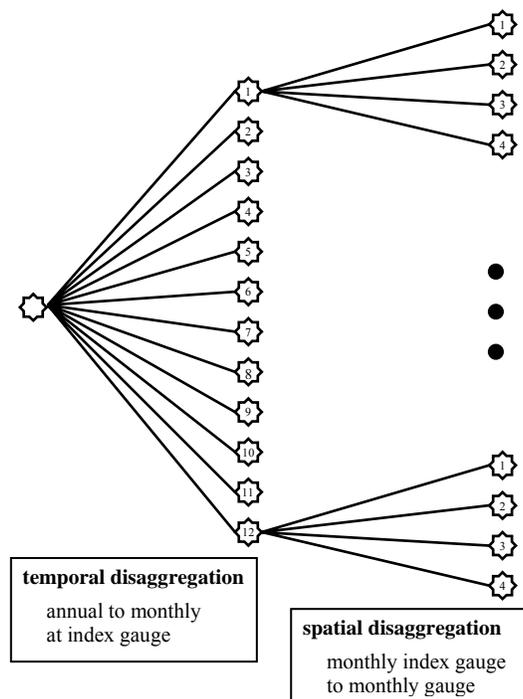
Figure 2-2 Schematic of space-time disaggregation.

2.4 Performance Statistics

These performance statistics include monthly and annual (i) mean, (ii) standard deviation, (iii) coefficient of skew, (iv) maximum, (v) minimum, (vi) backward lag-1 autocorrelation of the flows at the four locations, (vii) Probability Density Functions (PDF), (viii) correlation of flows between the locations, (ix) surplus and drought

statistics. Comparisons with a standard parametric alternative (Salas et al., 1980) are also provided.

2.5 Results

The results are displayed as boxplots where the box represents the interquartile range and whiskers extend to the 5th and 95th percentile of the simulations (note this is different from the standard boxplot definition). The statistics of the historic data are represented as a triangle connected by a solid line. Performance on a given statistic is judged as good when the historic value falls within the interquartile range of the boxplots, while increased variability is indicated by a wider boxplot.

The mean was well reproduced at all sites and therefore not include in the Figures. Performance statistics of Green River at Green River, Utah are shown in Figure 2-3.
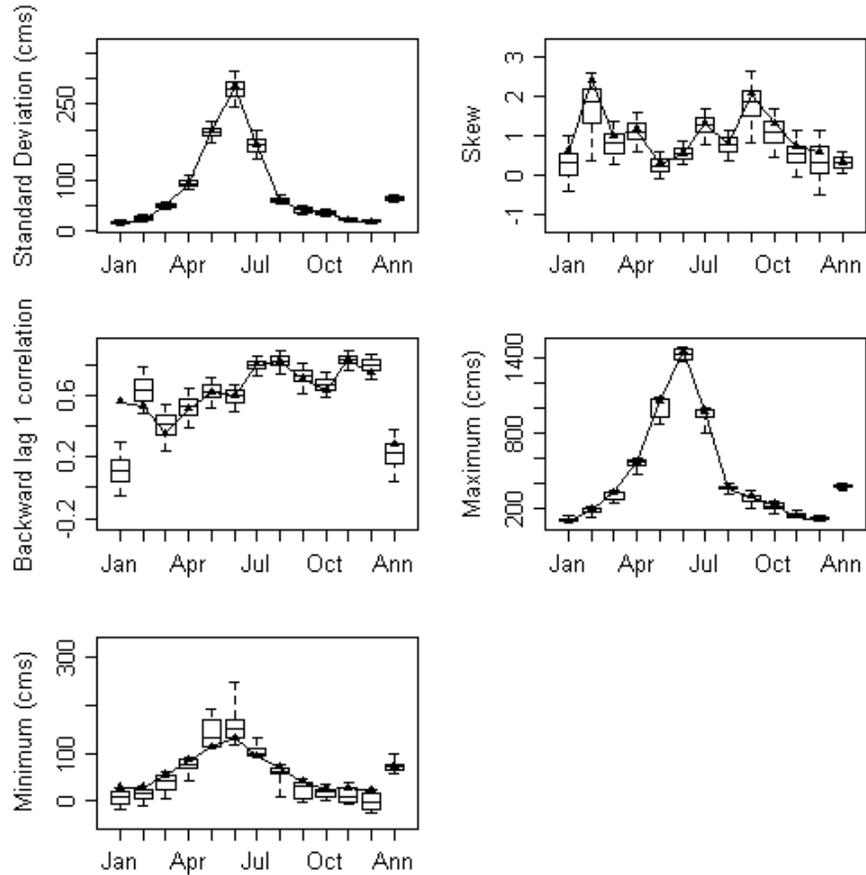
Figure 2-3 Boxplots of monthly and annual statistics for flow at Green River at Green River, Utah. The box represents the interquartile range, and whiskers extend to the 5th and 95th percentile of the simulations. The statistics of the historic data are represented as a triangle.

The standard deviation and skews are well preserved for most all the months and at the annual time step. The low flow months January and February skews are slightly under represented. The Lag-1 autocorrelations are also well simulated though February is slightly over correlated. However, the correlation between the first month of a year and the last month of the preceding year is not preserved. At the index gauge the temporal disaggregation does not incorporate this dependence; therefore, it is not captured in the simulations at the spatial locations. In this basin the flows are largely snow-melt driven and thus, the first (Jan) and last (Dec) months of the calendar year are part of the low flow season (accounting for about 4% on the annual flow); hence, capturing their correlation was not essential. However, we deemed it important to

21

capture the correlations in the remaining months especially during the high flow months, which are well preserved. Kumar et al. (2000) resolved this issue with their optimization framework but at a computation cost. Linear adjustment procedures have also been developed to capture the first month's correlation with the last month of the preceding year (Grygier and Stedinger, 1988; Lane and Frevert, 1990; Koutsoyiannus and Manetas. 1996 and; Koutsoyiannus, 2001). But they all involve estimating several additional parameters and can impact reproduction of other statistics.

The maximum and minimum flow statistics are also reasonably well simulated for most of the months. Extrapolation beyond the maximum historic flow occurs more extensively for some months (Jan, Feb, Aug-Sep, Nov), while other months (Mar-Jul, Oct, Dec) display limited to no extrapolation. A very small number (0.4%) of negative numbers were generated, mostly in low flow months, and had no significant impact on statistics. Similar results were obtained for the flows at Colorado River near Lees Ferry, Arizona (Figure 2-4) and also at the remaining two locations (figures not shown).
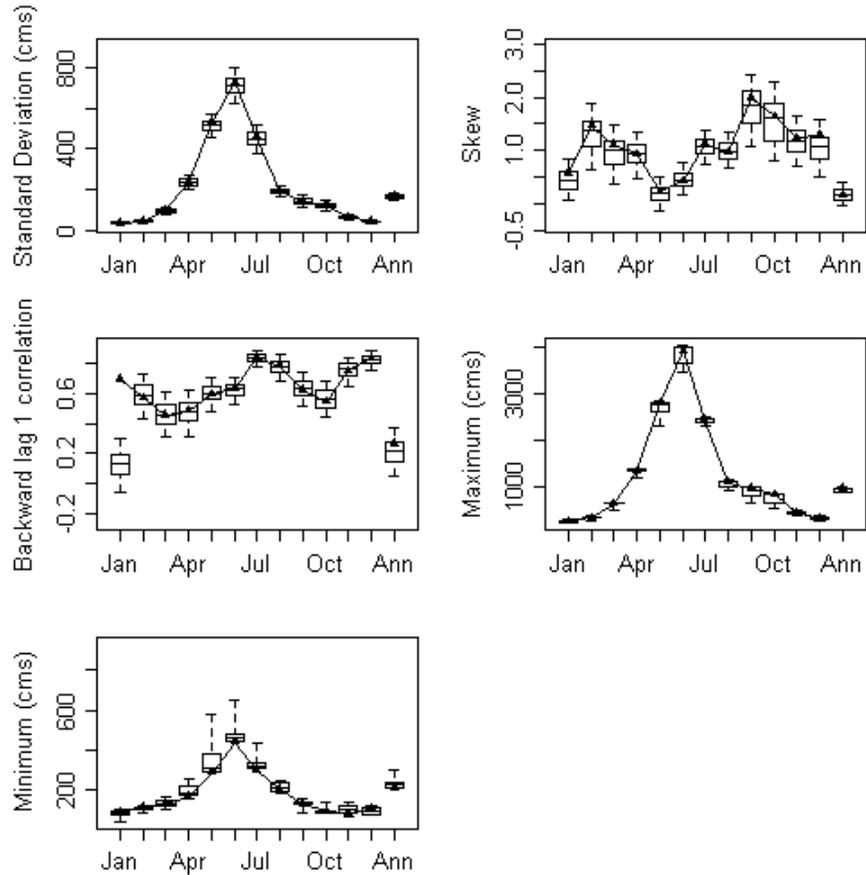
Figure 2-4 Same as Figure 3 but for flows at Colorado River at Lees Ferry, Arizona.

The spatial cross correlation between the monthly flows at Colorado River at Lees Ferry, Arizona (Site 4, the downstream location) and the other three gauges are shown in Figure 2-5. The cross correlations are very well captured during the spring months (the high flow season) and also during other months. There is a slight under simulation of the cross correlations during the low flow months of Jan-Mar and Nov-Dec.
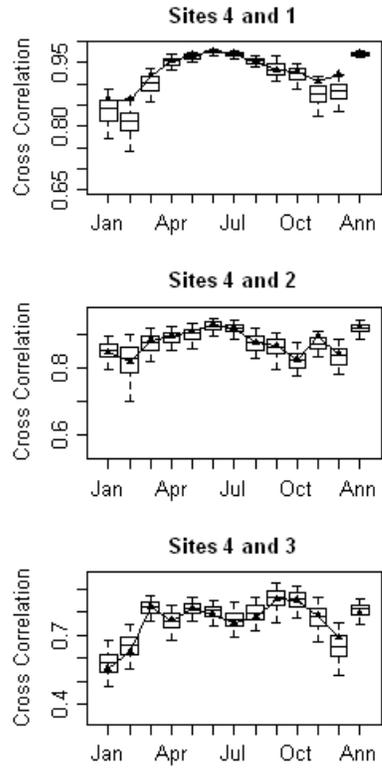
Figure 2-5 Boxplots of monthly and annual cross correlation between the streamflows at the four locations.

Figure 2-6 displays the temporal cross correlation of the monthly and annual flows at several lags for the Colorado River at Lees Ferry, Arizona (Site 4). These statistics are also very well simulated.
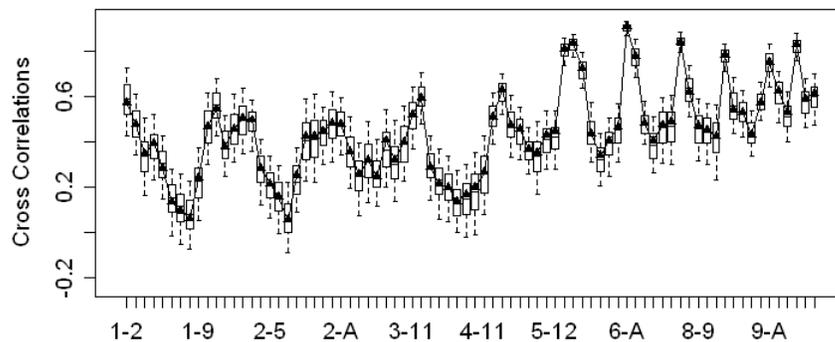
Figure 2-6 Temporal cross correlation pairs for streamflows at Colorado River at Lees Ferry, Arizona. The x-axis sequence is 1-2, 1-3 ···, 1-12, 1-A, 2-3, 2-4 ···, 2-12, 2-A, 3-4 ··· Months are numbered according to calendar year and 'A' represents annual.

24

As described earlier, one of the advantages of nonparametric methods is the ability to capture any arbitrary PDF structure. To test this we estimated the PDF from the simulations and compared them with those of the observed data. Figure 2-7 presents the PDF for June flows at San Juan River near Bluff, Utah. The PDF of the historic data is shown by the solid line and the boxes and whiskers are those of the simulations. The simulations capture the non-normal feature of the historic PDF very well. Nonparametric kernel density estimators are used to compute the PDF (Bowman and Azzalini, 1997). Similar performance was seen with PDFs from other months and locations.
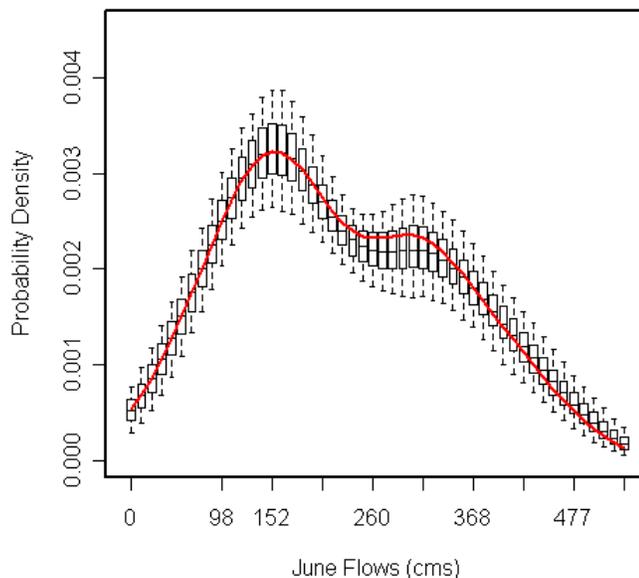
Figure 2-7 Boxplots of PDF of June flows from San Juan River near Bluff, Utah.

To evaluate the performance of the disaggregation approach in capturing longer temporal properties we calculated surplus and drought statistics which include the longest surplus (LS), the longest drought (LD), the maximum surplus (MS) and the maximum deficit (MD) based on the long-term mean as the threshold for drought. These statistics for the flows at Colorado River at Lees Ferry, are shown in Figure 2-

8. The LS statistics exactly reproduces the historic data. The LD statistic in captured within the inter-quartile range of the simulations though tends to be under represented. The MS is again captured within the inter-quartile range of the simulations and well represented. While the MD statistic shows the greatest variability of the all these statistics, though captured within the inter-quartile range the MD tends to be under represented. The simulations generate droughts that are longer in length and greater in magnitude than those in the observed record, though these are only generated for less that 25% of the simulations.
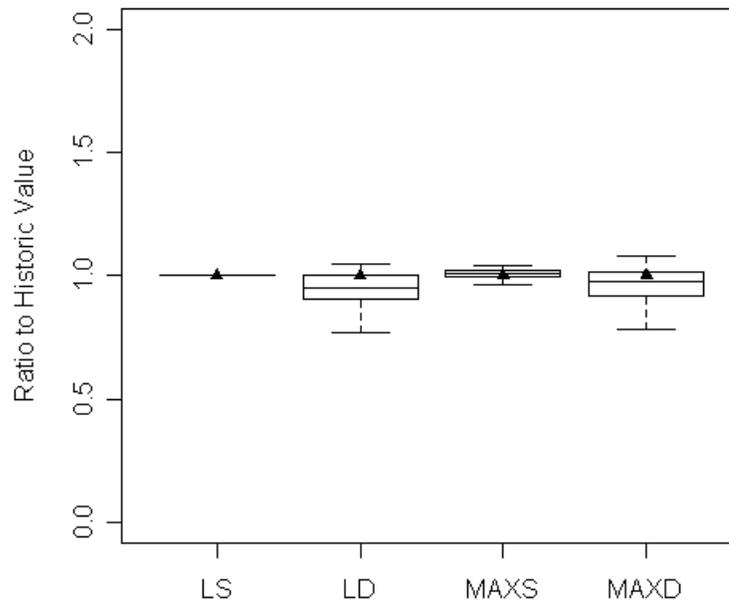
Figure 2-8 Boxplots of surplus and drought statistics for Colorado River at Lees Ferry, Arizona.

2.6 Comparison with a parametric model

We compared the simulations from the K-NN space-time disaggregation approach developed in this research to a traditional parametric model (of the form in equation 1) developed by Mejia and Rousselle (1976) and Salas et al. (1980). The parametric disaggregation models are designed to capture all the basic statistics, but as described

earlier, they have difficulty in capturing non-normal PDF structure and also the coefficient of skewness. These two statistics depend upon the appropriate transformations used to transform the data to a normal distribution. The transformations applied for the parametric model all passed a skewness test for normality, i.e., the transformed data had a coefficient of skew close to zero. Figures 2- 9 and 2-10 displays the PDF of May flows and the monthly and annual coefficient of skew for the Colorado River at Lees Ferry, Arizona. The historic PDF displays a clear bimodal feature which is extremely well preserved by the K-NN disaggregation model (Figure 2-9), while the parametric model (Figure 2-10) is unable to capture this feature instead reproducing a normal structure. Similar results are seen with the coefficient of skewness that is well represented by the non-parametric model but not by the parametric model. It should be noted that even though the transformations were effective the parametric model was fitted to the transformed data and hence, does not guarantee the reproduction of the statistics in the original space.
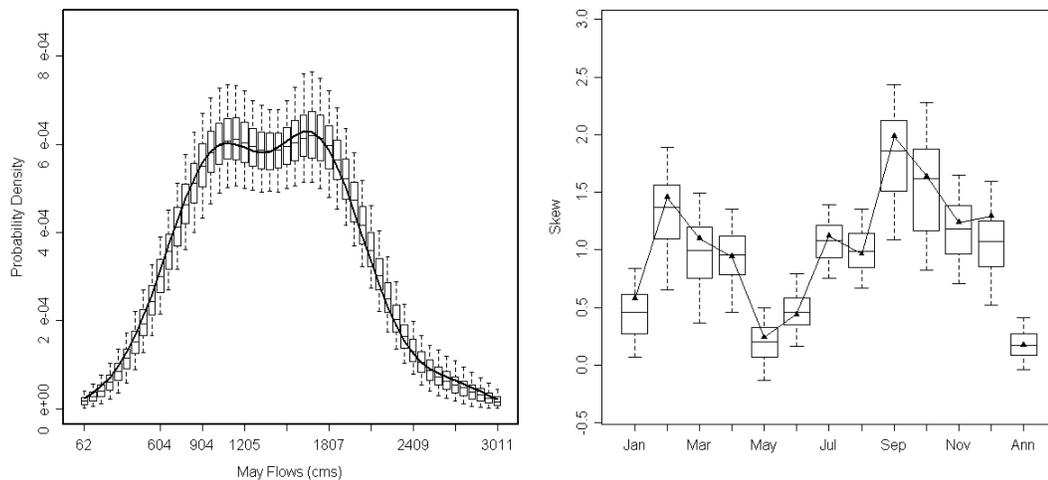
Figure 2-9 Boxplots from nonparametric disaggregation model of PDF (left column) of May flows and monthly and annual coefficient of skew (right column) from Colorado River at Lees Ferry, Arizona.

27
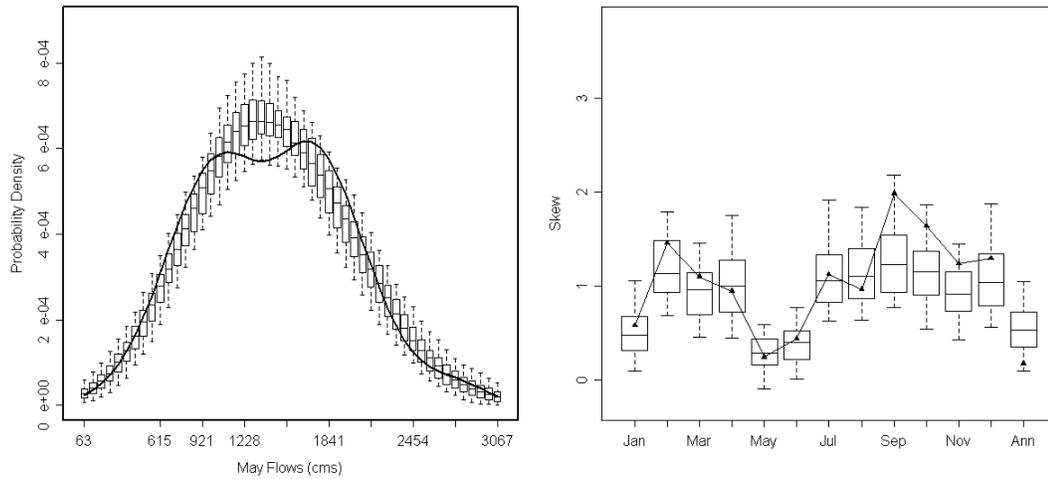
Figure 2-10 Boxplots from parametric disaggregation model of PDF (left column) of May flows and monthly and annual coefficient of skew (right column) from Colorado River at Lees Ferry, Arizona.

2.7   Summary and Discussion

We have presented a simple, robust, and parsimonious framework for space-time simulation of streamflows on a large river network. We adapted the Tarboton et al. (1998) framework but used a K-NN approach to construct and simulate from the conditional PDF. The model captures all the distributional properties and the spatial dependence of the flows at all the locations. Simulating space-time flow scenarios conditioned upon large-scale climate information (e.g., El Nino Southern Oscillation, etc.) for seasonal forecasts can be easily achieved.

A few limitations exist in the proposed nonparametric disaggregation approach, which should to be considered. An obvious limitation was the inability to capture the correlation between the first month of one year and the last month of the previous year. As presented, the proposed approach only solved for the monthly flows conditioned on the dependence structure for the current year. Incorporating the last months flow in the conditional function could be explored to remedy this limitation in scenarios where preserving this correlation is essential. Another, limitation in the

28

proposed approach can arise if extreme values are of interest. Extrapolation beyond observed monthly flows is limited in comparison with parametric counterparts. This limitation needs to be considered on a case by case basis. With the proposed approach this limitation can be addressed with the choice of a proper annual flow model. An annual simulation model that generates more extreme annual flows will in turn generate more extreme monthly values.

The proposed approach involved a two step process in which the temporal disaggregation was first performed followed by the spatial disaggregation. The work of Kumar et al (2000) considers a simultaneous space-time disaggregation based on the K-NN method in an optimization framework, albeit with significant computational effort. For the annual to monthly disaggregation, an approach that blends elements of the method presented here with the optimization approach of Kumar et al. (2000) may provide a means to perform a simultaneous disaggregation.

Efforts are under way to integrate this framework with a basin wide salinity model (Prairie et al., 2005) to generate salinity ensembles. Additionally, tree-ring reconstructions of past annual streamflows at Lees Ferry will be incorporated into this approach to simulate (i.e., reconstruct) monthly streamflows at all the locations in the Upper Colorado River basin that may include extreme events addressing the second aforementioned limitation. Together these projects will enable the evaluation of various policy strategies in the basin.

CHAPTER 3

A BASIN WIDE STOCHASTIC SALINITY MODEL

To simulate salinity as well as streamflow the salinity data must be correlated with the generated streamflow. Next we present our proposed method to generate salinity using components of the work presented in Chapter 2 and applying these to salinity generation. We also present an extension of a single site salinity model presented in Prairie et al. (2005) to multi-site for use throughout a river basin network.

3.1   Motivation and Background

Many rivers experience high salinity resulting from natural and anthropogenic sources, more so in the western United States and in the Colorado River basin. This impacts water quality and hence, is closely monitored. The salinity is closely linked with streamflow quantity in that, a higher flow brings with it more salinity but also provides substantial dilution to reduce the salt concentration and vice-versa during low flow regimes (DOI, 2003b). In the Colorado River basin, enactment of the Federal Water Pollution Control Act Amendments of 1972 led to the development of fixed numeric criteria for salinity levels at three key locations on the Lower Colorado River (DOI, 2003b). This coupled with year-to-year variability of streamflows and operating policy of reservoirs place unique constraints on salinity monitoring and mitigation and also on the management of reservoirs and usability of water for irrigation in the basin.

To evaluate policy options for water quantity and quality, Reclamation developed a basin wide decision support model for streamflows and salinity, named, the Colorado River System Simulation (CRSS) (DOI, 1987). In this model the river basin

is represented by 29 gauges spread throughout the basin, which are also observed gauges, along with operating rules and policies that represent the "Law of the River" (Nathanson, 1978). Basin scale decision support models, such as the CRSS, are typically employed to understand the complex interaction between water resources operating policy, planned strategies for salinity mitigation, and to evaluate impacts of water management policy options on salinity in the river system. These models require statistically consistent basin wide scenarios of streamflow and salinity, which are generated from stochastic models.

There is a rich history of stochastic models for generating basin wide streamflows that reproduce the statistical properties of the observed data. There are several linear time series models (Valencia and Schaake, 1973; Mejia and Rousselle, 1976; Lane, 1979; Salas et al., 1980; Stedinger and Vogel, 1984; Stedinger et al., 1985; Salas, 1985; Santos and Salas, 1992; Koutsoyiannis and Manetas, 1996; Koutsoyiannis, 2001) that have been developed over the years and are used widely. These models, however, can only capture linear relationships and thus, are somewhat limited in their abilities. Nonparametric methods, on the other hand, offer an attractive alternative with their simplicity, assumption free nature, ability to capture any arbitrary functional features (i.e., linear, nonlinear, Gaussian or non Gaussian) and portability across sites. There have been interesting advances in nonparametric time series models for streamflow simulation (Tarboton et al., 1998; Kumar et al., 2000; Srinivas and Srinivasan, 2005). More recently, we developed a robust nonparametric space-time streamflow simulation model (Prairie et al., 2006b) and applied it successfully to the Colorado River basin.

There is very little literature on stochastic generation of basin wide salinity. Malone et al. (1979) developed a linear deterministic model for characterizing and modeling salinity in the Colorado River basin, they subsequently added a stochastic component. The stochastic component was developed with the generation of moment equations technique (Schweppe, 1973). This technique has equations to describe the mean and variance but is limited as these are the only statistics guaranteed to be preserved by the model. Additionally, the model assumes a linear structure. Mass balance models for estimating salinity were developed by Wurbs and Karama (1995) and Wurbs et al. (1995), which could be coupled with a stochastic streamflow generator. Prairie et al. (2005) developed a nonparametric statistical model for simulating salt scenarios consistent with streamflows at a single location. In this, a nonparametric regression based on a local polynomial approach (Loader, 1999) is developed for the salt mass using the streamflow. This regression is used in conjunction with streamflows from a stochastic simulation model to generate salinity scenarios. The nonparametric model performed extremely well in being able to simulate the observed salinity at the Glenwood Springs gauge on the Colorado River and improved upon the USGS developed method (Mueller and Osen, 1988).

Clearly their does not exist a basin wide stochastic model for generating streamflow and salinity scenarios that can capture the spatial and temporal statistics of the observed salt and flow data. As noted above, such a model is crucial for evaluating various management policy options for effective planning, which motivates the present study. To this end, our recent developments of a stochastic nonparametric basin wide streamflow generation model (Prairie et al., 2006b) and

nonparametric single-site salinity model (Prairie et al., 2005), provides a unique opportunity to integrate these two into a basin wide salinity and flow generation framework. This forms the basis of the proposed research.

The chapter is organized as follows. The basin wide salinity generation framework is described along with two alternate approaches for salinity and flow generation. This is followed by their application to data from four locations in the Upper Colorado River Basin. Results from the application with a discussion and summary of our findings conclude the chapter.

3.2   Basin wide Salinity Generation Framework

As mentioned above the framework uses the stochastic nonparametric single-site salinity model and the basin wide streamflow generation model. Therefore, we first present these two models in brief for the benefit of the readers and refer to the respective papers for details, followed by the basin wide salinity generation framework.

3.2.1   Single Site Salinity Model

The single site salinity model presented in Prairie et al. (2005) uses a nonparametric regression method based on local polynomial estimation, which describes the variability of salt mass as a function of flow. The model is defined as:

$$Saltmass = f(streamflow) + error \qquad (1)$$

The main feature is that the function f is estimated locally (Loader, 1999). The implementation steps are as follows.

    i.   At any value of the streamflow, say x*, K-nearest neighbors (K-NN) are identified from the observations.

ii. To the K-NN a polynomial of order p is fit.

iii. The fitted polynomial is then used to estimate the salt mass corresponding to the streamflow x*.

