

STOCHASTIC STREAMFLOW SIMULATION AT INTERDECADAL TIME SCALES AND  
IMPLICATIONS TO WATER RESOURCES MANAGEMENT IN THE COLORADO RIVER  
BASIN

By

KENNETH C NOWAK

B.S., Rensselaer Polytechnic Institute, 2006

M.S., University of Colorado, 2008

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

2011

This thesis entitled:  
Stochastic Streamflow Simulation at Interdecadal Times Scales and  
Implications for Water Resources Management in the Colorado River Basin  
written by Kenneth C Nowak  
has been approved for the  
Department of Civil, Environmental, and Architectural Engineering

---

Balaji Rajagopalan

---

Edith Zagona

---

James Prairie

---

Martin Hoerling

---

Roseanna Neupauer

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.

Kenneth C Nowak (Ph.D., Civil, Environmental, and Architectural Engineering)

Stochastic Streamflow Simulation at Interdecadal Times Scales and

Implications for Water Resources Management in the Colorado River Basin

Thesis directed by Professor Balaji Rajagopalan

The severe and sustained nature of the recent drought in the southwestern United States, coupled with a growing collection of scientific literature indicating that anthropogenic climate change will further dry the region, have raised considerable concern about the reliability of its water resources. Undoubtedly, climate change poses a significant threat to water supplies in the Southwest. However, relative to this long-term shift in basin yield, natural variability is likely to be the key driver of supply risk over, say, a 20 year horizon. Hence, the focus of this work is on the development of tools and strategies to address streamflow variability, at the decadal time scale, with application for improved water management in the Colorado River Basin.

Streamflow simulation and disaggregation methods are first presented. These novel techniques offer the ability to simulate multi-site data at a variety of time scales – including daily, while also preserving non-stationary spectral properties seen in historic data. An analysis of Colorado River flow variability follows. Here, physical mechanisms are established for significant variability modes, and subsequently, links with large scale climate phenomena are proposed. Based on insight gained from the aforementioned work, a technique for decadal flow regime projections in the Colorado River Basin was established. The approach is demonstrated on paleo reconstructed data and then used to make actual future projections. Last, the Bureau of Reclamation Colorado River long-term planning model (CRSS) is used to assess adaptive operational policies. The objective is to improve Basin-wide supply reliability under various

potential future hydrologic conditions, including climate change scenarios derived from downscaled general circulation model output. These policies were developed to reduce system losses and capitalize on information provided by flow regime projections. Results show that under the variety of hydrologic conditions notable savings are seen, which help to reduce overall system risk of shortage conditions.

## **Dedication**

For my wife, Kristen. Words can not express my gratitude and love.

## **Acknowledgements**

This work would not have been possible without the assistance of many individuals. I would like to first thank my co-advisors, Dr. Balaji Rajagopalan and Dr. Edith Zagona, for their enthusiasm and guidance in undertaking this endeavor and seeing it through completion. I would also like to thank my committee members, Dr. Martin Hoerling, Dr. Roseanna Neupauer and Dr. Jim Prairie. Their insight and expertise is evident throughout this work.

I am indebted to Russ Callejo, Jim Prairie and Carly Jerla of the Bureau of Reclamation. Their willingness to time and again delve into details of the Colorado River Basin and CRSS was an essential resource in the final stages of this project. Thanks are also due to the staff at the Center for Advanced Decision Support for Water and Environmental Systems. I am especially grateful for technical assistance from Jim Pasquotto. For countless hours of discussion, feedback and model troubleshooting, I thank the other students at CADSWES, particularly Alan Butler and Cameron Bracken.

Finally, I thank my family; their encouragement and support continues to be my greatest asset.

## Contents

CHAPTER 1: INTRODUCTION.....	1
1.1 Background.....	1
1.2 Motivation.....	2
1.3 Outline .....	4
CHAPTER 2: A NONPARAMETRIC STOCHASTIC APPROACH FOR MULTISITE DISAGGREGATION OF ANNUAL TO DAILY STREAMFLOW.....	7
2.1 Introduction.....	7
2.2 Annual to Daily Disaggregation Method.....	11
2.2.1 Implementation Algorithm .....	11
2.2.2 Numerical Example .....	13
2.3 Model Evaluation .....	14
2.3.1 Performance Statistics .....	15
2.4 Results .....	15
2.4.1 Daily Results .....	15
2.4.2 Monthly Results .....	19
2.4.3 Comparison with Methodology of Prairie et al., (2007).....	22
2.4.4 Lag Structure and Additional Daily Results .....	26
2.5 Summary and Discussion .....	31
CHAPTER 3: A WAVELET AUTO-REGRESSIVE METHOD (WARM) FOR MULTI-SITE STREAMFLOW OF DATA WITH NON-STATIONARY SPECTRA .....	33
3.1 Introduction.....	33
3.2 Methodology.....	35

3.2.1 Brief Introduction to WARM.....	35
3.2.2 WARM Drawbacks .....	39
3.2.3 Proposed Enhancement to WARM .....	40
3.2.4 Multisite Streamflow Simulation .....	43
3.3 Results from Application to Colorado River Basin .....	44
3.3.1 Data and Study Area.....	44
3.3.2 Simulations at Lee’s Ferry, AZ.....	46
3.3.3 Multi-Site Simulations.....	50
3.4 Summary and Discussion .....	53
CHAPTER 4: COLORADO RIVER BASIN HYDRO-CLIMATIC VARIABILITY ....	57
4.1 Introduction.....	57
4.2 Data Sets .....	59
4.2.1 Colorado River Natural Streamflow.....	59
4.2.2 Climate Index Data.....	62
4.2.3 Upper Colorado River Basin Precipitation and Temperature Data.....	62
4.2.4 Global Sea Surface Temperatures .....	62
4.2.5 Paleo Reconstructed Data .....	63
4.3 Methods .....	63
4.3.1 Wavelet Spectral Analysis .....	64
4.4 Results .....	65
4.4.1 Dominant Modes of Variability .....	65
4.4.2 Links to Large Scale Climate Forcings: Low Frequency Mode .....	72
4.4.3 Links to Large Scale Climate Forcings: Decadal Mode.....	74

4.4.4 Temporal Variability over Paleo Time Period.....	78
4.4.5 Additional Climate Model Results.....	82
4.5 Summary.....	84
CHAPTER 5: AN APPROACH TO MULTI-DECADAL MEAN STREAMFLOW	
PROJECTION USING PALEO RECONSTRUCTED STREAMFLOW .....	86
5.1 Introduction.....	86
5.2 Data Set.....	88
5.2.1 Paleo Reconstructed Streamflow Data .....	88
5.3 Proposed Method .....	90
5.3.1 Introduction.....	90
5.3.2 Regression Model for Mean Flow Projections .....	92
5.3.3 Wavelet Based Resampling for Variance Projections.....	95
5.3.4 Combination of Mean and Variance Projections .....	99
5.4 Projection Results.....	101
5.5 Summary.....	109
CHAPTER 6: APPLICATION OF DECADAL PROJECTIONS TO MANAGEMENT OF THE	
COLORADO RIVER.....	111
6.1 Introduction.....	111
6.2 Colorado River System Overview .....	113
6.3 Adaptive Policies .....	116
6.4 Colorado River Basin Model .....	120
6.5 Model Input Data .....	122
6.6 Results and Discussion .....	127

CHAPTER 7: SUMMARY .....	144
7.1 Review of Work Presented and Considerations.....	144
7.1.1 Chapter 2.....	144
7.1.2 Chapter 3.....	145
7.1.3 Chapter 4.....	145
7.1.4 Chapter 5.....	146
7.1.5 Chapter 6.....	146
7.2 Future Direction of Research.....	147
7.3 Perspective on Colorado River Flow Variability .....	148
REFERENCES .....	150

## Tables

Table 1 Colorado River system water balance [ <i>Rajagopalan et al.</i> , 2009]. Flow values are based on the natural flow data set for the period of 1906-2006. Demands are present day estimates. Losses are average values based on operational data and modeling results also for the period 1906-2006[ <i>United States Department of the Interior</i> , 2005; , 2008].....	115
--	-----

## Figures

Figure 1 Colorado River Basin; green and pink dots are natural flow nodes in the upper and lower basins respectively.....	2
Figure 2 Simplified basin schematic with index gauge (IG), San Juan River near Carracas, CO (1), San Juan River near Pagosa Springs, CO (2) and Navajo River near Chromo, CO (3). ....	14
Figure 3 San Juan near Carracas, CO daily statistics; triangles represent values from observed data. ....	16
Figure 4 Daily lag-1 correlation by month for (a) San Juan River near Carracas, CO, (b) San Juan River near Pagosa Springs, CO and (c) Navajo River near Chromo, CO. Triangles represent values from observed data. ....	17
Figure 5 San Juan River near Carracas, CO, daily May flow PDF. Solid line is observed data, boxplots are simulated flow.....	18
Figure 6 Daily September flow PDF for Navajo River near Chromo, CO. Solid line is observed data, boxplots are simulated flow. ....	19
Figure 7 San Juan River near Carracas monthly statistics; triangles represent values from observed data. ....	20
Figure 8 Monthly cross correlations for San Juan near Pagosa Springs, CO. Triangles represent values from observed data. Labels refer to months for which the cross correlation is computed (e.g. 1-6 is the cross correlation between January and June).....	21
Figure 9 Monthly correlation between San Juan River near Carracas, CO and Navajo River near Chromo, CO. Triangles represent values from observed data.....	21
Figure 10 January Flow PDF for San Juan River near Carracas, CO. Solid line is observed data, boxplots are simulated flow.....	22

Figure 11 Colorado River at Lee's Ferry, AZ Distributional Statistics (12 months and annual) [based on proportional disaggregation method]. Triangles represent values from observed data. ....	23
Figure 12 Colorado River at Lee's Ferry, AZ Distributional Statistics (12 months and annual) [based on Prairie et al., 2007 method]. Triangles represent values from observed data.....	24
Figure 13 June Flow PDF for the San Juan River near Bluff, UT [based on proportional disaggregation method]. Solid line is observed data, boxplots are simulated flow.....	25
Figure 14 June Flow PDF for the San Juan River near Bluff, UT [based on Prairie et al., 2007 method]. Solid line is observed data, boxplots are simulated flow. ....	25
Figure 15 Navajo River near Chromo, CO average daily autocorrelation function for dry season (Jan-Mar) (left) and wet season (May-Jul) (right). Grey region is range for disaggregated data (averaged over each simulation) and solid line is from observed data (averaged over period of record). ....	26
Figure 16 San Juan River near Caraccas, CO daily statistics for the period of May 30 to June 12. Triangles represent values from observed data.....	27
Figure 17 Navajo River near Chromo, CO daily statistics for the period of May 30 to June 12. Triangles represent values from observed data.....	28
Figure 18 Daily flow ranges (solid grey) based on observed data with median values shown as dashed line for (a) San Juan River near Caraccas, CO, (b) San Juan River near Pagosa Springs, CO and (c) Navajo River near Chromo, CO.....	30
Figure 19 Global wavelet spectrum for synthetic flow data (black line) and the 5th to 95th percentiles of the global wavelet spectra from 1000 traces simulated with an Auto Regressive (AR) model. ....	34

Figure 20 Example diagram of WARM simulation procedure. The original data is decomposed via continuous wavelet transform using the Morlet wavelet and components are identified based on peaks in the global wavelet spectrum. Auto-Regressive models are fitted to each component as well as the residual noise term and each simulation is produced by combining the modeled component and noise signals. ....36

Figure 21 Wavelet spectrum (left) and global wavelet spectrum (right) for synthetic flow data. The grey line is the 95% significance white noise spectrum.....38

Figure 22 Average wavelet spectrum (left) and average global wavelet spectrum (right) from 1000 WARM simulations for synthetic flow data. The grey line is the 95% significance white noise spectrum. ....39

Figure 23 (a) 25 year component of synthetic flow data (b) scale averaged wavelet power (SAWP) for 25 year component of synthetic flow data. ....41

Figure 24 Schematic of improved WARM simulation procedure, shown for a single, non-stationary component. ....42

Figure 25 Colorado River Basin map with gauging locations for disaggregation annotated; Green River at Green River, UT, Colorado River near Cisco, UT, San Juan River near Bluff, UT and Colorado River at Lee’s Ferry, AZ. ....45

Figure 26 Wavelet spectrum (left) and global wavelet spectrum (right) for Lee’s Ferry, AZ natural flow data. The grey line is the 95% significance white noise spectrum.....46

Figure 27 Re-constructed components (a) low frequency signal, (b) decadal signal and (c) noise .....47

Figure 28 Average wavelet spectrum (left) and average global wavelet spectrum (right) from 1000 improved WARM simulations for natural flow at Lee’s Ferry, AZ. The grey line is the 95% significance white noise spectrum. ....48

Figure 29 Improved WARM Lee’s Ferry flow PDFs shown as boxplots with historical flow PDF as solid black line. ....49

Figure 30 Boxplots of mean, variance, skew, max and min values from 1000 improved WARM simulations for Lee’s Ferry, AZ natural flow. The box of the boxplot represents the interquartile range and whiskers extend from the 5<sup>th</sup> to 95<sup>th</sup> percentiles. Black triangles are values from observed data. ....49

Figure 31 Global wavelet spectra for (a) Green River at Green River, UT, (b) Colorado River near Cisco, UT and (c) San Juan River near Bluff, UT. Black line is historical spectrum and grey region is 5<sup>th</sup> to 95<sup>th</sup> percentile based on spectra from disaggregation of 1000 improved WARM traces. ....51

Figure 32 Disaggregated, improved WARM simulation flow PDFs for (a) Green River at Green River, UT, (b) Colorado River near Cisco, UT and (c) San Juan River near Bluff, UT. The black line is the observed PDF and the boxplots show the range of simulation PDFs. ....52

Figure 33 Statistics for disaggregated improved WARM simulations for (a) Green River at Green River, UT, (b) Colorado River near Cisco, UT and (c) San Juan River near Bluff, UT. Measures are mean, variance, skew, maximum and minimum (left to right). The box of the boxplot represents the interquartile range and whiskers extend from the 5<sup>th</sup> to 95<sup>th</sup> percentiles. Black triangles are values from observed data. ....53

Figure 34 Colorado River Basin. The tan region is the Upper Basin and the pink region is the Lower Basin. Green dots represent natural flow nodes in the Upper Basin, and purple dots indicate natural flow nodes in the Lower Basin. ....61

Figure 35 Local wavelet power spectrum (left) and global wavelet spectrum (right) of Lees Ferry flow. Features significant against a 90% white noise background spectrum are encircled in black for the local spectra. In the global spectrum, the grey line is a 90% confidence level against a white noise background spectrum. ....66

Figure 36 Local wavelet power spectrum (left) and global wavelet spectrum (right) of Upper Colorado River Basin PRISM annual water year precipitation data. Features significant against a 90% white noise background spectrum are encircled in black for the local spectra. In the global spectrum, the grey line is a 90% confidence level against a white noise background spectrum. ....67

Figure 37 Local wavelet power spectrum (left) and global wavelet spectrum (right) of Upper Colorado River Basin PRISM annual water year temperature data. Features significant against a 90% white noise background spectrum are encircled in black for the local spectra. In the global spectrum, the grey line is a 90% confidence level against a white noise background spectrum. ....68

Figure 38 Lees Ferry annual water year streamflow data (solid black line) with Upper Colorado River Basin annual water year PRISM precipitation data.....69

Figure 39 Lees Ferry annual water year streamflow data (solid black line) with Upper Colorado River Basin annual water year residual temperature anomalies. ....69

Figure 40 (Left) Flow vs. precipitation scatter plot with best-fit linear regression. (Right) Flow vs. residual temperature anomalies scatter plot with best-fit linear regression. ....70

Figure 41 Upper Colorado River Basin run-off efficiency time series.....	71
Figure 42 Run-off efficiency vs. precipitation scatter plot. ....	71
Figure 43 Run-off efficiency vs. residual temperature anomalies scatter plot.....	72
Figure 44 Upper Colorado River Basin low pass filtered residual temperature data (1906 – 2006) correlated with Kaplan global sea surface temperature anomalies. Colored regions are significant at the 90% confidence level (correlation > 0.16).....	73
Figure 45 Scaled Atlantic Multi-decadal Oscillation index (vertical red bars) with scaled low frequency band pass filtered Upper Colorado River residual temperature data (solid black line).....	74
Figure 46 Upper Colorado River Basin high pass filtered annual precipitation data (1906-2006) correlated with global Kaplan sea surface temperature data. Colored regions are significant at the 90% confidence level (correlation > 0.16).....	75
Figure 47 Upper Colorado River Basin high pass filtered annual precipitation data (1970-2006) correlated with global Kaplan sea surface temperature data. Colored regions are significant at the 90% confidence level (correlation > 0.26).....	76
Figure 48 Upper Colorado River Basin high pass filtered annual precipitation data (1906-1969) correlated with global Kaplan sea surface temperature data. Colored regions are significant at the 90% confidence level (correlation > 0.2).....	77
Figure 49 Local wavelet power spectrum for Woodhouse et al., 2006 Lees Ferry paleo reconstructed streamflow. Features significant against a white noise background spectrum at the 90% level are encircled in black. ....	79
Figure 50 Scaled, reconstructed AMO index from Gray et al., 2005 (black vertical bars) and scaled low pass filtered Lees Ferry streamflow data from Woodhouse et al., 2006 (red line)..	80

Figure 51 (a) Decadal (8-16 year) band Scale-Averaged Wavelet Power (SAWP) from the paleo reconstructed Lees Ferry streamflow. (b) 25 year variance moving window of paleo reconstructed Lees Ferry streamflow. ....	81
Figure 52 Global wavelet spectrum of the decadal SAWP. ....	82
Figure 53 Annual temperature anomalies ( $^{\circ}\text{C}$ ) (color bar) and 200mb geo-potential height anomalies (contours) from differencing results of five GCMs forced first with a warm Atlantic/neutral Pacific and then a cool Atlantic/neutral Pacific. ....	83
Figure 54 Annual Colorado River natural flow above Imperial Dam, AZ, (red line) and annual Basin consumptive use (blue line). ....	87
Figure 55 Annual Colorado River streamflow at Lees Ferry, AZ from the Woodhouse et al. (2006) paleo reconstruction (solid black line) and Bureau of Reclamation's natural flow data set (blue dashed line). ....	89
Figure 56 Idealized framework for adapting the Wavelet Auto Regressive Method for making flow projections. Dashed lines represent data from the historical signal used to initialize the AR modeling. ....	91
Figure 57 Lees Ferry 20 year mean flow plotted against preceding 20 year mean flow window for the paleo reconstructed flow data set. ....	93
Figure 58 R-Squared vs. preceding window length based on linear model to project future 20 year mean flow. ....	94
Figure 59 Lees Ferry 20 year mean flow plotted against preceding 53 year mean flow window for the paleo reconstructed flow data set with best fit line. ....	95
Figure 60 Flow variance projections for 20 year future windows for the Colorado River at Lees Ferry, AZ. ....	99

Figure 61 Normal Q-Q plot of regression residuals.....	100
Figure 62 Ensembles of 20 year mean projections for Colorado River flow at Lees Ferry, AZ, for (a) the projection method presented in this work and (b) a climatological projection approach. The solid red line is the actual value for each projection. ....	103
Figure 63 Boxplots of RPSS values for flow projections for the period 1939 to 1963 using (a) the projection method presented in this work and (b) a climatological projection approach. ....	104
Figure 64 Ensembles of 20 year mean projections for Colorado River flow at Lees Ferry, AZ, for (a) the projection method presented in this work and (b) a climatological projection approach. The solid red line is the actual value for each projection. ....	105
Figure 65 Boxplots of RPSS values for flow projections for the period 1779 to 1803 using (a) the projection method presented in this work and (b) a climatological projection approach. ....	106
Figure 66 Ensembles of 20 year mean projections for Colorado River flow at Lees Ferry, AZ, for (a) the projection method presented in this work and (b) a climatological projection approach. The solid red line is the actual value for each projection. ....	107
Figure 67 Boxplots of RPSS values for flow projections for the period 1584 to 1608 using (a) the projection method presented in this work and (b) a climatological projection approach. ....	108
Figure 68 Twenty year mean projections for Colorado River flow at Lees Ferry, AZ.....	109
Figure 69 United States population change map. (source: <a href="http://www.CensusScope.org">www.CensusScope.org</a> ).....	112
Figure 70 Colorado River natural flow (acre-ft) at Imperial Dam, AZ, which represents total Colorado River yield (1906-2006).....	114
Figure 71 Colorado River Basin. Nodes shown here are also the flow data input points in CRSS. .....	122

Figure 72 Boxplots of mean, variance, skew, max and min values for annual natural flow at Lees Ferry, AZ, from traces comprising the three input hydrology data sets. “I” denotes ISM, “P” denotes paleo and “C” denotes climate model hydrology data. ....	126
Figure 73 Boxplots representing the Lees Ferry, AZ, natural flow PDFs based on traces from the three input hydrology data sets. ....	127
Figure 74 Projected Colorado River Basin system demand (black), computed as sum of projected state depletions [ <i>United States Department of the Interior, 2007b</i> ], reservoir evaporation and losses below Hoover Dam. Also shown are Colorado River Basin average yields for the three input hydrology data sets, all estimated as annual natural flow at Imperial Dam. ....	128
Figure 75 Average annual combined storage for Lake Powell and Lake Mead based on ISM input data. ....	129
Figure 76 Average annual combined storage for Lake Powell and Lake Mead based on paleo input data. ....	130
Figure 77 Average annual combined storage for Lake Powell and Lake Mead based on climate model input data. ....	130
Figure 78 Average annual groundwater accumulation for paleo input hydrology, were it not applied to mitigate shortage conditions. ....	131
Figure 79 Average annual groundwater accumulation for climate model input hydrology, were it not applied to mitigate shortage conditions. ....	131
Figure 80 Average annual combined evaporation from Lake Powell and Lake Mead based on ISM input data. ....	132
Figure 81 Average annual combined evaporation for Lake Powell and Lake Mead based on paleo input data. ....	133

Figure 82 Average annual combined evaporation for Lake Powell and Lake Mead based on climate model input data. .... 133

Figure 83 Probability of Lake Mead pool elevation below 1000' in the month of July based on ISM input data..... 134

Figure 84 Probability of Lake Mead pool elevation below 1000' in the month of July based on paleo input data. .... 135

Figure 85 Probability of Lake Mead pool elevation below 1000' in the month of July based on climate model input data. .... 135

Figure 86 Probability in the Lower Basin of any shortage (upper left), tier 1 shortage (upper right), tier 2 shortage (lower left) and tier 3 shortage (lower right), based on ISM input data. .... 137

Figure 87 Probability in the Lower Basin of any shortage (upper left), tier 1 shortage (upper right), tier 2 shortage (lower left) and tier 3 shortage (lower right), based on paleo input data. .... 138

Figure 88 Probability in the Lower Basin of any shortage (upper left), tier 1 shortage (upper right), tier 2 shortage (lower left) and tier 3 shortage (lower right), based on climate model input data. .... 139

Figure 89 Average annual Lower Basin shortage magnitude based on ISM input data. .... 140

Figure 90 Average annual Lower Basin shortage magnitudes based on paleo input data. .... 140

Figure 91 Average annual Lower Basin shortage magnitudes based on climate model input data. .... 141

Figure 92 Average local and global wavelet power spectra of the 112 climate model hydrology traces for the period of 1950 to 2099. The grey line is a 90% confidence background white noise power spectrum..... 142

# **CHAPTER 1: INTRODUCTION**

## **1.1 Background**

The Colorado River is the lifeblood of the Southwestern United States. It stretches from the Rocky Mountains to the Gulf of California, traveling over 1000 miles, with a watershed that includes seven states and part of Mexico (Figure 1). Colorado River Basin water is relied on by approximately 30 million people and irrigates 3.5 million acres of farmland. The system is unique in that the reservoir storage capacities total roughly four times the mean annual flow of the river. Few systems in the world have such large storage capacities, which afford the ability to reliably provide water and “weather” extended dry periods. Even with this buffering capacity, the river is, at present, operating “on a knife’s edge” due to an over-allocation of water during the abnormally wet early 1900’s coupled with a severe and sustained dry period dating back to 2000. Furthermore, the compacts governing water apportionment allow for the continued growth of water use. As such, there is considerable concern regarding the reliability of water supply in the Colorado River Basin considering the impacts of natural variability, climate change and increasing demands.

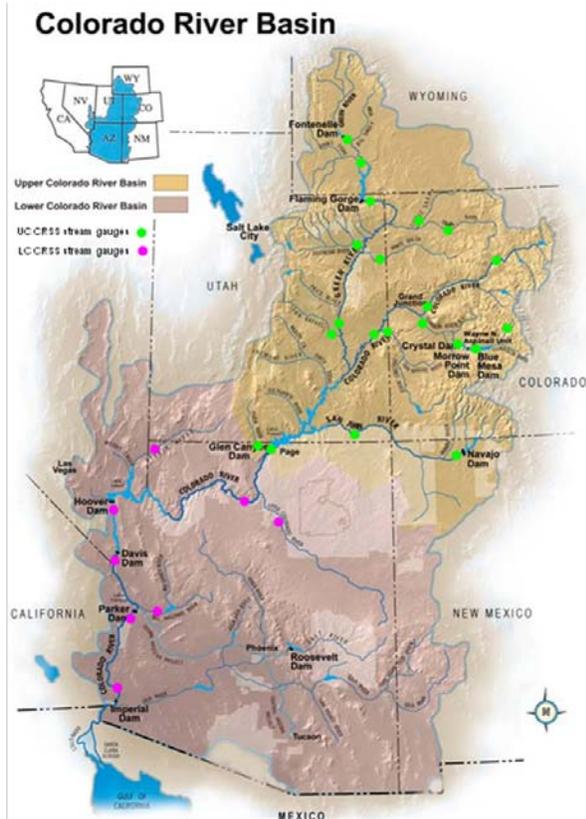


Figure 1 Colorado River Basin; green and pink dots are natural flow nodes in the upper and lower basins respectively.

## 1.2 Motivation

Conventional water resource planning and management are based upon the assumption that past run-off records are indicative of future hydrologic conditions. The severe and sustained nature of the recent drought in the southwestern United States has underscored the limitation of this planning approach [Milly *et al.*, 2008]. Furthermore, a growing collection of scientific literature indicates that anthropogenic climate change will further dry the region and strain its water resources [Christensen and Lettenmaier, 2007; Nash and Gleick, 1991; Ray *et al.*, 2008; Seager *et al.*, 2007]. While the impact of climate change is indeed relevant, consideration of natural variability is equally, if not more important over the next 20 years. For example, in the Colorado Basin, 10 year moving window flow means vary from 82% to 126% of the full record

mean. Such variations are likely to trump a slow flow reduction trend associated with anthropogenic climate change with regard to water supply reliability/strain. Thus, developing tools and strategies to address streamflow variability, in the context of historical and paleo reconstructed data, is critical for effective water management in the Colorado River Basin.

There is increasing evidence that the interannual (year to year) and longer scale variations in the western U.S. hydro-climate are driven by large scale climate features such as the El Niño Southern Oscillation (ENSO), Pacific Decadal Oscillation (PDO), Atlantic Multi-decadal Oscillation (AMO), etc. [Hunter *et al.*, 2006; McCabe and Dettinger, 1999; Piechota *et al.*, 1997; Tootle *et al.*, 2005]. Presently, stochastic simulation methods for the purpose of planning and water management aim to capture the distributional statistics (mean, variance, skew, etc.) of the observed record, but do not explicitly reflect the links between large scale, persistent climate features and streamflow. Furthermore, studies pertaining to simulation/projection of future hydrologic conditions often focus on the inter-annual time scale (1-2 years) or at the climate change scale (50-100 years), leaving a substantial gap in the decadal to inter-decadal time scales. However, much of water management and planning focuses on meeting demands and reducing risk over decadal time scales. Thus, there is a clear need for tools that can generate stochastic traces and hydrologic projections at this time scale.

Complimentary to this is the development of robust management policies and strategies that benefit from these methods. The recently signed Colorado River Shortage Environmental Impact Statement (EIS) established policies for distribution of water between the two major reservoirs on the system (Lakes Powell and Mead) and contains provisions for delivery curtailment in times of drought [United States Department of the Interior, 2007b]. These policies are a positive step toward developing adaptive management of Colorado River Basin

water resources. In addition, it is likely that given even modest insight regarding decadal scale future hydrologic conditions, policies could be crafted to benefit from this information, thus further improving overall system reliability/yield.

### **1.3 Outline**

In chapter 2, a novel non-parametric method for space-time disaggregation of streamflow is presented. Streamflow disaggregation is an important water planning and management tool. Often times, reservoir system modeling requires multiple inputs at specific timescales that otherwise may not have been available. This approach is computationally simple compared with both parametric and non-parametric alternatives. Furthermore, it is applicable to a variety of timescales, including daily, and is effective for multi-site disaggregation. The method is first demonstrated for multi-site daily disaggregation in the San Juan River Basin and later compared with the Prairie et al. [2007] method for generating monthly flow data in the Colorado River Basin. This work is published in Water Resources Research [Nowak et al., 2010].

Chapter 3 presents a stochastic streamflow simulation framework that provides the capacity to capture non-stationary spectral features of historical streamflow data in addition to traditional measures such as mean, variance, etc. This approach is based on the Wavelet Auto Regressive Method (WARM), originally proposed by Kwon et al. [2007]. As discussed earlier, there is considerable evidence of quasi-periodic climate phenomena (e.g. ENSO, PDO, etc.) associated with hydroclimatic variability in the Colorado River Basin. This spectral domain simulation approach captures these types of influences in the traces it generates. Thus, it is noteworthy that Colorado River streamflow exhibits several distinct spectral features that are not reflected in simulations from other stochastic methods. Additionally, when this approach is

coupled with the disaggregation presented in chapter 2, upstream spectral features are also reproduced without additional conditioning beyond the standard method.

Modeling spectral features is extremely useful in the realm of stochastic simulation, particularly when attempting to realistically reproduce variability structure for water management applications. Equally important is understanding the sources of the variance being modeled. In chapter 4, a systematic investigation into the sources of the major spectral features found in Colorado River Basin hydroclimatic data is presented. Physical processes are first identified for significant variability scales. Following, links with climatic phenomena are proposed. Additional perspective is offered in closing through the analysis of paleo reconstructed data and idealized climate modeling experiments.

Understanding historical variability from the simulation and attribution perspectives is critical to the consideration of possible future hydrologic conditions. Chapter 5 presents a method for making decadal scale Colorado River streamflow projections which utilizes the wealth of information contained within the paleo reconstructed streamflow record. The approach cues off previously identified key variability structure to produce a projection of the average hydrologic state over a future horizon of 20 years. It is demonstrated that this method has considerable skill over climatology by making projections for periods of the paleo and historical streamflow data.

In Chapter 6, the Colorado River Basin long-term planning model, Colorado River Simulation System (CRSS) is utilized to demonstrate the utility of decadal regime projections. In this, a baseline system is established for reliability and losses. Based on decadal projection information, regime specific policies are developed and employed. Results indicate that supply reliability is increased and system losses are reduced relative to the baseline.

Chapter 7 provides an overall summary of the work, addressing the methods and results presented. Additionally, a brief discussion of Colorado River system “robustness” is offered in light of the high run-off anticipated for water year 2011. This section concludes with recommendations for future work and improvements.

## CHAPTER 2: A NONPARAMETRIC STOCHASTIC APPROACH FOR MULTISITE DISAGGREGATION OF ANNUAL TO DAILY STREAMFLOW

This chapter has been reproduced by permission of American Geophysical Union [*Nowak et al.*, 2010].

### 2.1 Introduction

As the demand and use of water continues to increase, water management has become a complex task requiring detailed models and planning in order to effectively manage this valuable resource. Seasonal diversions, irrigation needs, instream flow requirements and hydropower all contribute to the challenge of present-day water administration. In order to maximize the utility of models with a high level of physical system detail, flow inputs need to be rich in variety and of an appropriate time step. These data should represent a range of hydrologic scenarios and be temporally consistent with the system obligations (e.g. daily hydropower demands, seasonal irrigation, etc.). Historical flows are not always available such that the data needs of a model are appropriately met. This highlights the utility of synthetic data techniques. Methods for generating synthetic streamflows provide the data needed for operations models. These methods are capable of producing extreme events (drought/surplus) that are greater in magnitude and/or duration compared to those of the observed record, yet still plausible, thus further exploring operations beyond the historical streamflows.

Methods to generate flows usually begin at the annual time-step for a location of particular significance, (e.g., due to interstate compacts, physical location, etc.) in order to ensure statistical properties are effectively preserved at this location. However, as eluded earlier, single-site, annual flow values may not be sufficient for detailed models. Thus, the need to disaggregate flow values to multiple sites and finer time scales arises. The task of appropriately

distributing flow in space and time requires that distributional statistics (i.e. mean, variance, skewness and maximum and minimum values) are reproduced for all sites, while also preserving summability, cross-correlations and continuity. Parametric approaches [Grygier and Stedinger, 1988; Stedinger and Vogel, 1984; Valencia and Schaake, 1973] for addressing disaggregation have traditionally been framed on a linear approach, similar in structure to an Auto-Regressive (AR) model. This approach is represented by Equation 1:

$$\mathbf{X}_t = \mathbf{A}\mathbf{z}_t + \mathbf{B}\mathbf{v}_t$$

**Equation 1**

where  $\mathbf{X}_t$  is the disaggregate variables (e.g., monthly flows) at time  $t$ ,  $\mathbf{z}_t$  is the aggregate variable (e.g., annual flow) and  $\mathbf{v}_t$  is a vector of random values from a normal distribution. The elements of  $\mathbf{X}_t$  sum to  $\mathbf{z}_t$ , also known as the summability criteria, which is an important element for generating space-time streamflows on a river network. Model parameters contained in matrix  $\mathbf{A}$  and  $\mathbf{B}$  are estimated such that the simulations preserve the cross correlation between and among the variables in addition to the summability property.

Main drawbacks to this general class of models are [1] the data are assumed to be normally distributed and [2] the  $\mathbf{A}$  and  $\mathbf{B}$  matrices used can quickly become unwieldy as the dimensions of the disaggregation increase (e.g. disaggregating to daily time scale and/or many spatial locations). Often the data at each time scale and location needs to be transformed to Normality using a log or power transform. However, this does not ensure that summability or the observed statistics are preserved when the results are returned to the original space. Efforts to reduce the computational intensity associated with large  $\mathbf{A}$  and  $\mathbf{B}$  matrices suggest a stepwise disaggregation approach [Santos and Salas, 1992](e.g., annual to seasonal to monthly, etc.) which results in additional challenges of continuity across the generated flows at the finer time

scale. In summary, this class of methods is inadequate for disaggregation finer than the monthly timescale.

Recently, several non-parametric techniques have been put forth that offer improvements, such as the ability to capture non-Gaussian distributions [*Prairie et al., 2007; Tarboton et al., 1998*]. The nonparametric methods cast the problem as simulating from the conditional Probability Density Function (PDF),  $f(\mathbf{X}_t / \mathbf{z}_t)$ . Tarboton et al. employ a kernel density estimation based approach to simulating from the conditional PDF. While this approach is effective at ameliorating the issues associated with non-normal distributions, it remains computationally intensive at even the annual to monthly scale disaggregation. Furthermore, kernel methods suffer from boundary biases which worsen with dimension. Prairie et al. (2007) circumvent this issue by replacing the kernel density estimation with a computationally faster K-Nearest Neighbor (K-NN) based time re-sampling approach [*Lall and Sharma, 1996*]. This retains the ability to preserve non-gaussian distributions while significantly reducing computing demands and also alleviates the boundary issue associated with kernel methods. However, both struggle with regard to basins with sites that vary significantly in flow magnitude and furthermore, these methods have only been shown effective for the disaggregation of annual to monthly flows.

Very few techniques have proven useful for disaggregation to daily values. This is generally due to a mismatch between the parametric distributions appropriate for daily flows versus monthly or annual flows, the high dimension of the disaggregation problem, compounded uncertainty in parameter estimation for multi-stage approaches and the inability to maintain flow continuity across disaggregation time period boundaries. In the realm of precipitation generation, there has been considerably more success in producing daily precipitation values from a wet spell total, which is largely attributed to the lack of a continuity requirement [*Bogardi*

*et al.*, 1993; *Chang et al.*, 1984; *Chin*, 1977; *Lall et al.*, 1996; *Srikanthan and McMahon*, 2001].

Kumar et al. (2000) have put forth an effective approach for disaggregating monthly to daily streamflow by casting the task as a linear programming problem which is then optimized in adherence to a number of constraints that preserve statistical relationships and continuity [*Kumar et al.*, 2000]. However, the method is very computationally intensive; for a 5 site, single month to daily disaggregation, there are more than 1500 decision variables which must be solved for in order to reach a solution. This poses obvious challenges for larger basins with many gauging locations.

Here we propose a non-parametric disaggregation approach that resamples a historical vector of daily proportions, conditioned on an annual flow, via K-NN resampling. The disaggregated values are obtained by projecting the aggregate flows on to the proportion vector. This ensures summability, continuity across the daily time scale and positive disaggregated values unlike all the method discussed above. Furthermore, it has the ability to capture all the distributional and cross-dependency properties. The method can be readily applied to disaggregation at multiple time and space scales. Independently, a similar approach has also been proposed by Lee (2008), which was applied to multi-site, annual to monthly disaggregation. The method employs an adjusting procedure for re-sampled historical data with a genetic algorithm to increase variability [*Lee*, 2008]. To demonstrate the utility and effectiveness of our proposed method, it is used to generate daily flows at three sites on the San Juan River, a tributary of the Colorado River located in southwestern Colorado. Additionally, a comparison with the results of *Prairie et al.* (2007) for locations in the Colorado River Basin is provided.

## 2.2 Annual to Daily Disaggregation Method

Non-parametric space-time streamflow disaggregation can be thought of as simulating from the conditional probability distribution function (PDF)  $f(\mathbf{X}_t / \mathbf{z}_t)$  where  $\mathbf{z}_t$  is a vector of annual aggregate flows to be disaggregated and  $\mathbf{X}_t$  is a matrix of flows that sum to  $\mathbf{z}_t$ .

Here we cast the problem as a conditional simulation of the proportion vector from the conditional PDF  $f(\mathbf{X}_t / \mathbf{p}_t)$ , where  $\mathbf{p}_t$  is the vector of daily proportions, whose elements sum to unity by definition, and  $z$  is the aggregate flow. The simulation uses a K-NN resampling approach [Lall and Sharma, 1996; Prairie et al., 2007] in that, K-nearest neighbors to  $z$  are identified from historical flows and one of them is resampled using a weight metric (see equation 2) that gives more weight to those neighbors closest in magnitude to  $z$ . This is akin to constructing the conditional PDF locally in the neighborhood of  $z$  and simulating from it. The daily proportion vector corresponding to the selected neighbor (i.e., one of the historical years) is selected and the aggregate flow  $z$  is projected on to this proportion vector of the resampled historical year to obtain the disaggregated flows. This simple yet robust and local approach provides the ability to capture nonlinear and non-Normal features. Additionally, the method always generates positive values and the summability of the disaggregated flows to the aggregate flow is automatically preserved. The disaggregated values have a rich variety, including flows outside the range of observed values. Daily values at a single location are further disaggregated in space by projecting them on to the appropriate proportion vector. The implementation algorithm is described below followed by an example.

### 2.2.1 Implementation Algorithm

For the purpose of describing the implementation of this technique, a single site, annual to daily disaggregation problem is considered. Extension of the method to a multi-site network is straight forward and will be discussed later.

1. Observed daily streamflow values are converted to a proportion of the total annual flow of that particular year (i.e. daily flows in a given year are divided by the total annual flow of that year), giving a matrix  $\mathbf{P}$ , with dimensions  $n \times 365$  ( $n$ = number of years of observed data). By construction, each row of this matrix sums to unity.

2. Suppose  $Z$  is the annual flow that needs to be disaggregated to daily flows. First,  $K$ -nearest neighbors of  $Z$  are identified from the historical annual streamflow vector ( $\mathbf{z}$ ). The neighbors are computed based on the distance between  $Z$  and all the historical annual streamflow values. The number of nearest neighbors,  $K = \sqrt{n}$ , based on heuristics [Lall and Sharma, 1996] is found to be very effective in a variety of applications. Each of the  $K$ -nearest neighbors is assigned a weight based on the weight function proposed by Lall and Sharma (1996), given as:

$$W(i) = \left( \frac{1}{i} \right) / \left( \sum_{i=1}^K \frac{1}{i} \right)$$

**Equation 2**

where  $K$  is the number of nearest neighbors and  $i$  refers to the “neighbor index,” with  $i=1$  being the closest of the nearest neighbors. Note the weights are normalized so that they sum to unity.

3. One of the  $K$  nearest neighbors (i.e., one of the historical years) is chosen based on a weighted resampling (i.e. probability of picking a given year is determined by the corresponding weight from Equation 2). The proportion vector corresponding to the picked year, say, ( $\mathbf{p}_y$ ) is selected. Next, the annual streamflow ( $Z$ ) is projected on to this proportion vector to obtain the daily streamflow vector ( $\mathbf{d}$ ), which is given by:

$$\mathbf{d} = Z * \mathbf{p}_y$$

**Equation 3**

Since the proportion vector sums to unity, the disaggregated daily streamflows sum to the aggregate annual flow.

4. Steps 2 and 3 are repeated to generate an ensemble of daily streamflows.

To extend the method to both space and time, the process simply gains another dimension. The matrix  $\mathbf{P}$  ( $365 \times n$ ) will become an array of dimension  $365 \times n \times s$ , where  $s$  indicates the number of locations for disaggregation. Thus, the resampled proportion vector ( $\mathbf{p}_y$ ) will now be a matrix of dimension  $365 \times s$ . For multi-site disaggregation, the annual flow vector corresponds to an aggregate location downstream or an index gauge created by addition of streamflows at all the locations.

This approach, as seen above, is simple to implement and can easily be applied to any space and time scales and obviates any need to adjust the post-disaggregation values, unlike methods developed previously. Also, it is equivalent to simulating from the conditional PDF  $f(x / z)$ .

### 2.2.2 Numerical Example

The following provides a simple numerical example of the technique, using limited data and a “4-day year.”

$$\mathbf{P} = \begin{bmatrix} 1967 & .1 & .3 & .4 & .2 \\ 1968 & .15 & .25 & .35 & .25 \\ 1969 & .1 & .2 & .5 & .2 \\ 1970 & .05 & .15 & .65 & .15 \\ 1971 & .2 & .2 & .4 & .2 \\ 1972 & .1 & .2 & .4 & .3 \\ 1973 & .15 & .2 & .4 & .25 \\ 1974 & .05 & .1 & .8 & .05 \\ 1975 & .2 & .2 & .5 & .1 \end{bmatrix} \quad \mathbf{z} = \begin{bmatrix} 1967 & 35 \\ 1968 & 40 \\ 1969 & 33 \\ 1970 & 52 \\ 1971 & 43 \\ 1972 & 56 \\ 1973 & 38 \\ 1974 & 49 \\ 1975 & 32 \end{bmatrix} \quad Z = 45$$

Based on the weighted resampling of the “K” nearest neighbors, say year 1968 ( $y = 1968$ ) is selected to be the nearest neighbor. Next, the simulated annual value ( $Z$ ) is applied to the vector  $\mathbf{p}_{1968}$  to produce the disaggregated values ( $\mathbf{d}$ ).

$$Z * \mathbf{p}_y = \mathbf{d}$$

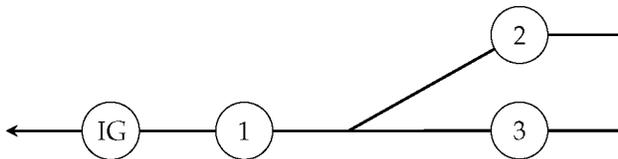
$$45 * [.15 \ .25 \ .35 \ .25] = [6.75 \ 11.25 \ 15.75 \ 11.25]$$

$$\sum_{i=1}^4 d_i = Z$$

From the example, it can be seen that the technique is quite parsimonious and easy to implement. Section 3 discusses the ability of the method to reproduce observed statistics.

### 2.3 Model Evaluation

The proportion based K-NN disaggregation method is evaluated by applying it to daily streamflow simulation at three gauges on the San Juan River. The selected sites are part of the United States Geological Survey (USGS) gauging network in the San Juan Basin, located in southwestern Colorado. Specifically, the gauging stations are San Juan River near Carracas, CO (09346400), San Juan River at Pagosa Springs (09342500) and Navajo River near Chromo, CO (09344400). Data for these sites were obtained from the USGS’s National Water Information System: Web Interface and span from 1972-1995. The gauge near Chromo, CO is no longer maintained, thus truncating the common record at 1995. Daily flow values were converted from cubic feet per second (cfs) to acre-ft/day. Figure 2 shows a schematic of these three gauges, the gauge IG is an index gauge constructed by aggregating the flows at the three locations.



**Figure 2 Simplified basin schematic with index gauge (IG), San Juan River near Caraccas, CO (1), San Juan River near Pagosa Springs, CO (2) and Navajo River near Chromo, CO (3).**

For evaluation, 100 traces each of the 50 years in length for annual flow at the index gauge were generated using the K-NN lag-1 [Lall and Sharma, 1996] approach. Alternatively, a traditional AR(1) model could also have been applied, as the disaggregation method is

independent of the annual simulation method. Thus, generated index gauge annual flows are disaggregated to daily streamflows at all the three locations using the methodology presented above. A suite of statistics are computed at monthly and daily time scales from the simulations and are presented as boxplots along with the corresponding value from the historical data for comparison. The box of the boxplot (see e.g., Figure 3) represents the interquartile range (IQR) with the horizontal line being the median and the whiskers extending to the 5th and 95th percentiles. Performance on a given statistic is generally considered “good” when the historical value falls within the IQR.

### **2.3.1 Performance Statistics**

For the purpose of this work, the term “distributional statistics” refers to mean, variance, coefficient of skew, maximum and minimum values, which are computed at the monthly and daily time scales for comparison. In addition, lag correlation, cross correlation between sites and Probability Density Functions (PDFs) are also used to evaluate the performance of the proposed method.

## **2.4 Results**

We describe the results of the daily statistics followed by the monthly. Then we implemented the proposed method for annual to monthly disaggregation at the three locations and compare them with the methodology proposed by Prairie et al. (2007).

### **2.4.1 Daily Results**

The daily statistics - mean, variance, skew, maximum and minimum values were computed for a single day at the middle of each month and are presented. Boxplots of all the basic distributional statistics for the San Juan River near Carracas, CO (Figure 3) indicate that the method effectively reproduces observed statistics at the daily timescale through all the

months.

