5-2011

Association of oceanic-atmospheric oscillations and hydroclimatic variables in the Colorado River Basin

Ajay Kalra
University of Nevada, Las Vegas, Ajay.Kalra@dri.edu

ASSOCIATION OF OCEANIC-ATMOSPHERIC OSCILLATIONS AND HYDROCLIMATIC VARIABLES IN THE COLORADO RIVER BASIN

by

Ajay Kalra

Bachelor of Engineering
Punjab Engineering College
2002

Master of Science
Utah State University
2005

A dissertation submitted in partial fulfillment
of the requirements for the

Doctor of Philosophy in Engineering
Department of Civil and Environmental Engineering
Howard R. Hughes College of Engineering

Graduate College
University of Nevada, Las Vegas
May 2011

UNLV
UNIVERSITY OF NEVADA LAS VEGAS

THE GRADUATE COLLEGE

We recommend the dissertation prepared under our supervision by

Ajay Kalra

entitled

Association of Oceanic-Atmospheric Oscillations and Hydroclimatic Variables in the Colorado River Basin

be accepted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Engineering
Department of Civil and Environmental Engineering

Sajjad Ahmad, Committee Chair

David E. James, Committee Member

Moses Karakouzian, Committee Member

Kumud Acharya, Committee Member

Ashok Singh, Graduate Faculty Representative

Ronald Smith, Ph. D., Vice President for Research and Graduate Studies
and Dean of the Graduate College

May 2011

ABSTRACT

Association of Oceanic-Atmospheric Oscillations and Hydroclimatic Variables in the Colorado River Basin

by

Ajay Kalra

Dr. Sajjad Ahmad, Examination Committee Chair
Associate Professor, Civil and Environmental Engineering
University of Nevada, Las Vegas

With increasing evidence of climatic variability, there is a need to improve forecast for hydroclimatic variables i.e., precipitation and streamflow preserving their spatial and temporal variability. Climatologists have identified different oceanic-atmospheric oscillations that seem to influence the behavior of these variables and in turn can be used to extend the forecast lead time. In the absence of a good physical understanding of the linkages between oceanic-atmospheric oscillations and hydrological processes, it is difficult to construct a physical model. An attractive alternative to physically based models are the Artificial Intelligence (AI) type models, also referred to as machine learning or data-driven models. These models do not employ traditional forms of equations common in physically based models, but instead have flexible and adaptive model structures that can extract the relationship from the data.

With this motivation this research focuses on increasing the precipitation and streamflow forecast lead times and enhancing the temporal resolution of precipitation within the Colorado River Basin (CRB). An AI-type data-driven model, Support Vector Machine (SVM), was developed incorporating oceanic-atmospheric oscillations to increase the precipitation and streamflow forecast lead times. The temporal resolution of precipitation was improved using the stochastic nonparametric K-Nearest Neighbor

(KNN) approach. The hydrologic data used in the dissertation comprised of climate division precipitation data and naturalized streamflow data for the Colorado River Basin. The interdecadal and interannual Pacific Ocean (Pacific Decadal Oscillation (PDO) and El Niño-Southern Oscillation(ENSO)) and Atlantic Ocean (Atlantic Multidecadal Oscillation (AMO) and North Atlantic Oscillation(NAO)) climatic variability was used in this dissertation.

Initially, the coupled and individual effect of oceanic-atmospheric oscillations in relation to annual precipitation within Colorado River Basin was investigated using the statistical SVM modeling approach. Next, the SVM modeling was used to investigate the coupled and individual effect of oceanic-atmospheric oscillations in relation to annual streamflow volume within Colorado River Basin. Finally, the long-term changes (Trend and Step) in seasonal precipitation within Colorado River Basin were analyzed using nonparametric statistical tests (Mann-Kendall, Spearman's Rho, and Rank Sum). Additionally, the temporal resolution of precipitation was enhanced from annual (water year) to seasonal precipitation (autumn, winter, spring, and summer) using the nonparametric K-Nearest Neighbor disaggregation approach.

The results indicated that annual precipitation predictions for 1-year lead time for the Upper Colorado River Basin can be successfully obtained using a combination of PDO, NAO, and AMO indices, whereas coupling AMO and ENSO results in improved precipitation predictions for the Lower Colorado River Basin. Satisfactory annual streamflow predictions for 3-year lead time for the Upper Colorado River Basin can be obtained using a combination of NAO and ENSO. The seasonal changes in precipitation indicated a decrease in the Upper Basin and increase in the Lower Basin winter

precipitation due to an abrupt step change. KNN disaggregation results indicated satisfactory seasonal precipitation estimates during winter and spring season compared to the autumn and summer season.