The number of nearest neighbors (K) and the order of polynomial p are estimated for the observed data using objective criteria, Generalized Cross Validation (GCV).

The local estimation of the function f provides the capability to capture any arbitrary features (linear or nonlinear) that might be present in the data; besides, this obviates making any assumptions as to the underlying form of the function f (linear in the case of traditional linear regression approach). For details on the methodology and its development for salinity modeling we refer the readers to Prairie et al. (2005). The local polynomial approach above also provides estimates of uncertainty assuming the errors to be Normally distributed (Loader, 1999) using regression theory. Prairie et al. (2005) suggested resampling of residuals within the neighborhood of $x*$ and adding them to the mean estimate from step (iii), as a way to obtain asymmetric confidence intervals and also better characterization of the error structure. As mentioned above, this approach was applied successfully in salt modeling at the Glenwood Springs gauge on the Colorado River.

3.2.2 Basin wide Streamflow Simulation Model

A stochastic nonparametric space-time disaggregation model for simulating streamflows simultaneously across the river basin was developed in Prairie et al. (2006b). In this, the streamflows are first generated at a downstream aggregate gauge which is then disaggregated to all the upstream locations, thus, preserving the summability, and spatial and temporal dependencies. The simulation is performed

from the conditional probability density function $f(X|Z)$ where in the case of temporal disaggregation X represents the seasonal or monthly (disaggregate) flows and Z is an annual (aggregate) of the seasonal or monthly flows. For spatial disaggregation X represents the flows at several spatial locations and Z denotes the flow at the aggregate location. The steps involved in performing the disaggregation are as follows:

1. The historic data of monthly flows **X** is rotated into **Y** by the rotation matrix **R** where,

$$\mathbf{Y} = \mathbf{RX} \qquad (2)$$

The procedure for obtaining the rotation matrix is described in detail in the appendix of Tarboton et al. (1998) and summarized in Prairie et al. (2006b) The **R** matrix is orthonormal and has the property $\mathbf{R}^{\mathrm{T}} = \mathbf{R}^{-1}$. The last column of the matrix **Y** is $Y_d = Z/\sqrt{d} = Z'$. The first $d-1$ components of the vector Y can be denoted as U and the last component is Z', i.e., $Y = (U, Z')$. Hence, the simulation involves re-sampling from the conditional PDF ($f(U|Z')$)

2. An aggregate flow (i.e., annual flow) z^* is generated from an appropriate model fitted to the annual flow data. This could be a traditional auto regressive model (Salas, 1985) or a K-NN bootstrap approach (Lall and Sharma, 1996) or a kernel density estimator based method (Sharma et al., 1997) or a modified K-NN bootstrap (Prairie et al., 2006a) or a block bootstrap resampling (Vogel and Shallcross, 1996). Here we used modified K-NN.

3. K nearest neighbors (corresponding to K historic years) of the generated

$z'_{sim} = z^* / \sqrt{d}$ are identified. The nearest neighbors are obtained by computing

the distance between the generated z'_{sim} and the historic Z'. The neighbors are

assigned weights based on the function

$$W(k) = \frac{1}{k \sum\limits_{i=1}^{K} \frac{1}{i}} \qquad \text{where } k = 1,2,..K \qquad (3)$$

This weight function gives more weight to the nearest neighbor and less to the

farthest neighbors. Using these weights as a probability metric, one of the

neighbors is resampled. Suppose the selected neighbor corresponds to year j of

the historic record.

The number of nearest neighbors, K is based on the prescriptive choice of square-

root of all possible candidates (i.e., $K = \sqrt{N}$) (Lall and Sharma, 1996;

Rajagopalan and Lall 1999; Yates et al., 2003) following the asymptotic

arguments of Fukunaga (1990). Objective criteria such as Generalized Cross

Validation (Lall and Sharma, 1996) can also be used but the above heuristic

scheme was found to perform quite well in a variety of applications (see the above

references).

4. The corresponding vector Y^* is created as

$$Y^* = (U_j, z') \qquad (4)$$

5. The final step is to rotate back to the original space

$$x^* = R^T Y^* \qquad (5)$$

x^* is the vector of disaggregated (i.e., monthly) flows that will sum to z^*.

Steps 2 though 5 are repeated to generate ensembles of monthly streamflows. The same steps are followed for spatial disaggregation, in which case the matrix **X** represents the spatial streamflows and Z represents the spatial aggregate flow.

Even though we resample historic data in the rotation space, steps 4 and 5 (back rotation) enable the simulation of values not seen in the historic record and also negative values. However, the negative values simulated are few in number. This approach was developed and applied for simulating streamflows at four gauges on the Colorado River basin to show the capability of the model in capturing all the distributional, dependence and non-Gaussian properties exhibited by the data (Prairie et al., 2006b).

3.2.3 Basin wide Salinity Generation

The above described basin wide streamflow generation model and the single site salinity model are combined to develop a basin wide flow and salt simulation framework. Under this framework we present two approaches to generating basin wide salinity. For ease of presentation, the description of the approaches below assumes that salinity is to be generated at annual and monthly time scales at four locations on a river basin. The same methodology carries over to any number of locations that might be present.

3.2.3.1 Approach I

The schematic of this approach is shown in Figure 3-1a. First an index (or spatial aggregate) site is created; this can be an imaginary downstream gauge whose values of salinity or flows are the sum of values of all the upstream gauges; or a downstream gauge whose values are an aggregation of the gauges above. Here we create an

imaginary index gauge that has the monthly flow and salinity values as the aggregation of the four gauge values upstream. The steps are as follows:

i. An annual flow value is generated at the index gauge using a modified K-NN lag-1 model (Prairie et al., 2006a). A traditional Auto Regressive (AR-1) model or any other approach can instead be used.

ii. A corresponding annual salinity is generated from the single-site salinity model.

iii. The annual salinity is then disaggregated to monthly values at the index gauge and subsequently, each monthly is disaggregated to four spatial locations; thus, obtaining all the monthly salinity values at all the four locations.

iv. Repeat steps (i) through (iii) to generate ensembles of salinity.

Step (iii) is also applied to the annual flow value generated in (i) to obtain the spatial and temporal flows at all the gauges.

a)

salt / flow

Single annual regression
Models natural salt at single gauge

temporal disaggregation
annual to monthly at index gauge

spatial disaggregation
monthly index gauge to monthly gauge

b)

spatial disaggregation
performed on annual flow

Four annual regressions
Models natural salt at each gauge

temporal disaggregation
annual to monthly at each gauge
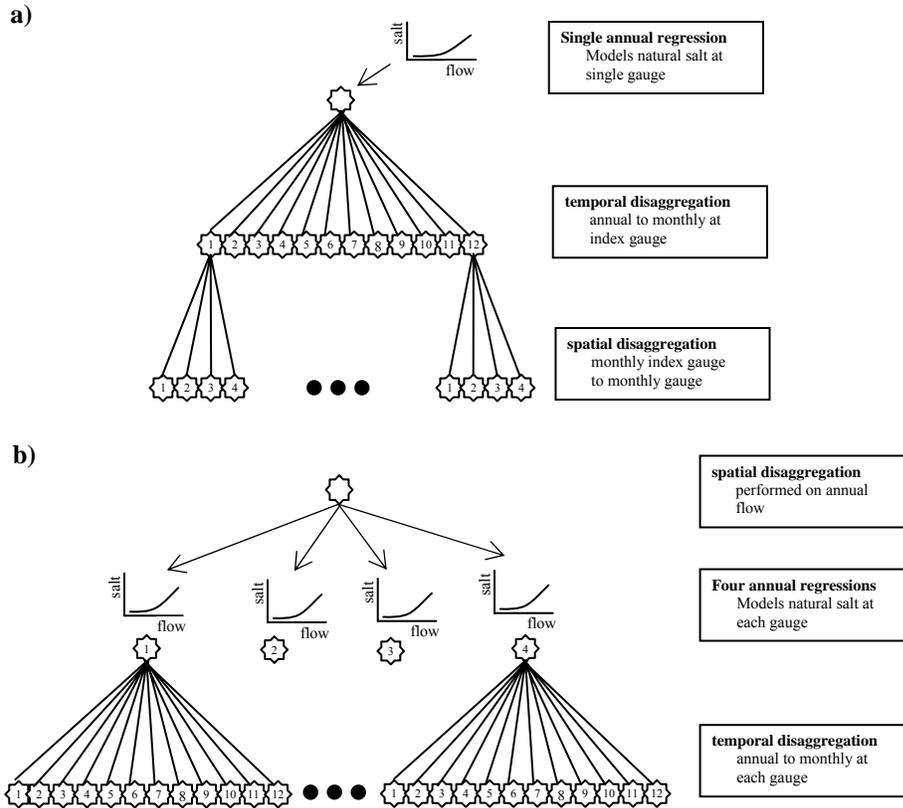
Figure 3-1 a) Schematic of space-time disaggregation with single regression. b) Schematic of space-time disaggregation with four regressions.

This approach preserves the summability of the salinity values at the annual and monthly time scale, i.e., for each month the salinity value at the index gauge will equal the sum of values at the four gauges and the same is true at the annual time scale. Spatial correlation of salinity between the gauges will also be preserved as the index gauge salt is disaggregated in space and time. In addition all the distributional properties of the salt at each gauge will be well captured. However, the correlation between salt and flows at any gauge are not guaranteed to be preserved. This is due to the fact that the salt and flows are disaggregated separately.

39

3.2.3.2 Approach II

It is well known that flow and salt are significantly related and in the context of policy evaluation and management joint ensembles of both flows and salinity are required. This implies that the generated ensembles have to capture the flow and salt relationship presented in the data at least annually, the time scale for salinity monitoring. To achieve this, a modified approach is proposed whose schematic is shown in Figure 3-1b. The steps are as follows:

i. An annual flow value is generated at the index gauge.

ii. This is disaggregated to annual flows at the four gauges using the disaggregation model.

iii. A corresponding annual salinity is generated from the single-site salinity model separately at each gauge.

iv. The annual salinity is then disaggregated to monthly values individually at each gauge; thus, obtaining all the monthly salinity values at all the four locations.

v. Repeat steps (i) through (iv) to generate ensembles of salt scenarios.

As in the first approach, the annual flow at the index gauge can be disaggregated in space and time to obtain the monthly flows at all the gauges.

Clearly, this approach preserves the annual flow and salt correlation at each gauge, as the salt is generated from the flow annually using the appropriate nonparametric flow and salt relationship at each gauge. The distributional properties of salt at each gauge will also be preserved. The spatial correlation of salinity between the gauges is not designed to be captured because the salinity is generated

40

independently at each gauge based on the streamflow, which is generated simultaneously at all the locations. However, much of the spatial dependence in salinity will be captured via the flows.

While we presented two alternate approaches several other variations are possible within this framework, depending on the requirements of each application. The performance of the presented approaches in capturing a suite of statistical properties is evaluated in the following sections.

3.3   Application and Model Evaluation

The basin wide salinity generation framework described above was again applied to the four gauges on the Upper Colorado River basin shown in Figure 2-1. These locations include the Colorado River near Cisco, Utah (site 1); Green River at Green River, Utah (site 2); San Juan River near Bluff, Utah (site 3); and Colorado River at Lees Ferry, Arizona (site 4). Monthly natural streamflows and salt data at these locations are available for the 33 year period spanning 1971 to 2003. The natural flow and salt data are computed by removing anthropogenic influences (e.g., reservoir regulation, consumptive water use, salt loading from agriculture, or salt removed with exports) for both flow and salt from observed gauge records. The natural flows are calculated from the observed streamflows at the gauges. Likewise, the natural salt is also calculated from the observed gauge salinity[2]. For ease of presentation these data are referred to as, historic flow and salt values, henceforth.

The annual flow at the index site is generated from the modified K-NN approach (Prairie et al., 2006a). We generated 500 simulations of flow and salt each of 33 years

---

[2] Further information on the natural flow and salt data and computation techniques can be obtained at http://www.usbr.gov/lc/region/g4000/NaturalFlow/index.html.

length. Salt is computed with the single site salinity model. The local polynomial for each salinity model is presented in Figure 3-2 (solid line) along with the historic flow and salt values (open circles).
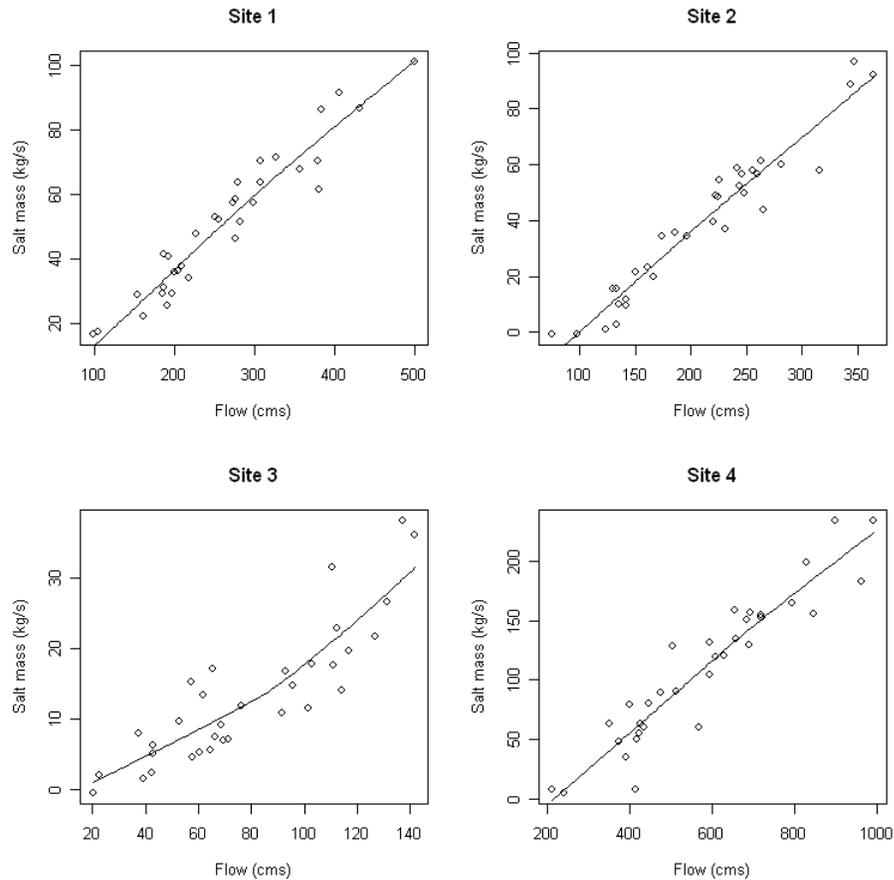
Figure 3-2 Annual regressions of natural flow versus salt at four sites.

To evaluate the performance of this salinity generation framework, a suite of statistics, listed below, is computed for simulations from each approach and select results are presented as boxplots along with the corresponding statistic from the observed values. Only limited statistics of the flow are presented as the flow statistics at these gauges have been extensively presented and studied in Prairie et al., (2006b).

42

3.3.1 Performance Statistics

The performance statistics include the monthly and annual (i) mean, (ii) standard deviation, (iii) coefficient of skew, (iv) maximum, (v) minimum, and (vi) lag-1 autocorrelation of the flows at the four sites. Annual and monthly (i) correlation between flow and salt, (ii) probability density function (PDF) of the annual flow and salinity at the locations, and (iii) cross correlation of salt between the four sites.

Although we compute monthly statistics, we focus on the annual statistics as it is the time scale for salinity monitoring and mitigation planning in the basin.

3.4 Results

The results are displayed as boxplots where the box represents the interquartile range and whiskers extend to the 5th and 95th percentile of the simulations. The statistics of the historic values are represented as a triangle. Performance on a given statistic is judged as good when the historic value falls within the interquartile range (IQR) of the boxplots, while increased variability is indicated by a wider boxplot.

The PDFs of annual flow at all sites (Figure 3-3) are preserved within the IQR except for site 3. Site 3 was well preserved when the entire available record for flow (1906-2003) were used, as in Prairie et al. (2006b). Because flow and salt data are only available back to 1971 we could not included the earlier set of flows. Limiting the sample size reduced the ability to capture the observed statistics within the IQR. Extending the available salt data may improve preservation of site 3. Still, the observed statistic is captured within the whiskers.
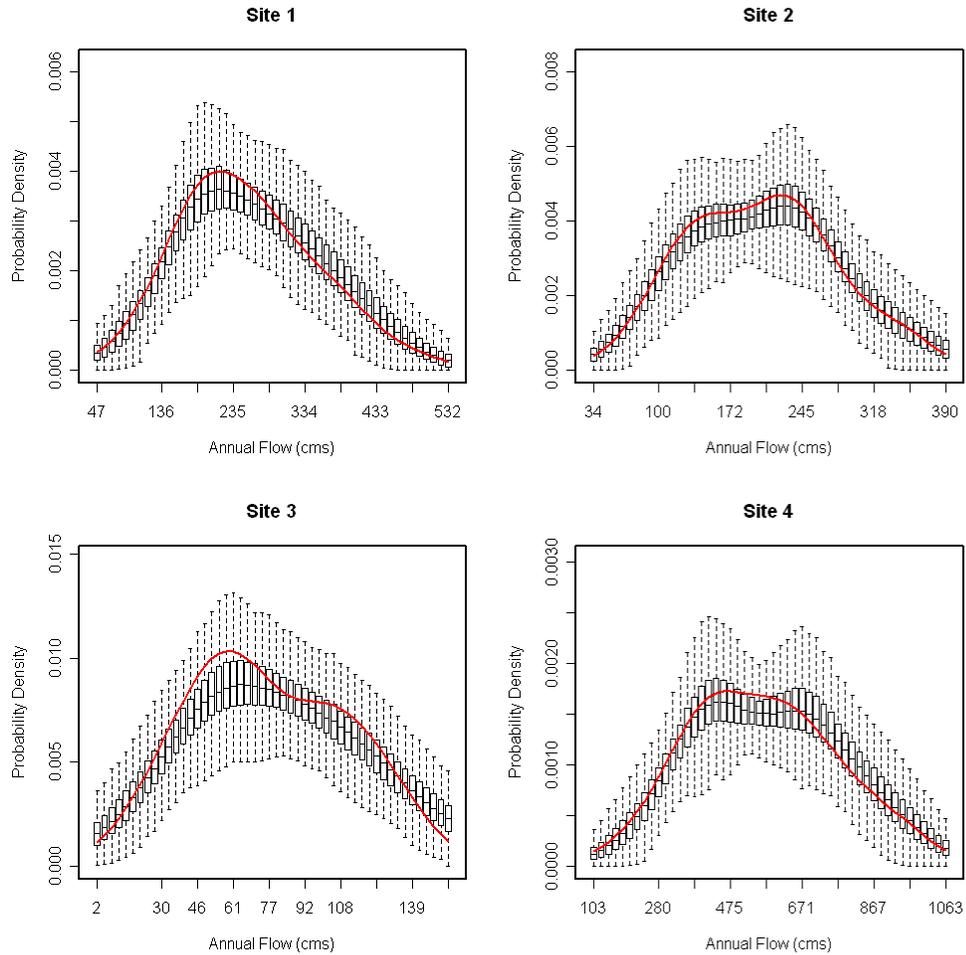
Figure 3-3 Probability density functions for annual flow.

The annual correlations between flow and salt at the four sites are shown in Figure 3-4. The left side results are from approach 1 - one single annual flow versus salt regression; while the right side results are from approach 2 - four annual flow versus salt regressions representing each site. The correlation between annual flow and salt generated from approach 1, left side, are under represented at the three upstream sites though the farthest downstream site, site 4, is preserved the best. Results from approach 2, right side, display an improvement in the natural flow and salt correlation; this is because, in this approach, the salt is generated from the individual site flow and salt regressions. Overall the flow and salt correlations are

better captured from the second approach, as expected. Site 3 is slightly over-correlated which may result from the limited sample size discussed earlier.
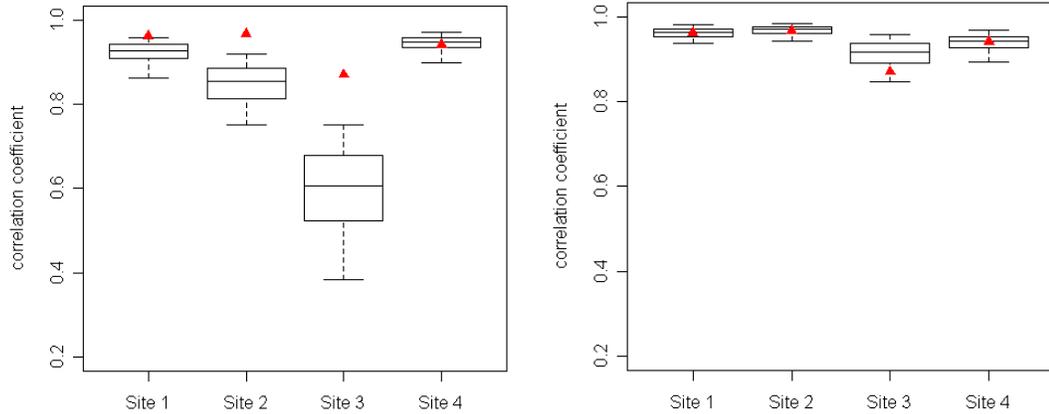
Figure 3-4 Correlation coefficient for annual flow and salt from single annual regression (left) compared with four annual regressions (right).

Figures 3-5 and 3-6 display annual salt PDFs for all sites from approaches 1 and 2, respectively. Approach 2 is more capable of fully preserving the observed PDF at each site though site 3 is not fully captured because the annual flow PDF (Figure 3-3) was not fully captured.
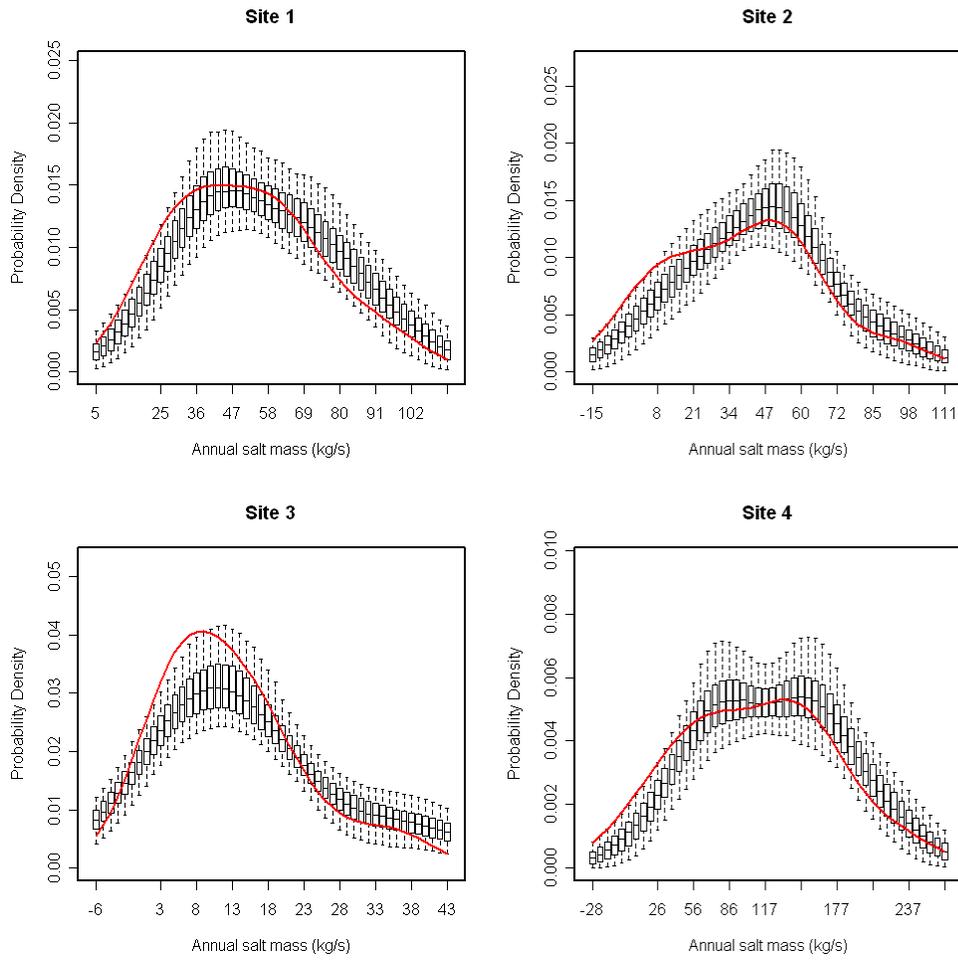
Figure 3-5 Probability density functions for annual salt mass from single annual regression.
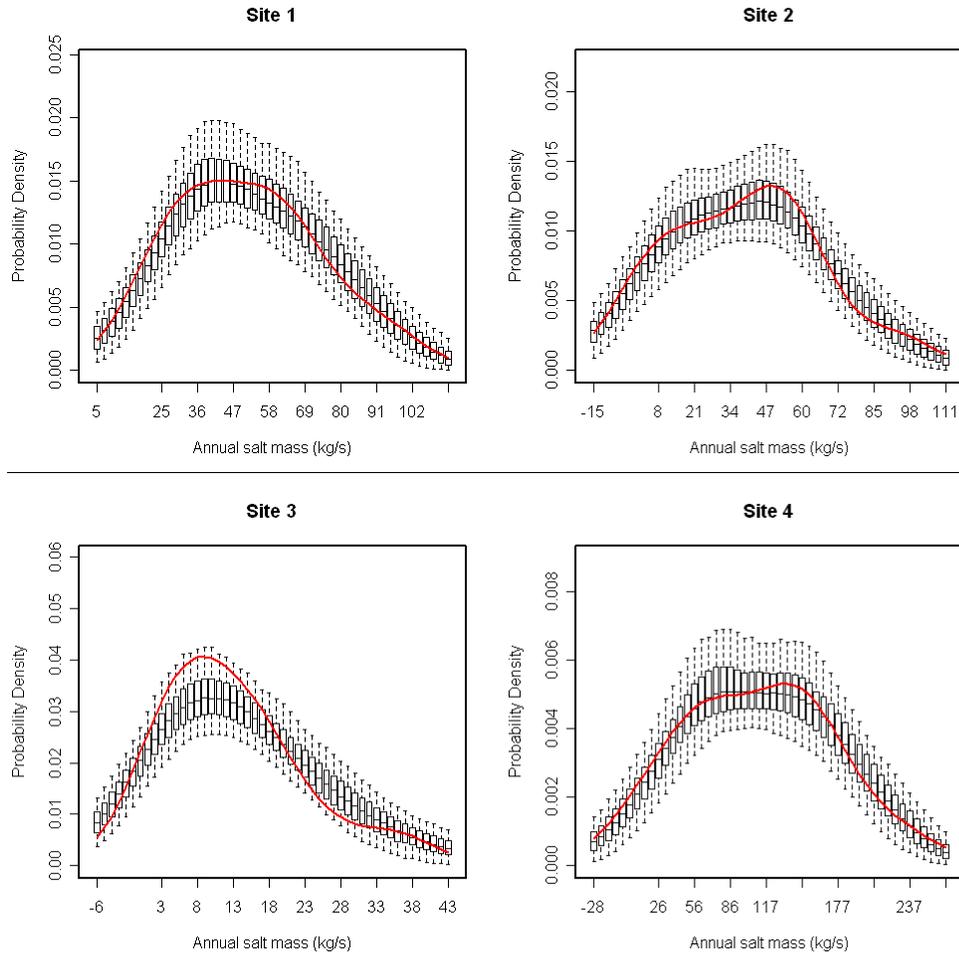
Figure 3-6 Probability density functions for annual salt mass from four annual regressions.