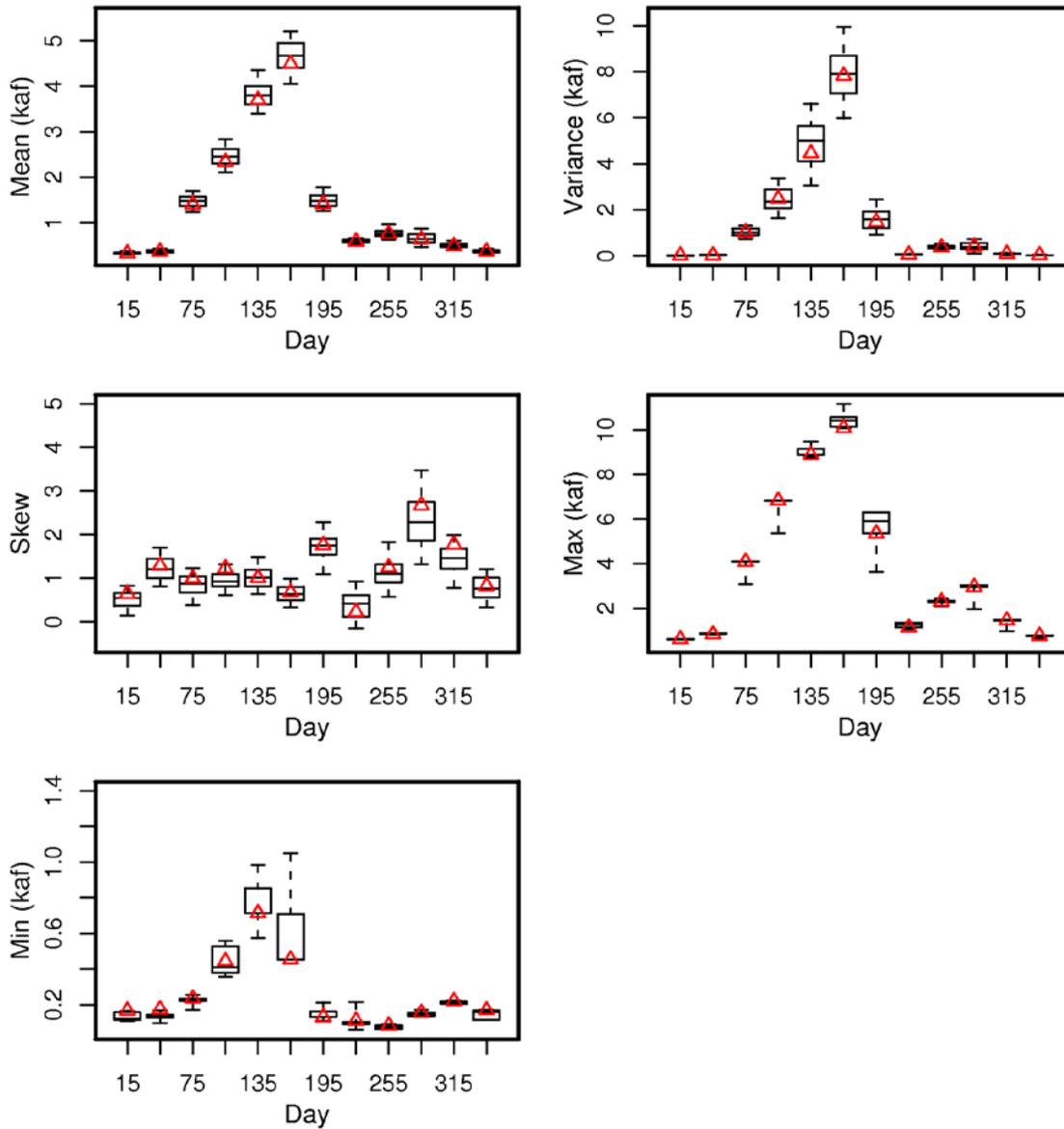
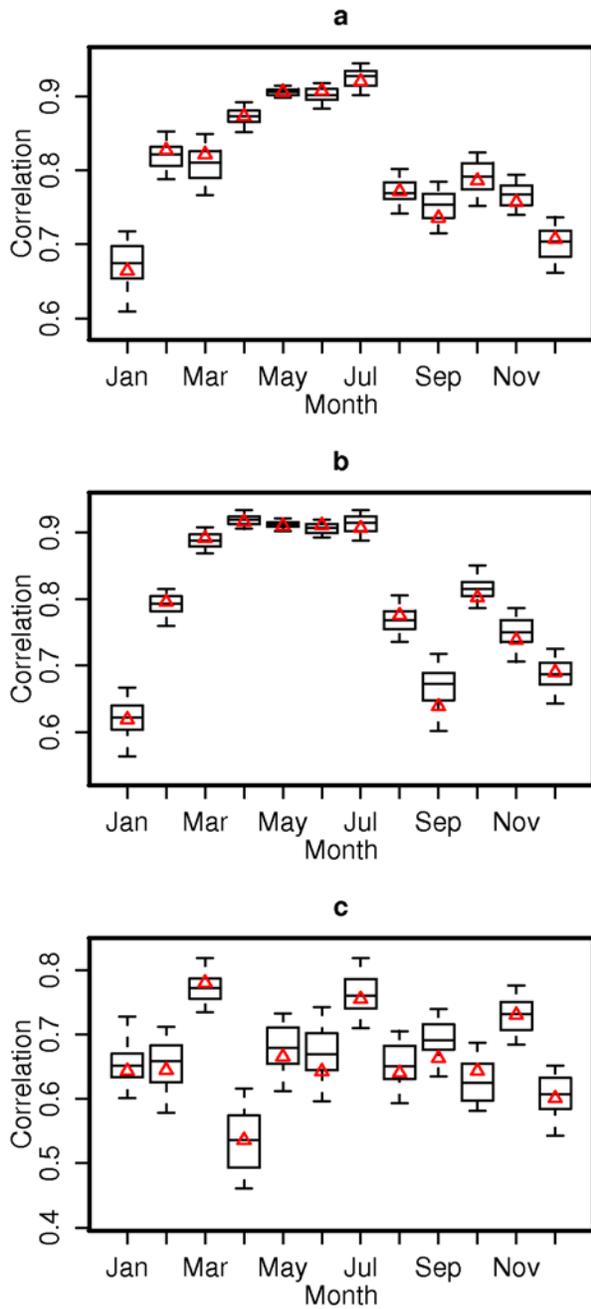


Figure 3 San Juan near Carracas, CO daily statistics; triangles represent values from observed data.

The lag-1 autocorrelation is also well captured (Figure 4) indicating that the simulated flows have realistic continuity. The results were similar at other locations and days (figures not shown).

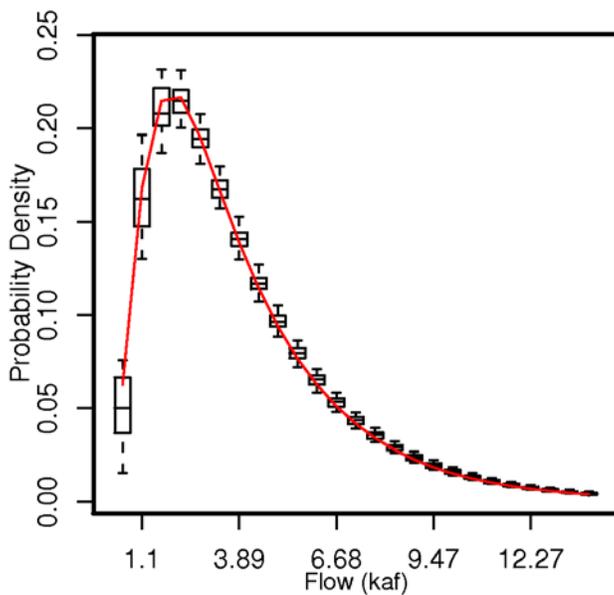


**Figure 4 Daily lag-1 correlation by month for (a) San Juan River near Caraccas, CO, (b) San Juan River near Pagosa Springs, CO and (c) Navajo River near Chromo, CO. Triangles represent values from observed data.**

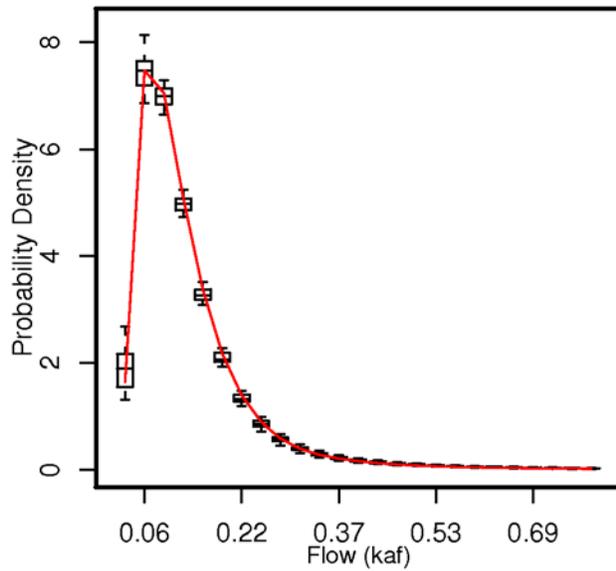
Since the disaggregation is performed for each year separately, as to be expected, the correlation between the flow of the last day of a year and the first day of the following year is not well captured. This can be improved upon by including the streamflow of the last day from the

previous year to the annual flow in computing the nearest neighbors. However, this can lead to deterioration of other statistics. Since the last and first months, December and January, are generally low flow months in this basin, we deem it not essential to explicitly preserve this autocorrelation so long as other key statistics are well preserved, especially during high flow months.

The boxplots of PDFs of daily flows from the simulations and the observed values for selected high flow month of May and low flow month of September are shown in Figure 5 and Figure 6. This provides insights into the ability of the method to capture the entire distribution, arguably of greater importance than point statistics shown in previous figures. It can be seen that the historical PDFs of daily streamflows, which are highly non-Normal, are very well captured by the model at different locations and months.



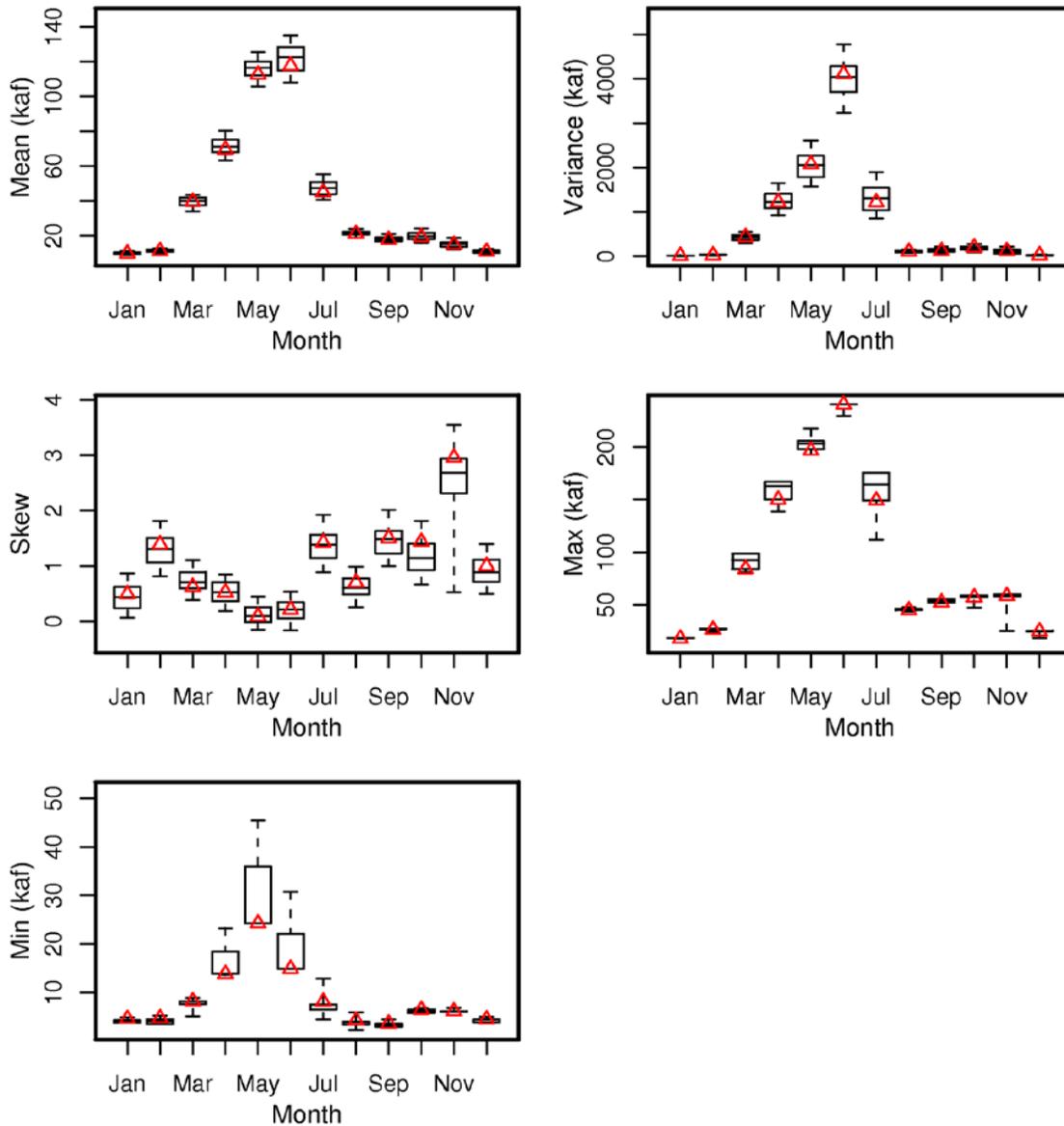
**Figure 5 San Juan River near Carracas, CO, daily May flow PDF. Solid line is observed data, boxplots are simulated flow.**



**Figure 6 Daily September flow PDF for Navajo River near Chromo, CO. Solid line is observed data, boxplots are simulated flow.**

### 2.4.2 Monthly Results

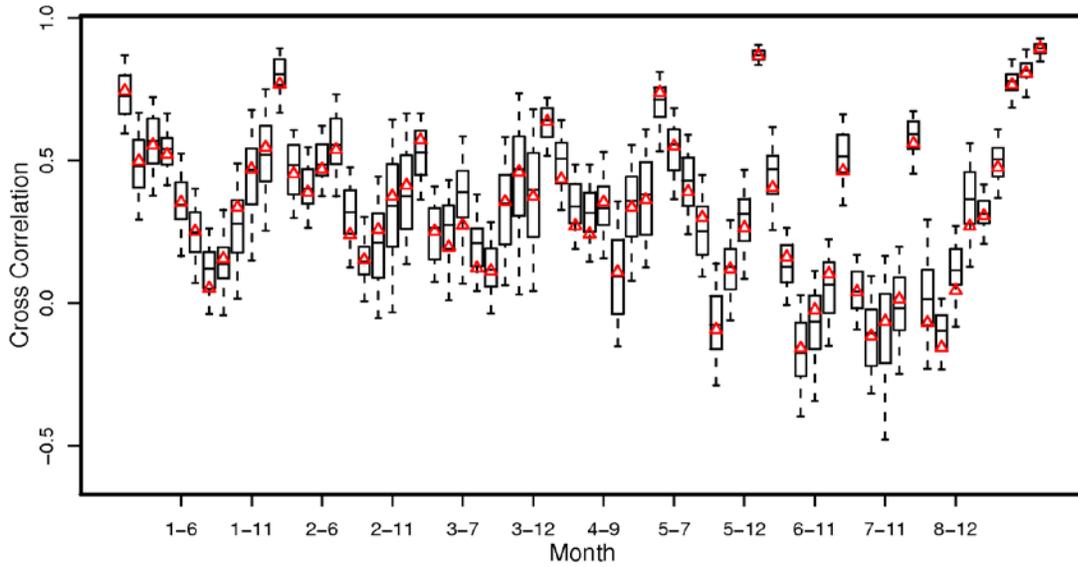
The disaggregated daily streamflows are aggregated to obtain monthly values and the suite of basic statistics is computed. Boxplots of these statistics at the monthly time-step are shown for the San Juan River near Carracas, CO (Figure 7). It can be seen that they are all well preserved. Also, the maximum and minimum values generated are beyond the range of the observed data, indicating that even though a resampling approach is used, it is possible to generate extremes outside the range of the data. Furthermore, all values are positive.



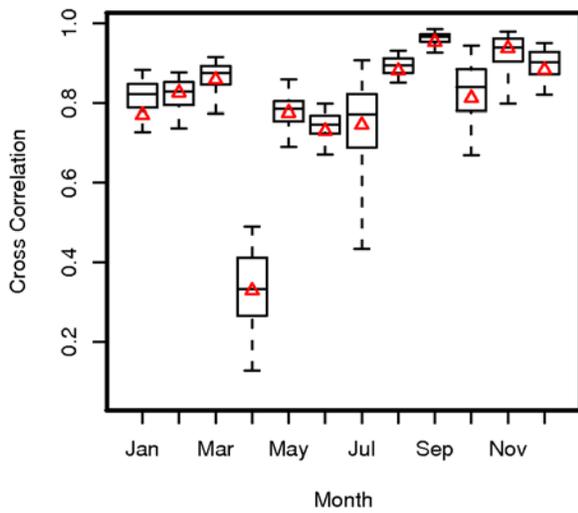
**Figure 7 San Juan River near Carracas monthly statistics; triangles represent values from observed data.**

Monthly cross-correlations at the same location are displayed in Figure 8, which are also well captured. Similar results were seen at the other locations (not shown). Cross correlation of monthly flows between Pagosa Springs and Chromo sites are shown in Figure 9. From the plot, it can be seen that the high correlation between monthly flows is reflected in the disaggregated simulations. The lower correlation in April may be due to spatial variability associated with the

start of snowmelt and high flows. Similar monthly cross-correlation results were seen for other gauge combinations.

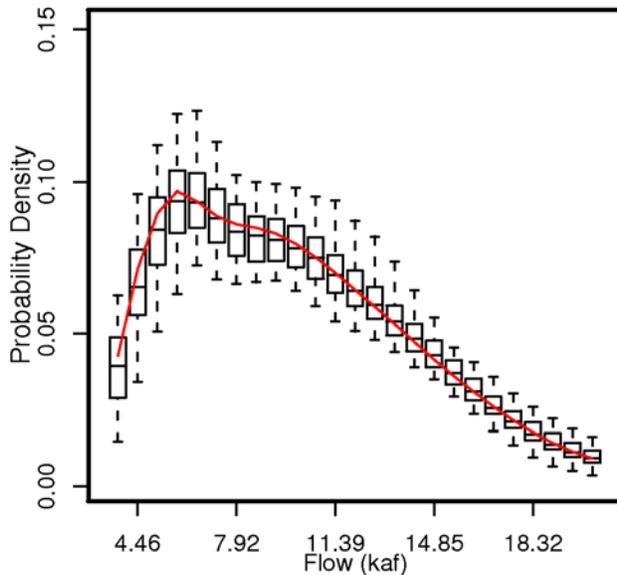


**Figure 8 Monthly cross correlations for San Juan near Pagosa Springs, CO. Triangles represent values from observed data. Labels refer to months for which the cross correlation is computed (e.g. 1-6 is the cross correlation between January and June).**



**Figure 9 Monthly correlation between San Juan River near Carracas, CO and Navajo River near Chromo, CO. Triangles represent values from observed data.**

Additionally, the simulated January flow PDFs for the San Juan near Carracas are presented as boxplots with the observed PDF overlaid (Figure 10). This demonstrates the ability to capture the entire distribution, which is non-normal. These results indicate that the statistics at scales different from the scale of disaggregation are very well reproduced by this methodology.

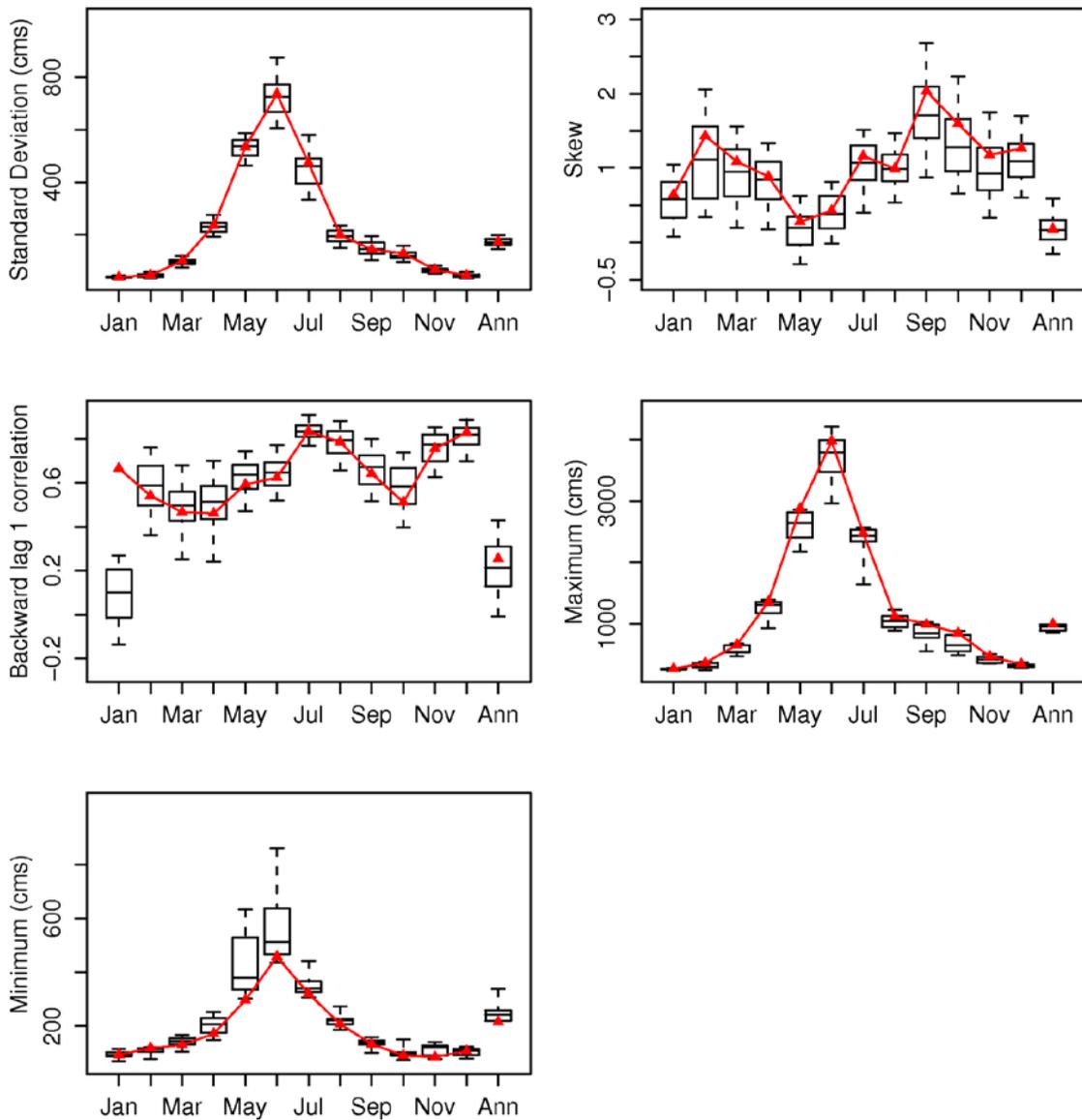


**Figure 10** January Flow PDF for San Juan River near Carracas, CO. Solid line is observed data, boxplots are simulated flow.

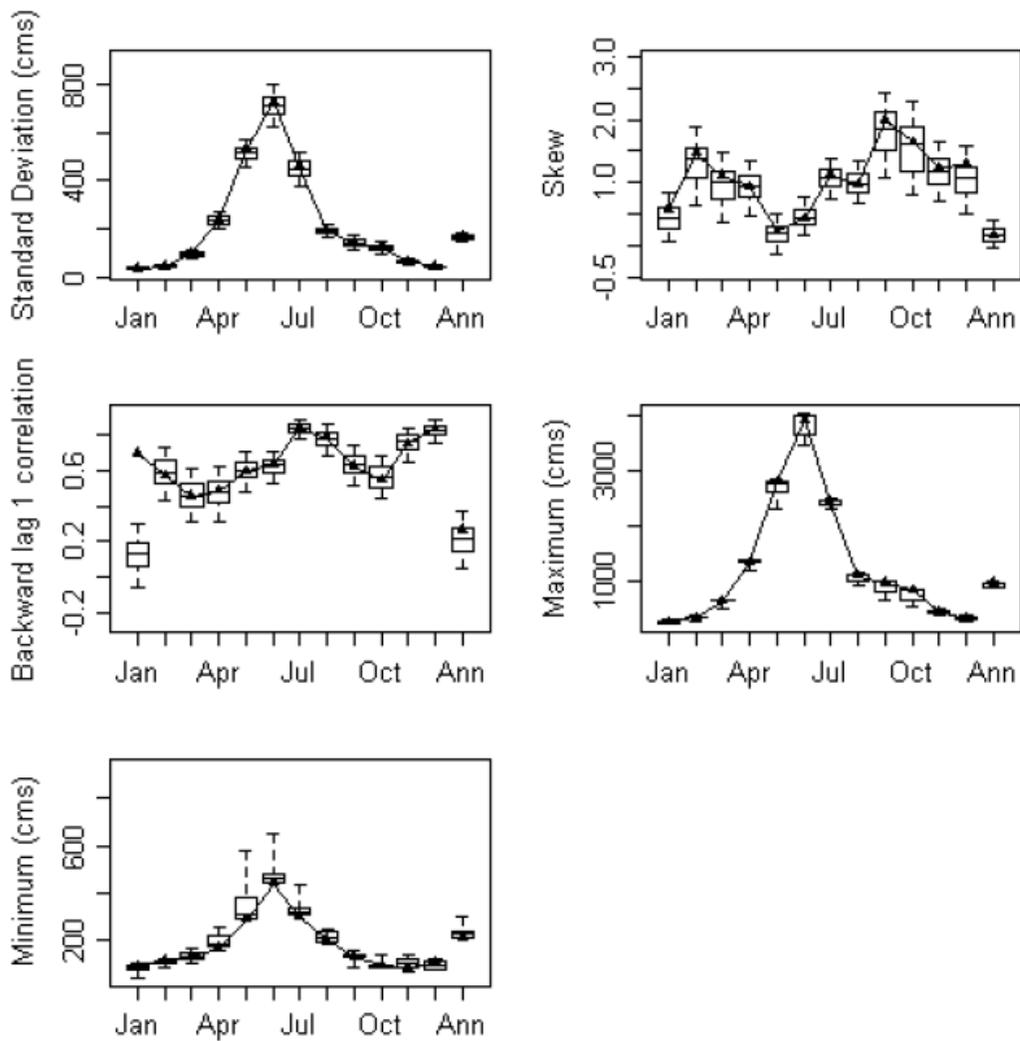
### 2.4.3 Comparison with Methodology of Prairie et al., (2007)

To demonstrate the wide applicability of the proposed technique to disaggregate flows at different temporal locations and basins, a brief comparison with the results of Prairie et al. (2007) on the Colorado River Basin is provided. As mentioned earlier, Prairie et al. (2007) present an effective non-parametric annual to monthly disaggregation which builds upon the earlier work of Tarboton et al. (1998). Four streamflow gauges (Colorado River near Cisco, UT, Green River at Green River, UT, San Juan River near Bluff, UT and Colorado River at Lee's Ferry, AZ see their Figure 1) in the Colorado River Basin were considered in their paper. We apply our methodology to these gauges in the same manner (i.e., annual to monthly

disaggregation) for direct comparison. Figure 11 and Figure 12 show boxplots of monthly and annual basic statistics at the Lees Ferry, AZ gauge, from the methodology proposed here and their work, respectively. It can be seen that both the methods capture the basic statistics very well. However, the proportional disaggregation approach has the ability to produce values beyond the observed data range and is guaranteed to never produce negative flows.

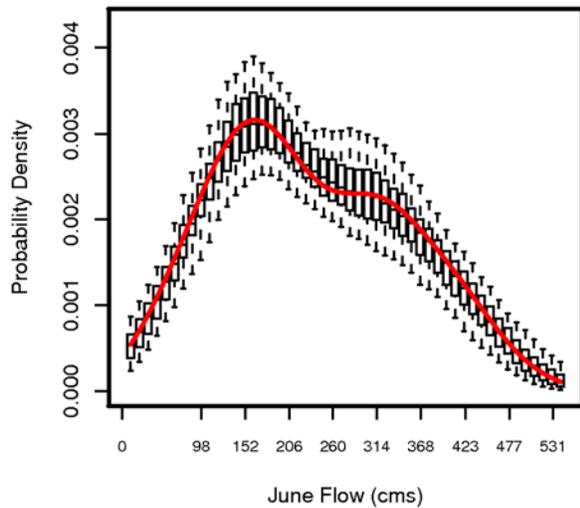


**Figure 11 Colorado River at Lee's Ferry, AZ Distributional Statistics (12 months and annual) [based on proportional disaggregation method]. Triangles represent values from observed data.**

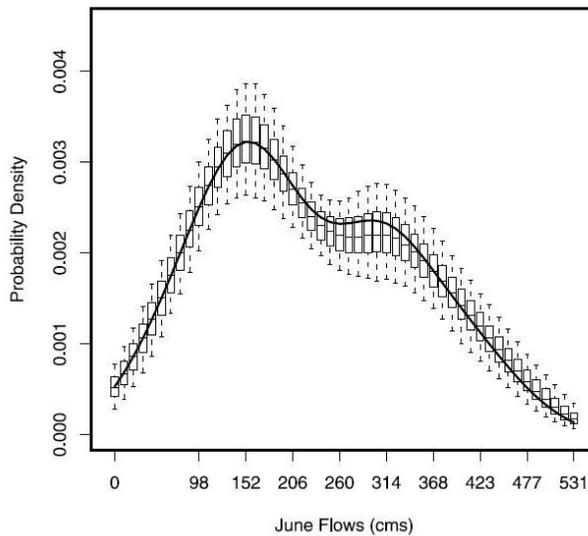


**Figure 12 Colorado River at Lee's Ferry, AZ Distributional Statistics (12 months and annual) [based on Prairie et al., 2007 method]. Triangles represent values from observed data.**

Furthermore, the methods have the ability to capture non-Normal distributions very effectively, as can be seen from Figure 13 and Figure 14, showing the June flow PDF for the San Juan River near Bluff, UT.



**Figure 13 June Flow PDF for the San Juan River near Bluff, UT [based on proportional disaggregation method]. Solid line is observed data, boxplots are simulated flow.**



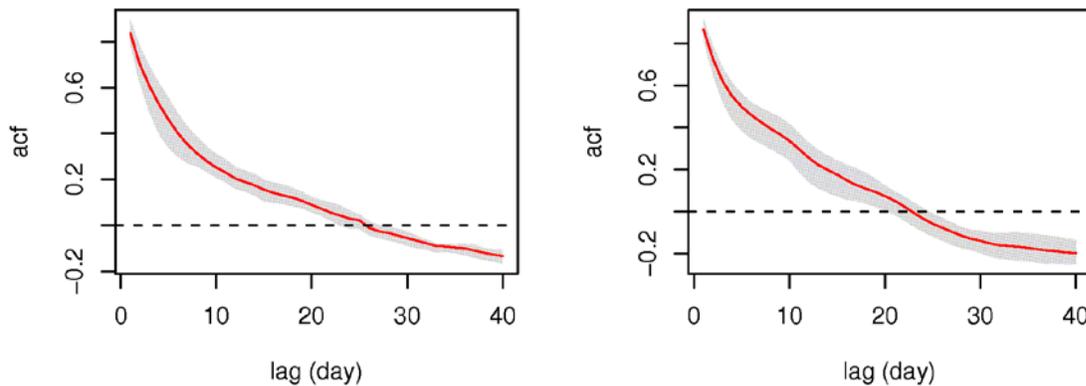
**Figure 14 June Flow PDF for the San Juan River near Bluff, UT [based on Prairie et al., 2007 method]. Solid line is observed data, boxplots are simulated flow.**

Regarding the utility of the Prairie et al. method for daily disaggregation, this technique has a tendency to produce negative flows at the location with the smallest flow magnitude, especially during the low flow season. This issue was also seen in the work of Lee (2008) and

further complicated when attempting to apply the genetic algorithm to daily data. However, at the monthly timescale, results were quite good and similar to those of Prairie et al. (2007).

#### 2.4.4 Lag Structure and Additional Daily Results

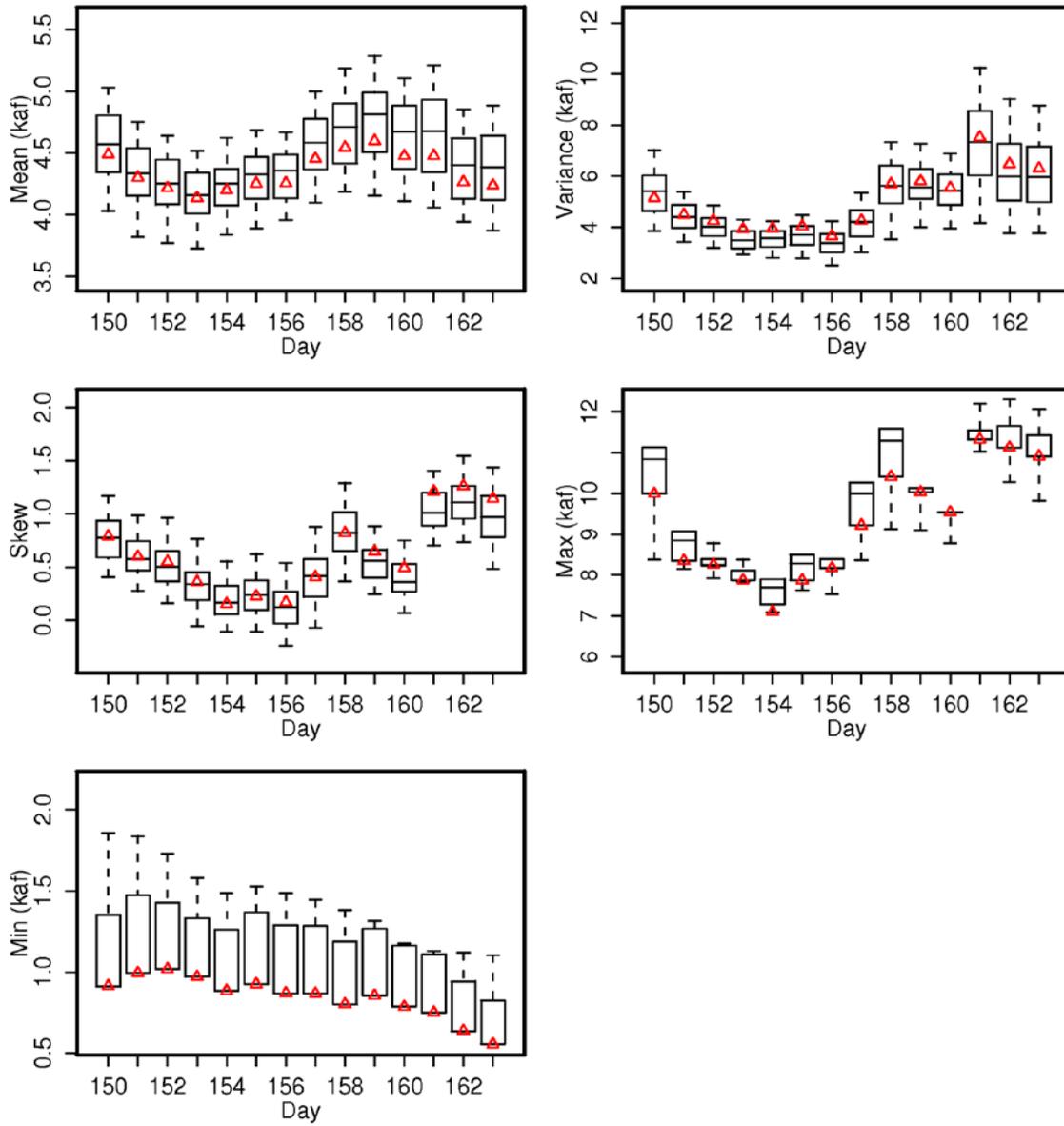
To examine the persistence structure we provide daily autocorrelation function (ACF) plots for wet (May-Jul) and dry (Jan-Mar) seasons at the Navajo River near Chromo, CO site. For each disaggregated simulation, an average ACF is computed for the two seasons, the range of which is shown by the grey region (Figure 15). Similarly, average ACFs are computed based on the observed data and are shown as solid lines. The observed curve falls within the range of the disaggregated results for both seasons.



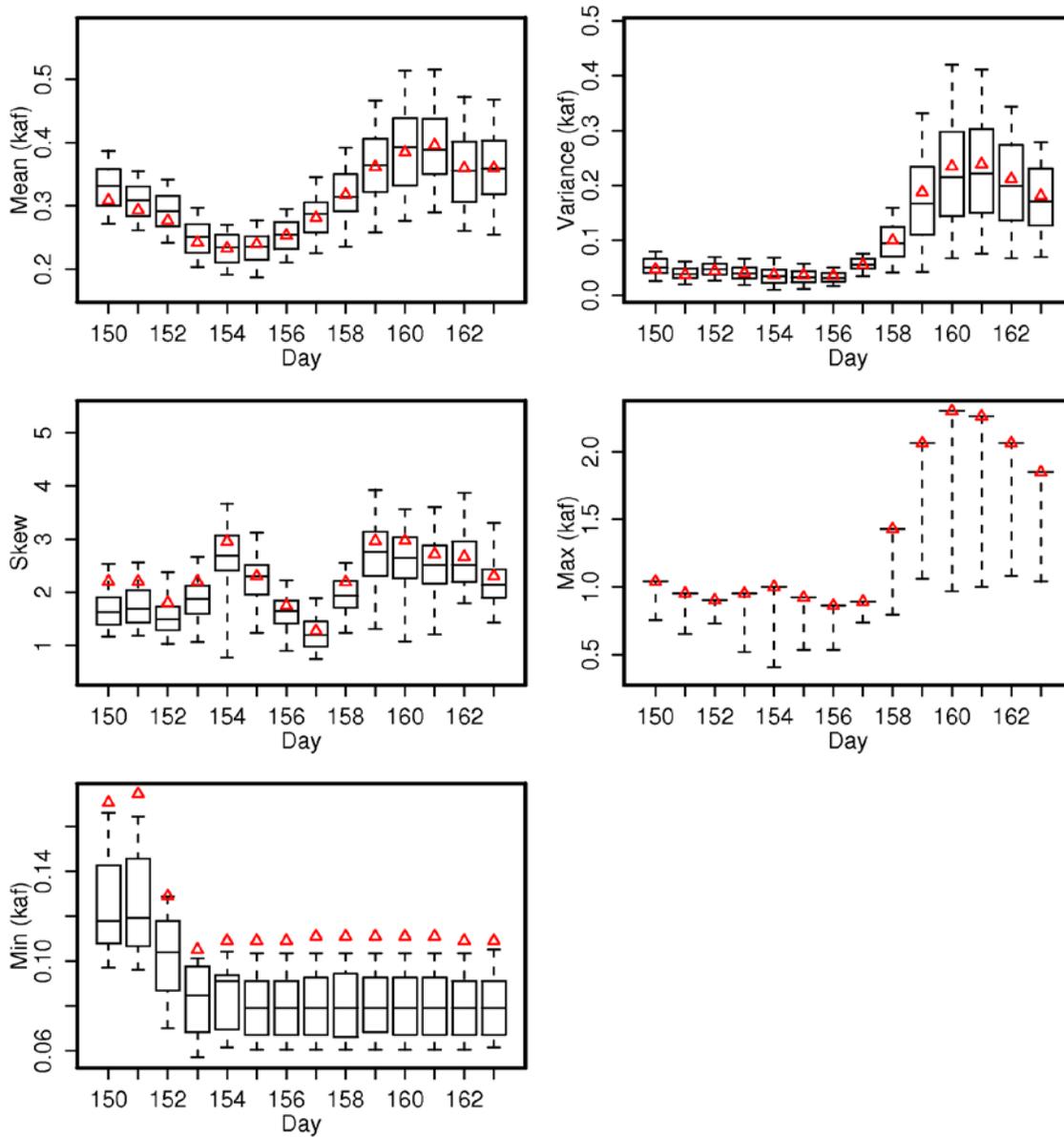
**Figure 15 Navajo River near Chromo, CO average daily autocorrelation function for dry season (Jan-Mar) (left) and wet season (May-Jul) (right). Grey region is range for disaggregated data (averaged over each simulation) and solid line is from observed data (averaged over period of record).**

Additionally, daily statistics for a fourteen-day period of the wet season are provided for the San Juan near Caraccas, CO and Navajo near Chromo, CO sites (Figure 16 and Figure 17). The mean, variance and skew of the observed data are well reproduced by the simulations.

However, simulated maximum and minimum distributions do not capture the respective observed values.



**Figure 16 San Juan River near Caraccas, CO daily statistics for the period of May 30 to June 12. Triangles represent values from observed data.**

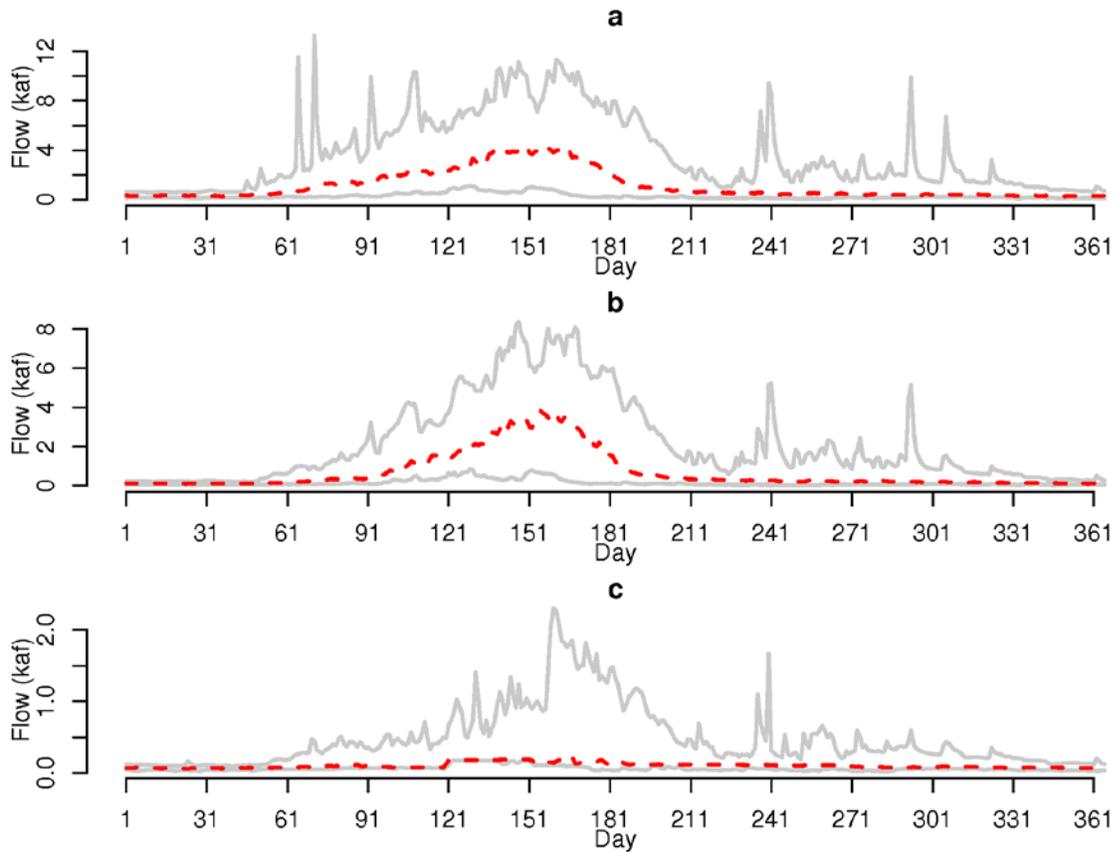


**Figure 17 Navajo River near Chromo, CO daily statistics for the period of May 30 to June 12. Triangles represent values from observed data.**

For the San Juan near Caraccas, few values are simulated below the observed minimum, but a variety extend beyond the observed maximum (Figure 16). In the case of the Navajo near Chromo (Figure 17), the opposite is seen; few values are larger than the observed maximum, yet a wealth of new minimum flows are simulated. These results were determined to be an artifact

of the observed data and index gauge simulation method. For Figure 16, when new minimum flows are not produced, this is due to the smallest proportion values coming from the same year as the index gauge annual minimum flow. While this is not often the case, it can occur. This prohibits any new minimum flows from being generated when the index simulation is a resampling of the observed record (e.g. index sequential method, K-NN). However, this can easily be remedied by employing a technique to generate annual values at the index location which extend beyond the observed extremes (e.g. AR-1, etc).

Figure 18 shows the daily range of observed values for all three locations as well as the median. In the case of the Navajo River, the observed minimum and median are quite close, indicating that at least 50% of the values for a given day are similar to the value corresponding to the historical low flow value. These consistently similar magnitudes translate to substantial variability in the proportion space; the local flow is fairly constant while the index flow changes considerably year to year (i.e. for a fixed release, the smallest proportion values are linked to the largest index flow). Therefore, there is a high propensity to resample a proportion vector that, when coupled with a below average annual flow, will generate unseen values.



**Figure 18 Daily flow ranges (solid grey) based on observed data with median values shown as dashed line for (a) San Juan River near Caraccas, CO, (b) San Juan River near Pagosa Springs, CO and (c) Navajo River near Chromo, CO.**

The ability to generate a wide range of flow magnitudes is generally considered a strength in stochastic simulation. However, a caveat exists; if there is a minimum flow requirement, this method has the potential to simulate flow values that violate such regulations. This is highlighted in Figure 17. Therefore, it is essential to emphasize the importance of understanding the basin in which flows are being disaggregated and any special needs or requirements that may exist. These issues could largely be circumvented by working with and simulating naturalized flows and subsequently adding the effects of regulation in a secondary modeling effort.

## 2.5 Summary and Discussion

A simple and highly adaptable disaggregation technique has been presented. The approach resamples historical flow proportion vectors conditioned on flow at an index location to produce disaggregated values that are continuous and are guaranteed to sum to the original value being disaggregated. The most novel application of the method is to produce daily flows from a simulated annual value. However, it has been demonstrated that the method is useful and effective at producing values at a variety of time scales. Distributional statistics of the historical data are reproduced in almost any space and time domain combination. Furthermore, the proportional disaggregation has the ability to generate extreme values previously unseen in the historical record, albeit with less frequency than some parametric approaches.

The main drawback to this technique is one that frequently plagues non-parametric disaggregation schemes: flow continuity between the end of one year and the start of a subsequent year. Lee (2008) has proposed a non-parametric disaggregation resampling approach that addresses this issue. However, it is mainly effective at the monthly time scale and for sites that dominate the index gauge. Proper lag relationships and continuity are not guaranteed for upstream locations that minimally contribute to the index gauge. Additionally, at the daily disaggregation resolution, the method of Lee (2008) offers little aid in preserving flow continuity.

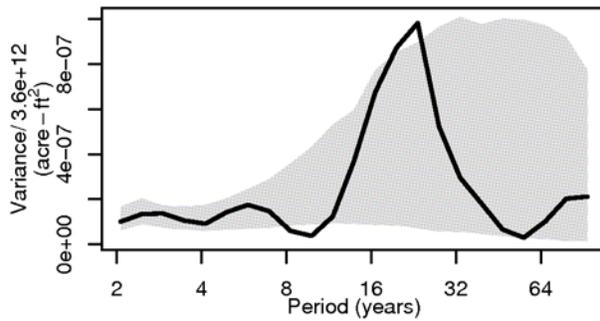
Additionally, as discussed earlier, results suffer in highly regulated reaches. As such, the method is best suited for locations with limited anthropogenic impact (e.g. USGS Hydro-Climatic Data Network) or completely naturalized flows. Once disaggregated, these data can be used as inputs to a water decision support tool, enabling the proper modeling of diversions, environmental flows, hydropower, reservoir levels, etc. without having to make such considerations in stochastic methods.

Kumar et al. (2000) successfully developed a daily disaggregation of monthly values through a kernel density estimation of the conditional PDF from which to resample, coupled with optimization to preserve continuity. The approach was demonstrated on some of the same locations within the San Juan Basin as utilized in this work. Results indicate that both have the ability to effectively disaggregate to daily values, however a major advantage of this work is the simplicity and limited computational requirements. Thus, given the parsimony and versatility of this approach, we believe it is a significant contribution to the field of stochastic flow simulation and will prove quite useful as a tool in water management and planning.