The major contributions of this research are threefold. First, this research is the first of its kind that used an AI-type SVM modeling approach to increase precipitation and streamflow forecast lead times using oceanic-atmospheric oscillations for the Colorado River Basin. Second, the results indicated that there is no single climate signal that can be used to explain the hydroclimatology within Colorado River Basin. Coupled response of oceanic-oscillations in relation to precipitation and streamflow is more pronounced in CRB compared to their individual effects. Finally, this is the first study that used a nonparametric KNN disaggregation approach for estimating seasonal precipitation for the Colorado River Basin. Other studies have focused on disaggregating streamflow within CRB from one scale to the other but no other study has attempted to disaggregate precipitation within the Colorado River Basin. Overall, this research improves the understanding of the relationship between climatic variables and hydrology within Colorado River Basin. The long lead time estimates of precipitation and streamflow developed in this research can help water managers in managing the water resources (e.g. reservoir releases, allocation of water contracts etc.) within the Colorado River Basin.

DEDICATION AND ACKNOWLEDGEMENTS

With any work of this scope there are always many people to thank for their advice and assistance. With great pleasure and deep sense of gratitude, I take this opportunity to express my sense of indebtedness to my major advisor, Dr. Sajjad Ahmad, for providing me invaluable guidance, constant encouragement, and cooperation to achieve this dissertation a part of my PhD degree program. He has been always supportive and understanding, which helped me in taking off the excess stress during my research. He provided support whenever I hit a brick wall and guided me through the maze of doctoral research. I have learned so much from him, not only about hydrology and engineering, but about being a professional and a teacher. I can not thank him enough for the opportunity he gave me; I don't know where I would be without it.

I would like to thank my committee Dr. David James, Dr. Moses Karakouzian, Dr. Ashok Singh, and Dr. Kumud Acharya for their advice, assistance, and patience throughout my research. I am grateful to Dr. Thomas Piechota for providing the learning experience that was pivotal in strengthening my foundation in the field. I would like to acknowledge Dr. Haroon Stephen for his assistance with computer programming related questions during the research. Thanks are due for my friends Dr. Navin K. Twarakavi and Dr. Kiran K. Chinnayakanahalli for their prompt and selfless help. I would like to acknowledge the financial support of SPGRA and Post Buckley Schuh and Jernigan Inc. (now Atkins) during the completion of this research.

Where would I be without my family? My family deserves special mention for their inseparable support and prayers. My father, Dr. Aditya M. Kalra in the first place is the person who put the fundamental learning character, showing me the joy of intellectual

pursuit ever since I was a child. My mother, Kusam Kalra, is the one who sincerely raised me with her caring and gentle love. My brother, Dr. Abhay Kalra has always meant more to me than I could ever describe. He is the best brother in every sense of the word, and my best friend. Many thanks goes to my wife Ritu and son Aryamun whose love and persistent confidence in me, has taken the load off my shoulders and have given me much joy to cherish. I love them all so much; I am too fortunate.

Life is incomplete without friends and fun so I would like to thank all my friends, here and elsewhere, for the support and their warm friendship that directly or indirectly made contribution to my research. Last but most importantly I would like to thank God for giving me the opportunity and the strength to complete this work.

I dedicate this dissertation to my father Dr. Aditya M. Kalra who's never ending support and encouragement helped me believe in myself and discover that I can complete anything I put my heart in. I could never thank him enough and can only hope to pay forward his never ending love and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Climate Variability

One of the most important elements essential to the persistence of human life on Earth and the maintenance of terrestrial biological systems is the availability of fresh water. However, rapidly growing human populations and the accompanying anthropogenic effects worldwide are increasing the stresses on currently available water supplies. This water stress is further aggravated by the anticipated effects of a changing climate induced as a function of natural climate variability. The climate variability has direct impacts, both socially and economically, on mankind (Redmond and Koch, 1991). The direct impacts occur through the hydrological cycle, and cause extreme events such as floods and droughts that in turn cause damage greater than other natural disasters in the United States (Wilhite, 1997; Pielke and Downtown, 2000; Lakshmi et al., 2004; Huntington, 2006). Hydroclimatic variables such as precipitation and streamflow are directly influenced by climatic variability. The climate variability causes a shift in the regime of these hydroclimatic variables. Due to this shift, the variables deviate from their normal trend and cause an increase/decrease in their magnitude compared to the normal. Large scale changes in these hydroclimatic variables occurring due to the changing climate have caused several catastrophic floods and drought events globally. These changes have caused large scale destruction both to the nature and mankind. A few examples of some catastrophic flooding events are the 1993 flooding events along the Mississippi, the 1996 autumn floods in New England, the winter floods of 1997 in Pacific Northwest and California, and the Ohio River and Red River valley floods during the

spring of 1997 (Karl and Knight, 1997). Similar to floods, a few notable drought events are the widespread 1988, 1995-1996 droughts in the Upper Midwest and the Ohio Valley, the 1991 drought of California, the 1987 droughts in India, the 1983 droughts in Brazil, the 2003 droughts in Europe, and the severe sustained drought within the Colorado River Basin since 2000. Moreover, the World Health Organization has estimated 150,000 deaths since 2000 due to the changing precipitation. These fluctuations in climate often force water managers to develop plans to mitigate these extremes that require forecasting of hydroclimatic variables.