Figure 3-7 displays the spatial cross correlation among the four sites for approach 1 (left side) and approach 2 (right side). Both the methods are able to preserve the annual cross correlation among sites. But the first approach is able to better preserve the monthly spatial cross correlation since the salt values are simulated from a space-time disaggregation of the salinity at the index gauge.

47
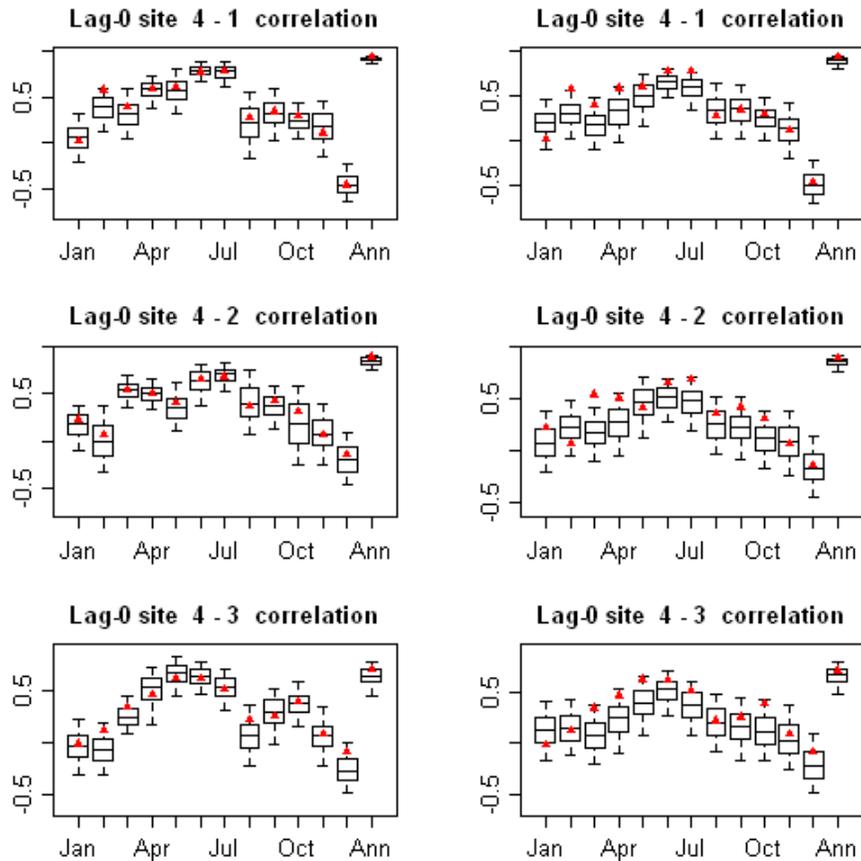
Figure 3-7 Single annual regression (left) compared with four annual regressions (right).

Basic performance statistics for approach 2 are shown in Figure 3-8. Both methods (approach 1 not shown) perform well in preserving the lower order statistics. Generally, almost all of the monthly statistics are well preserved within the IQR with a couple exceptions. Considering that the annual salt at each gauge is simulated from the flow-salt nonparametric regression and subsequently disaggregated to monthly values, this result was not guaranteed. The Lag-1 correlation for the January of the current year with December of the previous year is not explicitly preserved with the disaggregation algorithm and therefore, not well captured by either approach. For site 4 the maximum and minimum of the monthly salt simulations generally do not exceed the historic maximum or go below the historic minimum values. This results

48

mainly from a smaller K-NN size. With increased K the variability in the generated values from the annual natural flow generation can be increased though at a cost of reduced performance in other basic statistics. Negative values can be obtained since the historic salt mass is a computed value. This represents salt removed from the river system.
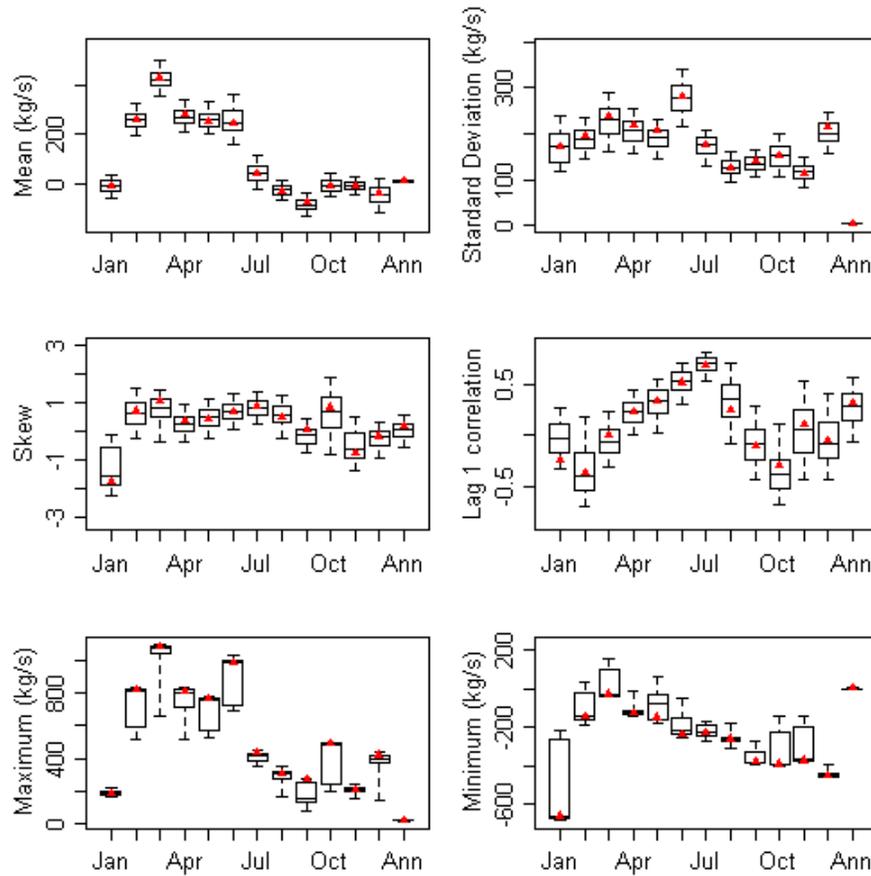
Figure 3-8 Boxplots of monthly and annual basic statistics for total salt load at site 4. Salt load modeled from four annual regressions i.e., approach 2.

An exceedance probability plot of violating a proposed threshold is presented in Figure 3-9 from simulations generated using approach 2. For comparison we generated flows from a parametric disaggregation method using an ARMA(1,0) (Salas, 1985) to model the index gauge. The disaggregated flows were then used to generate natural salt at each of the four sites with the site's respective regression. The

flows from the nonparametric method capture the observed exceedance probability well within the IQR while the parametric framework captures the observed at the limits of the IQR. This results because the nonparametric disaggregation method is able to fully capture the annual flow (Figure 3-3) and salt mass (Figure 3-6) PDF at site 4 while the parametric disaggregation method can only reproduce a Gaussian PDF of flow and resulting salt mass at site 4 (Figure 3-10).
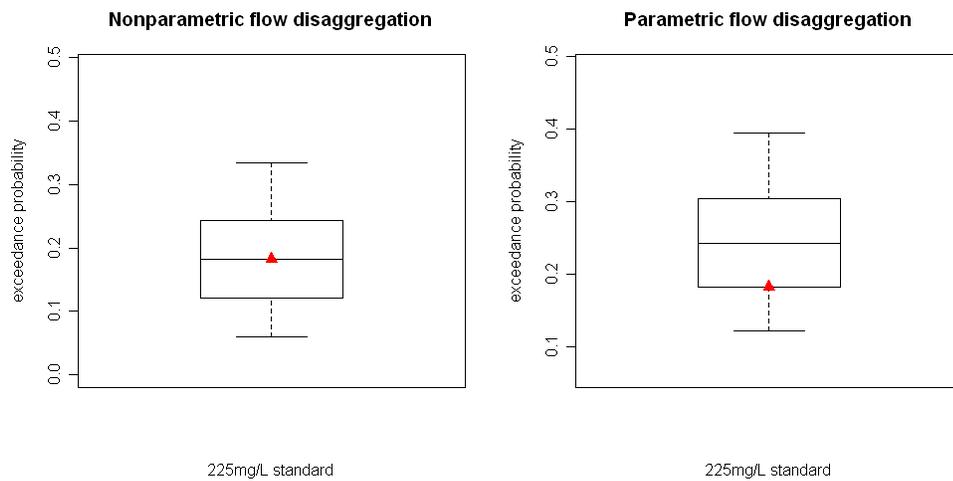
Figure 3-9 Exceedance plots for a salinity threshold of 225mg/L.
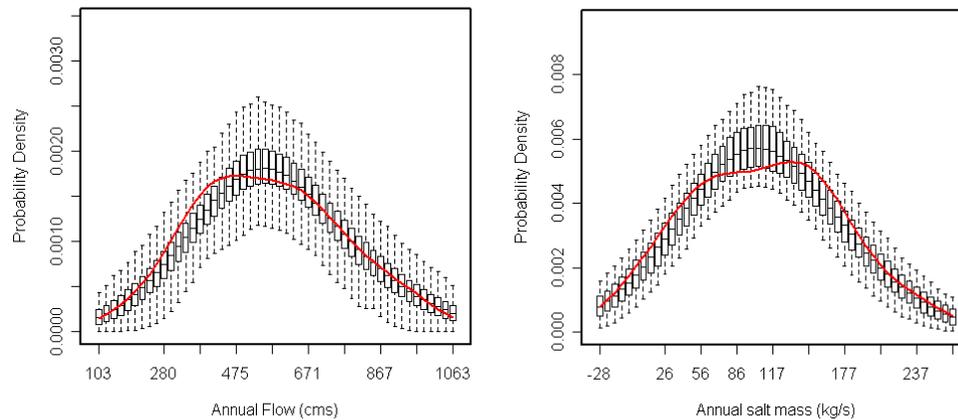
Figure 3-10 Probability density function for annual flow (left) generated from parametric disaggregation and the associated annual salt mass (right) from the flow-salt regression at site 4.

3.5 Summary and Discussion

We developed a framework to simulate basin wide salinity scenarios simultaneously that preserves the spatial and temporal statistics. This framework combines the nonparametric space-time disaggregation model (Prairie et al., 2006a) for basin wide flow generation and a local polynomial regression based approach (Prairie et al., 2005) for salt estimation given a streamflow value. Within this framework two approaches were presented.

The first approach coupled flow and salt through a single regression of annual salt as a function of flow at an index site. Once the natural salt was determined from the regression the annual natural salt at the index site was first spatially disaggregated to 4 sites then these annual values were temporally disaggregated from annual to monthly values.

In the second approach the annual flow at the index gauge was first spatially disaggregated to the 4 individual sites. Annual salt was then simulated at the four locations based on their respective annual flow-salt nonparametric regression. The last step disaggregated the annual salt values at each site to monthly values, completing the process.

We demonstrated that approach 1 was better at preserving the spatial cross correlation across the 4 sites and could capture most of the flow and salt correlation at each site, but not entirely. Approach 2 was not able to capture the spatial cross correlation as well, but was superior at capturing both the correlation between flow and salt at the annual and monthly temporal scales. Both approaches were able to capture lower order statistics such as the mean, standard deviation, coefficient of

skew, lag 1 correlation, maximum, and minimum well but framework 2 was slightly better.

The choice of one of the approaches will depend on each application. If the preservation of the spatial cross correlation is most important then approach 1 would be best suited and if, capturing the flow-salt correlation is desired then the second approach may be best to implement.

Numeric salinity criteria for the Colorado River Basin require the maintenance of specified average annual salinity levels at three key points in the Lower Colorado River. With the need to model the relationship between flow and salt as accurately as possible we deemed the need to capture the correlation between annual flow and salt to be most important for our application in the CRSS; hence, we recommend approach 2 for this purpose.

Paleo reconstructed flows can also be used in this framework to generate salinity scenarios conditioned on paleo reconstructed flows. This provides a richer variety of flow and salt scenarios that will be of great value in developing reservoir operational policies and salinity mitigation strategies.

There is increasing evidence of the influence of large-scale climate features on the streamflow variability in western United States. Generating flow and salt scenarios conditioned on large-scale climate information can be a daunting task with traditional methods. The framework developed here provides an attractive and simpler, yet robust alternative.

CHAPTER 4

A STOCHASTIC NONPARAMETRIC APPROACH FOR

STREAMFLOW GENERATION COMBINING OBSERVED

AND PALEO RECONSTRUCTED DATA

Observed streamflow records are frequently short in length for many locations. In the Colorado River Basin a record of 100 years is a blessing to most water managers though even this longer observed record can be inadequate for understanding the probability and magnitude of both drought and surplus characteristic in a river basin. Paleo reconstructed streamflow records can extend the understanding of observed past records five-fold leading to enhanced understanding of long-term basin flow characteristic. In this Chapter we present a nouvelle approach to condition synthetic streamflow generation on paleo reconstruction system state (i.e., wet or dry states).

4.1 Introduction

Effective long-term planning and management of water resources requires (i) a tool that can generate all plausible streamflow scenarios and (ii) a decision model to evaluate policy alternatives. Stochastic models are typically built on observed streamflow data, which are then used to generate flow scenarios that capture the statistics of the observed data. There is a rich literature on models to simulate basin wide flows in a linear (Valencia and Schaake, 1973; Mejia and Rousselle, 1976; Tao and Delleur, 1976; Lane, 1979; Salas et al., 1980; Todini, 1980; Stedinger and Vogel, 1984; Salas, 1985; Stedinger et al., 1985; Koutsoyiannis, 1992; Santos and Salas, 1992; Koutsoyiannis and Manetas, 1996; Koutsoyiannis, 2001) or nonlinear (Tarboton et al., 1998; Kumar et al., 2000; Sharma and O'Neil, 2002; Srinivas and

Srinivasan, 2005) framework. Observed data is usually limited in time and thus, the simulations have a limited variability; more so in the magnitude and frequency of the extremes, which are crucial for robust long-term planning. This was underscored on the Colorado River basin during the recent severe and sustained drought. The basin experienced the worst drought on record from 2000-2004. Though this drought was unprecedented in the observed record (1906-2004), paleo reconstructions of streamflow from tree-ring chronologies have shown droughts of greater magnitude and duration. A recent paleo reconstructed streamflow for the period 1490-1997, on the Colorado River at Lees Ferry, AZ, a key gauge on the River (Woodhouse et al., 2006), is shown in Figure 4-1 along with the observed flows. It is evident that the recent drought is unprecedented during the observed period but the reconstructed streamflows prior to 1906 show severe droughts of 5 years length at least four times over the approximate 500 year period, indicating that the recent drought is not unusual.
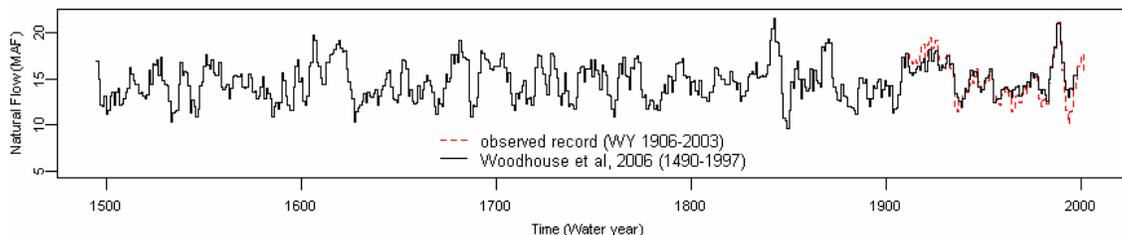
Figure 4-1 Five-year running means for historic and reconstructed streamflow.

Clearly, the rich information provided by paleo reconstructed streamflows has to be incorporated in stochastic streamflow models to enable the generation of a realistic variety of plausible flow scenarios. However, the magnitudes of streamflow reconstructions have a high degree of uncertainty due to the reconstruction methodology. Typically, a regression model is fit to the observed streamflow with a

suite of tree-ring observations as the predictors. This fitted model is then used to estimate streamflows in the pre-observed period using the tree-ring observations (Meko et al., 1995). The key assumptions that can be difficult to validate are (i) the tree-ring and streamflow relationship is assumed to be linear and stationary, (ii) the suite of tree-ring predictors in the regression model are also assumed to be stationary. Consequently, the reconstructed streamflows can be sensitive to the fitted model as demonstrated by Hidalgo et al. (2000). This apparent weakness of the paleo reconstructed flow data have made their use in a water resources planning context contentious, despite the availability of paleo reconstructed data for many decades. Despite these apparent weaknesses, few argue about the duration and frequency of dry and wet (i.e., the hydrologic state) periods from the reconstructions. The key question is how to combine the long paleo reconstructed streamflow information of lesser reliability with the shorter but reliable observed data to develop a framework for simulation of streamflow scenarios?

To address this question we propose a new framework in which the hydrologic state (i.e., wet or dry) is modeled using the paleo reconstructed data and the flow magnitudes derived from the observed data. Specifically, a nonhomogeneous Markov Chain model (Rajagopalan et al., 1996) is built on the paleo data that is then used to simulate the hydrologic state. The flow magnitudes are then generated conditioned on the simulated hydrologic state using a K-nearest neighbor (K-NN) conditional time series bootstrap (Lall and Sharma, 1996); thereby, using the strengths of both of these data sets. The framework is described in detail in the following sections.

The chapter is organized as follows. The data sets used are first described in detail followed by a description of the proposed framework and its implementation algorithm. Next the framework is applied to the Lees Ferry, AZ streamflow gauge on the Colorado River where both, paleo reconstructions and observed streamflow data are available and a suite of statistics is evaluated. Summary of the results and the potential applications for basin wide water resources planning are described, concluding the chapter.

4.2 Data Sets

As mentioned earlier, two data sets (i) paleo reconstructed streamflow and (ii) observed flows, are used in this study. These are described in detail below.

4.2.1 Natural Streamflow

The natural streamflow data for the Colorado River Basin are developed by Reclamation and updated regularly. Naturalized streamflows are computed by removing anthropogenic impacts (i.e., reservoir regulation, consumptive water use, etc.) from the recorded historic flows[3]. Prairie and Callejo (2005) present a detailed description of methods and data used for the computation of natural flows in the Colorado River Basin. For this study the annual water year (Sep-Oct) natural streamflow at Lees Ferry, AZ for the period 1906-2004 are used.

4.2.2 Paleo Reconstructed Streamflow

The latest annual water year streamflow reconstructions at the Lees Ferry, AZ gauge, by Woodhouse et al. (2006) for the period 1490-1997 are used in this study. These are streamflows reconstructed from tree-ring information. Tree-ring widths are

---

[3] The natural flow data and additional reports describing these data are available at http://www.usbr.gov/lc/region/g4000/NaturalFlow/index.html.

influenced by climate and available soil moisture and thus, are good integrators of the weather fluctuations; just as streamflow is a watershed integration of hydrologic and climatologic processes. Consequently, the tree-ring widths are well correlated with annual runoff. A series of trees are cored at multiple locations, chosen such that the tree species have annual rings sensitive to moisture availability. Selecting the species and the location is very important for this effort (Meko et al., 1995). Two core samples are taken from each tree for cross dating and the ring widths are measured, obtaining the chronology of tree-ring widths. The attractive aspect of tree-ring based reconstructions, unlike other paleo proxy data, is trees that put on annual rings have natural dating, with the outer ring corresponding to the current year and the subsequent inner rings to past years. A standard series of techniques (Stokes and Smiley, 1968; Swetnam et al., 1985) are employed to process the ring width series. Typically, the series is first detrended to remove the effects of reduced ring width with aging. Next, the ring width series from various cores at a single location are combined to develop a site chronology (Cook et al., 1990). The site chronology is related to observed streamflow during the overlap period; typically, a multiple linear regression model is fit (Weisberg, 1985). For the Colorado River at Lees Ferry, AZ, gauge the regression model developed by Woodhouse et al. (2006) using all the available pool of chronologies (30 in total), explains approximately 84% of the annual variance of the observed streamflow, which is very good. The fitted regression model is then used to estimate the streamflow during the pre-observation period when tree-ring information is available; thus, obtaining the reconstructed streamflow series.

57

Proxy data, such as tree-ring information, are no substitute for actual observations and therefore are not without fault. Especially, during high streamflow flow periods it is known that the tree-ring widths are influenced by variables other than moisture availability; thus, degrading their ability in accurately representing high flow years (Woodhouse and Brown, 2001). Further, different datasets and techniques to process tree-ring information can result in substantial differences in the reconstructed flows (Hidalgo et al., 2000). This can be seen in Figure 4-2, where four different streamflow reconstructions at the Lees Ferry, AZ gauge are shown, including the earliest reconstruction of Stockton and Jacoby (1976); later reconstructions by Hidalgo et al. (2000); that of Hirschboeck and Meko (2005) as part of the Salt River Project and, the most recent reconstruction by Woodhouse et al. (2006).
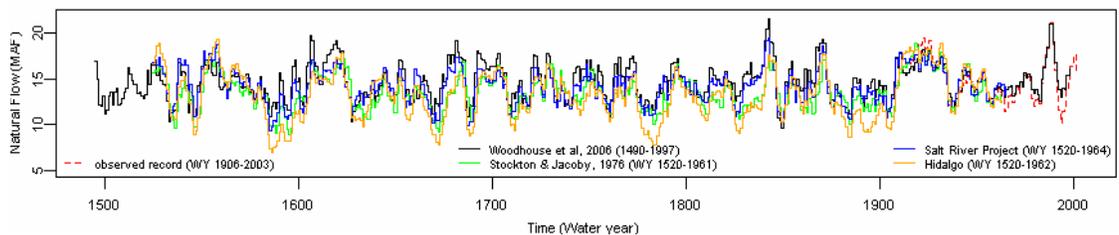
Figure 4-2 Five-year running means for recent and previous streamflow reconstructions at Lees Ferry.

Each reconstruction used a different set of tree-ring chronologies and different processing methods. Of particular interest is the increased severity of drought and reduced overall mean displayed by the Hidalgo reconstruction. Unfortunately, the variability among the reconstructions have not helped instill confidence in use of these data by policy makers and water managers in the Colorado River Basin, even with growing interest in wanting to use them. Despite the differences, all the reconstructions agree on wet and dry years exhibiting decreasing and increasing

58

streamflows during the dry and wet years, respectively. This is shown in Figure 4-3 where it is evident that the wet and dry periods (as defined with respect to the median streamflow of the observed period as the threshold) are consistent across the four reconstructions. This offered the potential to use the paleo reconstructed streamflows to model the hydrologic state (i.e., wet or dry) of the system and use the observed data for the flow magnitude. This forms the basis of our proposed framework.
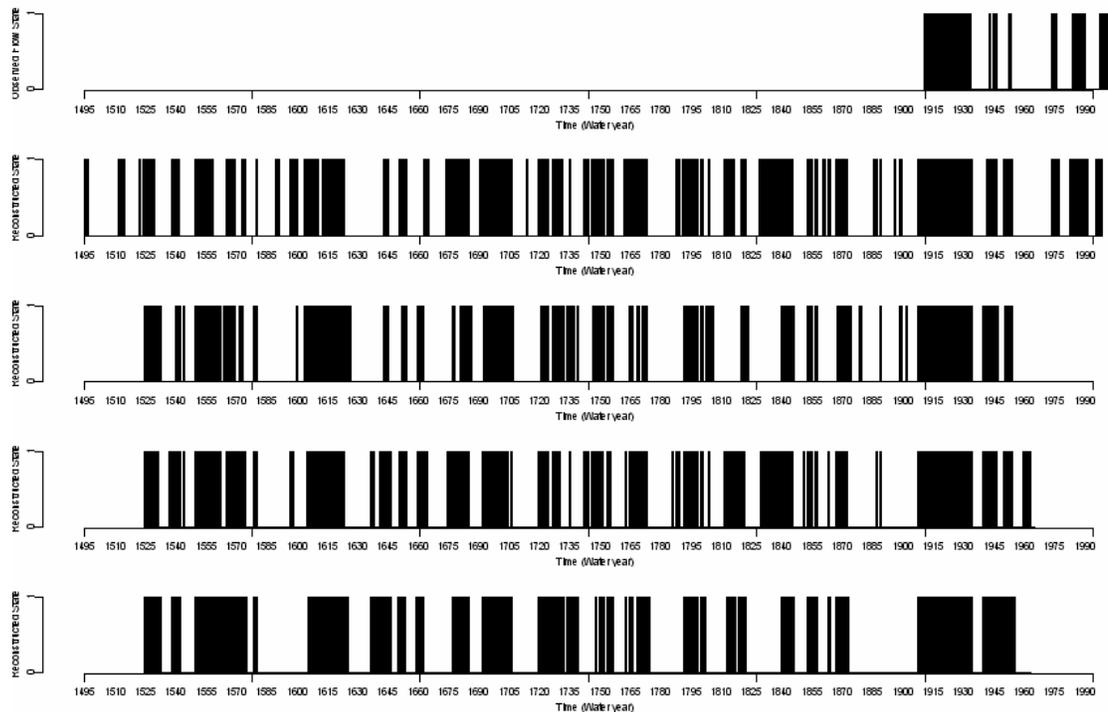
Figure 4-3 Five-year running means streamflow state for historic and reconstructed.

4.3   Proposed Framework

As mentioned above, the proposed framework combines the paleo reconstructed streamflows with the observed data in developing a framework for simulating robust streamflow scenarios for use in water resources management. The paleo reconstructed data is used to model the hydrologic state of the system. The median of the observed flows is considered the threshold to define periods as wet, if flow is greater than this

threshold, and dry, if flow is less than this threshold. From Figure 4-2 and 4-3 it can be seen that there are epochs of enhanced wet and dry periods and therefore, the propensity (i.e., the probability) of being in these states longer. Since, the state transition is varying through time a nonhomogeneous Markov Chain modeling approach is required. The streamflow magnitudes are then simulated from the conditional probability density function, using a nonparametric K-nearest neighbor bootstrap approach. The proposed framework is shown in Figure 4-4. The descriptions of these two components of the framework along with background information are provided below.

Nonhomogeneous Markov model
with smoothing

Generate system state
(S_t)

Generate flow conditionally
(K-NN resampling)
$f(x_t | S_t, S_{t-1}, x_{t-1})$
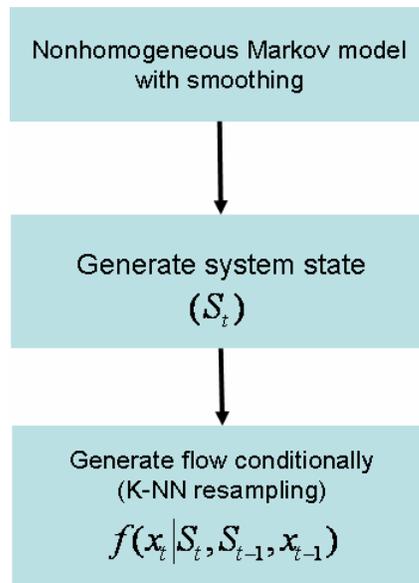
Figure 4-4 Modeling framework description.

4.3.1 Modeling the Hydrologic State

As mentioned above, Markov chains are the approach of choice for modeling the discrete state (wet or dry) transitions of the hydrologic state. Markov chains have been extensively used to model daily precipitation occurrence (Gabriel and Neuman, 1962; Todorovic and Woolhiser, 1975; Smith and Schreiber, 1974; Salas, 1985 and references within). Typically, for a two-state (wet, dry) first order (i.e., state transition

at the next time step depends on the current state), the transition probabilities are directly estimated from the data by counting the proportion of transitions to a wet year from a dry year, P_{dw}, and the probability of a wet year followed by a dry year, P_{wd}. The probability of a dry year followed by a dry year can be obtained as $P_{dd} = 1 - P_{dw}$ likewise, the probability of a wet year followed by a wet year as $P_{ww} = 1 - P_{wd}$. The transition probabilities can be readily used to simulate the hydrologic states and consequently, their frequencies. If these transition probabilities are assumed to be stationary and calculated from the entire data then it is a stationary Markov Chain. But here (Figure 4-2 and 4-3) the frequencies of wet and dry periods are varying (i.e., nonstationary) over time and hence, stationary transition probabilities are inadequate.