## CHAPTER 3: A WAVELET AUTO-REGRESSIVE METHOD (WARM) FOR MULTI-SITE STREAMFLOW OF DATA WITH NON-STATIONARY SPECTRA

### 3.1 Introduction

It has been recognized for some time that limited historical streamflow data are not sufficient information on which to base important water resources planning and management decisions. As such, approaches to identify a wider range of potential flow scenarios and variability have evolved over the years. Perhaps most rich in methods and application to water management is the field of stochastic streamflow simulation. These methods produce synthetic streamflow traces which reflect the statistical properties of the historical data, while often containing a wider range of flows and drought/surplus periods. Recently, studies examining the drivers of flow variability in the western United States have identified quasi-periodic climatic forcings (e.g. El Nino Southern Oscillation, etc) as having a role in modulating flow and other geophysical variability [McCabe *et al.*, 2007; Piechota and Dracup, 1996; Rajagopalan *et al.*, 2000; Tootle *et al.*, 2005]. While understanding of these links continues to evolve, it is important to consider their influences in stochastic simulation. Of particular importance for water planning and management is the link to persistence of wet/dry epochs. Popular parametric and non-parametric methods (e.g. Auto Regressive Moving Average, K-Nearest Neighbor, etc.) [Lall and Sharma, 1996; Ouarda *et al.*, 1997; Prairie *et al.*, 2006; Salas and Obeyseker, 1982; Thomas and Fiering, 1962] have been developed over the years to capture the distributional and dependence properties of the historical data, but they are poor at capturing the observed data spectrum (Figure 19) and consequently, the persistence of wet/dry epochs. This can lead to inaccurate estimation of water resources system reliability and thus affect management and planning decisions.



**Figure 19 Global wavelet spectrum for synthetic flow data (black line) and the 5th to 95th percentiles of the global wavelet spectra from 1000 traces simulated with an Auto Regressive (AR) model.**

The work of Kwon et al., 2007 offers a frequency domain simulation method that not only reproduces traditional statistical measures of the historical data, but captures spectral characteristics as well. In their paper, the frequency domain method, Wavelet Auto Regressive Method (WARM), is shown to well reproduce the global spectrum of the observed data along with distributional properties such as mean, variance, skew and probability density function (PDF). WARM offers an attractive alternative to traditional time domain stochastic simulation methods. However, for data with non-stationary spectral characteristics, we find that WARM cannot capture this feature and as a result, the traces lack the non-stationarity present in the historical data. Furthermore, there is also a need to provide simulations at multiple nodes in a streamflow network.

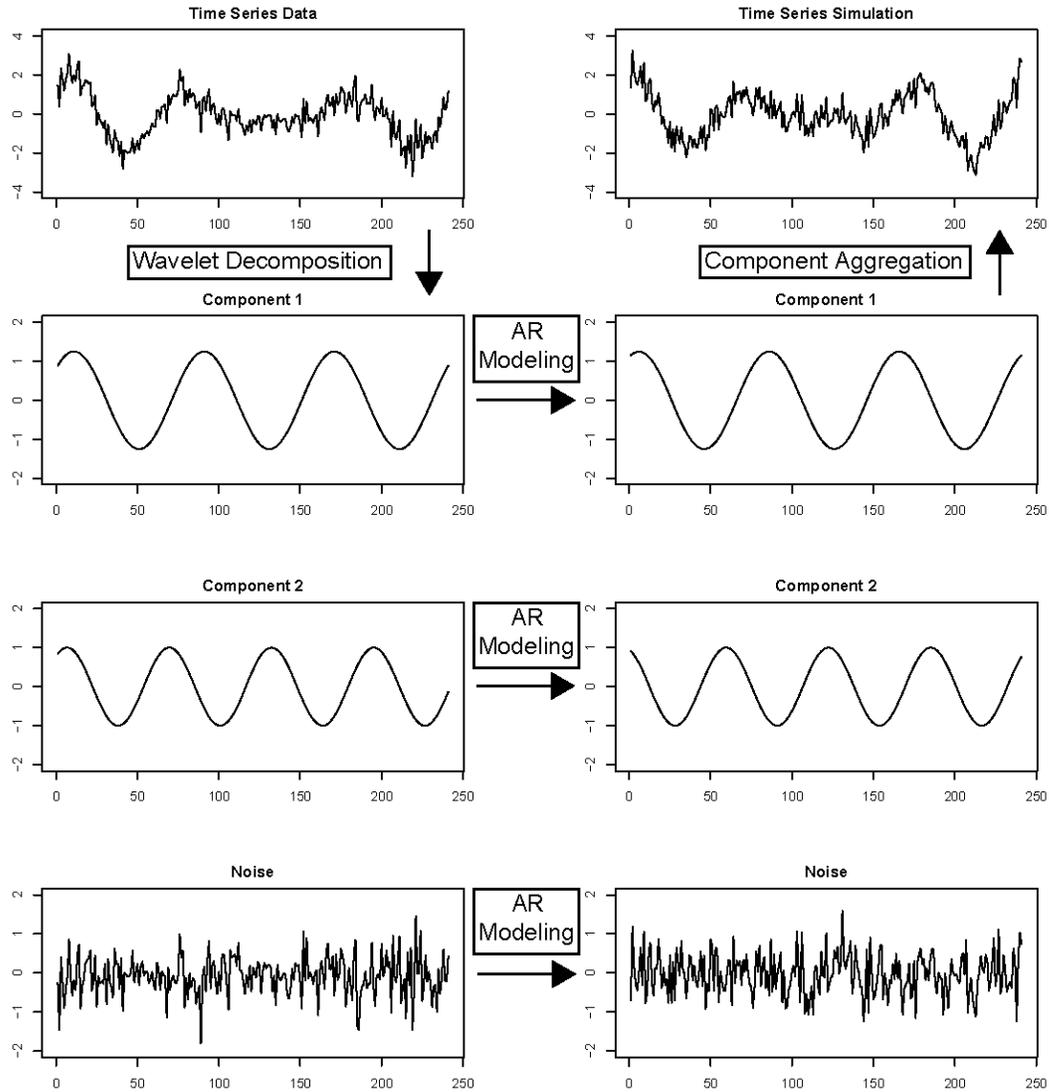
In this chapter we offer an enhanced WARM framework that provides stochastic simulations of streamflow traces at multiple locations, simultaneously capturing spectral and distributional properties present in the historical data. This generalized framework offers a flexible and robust multi-site streamflow simulation technique, which motivates this research. The chapter is organized as follows. Section 2 provides an overview of the original WARM framework, motivation for improvements and the proposed enhancements. Section 3 contains

results from application of the new method to streamflow in the Colorado River Basin. Section 4 is a summary discussion of the method and results.

## **3.2 Methodology**

### **3.2.1 Brief Introduction to WARM**

This section constitutes a brief overview of wavelet spectral analysis and the original WARM framework (see Figure 20). For additional detail, the reader is referred to Kwon et al., 2007. The use of wavelets in analyzing the time-frequency component of geophysical data has gained considerable popularity [Addison, 2002; Kwon et al., 2009; Torrence and Compo, 1998; Torrence and Webster, 1998]. Wavelet methods have several advantages over traditional spectral estimation techniques; an example is their ability to capture ‘local’ features such as amplitude modulation quite effectively. Additionally, wavelet-based methods have been shown to be efficient at revealing quasi-periodic signals in data with considerable noise – which is often the case with real data. For the purpose of introducing the WARM framework and motivating our enhancements, a synthetic flow data set is used, which by design contains underlying periodic variability at various time scales.



**Figure 20 Example diagram of WARM simulation procedure. The original data is decomposed via continuous wavelet transform using the Morlet wavelet and components are identified based on peaks in the global wavelet spectrum. Auto-Regressive models are fitted to each component as well as the residual noise term and each simulation is produced by combining the modeled component and noise signals.**

To begin, the WARM approach decomposes a time series into various components at a number of frequencies via the wavelet transform, thus giving the power (or variance) of the original data in the frequency and time domains. The continuous wavelet transform for some time series  $x_t$  is defined as:

$$X(a,b) = |a|^{-1/2} \int_{-\infty}^{+\infty} x_t \varphi^* \left( \frac{t-b}{a} \right) dt$$

**Equation 4**

where  $a$  is a scale parameter,  $b$  is the shift parameter and  $\varphi^*$  is the wavelet function [Kwon et al., 2007]. The (\*) denotes complex conjugate. We choose to employ the Morlet wavelet, given by:

$$\varphi(\eta) = \pi^{-1/4} \exp(i \omega_0 \eta) \exp(-\eta^2 / 2)$$

**Equation 5**

where  $\omega_0$  and  $\eta$  are non-dimensional frequency and time parameters respectively

[Torrence and Compo, 1998]. By substituting  $\left(\frac{t-b}{a}\right)$  for  $\eta$  in Equation 5, the shifted and dilated form of the mother wavelet is given [Addison, 2002; Kwon et al., 2009; Torrence and Compo, 1998; Torrence and Webster, 1998]. Equation 4 can be thought of as a series of convolutions between the wavelet function (Equation 5) and the original time series at all points for a variety of wavelet scales. To simplify the process, all convolutions can be completed simultaneously at a given scale by the convolution theorem. By doing so, the wavelet transform is defined as the inverse Fourier transform of the product of the data and the wave function in the Fourier space:

$$X_t(a) = \sum_{k=0}^{N-1} \hat{x}_k \hat{\varphi}^*(a \omega_k) \exp(i \omega_k t \delta_t)$$

**Equation 6**

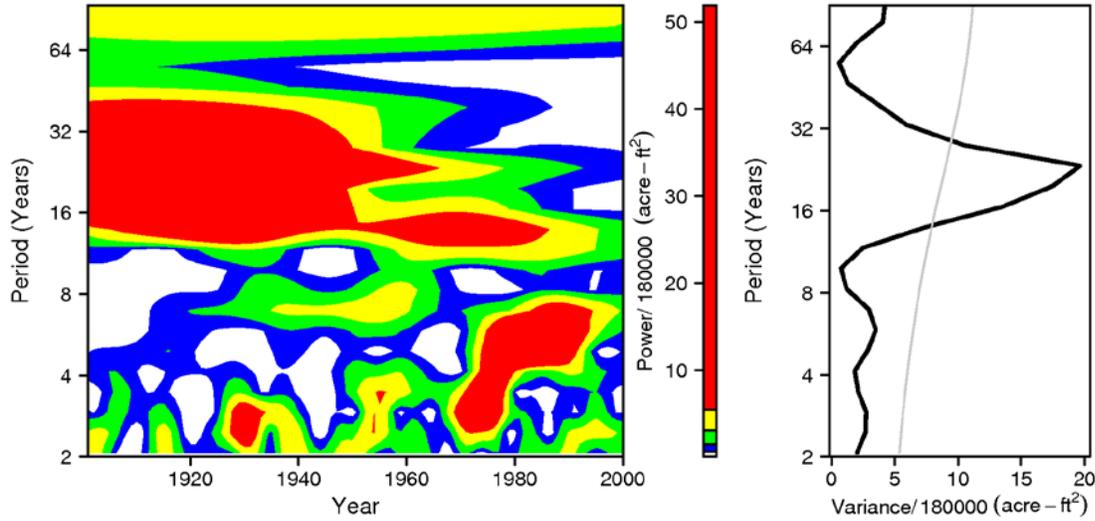
where  $\hat{x}_k$  is the discrete Fourier transform of  $x(t)$ ,  $\hat{\varphi}(a\omega)$  is the Fourier transform of the wavelet function,  $N$  is the number of points in the original data and  $\omega_k$  is the angular frequency.

The exponential term accomplishes the inverse transform [Torrence and Compo, 1998].

Additional details on wavelet-based time series spectral estimation can be found in Torrence and Compo (1998).

A contour plot of the wavelet transform ( $X_t(a)$ ) gives the wavelet spectrum at different frequencies over time (Figure 21). To the right in this figure is the global wavelet spectrum,

which shows average variance strength at each frequency across time. The global spectrum shows a strong peak in the 16-32 year period but it can be seen that this band is active only in the early epoch of the data.



**Figure 21** Wavelet spectrum (left) and global wavelet spectrum (right) for synthetic flow data. The grey line is the 95% significance white noise spectrum.

In the WARM approach, the global spectrum is next used to identify significant spectral components of the wavelet transformed data. Frequencies corresponding to statistically significant spectral peaks are combined together to create several ‘band passed’ time series which are then back-transformed to the original flow space. The same is done for the residual frequencies as ‘noise.’ The reconstructed (back-transformed) original time series is given as:

$$x_t = \frac{\delta_j \delta_t^{1/2}}{C_\delta \psi_0(0)} \sum_{j=0}^J \frac{\Re\{X_t(a_j)\}}{a_j^{1/2}}$$

**Equation 7**

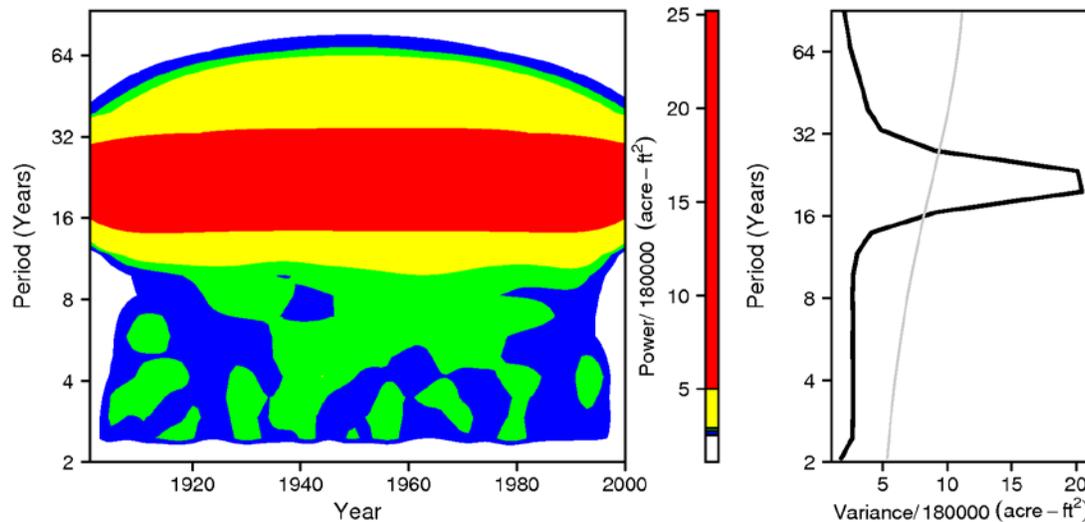
which constitutes the sum of the real part of the wavelet transform over all scales ( $j = 1:J$ ).  $\delta_j$  and  $\delta_t$  are the scale averaging coefficient and sampling period respectively. The remaining terms ( $C_s$  and  $\psi_0(0)$ ) are empirically derived reconstruction factors specific to the Morlet wavelet

[Torrence and Compo, 1998]. A “band specific” reconstruction is computed by simply limiting the range of scales ( $j$ ) over which the summation is performed (Equation 7).

By design, the ‘band passed’ time series components add up to the original data. To complete the WARM simulation, each of the components (including residual noise term) are modeled using traditional time series models such as an Auto-Regressive Moving Average (ARMA) model [Bras and Rodríguez-Iturbe, 1985; Salas, 1980]. The reconstructed components (except the noise) are smooth and hence can be effectively modeled as lower order AR models. A simulation is then produced by generating traces from each component model and then summing them together. This process is outlined in Figure 20.

### 3.2.2 WARM Drawbacks

From the results of Kwon et al. (2007) it can be seen that the WARM framework, reproduces the global wavelet spectrum quite well for a variety of datasets. In order to assess the efficacy of WARM simulations to reproduce key spectral features, a composite wavelet spectrum of 1000 WARM simulations based on the synthetic data is presented in Figure 22.

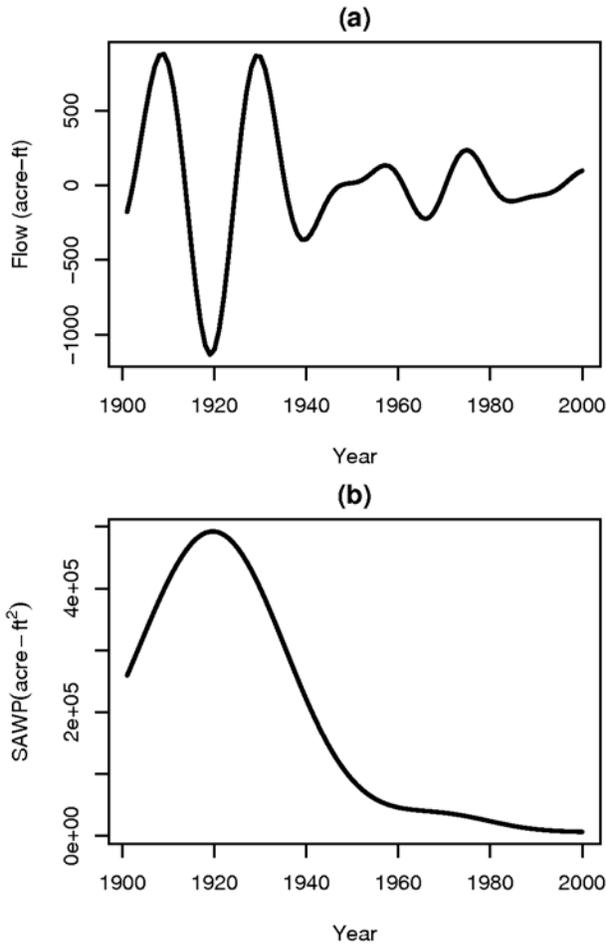


**Figure 22 Average wavelet spectrum (left) and average global wavelet spectrum (right) from 1000 WARM simulations for synthetic flow data. The grey line is the 95% significance white noise spectrum.**

It can be seen that the simulations capture the global spectrum very well but, the variability contributing to the significant peak in the global spectrum (which is only active in the first half of the century, see Figure 21) is smoothed across time, thus failing to capture the non-stationary spectral feature. This inability can have significant impact on the epochal behavior of the wet/dry sequences and consequently estimates of water resources system risk, when coupled with a decision model. Thus, enhancing the WARM approach to address this drawback is our first motivation of this study and our proposed enhancement is described below.

### **3.2.3 Proposed Enhancement to WARM**

The reconstruction of the example data in the dominant spectral band, (16-32 year period, Figure 23 ) using Equation 7, shows reduced variability in recent decades.



**Figure 23 (a) 25 year component of synthetic flow data (b) scale averaged wavelet power (SAWP) for 25 year component of synthetic flow data.**

Also shown in this figure is the Scale Averaged Wavelet Power (SAWP) [Kwon *et al.*, 2007] of this dominant band. The SAWP time series reflects the temporal variability of the strength of this band, which is given as:

$$\bar{X}_t^2 = \frac{\delta_j \delta_t}{C_\delta} \sum_{j=j_1}^{j_2} \frac{|X_t(a_j)|^2}{a_j}$$

**Equation 8**

where  $j_1$  and  $j_2$  are the scales over which averaging is computed.

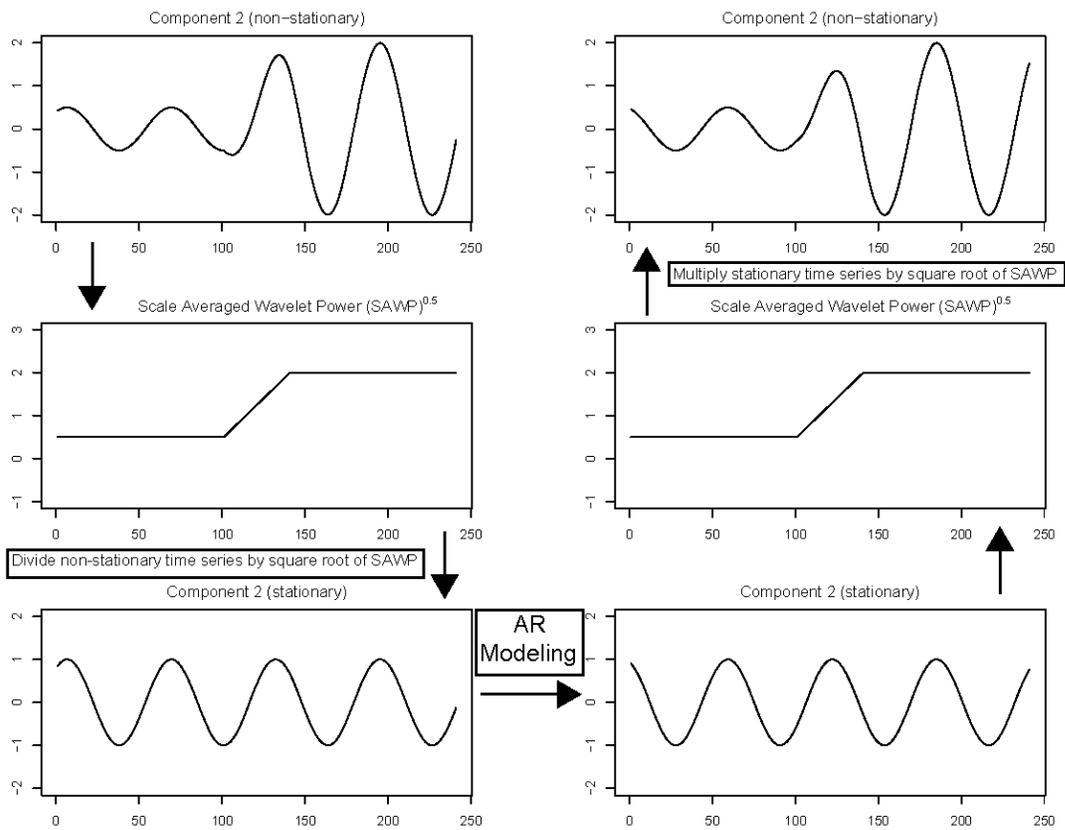
To provide the ability to capture non-stationary spectral features we propose the following enhancement to WARM using the SAWP.

(i) The reconstructed, non-stationary components, are divided by its  $\overline{X}_t$  (square root of SAWP from Equation 8), thus producing a new, stationary time series.

(ii) These new stationary versions can now be effectively modeled using AR models – as they satisfy the stationary assumptions. The fitted AR models are used to simulate the components – per the WARM approach.

(iii) The non-stationarity is re-introduced via multiplication with  $\overline{X}_t$ . –essentially reversing step (ii).

These enhancements to the WARM are outlined in Figure 24.



**Figure 24 Schematic of improved WARM simulation procedure, shown for a single, non-stationary component.**

### 3.2.4 Multisite Streamflow Simulation

It is well established that for the purpose of water management and planning, particularly on a large river network with competing demands, flow data is required at multiple locations that preserve spatial and temporal dependencies. As such, it is of considerable interest to extend the above described enhanced WARM to simulate streamflow at multiple sites. To accomplish this, the “enhanced WARM” is first applied to an aggregate streamflow (often the downstream gauge on a network). Next, the simulated flows are then disaggregated in space using the proportion method presented in chapter 2.

A brief summary of the disaggregation method is provided below.

(i) Flow at a downstream gauge is the ‘aggregate’ of the upstream flows. Sometimes a synthetic aggregate gauge is created by summing all the spatial flows.

(ii) For each historical year the ‘proportion’ vector – i.e., the proportion of aggregate flow at each spatial location, is computed.

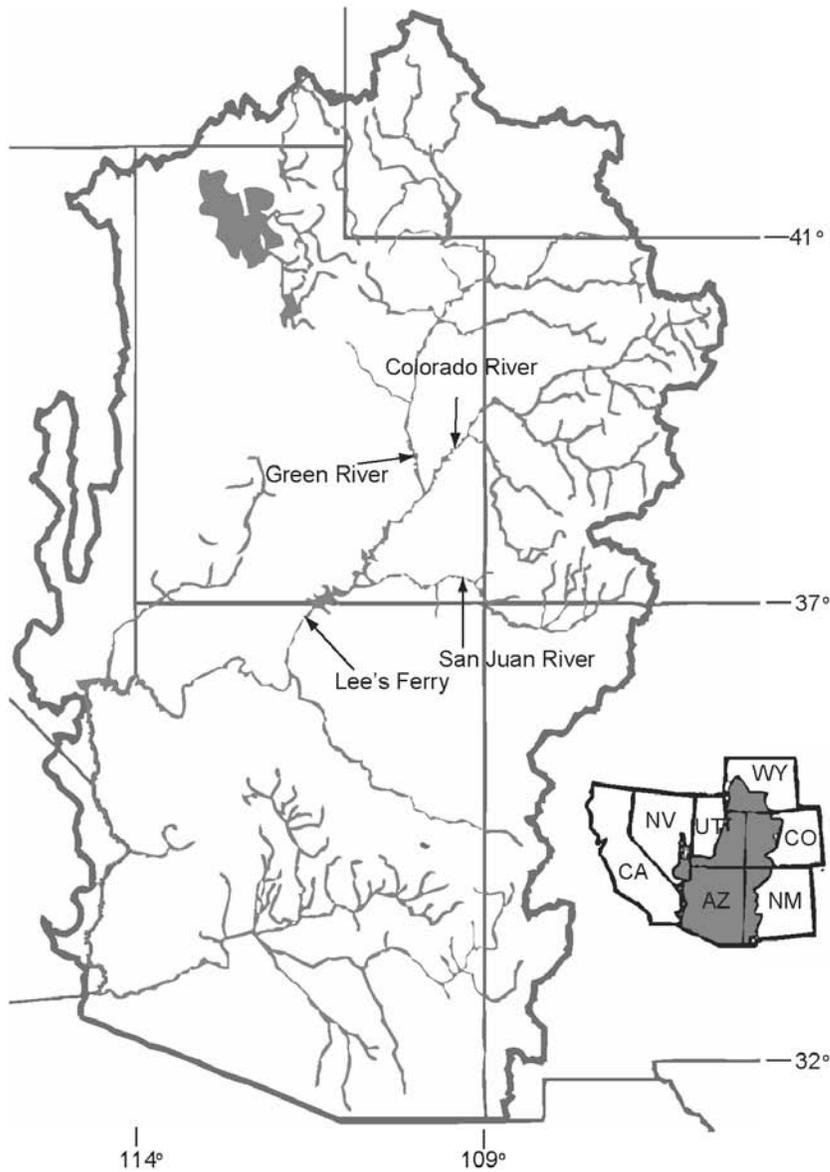
(iii) The aggregate flow (at the downstream location) is first generated, i.e., using the WARM approach described above. Analog years for each flow value are identified from the historical record and one is resampled using a probability metric that gives high weight to the nearest neighbors and least to the farthest. The proportion vector corresponding to the selected neighbor is then applied to the aggregate flow to obtain disaggregated flows at the spatial locations.

This produces streamflow scenarios at all the desired locations on the network, preserves summability, distributional and spectral properties. Application of this to streamflow from four locations on the Colorado River Basin is described below.

### **3.3 Results from Application to Colorado River Basin**

#### **3.3.1 Data and Study Area**

The study area for the demonstration of the enhanced WARM approach is the Upper Colorado River Basin (Figure 25). This region incorporates parts of Colorado, Utah, Wyoming and small portions of Arizona and New Mexico. The gauge at Lee's Ferry, AZ aggregates all of the Upper Basin flow. Our proposed methodology is first applied to simulate flow for this location, which is then disaggregated to three major upstream tributaries on the network. These locations are Green River at Green River, UT, Colorado River near Cisco, UT and San Juan River near Bluff, UT (Figure 25). Each represents a significant drainage of the Upper Colorado River Basin.



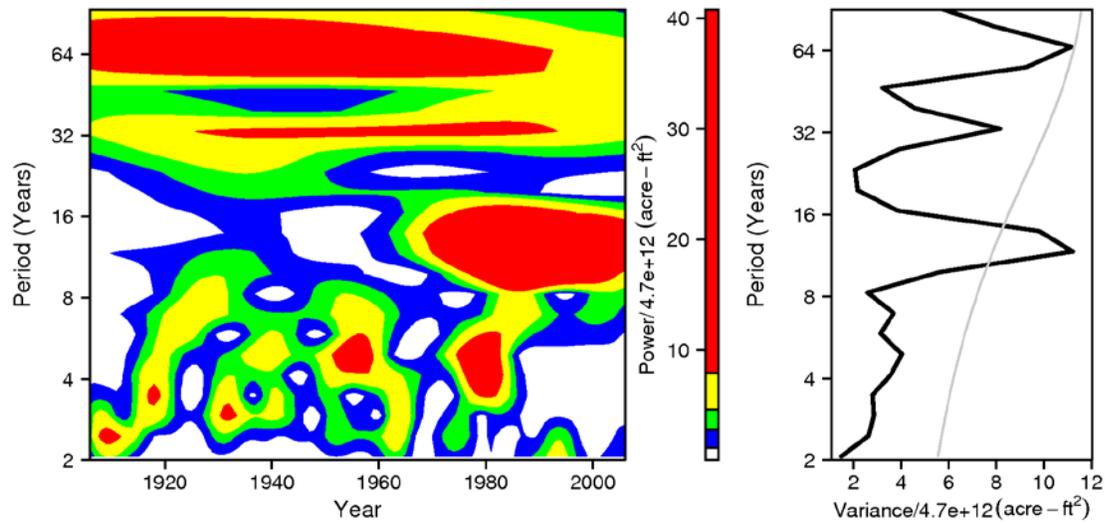
**Figure 25 Colorado River Basin map with gauging locations for disaggregation annotated; Green River at Green River, UT, Colorado River near Cisco, UT, San Juan River near Bluff, UT and Colorado River at Lee's Ferry, AZ.**

In this work, annual water year (Oct-Sep) natural streamflow data for the Colorado River Basin from the period 1906-2006 are employed. This dataset is developed and updated regularly by the Bureau of Reclamation. Naturalized streamflow are computed by removing anthropogenic impacts (e.g., reservoir regulation, consumptive water use, etc.) from the

historical flows. Prairie and Callejo [United States Department of the Interior, 2005] present a detailed description of methods and data used for the computation of natural flows in the Colorado River Basin.

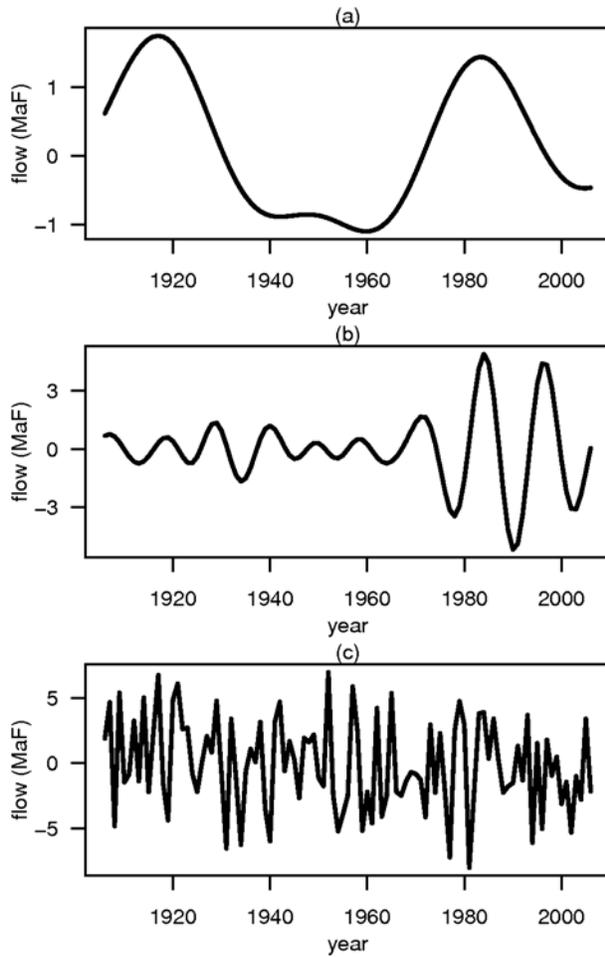
### 3.3.2 Simulations at Lee’s Ferry, AZ

The first step, as discussed earlier, is to compute the wavelet spectrum of Lee’s Ferry streamflow and identify the significant spectral bands. Figure 26 shows the wavelet spectrum and we identified significant bands in the 9-16 period and 50-70 year period.



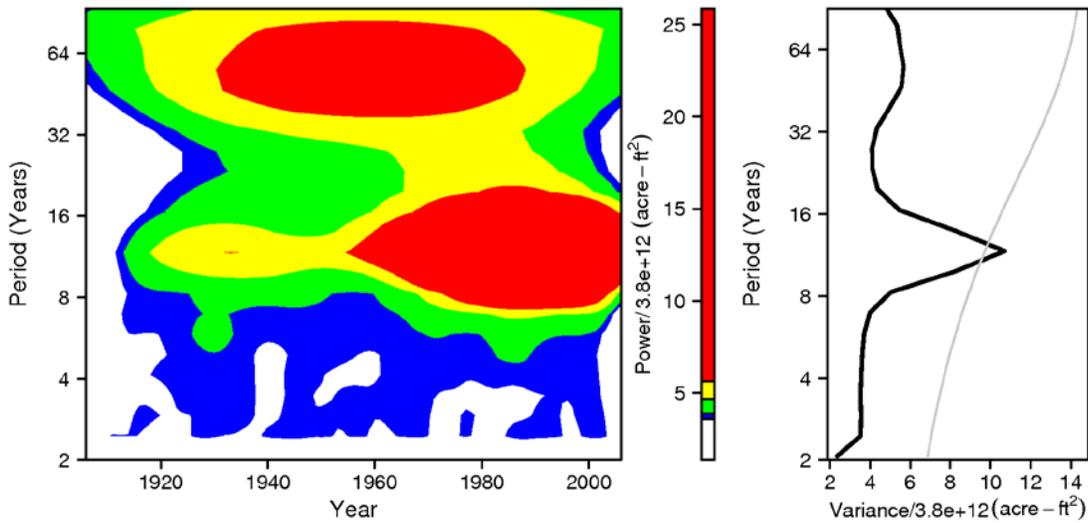
**Figure 26 Wavelet spectrum (left) and global wavelet spectrum (right) for Lee’s Ferry, AZ natural flow data. The grey line is the 95% significance white noise spectrum.**

The significance of the very low frequency band needs to be taken with caution due to smaller sample size and boundary effects. It can also be seen that the significant decadal band appears just in recent decades, indicating a strong non-stationarity in the spectrum, thus indicating the need for “enhanced WARM.” In all, we have three spectral bands including the noise component shown in Figure 27.



**Figure 27** Re-constructed components (a) low frequency signal, (b) decadal signal and (c) noise

The proposed enhanced WARM was applied and 1000 streamflow ensembles 101 years in length were generated (length of Lee's Ferry natural flow data set). The combined spectrum of these simulations can be seen to capture all the non-stationary spectral features presented in the historical data (Figure 28).



**Figure 28 Average wavelet spectrum (left) and average global wavelet spectrum (right) from 1000 improved WARM simulations for natural flow at Lee's Ferry, AZ. The grey line is the 95% significance white noise spectrum.**

Boxplots of probability density functions (PDFs) for annual flow simulations along with that of the historical data are shown in Figure 29 – it can be seen that they are very well captured by the simulations. Additionally, distributional statistics such as mean, variance, coefficient of skew and maximum/minimum values from the simulations are also well reproduced (see boxplots in Figure 30), with the exception of skew (which is very small) falling just outside the boxplot box. Nonlinear AR models for the components may help better capture the skew. Given the very small skewness in the data we did not deem it necessary to explore this.

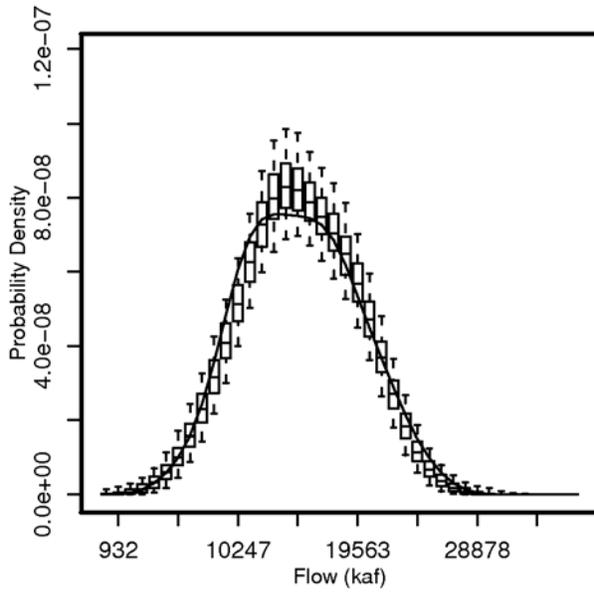


Figure 29 Improved WARM Lee's Ferry flow PDFs shown as boxplots with historical flow PDF as solid black line.

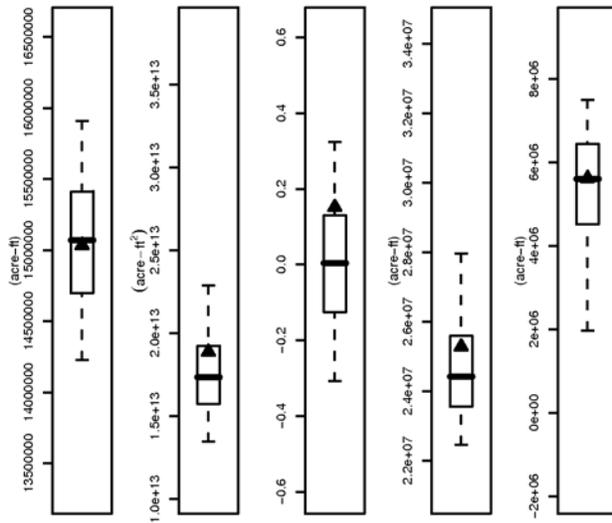
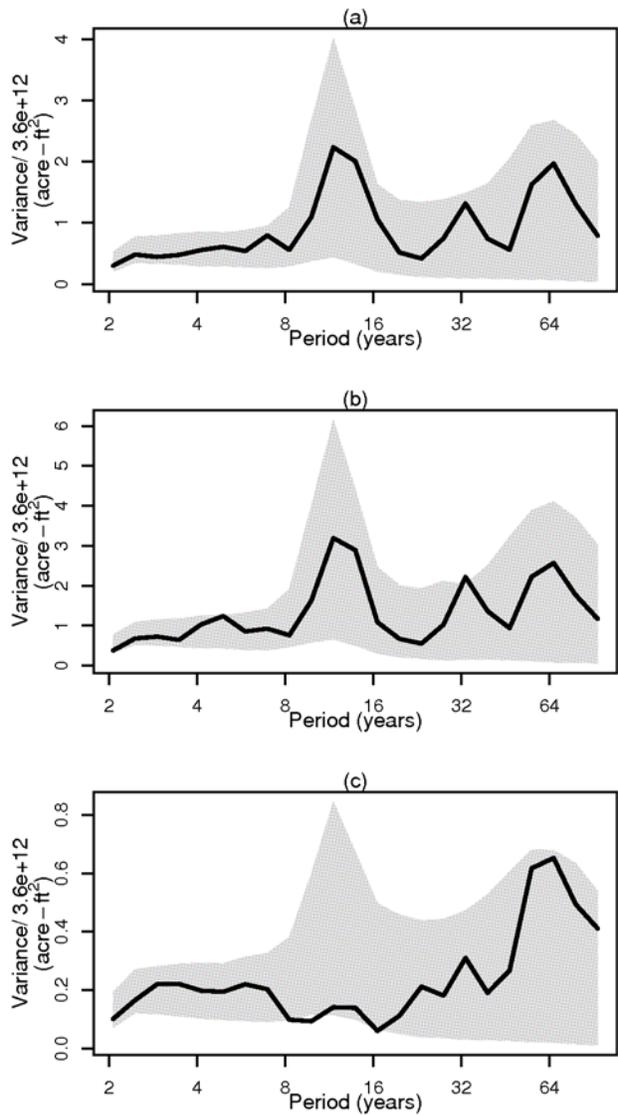


Figure 30 Boxplots of mean, variance, skew, max and min values from 1000 improved WARM simulations for Lee's Ferry, AZ natural flow. The box of the boxplot represents the interquartile range and whiskers extend from the 5<sup>th</sup> to 95<sup>th</sup> percentiles. Black triangles are values from observed data.

### **3.3.3 Multi-Site Simulations**

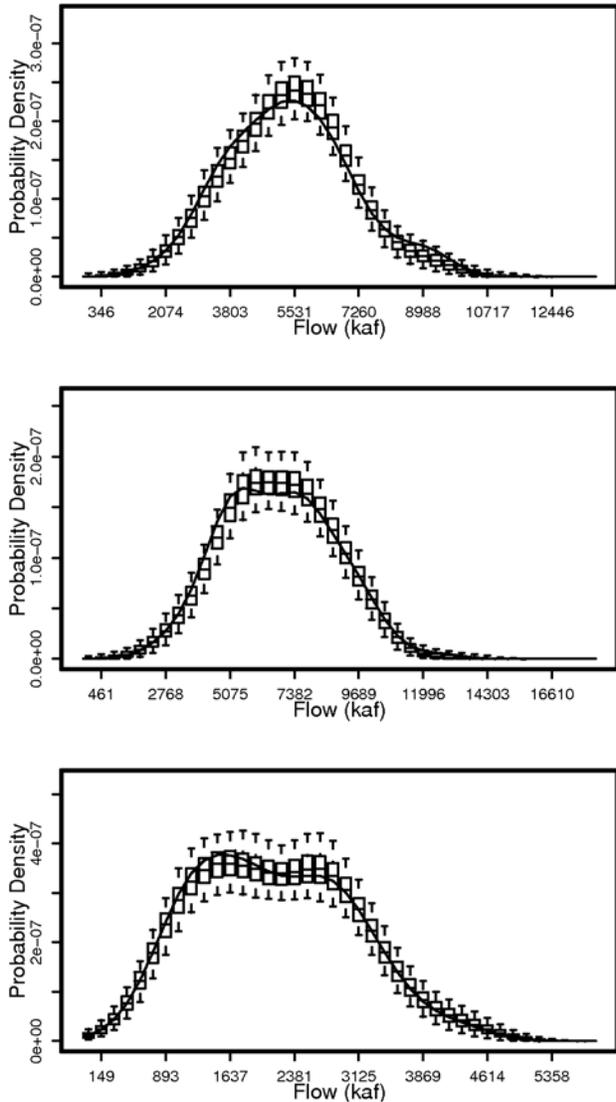
The simulated streamflow at Lee's Ferry from the previous step were disaggregated using the proportion disaggregation method presented in chapter 2 to generate streamflow ensembles at the three upstream locations on the network. The range of global wavelet spectra from the simulations are computed at the three locations and shown in Figure 31 along with the respective historical spectra. At the Green River and Cisco locations, the observed spectra are very well captured by the simulations.



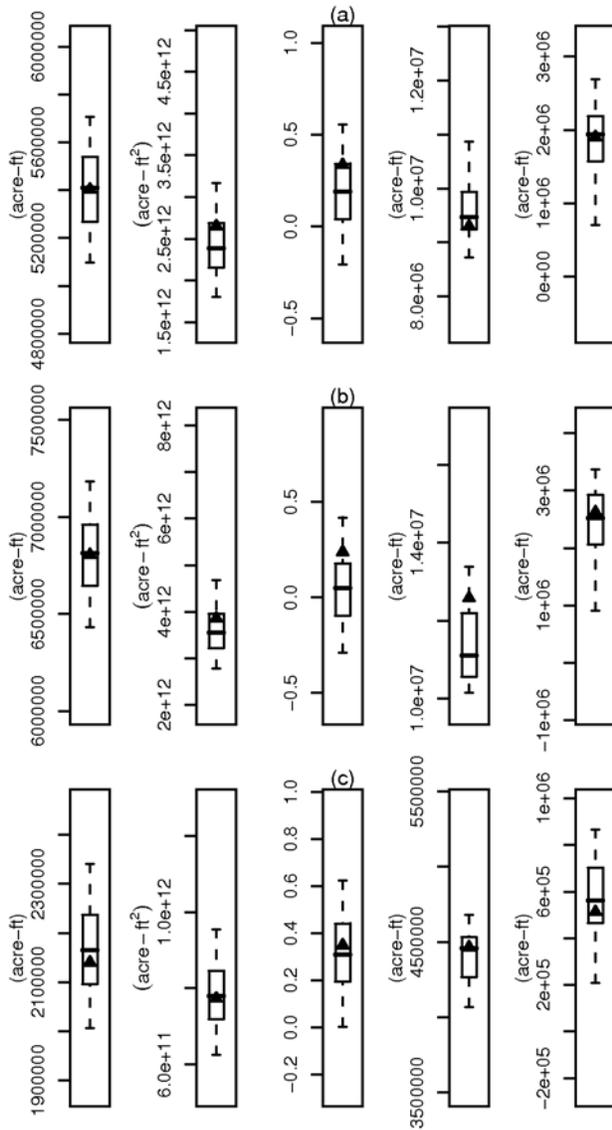
**Figure 31 Global wavelet spectra for (a) Green River at Green River, UT, (b) Colorado River near Cisco, UT and (c) San Juan River near Bluff, UT. Black line is historical spectrum and grey region is 5<sup>th</sup> to 95<sup>th</sup> percentile based on spectra from disaggregation of 1000 improved WARM traces.**

Furthermore, observed data for these sites exhibit the non-stationary decadal band seen in Figure 26 in Lee’s Ferry flow. These features are also reflected in the disaggregated traces (not shown). The San Juan spectrum, however, is not well reflected by the disaggregated results. The spectral properties of the disaggregated streamflow tend to be inherited from the aggregate flow location. The larger upstream tributaries typically share in the same spectral signature, but

smaller tributaries such as the San Juan can differ, particularly when the sub-basin is in a geographically or climatologically different region. The distributional properties on the other hand, are very well preserved at all the three locations as shown by both the PDFs (Figure 32) and suite of statistical measures (Figure 33).



**Figure 32 Disaggregated, improved WARM simulation flow PDFs for (a) Green River at Green River, UT, (b) Colorado River near Cisco, UT and (c) San Juan River near Bluff, UT. The black line is the observed PDF and the boxplots show the range of simulation PDFs.**



**Figure 33 Statistics for disaggregated improved WARM simulations for (a) Green River at Green River, UT, (b) Colorado River near Cisco, UT and (c) San Juan River near Bluff, UT. Measures are mean, variance, skew, maximum and minimum (left to right). The box of the boxplot represents the interquartile range and whiskers extend from the 5<sup>th</sup> to 95<sup>th</sup> percentiles. Black triangles are values from observed data.**

### 3.4 Summary and Discussion

An enhanced wavelet auto regressive method (WARM) is coupled with a flexible and proven disaggregation approach to produce stochastic streamflow at multiple locations that accurately reflect the historical spectral and distributional properties. The enhancements afford

the ability to model and simulate data with non-stationary spectral characteristics, which properly reflect the changing persistence of wet/dry conditions that are important for subsequent risk estimations.

From the results section, the disaggregated spectrum for the San Juan River merits further discussion. It is clear that the observed spectrum is not completely reflected in the simulations. From additional analysis, it was found that the spectral pattern common among the Lee's Ferry, Cisco and Green River locations is present throughout most of the upper Colorado Basin; they are likely impacted by similar large scale climate forcings. San Juan on the other hand, can exhibit variability patterns seen in both the Upper and Lower Basins and hence a different spectral signature. Thus, it is important to have some a priori knowledge of the characteristics of the basin in which the method is being applied. For most basins, annual flow variability is likely to be spatially homogeneous and thus tributaries should have similar spectral features. However, for a large basin, such as the upper Colorado Basin, which spans five states and encompasses extremely varied terrain, it is possible to have sub regions with substantially different spectral signatures. Apart from imparting "extra" decadal variability, the simulations capture all of the distributional properties and the additional low frequency persistence may be desirable in planning by generating longer wet and dry sequences.

From working with a variety of data and this simulation approach, a few additional implementation comments are offered. If the noise component exhibits non-normal features it might be better to use a bootstrap approach to capture its distribution. This was the case for the Lee's Ferry data used in this study and the bootstrap was quite effective at improving the overall result.

When combining the noise and component terms to produce a trace, it was found that overall variance can be slightly under-simulated. This was determined to be a result of weak correlation between the original components and noise. As such, when combined, the aggregate variance was greater than the sum of each term's variance. Thus, by simulating each term independently, the weak correlation is lost and total variance may have a tendency to be slightly low. None of the observed correlation values between terms proved to be significant but the small correlations are likely due to the periodic nature of the components. To address this, we look to basic probability theory for correlated random variables [*Montgomery and Runger, 2003*]. The variance of a time series produced by combining two correlated random variables is given by Equation 9:

$$\sigma_{x+y}^2 = \sigma_x^2 + \sigma_y^2 + 2\sigma_{xy}$$

**Equation 9**

where  $x$  and  $y$  are the correlated random variables,  $\sigma^2$  is variance and  $\sigma_{xy}$  is their covariance.