Additionally, changes in the frequency, intensity, and duration of droughts and precipitation events are being observed recently that can have a significant impact on water supplies both for human societies as well as the terrestrial and marine ecosystems (Karl and Knight, 1997). While the current climate change is bound to cause steady changes in climate profiles (such as temperature, precipitation) across the planet, IPCC (2007) also predicts an increased potential for the prevalence and severity of extreme events such as heat waves, cold waves, storms, floods and droughts. Changes in hydrology due to climate variability are likely to affect all regions in the United States. Of particular interest in the United States are the semi-arid regions of the Southwest including Colorado River Basin (CRB) that are projected to dry further with the suggested changes already assumed to be in progress (Sax et al., 2000). Several key issues have been identified with regards to climate variability in the southwest, these are: (a) increase stress on the availability of fresh water resources (Regonda et al., 2005), (b) increase in temperature, drought, wildfires and an accelerated transformation of the landscape (Mote et al., 2005; Stewart et al., 2005), and (c) increase in extreme flood

events that result in numerous socio-economic impacts (Clark et al., 2001). Given the high water stress in these areas, the future of water availability in the southwest and Colorado River Basin is a cause of serious concern.

Climatic fluctuations and their effects on precipitation and streamflow have received much attention in recent research because of their relation to floods and droughts and their possible relation to global warming. Several indices of atmospheric and oceanic processes have been developed, and these indices have been shown to be related to regular cycles in the magnitude and variability of precipitation and streamflow. The most commonly studied oceanic-atmospheric oscillations in the south western United States, particularly in the Colorado River Basin, are Pacific Decadal Oscillation (PDO), North Atlantic Oscillation (NAO), Atlantic Multidecadal Oscillation (AMO), and El Niño-Southern Oscillations (ENSO). These indices have been linked with hydrological variables for improving the reservoir operations and water resources planning and management (Pulwarty and Melis, 2001). Piechota and Dracup (1996) showed that ENSO events coincide with major dry and wet spells in the Lower Colorado River Basin (LCRB) evidenced by the Palmer Drought Severity Index (PDSI). Kahya and Dracup (1993) related the 1941 and 1983 heavy rainfall events in the Southwest U.S. including CRB with the ENSO phases. Cayan et al. (1998) showed that there has been a change in pattern and amount of precipitation within the Colorado River Basin. Merideth (2000) observed that 20th century was initially wetter than average, followed by a mid-century dry period followed by a wetter period at the end of the century within the Colorado Basin. Clark et al. (2001) showed the influence of ENSO on streamflow patterns over the United States. Kahya and Dracup (1993) studied the relationship between ENSO and

unimpaired streamflow over the conterminous U.S. and indicated a strong ENSO signal in the mid-latitudes of the United States. Tootle et al. (2005) evaluated the streamflow responses to coupled and individual effects of four oceanic-atmospheric modes i.e., PDO, NAO, AMO, and ENSO over the conterminous United States and found a well-established ENSO signal along with PDO, NAO, and AMO influencing the streamflow variability. Dettinger et al. (1998) studied multi-scale streamflow responses to ENSO phenomena for regions in America, Australia, Northern Europe, and parts of Africa and Asia and indicated that the streamflow changes are associated with the weakening ENSO signals for these regions.

1.2 Motivation for Current Research

The documented literature shows that oceanic-atmospheric oscillations do influence hydrologic variables i.e. precipitation and streamflow. However, the complex interaction between hydrologic variable and oceanic oscillation leads to many difficulties in constructing a physically based mathematical model (Lin et al., 2009). An attractive alternative to physically based models are the artificial intelligence (AI) models, also referred to as machine learning or data-driven models. AI models have gained popularity in the hydrologic modeling community because of their ease of use and their success in capturing the hydrologic process compared to physically based modeling approaches. In brief, AI models are used to determine the relationship between inputs and outputs in an empirical format. These models do not employ traditional forms of equations, as in physically based models, but instead they extract the relationship between input and output by employing flexible and adaptive model structures. Furthermore, if a hydrological variable such as precipitation or streamflow is estimated at an aggregate

scale e.g., annual stochastic disaggregation techniques can be used to improve the temporal resolution to a finer scale such as seasonal or monthly depending on the needs of end user.