The nonstationarity can be addressed in several ways. A moving window of some *W* time steps can be selected and the transition probabilities estimated for each time window and repeated by moving forward every time step. Resulting in transition probability estimates for each year, based on state observations present in the window length, this is a simple approach. Hidden Markov models are gaining popularity, in these the underlying epochal (or regime) changes are modeled probabilistically and the transition probabilities are then conditionally estimated based on the epoch. These models have been applied to precipitation, climate and streamflow data (see e.g., Zucchini and Guttorp, 1991; MacDonald and Zucchini, 1997; Lu and L.M. Berliner, 1999; Thyer and Kuczera, 2000, 2003a,b; Akıntuğ and Rasmussen, 2005). These Markov models require extensive fitting and calibration which can limit their use in operational studies. Another approach to deal with nonstationarity is the

nonhomogeneous Markov models (Hughes and Guttorp, 1994; Hughes et al., 1999; Bellone et al., 2000; Lambert et al., 2003). For example, Fourier series were fit to model the changing transition probability with season for precipitation (Woolhiser and Pegram, 1979; Roldan and Woolhiser, 1982; Feyerherm and Bark, 1965).

Nonparametric alternatives (e.g., Rajagopalan et al., 1996; Mehrotra et al., 2004; Mehrotra and Sharma, 2005) offer a more general approach in which the transition probability at any time t is estimated as a weighted average of the transitions within a window of size H centered on t. The window size H is obtained from objective criteria. Here we used the nonparametric nonhomogeneous Markov Model (NHM) developed by Rajagopalan et al., (1996). The development was for daily precipitation, which is adapted here for modeling the streamflow states described below.

The streamflow state occurrence process is illustrated in Figure 4-5. From this process four different aspects of the process can be extracted.
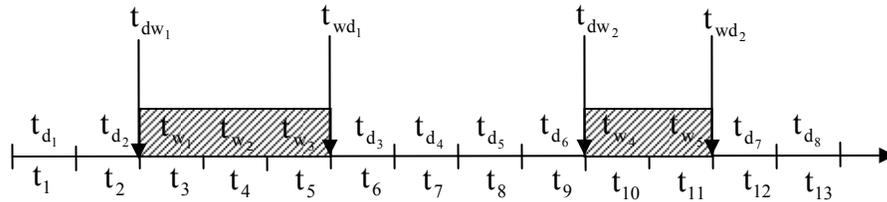
Figure 4-5 State occurrence process.

These include (1) the year indices $t_{d_1}, t_{d_2}, ..., t_{d_{nd}}$ for the nd dry years; (2) the year indices $t_{w_1}, t_{w_2}, ..., t_{w_{nw}}$ for nw wet years; (3) the year indices $t_{dw_1}, t_{dw_2}, ..., t_{dw_{mdw}}$ for ndw years on which the transition from dry to wet occurs; and (4) the year indices $t_{wd_1}, t_{wd_2}, ..., t_{wd_{mwd}}$ for nwd years on which the transition from wet to dry occurs. The transition probabilities, $P_{dw}(t)$ and $P_{wd}(t)$ for a given year are estimated by a discrete nonparametric kernel estimator given in Rajagopalan et al. (1996) as:

$$P_{dw}(t) = \frac{\sum\limits_{i=1}^{ndw} K\left(\dfrac{t - t_{dw_i}}{h_{dw}}\right)}{\sum\limits_{i=1}^{nd} K\left(\dfrac{t - t_{d_i}}{h_{dw}}\right)} \tag{1}$$

$$P_{wd}(t) = \frac{\sum\limits_{i=1}^{nwd} K\left(\dfrac{t - t_{wd_i}}{h_{wd}}\right)}{\sum\limits_{i=1}^{nw} K\left(\dfrac{t - t_{w_i}}{h_{wd}}\right)} \tag{2}$$

where, $K(\ )$ = the kernel function; $h_{()}$ = the kernel bandwidth; t = year of interest.

The discrete Quadratic Kernel function developed by Rajagopalan and Lall (1995) is used, which is given as:

$$K(x) = \frac{3h}{(1 - 4h^2)}(1 - x^2) \quad \text{for} \quad |x| \le 1 \tag{3}$$

where $x = \dfrac{t - t_{()}}{h_{()}}$ measures the distance for event $t_{()}$ from the year of interest t

within the bandwidth $h_{()}$, where $h_{()}$ is an integer. The weights from the Kernel

function are positive and sum to unity. It can be seen that the estimates of transition

probabilities at any year t are based only on the transitions within a window $t - h_{()}$ to

$t + h_{()}$.

The transition probability estimators (1) and (2) are fully defined once the

bandwidth $h_{()}$ is determined for each. An objective method based on a least square

cross validation (LSCV) procedure (Scott, 1992) is used to select the optimal

bandwidth that was also used by Rajagopalan et al., (1996) given as:

$$\text{LSCV}(h) = \frac{1}{n} \sum_{i=1}^{n} [1 - \hat{P}_{-t_i}(t_i)]^2 \tag{4}$$

where $n =$ the number of observations (ndw or nwd); $\hat{P}_{-t_i}(t_i) =$ the estimate of

the transition probability ($\hat{P}_{dw}$ or $\hat{P}_{wd}$) at year t dropping the information on year t.

The 1 in (4) results from an assumption that the prior probability of transition is 1 for

the years on which a transition has occurred. The minimizer of the LSCV is selected

as the optimal bandwidth. The bandwidths h_{dw} and h_{wd} are objectively determined

and subsequently used in the estimators (1) and (2) to estimate the transition

probabilities for each year.

Best Markov Chain model orders are generally selected as the minimizers of

AIC criteria (Gates and Tong, 1976). For the Lees Ferry paleo reconstructed data we

found the two state first order to be optimal.

4.3.2 Modeling the Flow Magnitudes

The streamflow magnitudes, as mentioned earlier, are modeled based on the

observed data and conditioned upon the hydrologic state. This model can be

described as the conditional probability density function (PDF):

$$f(x_t | S_t, S_{t-1}, x_{t-1}) \tag{5}$$

where the flow at the current time $t = x_t$ conditioned on the current system state $=$

S_t, previous system state $= S_{t-1}$, and previous flow $= x_{t-1}$.

Simulation from this conditional PDF is achieved by a K-NN bootstrap method

(Lall and Sharma, 1996; Rajagopalan and Lall, 1999). K-NN are identified in the

observed data of the current feature vector $[S_t, S_{t-1}, x_{t-1}]$. One of the neighbors is

selected, based on a metric that gives highest probability to the nearest neighbor and

lowest to the farthest. The corresponding streamflow of the selected neighbor is the simulated value for the current time.

This is a unique case where the feature vector includes discrete and continuous variables. Therefore, determination of the feature vector is split into two steps. The observed data is first split into the lag and state combination. For this example a lag one two state (wet and dry) system 4 categories are defined. In the second step, a given S_{t-1}, S_t combination defines a category from which the K neighbors are identified and one neighbor x_{t-1} is resampled as mentioned above, and the next year x_t from the observed record is chosen.

In this work K is based on the number of neighbor in each category, as the limited observed data limit K. The number of nearest neighbors can also be based on the heuristic scheme $K = \sqrt{N}$ where N equals the sample size (Lall and Sharma, 1996), following the asymptotic arguments of Fukunaga (1990). Objective criteria such as Generalized Cross Validation (GCV) can also be used.

4.3.2.1 Implementation Algorithm

The complete framework combines the two models. The simulation proceeds as follows. First a simulation horizon is identified, this is application dependent, suppose a 30-year horizon is chosen.

1. Randomly resample a block of 30-years from the paleo reconstructed streamflows, say 1651-1680.
2. Generate flow states $S(t)$ where $t = 1, 2, ..., 30$ using the transition probabilities of the resampled years from step 1 above.

3. Generate flow magnitudes $x(t)$ for each $t = 1, 2, ..., 30$ from the conditional PDF $f(x_t | S_t, S_{t-1}, x_{t-1})$ using the K-NN bootstrap approach described in the previous section.

4. Repeat steps 2 through 3 to obtain as many simulations as required.

4.4 Model Evaluation

The proposed framework is applied to the paleo reconstructed streamflows (1490-1997) and observed natural flows (1906-2003) at Lees Ferry, AZ on the Colorado River.

A suite of basic statistics are computed including the annual (i) mean, (ii) standard deviation, (iii) coefficient of skew, (iv) maximum, (v) minimum, and (vi) lag-1 autocorrelation. Surplus and drought statistics include the longest surplus (LS), longest drought (LD), maximum surplus (MS) volume, maximum deficit (MD) volume, average length surplus (avgLS), average length drought (avgLD), average surplus (avgS), and deficit (avgD) volume. Surplus (drought) is defined as values above (below) a threshold, here the median of the observed record. Figure 4-6 describes the computation of these surplus and drought statistics based on the threshold.
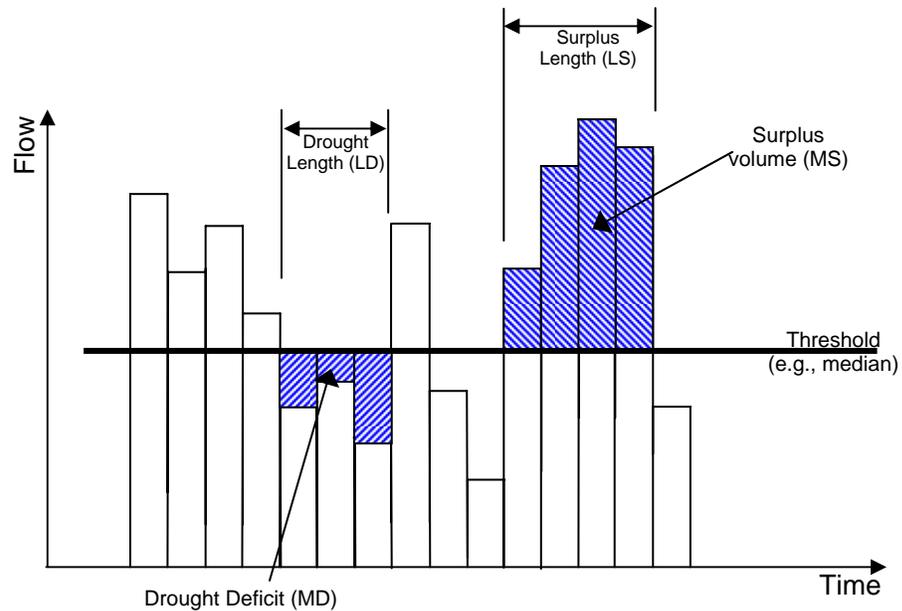
Figure 4-6 Definition of surplus and drought statistics.

The results are displayed as boxplots where the box represents the interquartile range and whiskers extend to the 5th and 95th percentile of the simulations and outliers are shown as points beyond the whiskers. The statistics of the observed data are represented as a triangle and paleo reconstructed data are represented as a circle. Performance on a given statistic is judged as good when the observed statistic falls within the interquartile range of the boxplots, while increased variability is indicated by a wider boxplot.

4.5 Results

To demonstrate the capability of the framework to capture the statistics of the observed period we did the following. The framework was applied using paleo data from the 92 year overlap with the observed period of 1906-1997, and 500 simulations were made each 98 years long. The statistics of the observed data should be well captured in this set up. The boxplots of the various basic statistics are shown in Figure 4-7. It can be seen that all the basic statistics are preserved well except the lag 1

67

correlation, which is slightly under simulated. The maximum and minimum are bounded by the observed record as expected since the K-NN bootstrap approach resamples only the observed data.
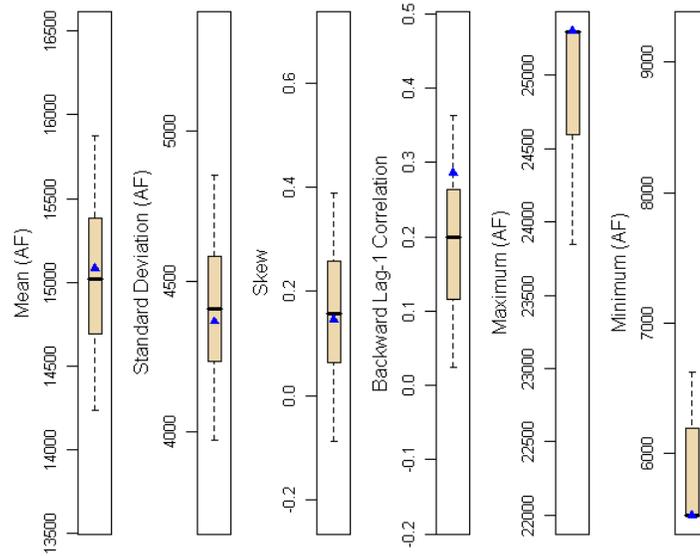
Figure 4-7 Basic statistics based on reconstruction subset 1906-1997 TP.

Boxplots of surplus and drought statistics are shown in Figure 4-8 along with the corresponding values from the observed record represented as a triangle and from the reconstructed record for the same as a circle. The boxplots over simulate LS and LD with the median of the surplus slightly higher than the drought. This is consistent with the fact that the observed period is generally wet compared to prior epochs. Other drought and surplus statistics were captured quite well by the simulations except MAXD. The framework is not designed to capture all the drought statistics, given this, the performance is quite good.
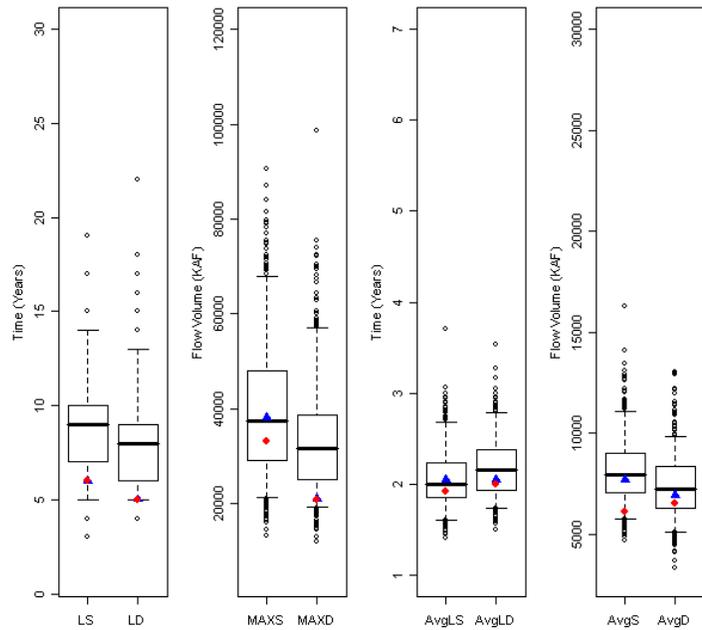
Figure 4-8 Drought statistics based on reconstruction subset 1906-1997 TP.

We next apply the framework to the entire paleo reconstructed streamflows and the observed data. Only the drought and surplus statistics are presented and discussed, as they are of interest. The simulations generated higher LS and LD and larger MS and MD statistics compared to that of the observed record (Figure 4-9) and also to those seen in Figure 4-8. The transition probabilities of the entire paleo reconstruction allow generation of a greater variety of drought and surplus lengths than those from the overlapping period (Figure 4-8).
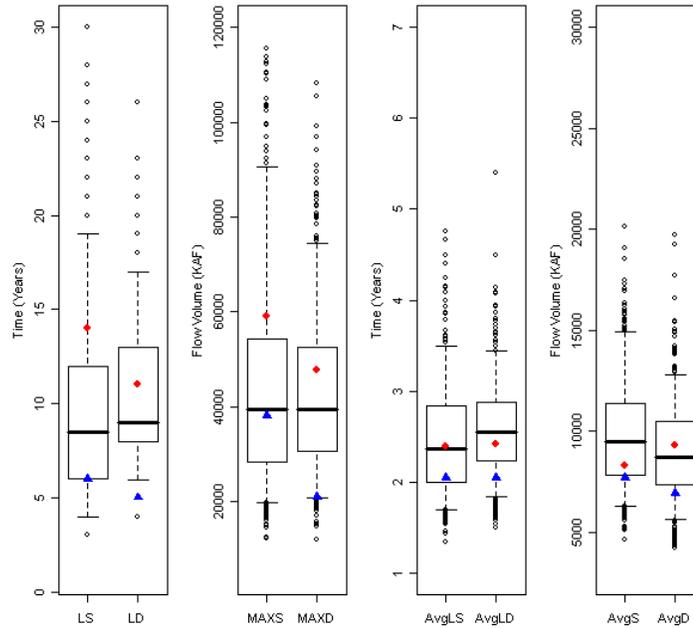
Figure 4-9 Drought statistics based on full reconstruction 1490-1997 TP.

In the Colorado River Basin the critical sequence of concern is a series of droughts connected over 12 years with surplus years interspersed. Such sequences are not represented in the drought statistics described above. Furthermore, the drought and surplus statistics estimated above are based on a pre-selected threshold (here it is the median streamflow of the observed period). Thus, the results are sensitive to this selected threshold. To avoid this, a better approach is to determine the required storage for a given streamflow sequence to meet various demand levels. This incorporates the effect of multiple linked droughts, thus is more realistic in representing critical droughts. The algorithm, termed the sequent peak algorithm (Loucks et al. 1981) used for this purpose is given as:

$$S_i' = \begin{cases} S_{i-1}' + d - y_i \\ 0 \end{cases} \tag{6}$$

$$S_c = \max\left[S_i', ..., S_N'\right] \tag{7}$$

Where S_i is the storage at time step i, d is the demand or yield, y_i is the streamflow from a sequence at time i, and S_c is the storage capacity. This is also widely used for designing reservoir capacities.

The algorithm is run for various demand (yield) levels with the historic flow (triangle) and each trace of the 500 simulations (boxplots) shown in Figure 4-10.
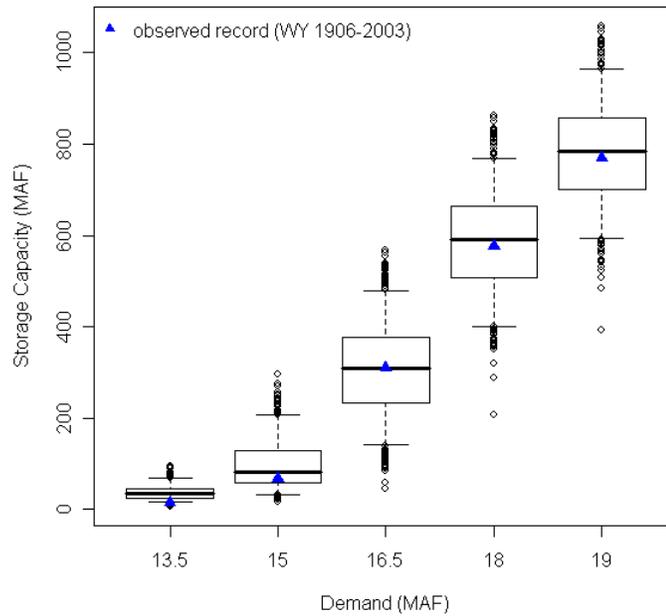
Figure 4-10 Sequent peak results based on full reconstruction TP.

Consider the boxplot corresponding to the demand of 16.5 MAF. It can be observed that to reliably meet this demand based on the historic inflow sequence a storage capacity of 325 MAF (triangle) is required. The boxplot indicates the variability in the required storage capacity based on 500 traces simulated from the combined framework. The boxplot, plotted as a PDF (Figure 4-11a) or a cumulative distribution function (CDF) (Figure 4-11b), can easily be used to find the reliability. For example, a demand of 16.5 MAF cannot reliability be met 99% of the time for a storage capacity of 60 MAF (the approximate current storage capacity of the

Colorado River basin). The reliability is the area under the PDF curve below 60MAF

which is 1 - area of the hatched region in Figure 4-11a, or $(1 - 0.99 = 0.01)$ as read

from the CDF. The reliability of alternate storage capacities can be found from Figure

4-11a or 4-11b in a similar manner.
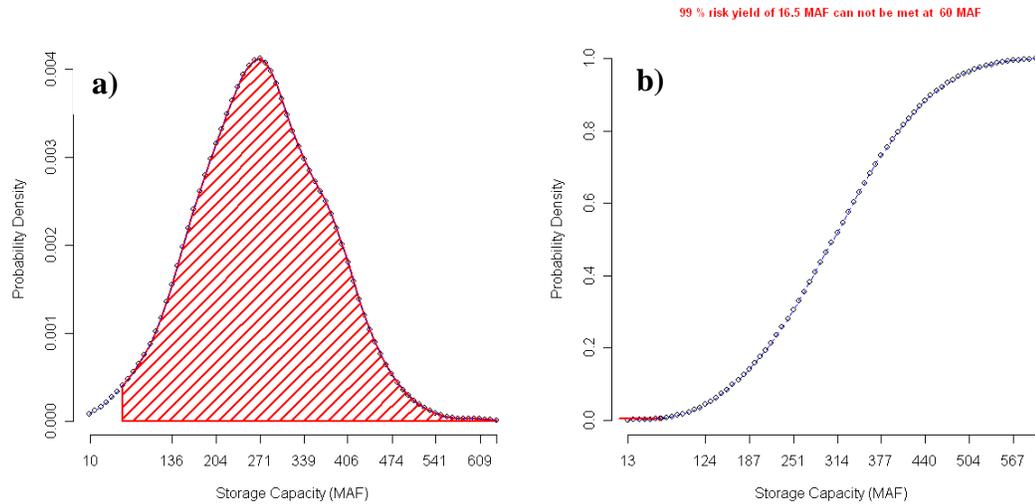
Figure 4-11 a) PDF and b) CDF for 16.5 MAF demand boxplot from Figure 4-10.

The sequent peak method assumes that the demand level is constant through time

and must be met in all years. However, in real operations this is not the case. As a

result the reliability estimates obtained above tend to be too simplistic and

conservative and provides only a coarse representation of the actual system reliability.

Therefore, we urge caution in using these results to read policy implications. To fully

appreciate the actual operations of the water resources in a river basin a decision

support system that incorporates variable demand schedules, proper topographic

layout for river system reservoir, diversion points, and operating policies must be

used. This will help provide realistic estimates of reliability for the various decision

components of the system.

4.6 Summary and Discussion

A nonparametric stochastic framework for streamflow simulation combining, the long paleo reconstructed streamflow information of lesser reliability with the shorter but reliable observed data was developed. The framework has two components (i) a nonhomogeneous Markov Chain model developed on the paleo data that is then used to simulate the hydrologic state and (ii) a K-nearest neighbor (K-NN) time series bootstrap to simulate the streamflow magnitude from the observed data conditioned on the hydrologic state and the previous flow magnitude. This new and unique framework combines the respective strengths of the two data sets. Furthermore, it is data driven, robust and parsimonious. The framework was applied to paleo reconstructed streamflow and observed data for the Lees Ferry, AZ, streamflow gauge on the Colorado River. The simulations showed the ability to capture all the distributional statistics of the observed period and also generate a rich variety of wet and dry sequences that will greatly benefit the sustainable management of water resources in the basin. Currently the threshold used to determine system state as well as drought and surplus statistics is based on the median of the observed flow. This threshold can be modified or more states could be included as required on a case by case basis.

The annual streamflow generated at Lees Ferry, AZ from this approach can be spatially and temporally disaggregated (Prairie et al., 2006) obtaining monthly flow scenarios at all the gauges in the basin. Such scenarios are used in a basin wide decision model in Chapter 5 and help determine realistic estimations for risk and reliability of various decision components in the water resources system, facilitating effective long-term planning. The developed framework will enable the water

resources planners to use the rich insights available from the paleo reconstructions by

alleviating its short comings; constituting a major contribution from this research.

CHAPTER 5

FRAMEWORK APPLICATION IN THE COLORADO RIVER

SIMULATION SYSTEM DECISION SUPPORT SYSTEM

Understanding the impact of reservoir operation policies to probable future streamflow scenarios is critical for effective sustainable water resources planning and management. This culminates the research components of Chapters 2 through 4 in a water management context. To achieve this, the stochastic streamflow simulation tools are integrated with the Colorado River Simulation System (CRSS), a decision model for the water resources management in the Colorado River Basin (CRB), thus developing a Decision Support System (DSS). Impact of policy alternatives on various components of the CRB water resources systems are evaluated under different streamflow scenarios. The integrated DSS framework, data sets, and operating policies considered are described in detail below, followed by results.

5.1 Introduction

In the previous chapters, stochastic nonparametric methods were developed to generate statistically consistent and richly varied streamflow and salinity scenarios that could also incorporate paleo reconstructed streamflow information. Statistically consistent scenarios display similar distributional properties with the observed record. While, richly varied scenarios include both flow magnitudes and sequences that are plausible by differ from those in the observed record. These stochastic methods were demonstrated by application to four key gauges in the upper Colorado River Basin. This work is extended to the entire Colorado basin of twenty-nine locations (Figure 5-1) on the stream network, required to drive the decision model.

The chapter is organized as follows. The observed streamflow data and three synthetic streamflow scenarios along with their corresponding salinity scenarios are first described followed by an overview of the decision model, CRSS. Then the normal reservoir operating policy focusing on Lakes Powell and Mead and an alternate operating policy that was developed from recent work completed by Jerla (2005) are described. The impact of these policies on the risk of shortages to water delivery obligations is investigated under three streamflow scenarios.
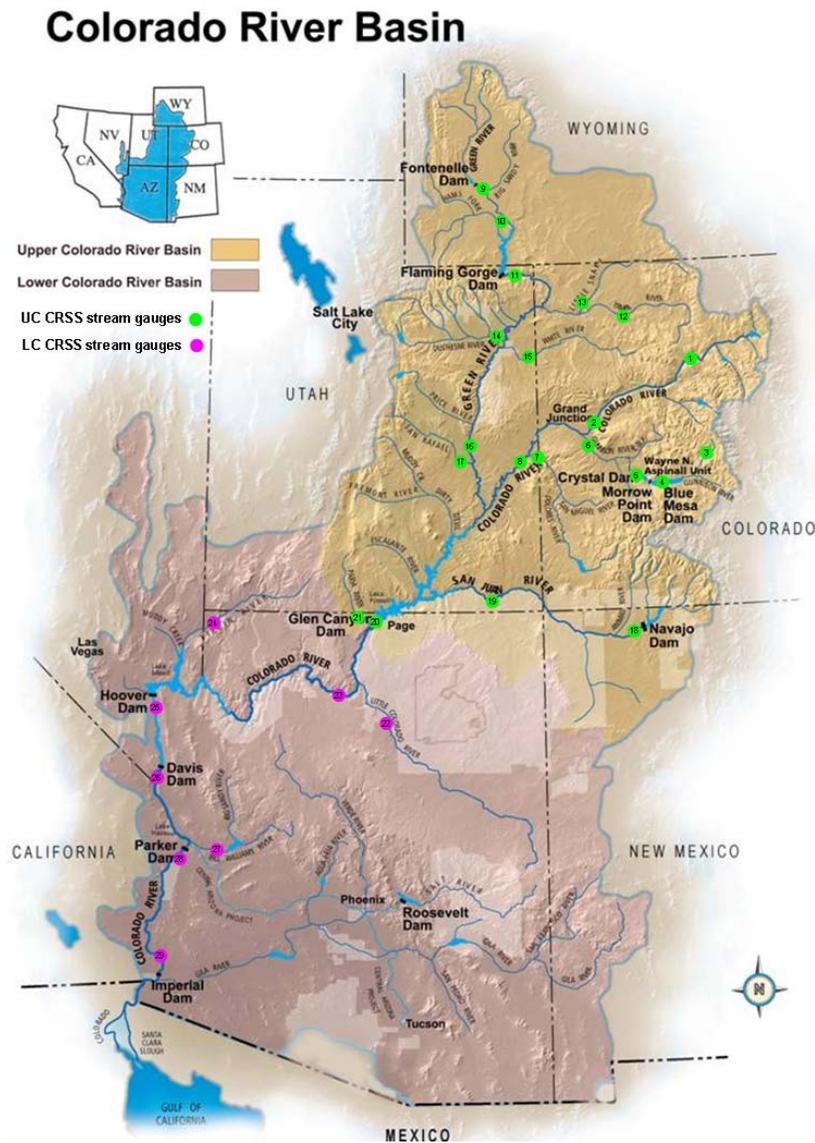
Figure 5-1 Colorado River Basin DSS input sites.