Thus, once the three simulated components are combined, the new time series is multiplied by a variance correction factor ( $v_f$ ) given by:

$$v_f = 1 + \left( 2\sigma_{xy} / \sigma_{x+y}^2 \right)$$

**Equation 10**

which corrects the variance to that of the original data. It is important to note that this may not always be a problem, and is likely dependant upon both the original data and the scales combined during the component reconstruction. However, should this issue present itself, we find that this is the most simple and effective way to address the problem – compared to other alternatives such as multivariate auto-regressive models which can be unwieldy.

The enhancements to the WARM approach offered in this work are unique and substantial improvements to the original framework, particularly for the purpose of water resources planning and management. Additionally the discussion provides insight on

applications and addressing potential issues such as under-simulation of variance and failure to re-produce spectral features when coupled with the disaggregation method. Last, the ability to properly model the time-varying strength of significant modes of variability constitutes an effective method by which to simulate changes in low frequency persistence.

## CHAPTER 4: COLORADO RIVER BASIN HYDRO-CLIMATIC VARIABILITY

### 4.1 Introduction

In many regions of the United States, water management and reservoir operations occur at the sub-annual timescale. These projects often have seasonal targets to accomplish objectives such as flood control, recreation and hydropower. Thus, for facilities with limited storage capacity, information beyond a seasonal forecast is of little use. On the other hand, the Colorado River Basin boasts a collective storage capacity of approximately four times the mean annual flow of the river. Furthermore, over 80% of that storage is found in Lakes Powell and Mead alone. Thus, understanding persistent modes of streamflow variability at the inter-annual and longer timescale is of great importance for water management in this system.

There is increasing evidence that the western U.S. hydro-climate exhibits significant inter-annual (year-to-year) variations, driven by large-scale climate features such as ENSO, PDO and AMO. Particularly, the links to ENSO are robust, widespread and have a long history of recognition. During El Niño events (warm sea surface temperature anomalies in the central and eastern equatorial Pacific Ocean), the winter subtropical jet over the southwestern U.S. strengthens [*Horel and Wallace, 1981*] and consequently, the Pacific Northwest experiences below-normal and the Southwest above-normal precipitation [*Cayan et al., 1998; Dettinger et al., 1998; Diaz and Markgraf, 1992; , 2000; Redmond and Koch, 1991*]. Generally the opposite effect is observed during La Niña events (cooler sea surface temperature in the central and eastern equatorial Pacific), but some non-linearities are present in this teleconnection [*Clark et al., 2001; Hoerling et al., 1997; Rajagopalan et al., 2000*]. Similar ENSO teleconnection patterns have been observed in the inter-annual variability of winter snow water equivalent [*Cayan, 1996*;

*Clark et al., 2001*], surface temperature [*Gershunov and Barnett, 1998; Higgins et al., 2002; Redmond and Koch, 1991*] and streamflow [*Dracup and Kahya, 1994; Grantz et al., 2005; Hamlet and Lettenmaier, 1999; Kahya and Dracup, 1994; Maurer et al., 2004; Piechota et al., 1997; Regonda et al., 2006; Tootle et al., 2005*]. Shifts in streamflow seasonality [*Cayan et al., 2001; Grantz et al., 2007; Regonda et al., 2005; Stewart et al., 2005*] have also been linked to ENSO and a general warming trend in recent decades. At decadal time scales, PDO is often considered the primary driver of hydrologic variability [*Brown and Comrie, 2004; Hidalgo and Dracup, 2003; McCabe and Dettinger, 1999; , 2002*], despite debate on the independence of PDO from ENSO [*Newman et al., 2003*].

Hunter et al. (2006) examined the link between April 1 snow water equivalent at Snotel<sup>1</sup> sites across the western U.S. and various climate indices. They found that for regions in the Upper Colorado River Basin (UCRB), La Niña years and the negative phase of AMO corresponded with above average snowpack. Additionally, the cold phase of PDO, when compared with the warm, was found to correspond with reduced snowpack in similar regions. Timilsena et al. (2009) noted similar climatic ties to streamflow in the UCRB, finding the warm phase of PDO to be associated with above average flow. Also, a weak inverse relationship between flow and AMO phase was noted. However, El Niño years were linked with above average flow for much of the Basin, albeit more pronounced in the Lower Basin. Timilsena et al. (2009) also found the effects of ENSO and PDO were intensified when in phase and suppressed when out of phase.

Thomas (2007) studied Lower Colorado River Basin (LCRB) flow ties to the same indices at seasonal and annual scales, for a variety of lead/lag times. ENSO and PDO were both found to have strong relationships for winter/spring and annual flow volumes at little to no lag.

---

<sup>1</sup> <http://www.wcc.nrcs.usda.gov/snow/about.html>

An inverse relationship was detected for annual and winter/spring volumes with AMO at a lag of 6 to 24 months. McCabe et al. (2007) examined decade-scale variability in global sea surface temperatures (SSTs) and UCRB flow, and they also found strong ties to AMO and PDO. Furthermore, from analysis of climate division temperature and precipitation data, for much of the United States, positive (negative) AMO corresponds with above (below) average temperatures and below (above) average precipitation.

While these works offer considerable insight into the influence of various climatic phenomena on the hydro-climate of the inter-mountain west region, little has been done from the perspective of detection and attribution for significant streamflow variability modes in the Colorado Basin. To this end, we performed an analysis of historical and paleo reconstructed Colorado River streamflow using wavelet spectral methods. Subsequently, physical processes and potential climate phenomena were identified and associated with inter-annual and longer modes of variability. This approach is similar to that employed by Coulibaly and Burn (2004) for basins across Canada and by Massei et al. (2010) in the Mississippi River drainage. The chapter is organized as follows: data sets and methods used are next presented, followed by results and discussion.

## **4.2 Data Sets**

The following provides an overview of the various data sets used in the subsequent analyses of this chapter.

### **4.2.1 Colorado River Natural Streamflow**

In this work, annual water year (Oct-Sept) natural streamflow data at Lees Ferry on the Colorado River for the period 1906-2006 are employed. Naturalized streamflow data are computed by removing anthropogenic impacts (i.e., reservoir regulation, consumptive water use,

etc.) from the recorded historical flows [*United States Department of the Interior*, 2005]. These data are developed and updated regularly by the Bureau of Reclamation. Figure 34 shows the Colorado River Basin and natural flow nodes. Of significance is the Lees Ferry gauge, which divides the Basin into upper and lower portions per the 1922 Colorado River Compact [*Colorado River Commission. et al.*, 1923]. More than 90% of the natural streamflow in the Basin passes through this point. As such, we chose to perform our analysis at this location [*United States Department of the Interior*, 2005].

# Colorado River Basin

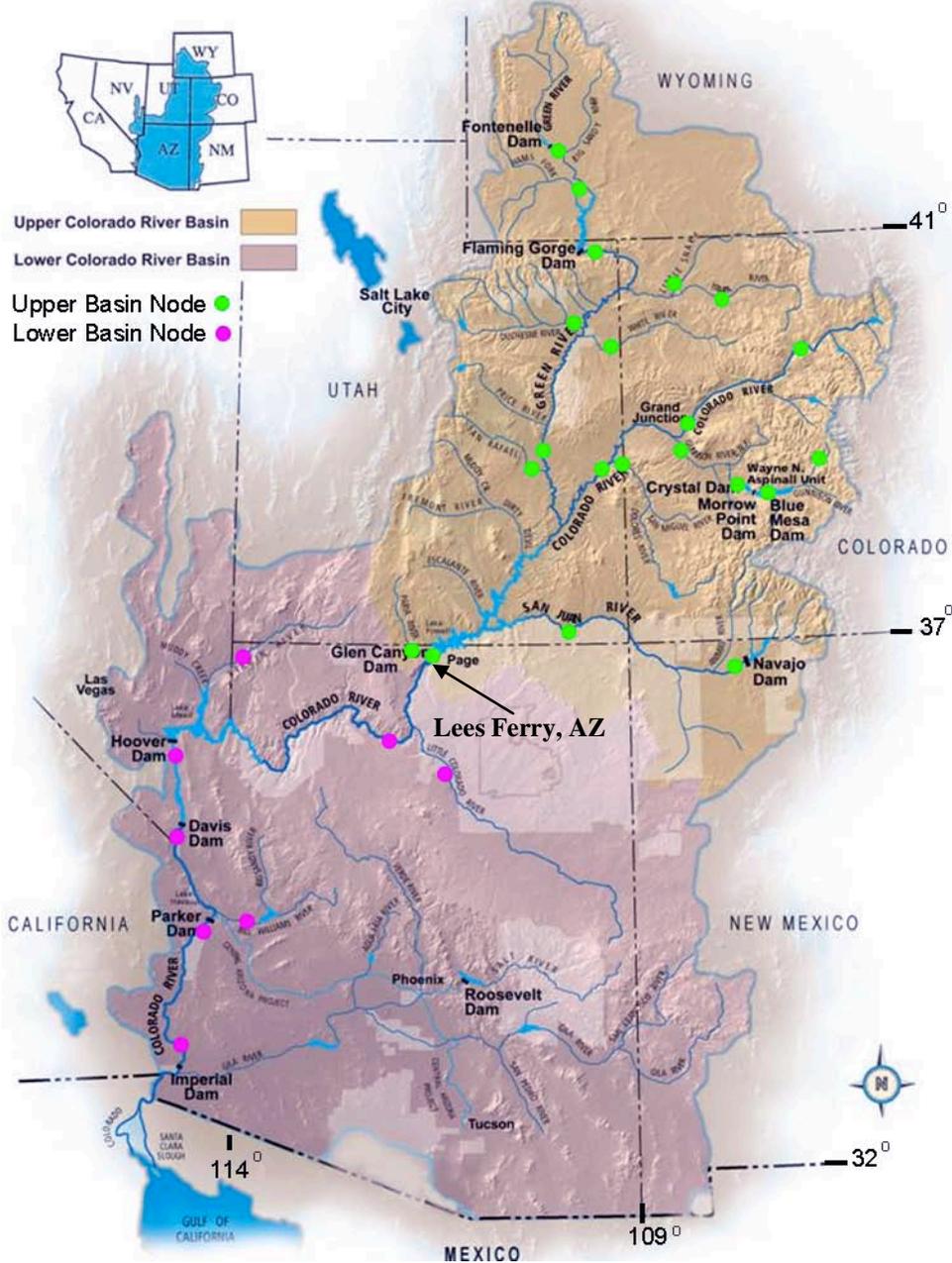


Figure 34 Colorado River Basin. The tan region is the Upper Basin and the pink region is the Lower Basin. Green dots represent natural flow nodes in the Upper Basin, and purple dots indicate natural flow nodes in the Lower Basin.

#### **4.2.2 Climate Index Data**

The AMO index is computed as a monthly area weighted average of North Atlantic (latitudes north of the equator) sea surface temperatures, which is subsequently de-trended. Gridded data used for this calculation are from the Kaplan SST data set. These data are updated monthly and have a spatial resolution of  $5^{\circ} \times 5^{\circ}$ . Values were accessed from the NOAA Physical Sciences data website<sup>2</sup> and are available for the period 1856 – present.

#### **4.2.3 Upper Colorado River Basin Precipitation and Temperature Data**

Monthly precipitation on a high resolution spatial grid ( $0.04^{\circ} \times 0.04^{\circ}$ ) is available for the period 1895 – present, known as the Precipitation-elevation Regression on Independent Slopes Model (PRISM) data set. The PRISM method is a statistical interpolation approach, which utilizes point measurements in conjunction with digital elevation data. While originally developed for precipitation data, the method has been successfully applied to other climate variables such as temperature and snowfall [*Daly et al.*, 1994]. Data and additional information are available at Oregon State University’s Oregon Climate Service website<sup>3</sup>.

#### **4.2.4 Global Sea Surface Temperatures**

Monthly global sea surface temperature anomalies on a  $5^{\circ} \times 5^{\circ}$  scale for the period 1856-present were obtained via the Lamont-Doherty Earth Observatory/International Research Institute website data library<sup>4</sup>. The data are produced from a ‘reduced space’ optimal smoother algorithm applied to global SST monthly anomalies obtained from the U.K. Hadley Center archives [*Bottomley et al.*, 1990; *Kaplan et al.*, 1998]. This data set is regularly updated and used extensively in climate diagnostics studies.

---

<sup>2</sup> <http://www.esrl.noaa.gov/psd/data/timeseries/AMO/>

<sup>3</sup> <http://www.prism.oregonstate.edu/>

<sup>4</sup> <http://portal.iri.columbia.edu/portal/server.pt>

#### 4.2.5 Paleo Reconstructed Data

Paleo reconstructed streamflow and climate data from tree-rings were employed in this work. Specifically, water year (Oct – Sep) streamflow at Lees Ferry, AZ [Woodhouse *et al.*, 2006] spanning the period 1490 – 1997 and a reconstructed Atlantic Multi-decadal Oscillation [Gray *et al.*, 2004] covering the period 1572 – 1985 were obtained from NOAA’s paleoclimatology web page<sup>5</sup>. The reconstruction process typically involves fitting a regression model to the tree ring data and the variable of interest (e.g., streamflow, AMO index, etc.) during the modern period when both data sets are available. The fitted regression model is then applied to tree ring data of the pre-observation period to obtain reconstructed estimates of the desired variables. The streamflow reconstruction was derived from tree-ring chronologies in the Colorado Basin, while the AMO reconstruction was developed with records from eastern North America, Western Europe, Scandinavia and the Middle East. For detailed discussion of the reconstruction methods, the readers are referred to the aforementioned references and others therein.

#### 4.3 Methods

The objective of this work was to obtain further insight into the source and characteristics of UCRB flow variability; thus spectral analyses were first performed. Specifically, wavelet analyses of streamflow, temperature and precipitation data were used to identify dominant modes of variability in the hydro-climate of the Upper Colorado River Basin. To highlight and isolate the impact of various factors on streamflow, a combination of wavelet band-pass filtering and simple linear regression were employed. Subsequently, the component time series were correlated with global sea surface temperature data to aid in identifying the drivers of the

---

<sup>5</sup> <http://www.ncdc.noaa.gov/paleo/paleo.html>

variability. Last, paleo reconstructed data were analyzed to investigate persistence of climatic ties and long-term fluctuations of the dominant periodicities.

#### **4.3.1 Wavelet Spectral Analysis**

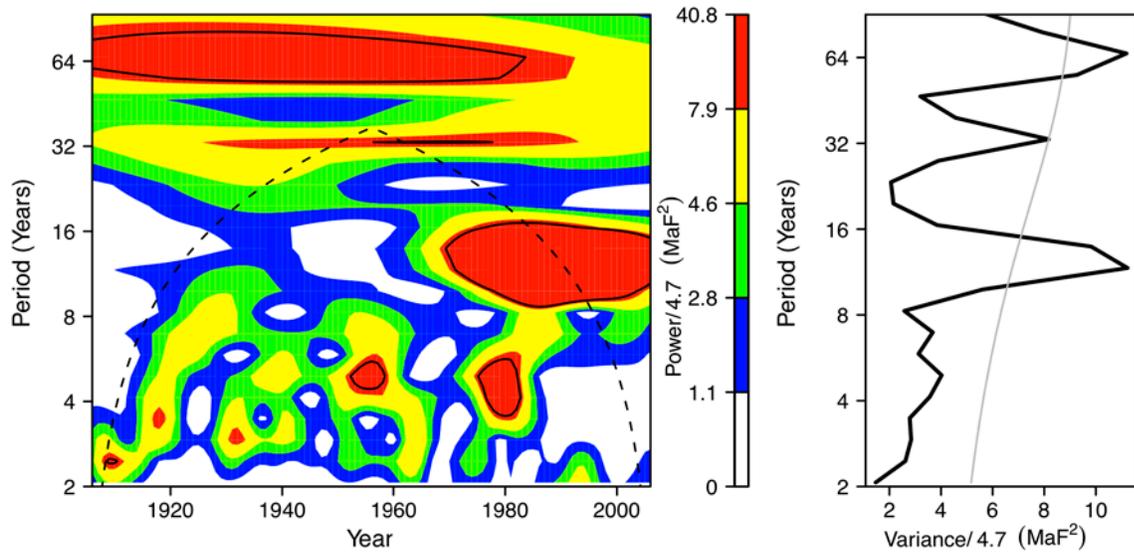
The use of wavelet spectral analysis to identify dominant spectral peaks and their temporal variations in geophysical data is increasingly popular [Addison, 2002; Kwon *et al.*, 2009; Torrence and Compo, 1998; Torrence and Webster, 1998]. Wavelet methods have several advantages over traditional spectral estimation techniques, such as their ability to effectively capture ‘local’ features resulting from amplitude modulation. Additionally, they are efficient at revealing quasi-periodic signals in data in the presence of considerable noise, which is often the case with geophysical data. In this work, several wavelet-based methods were employed, a brief overview of these methods is provided below.

For the purpose of investigating modes of variability, the local and global wavelet power spectra were analyzed. The local power spectrum shows variability in the two dimensional, frequency-time space. By averaging across the temporal domain at each scale, the global power spectrum is computed. For both the local and global spectra, a 90% significance white noise background spectra level was employed. Additionally, a type of wavelet band pass filtering was used to generate scale-specific components of the original time series. In simplest terms, this is a back-transformation of the wavelet-transformed data to the original space over a limited range of variability scales. For a detailed discussion of the wavelet methods, the reader is referred to Chapter 3 of this work.

## 4.4 Results

### 4.4.1 Dominant Modes of Variability

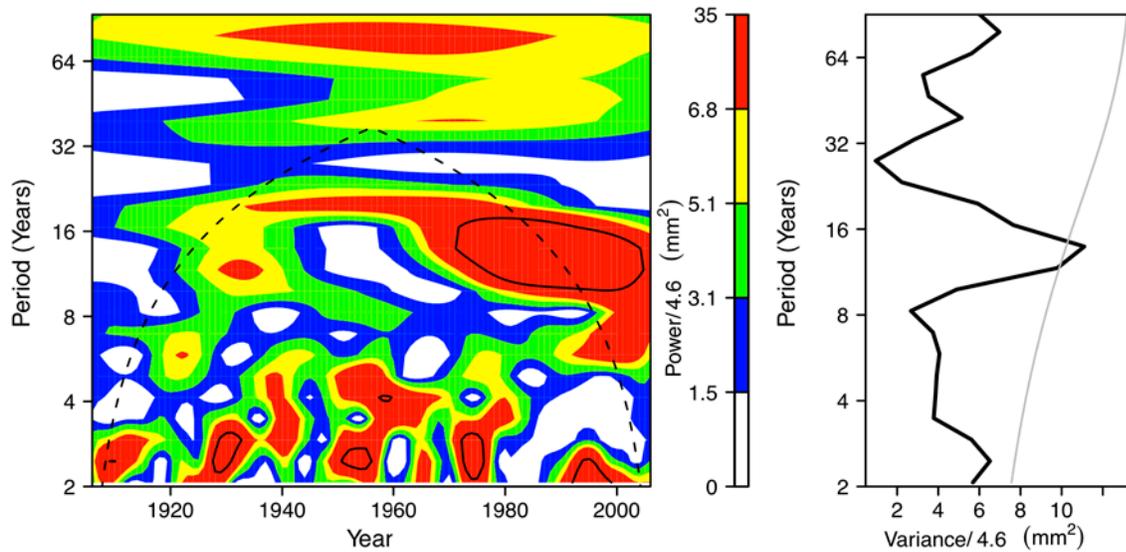
To begin, the global and local wavelet power spectra were computed for annual, water year flow at Lees Ferry, AZ (Figure 35). This location has two-fold significance; it serves as the delineation point between the Upper and Lower Basins per the 1922 Colorado River Compact [*Colorado River Commission. et al.*, 1923] and is the hydrologic aggregate for over 90% of the total Basin flow. Significant features in both the global and local spectra were identified against a white noise background spectrum at the 90% confidence level. Additionally, the cone of influence is shown in the local spectrum (dashed line); points beyond this region are influenced by boundary effects during the wavelet transform and should be interpreted with caution [*Torrence and Compo*, 1998]. From visual inspection of the global spectrum, there are two obvious, significant features – a peak in the 8-16 year period and a lower frequency (~64 year period) feature. An assessment of the local spectrum indicated that the low frequency variability is persistent throughout the time domain, while the “decadal” feature is only active in the most recent 30 years.



**Figure 35 Local wavelet power spectrum (left) and global wavelet power spectrum (right) of Lees Ferry flow.**

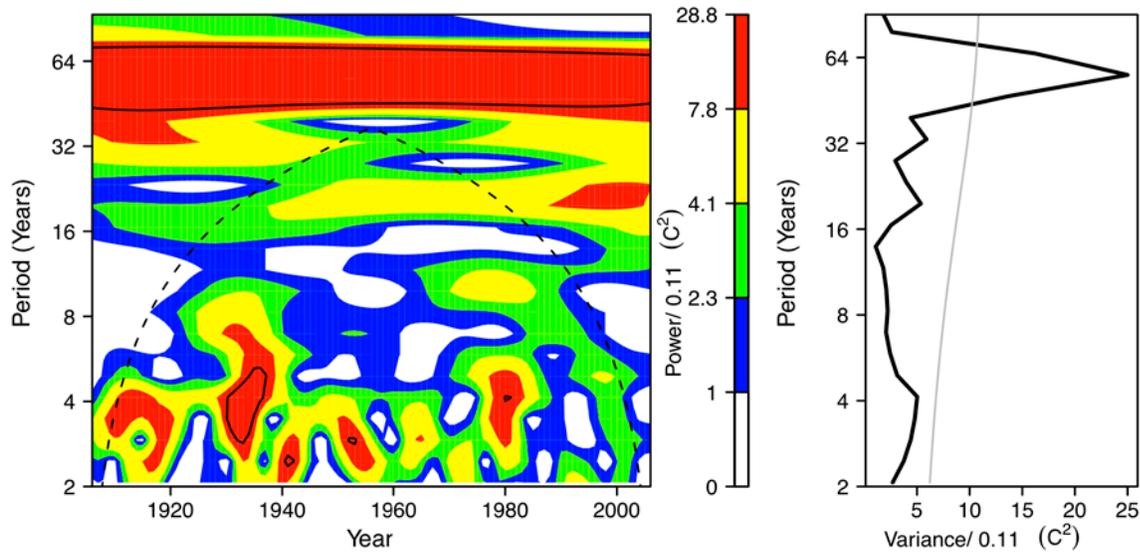
**Features significant against a 90% white noise background spectrum are encircled in black for the local spectra. In the global spectrum, the grey line is a 90% confidence level against a white noise background spectrum.**

In order to better understand the source of these variability modes, the same wavelet spectral analysis was performed for Upper Colorado River Basin temperature and precipitation data. Figure 36 shows the global and local wavelet power spectra for PRISM Upper Basin precipitation data. The major feature is power in the 8-16 year period, which is only active in the most recent 3 or 4 decades, and is strikingly similar to the decadal variability seen in the streamflow data.



**Figure 36 Local wavelet power spectrum (left) and global wavelet spectrum (right) of Upper Colorado River Basin PRISM annual water year precipitation data. Features significant against a 90% white noise background spectrum are encircled in black for the local spectra. In the global spectrum, the grey line is a 90% confidence level against a white noise background spectrum.**

As a compliment, Figure 37 shows the wavelet power spectra for Upper Basin PRISM temperature data. These data are characterized by persistent, low frequency (~64 year period) variability, and bear considerable similarity to the low frequency component of the streamflow data. Thus, subsequent analyses were based on the hypothesis that low frequency variability in Colorado River flow is associated with temperature and recent decadal variability with moisture delivery.

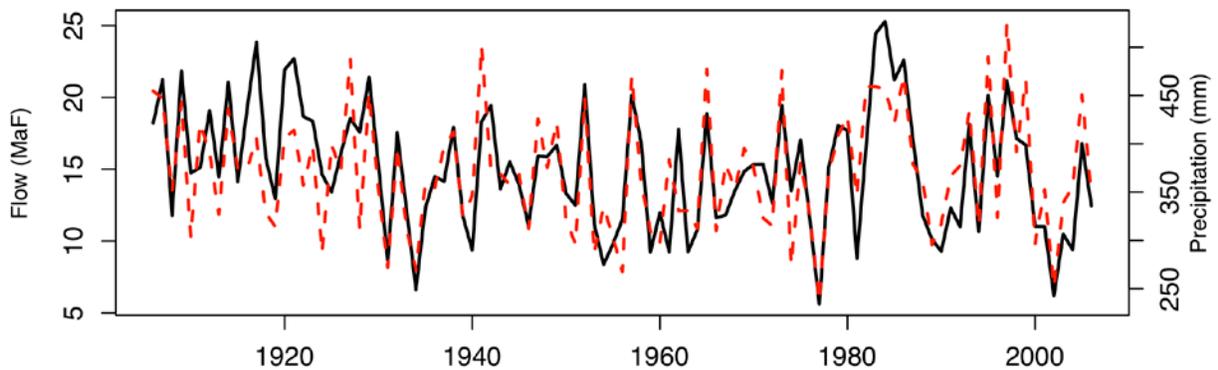


**Figure 37 Local wavelet power spectrum (left) and global wavelet spectrum (right) of Upper Colorado River Basin PRISM annual water year temperature data. Features significant against a 90% white noise background spectrum are encircled in black for the local spectra. In the global spectrum, the grey line is a 90% confidence level against a white noise background spectrum.**

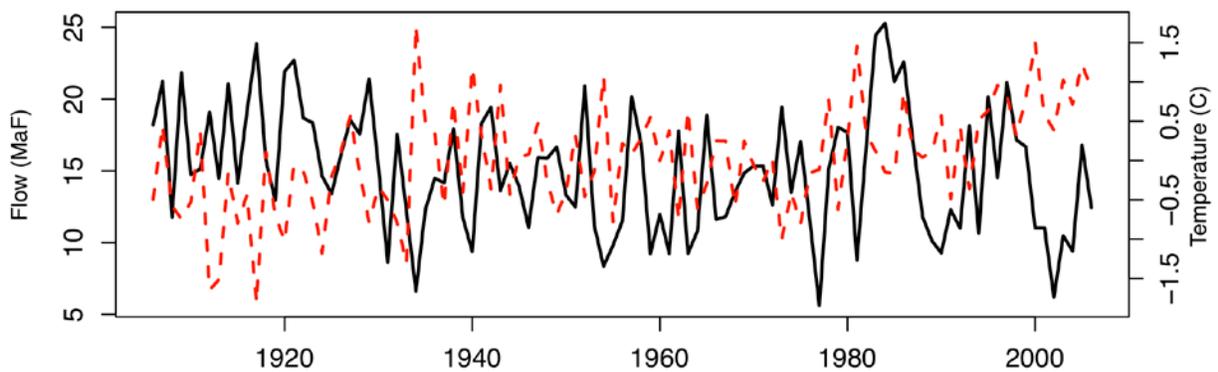
While investigating the link between precipitation and temperature with flow, it is important to recognize that observational temperature and precipitation data are not independent. Surface moisture availability is a key factor in the hydrologic energy balance. In simplest terms, evaporation has a cooling effect due to the energy expenditure associated with latent heat of vaporization. In a dry year, the evaporation component of the hydrologic cycle is moisture limited, and excess energy results in additional surface warming. Likewise, relative to years with below average precipitation, wet years tend to be cooler due to enhanced evaporation. Thus, to remove this interdependency, a linear regression was fit between temperature and precipitation. Residuals from the model were retained as temperature data independent of precipitation. Results indicated that 11% of the temperature variance is explained by precipitation. For all subsequent analyses and results, the temperature data employed have the influence of precipitation removed,

so as to avoid confounding the influences of these two variables and are referred to as “residual temperature.”

To highlight the relationship between flow, residual temperature and precipitation, time series plots are provided (Figure 38 and Figure 39). Precipitation and flow have a strong positive relationship (correlation = 0.77), while residual temperature has a slightly weaker and negative association (correlation = -0.32).

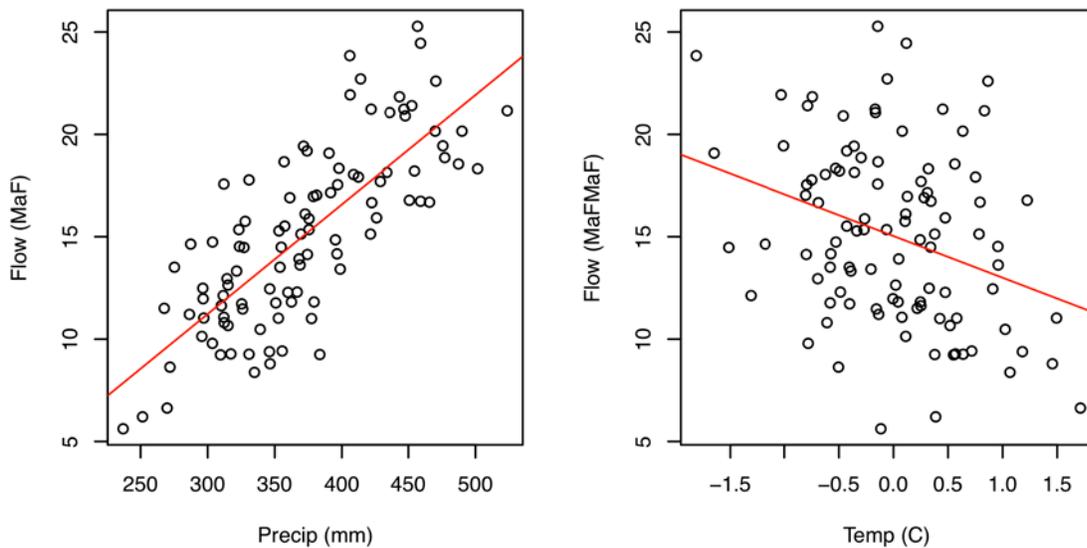


**Figure 38** Lees Ferry annual water year streamflow data (solid black line) with Upper Colorado River Basin annual water year PRISM precipitation data.



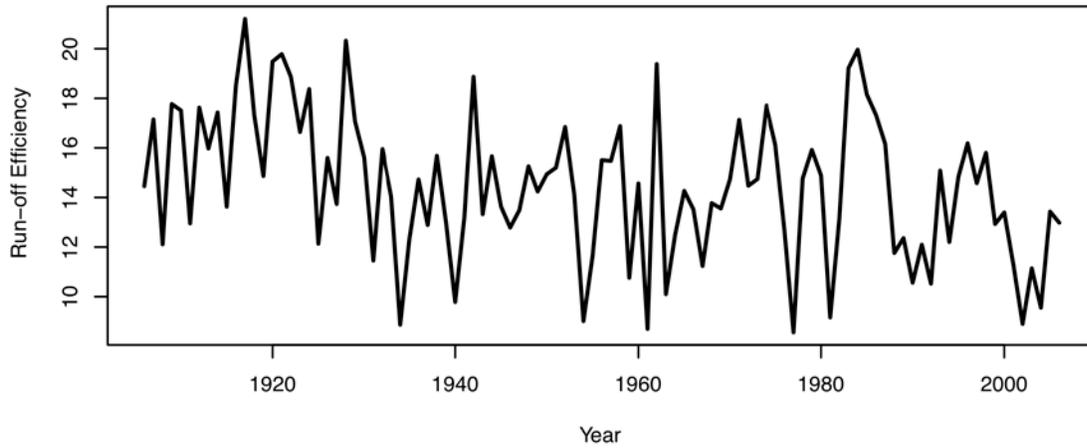
**Figure 39** Lees Ferry annual water year streamflow data (solid black line) with Upper Colorado River Basin annual water year residual temperature anomalies.

Figure 40 offers scatter plots of precipitation vs. flow and residual temperature vs. flow as additional evidence of the respective relationships. These links are quite significant; in fact a simple multivariate linear regression using residual temperature and precipitation as dependent variables explains about 70% of the flow variability.

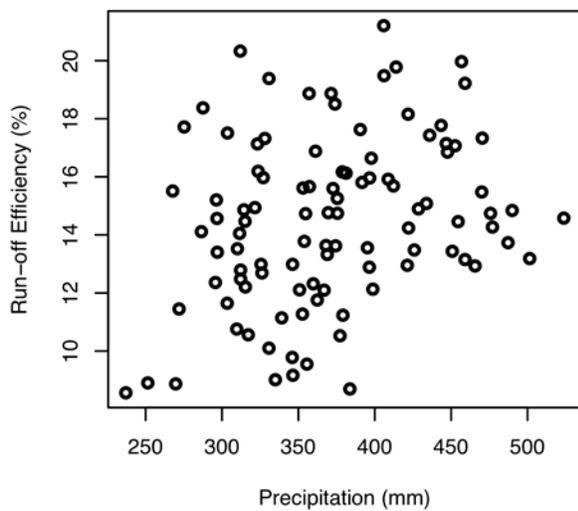


**Figure 40 (Left) Flow vs. precipitation scatter plot with best-fit linear regression. (Right) Flow vs. residual temperature anomalies scatter plot with best-fit linear regression.**

From a mass balance perspective, precipitation represents the maximum potential run-off (assuming 100% basin run-off efficiency) and residual temperature likely contributes to modulation of run-off efficiency. To explore this further, annual UCRB run-off efficiency was computed as annual flow volume divided by annual precipitation volume. This time series is shown in Figure 41. From analysis of these data, a weak tie to precipitation was seen (correlation = 0.29), which is likely due to a few “outlier” precipitation years. In these instances, an especially below average precipitation year can cause efficiency to suffer. However, for most years, including particularly wet ones, precipitation magnitude has little to no link with run-off efficiency (Figure 42).



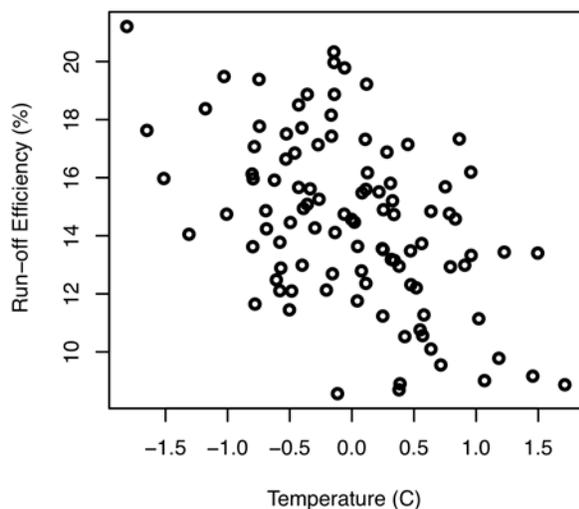
**Figure 41 Upper Colorado River Basin run-off efficiency time series.**



**Figure 42 Run-off efficiency vs. precipitation scatter plot.**

Residual temperature, on the other hand, has a much stronger and robust relationship with run-off efficiency (correlation = -0.49). As shown in Figure 43, efficiency steadily declines with increasing residual temperature. Furthermore, the slope from a linear regression fit to data in Figure 43 indicates that per degree warming, run-off percent efficiency decreases by 2. Thus, assuming a Basin average efficiency of 14.5%, one degree warming should be expected to

decrease annual streamflow by 13.8%, or approximately 2 MaF. A number of previous studies have investigated the impact of temperature on streamflow and run-off efficiency across the United States, many focusing on the Southwest and Colorado River Basin [*Christensen and Lettenmaier, 2007; Langbein, 1949; Nash and Gleick, 1991; Revelle and Waggoner, 1983*]. While the results range somewhat from study to study, the findings presented here are generally consistent with previous work. All of these investigations suggest that warming conditions will result in either reduced streamflow or reduced run-off efficiency. Thus, it should be quite evident that understanding temperature and precipitation variability modes is critical to the overall hydrology in the Upper Colorado River Basin.

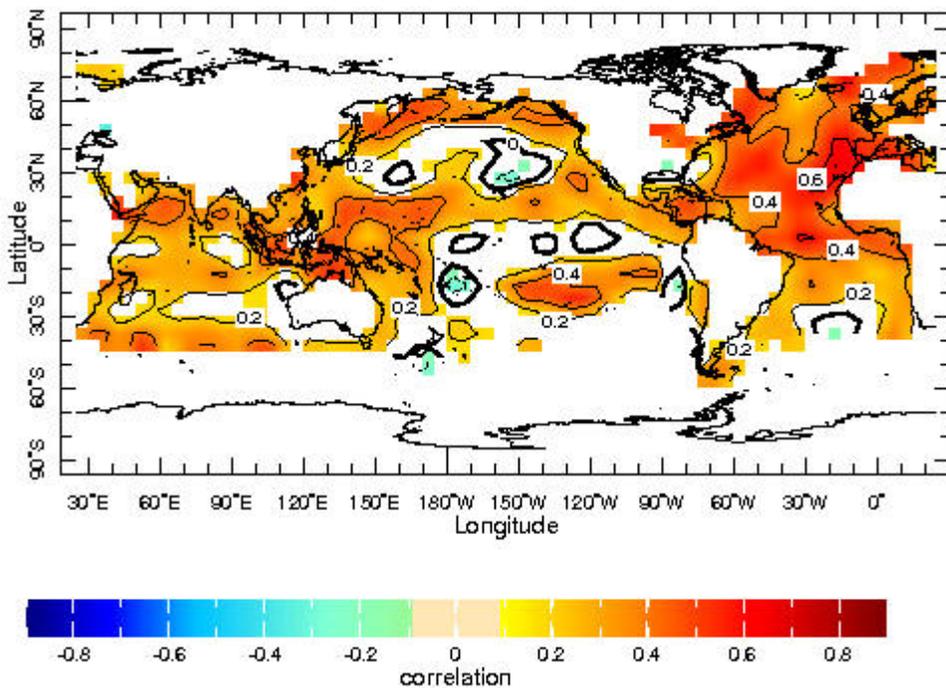


**Figure 43 Run-off efficiency vs. residual temperature anomalies scatter plot.**

#### **4.4.2 Links to Large Scale Climate Forcings: Low Frequency Mode**

To help understand the links between UCRB hydro-climatic variability and large-scale climate forcings, the previously identified major drivers of flow (water year temperature and precipitation) were correlated with winter season global sea surface temperature (SST)

anomalies<sup>6</sup>. Prior to the correlation analysis, residual temperature and precipitation data were filtered using wavelet methods to highlight their respective key variability scales. The residual temperature correlation was first performed using a low pass filter (periods > 32 years), and results are shown in Figure 44. Regions of correlation significant at the 90% confidence level are colored. This figure shows significant correlation in the North Atlantic and portions of the Pacific Ocean. This pattern is distinctly reminiscent of the Atlantic Multi-Decadal Oscillation (AMO) [Enfield et al., 2001]. Emerging research suggests an association between AMO and western U.S. hydroclimate [Hidalgo, 2004; McCabe et al., 2008; McCabe et al., 2007], raising the question about hopeful prospects for skillful long-lead projection of streamflow.

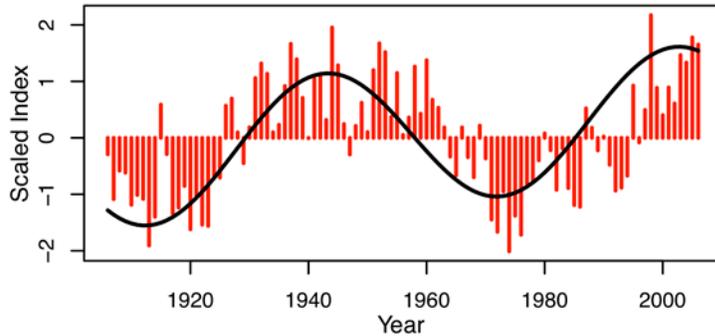


**Figure 44 Upper Colorado River Basin low pass filtered residual temperature data (1906 – 2006) correlated with Kaplan global sea surface temperature anomalies. Colored regions are significant at the 90% confidence level (correlation > 0.16).**

To further highlight this correlation, Figure 45 shows a scaled AMO index plotted with scaled low pass wavelet filtered UCRB residual temperature data. While the low frequency

<sup>6</sup> <http://iridl.ldeo.columbia.edu/index.html>

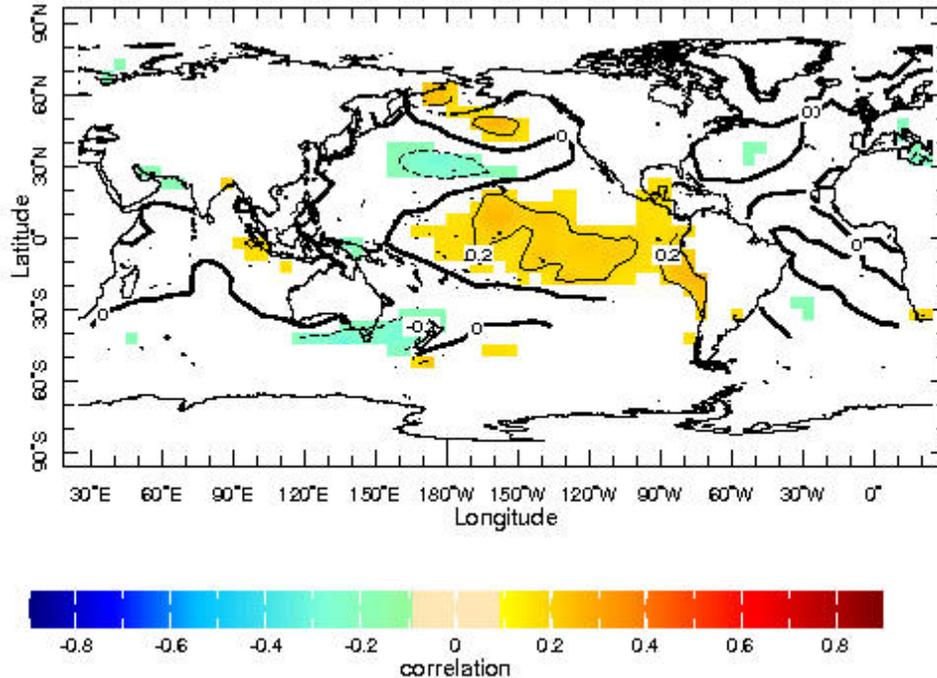
nature of the signal offers less than 2 cycles over the period of record, the tracking of the two data sets is quite striking. McCabe et al. (2008) showed similar covariance between the AMO and temperature variations across much of the continental United States.



**Figure 45 Scaled Atlantic Multi-decadal Oscillation index (vertical red bars) with scaled low frequency band pass filtered Upper Colorado River residual temperature data (solid black line).**

#### **4.4.3 Links to Large Scale Climate Forcings: Decadal Mode**

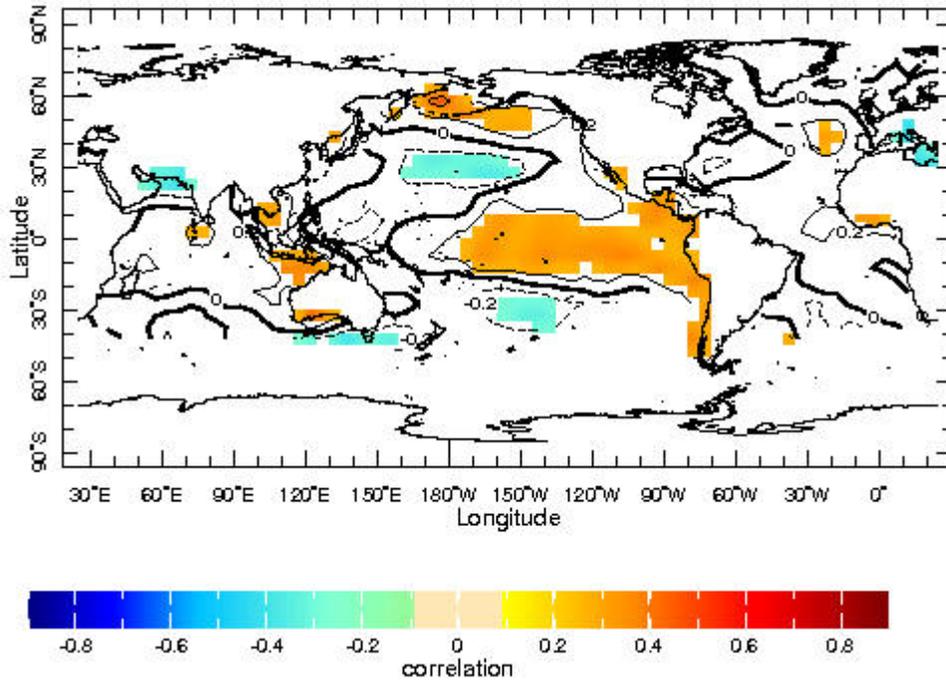
The same SST correlation was performed with the UCRB precipitation data with a high pass wavelet filter (periods < 20 years). It is well established that moisture delivery to the western United States is Pacific in origin. From Figure 46, the majority of regions with significant correlation are found in the Pacific, albeit relatively weak in magnitude and sparse in extent. The most noticeable feature in the map is a relatively large area of positive correlation in the equatorial Pacific, reminiscent of the El Niño Southern Oscillation (ENSO) pattern. This is not surprising; as previously discussed, there is a wealth of literature linking this phenomenon with the hydro-climate of the western United States.



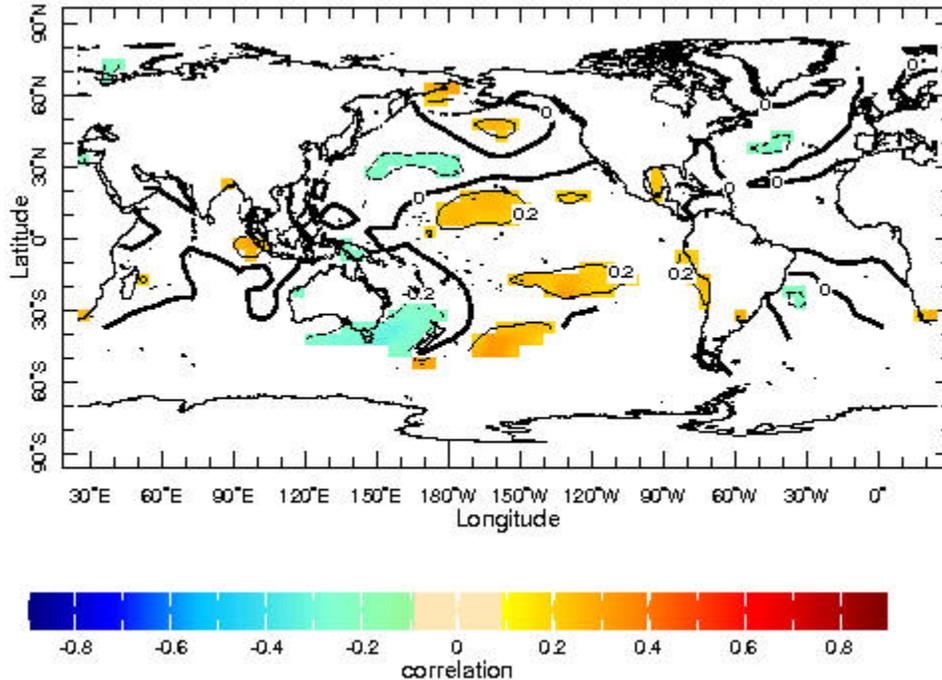
**Figure 46** Upper Colorado River Basin high pass filtered annual precipitation data (1906-2006) correlated with global Kaplan sea surface temperature data. Colored regions are significant at the 90% confidence level (correlation > 0.16).

Given the substantial non-stationary decadal variance seen in both precipitation and flow data, streamflow wavelet power spectra are revisited. Recall that the majority of precipitation-associated variability is seen in recent decades and not persistent throughout the time domain, in direct contrast with the lower frequency temperature variability. Thus, the SST/precipitation correlation was repeated for the period 1970-2006. The resulting correlation map shows heightened Pacific correlation – bearing semblance to ENSO and possibly the Pacific Decadal Oscillation (PDO) [*Diaz and Markgraf, 2000; Mantua and Hare, 2002; Mantua et al., 1997; White and Cayan, 2000*] (Figure 47). It should be noted that due to the shortened length of the sub-period, 90% significant correlation values are greater than 0.26, whereas for the entire period of record this value is 0.16. As a complimentary plot, Figure 48 shows the high pass filtered temperature data correlated with SST data for the period 1906-1969. There are few regions with

significant values, and they appear to be quite random. Suffice it to say, any link between UCRB precipitation and the equatorial Pacific is almost exclusively from the past ~30 years.