The stochastic disaggregation techniques help in establishing long range estimates from the historic data and generate synthetic values not seen in the historic records and also preserve the statistical properties such as mean, median, standard deviation, and skewness. With this motivation, this dissertation used statistical approaches to relate the climate variables i.e. oceanic-atmospheric oscillations to precipitation and streamflow to increase forecast lead time and attempted to improve the temporal resolution of precipitation. The ranges of statistical tools vary from simple regression-based approaches to pattern recognition methods such as Support Vector Machines (SVMs) and stochastic approaches such as K-Nearest Neighbor (KNN). At this juncture, it is important to note that all these approaches are statistical methods and cannot be used to establish any concrete physical mechanism. The complexity of the statistical methods is directly related to the order of the parameter space in which the regression is performed. For this dissertation, the Colorado River Basin of the southwest region of the United States is considered in the analysis because it is a hydrologically sensitive area for reasons explained earlier. The four most commonly studied oceanic-atmospheric oscillations i.e. PDO, NAO, AMO, and ENSO were used in the current research to show their interaction with the hydroclimatology of the Colorado River Basin.

1.3 Research Objective and Questions

The main objectives of the current research are to increase the precipitation and streamflow forecast lead times and to enhance the temporal resolution of precipitation.

This will lead to a better understanding of the relationship between ocean atmospheric indices and precipitation and streamflow and will potentially assist water managers in managing the water resource system (e.g. reservoir releases, allocation of water contracts, etc.). In order to achieve the desired objectives, the following research questions and related hypothesis are addressed.

Research Question # 1: What role do oceanic-atmospheric oscillations play in generating precipitation in the Colorado River Basin, and, can precipitation forecast lead time be increased using oceanic-atmospheric oscillations?

Hypothesis # 1: There is a strong linkage between oceanic-atmospheric oscillations and precipitation within Colorado River Basin and the linkage can be used to improve the annual precipitation forecast lead time.

Research Question # 2: What role do oceanic-atmospheric oscillations play in generating streamflow in the Upper Colorado River Basin and can streamflow forecast lead time be increased using oceanic-atmospheric oscillations?

Hypothesis # 2: There is a strong linkage between oceanic-atmospheric oscillations and streamflow within Upper Colorado River Basin and the linkage can be used to improve the annual streamflow forecast lead time.

Research Question # 3: How can temporal precipitation disaggregation be achieved when the data possess higher climate variability?

Hypothesis # 3: The nonparametric trend and step (Mann-Kendall, Spearman's Rho, and Rank Sum) change tests can be used as robust statistical methods to assess the variability in precipitation within the Colorado River Basin. Also, the KNN approach can be used to temporally disaggregate precipitation that exhibits high variability.

1.4 Research Outline

This dissertation follows a manuscript format and starts with this introduction. Chapter 2 is a manuscript titled "*Estimating Annual Precipitation for the Colorado River Basin using Oceanic-Atmospheric Oscillations"* that addresses the first research question. This work uses a state-of-the-art technique called the Support Vector Machines modeling approach to estimate annual precipitation for seventeen climate divisions within Colorado River Basin. The focus of this chapter is to explore for possible relationships between oceanic-atmospheric oscillations and annual precipitation within the Colorado River Basin. Chapter 3 is a manuscript titled *"Using Oceanic-Atmospheric Oscillations for Long Lead Time Streamflow Forecasting"* that addresses the second research question. Statistical data-driven SVM modeling approach is used for three naturalized gages in the UCRB and annual streamflow is estimated with up to 3 years of lead-time. Chapter 4 addresses research question 3 and is a manuscript titled *"Evaluating Changes and Estimating Seasonal Precipitation for Colorado River Basin using Stochastic Nonparametric Disaggregation Technique".* This chapter investigates the long-term changes in seasonal precipitation for 29 climate divisions encompassing Colorado River Basin using three nonparametric statistical tests. In addition to evaluating changes, water year precipitation is disaggregated into four seasonal values using KNN technique preserving their temporal dependences. Chapters 5 summarize the results and conclusions of this dissertation as well as provide recommendations for future research.

References

Cayan, D. R., M. D. Dettinger, H. F. Diaz, and N. E. Graham (1998), Decadal climate variability of precipitation over western North America, *Journal of Climate*, *11* (12), 3148-3166.

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.

Dettinger, M. D., 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*, 3095-4111.

Huntington, T. G. (2006), Evidence for intensification of the global water cycle: Review and synthesis, *Journal of Hydrology*, *319*, 83-95.

IPCC (2007), Climate Change 2007: The Physical Science Basis., *In* Report of the Intergovernmental Panel on Climate Change, edited by Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor, and H.L. Miller, Cambridge UK.