5.2 Streamflow data development

Observed natural flows through out the Colorado River Basin are available for 29 locations listed in Table 5-1. These include 21 sites in the upper Colorado River Basin and 8 locations in the lower Colorado River Basin. The Upper and Lower Basin are delineated in Figure 5-1.

Table 5-1 Natural flow locations required in CRSS.

Site No.	Gauge Number	Gauge Name
1	09211200	Green River below Fontenelle Reservoir, Wyoming
2	09217000	Green River near Green River, Wyoming
3	09234500	Green River near Greendale, Utah
4	09251000	Yampa River near Maybell, Colorado
5	09260000	Little Snake River near Lily, Colorado
6	09302000	Duchesne River near Randlett, Utah
7	09306500	White River near Watson, Utah
8	09315000	Green River at Green River, Utah
9	09328500	San Rafael River near Green River, Utah
10	09072500	Colorado River near Glenwood Springs, Colorado
11	09095500	Colorado River near Cameo, Colorado
12	09109000	Taylor River below Taylor Park Reservoir, Colorado
13	09124700	Gunnison River below Blue Mesa Reservoir, Colorado
14	09127800	Gunnison River at Crystal Reservoir
15	09152500	Gunnison River near Grand Junction, Colorado
16	09180000	Dolores River near Cisco, Utah
17	09180500	Colorado River near Cisco, Utah
18	09355500	San Juan River near Archuleta, New Mexico
19	09379500	San Juan River near Bluff, Utah
20	09380000	Colorado River at Lees Ferry, Arizona
21	09382000	Paria River At Lees Ferry, Arizona
22	09402000	Little Colorado River Near Cameron, Arizona
23	09402500	Colorado River Near Grand Canyon, Arizona
24	09415000	Virgin R At Littlefield, Arizona
25	09421500	Colorado River Below Hoover Dam, AZ-NV
26	09423000	Colorado River Below Davis Dam, AZ-NV
27	09426000	Bill Williams River Below Alamo Dam, Arizona
28	09427520	Colorado River Below Parker Dam, AZ-CA
29	09429490	Colorado River Above Imperial Dam, AZ-CA

Monthly natural streamflows at these locations are available for the 99 year period spanning 1906 – 2004. Naturalized streamflows are computed by removing anthropogenic impacts (i.e., reservoir regulation, consumptive water use, exports, etc.) from the recorded historic flows[4].

Streamflow scenarios from three stochastic methods were developed for use as input to CRSS. A simulation horizon of 53 years was selected projecting operation out to 2060. The first method used the index sequential method (ISM) (Ouarda et al., 1997). This method entails a sequential block bootstrap of the observed data, where the block size is determined by the simulation horizon. The ISM cycles through each year in the historic record generating 99 traces assuming that the record wraps around at the end (i.e., 2003, 1906, 1907, etc.). Each trace will only consist of flow magnitudes and sequences that have occurred in the observed record, with the exception of new sequences being generated as a result of the wrap. These drawbacks limit ISM's ability to model the variability of the streamflow and thus, possible future flow magnitudes and sequences. This method is easy to implement, understandable, and widely used by Reclamation on the Colorado River.

To address these drawbacks, a stochastic nonparametric space-time disaggregation method was developed in Chapter 2 and applied to four key gauges in Upper Colorado River. This was extended to the entire basin, the schematic of which is shown in Figure 5-2. The annual flows at Lees Ferry, AZ (site 20) are generated with the modified K-NN technique (Prairie et al. 2006a), which are then

---

[4] The natural flow data and additional reports describing these data are available at http://www.usbr.gov/lc/region/g4000/NaturalFlow/index.html.

disaggregated to intervening monthly flows through out the Upper Basin stream gauge sites 1-20 excluding site 3.

An issue at site 3 arises because the flow magnitudes at this site are much smaller than the remaining Upper Basin sites. When site 3 was initially included in the complete Upper Basin disaggregation the standard deviation was inflated. To address this issue site 4 is first modeled as the sum of flows at both sites 3 and 4. A second disaggregation is applied to the flows at site 4 to determine flows at site 3 subject to constraints: (i) if the simulated flow at site 4 is positive than it is disaggregated between sites 3 and 4 using a climatological monthly proportion (average), (ii) else, if site 4 is zero or negative then site 3 is set to zero and site 4 remains zero or the negative value.
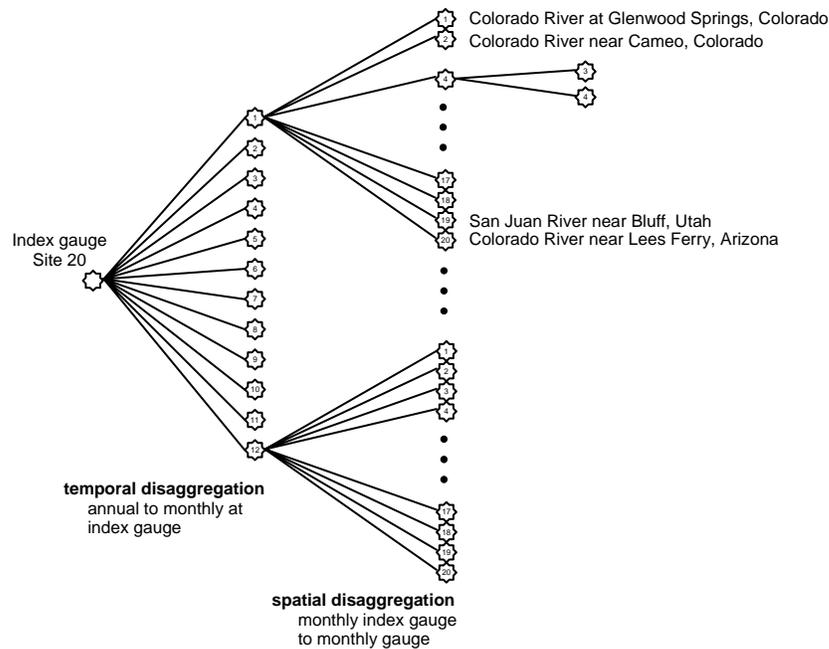
Figure 5-2 Disaggregation scheme.

Flows for the 9 gauges below site 20 are resampled from the observed natural flows based on the year that was chosen during step 3 (K-NN resampling) of the

temporal disaggregation algorithm, presented in Section 2.3. For example, if year 1954 was the year chosen in the temporal disaggregation during step 3 then the associated monthly flows for each of the 9 lower sites are resampled from 1954 observed natural flows. This method ensures the lower sites are both temporally and spatially correlated with each other and the upper sites. The lower sites 21-29 contribute significantly less flow (8% of the total calendar year flow) than the upper sites; therefore, resampling the direct observed natural flows does not adversely affect our ability to model unique and probable flows in the basin as a whole.

The third method was the combination approach developed in Chapter 4, combining the strengths of the paleo reconstructed streamflows and the observed data at the index gauge (site 20), where both the paleo reconstructions and observed data are available. Using this method and the disaggregation approach described above basin wide flow scenarios are generated based on the paleo reconstructed flows.

The three stochastic hydrologic scenarios are identified as (1) ISM, (2) K-NN no conditioning (KNN-NC), and (3) K-NN paleo conditioned (KNN-PC)

5.3 Salinity data development

Natural salinity data is available for 24 gauge locations listed in Table 5-2. The remaining 5 locations that have an associated natural flow but do not have salinity data are set to zero in CRSS. Natural salt is computed by removing anthropogenic influences (upstream reservoir regulation, salt loading from agriculture return flows, and salt removed with exports) effecting salt from observed historic data.

Table 5-2 Historic salt data available at selected sites.

Site No.	Gauge Number	Gauge Name
2	09217000	Green River near Green River, Wyoming
3	09234500	Green River near Greendale, Utah
4	09251000	Yampa River near Maybell, Colorado
6	09302000	Duchesne River near Randlett, Utah
7	09306500	White River near Watson, Utah
8	09315000	Green River at Green River, Utah
9	09328500	San Rafael River near Green River, Utah
10	09072500	Colorado River near Glenwood Springs, Colorado
11	09095500	Colorado River near Cameo, Colorado
15	09152500	Gunnison River near Grand Junction, Colorado
16	09180000	Dolores River near Cisco, Utah
17	09180500	Colorado River near Cisco, Utah
18	09355500	San Juan River near Archuleta, New Mexico
19	09379500	San Juan River near Bluff, Utah
20	09380000	Colorado River at Lees Ferry, Arizona
21	09382000	Paria River At Lees Ferry, Arizona
22	09402000	Little Colorado River Near Cameron, Arizona
23	09402500	Colorado River Near Grand Canyon, Arizona
24	09415000	Virgin R At Littlefield, Arizona
25	09421500	Colorado River Below Hoover Dam, AZ-NV
26	09423000	Colorado River Below Davis Dam, AZ-NV
28	09427520	Colorado River Below Parker Dam, AZ-CA
29	09429490	Colorado River Above Imperial Dam, AZ-CA

Salinity for the ISM hydrology was generated from monthly local polynomial regressions at each site, the current method used by Reclamation. The remaining two K-NN based hydrologies were develop from the methods presented in Chapter 4.

Approach 2 (shown in Figure 3-1b), that uses an annual local polynomial regressions of salt based on flow at each of the 15 sites above and including site 20, where natural salt data are available, was extended to the entire Upper Basin. Natural salt data is available from 1971-1995 throughout the basin and thus, this is the timeframe used to develop the regressions. For gauges below site 20 the historic year used for the lower sites streamflow as a result of resampling is also used for the salt

data. This ensures the consistency between flow and salinity generated in the lower sites.

5.4 Overview of the Colorado River decision support system

The Colorado River is managed by Reclamation in consultations with basin stakeholders. To help Reclamation and stakeholders evaluate and assess various operational policies under probable future hydrologies, the long-term planning model, CRSS is crucial. The model resides in the generalized river basin modeling software RiverWare (Zagona et al. 2001). This software includes a rule language that allows one to define reservoir operation policies and apply these polices in "what if" scenarios to better understand their impacts in a complex reservoir network. Such rules have been developed to represent the "Laws of the River", which govern the operation of reservoirs throughout the Colorado River Basin. Detailed descriptions of the design and implementation of these rules are described in Jerla (2005). Here attention is focused on the two largest reservoirs in the Colorado River system, Lakes Powell and Mead. In particular, the study focuses on these reservoirs in relation to development of operational criteria that govern their operation during low reservoir elevations.

5.4.1 General policy overview

Two alternate policy scenarios were analyzed for the three hydrology and salinity scenarios making a total of six combinations. The two policy scenarios include present reservoir operation for Lakes Powell and Mead, termed normal operation and an operation introduced in Jerla (2005), termed Balance Contents C2 operation, which is similar to the "reverse operation" first presented in the Severe and Sustained Drought (SSD) study performed by Harding, et al. (1995).

Normal operations at Lakes Powell and Mead consist of a series of prioritized objectives, which reservoir operations attempt to meet. At Powell the annual release is determined as follows.

- Make a minimum object release of 8.23 MAF over the water year.

- Release additional water when;

    o the end-of-water-year (EOWY) forecasted storage at Powell is greater than at Mead AND

    o the storage in the Upper Basin reservoirs is greater than 602(a) storage[5], in order to balance or "equalize" the storages by the EOWY.

- Release additional water if the reservoir is nearly full to avoid future spills.

At Mead the annual release is determined as follows.

- Set releases to meet downstream demands.

- Release additional water when prescribed by the flood control procedures.

C2 operation differs from normal operation in that the objective in C2 operation is to maintain Powell and Mead at equivalent storages, when possible throughout a broad range of Powell's operation. No power protection is provided for Powell. During C2 operations equalization releases are made from Powell when the forecasted EOWY storage at Powell is higher than the forecasted EOWY storage at Mead. However, when the forecasted EOWY storage at Powell results in an elevation below minimum power pool (3490 ft msl) the monthly releases revert to meeting the

---

[5] The 602(a) storage quantity is the storage in the Upper Basin necessary to assure Lower Basin delivery obligations can be met without impairing consumptive use in the Upper Basin. The Interim 602(a) Storage Guideline, in effect through 2016, established that Lake Powell's elevation must be above 3,630 feet msl (which corresponds to storage of approximately 14.85 maf) for equalization releases to occur.

minimum objective annual release of 8.23 MAF. Operations at Powell and Mead are depicted in Table 5-3.

Table 5-3 Operational Diagram for Lake Powell and Mead.

Lake Powell Elevation (ft)	Normal	C2 (balancing)	Lake Powell Storage (MAF)
3700	Equalize or Release 8.23 MAF	Equalize or Release 8.23 MAF	24.3
602(a)			602(a)
	Release 8.23 MAF[1]	Balance contents with a min/max release constraint from the Glen Canyon ROD	
3490			4.0
		Release 8.23 MAF[1]	
3370			0

[1]Below 3440 ft Powell outlet works cannot physically release 8.23 MAF as a result of reduced head behind the dam.

Lake Mead Elevation (ft)	Normal	C2 (balancing)	Lake Mead Storage (MAF)
1220	Flood Control/70R Surplus	Flood Control/70R Surplus	25.9
1200			22.9
	Full Domestic Surplus (through 2016)	Full Domestic Surplus (through 2016)	
1145			15.9
	Partial Domestic Surplus (through 2016)	Partial Domestic Surplus (through 2016)	
1125			13.9
	Normal Operations	Normal Operations	
	Shortage 80% Protection of 1050 (80P1050)	Shortage 80% Protection of 1050 (80P1050)	
1000			4.3
	Shortage Absolute Protection of 1000 (AbsPro 1000)	Shortage Absolute Protection of 1000 (AbsPro 1000)	
895			0

Detailed descriptions of both these policies along with explanations of all variables in these two policies are available in Jerla (2005).

5.4.2 Model setup

The model setup in CRSS is quite similar to that performed in Jerla (2005) though this study is performed at a monthly time step in CRSS rather than in the CRSSLite

model, developed in Jerla (2005). CRSSLite is a simplified annual time step model designed to screen multiple operation policies rapidly before running these policies in CRSS, which requires more run time and additional input data.

The CRSS model was updated to reflect current conditions in the Colorado River basin. Initial conditions for reservoir elevation (see Table 5-4) were set to the projected end of December 2007 elevations based on Reclamation's mid-term operations model (the 24 month study) used to project reservoir operations throughout the current and next operating years.

Table 5-4 Reservoir initial conditions from August 24 month study.

Reservoir name	December 2007 project pool elevation (*ft msl*)
Fontenelle	6486.29
Flaming Gorge	6029.67
Taylor Park	9308.32
Blue Mesa	7489.99
Morrow Point	7153.73
Crystal	6753.04
Starvation[6]	255000 (acre-ft storage)
Navajo	6080.33
Powell	3614.80
Mead	1116.53
Mohave	638.71
Havasu	445.80

Additionally, the Upper Basin depletions are consistent with the Surplus Interim Agreement Environmental Impact Statement (DOI, 2002) and the Lower Basin is consistent with the Water Delivery Agreement (DOI, 2003a).

One reservoir operational change recently incorporated in the official CRSS is also included in this research. At water surface elevations below 3440 feet Powell's

---

[6] Represents a conglomeration of 8 smaller reservoirs (DOI, 1987).

outlet works cannot physically release the minimum objective release of 8.23 MAF as a result of reduced head above the outlet works. The relation between pool elevation and the maximum possible release from Powell (Figure 5-3) between 3370 feet (top of dead pool) and 3490 feet (minimum power pool) has been included in the operational rules for these model runs. Similar restrictions do not exist at Mead. (C. S. Jerla, personal communication, November, 29, 2006).
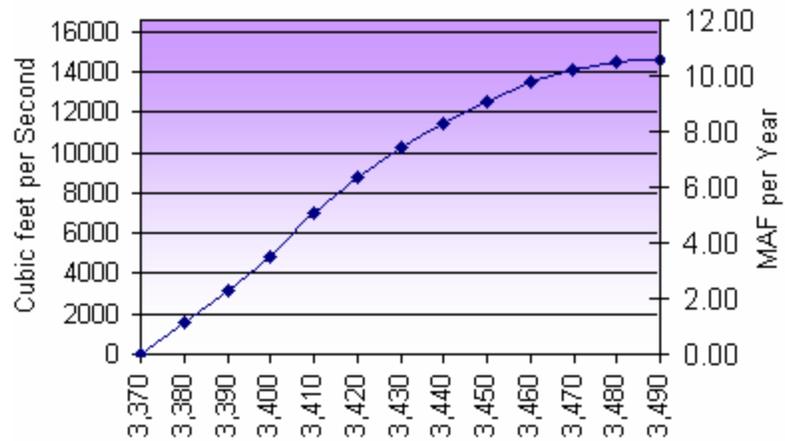
Figure 5-3 Powell elevation and maximum possible release.

Reservoir initial salinity concentrations were set based on the latest historic values available. These are the December 2005 values reported by the USGS with the exception of Davis and Parker Dam, which were assumed to be equivalent to Mead concentration since a December 2005 value is not available. Table 5-5 lists the initial concentrations assumed for each reservoir that model salinity in CRSS.

Table 5-5 Initial reservoir salt concentration.

Reservoir name	December 2005 historic salt concentration (*mg/L*)
Flaming Gorge	438.10
Starvation	716.00
Navajo	185.30
Powell	447.60
Mead	644.10
Mohave	644.10
Havasu	644.10

Additional salinity input data includes projected water quality improvement project (WQIP) control levels and agricultural salinity loading. Both these inputs are the same as those used in the recent 2005 Triennial Review of Water Quality Standards for Salinity Colorado River System (Colorado River Salinity Control Forum, 2005).

5.4.3 Lower Basin Shortage policy

Since much of the comparative analysis focuses on the risk of shortages to the Lower Division States (AZ, NV, CA) explanation of the assumed operation under shortage is warranted. The Secretary of the Interior as water master in the Lower Basin, determines the water supply available to the Lower Division States each year, through the development of the Annual Operating Plan (AOP). This determination is either normal, surplus, or shortage and results in either 7.5 MAF, greater than 7.5 MAF, or less than 7.5 MAF of water being made available for consumptive use in those states, respectively. Currently, there are no specific guidelines in place to assist the Secretary in determining when to declare a shortage or by how much deliveries would be reduced. To date no shortage has been declared. Reclamation recently initiated efforts to adopt shortage and additional operational guidelines for Lakes

Powell and Mead though a National Environmental Policy Act (NEPA) process (Fulp, 2005). Absent specific guidelines, Reclamation has used a set of assumptions to allow the model to determine when the river system is in shortage, who will be shorted, and by how much. The assumptions used here are the same under both policy alternatives considered in this research and are consistent with those used by Reclamation in recent NEPA compliance documents (DOI, 2000). These assumptions are based on a two tier or level approach. At the first level, elevation 1050 ft (the current minimum elevation for power generation at the Hoover powerplant) is protected approximately 80% of the time. At the second level, elevation 1000 ft (the minimum elevation for operation of Southern Nevada Water Authority's (SNWA) lower intake structure) is protected 100% of the time.

Level 1 shortage is declared based a series of trigger elevations (DOI, 2002) at Lake Mead referred to as 80P1050 protection elevations. These protection (trigger) elevations are presented in Figure 5-4.
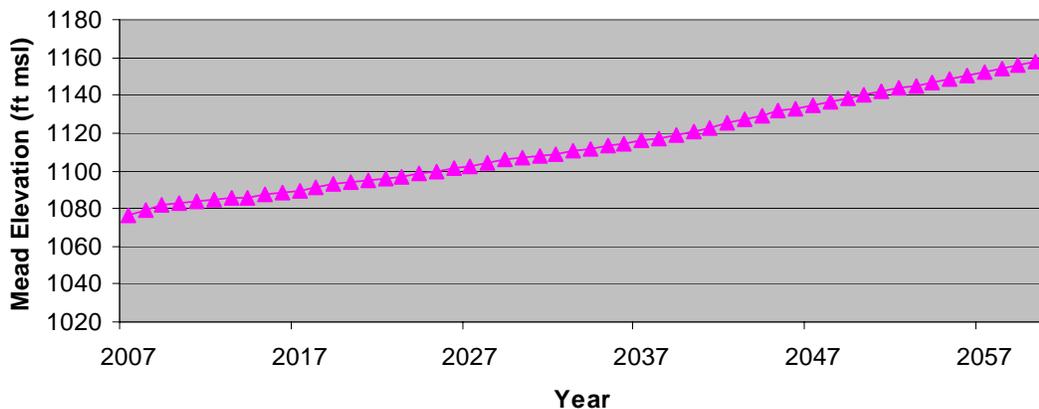
Figure 5-4 Mead 80P1050 protection elevations.

The trigger elevations increase because they are dependent upon increasing Upper Basin demands. With increased demands in the Upper Basin, Mead would expect less water from Powell equalization releases and would need to impose

shortages at higher levels in order to assure being above 1050 feet approximately 80% of the time.

If Mead's elevation at the beginning of the year is less than the trigger elevation a level 1 shortage is declared and shortages are administered to both Central Arizona Project (CAP) and SNWA, both junior rights to California's diversions. During level 1 shortage, CAP total use is set to 1.0 MAF (a reduction of about 300-400 KAF, depending upon the scheduled use) and SNWP is reduced 4% of the total shortage amount. No other Lower Basin diversions receive a shortage during a first level shortage.

A level 2 shortage is declared if the delivery reductions imposed under level 1 are not sufficient to protect Mead's elevation from falling below 1000 feet. A level 2 shortage is necessary when in January of the current year, the forecasted end-of-calendar-year (EOCY) storage in Mead, with level 1 shortages in place, would result in an elevation below 1000 feet.

When this occurs, CAP and SNWP depletions are further reduced to keep Mead at 1000 ft. If CAP depletions are reduced to zero and further reductions are required to keep Mead at 1000 feet, additional shortages are imposed on California and Mexico equally as required to maintain Mead at 1000 feet.

All other Lower Basin demands are provided with available storage from Mead even if Mead is below 1000 feet except for SNWA, which can no longer divert water because elevations less than 1000 feet are below the lower intake structure.

5.5 Model Results and Analysis

This section begins with a description of the decision variables that are compared followed by a description of the statistics used in the comparison. Results from the six

scenarios based on the three alternate future hydrologies and the two operating policies (ISM normal, ISM C2, KNN-NC normal, KNN-NC C2, KNN-PC normal, and KNN-PC C2) are compared.

5.5.1 <u>Presentation of results</u>

The modeling results are analyzed with a suite of statistics that include; percentiles, non-exceedance, cumulative density, and minimum values. These statistics are used to explore the impact of each scenario on reservoir elevation and release, probability of shortage and shortage volumes, and salinity concentration.

5.5.1.1 Percentile Statistics

Percentile statistics are used to understand each scenarios impact on reservoir elevation. The 90th, 50th, and 10th percentiles are plotted and generally allow an understanding of a reservoirs operation under high, normal, and low conditions, respectively. The percentile statistic is computed for each year of the simulation horizon displaying the statistics variation over the simulation horizon. This statistic computes the value for which a specified percentage of values are equal to or below. For example, the 50th percentile values indicate that 50% of the values are either equal to or below the 50th percentile value. To compute the percentile value in each year the 99 traces are ranked, and an index is assigned to each ranked value. If the user defined percentile does not fall on an indexed value than a weighted average is used to determine the user defined percentile value. Percentile values are independent across the simulation horizon and should not be interpreted as single trace operations over the simulation horizon.

5.5.1.2 Probability of non-exceedance

The probability of non-exceedance is computed by counting the number of times the values across traces did not exceed a value of interest and dividing this number by the total number of traces. This can be computed separately for each year in the simulation horizon or a user specified time period.

5.5.1.3 Cumulative density functions

The cumulative probability shows the probability of being at or below a given value. The probability is cumulative because it is the summation of the individual probabilities of all the values at or below a given probability. Mathematically the cumulative density function (CDF) is computed by integration of the probability density function, which provides the probability for each individual value over a given sample. The CDF included in these results is computed empirically. For a variable of interest all the values across traces and over the simulation horizon are ordered and the number of times a value is equal to or below a given value divided by the total number of values determines the cumulative probability. The computation is performed for each value in the set and the ordered probabilities are plotted to generate the empirical cumulative probability function (ECDF).

5.5.2 Stochastic Natural Flows

The ECDF of total natural flow for Colorado River at Lees Ferry from the three hydrologies is shown in Figure 5-5. The ISM hydrology shows the smallest range of flows while the K-NN paleo conditioned hydrology shows the largest range of flows. This supports previous statements that ISM hydrology exhibits reduced variability. The natural flows generated by the K-NN no conditioning hydrology generally show

an increased cumulative probability for a given flow evidenced by the ECDF for this hydrology being shift up in relation to either of the other two hydrology scenarios for a given flow. Of the other two hydrologies (ISM and K-NN paleo conditioned) the ISM hydrology generally displays the lowest cumulative probability for a given flow.
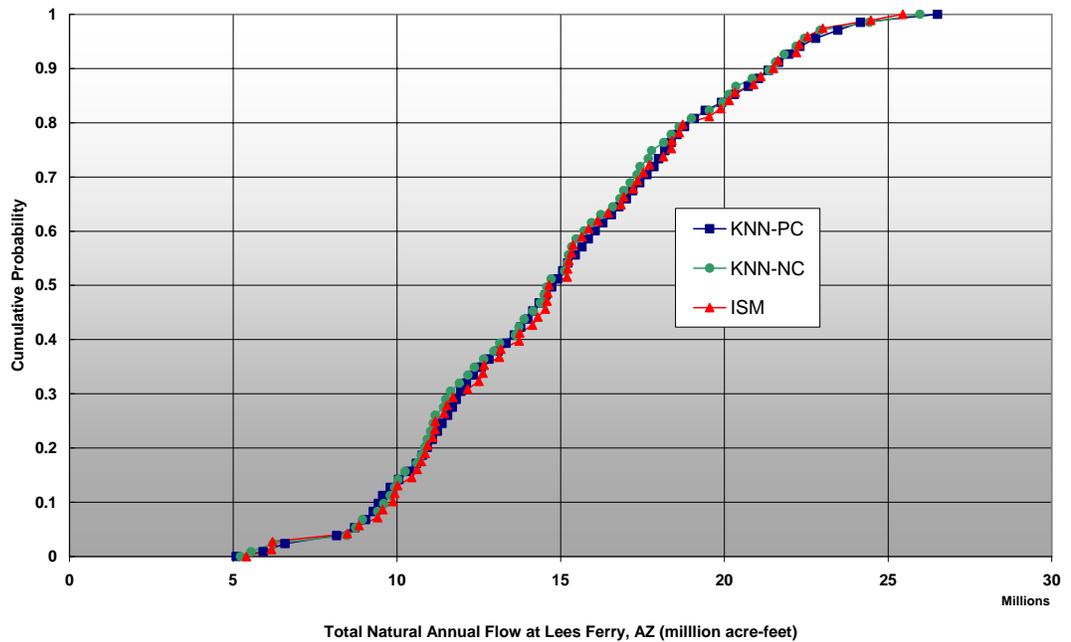
Figure 5-5 ECDF of total natural flow for Colorado River at Lees Ferry, AZ.