**Figure 47** Upper Colorado River Basin high pass filtered annual precipitation data (1970-2006) correlated with global Kaplan sea surface temperature data. Colored regions are significant at the 90% confidence level (correlation > 0.26).

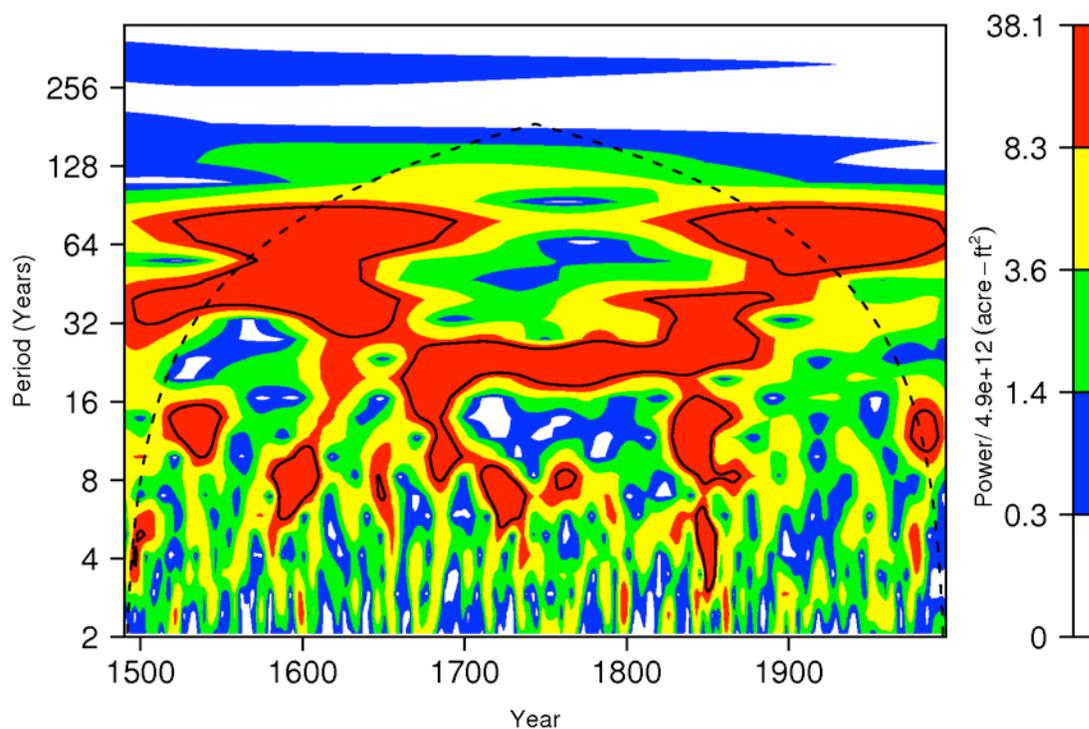


**Figure 48** Upper Colorado River Basin high pass filtered annual precipitation data (1906-1969) correlated with global Kaplan sea surface temperature data. Colored regions are significant at the 90% confidence level (correlation > 0.2).

Interestingly, there is strong evidence of heightened activity in the Pacific beginning in the early to mid-1970s [Graham, 1994; Hare and Mantua, 2000; Miller et al., 1994; Trenberth and Hurrell, 1994]. This period is characterized by a shift in ENSO characteristic, a strong, positive PDO regime and a persistent negative North Pacific Index (NPI) [An and Wang, 2000; Diaz et al., 2001; Latif and Barnett, 1994; Rajagopalan et al., 1997; Torrence and Webster, 1999; Wang, 1995]. Furthermore, post-1975, precipitation variance in the UCRB increased by 50% relative to pre-1975 period, and the relationship between flow and precipitation strengthened post 1975. For the period prior to 1975, precipitation explains 55% of the flow variance, while post 1975 it explains over 70%. In light of this, understanding duration and frequency of periods characterized by heightened Pacific activity, and thus precipitation, is of great interest.

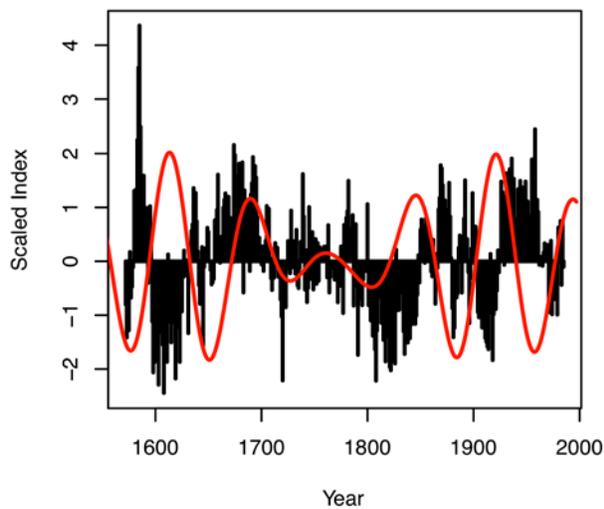
#### **4.4.4 Temporal Variability over Paleo Time Period**

Confident detection and attribution of spectral features, particularly lower frequency (such as the ~64-year period) can be challenging due to limited data, boundary effects and fewer degrees of freedom. In the case of the recent period of strong decadal variance, it is difficult to know if this is an anomalous epoch or a feature that waxes and wanes over time. Paleo reconstructed data covering a longer period provide a good source to further investigate these uncertainties. The Woodhouse et al. (2006) reconstruction of Upper Basin streamflow was utilized. The wavelet spectrum of this paleo streamflow (Figure 49) shows similar spectral signals as that of the dominant mode of streamflow (Figure 35). In particular, two interesting features: (i) the significant power of an ~64-year period is robust over the entire duration except during 1700-1800 when it exhibits an ~32-year period, and (ii) the significant decadal signal (~8-16 year period) exhibits epochal behavior.



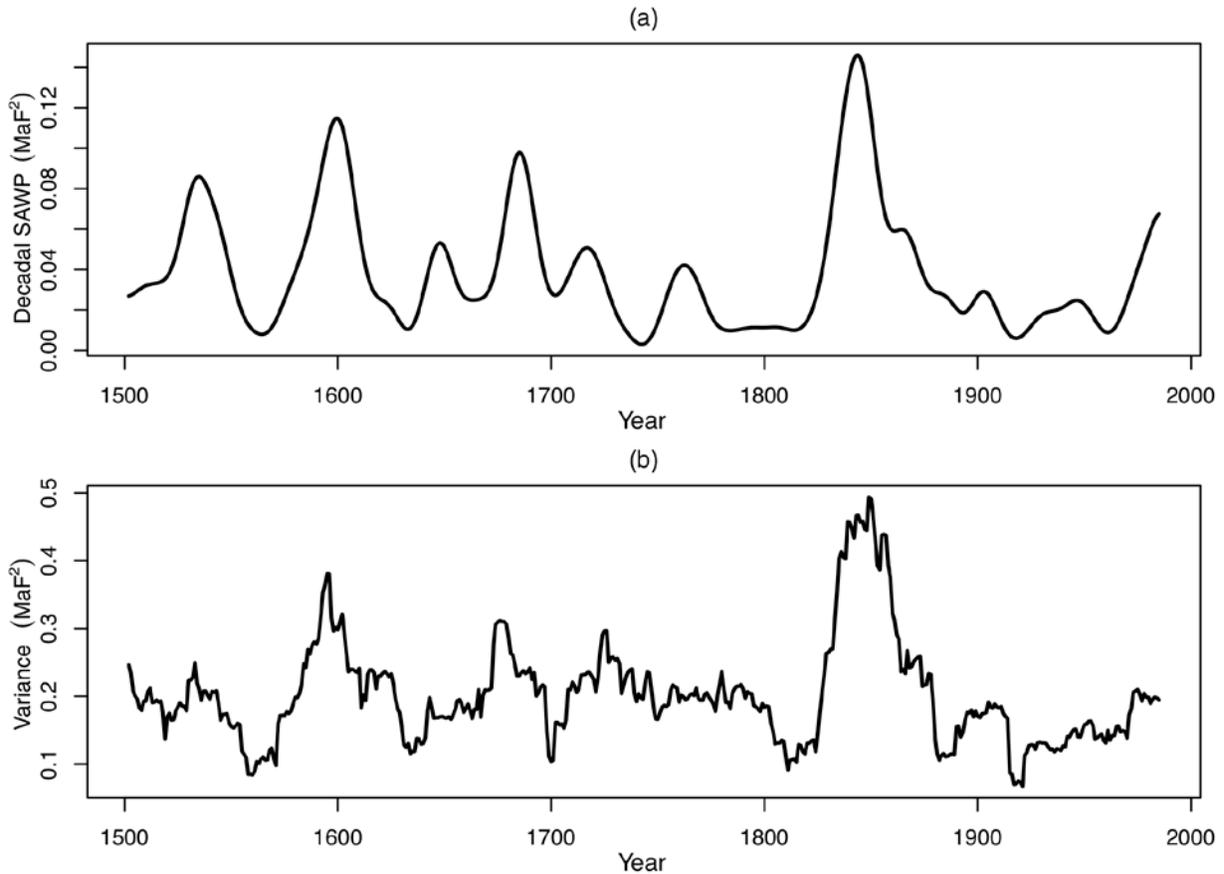
**Figure 49** Local wavelet power spectrum for Woodhouse et al., 2006 Lees Ferry paleo reconstructed streamflow. Features significant against a white noise background spectrum at the 90% level are encircled in black.

Low frequency reconstructions of the paleo streamflow in the 64-80 year band were computed and shown with the paleo AMO index [Gray et al., 2004] in Figure 50. They co-vary with an inverse relationship reasonably well over the entire period, with increased amplitudes in the recent centuries and a general weakening during the 18<sup>th</sup> century. This provides further corroboration of the low frequency covariance of streamflow with AMO identified earlier in this research and by others [Hunter et al., 2006; McCabe et al., 2008].



**Figure 50 Scaled, reconstructed AMO index from Gray et al., 2005 (black vertical bars) and scaled low pass filtered Lees Ferry streamflow data from Woodhouse et al., 2006 (red line).**

To investigate the temporal variability of the decadal band ( $\sim 8 - 16$  year period) signal, the Scale Averaged Wavelet Power (SAWP) was computed, which represents the strength of this signal over the time domain (Figure 51a). The epochal nature of the variability is quite apparent, and a 30-year moving window variance of the paleo streamflow (Figure 51b) shows that epochs with enhanced SAWP are consistent with increased overall streamflow variance (correlation = 0.75). The recent decades appear to be one such epoch.



**Figure 51 (a) Decadal (8-16 year) band Scale-Averaged Wavelet Power (SAWP) from the paleo reconstructed Lees Ferry streamflow. (b) 25 year variance moving window of paleo reconstructed Lees Ferry streamflow.**

A global spectrum of the SAWP (Figure 52) shows two distinct peaks, suggesting that the decadal variability signal tends to be modulated at an  $\sim 75$  and  $\sim 150$  year period. Perhaps only coincidence, it is worth noting that the decadal variability strength is modulated at approximately the same timescale as the overall low frequency component (64-80 year period). It is possible that there may be some interplay between these components, but without additional work this is largely speculative. The  $\sim 150$  year peak is more difficult to attribute and may be largely due to the magnitude of the peak in decadal strength around 1850.

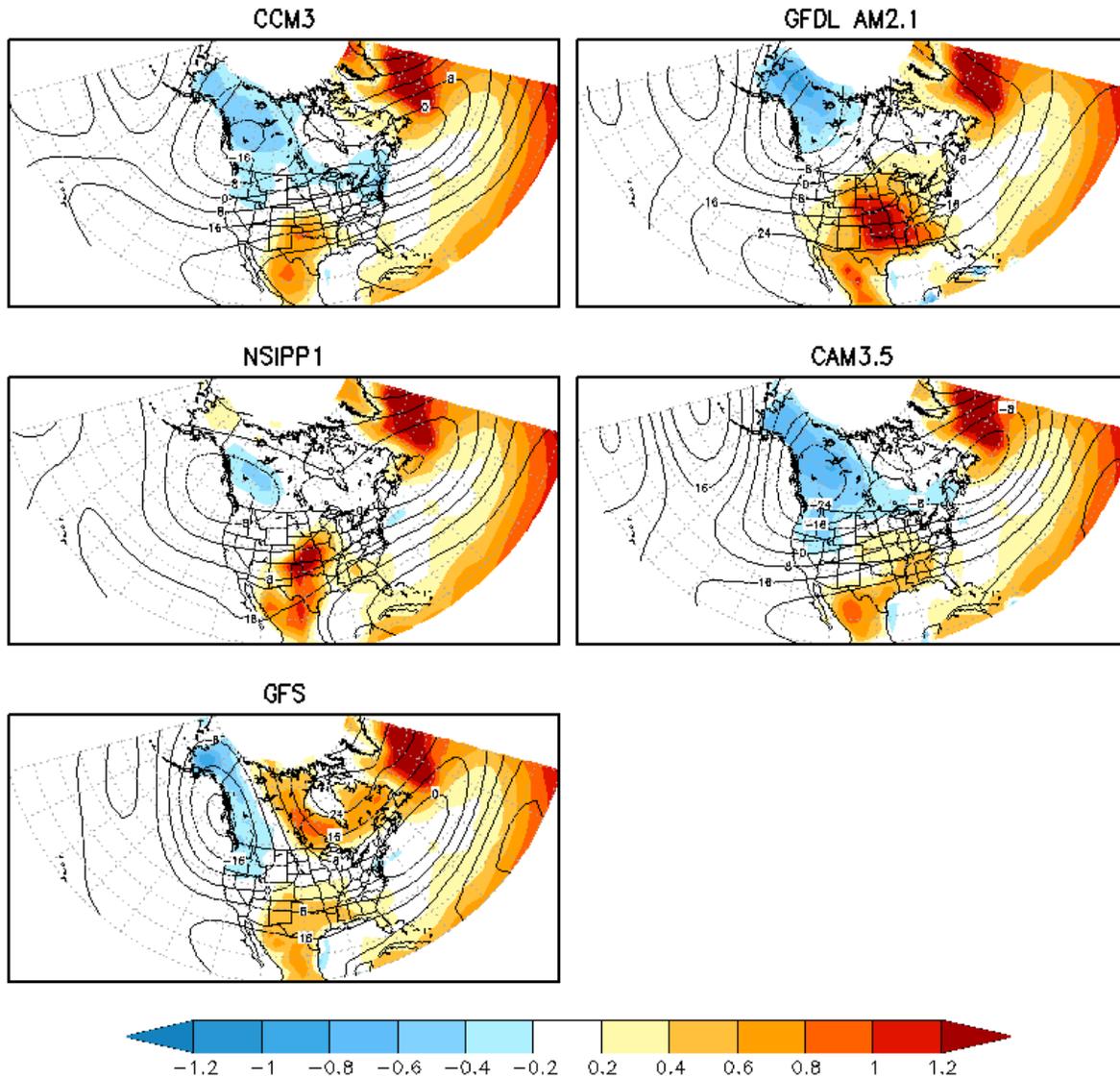


**Figure 52 Global wavelet spectrum of the decadal SAWP.**

#### **4.4.5 Additional Climate Model Results**

To further investigate the low frequency relationship between Atlantic sea surface temperatures and the Colorado River Basin, results from idealized climate model experiments are presented. Schubert et al. (2009) used five general circulation models (GCMs) to examine the impact of large-scale sea surface temperature patterns on drought in the United States. Specifically, the National Center for Atmospheric Research’s Community Climate Model version 3.0 (CCM3) and Community Atmospheric Model version 3.5 (CAM3.5) were both used in addition to NASA’s Seasonal-to-Interannual Prediction Project model version 1 (NSIPP-1). Also, NOAA’s Global Forecast System model (GFS) and Atmospheric Model version 2.1 (AM2.1) were employed. For model runs forced with a neutral Pacific and warm Atlantic Ocean pattern, four of the five GCMs showed above average annual temperatures for the continental U.S., with these anomalies ranging from 0.15°C to 0.4°C. The Atlantic SST forcing pattern for these experiments (not shown) strongly resembles the positive phase of the Atlantic Multi-decadal Oscillation. Figure 53 shows additional results based on the original work of Schubert et al. (2009), highlighting the impact of a warm versus cool Atlantic Ocean on annual temperature

and 200mb geo-potential height anomalies. Warming is most elevated across the Great Plains region, but moderate increases are noted for the Colorado Basin headwaters. These results offer some support for the hypothesis that the AMO may have a hand in UCRB low frequency temperature variance and consequently run-off efficiency. However, additional work is needed to establish the explicit physical mechanism of such a relationship.



**Figure 53 Annual temperature anomalies (°C) (color bar) and 200mb geo-potential height anomalies (contours) from differencing results of five GCMs forced first with a warm Atlantic/neutral Pacific and then a cool Atlantic/neutral Pacific.**

## 4.5 Summary

An analysis of UCRB hydrologic variability was performed to better understand the dominant modes of streamflow variability. Wavelet spectral analysis of annual Lees Ferry streamflow indicated two dominant variability scales – a persistent low frequency component (~64 year period) and a non-stationary decadal signal (8-16 year period). Subsequent spectral analysis suggested that the low frequency component of flow is associated more with temperature variability, while the higher frequency is associated more with precipitation variability.

A strong decadal signal was seen in recent decades for both flow and precipitation, and was quite weak for the first two-thirds of the historical record. Correlation of high pass filtered precipitation data with global sea surface temperatures over the entire record (1906-2006) indicated a weak, but significant tie to the equatorial Pacific. The same analysis was performed for the period of high decadal variability (1970-2006) and showed a heightened equatorial Pacific link, while there was little to report for 1906-1969. Enhanced decadal variability in precipitation and streamflow coincided with well-documented shifts in Pacific Ocean characteristics; PDO switched from a period of negative to positive phase, while ENSO had a tendency toward stronger, persistent El Niño events and weaker, less frequent La Niñas.

Temperature oscillations on the 64-year period scale were shown to modulate run-off efficiency. Per degree warming, a 14% decrease in UCRB discharge was implied by the diagnosis of historical Lees Ferry streamflow and climate data. From correlation of low pass filtered temperature data with global SST data, a strong link to the Atlantic was noted. Further investigation showed distinct covariance with the AMO index, also supported by results from climate model experiments. As no causal relationship has been established at this time, further research is needed to assess the reliability of the aforementioned association. Specifically,

identification of a physical mechanism by which the Colorado River Basin is influenced by Atlantic SSTs will solidify the proposed connection. As anthropogenic climate change is all but certain to warm the UCRB, continued study of this variability mode will be critical for understanding future flow characteristics, in terms of both magnitude and variability.

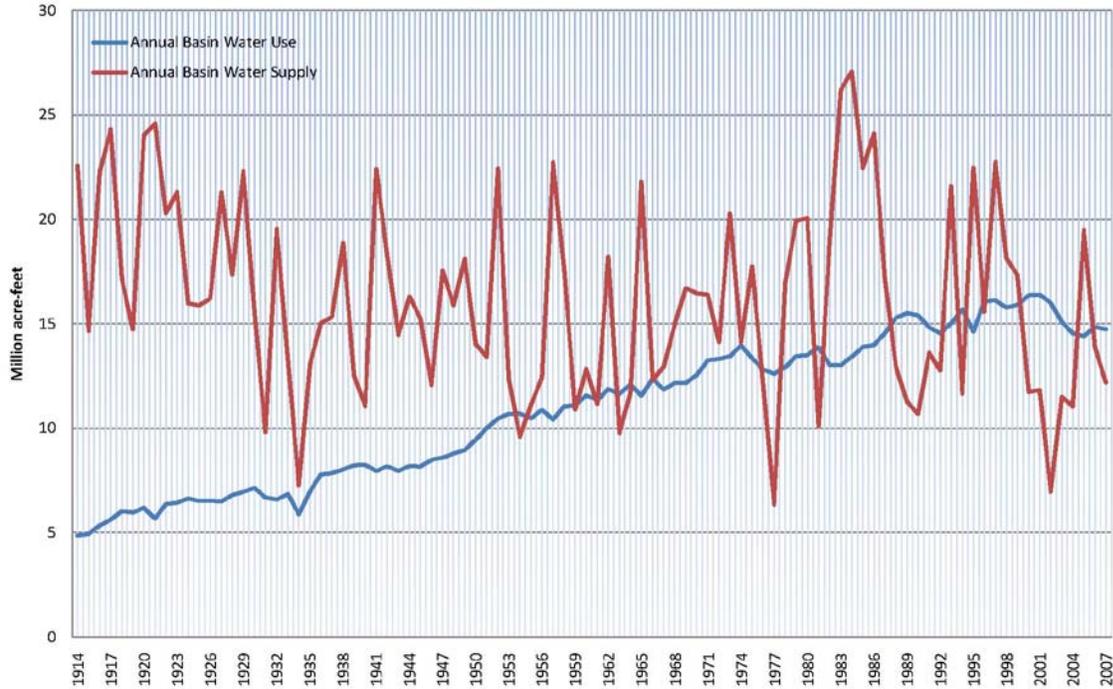
Last, paleo reconstructed streamflow at Lees Ferry were analyzed to gain insight to the persistence and strength of variability features identified over the historical period. Decadal variance appears to wax and wane in strength at roughly a ~75-year time scale. Furthermore, the epochal variations of the strength of the decadal signal are consistent with epochal variations of overall streamflow variance. When the decadal signal is active, the streamflow exhibits enhanced variance. Low frequency variance was also seen throughout much of the paleo data, with a lull/shift to higher frequency variability in the 1700's. A comparison of the AMO reconstruction of Grey et al., 2004 and low pass filtered paleo flow at Lees Ferry showed reasonable history of the inverse flow/AMO relationship from the observed period.

In closing, many operational and planning decisions for water resources management are at the multi-decadal horizon. Thus, robust flow scenarios at these timescales are important for providing accurate estimates of system risk and to devise optimal management strategies. The findings from this research could lead to improved stochastic streamflow projection techniques. Stochastic flow generation methods for near term (i.e. 2-3 decades) do not incorporate spectral characteristics, and consequently, the scenarios tend to provide an inaccurate estimate of system risk. Chapter 5 will address decadal projections for flow magnitude and variance in the Colorado River Basin.

## CHAPTER 5: AN APPROACH TO MULTI-DECADAL MEAN STREAMFLOW PROJECTION USING PALEO RECONSTRUCTED STREAMFLOW

### 5.1 Introduction

As early as the 1950's, demand in the Colorado River Basin has exceeded the annual flow of the river (Figure 54). However, the system has been able to reliably provide water despite significant inter-annual streamflow variability. This is largely due to a total system storage capacity roughly four times the mean annual flow. With such long-term storage, decadal scale flow regime changes are operationally relevant. This is in contrast with most other river systems that have limited storage capacities, thus prohibiting the consideration of such a horizon. As a result, most studies pertaining to simulation/projection of future hydrologic conditions tend to focus on the inter-annual time scale (1-2 years) [*Grantz et al.*, 2005; *Regonda et al.*, 2006]. These consider, for example, guiding year to year reservoir drawdown to better meet storage targets, providing effective flood control and accomplishing a myriad of other objectives. Additionally, there have been a number of studies assessing the impacts of climate change on water availability over the next 50 to 100 years in the Colorado River Basin [*Christensen and Lettenmaier*, 2007; *Milly et al.*, 2005; *Seager et al.*, 2007]. For the southwestern United States, an anthropogenic climate change induced streamflow reduction of 10-20% is anticipated over approximately the next 50 years [*Ray et al.*, 2008]. Undoubtedly, this is pertinent information for planning and sustainability in the region. However, there is a substantial gap in hydro-climatic research pertaining to the decadal to inter-decadal time scales.



**Figure 54 Annual Colorado River natural flow above Imperial Dam, AZ, (red line) and annual Basin consumptive use (blue line).**

While the impact of climate change is indeed relevant, consideration of natural variability is equally, if not more important. The annual standard deviation of Colorado River flow represents a 30% departure from the long-term average. Furthermore, 10 year moving window flow means vary from 82% (12.4 MaF) to 125% (18.9 MaF) of the full record mean. Such variations are likely to strain a reservoir system more than a slow flow reduction trend associated with anthropogenic climate change over a decadal horizon. In the Colorado Basin, the decadal horizon is quite relevant. As the system has a total storage roughly four times the mean annual flow of the river, the management outlook is considerably longer than many other basins. This underscores the potential utility of decadal scale hydro-climate information. Recent policies such as the Interim Surplus and Shortage Guidelines [United States Department of the Interior, 2001; , 2007b] are further evidence of the decadal horizon relevance in the Colorado River Basin; their implementation periods are 15 and 20 years respectively. Thus, decadal projections of similar

time scale are likely to be valuable when these guidelines are revisited or new polices enacted. The need for information at the decadal timescale has been widely recognized in recent years by the climate community and many efforts are underway to address the deficiency [*Keenlyside et al.*, 2008; *Meehl et al.*, 2009; *Mehta et al.*, 2011; *Solomon et al.*, 2011]. However, at present, there is little research that could be used to inform decision making in the Colorado Basin.

It is also important to note that the general circulation models (GCMs) used in climate studies are largely ineffective for explicit prediction of inter-annual to inter-decadal phenomena (e.g. El Niño Southern Oscillation, Atlantic Multi-decadal Oscillation, etc.), many of which have been identified as drivers of hydroclimatic variability [*Cayan et al.*, 1999; *Hidalgo and Dracup*, 2003; *McCabe et al.*, 2007; *McCabe and Dettinger*, 1999; *Piechota et al.*, 1997; *Timilsena et al.*, 2009]. Hence, alternate approaches are required for informing policy decisions at the decadal scale.

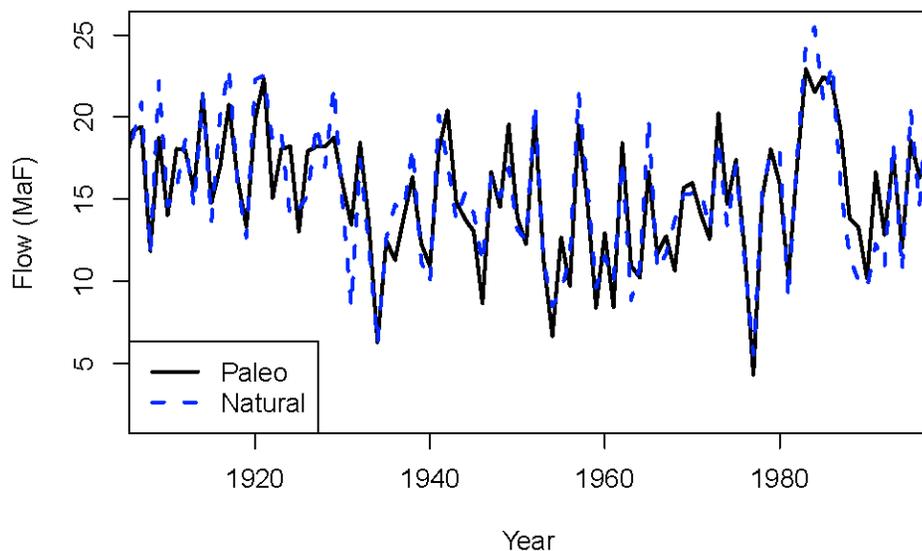
Given the need for robust projections of mean flow at multi-decadal time scales and the lack of existing methods, an approach is proposed in this chapter that uses paleo reconstructed flow at Lees Ferry, AZ on the Colorado River. Following the precedent set by the current Interim Shortage Guidelines length, the projection horizon in this approach was chosen to be 20 years. The utility of projection information is demonstrated in chapter 6 as part of potential water resources management in the Basin. The chapter is organized as follows: the data set used is first presented, followed by the proposed methodology and then results and discussion.

## **5.2 Data Set**

### **5.2.1 Paleo Reconstructed Streamflow Data**

Paleo reconstructions of water year (Oct–Sept) streamflow at Lees Ferry, AZ, from Woodhouse et al. [2006], spanning the period 1490 – 1997, were obtained from NOAA’s

Paleoclimatology web page<sup>7</sup> and used to develop the proposed methodology. Besides, for the overlapping observational period 1906 – 1997 the paleo reconstructions and natural flow [United States Department of the Interior, 2005] are virtually the same [Woodhouse et al., 2006](Figure 55). This reconstruction is quite robust and is widely used in past studies[Prairie et al., 2008; Rajagopalan et al., 2009; United States Department of the Interior, 2007b] .The reconstruction process typically involves fitting a regression model to the tree ring data and the variable of interest (e.g. streamflow) during the modern period when both data are available. The fitted regression model is then applied to tree ring data of the pre-observation period to obtain reconstructed estimates of the desired variables. The streamflow reconstruction was derived from select tree-ring chronologies throughout the Colorado River Basin. For a detailed discussion of the reconstruction methods the reader is referred to the aforementioned references and others therein.



**Figure 55 Annual Colorado River streamflow at Lees Ferry, AZ from the Woodhouse et al. (2006) paleo reconstruction (solid black line) and Bureau of Reclamation’s natural flow data set (blue dashed line).**

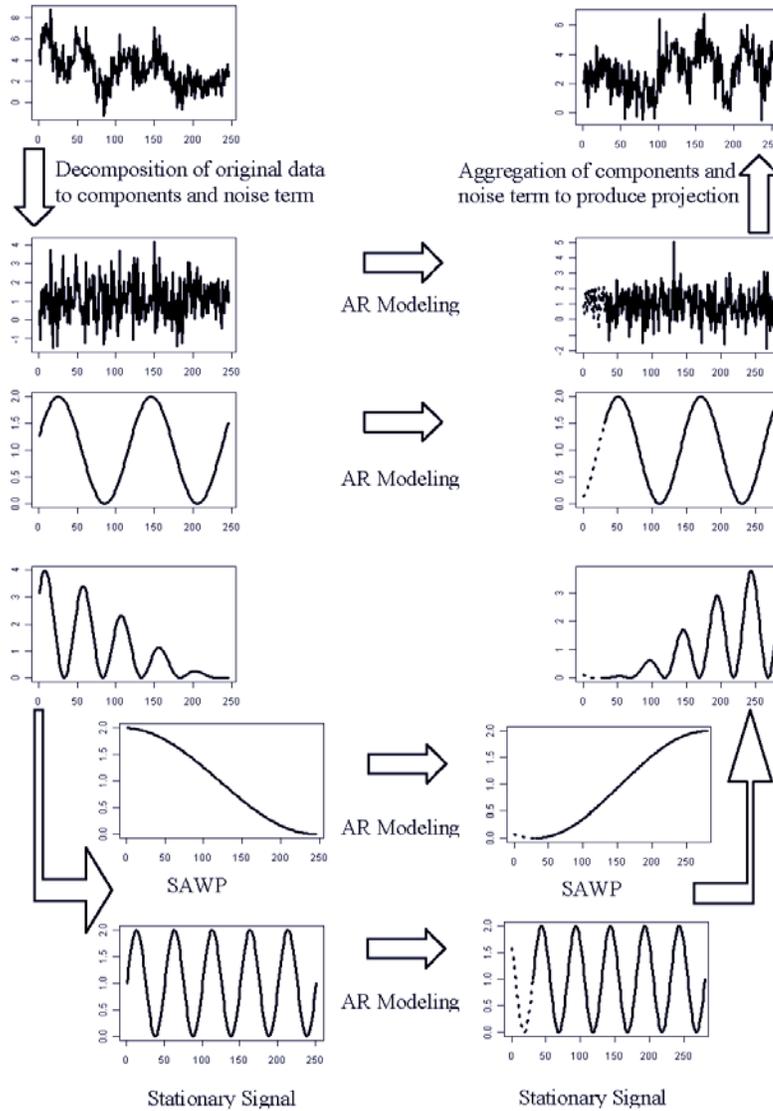
<sup>7</sup> <http://www.ncdc.noaa.gov/paleo/paleo.html>

## **5.3 Proposed Method**

### **5.3.1 Introduction**

As established in earlier chapters, the wavelet power spectrum of streamflow offers much information about the temporal evolution of dominant modes of variability. In the case of Lees Ferry flow, these modes are at the decadal (8-16 year period) and low frequency (~64 year period) time scales. (See Figure 35, chapter 4.) The “Improved Wavelet Auto-Regressive Model” proposed in Chapter 3 offers an effective framework for streamflow simulation that captures not only the global, but also local (or epochal) spectral features of a given data set. As a logical progression, utilizing this framework as the basis for projecting streamflow was considered.

The improved WARM approach described in chapter 3 was first considered from a “current conditions” starting point to project flows over a desired future horizon. To summarize this, auto-regressive (AR) models were fit to the significant spectral components (low frequency and decadal) and the noise bands based on historic data. These models were then evolved forward based on the current phase of the respective signals to offer a projection instead of a simulation. An AR model was also fitted to the scale averaged wavelet power (SAWP) of the decadal band, which exhibited a strong non-stationarity and also evolved forward in the same manner. The projections of the spectral components from the corresponding AR model were combined with the projected evolution of the SAWP to produce flow projections for the 20 year horizon. This procedure is outlined in Figure 56.



**Figure 56 Idealized framework for adapting the Wavelet Auto Regressive Method for making flow projections. Dashed lines represent data from the historical signal used to initialize the AR modeling.**

Testing this framework on historical and paleo reconstructed data revealed it to offer little or no information beyond climatology. This is attributed to several factors but the two main ones are as follows: (i) Component traces from the respective AR models tend to converge to their long-term mean fairly quickly, thus producing limited multi-decadal variability, resulting in poor skills in the projections. (ii) Boundary effects also presented challenges for this method. As a result of padding the data with zeros for the wavelet transform, variability strength always

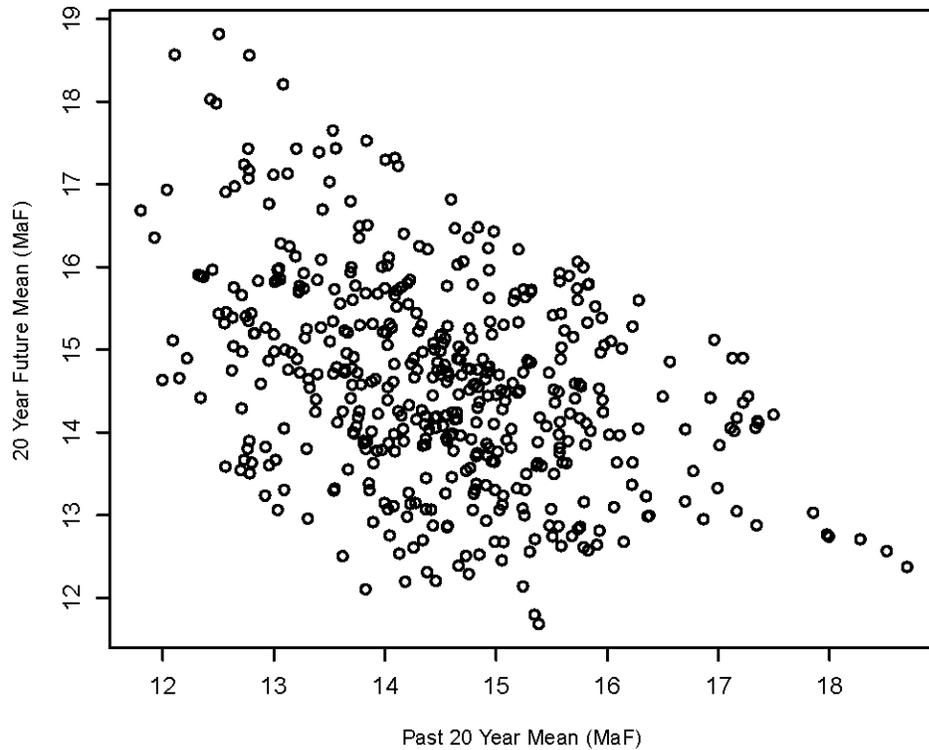
appears to be reduced as the boundary is approached. Thus, values used to initialize the AR models tended to bias the projections toward reduced variability.

After trying variations of the above methods the following approach is proposed. The problem is split into mean and variance projections, with a regression model for the mean component and a spectral re-sampling for the variance component. The motivation for this came from the finding that the wavelet based spectral approach captures the temporal evolution of the variance much better (see Figure 51, chapter 4), while the mean component can be captured using a linear regression based on the low frequency variability seen in the paleo streamflow.

These two components of the methodology are described below and are subsequently combined to form the overall approach for 20 year streamflow projections

### **5.3.2 Regression Model for Mean Flow Projections**

Figure 57 shows a scatter plot of past 20 year mean flow against the future 20 year mean flow values from the paleo reconstructed flow data. The 20-year time window is chosen based on management and planning considerations motivated in the introduction. The methodology is described for this time window; however, testing showed little sensitivity to future horizons in the 15-25 year range. The long paleo flow record allows for a more robust investigation of this relationship compared to the historical record. The scatter plot shows an inverse relationship significant at 95% confidence level (correlation  $\sim -0.35$ ). Although the significance level is quite high, the overall fraction of variance explained by this relationship is only about 12%.



**Figure 57 Lees Ferry 20 year mean flow plotted against preceding 20 year mean flow window for the paleo reconstructed flow data set.**

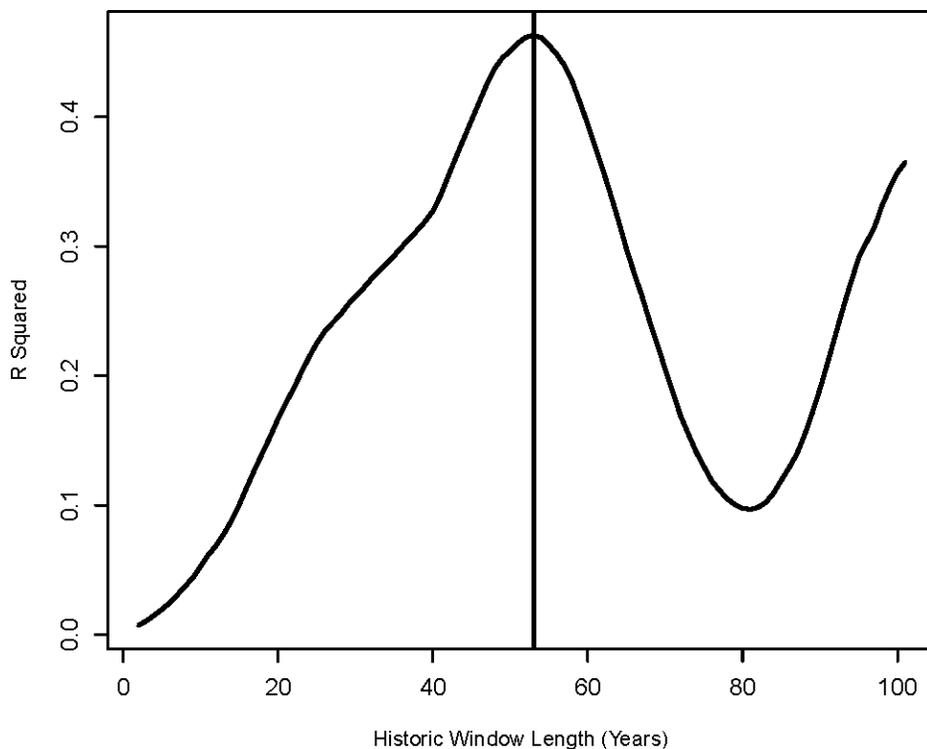
To investigate the variability of this relationship, the preceding window length was varied from 2 years to 100 years. For each preceding window length, a linear regression is fitted and R-square computed. The R-squared values are plotted against the associated preceding window length (Figure 58), which indicates an optimal (highest R-square) window length of about 53 years for a 20 year future mean projection. Figure 59 shows the scatter plot using this preceding window. Relative to Figure 57, the relationship is stronger and scatter is reduced. This regression is represented as:

$$\bar{F}_{20} = \bar{P}_{53}m + b$$

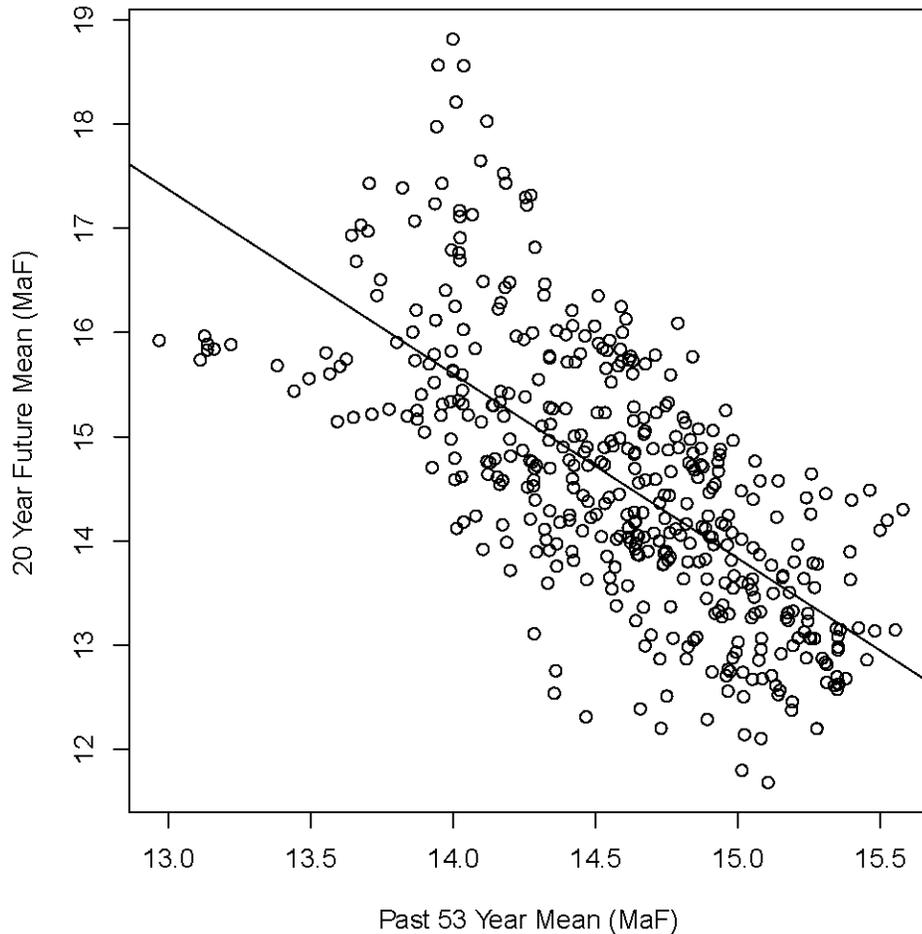
**Equation 11**

where  $\bar{F}_{20}$  is the future 20 year mean flow,  $\bar{P}_{53}$  is the mean from the preceding 53 years of data,  $m$  and  $b$  are the regression coefficients (-1.5 and 36.3 respectively).

The low frequency variability mode seen in Lees Ferry streamflow (see Figure 35, chapter 4) has a period of about 65 years. Relative to that period length, the 20 year projection window is about 1/3 of this cycle and the best preceding window length of ~50 years identified above for the regression is roughly 2/3 of the cycle. From empirical results using idealized periodic functions, it was seen that for a future projection horizon that is 1/3 of the function's characteristic period, a preceding window length of approximately 2/3 the period is optimal. Thus, it seems that the regression capitalizes on the low frequency signal found in both the paleo and historic flow data.



**Figure 58 R-Squared vs. preceding window length based on linear model to project future 20 year mean flow.**



**Figure 59** Lees Ferry 20 year mean flow plotted against preceding 53 year mean flow window for the paleo reconstructed flow data set with best fit line.

### 5.3.3 Wavelet Based Resampling for Variance Projections

The regression provides the mean flow for the future time horizon and to offer a more complete projection of future conditions, variability information is needed. To accomplish this, a wavelet based resampling method was developed, whose steps are described below.

(i) At current time  $t$  from which the future projections are desired, the wavelet transform is computed for the time series of flows from the recent preceding window (say,  $t-h$  through  $t$ ). In order to capture a range of variability scales, particularly the low frequency mode of 64-yr identified earlier, an 80 year past time window (i.e.,  $h$ ) was established (e.g., 1921-2000 data is used in the wavelet for projection period of 2001-2020). An 80 year past window is the

minimum length to resolve the low frequency. Using a red noise background spectrum, ‘significant variability’ scales are identified at 90% confidence (i.e., the significant spectral periods). It is worthwhile to note that if multiple sequential scales are deemed significant, they are retained as one “band,” just as was done in the WARM framework (Chapter 3).

(ii) The wavelet transform is computed for all possible 80 year blocks of the paleo data prior to time  $t$ . For each block, information pertaining to the ‘significant variability’ scales/bands from above is saved as ‘feature vectors.’ This process proceeds sequentially, moving the block window one year at a time through the paleo data set, such that a ‘feature vector’ is recorded for each paleo block. The contents of the feature vector are two measures, phase and amplitude, for each significant variability band identified above. Thus, feature vector length is a function of the number of identified bands.

(iii) The amplitude similarity metric (ASM) and phase similarity metric (PSM) are computed for each 80 year block window, quantifying the spectral similarity of the block windows with respect to the recent 80 year block. The ASM between any two periods is computed as:

$$ASM(s) = \frac{\sum_{i=1}^n abs\{[SAWP_i^a(s) - SAWP_i^h(s)] * W(i)\}}{n}$$

**Equation 12**

where  $s$  is the scale or band of scales,  $i$  represents position in time through the block such that  $i=1$  is the most recent value,  $n$  is the block length,  $a$  indicates the analog period,  $h$  denotes the historical period and  $W(i)$  is a weight function given by:

$$W(i) = 1/i$$

**Equation 13**

such that values closer to the projection boundary are given more emphasis.

To compute the PSM, the cross wavelet transform is employed [Torrence and Compo, 1998]. For two time series,  $X$  and  $Y$ , with wavelet transforms  $W_i^X(s)$  and  $W_i^Y(s)$ , their cross wavelet transform is given as:

$$W_i^{XY}(s) = W_i^X(s)W_i^{Y*}(s)$$

**Equation 14**

where  $W_i^{Y*}(s)$  is the complex conjugate of  $W_i^Y(s)$  with  $i$  representing position in the time domain and  $s$  the wavelet scale [Grinsted et al., 2004]. From this, the cross wavelet phase is denoted as:

$$\varphi_i^{XY}(s) = \tan^{-1}[\Im\{W_i^{XY}(s)\}/\Re\{W_i^{XY}(s)\}]$$

**Equation 15**

where  $\Im$  and  $\Re$  are the imaginary and real operators, respectively. The phase similarity metric (PSM) is obtained as:

$$PSM(s) = \frac{\sum_{i=1}^n \text{abs}\{\varphi_i^{ah}(s) * W(i)\}}{n}$$

**Equation 16**

where  $\varphi_i^{ah}(s)$  is the cross wavelet phase angle between the analog ( $a$ ) and historical ( $h$ ) periods at scale  $s$  and position  $i$ .