Kahya, E., and J. A. Dracup (1993), U.S. streamflow patterns in relation to the El-Niño/southern oscillation, *Water Resources Research*, *29* (8), 2941-2503.

Karl, T. R., and R. W. Knight (1997), Secular trends of precipitation amount, frequency, and intensity in the United States, *Bulletin of the American Meteorological Society*, *79* (2), 231-241.

Lakshmi, V., T. C. Piechota, U. Narayan, and C. Tang (2004), Soil moisture as an indicator of weather extremes, *Geophysical Research Letters*, *31* (L11401), 1-4.

Lin, G.-F., G.-R. Chen, M.-C. Wu, and Y.-C. Chou (2009), Effective forecasting of hourly typhoon rainfall using support vector machines, *Water Resources Research*, *45*, W08440, doi:10,1029/2009WR007911.

Merideth, R. 2000. A primer on climatic variability and change in the southwest. Udall Center for Studies in Public Policy and the Institute of the Study of Plant Earth, University of Arizona, Tucson, AZ. pp.28.

Mote, P. W., A. F. Hamlet, M. P. Clark, and D. P. Lettenmaier (2005), Declining mountain snowpack in Western North America, *Bulletin of American Meteorological Society*, *86*, 39-49.

Piechota, T. C., and J. A. Dracup (1996), Drought and regional hydrologic variation in the United States: Association with the El. Nino-Southern Oscillation, *Water Resources Research*, *32* (5), 1359-1373.

Pielke, R. A., and M. W. Downtown (2000), Precipitation and damaging floods: Trends in the United States, *Bulletin of the American Meteorological Society*, *13*, 3625-3637.

Pulwarty, R. S., and T. S. Melis (2001), Climate extremes and adaptive management on the Colorado River: lessons from the 1997-1998 ENSO event, *Journal of Environmental Management*, *63*, 207-324.

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 indices, *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.

Sax, J. L., B. H. Thompson, J. D. Leshy, and R. H. Abrams (2000), Legal control of water resources: Cases and Materials, West Group, pp: 956.

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.

Tootle, G. A., T. C. Piechota, and A. Singh (2005), Coupled oceanic-atmospheric variability and U.S. streamflow, *Water Resources Research*, *41*, W12408, doi:10.1029/2005WR004381.

Wilhite, D. A. (1997), State actions to mitigate drought: Lessons learned*, Journal of American Water Resources Association*, *33* (5), 961-968.

CHAPTER 2

ESTIMATING ANNUAL PRECIPITATION FOR THE COLORADO RIVER BASIN USING OCEANIC-ATMOSPHERIC OSCILLATIONS (SUBMITTED TO WATER RESOURCES RESEARCH)

Abstract

Estimating long-lead time precipitation under the stress of increased climatic variability is a challenging task in the field of hydrology. A moving period data-driven precipitation estimation model using Support Vector Machine (SVM) is presented, which uses ocean-atmospheric oscillations to estimate annual precipitation. SVM's are a class of neural networks based on statistical learning theory. Oceanic-atmospheric oscillations, comprising of Pacific Decadal Oscillation (PDO), North Atlantic Oscillation (NAO), Atlantic Multidecadal Oscillation (AMO), and El Niño-Southern Oscillations (ENSO) for a period of 1900–2007 are used to generate annual precipitation estimates with a 1-year lead time. The SVM model is applied to seventeen climate divisions encompassing the Colorado River Basin in the western United States. The results indicate that long-term precipitation predictions for the Upper Colorado River Basin can be successfully obtained using a combination of PDO, NAO, and AMO indices, whereas coupling AMO and ENSO results in improved precipitation predictions for the Lower Colorado River Basin. The results also show that SVM model provides better precipitation estimates compared to the Artificial Neural Network and Multivariate Linear Regression models. The overall results revealed that the annual precipitation in the Colorado River Basin is significantly influenced by oceanic-atmospheric oscillations. The annual precipitation

estimates obtained using SVM modeling technique will be useful for the long-term management of the water resources within the Colorado River Basin.

Kalra, A., and S. Ahmad (2011), Estimating annual precipitation for the Colorado River Basin using oceanic-atmospheric oscillations, *Water Resources Research* (in review).