Table 5-6 presents basic statistics for the observed record and the three future hydrologies. The basic statistics include the mean, standard deviation, coefficient of skewness, lag 1 autocorrelation, maximum, and minimum values. The average value for each statistic across all simulations is provided except for the maximum and minimum, for which the actual maximum or minimum are provided. Though a single value for each statistic is provided in Table 5-6 the three alternate future hydrologies generate a range of values for each statistic which are not shown in the table. The observed basic statistics is always captured within the range of values from each

future hydrology. The ISM hydrology generates the same mean, maximum, and minimum as the observed data since it only generates flow magnitudes and sequences that are represented in the observed data but has a standard deviation slightly less than the observed resulting of sampling error. The sampling error arises from only resampling a 53 year long sequence, which is a subset of the 99 observed years. The skew and autocorrelation are also unrepresented. The K-NN no conditioning has a slightly reduced mean, standard deviation, and autocorrelation. The coefficient of skewness is slightly increased. Similar results are seen for the K-NN paleo conditioned though the coefficient of skewness and autocorrelation are slightly less than the observed. The K-NN hydrologies are able to generate higher and lower flows than observed. Though the K-NN paleo conditioned hydrology resamples magnitudes from the observed record this is done on a water year basis, thereby allowing magnitudes never seen before on the calendar year basis to be generated. Both K-NN hydrologies can generate calendar year annual flow magnitudes and flow sequences not seen in the observed record.

Table 5-6 Basic statistics for observed and three future hydrologies.

Hydrology	Mean (*acre-ft/yr*)	Standard deviation (*acre-ft/yr*)	Skew	r(1)	Maximum (*acre-ft/yr*)	Minimum (*acre-ft/yr*)
observed	15,024,444	4,443,956	0.18	0.27	25,432,344	5,398,985
ISM	15,024,444	4,404,577	0.15	0.22	25,432,344	5,398,985
KNN-NC	14,844,457	4,357,936	0.19	0.21	25,975,260	5,219,712
KNN-PC	14,988,921	4,399,353	0.16	0.26	26,507,364	5,094,739

5.5.3 Reservoirs Powell and Mead elevation

To understand changes in Powell's storage as a result of each of the six scenarios we first discuss percentile plots for Powell. Figures 5-6, 5-7, and 5-8 respectively show the 90th, 50th, and 10th percentile plots. Each plot compares the six scenarios. At

the 90th percentile the six scenarios lie over each other after 2016 indicating that the two policy alternatives do not operate Powell differently at the higher elevations. At these higher elevations Powell is either in equalization or in spill avoidance operations and these operations are the same across the two policy alternatives. Before 2016 both policy alternatives for the K-NN no conditioning hydrology display lower 90th percentile elevations at Powell than the other scenarios. This is a result of relatively higher flows produced by this hydrology scenario producing a higher probability for spill avoidance and/or equalization releases at Powell.
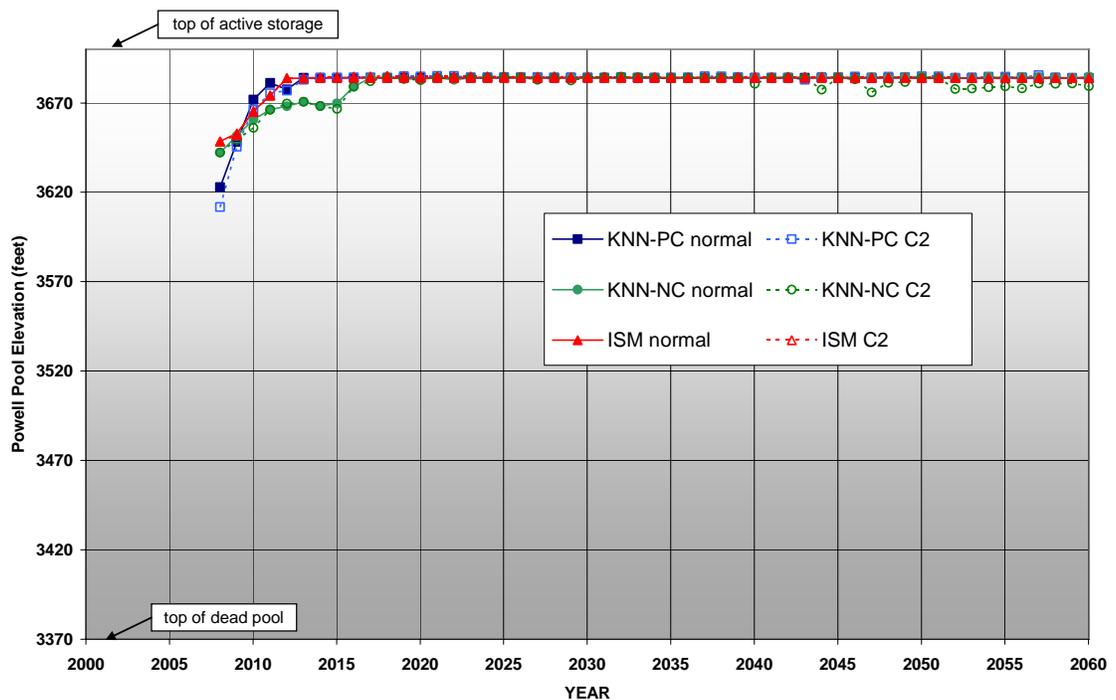
Figure 5-6 Powell EOCY 90th percentile elevation.

At the 50th percentile the scenarios are first split by the different reservoir operation policies. The normal operations tend to hold more water in Powell while the C2 operations tend to result in lower average storages in Powell. This results from the C2 policy pushing more water to Mead during average contents to balance the two reservoirs. The ISM hydrologies tend to produce lower elevation in Powell for both

alternative policies in general over the entire run. Also, it is evident that the percentile values tend to fluctuate much less over time than with either of the K-NN hydrologies. This is a factor of ISM using the same set of hydrologies for every year in the simulation horizon, thereby reducing the random nature of the synthetic hydrologies. This is a drawback of ISM, which is limited to only producing flows that were seen before and using the same set of flows when computing statistics over the traces. The K-NN hydrologies tend to dip above or below each other for each reservoir operating policy, a function of their improved random nature across traces. It is not evident that either K-NN hydrology performs different in the long run. During the initial years 2008-2020 the paleo conditioned hydrology result in lower initial Powell elevations but are higher after 2012. The reverse is seen in the non conditioned hydrology. This is a result of the paleo conditioned traces initially tending to show a propensity for drought but also have a higher probability of surplus, appearing after 2012, than the non conditioned hydrology.
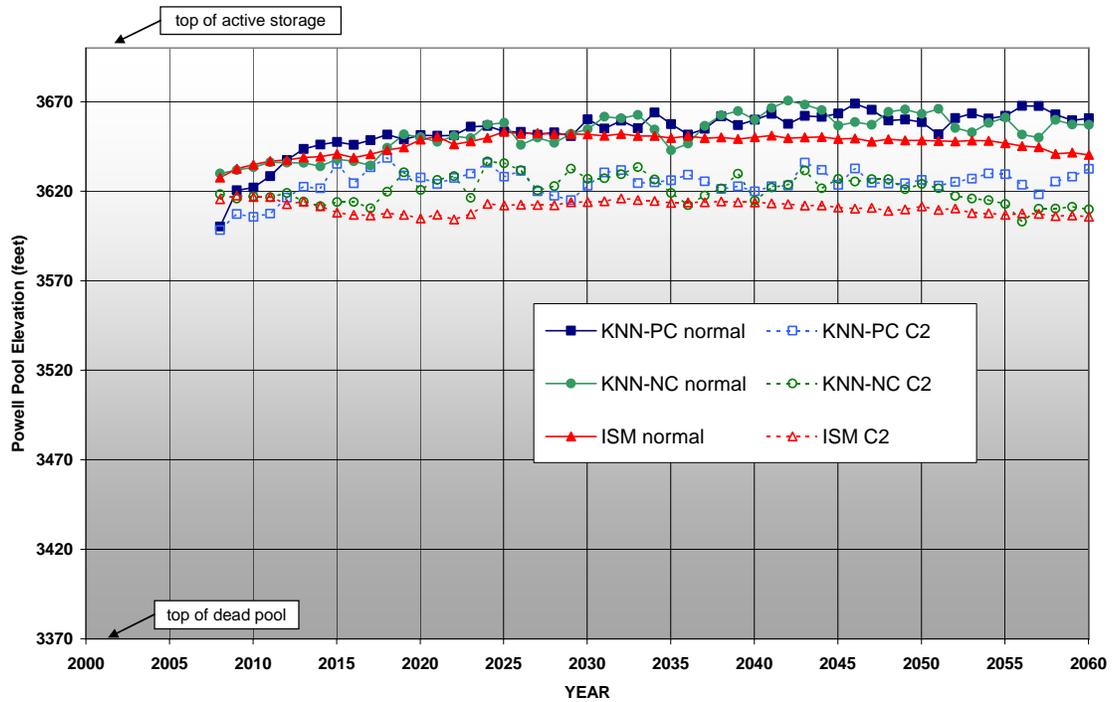
Figure 5-7 50th percentile of Powell's EOCY elevation.

At the 10th percentile marked differences between both the operating policies and hydrologies appear. Overall the C2 policy tends to produce lower pool elevations at these lower operating elevations. The ISM hydrology again shows little variation over time. In this plot the elevation of 3490 feet is important to the C2 operating policy as it triggers balancing operations to stop and Powell resumes releasing 8.23 MAF. The ISM hydrology barely reaches this point while the K-NN paleo conditioned hydrology demonstrates large dips in elevation (2027 and 2052) indicating that Powell is releasing 8.23 MAF and rapidly losing storage. The K-NN no conditioning hydrology does not display these dips as the inflows to Powell are not as reduced as in the paleo conditioned traces and Powell is able to better maintain storage. But we see that for elevation below 3490 feet there is no longer a distinction between the two reservoir operating policies as they are both releasing 8.23 MAF annually from

96

Powell. Though minimum elevations are not shown Lake Powell never reached dead storage (3370 feet) elevation for any of the six scenarios.
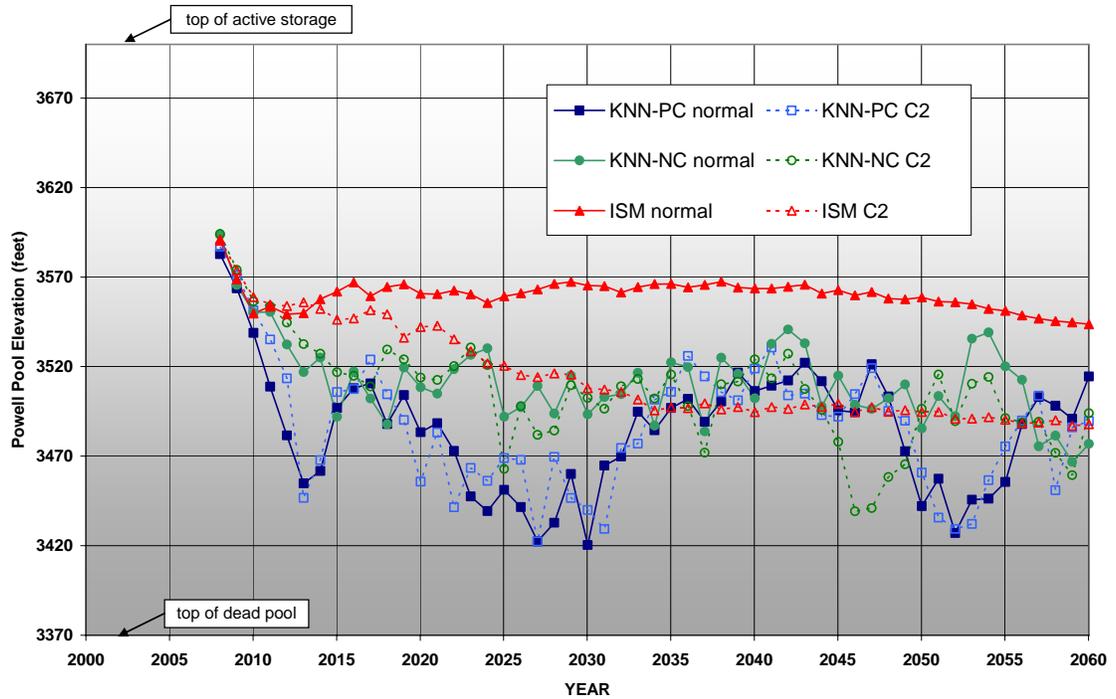
Figure 5-8 Powell EOCY 10th percentile elevation.

Figures 5-9, 5-10, and 5-11 show Mead elevation 90th, 50th, and 10th percentile plots, respectively. As with Powell, Mead does not display much difference at the 90th percentile over the 6 scenarios because at higher elevations flood control or surplus operations are occurring and are nearly the same for both operating policies. There is a tendency for the K-NN paleo conditioned hydrologies to generate higher elevations, a function of the higher surplus probability in the K-NN paleo conditioned hydrologies. The K-NN non conditioned hydrologies tend to produce slightly lower elevations than either the ISM or K-NN paleo conditioned hydrologies.
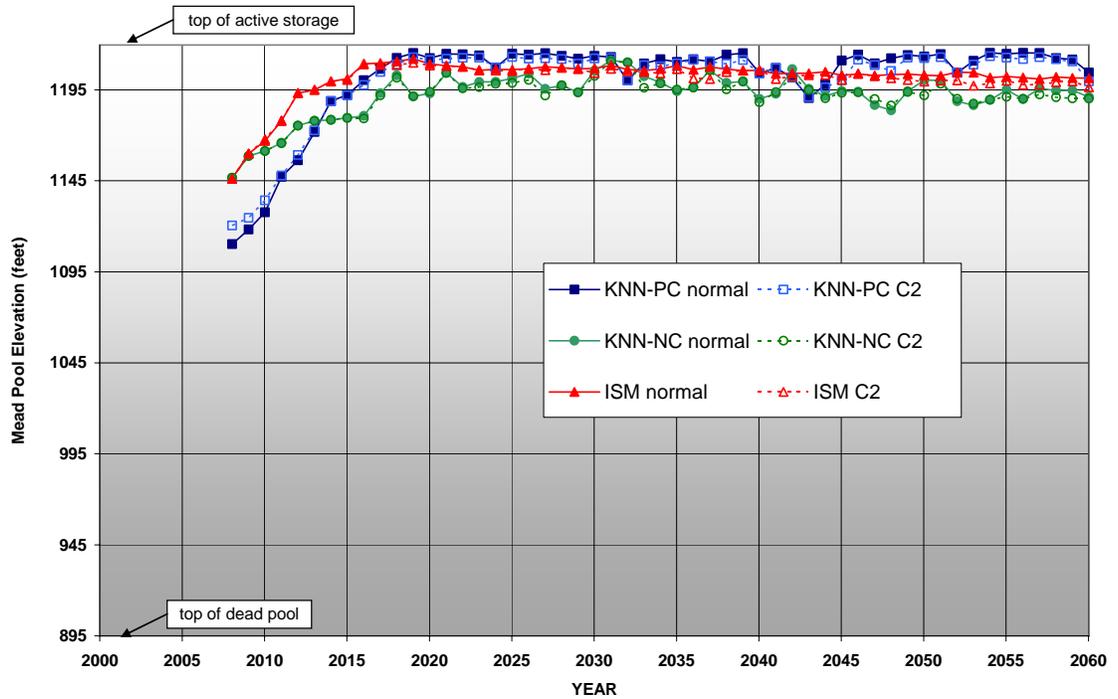
Figure 5-9 Mead EOCY 90th percentile elevation.

At the 50th percentile the C2 operation tends to provide more water in storage as a result of the balancing release from Powell. The two K-NN hydrologies perform similarly at these median storage volumes; though the ISM hydrologies tend to indicate lower storages than the K-NN hydrologies. Most likely this is a function of the reduced variability demonstrated in the ISM hydrologies.
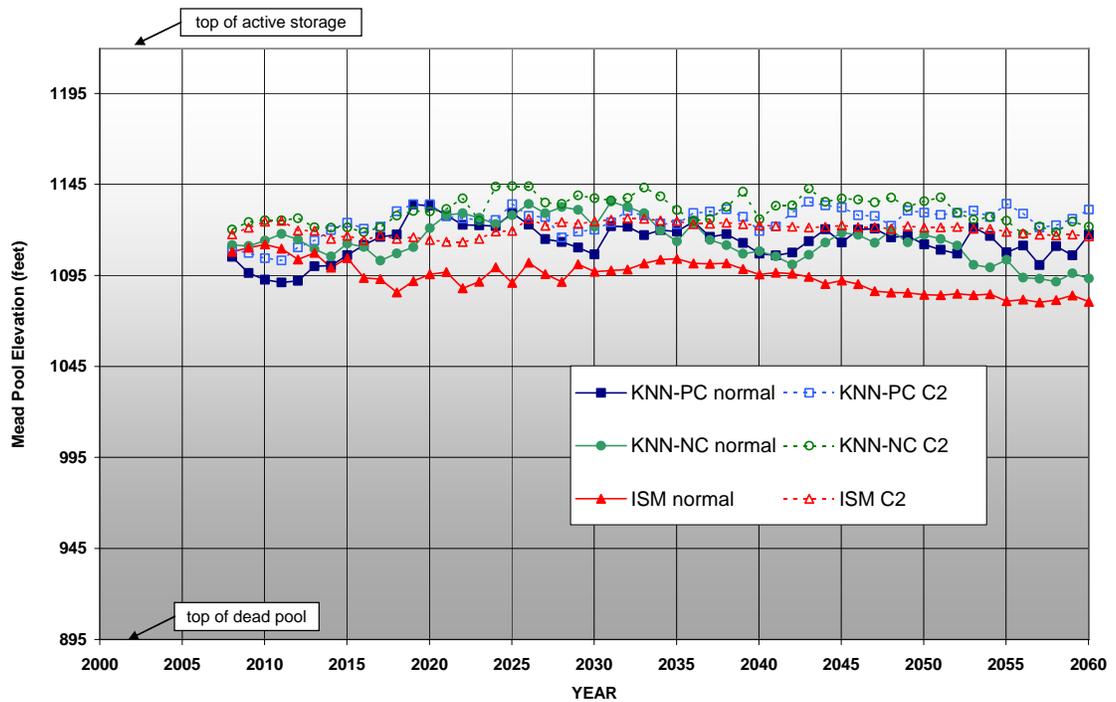
Figure 5-10 Mead EOCY 50th percentile elevation.

At the 10th percentile more water is stored in Mead under C2 operations, again a function of the help from the balancing releases from Powell. The K-NN hydrologies tend to produce higher storage volumes at these lower elevations than the ISM hydrologies. At this percentile we see Mead operations are attempting to protect 1000 feet elevation as 10% of the simulations for all years and hydrologies under both operating scenarios are over 1000 feet. However, given the Lower Basin shortage policy implemented for these scenarios, the maximum shortage that can be applied is 3.8-4.0 MAF, depending upon CAP's schedule. In some cases additional shortage is necessary to keep Mead above 1000 feet. Additional analysis found the maximum shortage was applied approximately 5% of the time in any given year of the simulation horizon. During these times Mead can drop below 1000 feet even though

99

the maximum shortage is applied because additional shortage beyond the current maximum is required to keep Mead above 1000 feet.
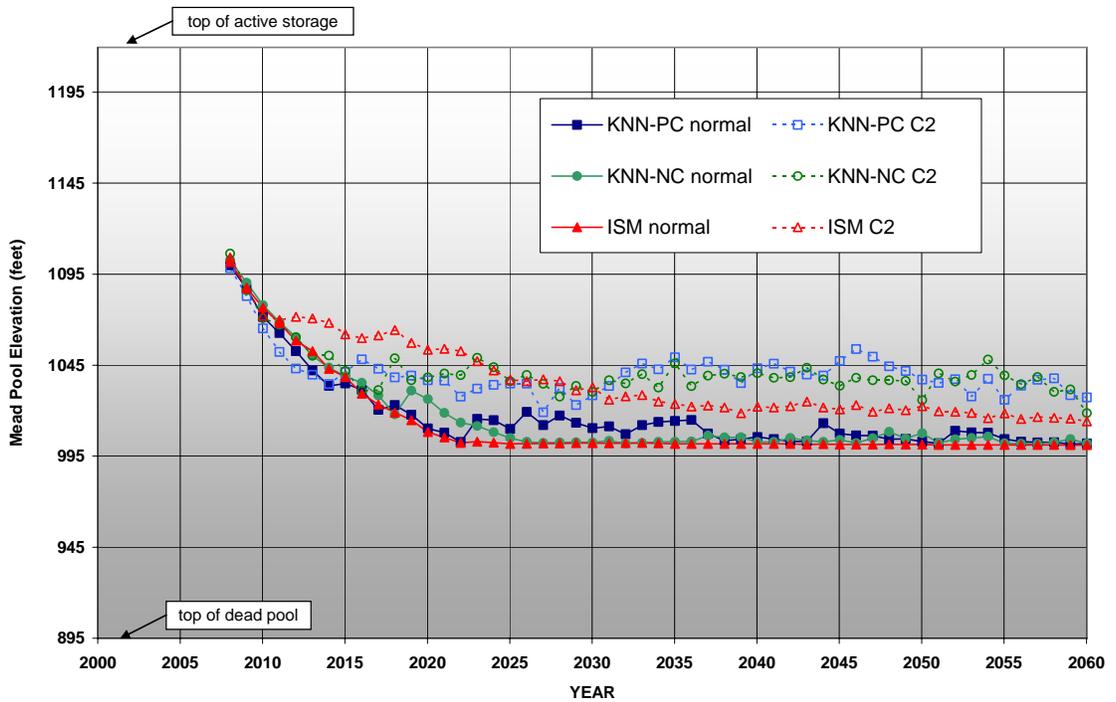
Figure 5-11 Mead EOCY 10th percentile elevation.

Figure 5-12 shows the probability of being below 1000 feet for each year of the simulation horizon. The ISM C2 scenario shows the lowest percentage (0%-1%) of years below 1000 ft. While the K-NN no conditioning C2 scenario has one year at 6% in 2060, but generally this scenario is more often at 3%. The ISM hydrologies tend to show less chance of being below 1000 ft while both K-NN hydrologies show more probability, a result of generating new sequences for plausible future streamflows. Overall C2 operations tend to show lower probability than normal operations. Though minimum elevations are not shown Lake Mead never reached dead storage (895 feet) elevation for any of the six scenarios.
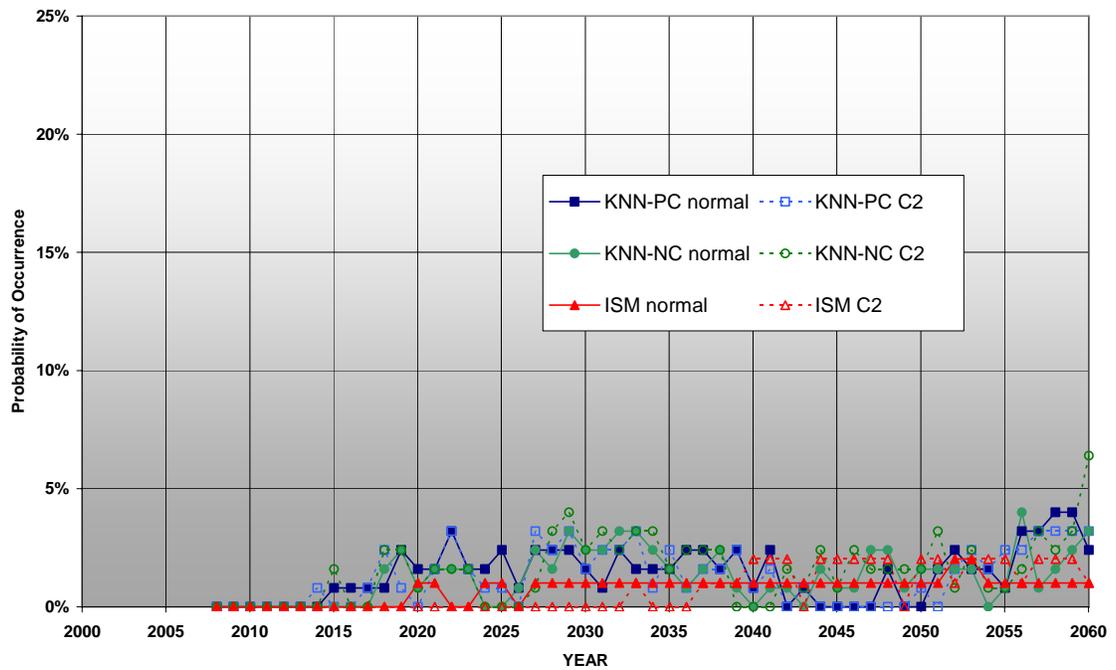
Figure 5-12 Probability of not exceeding 1000 ft elevation at Mead.

5.5.4 Lower Basin shortage statistics

We next present a suite of figures describing shortage probability, shortage volume, and reduced reservoir releases. Figure 5-13 shows the probability of shortage for each year in the simulation horizon. Initially, the probability of shortage is highest for the K-NN paleo conditioned hydrology under normal operations, but as time progresses the ISM hydrologies produce the highest probability of shortage. Overall the normal operations display a higher probability of shortage than the C2 operations. The two K-NN hydrologies do not tend to show a significantly different shortage probability over time.
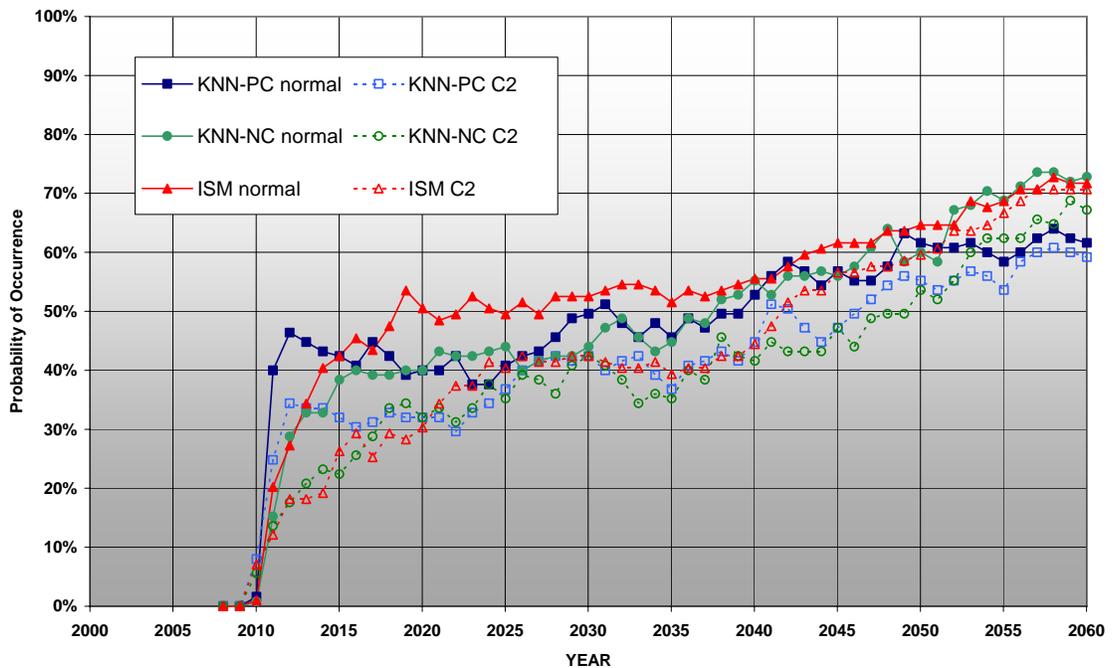
Figure 5-13 Probability of any shortage.