(iv) Euclidean distances are computed using the ASM and PSM to arrive at a single similarity value from each vector – i.e., similarity of each 80 year block with the most recent. However, before this can be done, the ASM and PSM values in the feature vectors must be normalized such that their magnitudes are comparable. The phase angles range from zero to pi, while the SAWP values are quite large (e.g.  $10^{13}$ ). To accomplish this, all PSM values are divided by the largest PSM computed from any feature vector. The same is done for the ASM values, thus ensuring that the range of both metrics is zero to one. With the ASM and PSM values normalized, the Euclidean distance is calculated for each feature vector as:

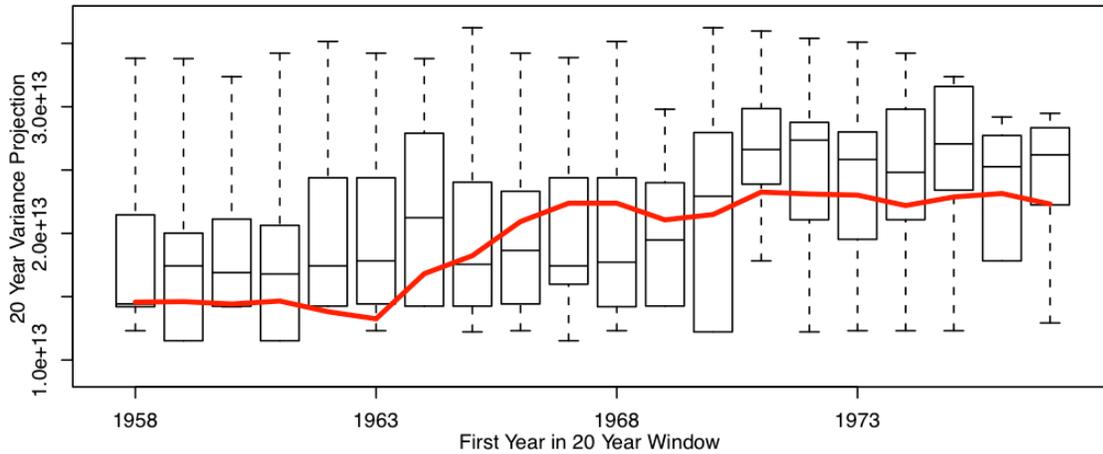
$$d(a, h) = \sqrt{(ASM_{a1} - ASM_{h1})^2 + (PSM_{a1} - PSM_{h1})^2 \dots (ASM_{an} - ASM_{hn})^2 + (PSM_{an} - PSM_{hn})^2}$$

**Equation 17**

where  $a$  represents the analog period,  $h$  represents the historic period, and  $n$  is the number of significant scales for which the similarity metrics are computed.

(v)Based on these distances, the paleo blocks are ordered from most similar (smallest distance) to least similar (largest distance). The blocks are weighted inversely to their distance and are then resampled, similar to the K-nearest neighbor (K-NN) time series resampling approach discussed in chapter 2 [Lall and Sharma, 1996]. For each resampled block, the variance of the following 20 years of flow data is computed. In this work, 100 traces were generated for each ensemble. Thus an ensemble of variances is generated. Steps (i) through (v) are repeated for each year of the 20-year time horizon, providing an ensemble of 20-year variance projections.

Figure 60 shows the application of this approach for making 20 year variance projections at Lees Ferry for time horizon of 1958-1977. The 20 year variance of the observed flow is shown as a red line. The ensemble variance in 1958 corresponds to the projected variance of mean streamflow for the 20-year period 1959-1978 and so on. The shift to a higher variability regime in the later part of the 20<sup>th</sup> century is reflected well by the projections. This demonstrates the ability of the wavelet based resampling method to provide skillful variance projections.

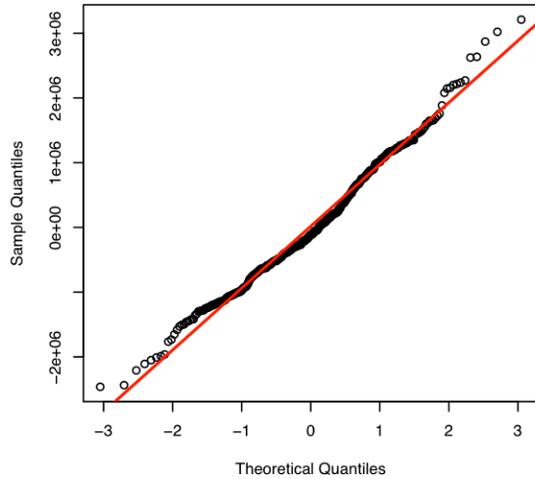


**Figure 60 Flow variance projections for 20 year future windows for the Colorado River at Lees Ferry, AZ.**

### **5.3.4 Combination of Mean and Variance Projections**

The decadal projection of mean streamflow requires the combination of the mean and variance components. To this end, the mean projection from the regression model is combined with the variance projection from the wavelet based approach. In addition, the mean projection from the regression model includes uncertainty which is also incorporated in this combination.

The errors from the regression model show a Normal distribution with no autocorrelation (Figure 61) – which satisfies the assumptions of regression theory, thus normal distribution can be used to generate uncertainties. The combination steps are described below.



**Figure 61** Normal Q-Q plot of regression residuals.

(i) The regression for the mean can be expressed as:

$$\bar{F}_{20} = \bar{P}_{53}m + b + \varepsilon$$

**Equation 18**

where  $\varepsilon$  is Normally distributed with mean zero and standard deviation  $\sigma_1$ ,  $N(0, \sigma_1)$ . This captures uncertainty in the linear regression and is obtained from the standard regression theory [Montgomery and Runger, 2003].

(ii) The mean variance generated from the wavelet resampling ensemble ( $\overline{\sigma^2}$ ) represents the variance estimate of the 20 year flow window. Thus, from sampling theory [Montgomery and Runger, 2003], the sampling distribution of the 20-year mean flows is Normally distributed with some mean and a standard deviation given by:

$$\sigma_2 = \frac{\bar{\sigma}}{\sqrt{n}}$$

**Equation 19**

where  $n$  is equal to 20.

(iii) The above two sources of uncertainty can be incorporated in a Monte Carlo approach where in an ensemble of standard Normal deviates,  $R$ , are generated with standard deviations  $\sigma_1$  and  $\sigma_2$  and added to the estimate of 20-year mean flow from the regression. In other words, the 20-year mean flow ensembles are generated as:

$$\bar{F}_{20} = \bar{P}_{53}m + b + R$$

**Equation 20**

Another approach is to combine the two sources of uncertainty and generate standard normal deviates,  $R$ , with standard deviation  $(\sqrt{\sigma_1 + \sigma_2})$ .

#### 5.4 Projection Results

Using the long paleo streamflow record, we apply the proposed methodology to generate ensembles of 20 year mean flow projections. Results for three periods with different flow characteristics are presented. The periods are 25 years in length and an ensemble of future 20 year mean flow projections is made for each one of those years. For example, consider the period of 1925 to 1949. The first ensemble of flow projections (1925) represents the average flow for 1926 to 1945 and so on such that the last ensemble (1949) represents the average flow for 1950 to 1974.

Validations of the projections are made as follows. Results are presented as boxplots, where the whiskers extend from the 5<sup>th</sup> to 95<sup>th</sup> percentile, and the box represents the inter-quartile range. The actual mean values for each 20 year window are shown as a solid line, offering a visual inspection of the overall performance. For comparison, ensembles of flow climatology are also shown. This is similar to the Index Sequential Method (ISM), which is widely used in the water resources field [Ouarda *et al.*, 1997]. Correlation values between the actual 20 year means and projected medians are provided in addition to the Rank Probability Skill Score (RPSS) [Wilks, 1995].

RPSS quantifies categorical skill and is widely used measure for evaluating climate and weather forecasts. Here we consider three categories at the terciles such that a ‘climatological’ projection has equal probability (1/3) of being in any of the three categories. From each

projection ensemble, the probability of being in the three categories is also computed. The RPSS is then computed as:

$$RPSS = 1 - \left( \frac{RPS(\text{forecast})}{RPS(\text{climatology})} \right)$$

**Equation 21**

Where RPS (rank probability score) is given by:

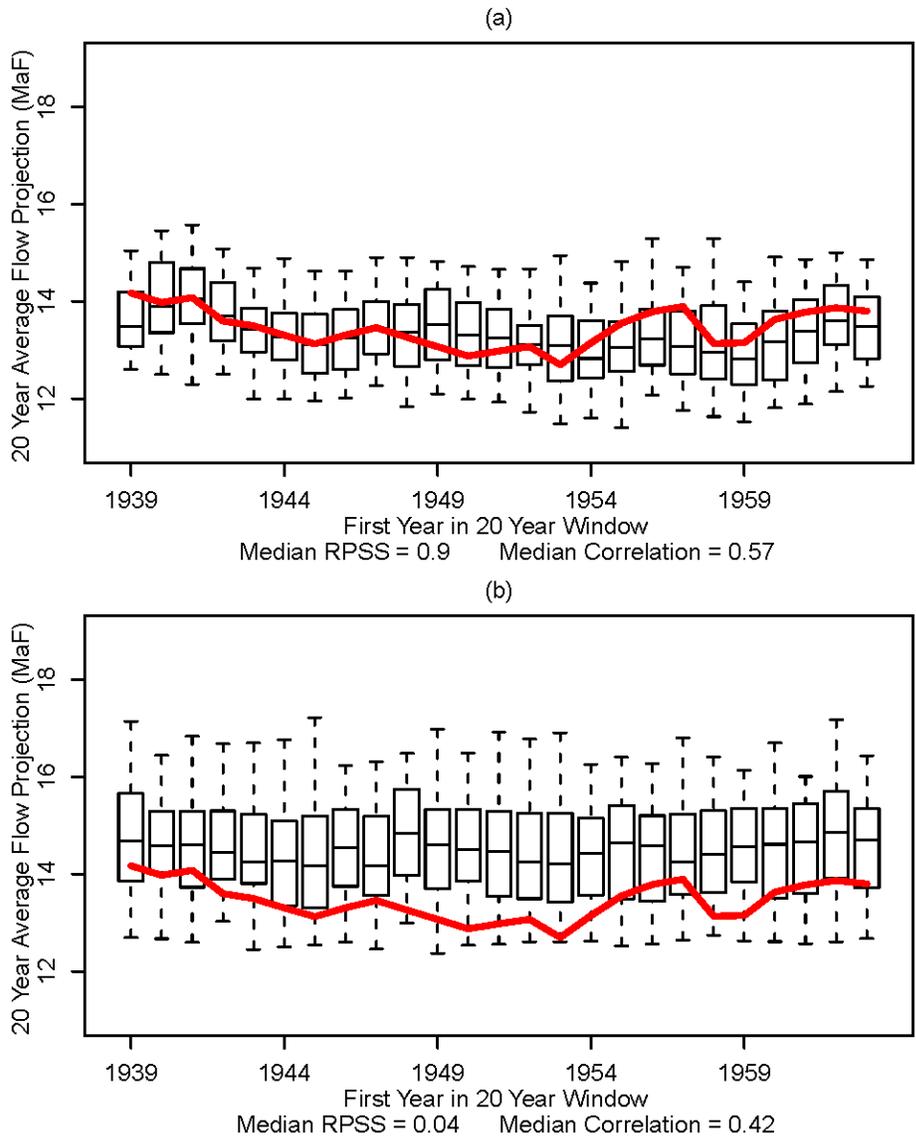
$$RPS(\text{forecast}) = \sum_{m=1}^k \left[ \left( \sum_{j=1}^m P_j - \sum_{j=1}^m d_j \right)^2 \right]$$

**Equation 22**

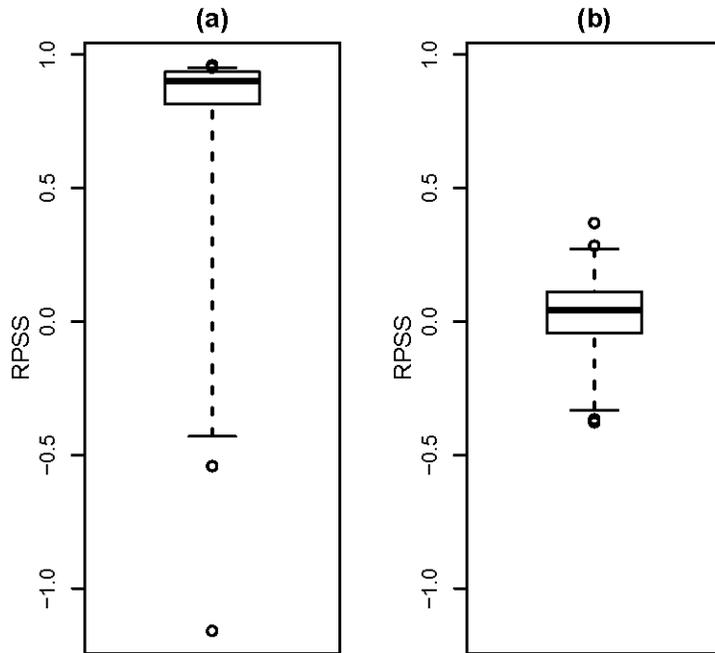
$k$  represents the number of categories (in this case,  $k=3$ ),  $P$  is the probability of being in category ' $j$ ' and  $d_j$  is a binary variable of category ' $j$ ' that is set to zero if the actual is not in this category.

RPSS scores range from 1 (when the projections are accurate and highly skillful relative to climatology in projecting the correct category) to negative infinity (when projections are worse than climatology) with zero being the point where projections are no more skillful than climatology. The RPSS is computed for each year of the projection horizon and displayed as boxplots. Also, the median RPSS is computed as the representative value for the period.

Results are presented for the three projection horizons (1939-1963, 1779-1803 and 1584-1608), which were chosen to be representative of the flow variations over the paleo record and include both wet and dry epochs. Figure 62 shows the projection ensembles for the 25 year period of 1939 to 1963, which is characterized by persistent low flows. For this period, the median RPSS is 0.9 and the median correlation is 0.57. The ensembles follow the historic values (red line) throughout the period and reflect the below average flows. Ensembles from the climatology projections are consistently above average and this is reflected by the median RPSS. The median correlations however, are similar due to the relatively stable values throughout this period. Figure 63 shows boxplots of the RPSS values from the two projection methods, further highlighting the skill over climatology.

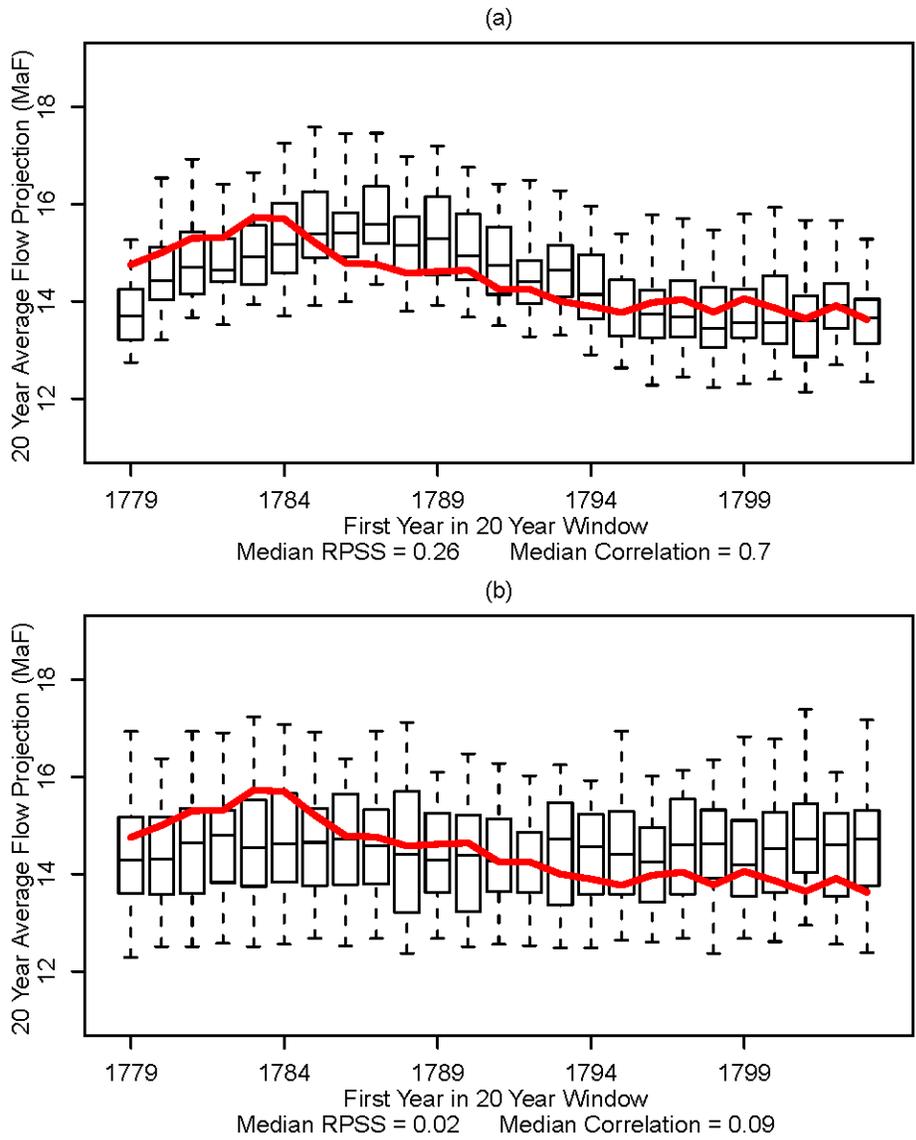


**Figure 62 Ensembles of 20 year mean projections for Colorado River flow at Lees Ferry, AZ, for (a) the projection method presented in this work and (b) a climatological projection approach. The solid red line is the actual value for each projection.**

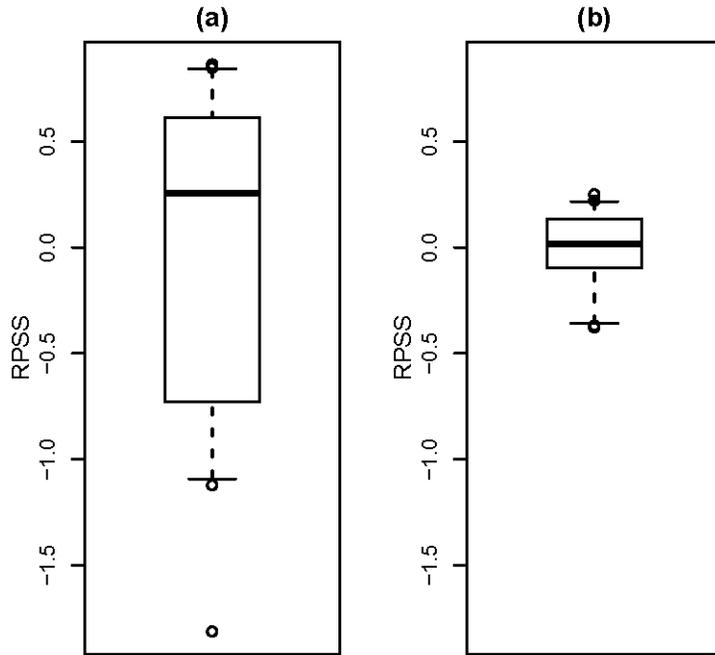


**Figure 63** Boxplots of RPSS values for flow projections for the period 1939 to 1963 using (a) the projection method presented in this work and (b) a climatological projection approach.

Figure 64 shows the projections for the time horizon of 1779-1803. For this period, the historic 20 year mean values tend to fall in the middle tercile, indicating an “average” flow period. The median RPSS is 0.26 and the median correlation is 0.7 for decadal projection method presented in this work. On a whole, the projections track the actual variability very well, even capturing the minor local variations around 1785. The climatological projections do not exhibit any temporal variability. However, the boxplots of RPSS values from the two projection approaches show a considerably wider range for the decadal projection framework (Figure 65). Upon review of the results, several of the historic 20 year means for this period are close to the tercile divisions, causing a small difference between the projected and actual values to have a disproportionate effect on the RPSS for that year. Recall that RPSS is a categorical metric and the “coarse” divisions can at times fail to resolve finer details. This is why a variety of evaluation metrics are presented.

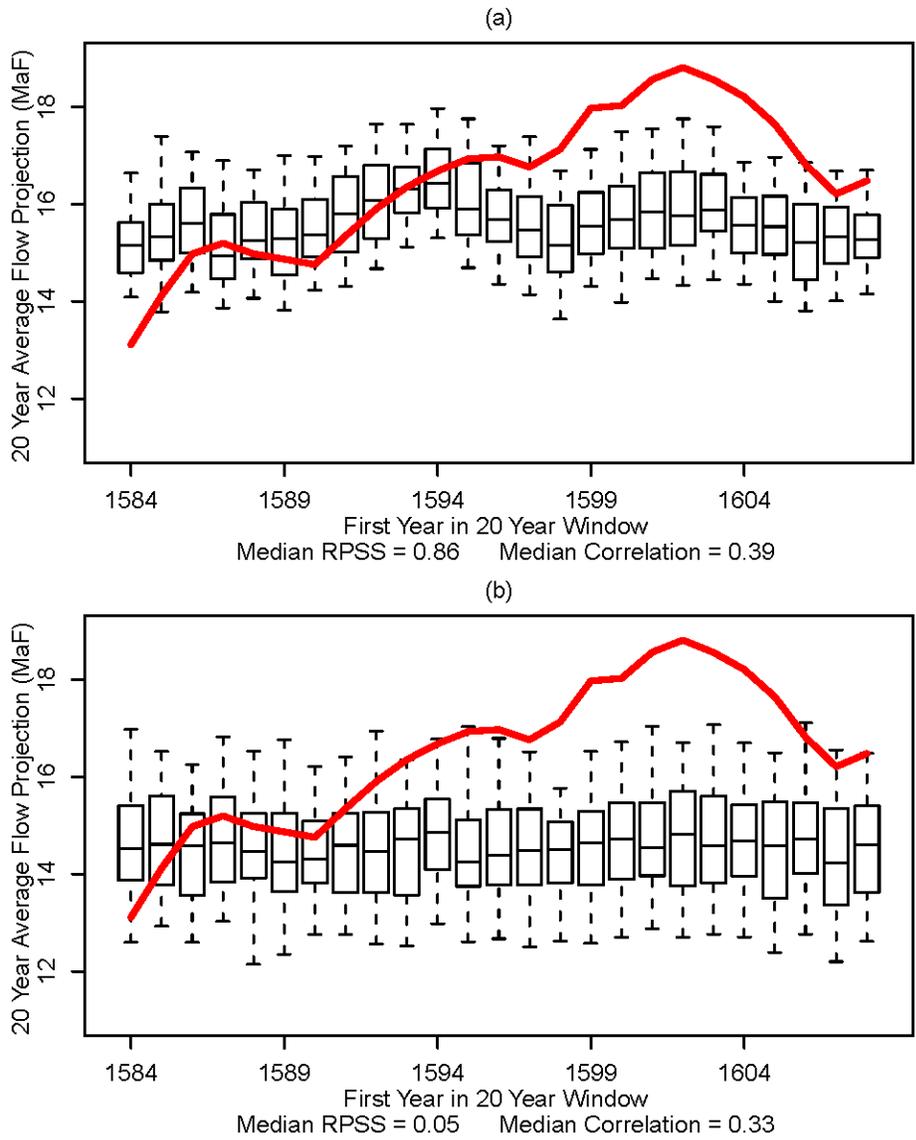


**Figure 64 Ensembles of 20 year mean projections for Colorado River flow at Lees Ferry, AZ, for (a) the projection method presented in this work and (b) a climatological projection approach. The solid red line is the actual value for each projection.**

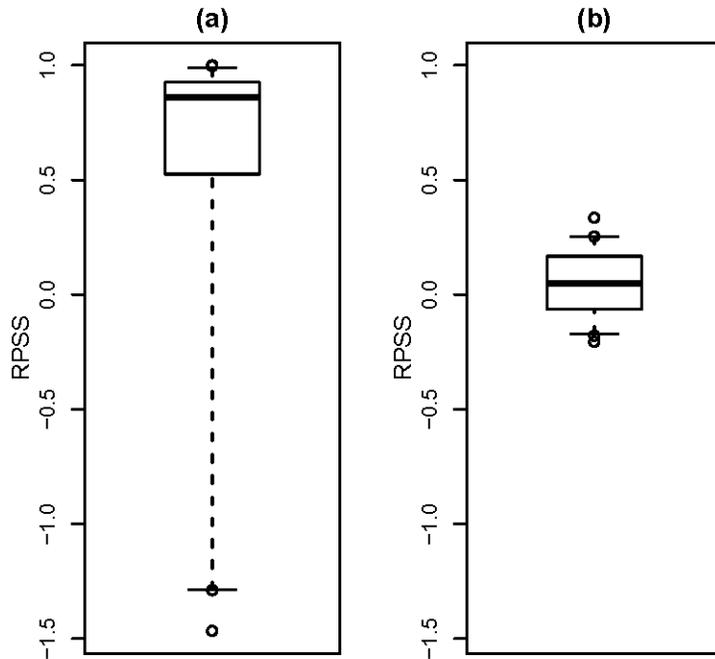


**Figure 65** Boxplots of RPSS values for flow projections for the period 1779 to 1803 using (a) the projection method presented in this work and (b) a climatological projection approach.

Last, projections were made for a high flow period in the paleo record (1584-1608). Results from Figure 66a show high RPSS score, but lower median correlation relative to the values from the “average” flow period (Figure 64). For this period, the projections are all largely in the high flow tercile, consistently producing strong RPSS values (Figure 67). However, local variations are not well tracked, causing lower median correlation values. Again, the climatological projections show no temporal variation from one window to the next and have considerably lower RPSS scores.

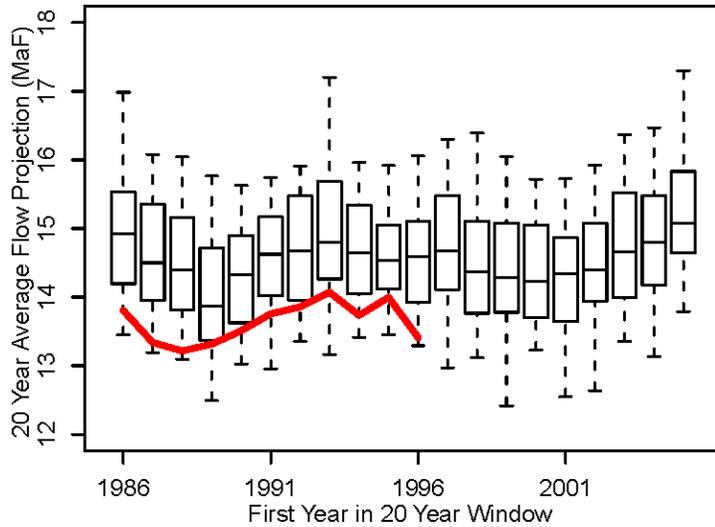


**Figure 66 Ensembles of 20 year mean projections for Colorado River flow at Lees Ferry, AZ, for (a) the projection method presented in this work and (b) a climatological projection approach. The solid red line is the actual value for each projection.**



**Figure 67** Boxplots of RPSS values for flow projections for the period 1584 to 1608 using (a) the projection method presented in this work and (b) a climatological projection approach.

Figure 68 shows 20 year mean flow projections beginning in 1986. The observed natural flow data extends to 2006 so the ‘actual’ 20-year mean values beyond 1986 are based on less than 20-years of flow values. Hence they are rough estimates and we show until 1996 for which 10-years (1997-2006) of flows are used. The median of the projections has the same temporal pattern as the actual but is over simulated. The projections show a decline during 1996 – 2002 year period, coinciding with the low flow period. The skills can be better estimated with more data but the results are quite encouraging.



**Figure 68 Twenty year mean projections for Colorado River flow at Lees Ferry, AZ.**

## 5.5 Summary

The new methodology presented provides the ability to make skillful ensemble projections about the characteristics of future mean flow in the Colorado River Basin. The projections are combination of a mean estimate using a regression model and a variance component from wavelet based resampling approach. We applied this to making projections of 20-year mean flow, a window chosen based on water management and reservoir operations relevance in the Colorado Basin. The ensemble projections from the methodology for a suite of wet and dry epochs were well simulated based on a combination of evaluation metrics. Conditional simulation methods can be developed to generate ensembles of streamflows based on the projected mean flow.

A few limitations do exist in the presented approach. Especially, the assumption that frequencies identified in the recent time window (in this case ~80 years) will maintain their relationship in terms of the variability for the projection window (in this case 20-years). If the projection window is longer, then the relationship might not be robust and the projection skills are likely to drop. Additionally, long records (paleo or otherwise) are needed for robust

development of the regression model and for the wavelet resampling. The method seems to be under simulating the high flow epochs, perhaps due to the fact that the paleo record has many dry epochs, which could bias the regression component of the model.

## **CHAPTER 6: APPLICATION OF DECADAL PROJECTIONS TO MANAGEMENT OF THE COLORADO RIVER**

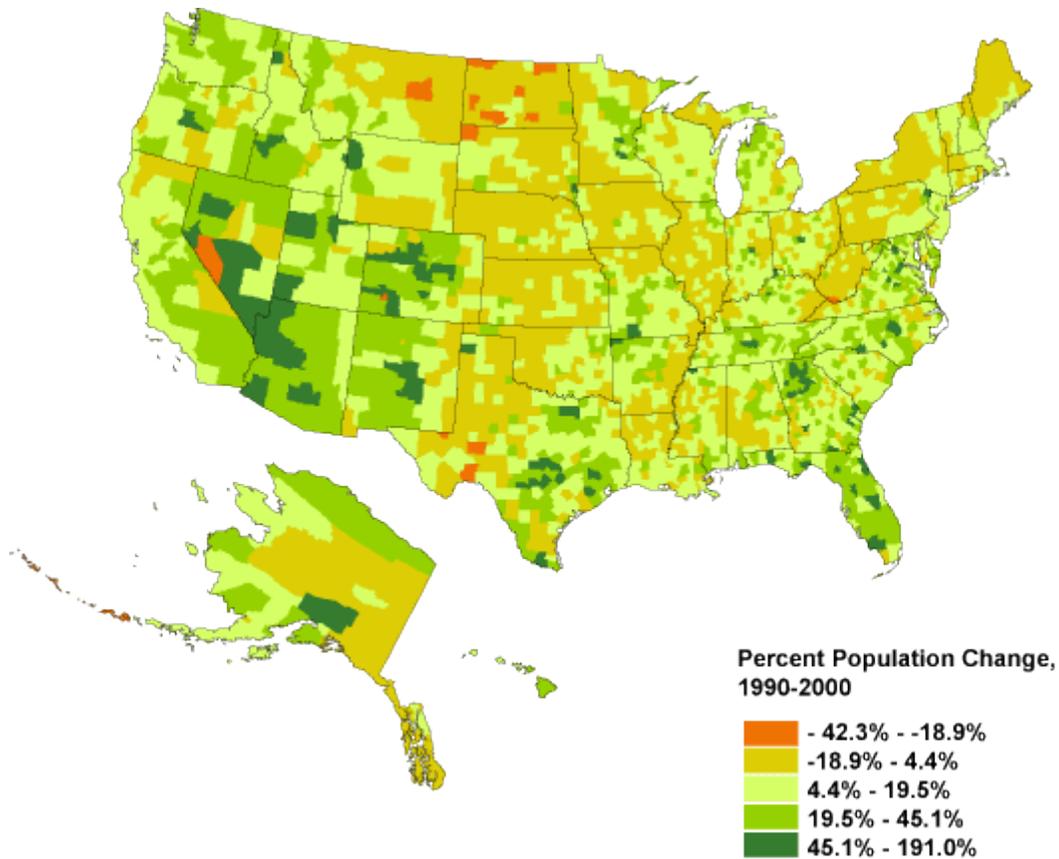
In chapter 5, a unique framework for making decadal flow regime projections was developed and demonstrated for the Colorado River. As water resources in this Basin become increasingly strained by flow variability, climate change and increased demand, alternate management practices may be considered to improve reliability. Given the phenomenal storage capacity and high streamflow variability of the system, these decadal projections should have considerable utility in crafting adaptation strategies for changing conditions.

### **6.1 Introduction**

Adaptation in water management and reoperation of reservoir systems are garnering increased consideration in light of both present and anticipated stress on water resources globally. The objectives of adaptation in these studies are quite varied. In some regions, hydropower may be the most important system component [*Minville et al.*, 2009]. For other areas, concern over increased flood risk due to climate change reigns supreme [*Veraart et al.*, 2010]. In the case of the southwestern United States, water supply reliability is the main focus of adaptation strategies [*Gober and Kirkwood*, 2010; *Miller et al.*, 1994; *Rajagopalan et al.*, 2009]. This work seeks to demonstrate that even though the Colorado River system is operated on a tight margin, due at times to conflicting objectives and an over allocation of resources, information from decadal projections coupled with adaptation strategies can be used for the benefit of supply reliability.

Since the early 1900's, the Southwest has grown over 1500% compared to the United States national average of approximately 225%. Figure 69 depicts this trend over a ten year period spanning 1990-2000. Despite this rapidly increasing strain on water resources and significant hydrologic variability, the system has always been able to meet demands. Reservoir

storage is the backbone of this reliability; the standard deviation of flow in the Colorado Basin is 30% of the average. During high flow years, the phenomenal capacity of the system can capture otherwise excess water and hold it over for dry periods. However, as the system becomes more stressed, there will undoubtedly be a role for additional management strategies to further shore up the system reliability and mitigate shortages.



**Figure 69 United States population change map. (source: [www.CensusScope.org](http://www.CensusScope.org))**

One component of the system for consideration in alternate management approaches is storage distribution across the system's reservoirs. While these facilities are indeed critical for reliability, their large surface areas, coupled with the arid nature of the Southwest, creates a substantial evaporative loss each year. Savings can be achieved by storing more water in lower evaporation rate reservoirs, thus improving reliability and reducing shortage. Another system

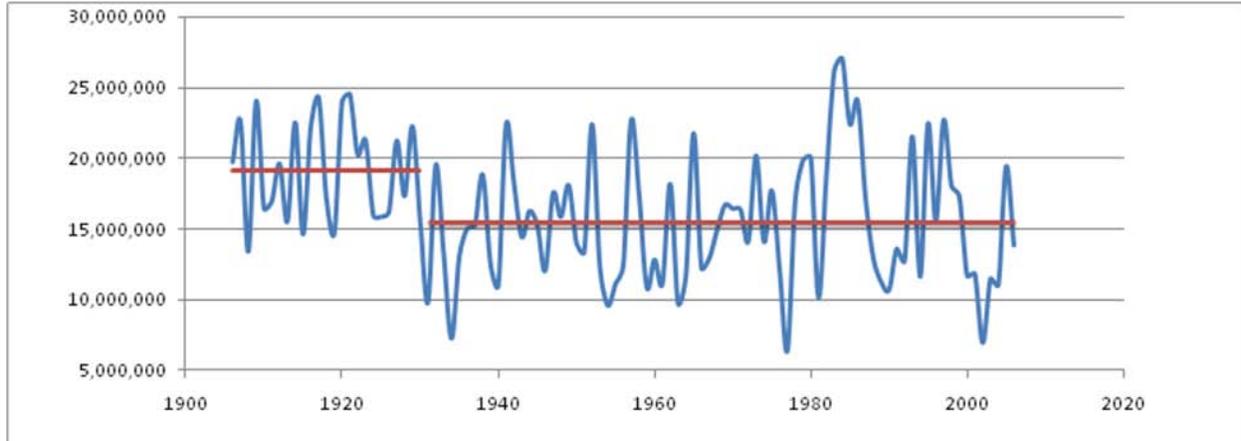
trait for improving reliability is the overall storage capacity of the system. Even with a total storage roughly four times the mean annual flow of the river, the system has occasionally been so full that reservoirs were forced to spill excess water. This represents another system loss that will likely persist from time to time in the future; although climate change is expected to reduce mean annual flow magnitude, variability is thought to be increasing. In fact, some magnitudes may occasionally double the current mean. As such, additional storage may help to capture and carryover some of these extremely high flows. Decadal projection information could be used to inform the decision to begin utilizing these supplemental capacities to reduce “spill related” losses at the major reservoirs. During dry epochs, this reserve could then be used to mitigate shortage conditions.

To this end, the Basin long term planning model, Colorado River Simulation System (CRSS), is used to demonstrate examples of how system flexibilities might be exploited, particularly in light of decadal flow projections, to improve the overall water supply reliability

## **6.2 Colorado River System Overview**

To provide perspective and background, a brief overview of the Colorado River system follows. On a macro scale, water in the Colorado River Basin is allocated by the 1922 Colorado River Compact and 1944 Treaty with Mexico. The 1922 agreement divides the U.S. portion of the system into the Upper and Lower Basins, each allocated 7.5 million acre-ft (MaF) of Colorado River water annually. Furthermore, it is stated that the Upper Basin will deliver no less than 75 MaF of water over any 10 year period to the Lower Basin at Lee Ferry, AZ. In addition to the 15 MaF/yr allocated to the Upper and Lower Basins, the treaty with Mexico guarantees delivery of 1.5 MaF/yr. This brings the total Basin allocation to 16.5 MaF/yr [Verberg, 2010]. Importantly, the Colorado River Compact was based on limited data (~1900 -1922) that

suggested a much wetter climate and overall greater availability of water (Figure 70).



**Figure 70 Colorado River natural flow (acre-ft) at Imperial Dam, AZ, which represents total Colorado River yield (1906-2006).**

The average annual Basin yield is about 16.3 MaF/yr, just short of the total allocation. However, a number of system losses, largely evaporation related, create a greater supply/allocation disjoint. While the Lower Basin has fully developed its apportionment, the Upper Basin has been slower to grow. At present, Upper Basin demand is estimated to be approximately 4.5 MaF/yr, which results in a net system surplus of about 400,000 acre-ft annually (Table 1).

<b>System Component</b>	<b>Value (MaF)</b>
Upper Basin Natural Flow (Lees Ferry)	15.0
Demands (7.5 LB, 4.5 UB & 1.5 Mex)	-13.5
Reservoir Evaporation	-1.4
Inflow Between Powell and Mead	0.86
Losses Below Hoover Dam	-1.0
Inflow Below Hoover Dam	0.45
<b>Net System Balance</b>	<b>0.4</b>

**Table 1 Colorado River system water balance [Rajagopalan et al., 2009]. Flow values are based on the natural flow data set for the period of 1906-2006. Demands are present day estimates. Losses are average values based on operational data and modeling results also for the period 1906-2006[United States Department of the Interior, 2005; , 2008].**

The physical system which constitutes the Colorado River Basin is also noteworthy. The Colorado River and its major tributaries originate high in the Rocky Mountains of Colorado, Wyoming and Utah. Over 90% of the total system yield originates in the Upper Basin, largely as snowpack run-off. Across the Upper Basin tributaries, there are several medium sized storage projects with a collective capacity of almost 5 MaF. Just upstream of the Lee Ferry Upper/Lower Basin divide sits the Glen Canyon Dam that creates Lake Powell. This large reservoir (24.4 MaF live storage capacity) allows the Upper Basin to capture and regulate high run-off season flows, thus ensuring compact obligations to the Lower Basin are met. Outside of Las Vegas and downstream from the Grand Canyon, the Hoover Dam and Lake Mead (26.2 MaF live storage capacity) store deliveries from the Upper Basin and reregulate flow of the Colorado River to reliably deliver water for downstream demands throughout the year. In the Lower Basin, water is allocated as follows: 2.8 MaF to Arizona, 4.4 MaF to California and 0.3 MaF to Nevada. The

Upper Basin apportionment is divided amongst the states via the following percentages: 51.75 for Colorado, 11.25 for New Mexico, 23 for Utah and 14 for Wyoming. Additionally, Arizona is entitled to 50,000 acre-ft of Upper Basin water due to gains made by the stretch of the San Juan River in northwest Arizona [Verberg, 2010].

### **6.3 Adaptive Policies**

As a response to the ongoing, unprecedented drought, in 2005, the Bureau of Reclamation, in coordination with the Basin states, began work to establish reservoir operational policies in the Colorado Basin for periods of drought and reduced reservoir storage. Through this process, based on input from the States, non-governmental organizations and other stakeholder groups, several policy alternatives were developed and examined. Out of this, a “preferred alternative” was identified and subsequently adopted as part of the “Law of the River” with the 2007 signing of the Record of Decision for the Colorado River Interim Guidelines for Lower Basin Shortage and Coordinated Operations for Lake Powell and Lake Mead Final Environmental Impact Statement. These “Interim Guidelines” extend through 2026 and include three main components aimed at mitigating the effects of drought: 1) coordinated operations of Lakes Powell and Mead, 2) a mechanism by which states can store excess water in Lake Mead for later use – termed “Intentionally Created Surplus” and 3) shortage policies for the Lower Basin in times of low reservoir storage.

For the purpose of this modeling, the Interim Guidelines were utilized as a baseline policy with which alternative operations were compared. By doing so, the objective was not to undermine the progress represented by this historic agreement, but to show that given a particular goal, in this case overall system reliability, flexibilities exist and can be exploited to meet the desired end.

Two policy alternatives were developed in this work, such that the second policy (referred to as step 2) incorporates and builds upon the first policy (referred to as step 1). This offers perspective on the relative benefits of the policy components. In step 1, storage distribution between Powell and Mead is reconfigured to reduce evaporative losses. Step 2 utilizes information from the decadal projections to preemptively begin storing water in offline storage ahead of wet epochs. This aims to reduce spill related losses and build a water reserve. As the Interim Guidelines are in place through 2026, all model runs maintain the current policies through the interim period. However, beyond this horizon, the policy alternatives considered were quite varied, so as to not constrain options by the current system paradigms. One way to improve overall system reliability is to reduce losses. In the Colorado Basin, these losses are largely due to reservoir evaporation. Between Lakes Powell and Mead alone, over 1 MaF are lost, on average, each year. While this constitutes only about 6% of the mean annual Basin yield, it is quite substantial given the narrow operational margin. Thus, “step 1” in the policy alternative development focuses on reducing evaporative losses of Powell and Mead. The evaporation rate from Lake Mead is almost double that of Lake Powell. The locations of these reservoirs differ by over 2000ft in elevation and almost 10°F annually<sup>8</sup>. From the Interim Guidelines, reservoir operations are “coordinated,” which as a gross simplification, affects a somewhat equitable storage of water between Powell and Mead. There are many benefits to this approach, ranging from hydropower considerations to flood control and a number of others. However, in doing so, water which could have been stored in a lower evaporation rate reservoir (Powell) is stored at a higher loss rate (Mead). Thus, the crux of the step 1 policy alternative is to preferentially store water in Lake Powell, thereby decreasing the surface area of Lake Mead and producing an “evaporation savings.” To elaborate, the policy is implemented as follows.

---

<sup>8</sup> <http://www.wrcc.dri.edu/>

Storage in Lake Mead is first utilized to meet downstream demands in the Lower Basin. If Lake Mead does not have water enough to meet these demands and maintain a minimum pool elevation of 1000', supplemental water is released from Lake Powell to correct the imbalance. If no supplemental water is required from Lake Powell, a minimum environmental release is made. Or, if necessary, space building releases are made to accommodate spring run-off. The benefit of this approach is that more water is stored in Lake Powell at a lower evaporative loss rate. This may be in conflict, at times, with environmental flow requirements, particularly in the case of high flow releases. However, for this work, some flexibility is assumed with regard to these environmental constraints. The decision to limit "drawdown" of Lake Mead to a minimum pool elevation of 1000' is to ensure reliable pumping operation at the intake for the Southern Nevada Water Authority (SNWA). The SNWA diversion is approximately 90% of the entire Nevada allocation of Colorado River water.

The shortage component of the interim guidelines is retained beyond the 2026 interim period. However, the operational changes implemented pose a problem for the shortage triggers established for the interim period. Shortage declaration and magnitude are based on Lake Mead pool elevations. Thus, left unchanged, under the step 1 policy alternative, shortage would be more frequent due to the deliberate reduction in Lake Mead storage. To circumvent this issue, an empirical analysis of shortage declaration and combined Lake Mead and Powell storage was conducted. The result provided an estimate of the total water remaining between both reservoirs at the various shortage levels. Thus, beyond 2026, volumetric shortage criteria were implemented. While these offered a decent estimation, it was far from exact. As such, when the interim guidelines were extended beyond the 2026 period for the operational baseline scenario,

the shortage triggers were also switched to volumetric criteria. This ensured that the results presented are not confounded by inconsistent shortage policies.

In policy step 2, a hypothetical groundwater bank reservoir is added to the model and policies outlined in step 1. The principal objective of this strategy is to introduce additional system storage with relatively low long term losses. Though storage capacity beyond that of Lakes Powell and Mead will be needed somewhat infrequently, given the increasing demands and potential for reduced Basin yield it is reasonable to believe that such a scenario will, from time to time, present itself. This is particularly salient in light of the occasional extremely high flows anticipated under climate change. Thus, capturing as much of the excess water as possible is in the best interest of overall reliability.

The Lower Basin geology is quite conducive to groundwater banking and recharge. Many regions have high porosity zones that are hundreds of feet thick [Fetter, 2001]. In fact, Nevada and Arizona are currently using this approach to help ensure their long term water resource reliability<sup>9</sup>. To avoid being overly restrictive, the groundwater bank added to the model is given near limitless storage capacity and infiltration rates. This decision was somewhat unrealistic, but affords the ability to estimate benefits at a variety of bank sizes, instead of attempting to impose “reasonable” physical limitations. It is important to keep in mind that although the groundwater bank is treated and discussed as a single reservoir, it is not implied that it would actually be added in the system as such.

The diversion criteria for water to be stored in the groundwater bank queue off decadal flow projections made using the method from chapter 5. Utilizing the median flow projection for the next twenty years, a coarse system water balance is computed. If Lake Mead is 2/3 or more full (indicating nearly full total system capacity), and there is a projected surplus for the period,

---

<sup>9</sup> [http://www.lvwvd.com/about/wr\\_groundwater\\_bank.html](http://www.lvwvd.com/about/wr_groundwater_bank.html)

water is diverted to the groundwater bank at a rate aimed to only “bank” water that would have otherwise been excess release from Lake Mead. Additionally, if surplus conditions are declared, water is also diverted to the groundwater bank. A policy for the use of this water is not developed, in the interest of avoiding particulars, which would require specific details and ultimately detract from the overall system focus of the work. The goal is not to prescribe a specific way to improve the system, but to demonstrate the potential utility of decadal information for meeting an objective.

While these policies may not be realistic under today’s operating policies, they provide an example of the benefits possible through relaxing what may be perceived as a rigid fixture in the interest of increasing overall system yield. The following section details the model used to simulate operations of the Colorado River system.

#### **6.4 Colorado River Basin Model**

To quantify the benefits of these policies, they were implemented as a change to Reclamation’s long term planning model, Colorado River Simulation System (CRSS). CRSS operates at a monthly time step and is developed in the RiverWare modeling software [*Zagona et al.*, 2001]. CRSS employs rule-based simulation, in which the modeler represents management strategies through the RiverWare Policy Language (RPL) with the ability to rank rules based on priority. The RPL interface employs a GUI structure for selecting expressions and functions which are based on “if, then” logic. Rules assign slot values (e.g. reservoir releases, etc.) during simulation such that objects have adequate information to be solved. These values propagate through the basin or system via the model links. This capability allows expression of alternative operating policies and has been used in numerous policy and environmental impact studies on the Colorado River by both Reclamation and stakeholders.

CRSS is the best available tool for representing the Colorado River Basin, both from the perspective of operational policies and the physical system. The model was established in RiverWare in 1996 and previously existed in FORTRAN. The experience from years of use and development are reflected by the model's effectiveness. Furthermore, Reclamation's team of engineers is constantly updating and maintaining CRSS so that it is as accurate as possible. As such, it is the perfect tool for demonstrating the ability to increase Basin yield and reduce shortage through alternative management policies.

With regard to model detail, a total of 12 reservoirs throughout the entire Basin are included. They are: Fontenelle, Flaming Gorge, Starvation, Taylor Park, Blue Mesa, Morrow Point, Crystal, Navajo, Powell, Mead, Mohave and Havasu. Each reservoir has unique operational rules. In the alternative policies presented, only the operations of Powell and Mead are modified. Additionally, over 100 water users are represented, including their diversions and return flows.