2.1 Introduction

2.1.1 Background

Climatic fluctuations and increasing water demand in growing regions have captured the attention of scientific communities to study and predict regional and global precipitation variations of interannual and longer time scales (Karl and Knight, 1997; Hidalgo and Dracup, 2003; Nayak et al., 2008 and 2010; Kim et al., 2006 and 2008). Although precipitation is predominantly episodic, changes in climate often cause a shift in the regime of precipitation, which results in such catastrophic events as floods and drought. The impact of these catastrophic events on the agriculture, water resources, and the environment has been studied by many researchers. With growing water demand in many parts of the world, efforts have increased to study and predict regional and global precipitation variation in relation to interannual, decadal, and multidecadal climatic signals, such as oceanic-atmospheric oscillations. Oceanic-atmospheric oscillations, often termed as teleconnections, have been used by researchers across the world to study their relationship with precipitation over land surfaces at interannual and longer time scales; this relationship, in turn, controls such key components as streamflow, soil moisture, and evaporation of the hydrological cycle (Ropelewski and Halpert, 1986; Redmond and Koch, 1991; Diaz and Kiladis, 1992; Rajagopalan and Lall, 1998; McCabe et al., 2004; Piechota and Dracup, 1996; Kahya and Dracup, 1993; Cayan et al., 1998, and1999;

Dracup and Kahya, 1994; Viles and Goudie, 2003; Kim et al., 2006 and 2008; Hidalgo and Dracup, 2003; Diaz et al., 2001; Gutzler et al., 2002; Xu et al., 2004; Hu and Feng, 2001; Kumar and Duffy, 2009). According to Allan et al. (1996), the first use of the word 'teleconnection' in a climate change context was in a paper by Angstrom (1935), whose focus was on climate change in the North Atlantic and Europe. Later, the term 'teleconnections' was used by Namias (1963) to evaluate the connections between weather phenomena in different parts of the world. Increasingly, climatologists have become aware of the modes of climatic variability operating over a range of temporal and spatial scales and their interactions with different hydroclimatic variables, including precipitation.

2.1.2 Oceanic Variability and Precipitation

Recently, much attention has been devoted to how and why precipitation varies in association with El Niño-Southern Oscillation (ENSO) events (Kane, 1999; Ropelewski and Halpert, 1986; Redmond and Koch, 1991; Piechota and Dracup, 1996; Barlow et al., 2002). Ropelewski and Halpert (1986) identified the regions in the continental U.S. where precipitation and temperature are related to the occurrences of ENSO events. Kane (1999) analyzed the association between El Niño and droughts in South Asia and China, and indicated that the long-term droughts in Singapore, Brunei, Indonesia, and East Asia showed good association with El Niño. Barlow et al. (2002) investigated droughts in central and southwest Asia, and observed that regional out-of-phase precipitation is related to the large scale climate variability induced by ENSO events. Wang et al. (2000) observed positive anomalies of precipitation in the central Pacific and eastern Asia during extreme phases of ENSO cycles. Lau and Wu (2001) found that ENSO accounts for 30%

of climate-related variability in Asian summer monsoon rainfall. Hu and Feng (2001) revealed that warmer phases of ENSO are related to increased summer precipitation within the central United States.

In the southwestern United States, particularly in the Colorado River Basin (CRB), various types of climate information have been identified for reservoir operations and water resources planning and management (Pulwarty and Melis, 2001). Piechota and Dracup (1996) showed that ENSO events coincide with major dry and wet spells in the Lower Colorado River Basin (LCRB), as evidenced by the Palmer Drought Severity Index (PDSI). Kahya and Dracup (1993) related that the 1941 and 1983 heavy rainfall events in the southwestern U.S. including CRB with the ENSO phases. Cayan et al. (1998) showed that there has been a change in the pattern and amount of precipitation within the Colorado River Basin. Merideth (2000) observed that during the 20th century, the Colorado Basin initially was wetter than average, followed by a mid-century dry period, followed by a wetter period at the end of the century. Piechota and Dracup (1996) and Cayan et al. (1999) have established that wet conditions during El Niño induce high flows in Colorado River; similarly, dry conditions during La Niña result in low flows in the southwest. Moreover, the correlation between ENSO and southwest U.S. precipitation is now routinely considered each year by the National Oceanic and Atmospheric Agency's (NOAA) Climate Prediction Center as well as other forecast centers to prepare long-lead time winter outlooks (Colorado River Basin Climate, 2005).

While the principal focus of the majority of studies has related ENSO with precipitation, there are other modes of atmospheric oscillations, for instance, Pacific Decadal Oscillations (PDO), North Atlantic Oscillations (NAO), Atlantic Multidecadal