The ECDF of shortage volume is shown in Figure 5-14. There are interesting probabilities which can be learned from this plot. To begin, there is a distinct line at about 3 MAF which is the approximate point where level 1 shortage is not sufficient and level 2 must be imposed. Next, the ISM hydrologies tended to generate the highest probability of shortage in the previous plot but we see here that those shortages are the smallest in magnitude. ISM hydrologies are in level 1 shortage about 52% of the time under normal operation and 43% of the time under C2. Both K-NN hydrologies begin level 1 shortage a few percentage points below ISM for their respective operations. Level 2 shortage begins first (22% of the time) with ISM hydrologies under normal operations but the maximum shortage is only 1.4 MAF. While ISM under C2 imposes level 2 8% of the time but the maximum shortage is 3.9 MAF. Similar results follow for the K-NN hydrologies with the paleo conditioned hydrologies showing the greatest probability of level 2.
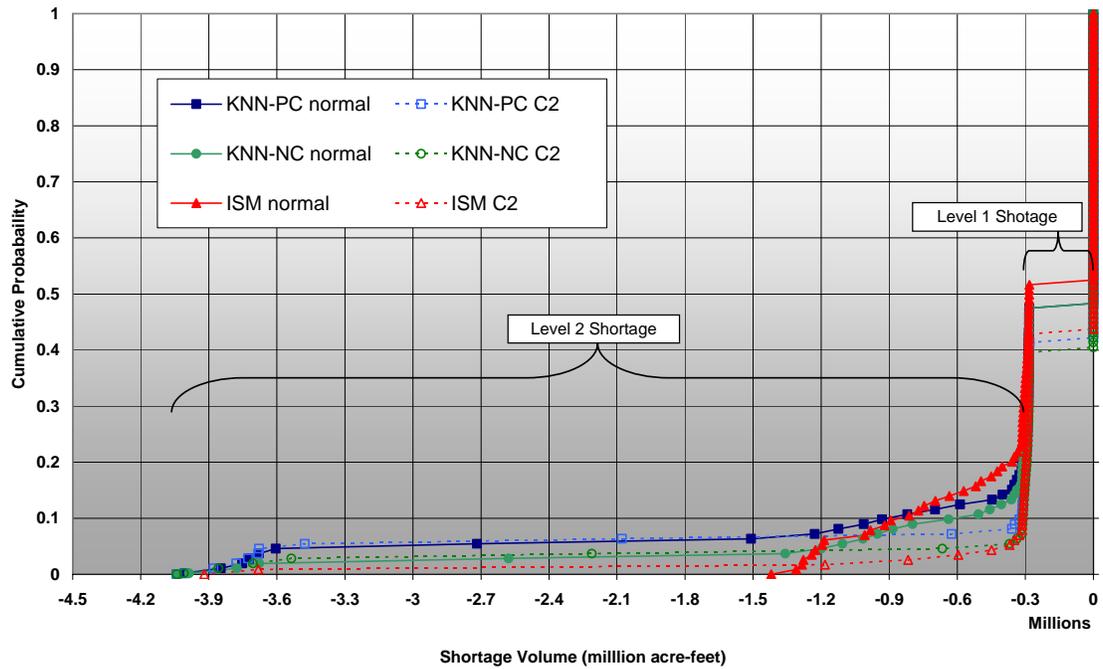
Figure 5-14 ECDF of shortage volume.

5.5.5 Powell's water year release

Powell's water year release is of interest to understand the probability of Powell not being able to meet the minimum objective release of 8.23 MAF. Figure 5-15 shows the probability of not meeting the minimum objective years in each year of the simulation horizon. It is evident that the C2 policy greatly increases the chances Powell will not release the minimum objective release as this requirement is relaxed under the balancing policy for a specified range of elevations. There is not a distinct difference between the alternate hydrologies under the C2 operations. Under normal operation ISM hydrologies always meet the minimum objective release except in 2012 and 2013. Of the K-NN hydrologies under normal operation the paleo conditioned hydrologies have a greater probability of not meeting the minimum

objective release, as result of increased probability for longer droughts when a drought occurs.
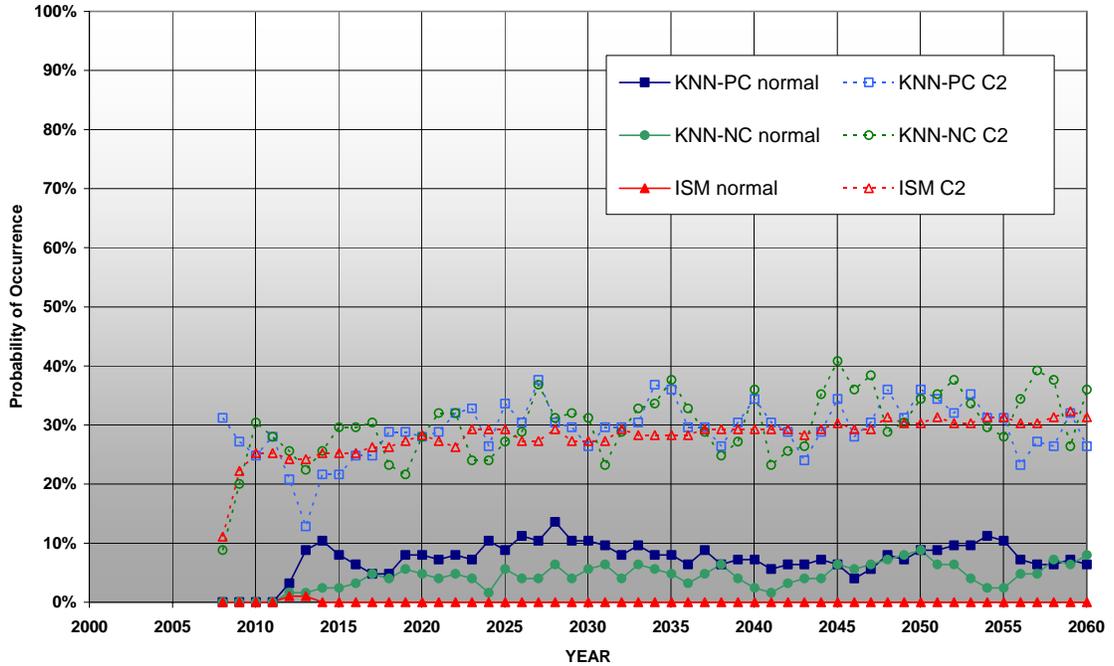
Figure 5-15 Probability of not exceeding an 8.23 MAF WY release from Powell.

The ECDF of Powell water year release is shown in Figure 5-16. The figure displays multiple characteristics of both the alternate hydrologies and operating policies. The straight portion of the ECDF for normal operation indicates that between 62% and 52% of the time Powell is making at least the minimum objective release and that between 40% and 38% of the time the releases are higher than the minimum objective release. At the tails, the K-NN paleo conditioned hydrologies release less than 8.23 MAF 8% of the time and also produce the lowest release but conversely they release more the 8.23 the most at 40% of the time and produce the highest flows, indicative of generating a greater variety of drought and surplus sequences. The C2 policy generates a wide range of releases in a more gradual increasing trend. With the release crossing the 8.23 MAF release at about 30%

probability indicating that 70% of the time the release is above 8.23 MAF while 30% of the time the release is below 8.23 MAF. The full range of possible releases is similar for both policies.
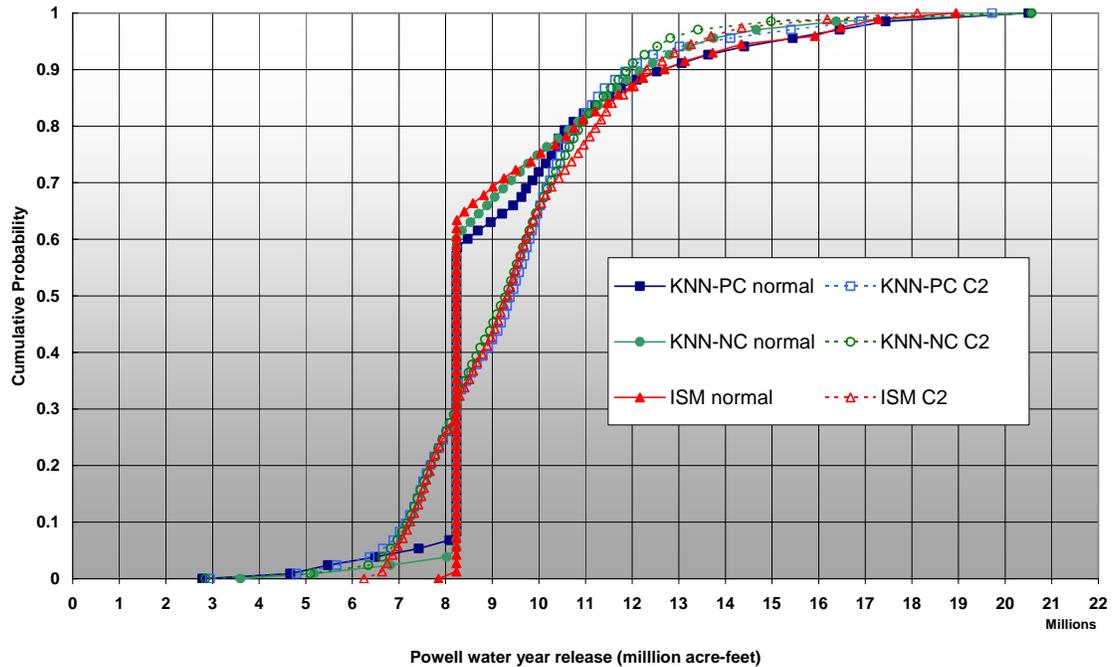
Figure 5-16 ECDF of Powell WY release.

To determine the probability of a reduced 10 year running sum annual water year release from Powell we display an ECDF of Powell 10 year running sum water year release (Figure 5-17) and the minimum of Powell 10 year running sum water year release (Figure 5-18). A line depicting 75 MAF is provided on both plots to easily determine when this threshold is passed. For both operation policies under ISM hydrologies the results never fall below the threshold. But under the K-NN hydrologies the paleo conditioned hydrologies fall below the threshold 7% of the time and the non conditioned hydrologies fall below the threshold 4% of the time. The paleo conditioned hydrologies generate longer drought sequences accounting for the increased probability and magnitude of shortage. The larger range of possible flows

from the K-NN hydrologies (48-170 MAF) versus ISM hydrologies (78 -130 MAF)
displays the effect of increased variety in flow and in flow sequences. The minimum
10 year running sum water year release from Powell is lowest with the K-NN paleo
conditioned hydrologies under normal operations followed by C2 operations. K-NN
non conditioned hydrologies under normal operations are followed by operations
under C2. A 10 year running sum water year release of only 48 MAF is seen for a K-
NN paleo conditioned hydrology under normal operations in 2040 and 2058.
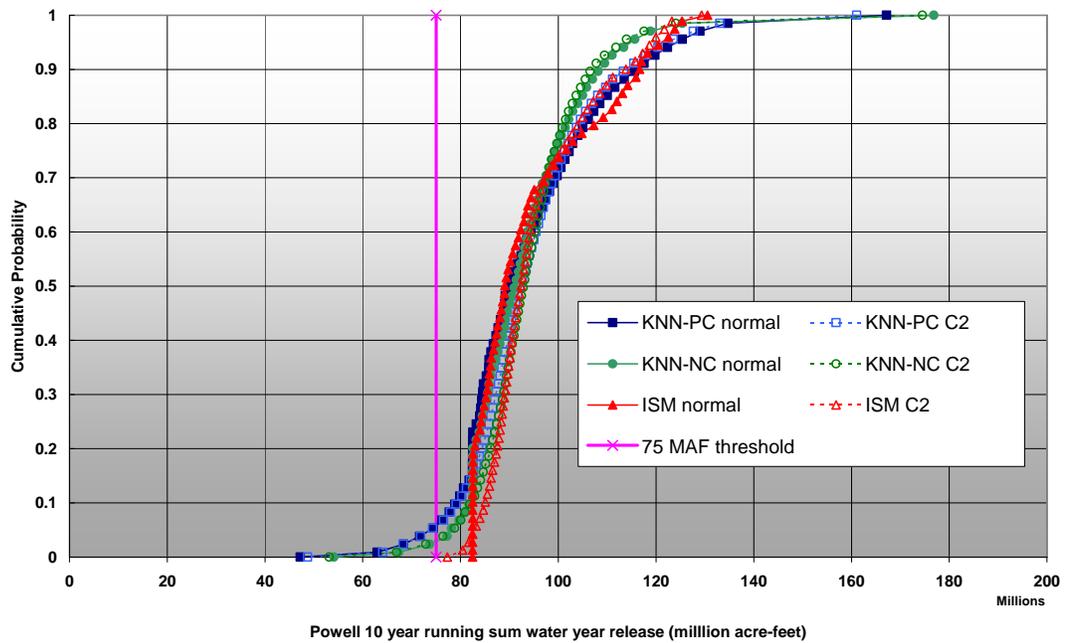
Figure 5-17 ECDF of 10 year running sum WY release from Powell.
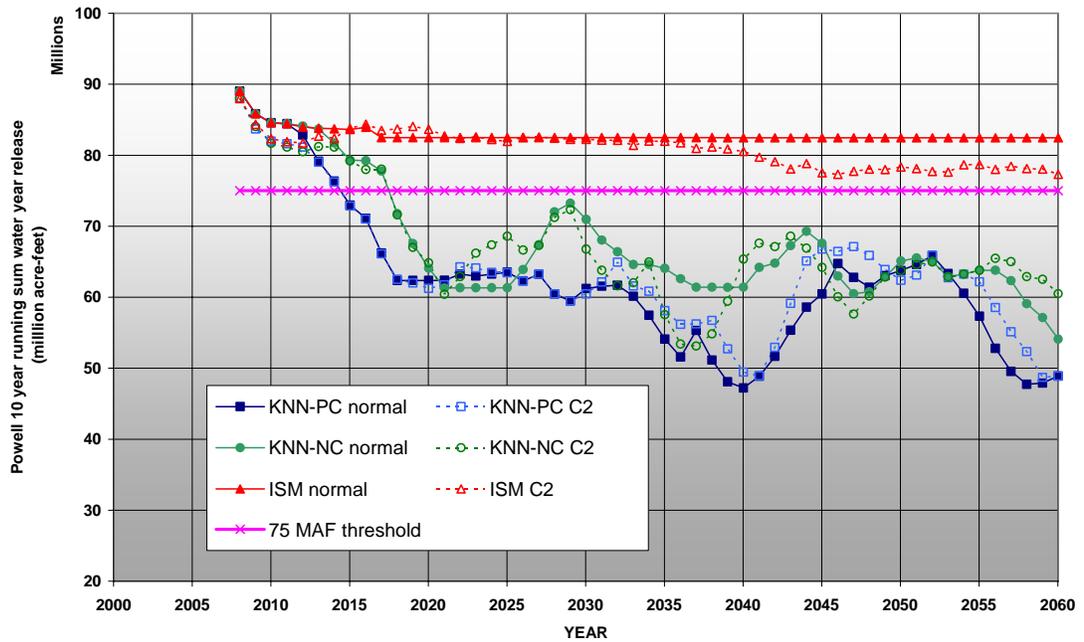
Figure 5-18 Minimum 10 year running sum WY release from Powell.

5.6 Salinity Concentration at Numeric Criteria Stations

The Federal Water Pollution Control Act Amendments of 1972 required development of fixed point numeric criteria for salinity in the Colorado River Basin. The fixed point numeric criteria were set in 1975 at an average annual salinity concentration of 723 mg/L below Hoover Dam; 747 mg/L below Parker Dam; 879 mg/L at Imperial Dam. The basin wide salinity model is evaluated at these three numeric criteria stations. To evaluate whether the numeric criteria are violated in any year over the simulation horizon the 50th percentile concentration at each station is compared with its respective numeric criteria.

Figure 5-19 presents the salinity concentrations resulting from the 6 scenarios below Hoover Dam. The ISM scenarios develop salinity based on monthly regressions while the K-NN scenarios are based on approach 2 presented in Chapter 3 (the annual regression at each site followed by temporal disaggregation of annual salt

107

to monthly). The ISM scenarios produce the lowest concentrations with little difference between the two reservoir policies. This results from the fully mixed assumption for salinity in the reservoirs that causes little appreciable change in average reservoir concentration for the two policies, though Figure 5-10 did indicate different storage levels in Mead for the two policies. Both K-NN based hydrologies generated higher salinity concentration than the ISM hydrologies; though again, both policies perform similarly. The paleo conditioned hydrologies tend to have a higher salinity than the non conditioned in the initial years of the simulation but are lower in the final years. This is a result of lower reservoir levels in the initial year for paleo conditioned hydrologies but higher reservoir levels in the long run due to increased surpluses. For all years and scenarios the simulations are well below the numeric criteria indicated at the top of the plot.
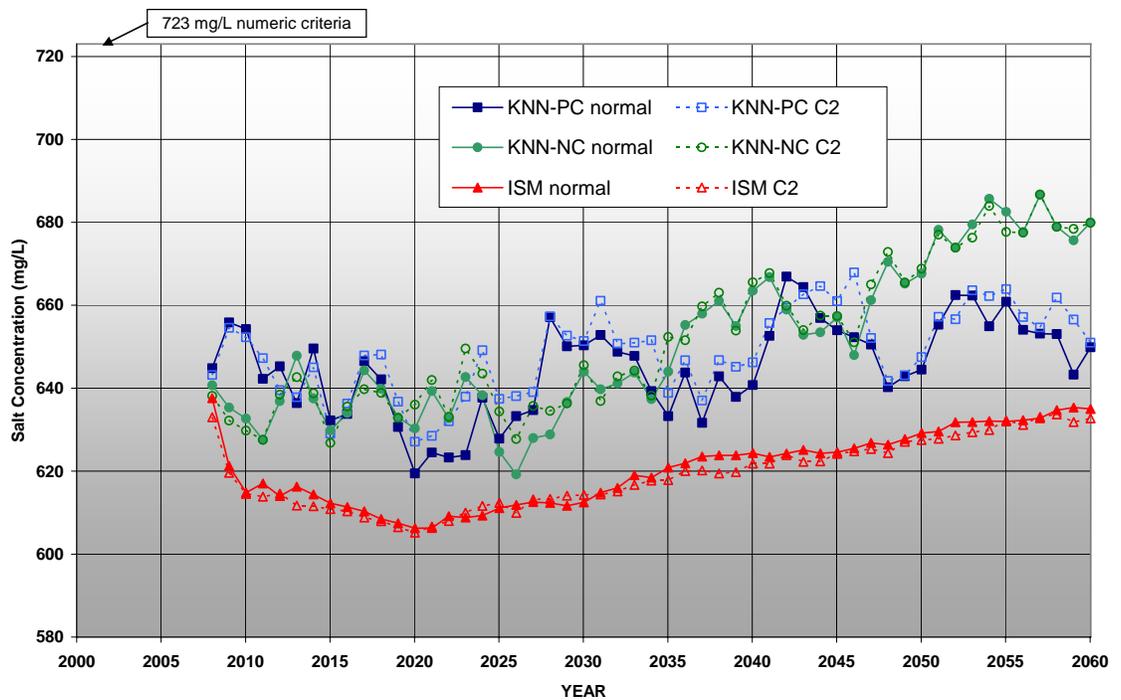
Figure 5-19 Salinity concentration 50th percentile below Hoover Dam.

Figure 5-20 presents the salinity concentrations resulting from the 6 scenarios below Parker Dam. Similar results to those seen below Mead are present below Parker. The separation between the ISM hydrology scenarios and the K-NN scenarios has been reduced. With further mixing, having routed the salt through both Lake Mohave and Havasu, the scenarios begin to appear more similar. Again we are below the numeric criteria for all scenarios.
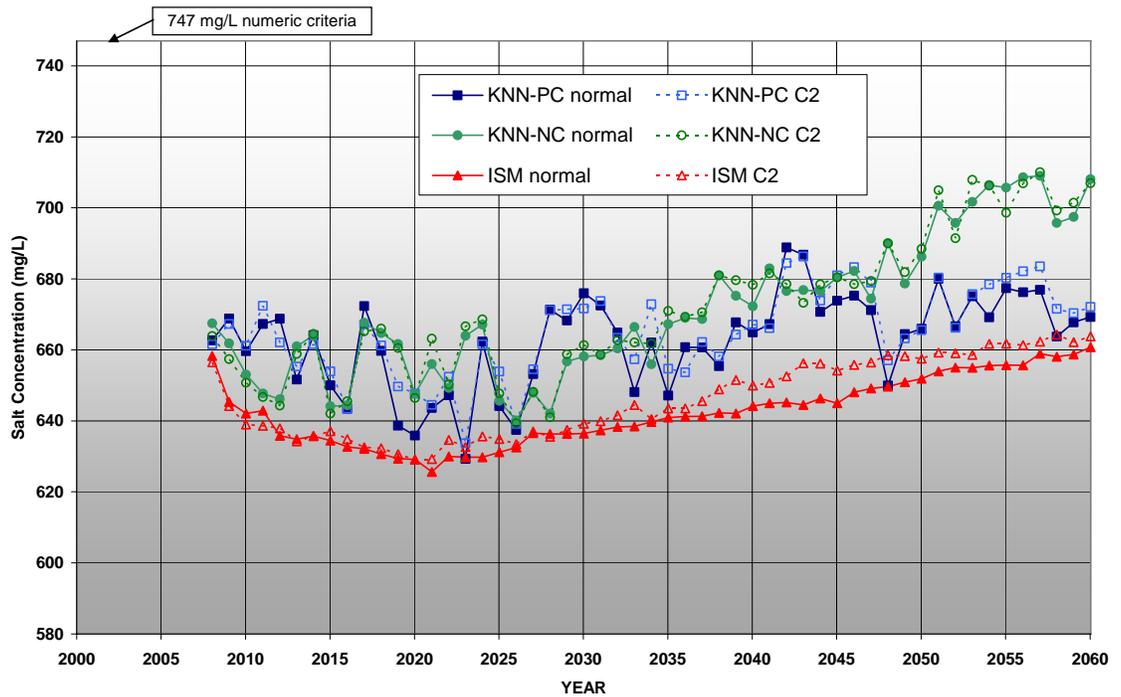
Figure 5-20 Salinity concentration 50th percentile below Parker Dam.

Figure 5-21 presents the salinity concentrations resulting from the 6 scenarios at Imperial Dam. Similar results are seen, though after passing through most all the Lower Basin the difference between the two K-NN hydrologies in the final years is basically gone. While in the initial years the paleo conditioned flows are consistently exhibiting the highest salinity concentration as a result of reduced flows, having diverted much of the required Lower Basin demands, and therefore increasing concentration. We are again well below the numeric criteria for the station.
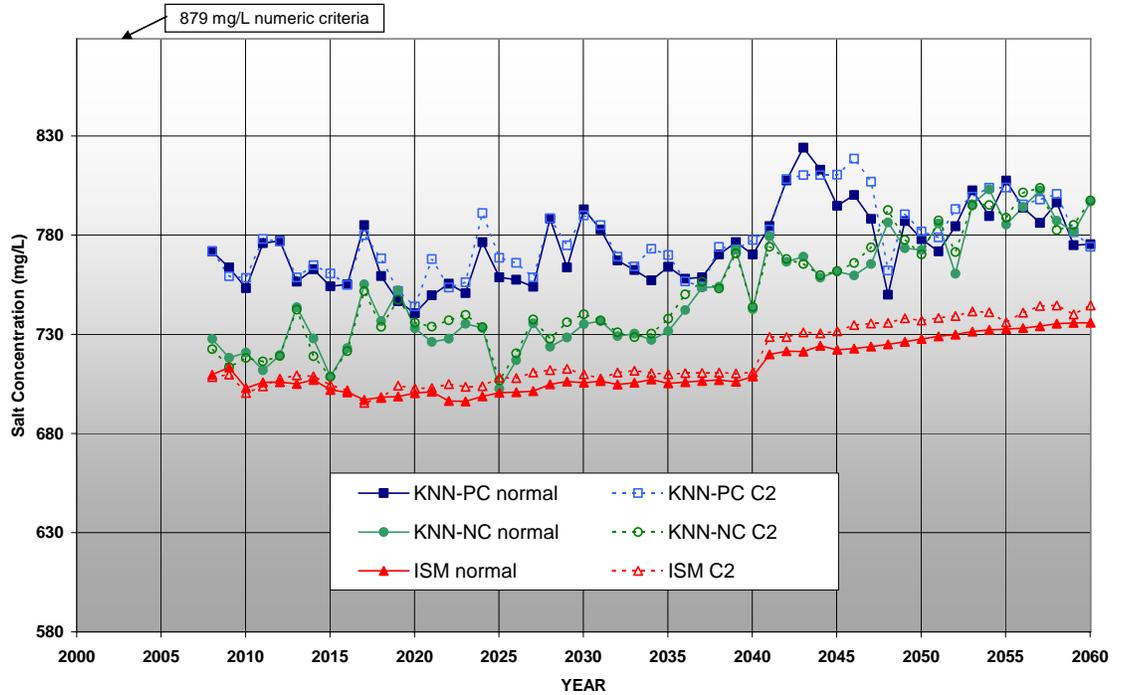
Figure 5-21 Salinity concentration 50th percentile at Imperial Dam.

5.7 Summary and discussion

Three alternate hydrologies and associated salinity and two alternate reservoir operating policies were developed and passed through the CRSS, the long-term planning model for the Colorado River Basin. The resulting model outputs were analyzed for several key decision variable relating to both flow and salinity concentration with an emphasis on variables that analyze the probability and magnitude of shortage.

The ISM hydrologies represent the current stochastic streamflow inputs used in CRSS. These streamflows are limited to only representing both magnitudes and sequences of streamflow that have occurred in the past 99 years. These limitations result in reduced variability in model outputs and detract from a water manager's ability to fully appreciate possible scenarios for streamflow that may occur in the future. This became painfully apparent with the onset of the recent drought (2000-

2004) that was never considered a possibility in previous model runs because such a sequence of streamflows had not previously occurred in the measured record.

Two alternate hydrologies were generated that included the ability to generate new monthly flow magnitudes and sequences. The first alternate hydrology depends on a modified K-NN lag-1 model generating stochastic streamflows that may not have actually occurred in the historic observed record (available form 1906-2004) but are statistically plausible based on the properties of the observed record. These data will allow basin mangers to better assess the risk of events such as shortage and reduced reservoir releases.

To extend our knowledge of the risks inherent as a result of climate variability, with regards to reservoir operations, the second alternate hydrology was conditioned on system state (i.e., wet or dry) information extracted from recent paleo streamflow reconstructions (Woodhouse et al., 2006) completed on the Colorado River at Lees Ferry (available from 1490-1997). Based on the reconstructed streamflows a nonhomogeneous Markov chain was used to simulate future system states. These system states where than used to conditionally resample an annual flow from the observed period with a traditional K-NN model. Together the Markov model and the K-NN resampling generate flows that were seen in the observed record but sequenced based on system states seen over the entire paleo reconstruction, This allows generation of drought and surplus sequences and magnitude that may not have occurred in the recent past; for example, the drought we recently experienced.

Addressing a shortcoming of ISM discussed in Jerla (2005) we were able to assess the risk of both shortage and reduced reservoir releases under normal and C2

or "balancing" operations with academically accepted stochastic streamflow generation techniques. Under normal operating conditions using the alternate hydrologies we found a 48% probability for any shortage as compared to a 52% probability with ISM. The probability of level 2 was 18% for K-NN with no conditioning and 17% with conditioning while ISM was 22%. Though ISM has an increased probability of level 2 shortage the magnitude is greatly reduced in comparison to both K-NN hydrologies. We also found that the K-NN non conditioned scenario had a 2.5% chance of shortage with a magnitude greater than 3.6 MAF while the paleo conditioned had a 5% chance of shortage greater than 3.6 MAF. The maximum shortage magnitude under ISM was only 1.45 MAF.