# Colorado River Basin

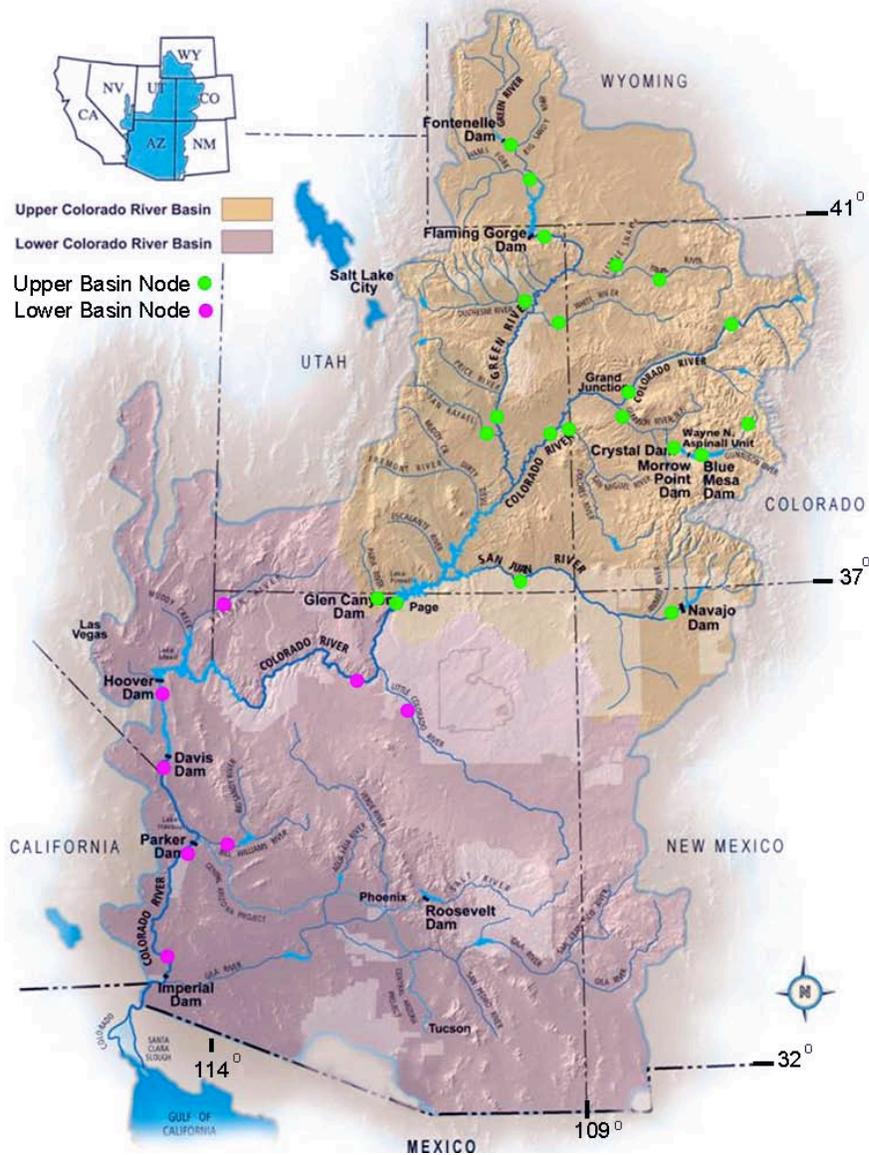


Figure 71 Colorado River Basin. Nodes shown here are also the flow data input points in CRSS.

## 6.5 Model Input Data

In this work, the model is initialized with reservoir conditions from the end of CY 2009. Therefore, the modeling horizon begins in CY 2010 and extends to the end of CY 2060. Thus, for each model trace, 51 years of monthly flow input data are required at 29 locations across the entire Basin. These are the same nodes at which the Bureau’s natural flow data set is computed

(Figure 71) [*United States Department of the Interior, 2005*]. Furthermore, these data must be intervening values.

For the modeling conducted in this chapter, three input hydrology data sets were employed. This provided the opportunity to assess the efficacy of the alternative management policies under a variety of potential future conditions. The first data set was developed using the Index Sequential Method (ISM) and Reclamation's natural flow data set [*Ouarda et al., 1997; United States Department of the Interior, 2005*]. ISM is a widely used, simple, non-parametric approach for generating streamflow data. The method uses a sequential block bootstrap of observed data in order to generate flow sequences. Each sequence is of the desired simulation length (in this case, 51 years of monthly flows). For the 102 years of available natural flow data (1906-2008), 102 traces 51 years in length can be generated (when the end of the data record is reached, it is wrapped around to the start of the record). To illustrate, the first sequence will be data from 1906 to 1956; the second from 1907 to 1957, and so on; and the last sequence will be comprised of data from 2008 along with 1906 to 1955. By resampling the observed data in sequential blocks with wrapping, ISM can potentially generate events (drought or surplus) longer than that of the observed. While this method is simple to implement and captures all the statistics and features of the observed data quite well (by design), it is limited to 102 specific sequences, most of which are seen in the observed data. To produce data for all 29 model nodes, this resampling scheme was simply applied to the natural flow data available at each input point. It is important to note, flow projections can not be made for ISM hydrologies. No traces generated via this method have the requisite 80 years of preceding flow data available to drive the projection model. Recall from chapter 5, this length is needed in order to capture low frequency variability modes in the wavelet transform part of the projection framework.

The second data set was derived from the Lees Ferry streamflow reconstruction of Woodhouse et al. [2006]. This long streamflow record was broken into the desired 51 year trace lengths via a block bootstrapping similar to the ISM, with two differences. To ensure flow projections could be made for the paleo based input hydrology, the first block in the resampling begins at the 81<sup>st</sup> year of the paleo data (1561). This provides 80 years of preceding input data for the projection model. Additionally, the block resampling is not wrapped from the end of the data back to the first year (i.e. 1997, 1490, 1491), as it is unlikely that there would be much skill in a projection made for traces that include the wrap. Even with these resampling modifications, the paleo record still produced 377 input traces. After the trace resampling was completed, the annual flows at Lees Ferry were disaggregated to intervening monthly values at the 29 input nodes via the disaggregation method developed in chapter 2.

The final input hydrology data were derived from climate models. The projections from 16 General Circulation Models (GCM) for 3 future emission scenarios (with multiple initial conditions) were downscaled, bias-corrected [Wood et al., 2004] and moved through the Variable Infiltration Capacity (VIC) hydrologic model [Liang et al., 1994; Liang et al., 1996] to produce 112 independent hydrologic inflow traces. For each trace, the simulation period was 1950 to 2099. For this work, the portion of each trace extending from 2010 to 2060 was used as CRSS input data. The results were produced as monthly, intervening values at the 29 CRSS input nodes. The transient nature of these data represents a significant departure from the relatively stationary characteristics of the ISM and paleo based inputs. As discussed in earlier chapters, anthropogenic climate change is expected to reduce average flows in the Colorado Basin over time [Christensen and Lettenmaier, 2007; Milly et al., 2005; Seager et al., 2007]. This is indeed reflected in the climate model derived hydrologies. These inflow traces were developed by the

Bureau of Reclamation and are still considered to be preliminary. It is likely that they will be revised before being used by Reclamation in any official studies.

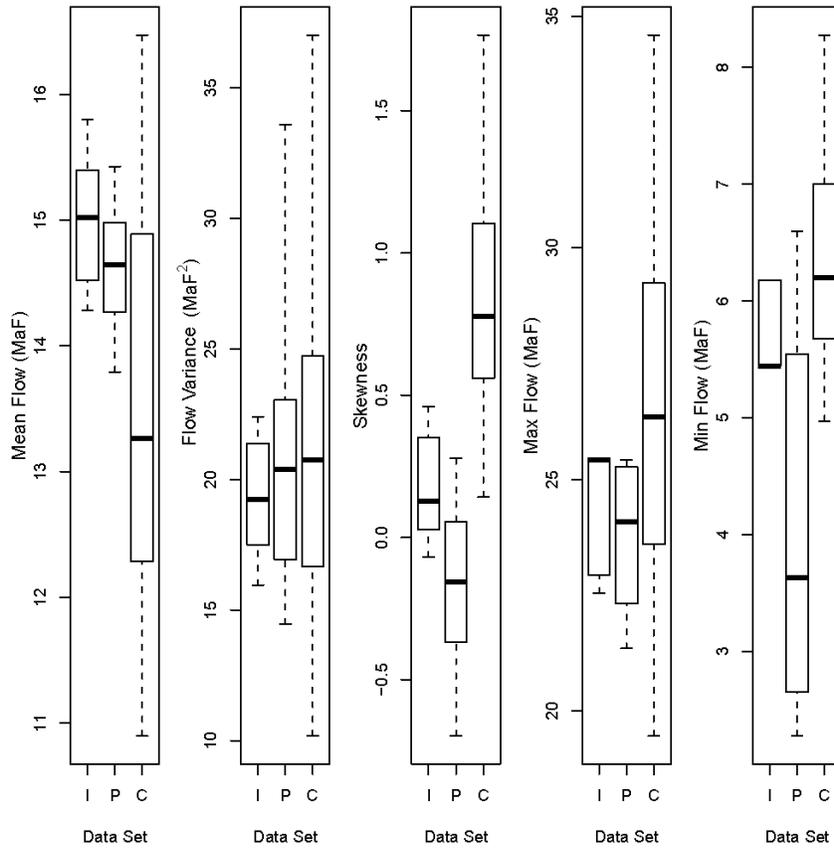
The three input hydrology data sets have different characteristics, and as such, will likely yield different outcomes in the modeling process. Figure 72 shows basic statistics for these data sets as annual, natural Lees Ferry flow. The data are presented as boxplots, where the box represents the inter-quartile range (IQR) and the whiskers extend from the 5<sup>th</sup> to 95<sup>th</sup> percentiles. Relative to ISM, both the climate and paleo traces represent a lower mean flow and greater variability; however, the differences are more substantial for the climate traces. Figure 73 shows probability density functions (PDFs) from the various traces, presented as boxplots. The paleo data include the most extreme low flow values, while the climate traces occasionally include some rather large flow magnitudes. Thus, these three data sets provide a diverse mixture of input traces for the modeling. In the following section modeling results for the three hydrologic input scenarios, coupled with the baseline, step 1 and step 2 operational policies are presented.

For model runs using the input hydrology data based on climate model output, reservoir evaporation across the Basin is linearly increased by 14.5% from 2010 to 2060. This is based on an assumed Basin-wide warming of 2°F over the model time horizon. Specifically, historical annual temperature and evaporation rates from Lakes Powell and Mead were compared. Temperature data from the Western Resource Climate Center<sup>10</sup> show an annual average temperature difference of 8.8°F between these two locations and evaporation data from the Bureau of Reclamation indicates 2.52 ft/year greater evaporation at Lake Mead [*United States Department of the Interior*, 2007a]. From this, a coarse estimate of 0.286 ft/yr increased evaporation per degree warming was computed. Thus a two degree warming suggests an increase in evaporation of 0.573 ft/year by 2060. Relative to the current annual rate at Lake Powell (3.98

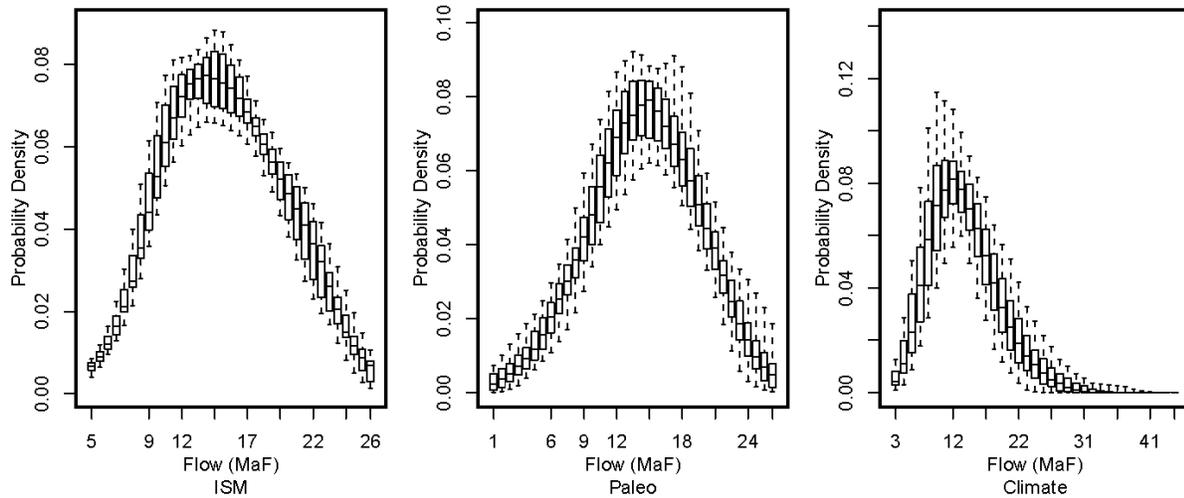
---

<sup>10</sup> <http://wrcc.dri.edu/>

ft/year), this magnitude constitutes a 14.5% increase in evaporation. While this is a gross oversimplification, it provides a semi-realistic estimation of increased reservoir losses due to warming.



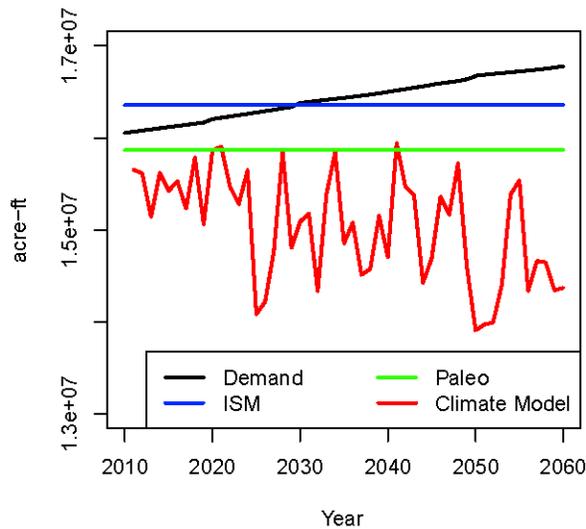
**Figure 72** Boxplots of mean, variance, skew, max and min values for annual natural flow at Lees Ferry, AZ, from traces comprising the three input hydrology data sets. “I” denotes ISM, “P” denotes paleo and “C” denotes climate model hydrology data.



**Figure 73** Boxplots representing the Lees Ferry, AZ, natural flow PDFs based on traces from the three input hydrology data sets.

## 6.6 Results and Discussion

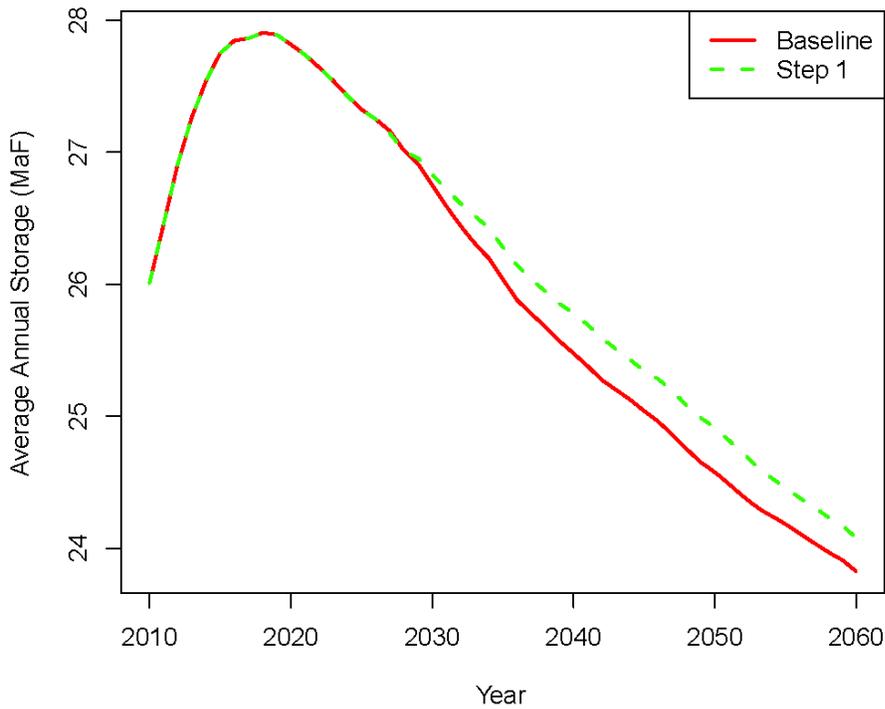
The results from the modeling conducted show an overall net system benefit as a result of the alternate management policies. From the step 1 improvements, system storage is increased through evaporative savings; Lake Mead has a lower probability of being below the SNWA minimum intake elevation, and the magnitude and probability of shortages is reduced. Furthermore, the decadal projections, coupled with the groundwater bank offline storage, provide an additional reserve of water at effectively no impact on other outcomes (e.g. Powell and Mead storage, etc.). This speaks to the effectiveness of capturing otherwise lost water relative to the step 1 outcomes. However, for all the results presented, system reliability overall is decreasing across the model horizon. Shortages are more frequent, overall storage is reduced and so on. For the ISM and paleo input hydrology data, this is due to increasing demands as the Upper Basin continues to develop towards its full allocation. In the case of the climate model hydrology input, this is compounded upon by a reduction in mean flow through time. The disconnect between supply and demand for the various hydrology inputs is estimated in Figure 74.



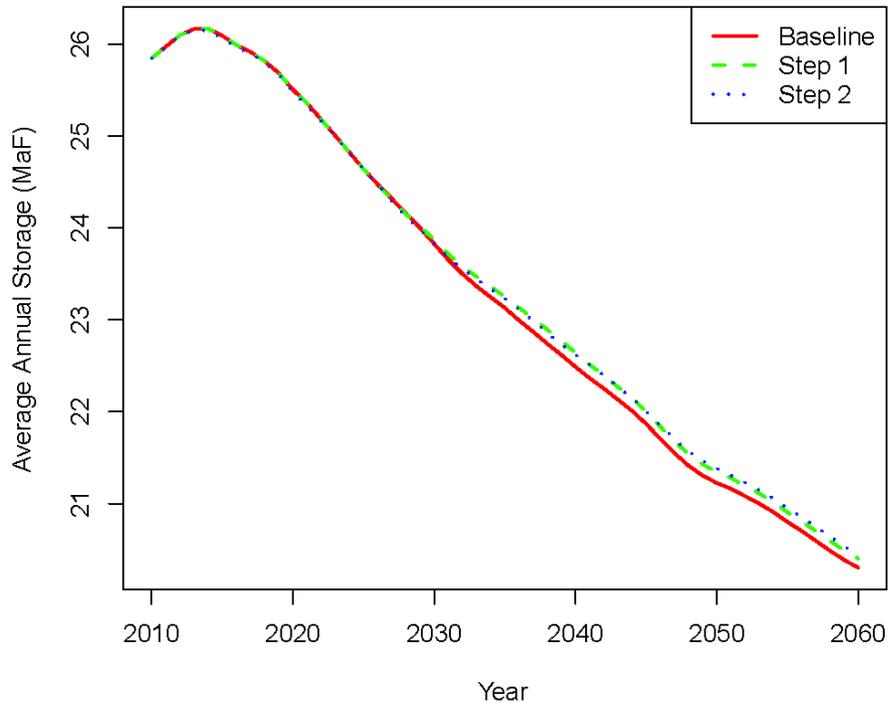
**Figure 74 Projected Colorado River Basin system demand (black), computed as sum of projected state depletions [United States Department of the Interior, 2007b], reservoir evaporation and losses below Hoover Dam. Also shown are Colorado River Basin average yields for the three input hydrology data sets, all estimated as annual natural flow at Imperial Dam.**

Figure 75, Figure 76 and Figure 77 show combined storage of Lakes Powell and Mead over the three input hydrologic scenarios. Recall that for the ISM hydrology only the baseline and step 1 policies could be applied. Relative to the baseline, the policy alternatives result in higher average storage over time. At the end of the model horizon this reaches as much as 0.5 MaF. While the increases in storage are somewhat modest, it is important to note that benefits of the alternate policies are largely devoted to reducing shortage magnitude and frequency, not increasing stored water. Additionally, the differences between policy step 1 and step 2 are negligible, which serves as a check that the introduction of the offline groundwater bank does not have an adverse effect on the storage in Powell and Mead. Additionally, Figure 78 and Figure 79 show average annual accumulation of storage in the groundwater bank for the paleo and climate model input hydrology data respectively. The accumulation rates are shown as though

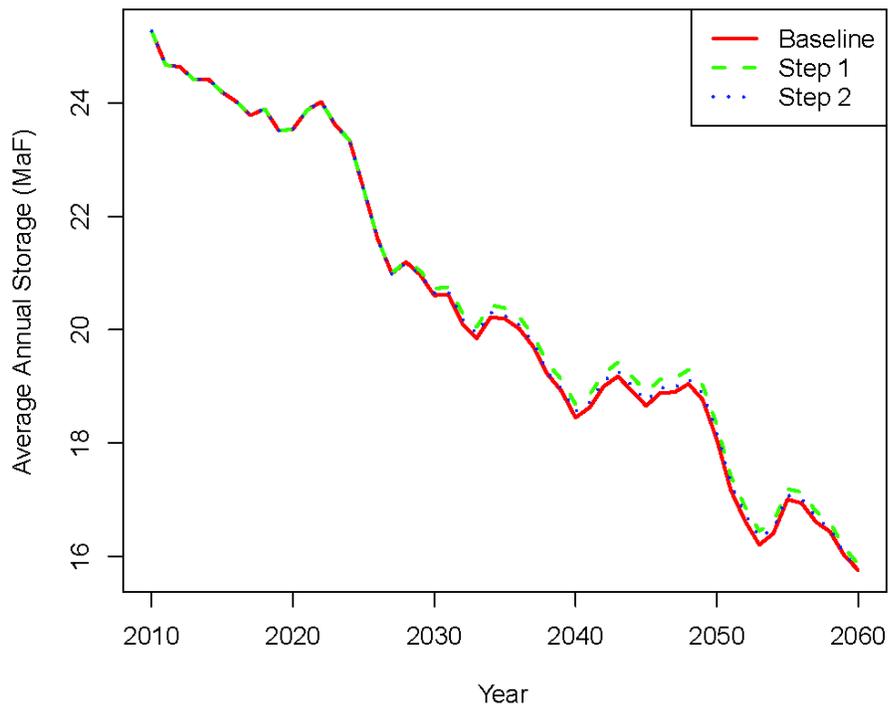
the storage was never used to mitigate shortage, in order to highlight the average amount of additional water that is retained in storage across a trace. For the paleo data, approximately 800,000 acre-ft on average is retained. For the climate model data, the total is almost 1,300,000 acre-ft, even though the mean flow is decreasing in that data set. The difference is attributed to the greater variance of the climate model hydrology data, such that extra system storage, in the form of a groundwater bank, is even more valuable.



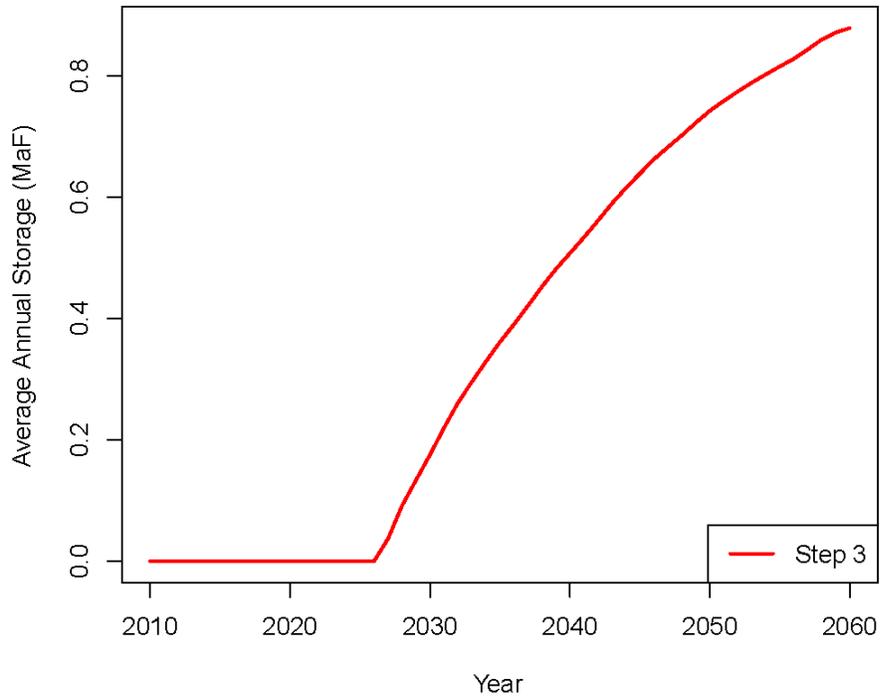
**Figure 75 Average annual combined storage for Lake Powell and Lake Mead based on ISM input data.**



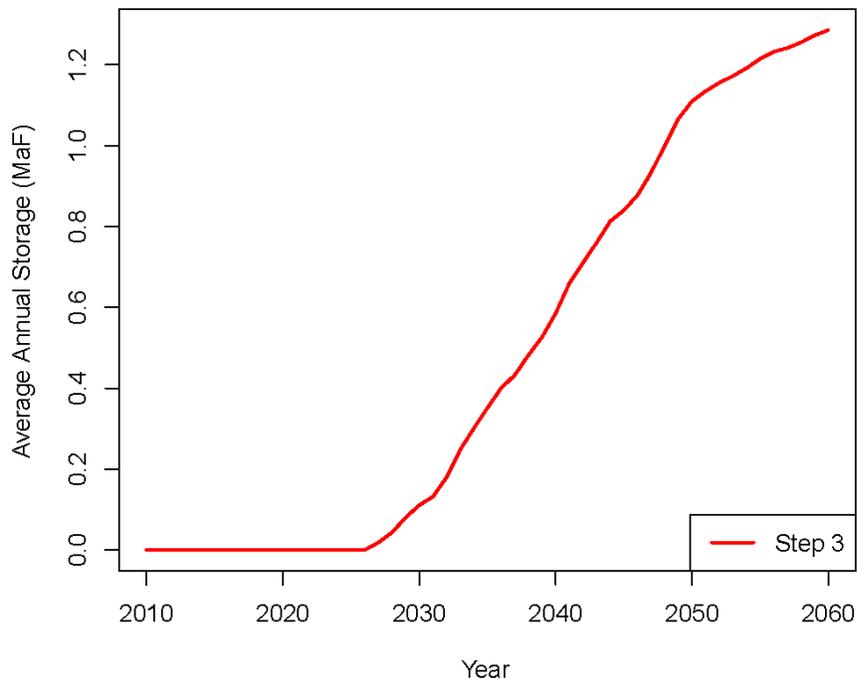
**Figure 76 Average annual combined storage for Lake Powell and Lake Mead based on paleo input data.**



**Figure 77 Average annual combined storage for Lake Powell and Lake Mead based on climate model input data.**

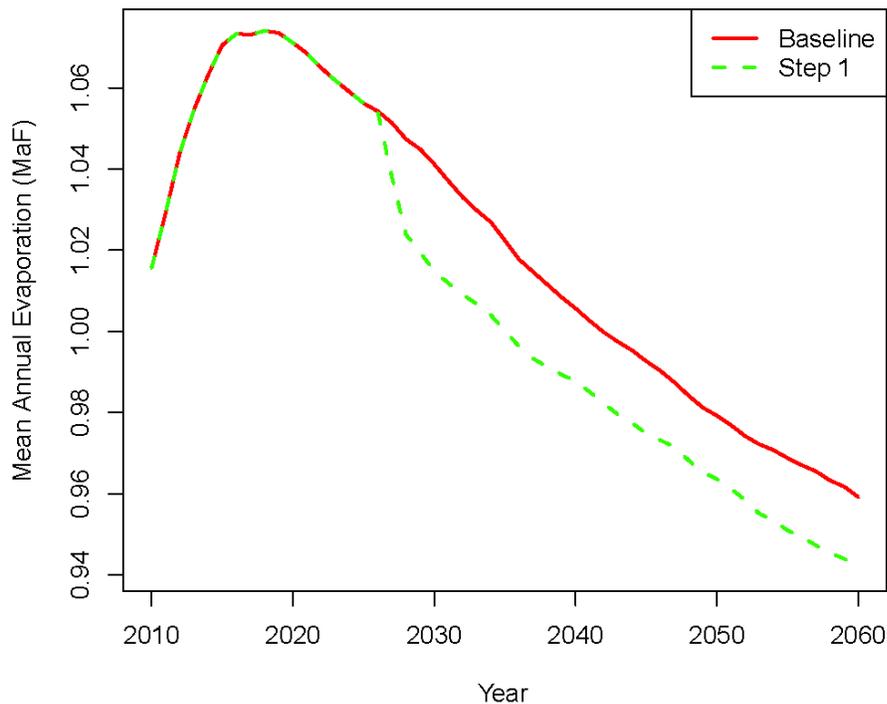


**Figure 78 Average annual groundwater accumulation for paleo input hydrology, were it not applied to mitigate shortage conditions.**

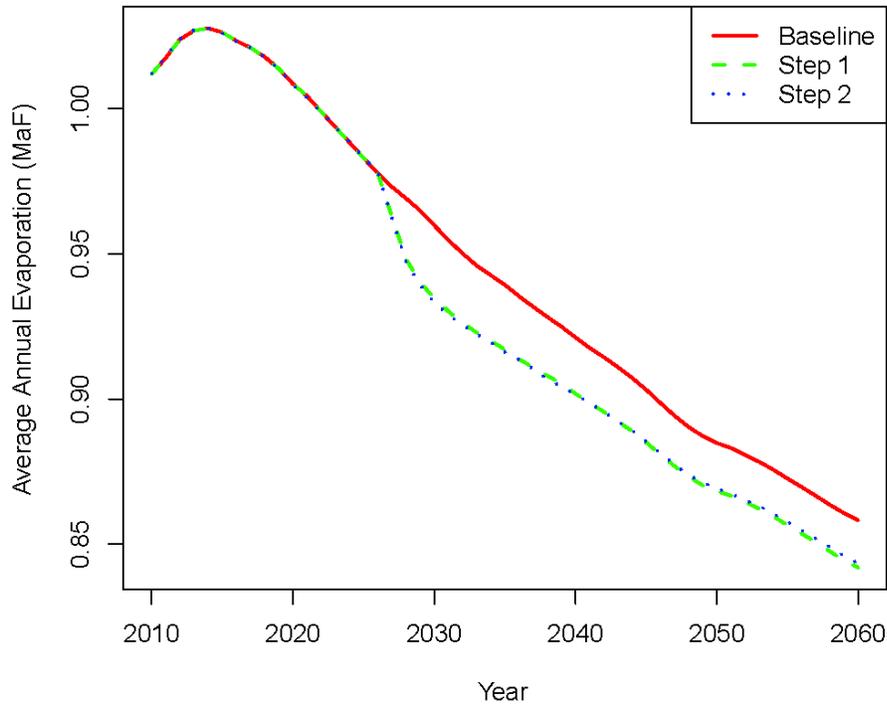


**Figure 79 Average annual groundwater accumulation for climate model input hydrology, were it not applied to mitigate shortage conditions.**

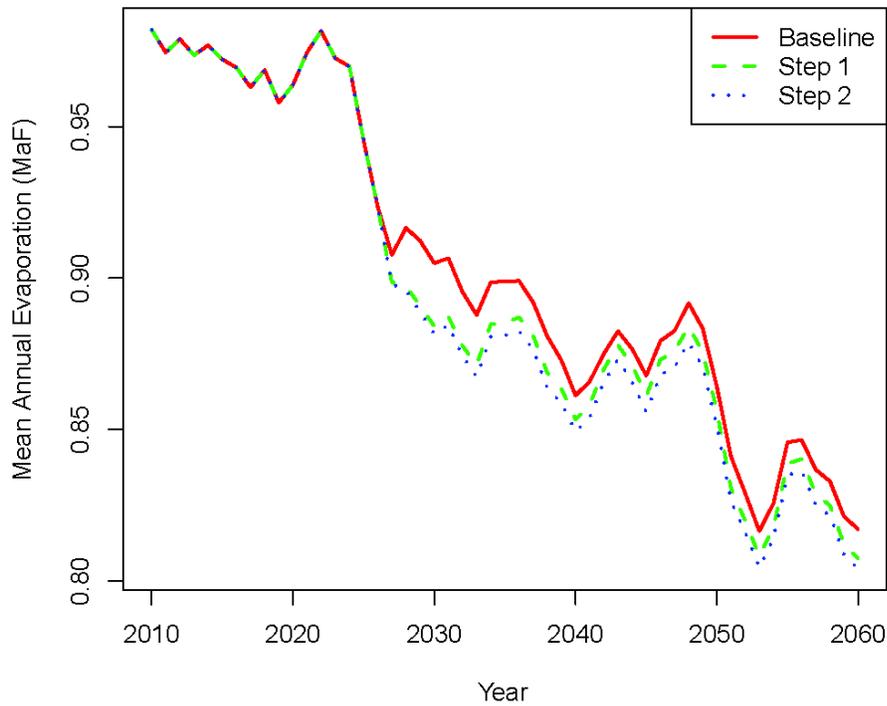
Related to the increased average storage are the evaporative losses associated with the different operational policies. Figure 80, Figure 81 and Figure 82 show average evaporative losses across the modeling horizon for the various policies and input hydrology data. In Figure 80, which shows the results from the ISM data, it can be seen that after the interim period the step 1 policy quickly reduces evaporation relative to the baseline. This is due to the shift towards storing water preferentially in Lake Powell at a lower evaporation rate. Similar results are seen for the paleo input data. For the climate model based data, note that the evaporation savings appear to be diminishing through time. While it may appear counter intuitive, this is simply the result of evaporation savings accumulating over time. As storage increases with these savings, there is a larger reservoir from which to evaporate, so evaporation slowly increases accordingly. Recall also that as these traces contain an implicit anthropogenic warming trend; modeled evaporation rates are increased to reflect this warming. These higher evaporation rates make storing water in lower loss reservoirs even more beneficial.



**Figure 80 Average annual combined evaporation from Lake Powell and Lake Mead based on ISM input data.**

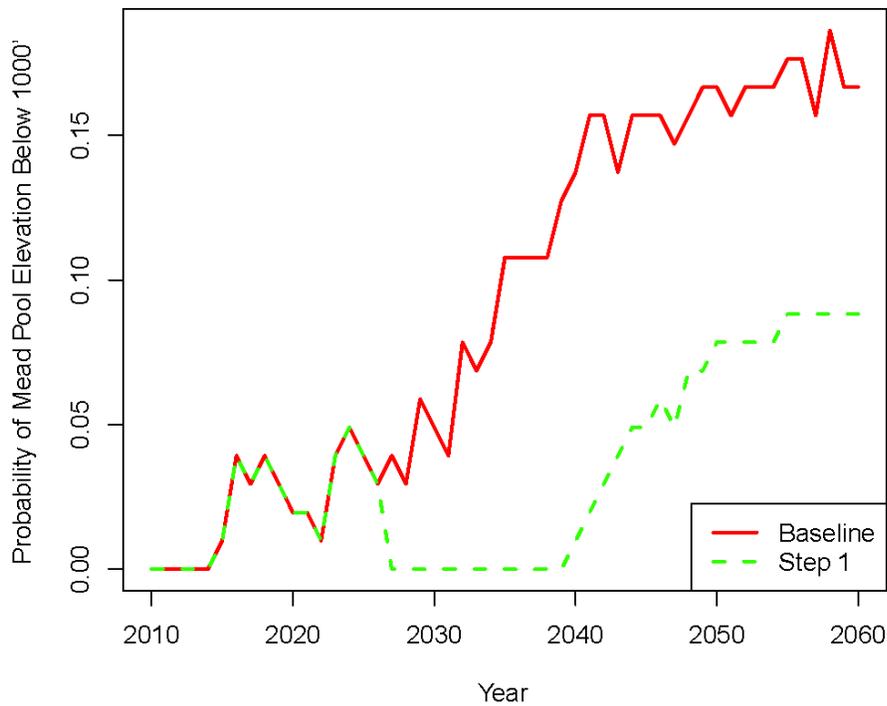


**Figure 81 Average annual combined evaporation for Lake Powell and Lake Mead based on paleo input data.**

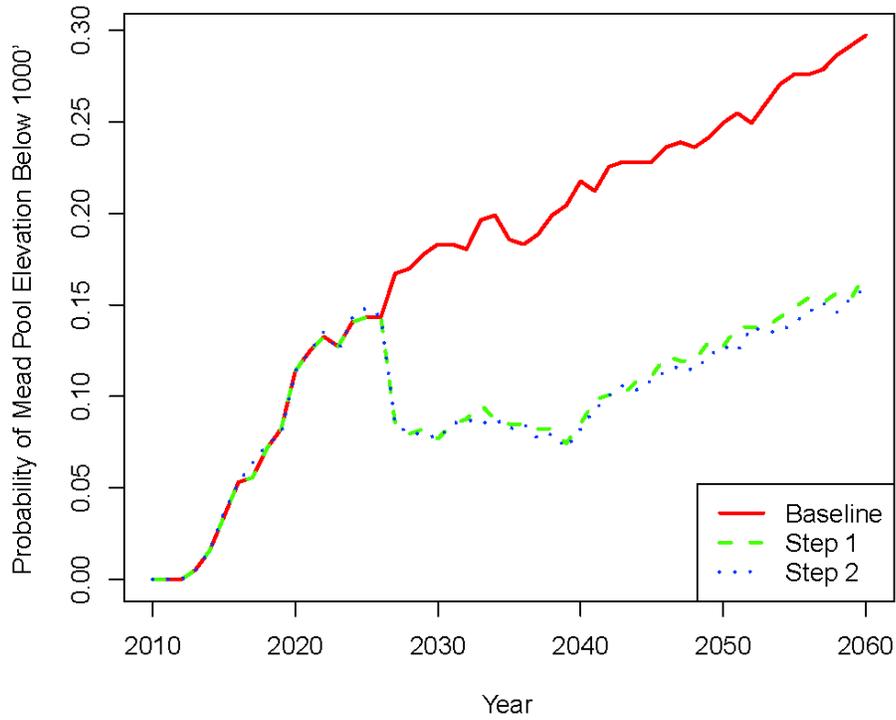


**Figure 82 Average annual combined evaporation for Lake Powell and Lake Mead based on climate model input data.**

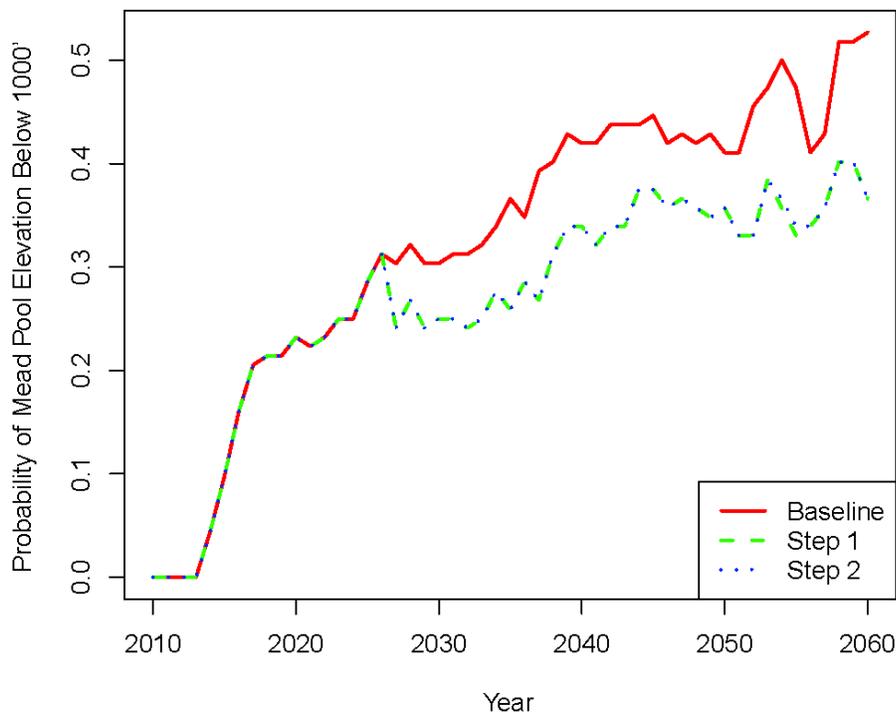
By saving water through lower evaporation rates, more water is available to improve system reliability. As discussed earlier, one objective of the alternate management policies is to maintain a minimum Lake Mead pool elevation of 1000' to ensure reliable pumping conditions for the SNWA intake structure. The additional system storage, coupled with this explicit objective, reduced the probability of the pool elevation dropping below the 1000' threshold for all three hydrology inputs. Specifically, the month of July is presented, as it is traditionally a lower storage month relative to the annual cycle (Figure 83, Figure 84 and Figure 85). These are quite notable, particularly for the paleo and climate input data; at the end of the modeling horizon, risk is reduced by more than 15%.



**Figure 83 Probability of Lake Mead pool elevation below 1000' in the month of July based on ISM input data.**



**Figure 84 Probability of Lake Mead pool elevation below 1000' in the month of July based on paleo input data.**

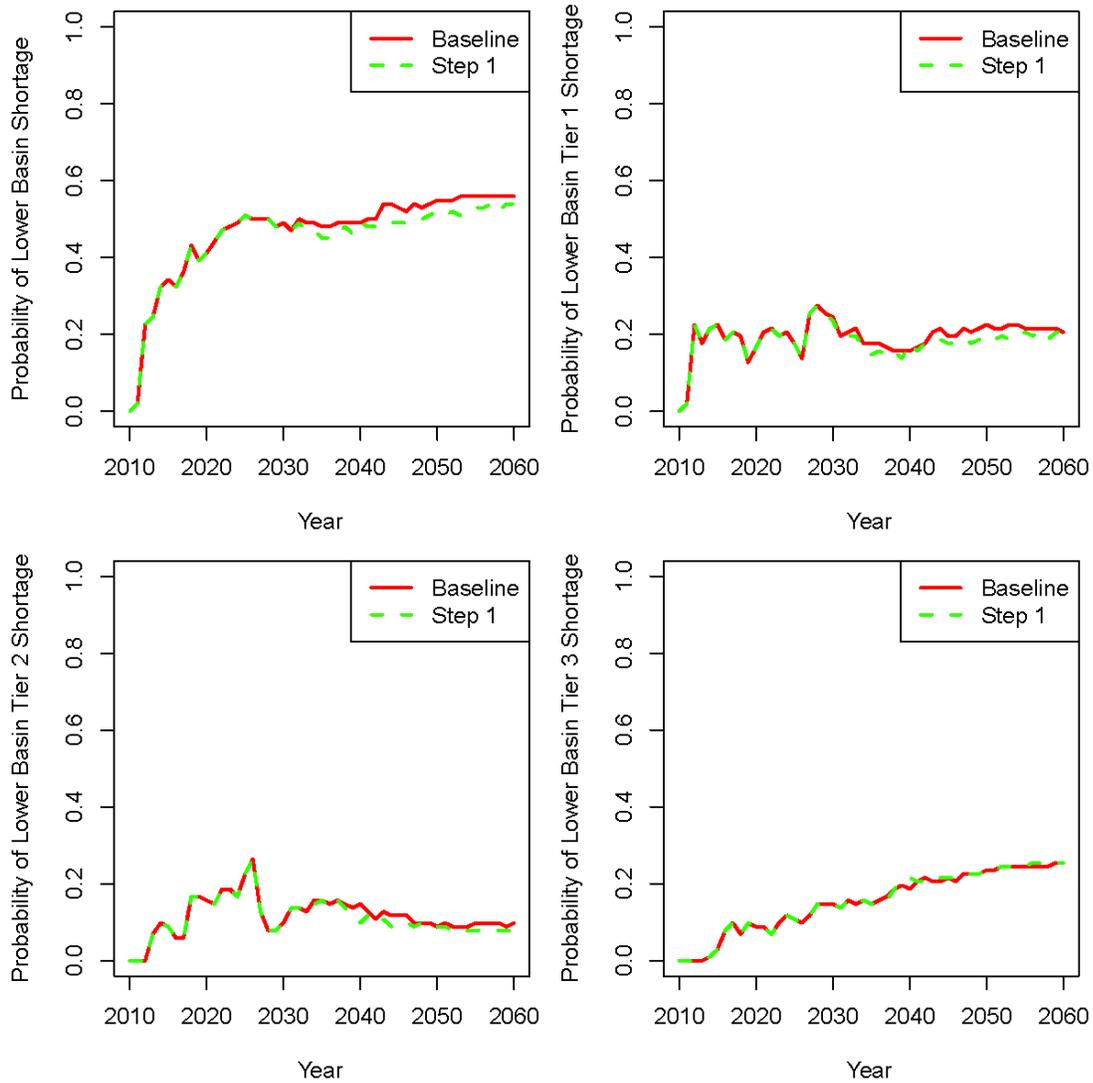


**Figure 85 Probability of Lake Mead pool elevation below 1000' in the month of July based on climate model input data.**

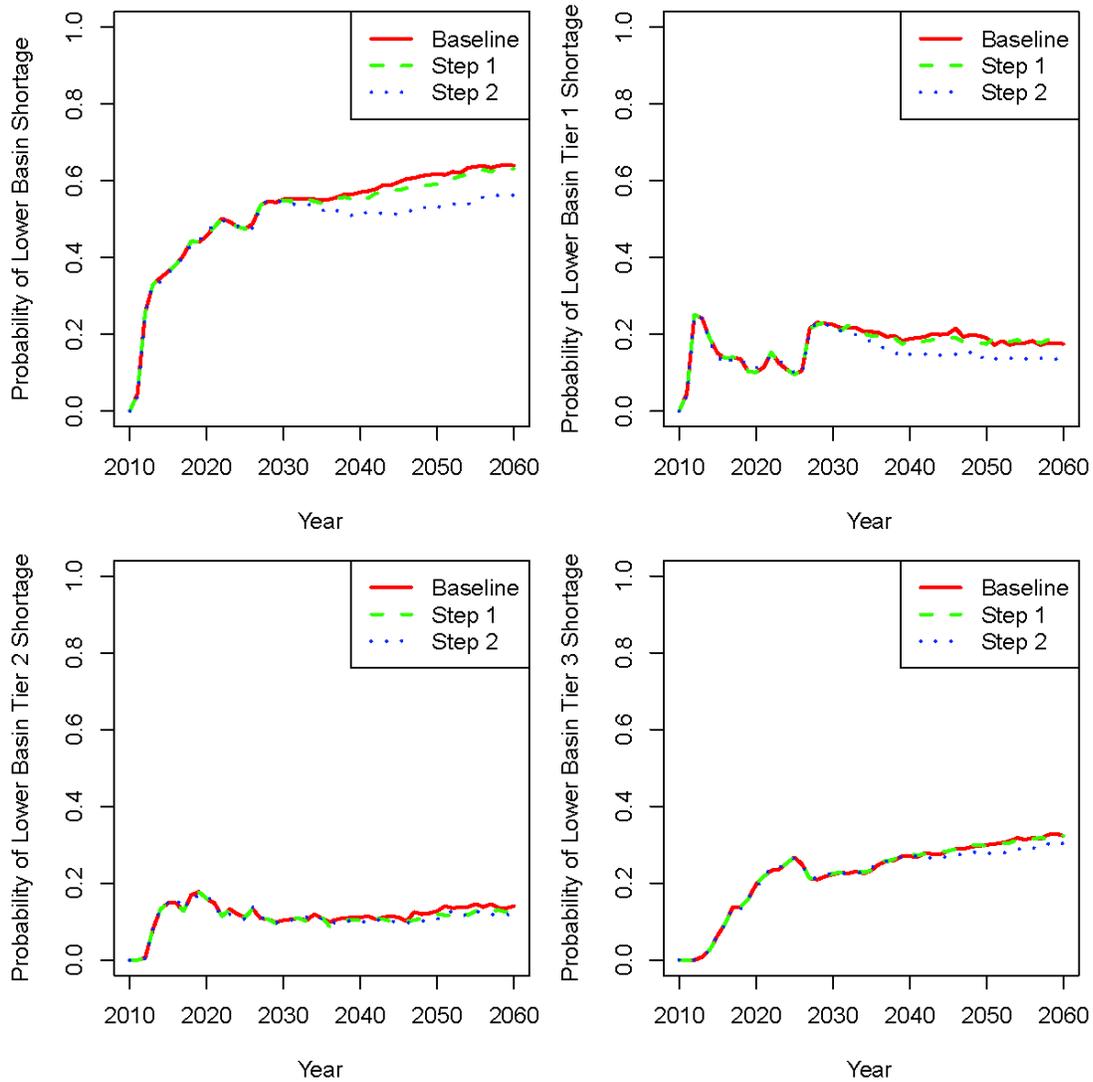
As a compliment to the figures shown above, the probabilities of Lower Basin shortage conditions were also computed. The shortage policies established as part of the interim guidelines contain provisions for three tiers of delivery reductions in the Lower Basin. Tier 1 represents the smallest curtailment, and magnitudes progressively increase, with tier 3 being the largest shortage size. In the figures that follow, probability of any tier shortage is first computed, followed by probabilities specific to the three tiers. In this work, shortage is computed as the difference between annual normally scheduled depletion and actual depletion for the three Lower Basin States (AZ, CA and NV) and Mexico. For all three input hydrology data sets, an overall reduction in shortage probability is seen, albeit small in some instances. Figure 86 shows the results for ISM, and the overall reduction of shortage risk is spread across the three tiers. In Figure 87 (paleo input data), the majority of the risk reduction in the step 1 policy comes from lessening the probability of tier 1 and 2 shortages. The step 2 probabilities show additional reduction of shortage risk by, when possible, supplementing the release from Lake Mead using supply accumulated in the groundwater bank.