Oscillations (AMO), Arctic Oscillation (AO), and Pacific North American (PNA) index. These are also important indicators of climate variability and have been linked with precipitation both individually and in conjunction with ENSO (Bjerknes, 1966; Giannini et al., 2001; Gershunov and Barnett, 1998; Higgins et al., 2000; Brito-Castillo et al., 2002; McCabe et al., 2004; Wang and Swail, 2001; Dickson et al., 2000). Bjerknes (1966) associated equatorial Pacific rainfall patterns with the 1957-1958 El Niño episodes. Brito-Castillo et al. (2002) correlated PDO with precipitation along northwestern Mexico. Higgins et al. (2000) prepared maps showing PDO-like variability affecting the interannual variation across the southwestern United States. Gershunov and Barnett (1998) found that common positive and negative phases of ENSO-PDO tend to strengthen each other when they are in phase and tend to be weak, unstable, and spatially incoherent when they are in opposite phases. Hidalgo and Dracup (2003) analyzed the effects of ENSO and PDO on streamflow and precipitation in the Upper Colorado River Basin (UCRB), using principal component analysis and wavelets. They found that ENSO impacts warm season precipitation, and that PDO shifts are coincident with changes in the mean of streamflow and precipitation. McCabe and Dettinger (1999) showed a pronounced decadal variation between the Southern Oscillation Index (SOI) and winter precipitation across the North America; they related these decadal variations to PDO. The AMO and NAO indexes exhibit considerable long-term variability, compared to ENSO and PDO. Cancelliere et al. (2007) used a stochastic model of ENSO, NAO, and European Blocking (EB) to monitor and forecast droughts in Sicily, Italy. In evaluating the climate variability, Giannini et al. (2001) observed that during strong periods of southern oscillations, NAO can be in phase and out of phase with ENSO. The responses

to changes on NAO over the northern hemisphere oceans during the winter and spring seasons and along the Arctic Ocean have been linked with the intensity, distribution, and prevalence of storms; wave climate (Wang and Swail, 2001); sea volume and iceberg flux (Dickson et al., 2000). Several studies have established links between NAO phases and precipitation over the Europe and the Mediterranean (Hurrell, 1995; Qian et al., 2000). Wedgbrow et al. (2002) found that positive winter anomalies of the NAO index are associated with a negative PDSI across eastern parts of the British Isles. Enfield et al. (2001) found that AMO has a strong influence on summer rainfall over the continental U.S.; this may adjust the strength of the teleconnection between ENSO and winter precipitation. McCabe et al. (2004) found that more than 52% of the spatial and temporal variation in multidecadal drought frequency over the continental U.S. is attributed to PDO and AMO.

From the past studies, it is evident that oceanic-atmospheric oscillations do influence precipitation. In fact, there have been attempts to use oscillations as predictors to estimate precipitation. However, the complex interaction between precipitation and oceanic oscillation leads to many difficulties in constructing a physically based mathematical model (Lin et al., 2009). An attractive alternative to physically based models are the artificial intelligence (AI) models, also referred to as machine learning or data-driven models. AI models have gained popularity in the hydrologic modeling community because of their ease of use and their success in capturing the hydrologic process compared to physically based modeling approaches. In brief, AI models are used to determine the relationship between inputs and outputs in an empirical format. These models do not employ traditional forms of equations, as in physically based models, but

instead they extract the relationship between input and output by employing flexible and adaptive model structures. By far, the most popular AI type of models are the Artificial Neural Network (ANN) models, which are flexible mathematical structures capable of identifying complex non-linear relationships between input and output (Tokar and Markus, 2000). The concept of ANN was first introduced by McCulloch and Pitts (1943), with the objective being to understand the human brain and emulate its functioning. Since then, ANNs have been used extensively in a variety of physical science applications, including hydrology. Several researchers have used different types of ANN algorithms to forecast precipitation (Raman and Sunilkumar, 1995; Kuligowski and Barros, 1998; Tokar and Johnson, 1999; Hsu et al., 1995 and 1997; Tokar and Markus; 2000; French et al.; 1992; Luk et al., 2000). Raman and Sunilkumar (1995) used ANN to model rainfall time series, and showed the superiority of ANN approach over the Autoregressive Moving Average (ARMA) modeling results. Kuligowski and Barros (1998) were able to make satisfactory short-term (0-6 h) precipitation forecasts for gages in Pennsylvania, using wind direction and antecedent precipitation data as input in an ANN model. A detailed review of the ANN applications in hydrology is available in the ASCE Task Committee report (2000b).

Despite having number of advantages, traditional neural network models have several drawbacks, including the possibility of getting trapped in local minima and also subjectivity in the choice of model architecture (Suykens, 2001). Additionally, the model architecture and weights are determined by a trial and error procedure and an iterative process, both of which are time consuming (Lin et al., 2009). Due to these drawbacks, there is a need for a more sophisticated AI-type data-driven model that is capable of

efficiently representing the multifaceted interaction between oceanic-atmospheric oscillations and precipitation.