Under C2 operations similar results were found but all scenarios displayed an approximate 8% reduction in the chance of level 1 shortage. While the probability of a shortage greater than 3.6 MAF had the same probability, though the ISM hydrologies had an increased chance of shortages as large as the two alternate hydrologies. This indicates that under C2 though the risk of needing level 2 is lower, once in level 2 the magnitudes of these shortages are similar under both operation alternatives.

Analysis of the probability of reduced reservoir releases under the alternate hydrologies indicated that ISM hydrologies never fall below the 75 MAF threshold for 10 year running sum water year releases from Powell. The alternate hydrologies with paleo conditioning indicate a 4% chance the 10 year running sum water year releases from Powell could fall below 75 MAF; while generating sequences of flows not seen in the observed streamflows but using magnitudes seen within the observed

record indicates a 7% chance. These results were seen under both normal and C2 operations.

The new basin wide salinity model generated higher salt mass and concentration than simulated with the monthly flow and salt relationships. This results from a better representation of the flow and salt relationship at the annual time step. Poor regression relationships in low flow months lead to reduced salinity values in these months. Though even with increased values the salinity concentrations were all found to be well below the numeric criteria at all locations for all the flow simulations. The increased salinity concentration demonstrated with the new basin wide salt model does indicate the possibility for increased economic damages resulting from salinity.

Clearly, the development of hydrologies that are statistically plausible but have not recently been observed is an essential addition to a basin manger's tools when evaluating the risks to stakeholders regarding reservoir operating strategies and should be included in all analysis to gain an appreciation of shortage risks under changing hydrologies.

CHAPTER 6

SUMMARY AND CONCLUSIONS

This chapter summarizes the key findings and conclusions from each component of the integrated framework developed in this research, followed by suggestions for future extensions.

6.1 Summary

The Colorado River system recently experienced the worst five year drought in the 100 years of measured streamflows. The vast networks of reservoirs that sustain and fuel development were drawn below 50% of system capacities to historic lows. The dry period also impacted water quality, increasing salinity concentration as a result of low flows. Paleo reconstructions of streamflow in the basin indicate that such dry spells are not uncommon. Interestingly, a compact to share water resources among the basin states, developed during wet periods of the early 1920s, is under stress as the states confront economic growth under limited water resources (Kuhn, 2005).

Given these factors, the key question for this research was how to plan for effective and sustainable management of water resources in the basin? This requires two key components; (i) a robust framework to generate realistic future basin wide streamflow and associated salinity scenarios and (ii) a decision making model to evaluate operating policy alternatives for efficient management and sustainability of water resources in the basin. To achieve this, three inter-related modules were developed including. (i) A stochastic nonparametric model for basin wide streamflow generation based on historic observations combined with paleo reconstructed

streamflow. This new model is data driven and improves considerably upon traditional approaches, besides being simple, robust, and flexible. A nonhomogeneous Markov Chain based approach to combine hydrologic state information (i.e., wet or dry) from the paleo streamflow and the flow magnitude from the historic data is a unique and novel method to combine the strengths of these two different data sets. Together, these constitute a significant methodological contribution from this research. (ii) A basin wide stochastic model for generating salinity scenarios, extending the nonparametric flow model for salinity simulation. (iii) A realistic decision model of the basin to evaluate policy alternatives under various flow and salinity scenarios.

6.2   Conclusions

6.2.1   A Stochastic Nonparametric Technique For Space-Time Disaggregation Of Streamflows

A simple, robust, and parsimonious stochastic nonparametric space-time framework for large river networks was developed. The framework builds on work presented by Tarboton et al. (1998) but adopts a K-NN approach to construct and resample from a conditional PDF. This generates consistent monthly streamflow scenarios across all the locations in the Colorado River basin from annual streamflows at a single aggregate site. Application of the framework to streamflows in the upper Colorado River basin showed faithful reproduction of the spatial and temporal distributional statistics of the observed flows and, also simulation of streamflow magnitudes and sequences not seen in the observed data, unlike the Index Sequential Method (ISM) widely used by Reclamation.

Traditional stochastic disaggregation methods have several limitations including (i) assumption of Normal distribution, (ii) assumption of linear relationship besides, being mathematically complex. As a result, they cannot capture non-Normal and nonlinear relationships that might be present in the data and, they are difficult for agencies such as Reclamation to readily adopt in practice. The presented disaggregation model can generate streamflow ensembles across the basin at short (seasonal) and long (multi-year) time scales.

6.2.2 A Basin wide Stochastic Salinity Model

A framework to generate stochastic salinity scenarios consistent with stochastic streamflow scenarios was presented. In this, annual salinity at the aggregate location is first generated from the annual flow, using a local polynomial regression model developed by Prairie et al. (2005). These are then disaggregated to salinity at all the spatial location using the disaggregation approach developed in Chapter 2. Other variations were also proposed where the salinity is generated from the local polynomial regression at each location separately, based on the disaggregated streamflows. This ensures the capture of the flow and salt relationship at each gauge, which is desired in the application, hence is recommended. This methodology to generate basin wide salinity is a significant contribution.

6.2.3 Stochastic Streamflow Simulation Incorporating Paleo Reconstruction

As mentioned earlier, the Colorado River Basin experienced the worst 5 year drought on record during 2000-2004. But paleo reconstructions of streamflow of the pre-observed period, from tree-ring chronologies, show droughts of greater magnitude and duration indicating that the recent drought is not unusual. Clearly, the

rich information provided by paleo reconstructions have to be incorporated in any stochastic streamflow models to enable the generation of a realistic variety of plausible flow scenarios for robust planning and management of water resources in the basin. However, the magnitudes of reconstructed streamflow reconstructions have high degree of uncertainty due to the reconstruction methodology. This apparent weakness of the paleo reconstructed flow data has made their use in a water resources planning context contentious, despite their availability for many decades. Though this weakness exists few argue about the duration and frequency of dry and wet (i.e., the hydrologic state) periods from the reconstructions. The key question is how to combine the long paleo reconstructed streamflow information of lesser reliability with the shorter but reliable observed data to develop a framework for streamflow simulation?

To address this, a nonparametric stochastic framework for streamflow simulation combining the long paleo reconstructed streamflow information of lesser reliability with the shorter but reliable observed data was developed. The framework has two components (i) a nonhomogeneous Markov Chain model developed on the paleo data, which is then used to simulate the hydrologic state, and (ii) a K-nearest neighbor (K-NN) time series bootstrap to simulate the streamflow magnitude from the observed data conditioned on the hydrologic state and the previous flow magnitude. This new and unique framework combines the respective strengths of the two data sets. Furthermore, it is data driven, robust, and parsimonious. The framework was applied to paleo reconstructed streamflow and observed data for the Lees Ferry, AZ, streamflow gauge on the Colorado River. The simulations showed the ability to

capture the distributional statistics of the observed period and also generate a rich variety of wet and dry sequences that will greatly benefit the sustainable management of water resources in the basin. The annual streamflow generated from Lees Ferry, AZ from this approach can be spatially and temporally disaggregated (Prairie et al., 2006b) to obtain monthly flow scenarios at all the gauges in the basin, which will help drive a basin wide decision model and consequently, the realistic estimation of risk and reliability of various decision components in the water resources system. The developed framework will enable the water resources planners to use the rich insights from the paleo reconstructions by alleviating its short comings. This is the major contribution from this research. Extension of this method can also be used to generate streamflow sequences conditioned on climate change scenarios.

6.2.4 Framework Application In The Colorado River Simulation System Decision Support System

Effective planning and management for sustainable water resources in the basin requires evaluating the impact of operating policies on the water quantity and salinity in the various system components based on a rich variety of plausible streamflow scenarios. Three hydrologic scenarios were used from (i) the ISM, the current method used by Reclamation, (ii) the nonparametric space-time disaggregation framework developed in Chapter 2, based on the observed data and, (iii) the coupled methodology combining paleo reconstructed streamflows and observed data (developed in Chapter 4) in conjunction with the space-time disaggregation method. Corresponding salinity scenarios were generated from the method described in Chapter 3.

118

The flow and salinity scenarios were used to drive the long-term planning model of the water resources system in the Colorado River basin called, CRSS. Impact of two alternate operating policies on the quantity and quality (i.e., salinity) of the various subsystems, are evaluated. The first policy was based on current operation in the Colorado River basin while the second represented a policy to balance the volume of Lake Powell and Mead over the water year to better respond to the dry conditions. Under normal operation the primary objective for Powell is to provide a minimum objective release of 8.23 MAF over the water year and Powell further is required to balance (equalize) with Mead during high reservoir levels. While under balancing the primary objective is to keep Powell and Mead balanced. This may mean relaxing the 8.23 MAF minimum objective releases and balancing at lower reservoir levels. Shortage guidelines under both policies are identical and attempt to protect the Mead elevation of 1050 feet approximately 80% of the time under level 1 shortage and attempts to protect the Mead elevation of 1000 feet 100% of the time under level 2 shortage. General reservoir response to both the alternate scenarios and policies were compared with an emphasis on shortage probability and volume.

The ISM scenarios, as expected, had the least variability and sequences only from the observed data. Scenarios using the paleo reconstructed flows had the most variability across all flow regimes including longer surplus and drought lengths and volumes. Shortage lengths and volumes were also found to be most severe under the paleo conditioned scenarios. While under normal reservoir operations the shortage were found to occur more frequently but with reduced magnitudes in comparison with the balancing policy.

Of particular interest was the probability of releasing a 10 year running sum water release of less than 75 MAF from Powell. The ISM scenario indicated that the 10 year running sum water year release from Powell would always exceed 75 MAF in the next 53 years. But, flow scenarios from the two other methods indicated 4% to 7% probability of releases less than 75 MAF in the next 53 years. The balancing policy seems to help reduce the vulnerability to flow variability. The salinity concentrations were all found to be well below the numeric criteria at all locations for all the flow simulations. Though increased salinity concentration with the new basin wide salt model indicates the possibility of increased economic damages. These insights are very helpful in devising effective management strategies for sustainable water resources in the Colorado River basin.

6.3 Future Work

Several potential extensions and applications of the research developed in this thesis are possible, which are listed below.

1. Alternate models (e.g., parametric, semi-parametric) for annual flow generation at the aggregate gauge needs to be explored to generated a wider range of flows. These could then be run through the nonparametric disaggregation framework. Also, methods to capture inter-annual variability and correlation between the first month of the current year and the last month of the previous year have to be developed.

2. In the nonhomogenous Markov chain model the number of hydrologic states are defined based on a pre-determined threshold which is a subjective choice. A hidden Markov chain framework would alleviate this subjectivity and better model regimes objectively.

3. Monthly salinity scenarios are generated to drive the decision model, despite policy evaluations at the annual time scale. Besides, the annual flow and salt relationships are much robust (Prairie et al., 2005). Jerla (2005) demonstrated that policy in the Colorado River basin can be adequately represented at an annual time step. Therefore, it seems appropriate to develop a decision support model at an annual time step, which would eliminate the uncertainty added by generating monthly flow and salinity scenarios.

4. Other alternate policies need to be explored. Also, this can be cast as an optimization problem to obtain optimal policy solutions, under different objective functions (i.e., economic, utility etc.).

5. Streamflow scenarios consistent with climate change projections can be used to investigate the policy implications under a changing climate.

REFERENCES

Akıntuğ B. and P.F. Rasmussen (2005), A Markov-Switching model for annual hydrologic time series, *Water Resources Research,* 41, W09424, doi:10.1029/2004WR003605.

Andreadis, K. M. and D. P. Lettenmaier (2006), Trends in 20th century drought over the continental United States, *Geophysical Research Letters*, 35, L10403, doi:10.1029/2006GL025711.

Bellone, E., J.P. Hughes, and P. Guttorp (2000), A hidden Markov model for downscaling synoptic atmospheric patterns to precipitation amounts, *Clim. Res.*, 15, 1-12.

Bowman, A. W. and A. Azzalini (1997), *Applied smoothing techniques for data analysis*, Clarendon.

Cayan, D. R., S. A. Kammerdiener, M. D. Dettinger, J. M. Caprio, and D.H. Peterson (2001), Changes in the onset of spring in the western United States, *Bull. Amer. Meteor. Soc.,* 82, 399-416.

Colorado River Basin Salinity Control Forum (2005), *Water Quality Standards For Salinity Colorado River System*, Bountiful, Utah.

Cook, E.R., K. Briffa, S. Shiyatov, and V. Mazepa (1990), Tree-ring standardization and growth-trend estimation, *in Methods of Dendrochronlogy: Applcations in the Environmental Sciences*, edited by E.R. Cook and L.A. Kairiukstis, pp. 153-162, Springer, New York.

Efron, B. (1982), *The Jackknife, the Bootstrap and Other Resampling Plans*, Society for Industrial and Applied Mathematics, Providence, RI.

Feyerherm, A.M. and L.D. Bark (1965), Statistical methods for persistent precipitation patterns, *Journal of Applied Meteorology*, 4, 320-328.

Fukunaga, K. (1990), *Introduction to statistical pattern recognition*, Academic, San Diego, California.

Fulp, T. (2005), How low can it go, *Southwest Hydrology*, 4(2), 28.

Gabriel, K.R. and J. Neumann (1962), A Markov chain model for daily rainfall occurrence at Tel Aviv, *Quart. J. Roy. Meteor. Soc.*, 88, 90-95.

Gates, P., and H. Tong (1976), On Markov chain modeling to some weather data, *Journal of Applied Meteorology*, 15, 1145-1151.

Grantz, K., B. Rajagopalan, M. Clark, and E. Zagona (2005), A Technique for Incorporating Large-Scale Climate Information in Basin-Scale Ensemble Streamflow Forecasts, *Water Resources Research*, 41, W10410, doi:10.1029/2004WR003467.

Grygier, J.C. and J.R. Stedinger (1988), Condensed disaggregation procedures and conservation corrections for stochastic hydrology, *Water Resources Research,* 24(10), 1574-1584.

Harding, B.L., T.B. Sangoyomi, E.A. Payton (1995), Impacts of Severe Sustained Drought on the Colorado River Water Resources, *Water Resources Bulletin*, 31(5), 815-824.

Hidalgo, H.G., T.C. Piechota, and J.A. Dracup (2000), Alternative principal components regression procedures for dendrohydrologic reconstructions, *Water Resources Research*, 36(11), 3241-3249.

Hirschboeck, K.K. and D. M. Meko (2005), *A Tree-Ring Based Assessment of Synchronous Extreme Streamflow Episodes in the Upper Colorado & Salt-Verde-Tonto River Basins- Final Report*, Laboratory of Tree-Ring Research, University of Arizona.

Hoerling, M. and A. Kumar (2003), The perfect ocean for drought, *Science*, 299(5607), 691-694.

Hughes, J.P. and P. Guttorp (1994), A class of stochastic models for relating synoptic scale atmospheric patterns to regional hydrologic phenomena, *Water Resources Research*, 30(5), 1535-1546.

Hughes, J.P., P. Guttorp, and S.P. Charles (1999), A non-homogeneous hidden Markov model for precipitation occurrence, *Appl. Stat.*, 48(1), 15-30

Jerla, C.S. (2005), An analysis of coordinated operation of Lakes Powell and Mead under lower reservoir conditions, Masters thesis, University of Colorado.

Koutsoyiannis, D. (1992), A nonlinear disaggregation method with a reduced parameter set for simulation of hydrologic series, *Water Resources Research*, 28(12), 3175-3191.

Koutsoyiannis, D. (1999), Optimal decomposition of covariance matrices for multivariate stochastic models in hydrology, *Water Resources Research*, 35(4), 1219-1229.

Koutsoyiannis, D. (2001), Coupling stochastic models of different timescales, *Water Resources Research*, 37(2), 379-391.

Koutsoyiannis, D. and A. Manetas (1996), Simple disaggregation by accurate adjusting procedures, *Water Resources Research*, 32(7) 2105-2117.

Kuhn, E. (2005), Colorado River water supplies: back to the future, *Southwest Hydrology*, 4(2), 28.

Kumar, D.N., U. Lall, and M.R. Peterson (2000), Multisite disaggregation of monthly to daily streamflow, *Water Resources Research*, 36(7), 1823-1833.

Lall, U. (1995), Recent advances in nonparametric function estimation: hydraulic applications, *U.S. Natl. Rep. Int. Union Geod. Geophys. 1991-1994, Rev. Geophys.*, 33, 1093-1102.

Lall, U. and A. Sharma (1996), A nearest neighbor bootstrap for resampling hydrologic time series, *Water Resources Research*, 32(3), 679-693.

Lambert, M.F., J.P. Whiting, and A.V. Metcalfe (2003), A non-parametric hidden Markov model for climate state identification, *Hydrology and Earth System Sciences*, 7(5), 652-667.

Lane, W. L. (1979), *Applied stochastic techniques, users manual*, Engineering and Resource Center, Bureau of Reclamation, Denver, Colo.

Lane, W.L. (1982), Corrected parameter estimates for disaggregation schemes, in *Statistical Analysis of Rainfall and Runoff*, edited by V. P. Singh, Water Resources Publications, Littleton, Colo.

Lane, W.L. and D.K. Frevert (1990), *Applied stochastic techniques, personal computer version 5.2, users manual*, Earth Sciences Division, Bureau of Reclamation, Denver, Colo.

Loader, C. (1999), *Local regression and likelihood*, Springer, New York.

Loucks, D.P., J.R. Stedinger, and D.A. Haith (1981), *Water Resource System Planning and Analysis*, Prentice-Hall, Inc., New Jersey.

Lu, Z.Q. and L.M. Berliner (1999), Markov switching time series models with application to daily runoff series, *Water Resources Research*, 35, 523-534

MacDonald, I.L., and W. Zucchini (1997), *Hidden Markov and other models for discrete-valued time series*, Chapman and Hall.

Malone, R. F., D.S. Bowles, W.J. Grenney, and M.P. Windham (1979), *Stochastic analysis of water quality*, Utah Water Research Laboratory, Utah State Univ., Logan, Utah.

Mehrotra, R. and A. Sharma (2005), A nonparametric nonhomogeneous hidden Markov model for downscaling of multisite daily rainfall occurrences, *Journal of Geophysical Research*, 110, D16108, doi:10.1029/2004JD005677.

Mehrotra, R., A. Sharma, and I. Cordery (2004), Comparison of two approaches for downscaling synoptic atmospheric patterns to multisite precipitation occurrences, *Journal of Geophysical Research*, 109, D14107, doi:10.1029/2004JD004823.

Mejia, J. M. and J. Rousselle (1976), Disaggregation models in hydrology revisited, *Water Resources Research*, 12(2), 185-186.

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

Moon, Y.-I., and U. Lall (1994), Kernel function estimator for flood frequency analysis, *Water Resources Research*, 30(11), 3095-3103.

Mote, P.W. (2003), Trends in snow water equivalent in the Pacific Northwest and their climatic causes, *Geophysical Research Letters*, 30(12), Art. No. 1601.

Mueller, D.K. and L.L. Osen (1988), *Estimation of natural dissolved-solids for the upper Colorado River basin*, U.S. Geological Survey, Water Resources Investigation Report 87-4069, Denver, Colorado.

Nathanson, M. N. (1978), *Updating the Hoover Dam documents*, Bureau of Reclamation, United Stated Department of the Interior, Denver.

Ouarda, T., J.W. Labadie, and D.G. Fontane (1997), Index sequential hydrologic modeling for hydropower capacity estimation, *J. of the American Water Resources Association*, 33(6), 1337-1349.

Prairie, J. and R. Callejo (2005), *Natural flow and salt computation methods*, U.S. Dept. of Interior, Salt Lake City, Utah.

Prairie, J.R., B. Rajagopalan, T.J. Fulp, and E.A. Zagona (2005), Statistical nonparametric model for natural salt estimation, *ASCE Journal of Environmental Engineering*, 131(1), 130-138.

Prairie, J.R., B. Rajagopalan, T.J. Fulp, and E.A. Zagona (2006a), Modified K-NN model for stochastic streamflow simulation, *ASCE Journal of Hydrologic Engineering*, 11(4), 371-378.

Prairie, J.R., B. Rajagopalan, U. Lall, and T.J. Fulp (2006b), A stochastic nonparametric technique for space-time disaggregation of streamflows, *Water Resources Research* (in press).

Rajagopalan, B. and U. Lall (1999), A k-nearest-neighbor simulator for daily precipitation and other weather variables, *Water Resources Research*, 35(10), 3089-3101.

Rajagopalan, B., U. Lall, and D.G. Tarboton (1996), Nonhomogeneous Markov model for daily precipitation, *Journal of Hydrologic Engineering*, 1(1), 33-39.

Regonda, S.K., B. Rajagopalan, M. Clark, and J. Pitlick (2005), Seasonal cycle shifts in hydroclimatology over the Western United States, *Journal of Climate*, 18, 372-384.

Roldan, J. and D.A. Woolhiser (1982), Stochastic daily precipitation models. 1. A comparison of occurrence processes, *Water Resource Research*, 18(5), 1451-1459.

Salas, J.D. (1985), Analysis and modeling of hydrologic time series. In: *Handbook of Hydrology*. Edited by D. R. Maidment, McGraw-Hill, New York, 19.1-19.72.

Salas, J.D., J.W. Delleur, V. Yevjevich, and W.L. Lane (1980), *Applied Modeling of Hydrologic Time Series*, 484 pp., Water Resources, Littleton, Colo.

Santos, E.G. and J.D. Salas (1992), Stepwise disaggregation scheme for synthetic hydrology, *Journal of Hydraulic Engineering, ASCE*, 118 (5), 765–784.

Schweppe, F. C. (1973), *Uncertain dynamic systems*, Prentice-Hall, Inc. Englewood Cliffs, New Jersey.

Scott, D.W. (1992), *Multivariate density estimation: Theory, practice, and visualization*, Wiley Ser. In Probability and Math. Statistics. John Wiley and Sons, New York, N.Y.

Sharma, A. and R. O'Neill (2002), A nonparametric approach for representing interannual dependence in monthly streamflow sequences, *Water Resources Research*, 38(7), 1100-1100.

Sharma, A., D.G. Tarboton, and U. Lall (1997), Streamflow simulation: a nonparametric approach, *Water Resources Research*, 33(2), 291-308.

Singhrattna, N., B. Rajagopalan, M. Clark, and K. Krishna Kumar (2005), Forecasting Thailand summer monsoon rainfall, *International Journal of Climatology*, 25(5), 649-664.

Smith, J.A. and H.A. Schreiber (1974), Point processes of seasonal thunderstorm rainfall. 1. Distribution of rainfall event, *Water Resources Research*, 10(3), 418-423.

Souza Filho, F.A., and U. Lall (2003), Seasonal to interannual ensemble streamflow

forecasts for Ceara, Brazil: Applications of a mutlivariate, semi-parametric algorithm, *Water Resources Research*, 39(11), 1307-1320.

Srinivas, V.V and K. Srinivasan (2001), Post-blackening approach for modeling periodic streamflows, *Journal of Hydrology*, 241, 221–269.

Srinivas, V.V. and K. Srinivasan (2005), Hybrid moving block bootstrap for stochastic simulation of multi-site multi-season streamflows, *Journal of Hydrology*, 302, 307-330.

Stedinger, J.R., and R.M. Vogel (1984), Disaggregation procedures for generating serially correlated flow vectors, *Water Resources Research*, 20(1), 47-56.

Stedinger, J.R., D. Pei and T.A. Cohn (1985), A condensed disaggregation model for incorporating parameter uncertainty into monthly reservoir simulations, *Water Resources Research*, 21(5), 665-675.

Stockton, C.W. and G.C. Jacoby (1976), Long-term surface-water supply and streamflow trends in the Upper Colorado River Basin, *Lake Powell Res. Proj. Bull.* 18, Natl. Sci. Found., Arlington, Va.

Stokes, M.A. and T.L. Smiley (1968), *An Introduction to Tree-Ring Dating*, Univ. of Ariz. Press, Tucson.

Swetnam, T.W., M.A. Thompson, and E.K. Sutherland (1985), *Using Dendrochronology to Measure Radial Growth of Defoliated Trees, Agric. Handb.*, vol. 639, For. Serv., U.S. Dep. Of Agric., Washington, D.C.

Tao, P.C. and J.W. Delleur (1976), Multistation, multiyear synthesis of hydrologic time series by disaggregation, *Water Resources Research*, 12(6), 1303-1312.

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

Thyer, M. and G. Kuczera (2000), Modleing long-term persistence in hydroclimatic time series using a hidden state Markov model, *Water Resources Research*, 36, 3301-3310.

Thyer, M. and G. Kuczera (2003a), A hidden Markov model for modeling long-term persistence in multi-site rainfall time series. 1, Model calibration using a Bayesian approach, *Journal of Hydrology*, 275, 12-26.

Thyer, M. and G. Kuczera (2003b), A hidden Markov model for modeling long-term persistence in multi-site rainfall time series. 2, Real data analysis, *Journal of Hydrology*, 275, 27-48.

Todini, E. (1980), The preservation of skewness in linear disaggregation schemes, *Journal of Hydrology*, 47, 199-214.

Todorovic, P. and D.A. Woolhiser (1975), Stochastic model of n-day precipitation, *J. Appl. Meteorology*, 14(1), 17-24.

U.S. Department of the Interior (1987), *Colorado River simulation system: system overview*, Bureau of Reclamation., Denver, Colorado.

U.S. Department of the Interior (2000), *Colorado River interim surplus criteria, final environmental impact statement*, Bureau of Reclamation., Lower Colorado Region, Boulder City, Nevada.

U.S. Department of the Interior (2002), I*mplementation Agreement, Inadvertent Overrun and Payback Policy, and Related Federal Actions, Final Environmental Impact Statement*, Bureau of Reclamation., Lower Colorado Region, Boulder City, Nevada

U.S. Department of the Interior (2003a), *Colorado River Water Delivery Agreement,* I*mplementation Agreement, Inadvertent Overrun and Payback Policy, and Related Federal Actions, Final Environmental Impact Statement*, Bureau of Reclamation, Lower Colorado Region, Boulder City, Nevada.

U.S. Department of the Interior (2003b), *Quality of Water Progress Report 21*, Bureau of Reclamation, Upper Colorado Region, Salt Lake City, Utah.

Valencia, D. R., and J. C. Schaake (1973), Disaggregation processes in stochastic hydrology, *Water Resources Research*, 9(3), 580-585.

Vogel, R.M. and A.L. Shallcross (1996), The moving blocks bootstrap versus parametric time series models, *Water Resources Research*, 32(6), 1875-1882.

Weisberg, S. (1985), *Applied Linear Regression*, 2nd ed., John Wiley, Hoboken, N.J.

Woodhouse, C.A. and P.M. Brown (2001), Tree-ring evidence for Great Plains drought, *Tree-Ring Research*, 57, 89-103.

Woodhouse, C.A., S.T. Gray, and D.M. Meko (2006), Updated streamflow reconstructions for the Upper Colorado River Basin, *Water Resource Research*, 42, W05415, doi:10.1029/2005WR004455.

Wurbs, R. A. and A.S. Karma (1995), Salinity and water-supply reliability, *Journal of Water Resource Planning Management*, 121(5), 352–358.

Wurbs, R. A., I. Saleh, and A.S. Karma (1995), Reservoir system reliability constrained by salinity, *Water Resour*ces *Development*, 11(3), 273–287.

Yates, D., S. Gangopadhyay, B. Rajagopalan, and K. Strzepek (2003), A technique for generating regional climate scenarios using a nearest neighbor bootstrap, *Water Resour. Res.*, 39(7), 1199.

Zagona, E.A., T.J. Fulp, R. Shane, T.M. Magee, and H.M. Goranflo (2001), RiverWare: A generalized tool for complex reservoir system modeling, *Journal of American Water Resources Association*, 37(4), 913-929.

Zucchini, W. and P. Guttorp (1991), A hidden Markov model for space-time precipitation, *Water Resources Research*, 27(8), 1917-1923.