Figure 88 shows the same plots for the climate change input data. Shortage probabilities are lower for the step 1 policy relative to the baseline, but only marginally. This is believed to be a product of the substantially reduced flow of the climate model traces, such that even moderate evaporation savings have little mitigating effect relative to the large supply demand disconnect (Figure 74). By capturing excess water from the occasional high volume years, the groundwater bank (policy step 2) offers a slightly larger reduction in shortage risk, mostly by negating some of the tier 1 shortages. For both the paleo and climate change input data, the groundwater bank was not always exhausted by the shortage mitigation efforts. Recall that policy step 1 and step 2 only take effect for the last 34 years of the model horizon. This period may not be long enough to

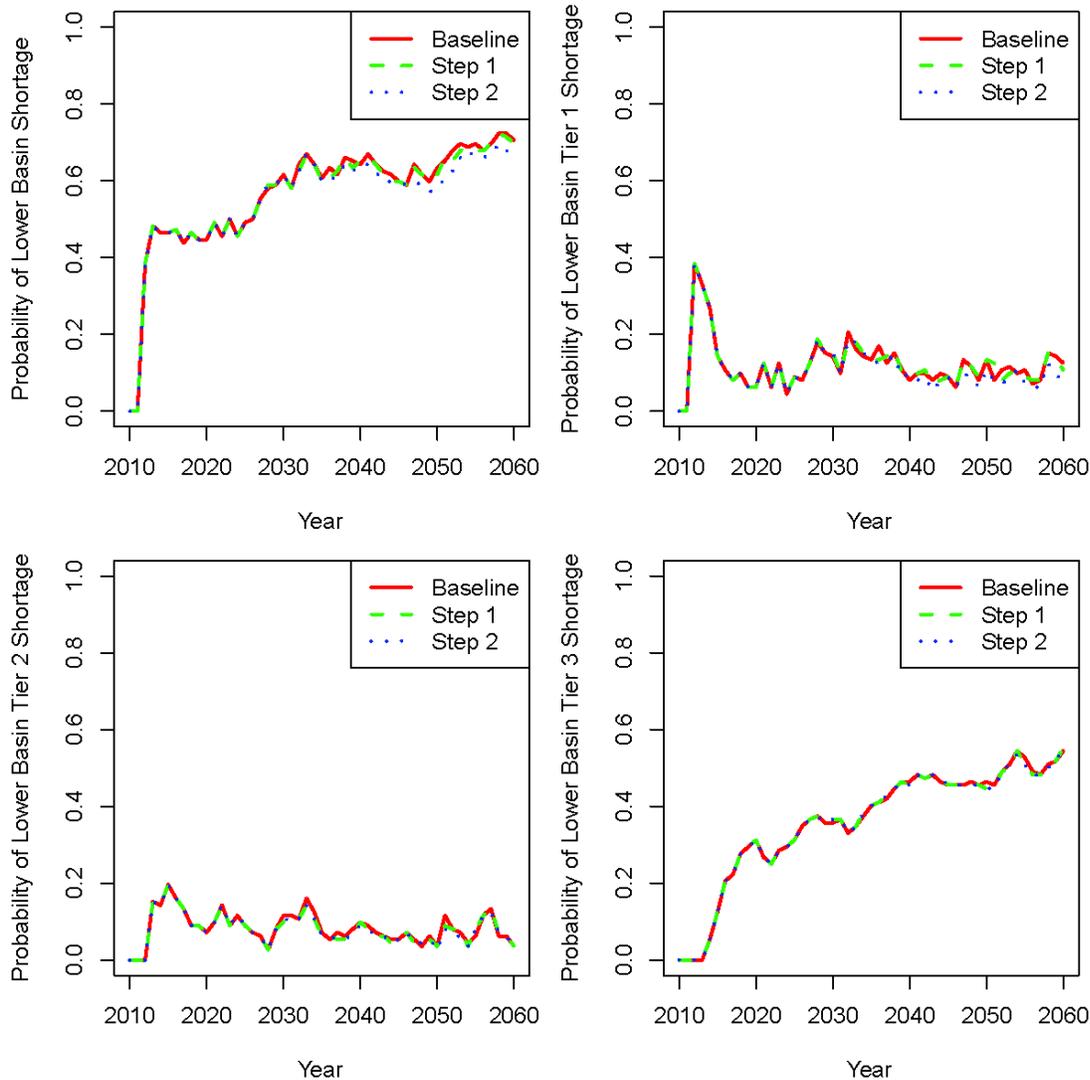
fully experience a surplus period followed by substantial shortage conditions. As such, more benefit from the offline storage concept might be seen with a longer horizon that fully allows groundwater to accumulate and later be depleted as part of the shortage mitigation tactics.



**Figure 86 Probability in the Lower Basin of any shortage (upper left), tier 1 shortage (upper right), tier 2 shortage (lower left) and tier 3 shortage (lower right), based on ISM input data.**



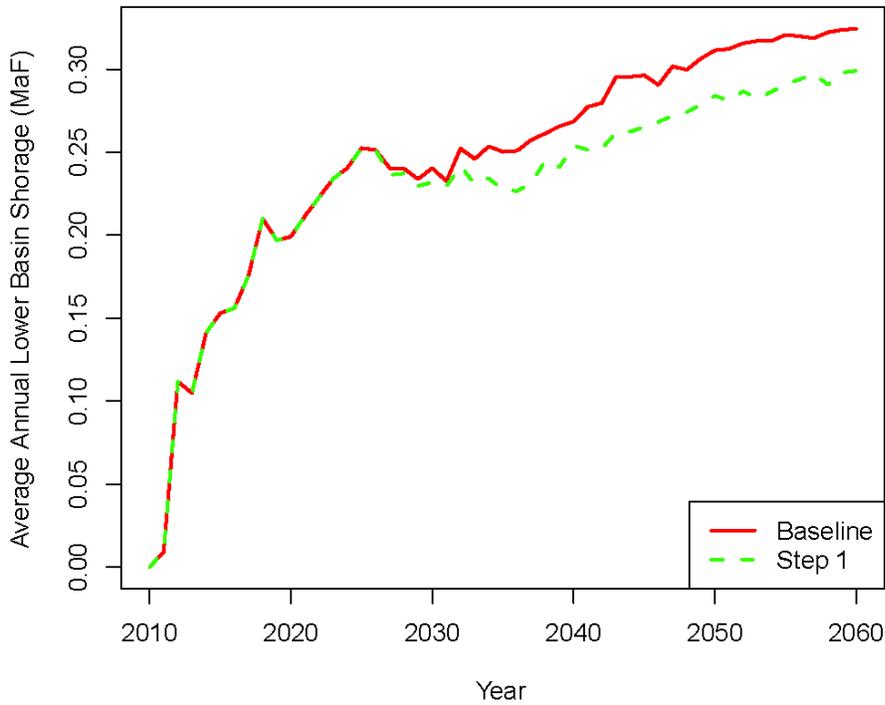
**Figure 87 Probability in the Lower Basin of any shortage (upper left), tier 1 shortage (upper right), tier 2 shortage (lower left) and tier 3 shortage (lower right), based on paleo input data.**



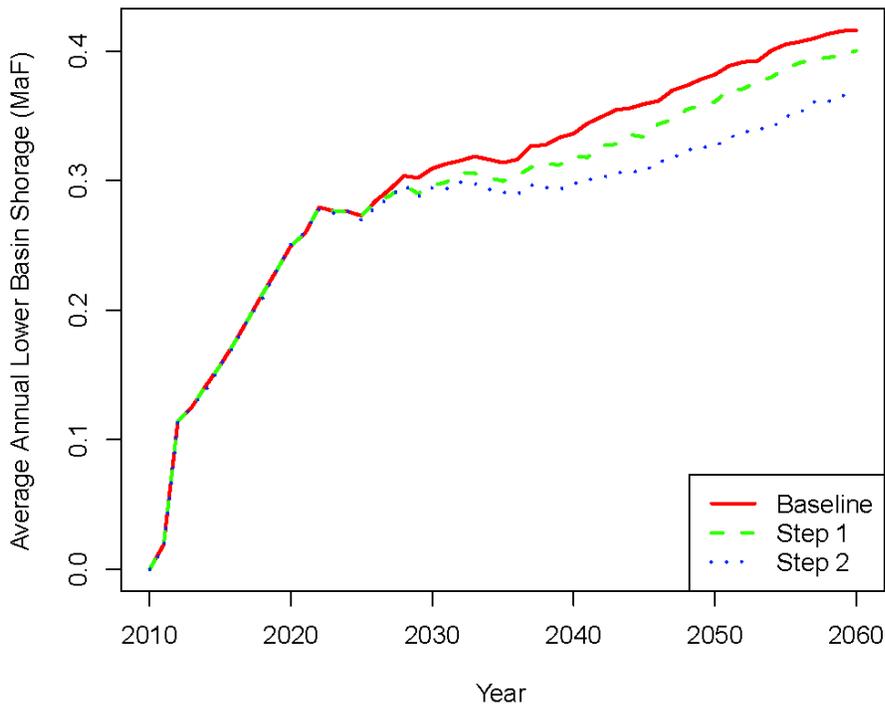
**Figure 88 Probability in the Lower Basin of any shortage (upper left), tier 1 shortage (upper right), tier 2 shortage (lower left) and tier 3 shortage (lower right), based on climate model input data.**

Last, average shortage magnitudes are shown over the modeling horizon. Shortage magnitude reductions under the various input hydrology data are consistent with the risk reductions shown in earlier figures (see Figure 89, Figure 90 and Figure 91). As was the case for the earlier shortage figures, when possible water from the groundwater bank was used to mitigate shortage conditions in policy step 2. For the ISM and paleo data, annual average shortage reductions are as much as 50,000 acre-ft. Results for the climate model input data also show reductions in average shortage, but are marginal relative to the baseline. As these shortages are

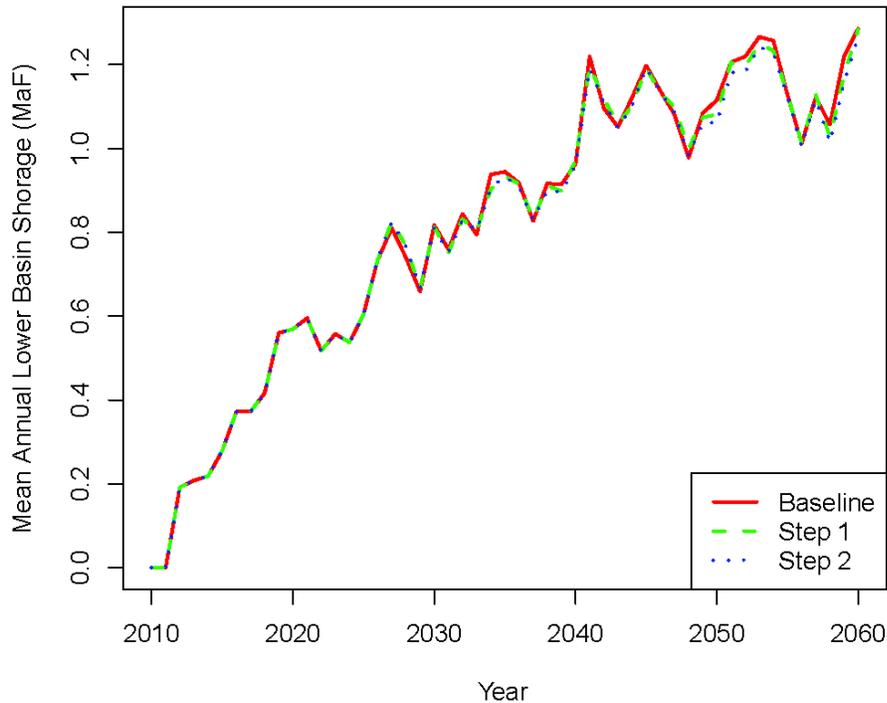
roughly four times those of the ISM or paleo input data, similar evaporative and “banked” savings have a considerably reduced mitigating effect.



**Figure 89** Average annual Lower Basin shortage magnitude based on ISM input data.



**Figure 90** Average annual Lower Basin shortage magnitudes based on paleo input data.

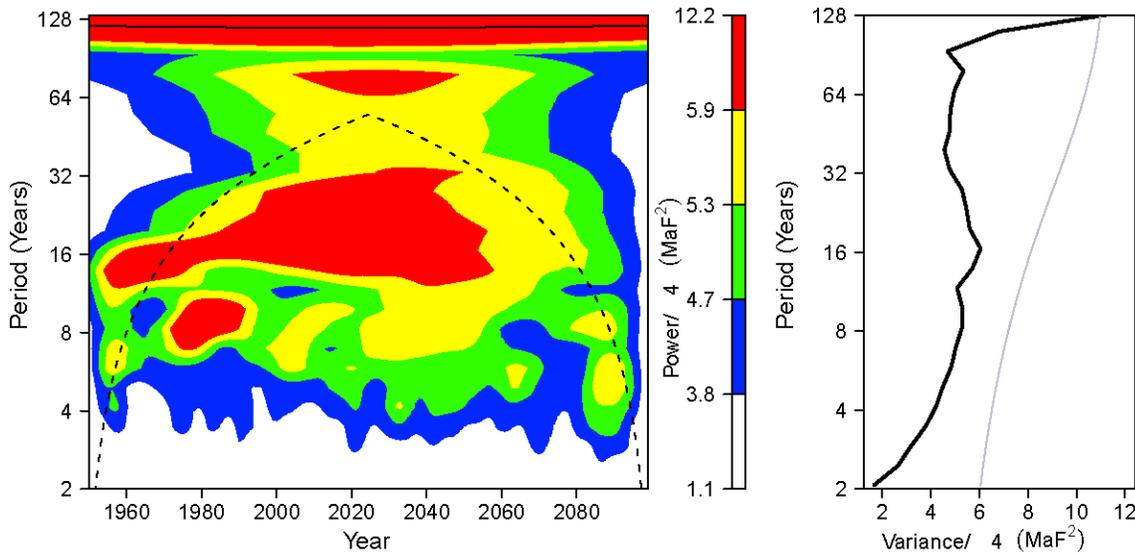


**Figure 91 Average annual Lower Basin shortage magnitudes based on climate model input data.**

In conclusion, the results of this chapter effectively demonstrated the utility of alternate management policies in the Colorado River Basin for the purpose of mitigating shortage and increasing supply reliability. Furthermore, it was shown that there is a high potential to utilize decadal flow regime projections as guidance in these policies, thus yielding improved outcomes. This is particularly notable in future scenarios containing heightened flow variability, where water can be preemptively diverted to offline supplemental storage ahead of a wet event to be later used to mitigate shortage conditions.

While the results shown support this, it is important to recognize that the projection framework relies largely on the persistence of a low frequency variability signal. From spectral analysis of the climate model hydrology data, there was little agreement between traces regarding significant future modes of variability. As such, the effectiveness of the projection method could be underrepresented in CRSS results driven by the climate model hydrology data. Alternatively, this could signal decreased skill for this projection framework if low frequency

variance fails to persist. Figure 92 shows the average local and global wavelet power spectra for the 112 climate model hydrology traces. The local and global spectra were computed for each trace and subsequently averaged together. The only significant feature is at the 128 year period and is the decreasing flow trend through time. The lack of any other significant features highlights the little persistence of variability modes across the different traces.



**Figure 92 Average local and global wavelet power spectra of the 112 climate model hydrology traces for the period of 1950 to 2099. The grey line is a 90% confidence background white noise power spectrum.**

Simplifications, such as electing not to consider implications for hydropower, recreation or other system objectives were decided upon in the interest of clearly highlighting benefits of the adaptation strategies through the isolation of a single outcome – supply reliability. Additionally, the unrestricted groundwater bank size and infiltration capacity with zero loss is unrealistic, but helps to showcase potential benefits.

Although the shortage probability and magnitude reductions were not as sizeable as had been hoped for, particularly in the case of the climate model based input data, it is important to remember that other outcomes had substantial improvements. Particularly, the probability of

Mead pool elevation below 1000' is greatly reduced. This clearly indicates that the policies are producing improved system operations. The objective of maintaining a 1000' pool elevation is to ensure that Las Vegas can reliably divert their apportioned water, which is a key, albeit small component of the Lower Basin demand. If water could be extracted from Lake Mead at lower pool elevations, greater evaporative savings would undoubtedly be effected under the proposed adaptive policies.

## **CHAPTER 7: SUMMARY**

The central theme of the work presented in this dissertation is the variability, particularly at the decadal timescale, in the Colorado River Basin as it pertains to water resources planning and management. As established throughout the chapters, this scale variability has and will continue to be the leading driver of water supply risk for the Basin. Although climate change may slowly decrease the annual yield, this trend is superseded in magnitude to the flow variability. Additionally, climate models offer little skill in short to mid-term outlook projections. To this end, stochastic simulation methods aimed at accurately representing multi-site spectral properties of historical data were developed and presented in chapters 2 and 3. Following, in chapter 4, a systematic analysis of the significant variability modes for flow at Lees Ferry offered physical mechanisms associated with the spectral features and further proposed links with large-scale climate phenomena. From the insights gained in earlier chapters, a robust method for producing decadal streamflow regime projections was developed and demonstrated in chapter 5. Last, chapter 6 highlighted the utility of these projections, with application to the Colorado River Basin, as part of adaptive management policies in a system modeling exercise.

### **7.1 Review of Work Presented and Considerations**

The sections that follow provide a brief recap of the key contributions and limitations of the work presented in each chapter.

#### **7.1.1 Chapter 2**

In chapter 2, a new non-parametric disaggregation method was presented. The method is effective for multi-site disaggregation of streamflow at a variety of time scales. Most notable is the ability to produce daily flows at multiple locations from a single annual value. Additionally,

relative to other parametric and nonparametric methods, this approach is simple and computationally efficient. A wide range of validation metrics was offered, and for most, results were quite good. One measure with area for improvement was the lag correlation structure from year to year, be it at the daily or monthly timescale. However, this is a known drawback associated with most disaggregation approaches and is often considered a small detractor in the overall results evaluation.

### **7.1.2 Chapter 3**

Chapter 3 presented a spectral domain stochastic streamflow simulation tool for data with non-stationary spectral properties. From analysis of flow data in the Colorado River Basin, key variability scales were noted and failed to be reproduced by traditional stochastic methods. Thus, the approach developed utilizes the wavelet transform to identify and model key variability features individually, such that the global and local wavelet spectra are reproduced. Other statistical measures of the historical data such as mean, variance, etc. were also well captured by the simulations. Last, by coupling this method with the disaggregation of chapter 2, spectral properties of upstream locations from Lees Ferry were reproduced without any additional conditioning. As the periodic spectral components were simulated with auto-regressive models, an underlying assumption of normality was implied. While this is not exactly the case for streamflow at Lees Ferry, the ability to accurately simulate the time-evolution of variability modes likely outweighs this drawback.

### **7.1.3 Chapter 4**

In this chapter, spectral and other methods of analysis were used to detect and attribute the key streamflow variability modes of the Colorado River with components of the hydrologic system; namely temperature and precipitation. Low frequency variability had strong ties with

temperature fluctuations, while decadal variability over the past 30 years was associated with precipitation. Links between these variability modes and climate indices were also proposed but require additional study to solidify the relationships. Furthermore, paleo reconstructed data and idealized climate model experiments were examined to provide additional perspective. While promising empirical associations were presented, particularly with regard to the links to climatic phenomena, additional work remains to establish physical mechanisms responsible for the relation.

#### **7.1.4 Chapter 5**

In chapter 5, a decadal flow regime projection approach was developed and validated. The method builds from the flow variability understanding from earlier chapters and is conditioned upon features in the paleo reconstructed streamflow data. Specifically, the method is a two component hybrid that 1) resamples from the paleo data conditioned on spectral features to project future variance and 2) employs regression to relate average flow from a previous period with that of a future window. While the projections were not always effective at capturing the exact magnitude of a future epoch, the general trends of regime shifts were well reflected in the results. This method is unique to the Colorado Basin as it queues off of key aspects of the long paleo reconstructed streamflow record available for the Basin. Additional research may improve the resolution of these projections, as it has been demonstrated that a wealth of data is contained within the paleo record.

#### **7.1.5 Chapter 6**

In this chapter, a system level analysis of Colorado River Basin was conducted under a combination of input flow hydrology data and operational policies. The focus was to use alternate management strategies that capitalize on decadal projection information to improve

supply reliability and mitigate the effects of shortage. The policies first reduced system losses due to evaporation by storing more water in a reservoir with lower evaporative losses. Next, a strategic groundwater reserve was introduced and diversion of water to this offline storage was guided by decadal flow projections with the goal of capturing flows that would have otherwise been lost as spill. This groundwater bank was used to offset shortages associated with declining storage across the various model runs. Both aspects of the adaptive management policies were effective at reducing the impact of shortages and increasing overall reliability. For these system modeling experiments, many liberties were taken to simplify the objectives and highlight potential benefits of alternate management strategies, specifically the decadal projection information. Obviously, this was not a realistic representation of the system, nor was it meant to be.

## **7.2 Future Direction of Research**

Across the work presented in this dissertation, there exist several opportunities for additional research and further work, which are listed below.

- Additional climate modeling experiments are needed to better understand and establish the proposed links between large scale climate features and the hydro-climate of the Colorado Basin.
- Improved resolution in the decadal regime projections should be investigated to possibly offer information that can be used in better managing smaller storage systems.
- A more detailed investigation of the Colorado River Basin, considering the various system objectives and dynamics is needed for a more realistic perspective on potential benefits of adaptation strategies.

In addition to the aforementioned further research topics, one large uncertainty, in light of expected climate change, is the future of Colorado River flow variability. Based on the mostly stationary paleo reconstructed and historical flow data, a variability “domain” has been established and several scales of variance are noteworthy. However, will these features persist in the future? General Circulation Models (GCMs) struggle to produce quasi-periodic climate features such as El Niño Southern Oscillation. If these indices are responsible, in part, for the hydroclimatic variability in the Colorado Basin, it is important to consider if they are 1) reflected in the results of the GCMs and 2) if the nature of these phenomena may alter due to anthropogenic climate change. The climate model based input hydrology data used in chapter 6 modeling indicates greater future variability. However, from spectral analysis of these data, no persistent variability structure or features were found. Thus, understanding the impact of climate change on Colorado River Basin streamflow variability is a logical progression of the work presented in this dissertation.

### **7.3 Perspective on Colorado River Flow Variability**

Recently, a flurry of publications investigated future prospects for the Colorado River Basin under climate change [*Barnett and Pierce, 2008; Barnett and Pierce, 2009; Barsugli et al., 2009; Rajagopalan et al., 2009*]. The projections ranged from dire – Lake Mead possibly going dry by 2021 [*Barnett and Pierce, 2008*] to more tempered, but still raise considerable concern [*Rajagopalan et al., 2009*]. In part, these ominous outlooks were heightened by the depressed initial reservoir conditions related to the recent drought. Reservoir capacity has been highlighted over and over in this work as the backbone of Colorado River supply reliability due to the sizeable year to year and multi-year variability seen in streamflow records. While the recent drought has strained the Colorado Basin more so than any other period, it has also showcased the

system resiliency. Throughout this unprecedented string of low flow years, the system never failed to meet delivery obligations in the Lower Basin.

Furthermore, at the time of this writing, the Upper Colorado River Basin snowpack is quite substantial and the April through July unregulated inflow to Lake Powell is forecasted to be 140% of average. Although this one year alone will not “reset” the effects of the drought, it will raise reservoir levels considerably. The result is that the probability of Lower Basin shortage conditions is effectively reduced to zero through 2015 and remains fairly moderate for the rest of the decade. This is quite a different story compared to the projection of Lake Mead possibly going dry by 2021. The takeaway from this, and the dissertation as a whole, can be summarized as follows: Colorado River flow is notoriously variable and it is this variability which will continue to stress resources in the Basin. However, the same variability will also replenish and restore the system with occasional high flow periods. Thus, robust management of this resource requires a keen physical and statistical understanding of flow variability which can be used in adaptation to constantly changing conditions and objectives.

## REFERENCES

- Addison, P. (2002), *The Illustrated Wavelet Transform Handbook*, Taylor & Francis, New York.
- An, S. I., and B. Wang (2000), Interdecadal change of the structure of the ENSO mode and its impact on the ENSO frequency, *Journal of Climate*, *13*, 2044-2055.
- Barnett, T. P., and D. W. Pierce (2008), When will Lake Mead go dry?, *Water Resources Research*, *44*.
- Barnett, T. P., and D. W. Pierce (2009), Sustainable water deliveries from the Colorado River in a changing climate, *Proceedings of the National Academy of Sciences*, *106*, 7334-7338.
- Barsugli, J. J., K. Nowak, B. Rajagopalan, J. R. Prairie, and B. Harding (2009), Comment on "When will Lake Mead go dry?" by T. P. Barnett and D. W. Pierce, *Water Resources Research*, *45*, -.
- Bogardi, I., I. Matyasovszky, A. Bardossy, and L. Duckstein (1993), Application of a Space-Time Stochastic-Model for Daily Precipitation Using Atmospheric Circulation Patterns, *Journal of Geophysical Research-Atmospheres*, *98*, 16653-16667.
- Bottomley, M., O. Great Britain. Meteorological, and A. a. P. S. Massachusetts Institute of Technology. Dept. of Earth (1990), Global ocean surface temperature atlas "Gosta", Meteorological Office ; Massachusetts Institute of Technology, Bracknell [England] U.K.; [Cambridge].
- Bras, R. L., and I. Rodríguez-Iturbe (1985), *Random functions and hydrology*, xv, 559 p. pp., Addison-Wesley, Reading, Mass.
- Brown, D. P., and A. C. Comrie (2004), A winter precipitation 'dipole' in the western United States associated with multidecadal ENSO variability, *Geophysical Research Letters*, *31*, -.
- Cayan, D. R. (1996), Interannual climate variability and snowpack in the western United States, *Journal of Climate*, *9*, 928-948.
- Cayan, D. R., M. D. Dettinger, H. F. Diaz, and N. E. Graham (1998), Decadal Variability of Precipitation over Western North America, *Journal of Climate*, *11*, 3148-3166.
- 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, *Bulletin of the American Meteorological Society*, *82*, 399-415.
- Cayan, D. R., K. T. Redmond, and L. G. Riddle (1999), ENSO and hydrologic extremes in the western United States, *Journal of Climate*, *12*, 2881-2893.
- Chang, T. J., M. L. Kavvas, and J. W. Delleur (1984), Daily Precipitation Modeling by Discrete

Autoregressive Moving Average Processes, *Water Resources Research*, 20, 565-580.

Chin, E. H. (1977), Modeling Daily Precipitation Occurrence Process with Markov-Chain, *Water Resources Research*, 13, 949-956.

Christensen, N. S., and D. P. Lettenmaier (2007), A multimodel ensemble approach to assessment of climate change impacts on the hydrology and water resources of the Colorado River Basin, *Hydrology and Earth System Sciences*, 11, 1417-1434.

Clark, M. P., M. C. Serreze, and G. J. McCabe (2001), Historical effects of El Nino and La Nina events on the seasonal evolution of the montane snowpack in the Columbia and Colorado River Basins, *Water Resources Research*, 37, 741-757.

Colorado River Commission., Arizona., California., Colorado., Nevada., New Mexico., Utah., Wyoming., and H. Hoover (1923), *Colorado River compact*, ii, 5 p. pp., Govt. print. off., Washington,.

Coulibaly, P., and D. H. Burn (2004), Wavelet analysis of variability in annual Canadian streamflows, *Water Resources Research*, 40, -.

Daly, C., R. P. Neilson, and D. L. Phillips (1994), A Statistical Topographic Model for Mapping Climatological Precipitation over Mountainous Terrain, *Journal of Applied Meteorology*, 33, 140-158.

Dettinger, M. D., D. R. Cayan, H. F. Diaz, and D. M. Meko (1998), North-South Precipitation Patterns in Western North America on Interannual-to-Decadal Timescales, *Journal of Climate*, 11, 3095-3111.

Diaz, H. F., M. P. Hoerling, and J. K. Eischeid (2001), ENSO variability, teleconnections and climate change, *International Journal of Climatology*, 21, 1845-1862.

Diaz, H. F., and V. Markgraf (1992), *El Niño : historical and paleoclimatic aspects of the southern oscillation*, xii, 476 p. pp., Cambridge University Press, Cambridge [England] ; New York, NY, USA.

Diaz, H. F., and V. Markgraf (2000), *El Niño and the southern oscillation : multiscale variability and global and regional impacts*, xv, 496 pp., Cambridge University Press, Cambridge, U.K. ; New York.

Dracup, J. A., and E. Kahya (1994), The Relationships between United-States Streamflow and La-Nina Events, *Water Resources Research*, 30, 2133-2141.

Enfield, D. B., A. M. Mestas-Nunez, and P. J. Trimble (2001), The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental US, *Geophysical Research Letters*, 28, 2077-2080.

Fetter, C. W. (2001), *Applied Hydrogeology*, 4th ed., Prentice Hall, Upper Saddle River.

Gershunov, A., and T. P. Barnett (1998), ENSO influence on intraseasonal extreme rainfall and temperature frequencies in the contiguous United States: Observations and model results (vol 11, pg 1575, 1998), *Journal of Climate*, 11, 3062-3065.

Gober, P., and C. W. Kirkwood (2010), Vulnerability assessment of climate-induced water shortage in Phoenix, *Proceedings of the National Academy of Sciences of the United States of America*, 107, 21295-21299.

Graham, N. E. (1994), Decadal-Scale Climate Variability in the Tropical and North Pacific during the 1970S and 1980S - Observations and Model Results, *Climate Dynamics*, 10, 135-162.

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.

Grantz, K., B. Rajagopalan, M. Clark, and E. Zagona (2007), Seasonal shifts in the North American monsoon, *Journal of Climate*, 20, 1923-1935.

Gray, S. T., L. J. Graumlich, J. L. Betancourt, and G. T. Pederson (2004), A tree-ring based reconstruction of the Atlantic Multidecadal Oscillation since 1567 AD, *Geophysical Research Letters*, 31, -.

Grinsted, A., J. C. Moore, and S. Jevrejeva (2004), Application of the cross wavelet transform and wavelet coherence to geophysical time series, *Nonlinear Processes in Geophysics*, 11, 561-566.

Grygier, J. C., and J. R. Stedinger (1988), Condensed Disaggregation Procedures and Conservation Corrections for Stochastic Hydrology, *Water Resources Research*, 24, 1574-1584.

Hamlet, A. F., and D. P. Lettenmaier (1999), Columbia River streamflow forecasting based on ENSO and PDO climate signals, *Journal of Water Resources Planning and Management-Asce*, 125, 333-341.

Hare, S. R., and N. J. Mantua (2000), Empirical evidence for North Pacific regime shifts in 1977 and 1989, *Progress in Oceanography*, 47, 103-145.

Hidalgo, H. G. (2004), Climate precursors of multidecadal drought variability in the western United States, *Water Resources Research*, 40, -.

Hidalgo, H. G., and J. A. Dracup (2003), ENSO and PDO effects on hydroclimatic variations of the Upper Colorado River basin, *Journal of Hydrometeorology*, 4, 5-23.

Higgins, R. W., A. Leetmaa, and V. E. Kousky (2002), Relationships between Climate Variability and Winter Temperature Extremes in the United States, *Journal of Climate*, 15, 1555-

1572.

Hoerling, M. P., A. Kumar, and M. Zhong (1997), El Nino, La Nina, and the nonlinearity of their teleconnections, *Journal of Climate*, *10*, 1769-1786.

Horel, J. D., and J. M. Wallace (1981), Planetary-Scale Atmospheric Phenomena Associated with the Southern Oscillation, *Monthly Weather Review*, *109*, 813-829.

Hunter, T., G. Tootle, and T. Piechota (2006), Oceanic-atmospheric variability and western US snowfall, *Geophysical Research Letters*, *33*, -.

Kahya, E., and J. A. Dracup (1994), The Influences of Type-1 El-Nino and La-Nina Events on Streamflows in the Pacific-Southwest of the United-States, *Journal of Climate*, *7*, 965-976.

Kaplan, A., M. A. Cane, Y. Kushnir, A. C. Clement, M. B. Blumenthal, and B. Rajagopalan (1998), Analyses of global sea surface temperature 1856-1991, *Journal of Geophysical Research-Oceans*, *103*, 18567-18589.

Keenlyside, N. S., M. Latif, J. Jungclaus, L. Kornblueh, and E. Roeckner (2008), Advancing decadal-scale climate prediction in the North Atlantic sector, *Nature*, *453*, 84-88.

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

Kwon, H. H., U. Lall, and A. F. Khalil (2007), Stochastic simulation model for nonstationary time series using an autoregressive wavelet decomposition: Applications to rainfall and temperature, *Water Resources Research*, *43*.

Kwon, H. H., U. Lall, and J. Obeysekera (2009), Simulation of daily rainfall scenarios with interannual and multidecadal climate cycles for South Florida, *Stochastic Environmental Research and Risk Assessment*, *23*, 879-896.

Lall, U., B. Rajagopalan, and D. G. Tarboton (1996), A nonparametric wet/dry spell model for resampling daily precipitation, *Water Resources Research*, *32*, 2803-2823.

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

Langbein, W. B. (1949), *Annual runoff in the United States*, 14 p. pp., [Washington].

Latif, M., and T. P. Barnett (1994), Causes of Decadal Climate Variability over the North Pacific and North-America, *Science*, *266*, 634-637.

Lee, T. (2008), Stochastic simulation of hydrologic data based on nonparametric approaches Colorado State University, Fort Collins.

Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges (1994), A Simple Hydrologically Based Model of Land-Surface Water and Energy Fluxes for General-Circulation Models, *Journal of Geophysical Research-Atmospheres*, 99, 14415-14428.

Liang, X., E. F. Wood, and D. P. Lettenmaier (1996), Surface soil moisture parameterization of the VIC-2L model: Evaluation and modification, *Global and Planetary Change*, 13, 195-206.

Mantua, N. J., and S. R. Hare (2002), The Pacific decadal oscillation, *Journal of Oceanography*, 58, 35-44.

Mantua, N. J., S. R. Hare, Y. Zhang, J. M. Wallace, and R. C. Francis (1997), A Pacific interdecadal climate oscillation with impacts on salmon production, *Bulletin of the American Meteorological Society*, 78, 1069-1079.

Massei, N., B. Laignel, E. Rosero, A. Motelay-massei, J. Deloffre, Z. L. Yang, and A. Rossi (2011), A wavelet approach to the short-term to pluri-decennial variability of streamflow in the Mississippi river basin from 1934 to 1998, *International Journal of Climatology*, 31, 31-43.

Maurer, E. P., D. P. Lettenmaier, and N. J. Mantua (2004), Variability and potential sources of predictability of North American runoff, *Water Resources Research*, 40, -.

McCabe, G. J., J. L. Betancourt, S. T. Gray, M. A. Palecki, and H. G. Hidalgo (2008), Associations of multi-decadal sea-surface temperature variability with US drought, *Quaternary International*, 188, 31-40.

McCabe, G. J., J. L. Betancourt, and H. G. Hidalgo (2007), Associations of decadal to multidecadal sea-surface temperature variability with Upper Colorado River flow, *Journal of the American Water Resources Association*, 43, 183-192.

McCabe, G. J., and M. D. Dettinger (1999), Decadal variations in the strength of ENSO teleconnections with precipitation in the western United States, *International Journal of Climatology*, 19, 1399-1410.

McCabe, G. J., and M. D. Dettinger (2002), Primary modes and predictability of year-to-year snowpack variations in the western United States from teleconnections with Pacific Ocean climate, *Journal of Hydrometeorology*, 3, 13-25.

Meehl, G. A., L. Goddard, J. Murphy, R. J. Stouffer, G. Boer, G. Danabasoglu, K. Dixon, M. A. Giorgetta, A. M. Greene, E. Hawkins, G. Hegerl, D. Karoly, N. Keenlyside, M. Kimoto, B. Kirtman, A. Navarra, R. Pulwarty, D. Smith, D. Stammer, and T. Stockdale (2009), Decadal Prediction Can It Be Skillful?, *Bulletin of the American Meteorological Society*, 90, 1467-+.

Mehta, V., G. Meehl, L. Goddard, J. Knight, A. Kumar, M. Latif, T. Lee, A. Rosati, and D. Stammer (2011), Decadal Climate Predictability and Prediction Where Are We?, *Bulletin of the American Meteorological Society*, 92, 637-640.

- Miller, A. J., D. R. Cayan, T. P. Barnett, N. E. Graham, and J. M. Oberhuber (1994), The 1976-77 Climate Shift of the Pacific Ocean, *Oceanography*, 7, 21-26.
- Milly, P. C. D., J. Betancourt, M. Falkenmark, R. M. Hirsch, Z. W. Kundzewicz, D. P. Lettenmaier, and R. J. Stouffer (2008), Climate change - Stationarity is dead: Whither water management?, *Science*, 319, 573-574.
- Milly, P. C. D., K. A. Dunne, and A. V. Vecchia (2005), Global pattern of trends in streamflow and water availability in a changing climate, *Nature*, 438, 347-350.
- Minville, M., F. Brissette, S. Krau, and R. Leconte (2009), Adaptation to Climate Change in the Management of a Canadian Water-Resources System Exploited for Hydropower, *Water Resources Management*, 23, 2965-2986.
- Montgomery, D. C., and G. C. Runger (2003), *Applied statistics and probability for engineers*, 3rd ed., xiv, 706 p. pp., Wiley, New York.
- Nash, L. L., and P. H. Gleick (1991), Sensitivity of Streamflow in the Colorado Basin to Climatic Changes, *Journal of Hydrology*, 125, 221-241.
- Newman, M., G. P. Compo, and M. A. Alexander (2003), ENSO-forced variability of the Pacific decadal oscillation, *Journal of Climate*, 16, 3853-3857.
- Nowak, K., J. Prairie, B. Rajagopalan, and U. Lall (2010), A nonparametric stochastic approach for multisite disaggregation of annual to daily streamflow, *Water Resources Research*, 46, -.
- Ouarda, T., J. Labadie, and D. Fontane (1997), Indexed Sequential Hydrologic Modelin For Hydropower Capacity Estimation, *Journal of the American Water Resources Association*, 33, 1337-1349.
- Piechota, T. C., and J. A. Dracup (1996), Drought and regional hydrologic variation in the United States: Associations with the El Nino Southern Oscillation, *Water Resources Research*, 32, 1359-1373.
- Piechota, T. C., J. A. Dracup, and R. G. Fovell (1997), Western US streamflow and atmospheric circulation patterns during El Nino Southern Oscillation, *Journal of Hydrology*, 201, 249-271.
- Prairie, J., K. Nowak, B. Rajagopalan, U. Lall, and T. Fulp (2008), A stochastic nonparametric approach for streamflow generation combining observational and paleoreconstructed data, *Water Resources Research*, 44.
- Prairie, J., B. Rajagopalan, U. Lall, and T. Fulp (2007), A stochastic nonparametric technique for space-time disaggregation of streamflows, *Water Resources Research*, 43, -.
- Prairie, J. R., B. Rajagopalan, T. J. Fulp, and E. A. Zagona (2006), Modified K-NN model for stochastic streamflow simulation, *Journal of Hydrologic Engineering*, 11, 371-378.

Rajagopalan, B., E. Cook, U. Lall, and B. K. Ray (2000), Spatiotemporal variability of ENSO and SST teleconnections to summer drought over the United States during the twentieth century, *Journal of Climate*, 13, 4244-4255.

Rajagopalan, B., U. Lall, and M. A. Cane (1997), Anomalous ENSO occurrences: An alternate view, *Journal of Climate*, 10, 2351-2357.

Rajagopalan, B., K. Nowak, J. Prairie, M. Hoerling, B. Harding, J. Barsugli, A. Ray, and B. Udall (2009), Water supply risk on the Colorado River: Can management mitigate?, *Water Resources Research*, 45, -.

Ray, A. J., J. J. Barsugli, and K. B. Averyt (2008), Climate Change in Colorado; A Synthesis to Support Water Resources Management and Adaptation, CU-NOAA Western Water Assessment, Boulder, CO.

Redmond, K. T., and R. W. Koch (1991), Surface Climate and Streamflow Variability in the Western United-States and Their Relationship to Large-Scale Circulation Indexes, *Water Resources Research*, 27, 2381-2399.

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.

Regonda, S. K., B. Rajagopalan, M. Clark, and E. Zagona (2006), A multimodel ensemble forecast framework: Application to spring seasonal flows in the Gunnison River Basin, *Water Resources Research*, 42, -.

Revelle, R. R., and P. E. Waggoner (1983), Effects of a carbon dioxide-induced climatic change on water supplies in the western United States, in *Changing Climate*, edited, pp. 419-432, National Academy Press, Washington, DC.

Salas, J. D. (1980), *Applied modeling of hydrologic time series*, xiv, 484 p. pp., Water Resources Publications, Littleton, Colo.

Salas, J. D., and J. T. B. Obeysekera (1982), Arma Model Identification of Hydrologic Time-Series, *Water Resources Research*, 18, 1011-1021.

Santos, E. G., and J. D. Salas (1992), Stepwise Disaggregation Scheme for Synthetic Hydrology, *Journal of Hydraulic Engineering-Asce*, 118, 765-784.

Schubert, S., D. Gutzler, H. L. Wang, A. Dai, T. Delworth, C. Deser, K. Findell, R. Fu, W. Higgins, M. Hoerling, B. Kirtman, R. Koster, A. Kumar, D. Legler, D. Lettenmaier, B. Lyon, V. Magana, K. Mo, S. Nigam, P. Pegion, A. Phillips, R. Pulwarty, D. Rind, A. Ruiz-Barradas, J. Schemm, R. Seager, R. Stewart, M. Suarez, J. Syktus, M. F. Ting, C. Z. Wang, S. Weaver, and N. Zeng (2009), A US CLIVAR Project to Assess and Compare the Responses of Global Climate Models to Drought-Related SST Forcing Patterns: Overview and Results, *Journal of Climate*, 22, 5251-5272.

Seager, R., M. F. Ting, I. Held, Y. Kushnir, J. Lu, G. Vecchi, H. P. Huang, N. Harnik, A. Leetmaa, N. C. Lau, C. H. Li, J. Velez, and N. Naik (2007), Model projections of an imminent transition to a more arid climate in southwestern North America, *Science*, 316, 1181-1184.

Solomon, A., L. Goddard, A. Kumar, J. Carton, C. Deser, I. Fukumori, A. M. Greene, G. Hegerl, B. Kirtman, Y. Kushnir, M. Newman, D. Smith, D. Vimont, T. Delworth, G. A. Meehl, and T. Stockdale (2011), Distinguishing the Roles of Natural and Anthropogenically Forced Decadal Climate Variability Implications for Prediction, *Bulletin of the American Meteorological Society*, 92, 141-+.

Srikanthan, R., and T. A. McMahon (2001), Stochastic generation of annual, monthly and daily climate data: A review, *Hydrology and Earth System Sciences*, 5, 653-670.

Stedinger, J. R., and R. M. Vogel (1984), Disaggregation Procedures for Generating Serially Correlated Flow Vectors, *Water Resources Research*, 20, 47-56.

Stewart, I. T., D. R. Cayan, and M. D. Dettinger (2005), Changes toward earlier streamflow timing across western North America, *Journal of Climate*, 18, 1136-1155.

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

Thomas, B. E. (2007), Climatic Fluctuations and Forecasting of Streamflow in the Lower Colorado River Basin, *JAWRA Journal of the American Water Resources Association*, 43, 1550-1569.

Thomas, H., and M. Fiering (1962), Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation, in *Design of water resource systems*, edited by A. Maasset al., Harvard University Press, Cambridge.

Timilsena, J., T. Piechota, G. Tootle, and A. Singh (2009), Associations of interdecadal/interannual climate variability and long-term colorado river basin streamflow, *Journal of Hydrology*, 365, 289-301.

Tootle, G. A., T. C. Piechota, and A. Singh (2005), Coupled oceanic-atmospheric variability and U.S. streamflow, *Water Resources Research*, 41, -.

Torrence, C., and G. P. Compo (1998), A practical guide to wavelet analysis, *Bulletin of the American Meteorological Society*, 79, 61-78.

Torrence, C., and P. J. Webster (1998), The annual cycle of persistence in the El Nino Southern Oscillation, *Quarterly Journal of the Royal Meteorological Society*, 124, 1985-2004.

Torrence, C., and P. J. Webster (1999), Interdecadal changes in the ENSO-monsoon system, *Journal of Climate*, 12, 2679-2690.

Trenberth, K. E., and J. W. Hurrell (1994), Decadal Atmosphere-Ocean Variations in the Pacific, *Climate Dynamics*, 9, 303-319.

United States Department of the Interior (2001), Record of Decision Colorado River Interim Surplus Guidelines Final Environmental Impact Statement Bureau of Reclamation, Boulder City.

United States Department of the Interior (2005), Natural flow and salt computation methods., Bureau of Reclamation, Salt Lake City, UT.

United States Department of the Interior (2007a), Appendix A, Final Environmental Impact Statement Colorado River Interim Guidelines for Lower Basin Shortages and Coordinated Operations for Lakes Powell and Mead, Bureau of Reclamation, Boulder City, NV, Boulder City.

United States Department of the Interior (2007b), Final Environmental Impact Statement Colorado River Interim Guidelines for Lower Basin Shortages and Coordinated Operations for Lakes Powell and Mead, Bureau of Reclamation, Boulder City.

United States Department of the Interior (2008), Upper Colorado River Basin Consumptive Uses and Losses Report: 2006-2010 (Provisional), Bureau of Reclamation, Salt Lake City, UT.

Valencia, D., and J. C. Schaake (1973), Disaggregation Processes in Stochastic Hydrology, *Water Resources Research*, 9, 580-585.

Veraart, J. A., E. C. van Ierland, S. E. Werners, A. Verhagen, R. S. de Groot, P. J. Kuikman, and P. Kabat (2010), Climate Change Impacts on Water Management and Adaptation Strategies in The Netherlands: Stakeholder and Scientific Expert Judgements, *Journal of Environmental Policy & Planning*, 12, 179-200.

Verberg, K. O. (2010), *The Colorado River Documents 2008.*, U.S. Department of the Interior, Bureau of Reclamation, Lower Colorado Region. .

Wang, B. (1995), Interdecadal Changes in El-Nino Onset in the Last 4 Decades, *Journal of Climate*, 8, 267-285.

White, W. B., and D. R. Cayan (2000), A global El Nino-Southern Oscillation wave in surface temperature and pressure and its interdecadal modulation from 1900 to 1997, *Journal of Geophysical Research-Oceans*, 105, 11223-11242.

Wilks, D. S. (1995), *Statistical methods in the atmospheric sciences : an introduction*, xi, 467 pp., Academic Press, San Diego.

Wood, A. W., L. R. Leung, V. Sridhar, and D. P. Lettenmaier (2004), Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs, *Climatic Change*, 62, 189-216.

Woodhouse, C. A., S. T. Gray, and D. M. Meko (2006), Updated streamflow reconstructions for the Upper Colorado River Basin, *Water Resources Research*, 42.

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