2.1.3 Motivation for Current Research

To estimate precipitation, the current study uses a different AI-type neural network, known as the Support Vector Machine (SVM). Support Vector Machine is an approximate implementation of the structural risk minimization (SRM) principle, which helps it to generalize well on unseen data (Vapnik, 1995 and 1998). The ability of SVM to generalize depends more on either the capacity concepts -- the ability of the machine to learn any training set without error than merely on the dimensionality of the space (a measure of complexity and expressive power) or else the number of free parameters of the loss function (Haykin, 2003). In other words, SVM is a statistical tool that approaches the problem of training polynomial function, radial basis function, or neural network regression estimators in a way that is similar to neural networks but at the same time by using a new approach (Liong and Sivapragasam, 2002). SVM originally was developed to solve classification problems; however, it was later used to solve regression problems. In comparing SVMs and ANNs, there are two major differences. First, ANN is constructed based on the principle of empirical risk minimization (ERM), whereas SVMs are based on the SRM principle, which simultaneously minimizes both the empirical risk and the complexity of the model. This results in better generalization ability of the model. Second, the architecture and weights of ANN are determined by trial and error procedure as well as an iterative process that is time consuming. In case of SVMs, the architecture and weight are expressed in terms of a quadratic optimization problem, which can be rapidly solved by a standard programming algorithm. Due to the stronger optimization

algorithm and better generalization ability, SVMs have been successfully applied in hydrology to forecast streamflow (Lin et al., 2010; Kalra and Ahmad, 2009; Asefa et al., 2006; Yu and Liong, 2007; Liong and Sivapragasam, 2002; Dibike et al., 2001), precipitation (Lin et al., 2009; Tripathi et al., 2006), and soil moisture (Ahmad et al., 2010; Gill et al., 2006). A majority of these applications have shown the superiority of SVM over the traditional ANN modeling approach. However, even though SVMs are considered superior to ANNs or other regression methods, they still are statistical data-driven models. A number of factors determine the accuracy of a SVM: (a) the choice of kernel, which is responsible for the data transformation into the high-dimensional space in which SVM performs regression; (b) the magnitudes of Cost (C) and Epsilon, which can result in overfitting or underfitting and affect the accuracy of the model's predictions. To overcome these issues, several algorithms and methods have been developed, and are available in the literature.

Climatic fluctuations and their effects on precipitation and streamflow have received much attention in recent research because of their relationship to floods and droughts as well as their possible relationship to global warming. Several indices of atmospheric and oceanic processes have been developed; these indices have been shown to be related to regular cycles in terms of the magnitude and variability of precipitation. The current study uses the four oceanic-atmospheric oscillations; PDO, NAO, AMO, and ENSO to estimate annual precipitation one year ahead for 17 climate divisions encompassing the Colorado River Basin. These oscillations have been studied by numerous researchers, using different time scales to show their interaction with the western U.S. hydroclimatology, particularly the Colorado River Basin. Moreover, these are the only

four oceanic indices for which reconstructed data has been developed using tree ring information. Additionally, the qualitative relationship between oceanic-atmospheric oscillations and precipitation has been studied extensively (Piechota and Dracup, 1996; Piechota et al., 1997; Cayan et al., 1998; McCabe and Dettinger, 1999; Hidalgo and Dracup, 2003; McCabe et al., 2004; Tootle et al., 2005). However, little attention has been paid to understanding the coupled impact of these oscillations on precipitation for the Colorado River Basin by using an AI-type data-driven model. Although there have been no conclusive answers on the physical processes that link the oceanic-oscillations with regional scale hydrologic processes, the oscillations are still believed to be statistically teleconnected to these hydrologic processes to a significant degree (Pulwarty and Melis, 2001; Kim et al., 2008). The use of climate oscillations information has the potential to improve hydrologic forecasts within a basin (Kim et al., 2008). This possible connection is the basis for performing the current study. Therefore, this research is an attempt to perform a quantitative analysis by incorporating oceanic-atmospheric oscillations in a SVM model to estimate annual precipitation with a 1-year lead time for the Colorado River Basin. Once annual precipitation with a 1 year lead time has been estimated, stochastic techniques, such as K nearest neighbor (KNN), can be used to temporally disaggregate the precipitation into seasonal, monthly, or weekly rainfall depending on the need of the end user (Kalra and Ahmad, 2011). Furthermore, annual precipitation estimates are helpful in analyzing the sediment yield within the basin, which varies as a function of annual precipitation (Wilson, 1973). Precipitation is vital for tree growth. Annual precipitation estimates are used in paleo-climate studies to reconstruct

past hydrology and to evaluate floods and droughts within the CRB (Stockton and Jacoby, 1976; Tarboton, 1994; Gray et al., 2003).

The current study aims at providing annual precipitation totals with a 1-year lead time using oceanic-atmospheric indices in a data-driven model. When estimating annual cumulative precipitation, as done in the current study, the standard definition of lead time i.e., lags between the predictors and the predictand, may not be appropriate. The current study uses the previous year oceanic-oscillation values and estimates annual precipitation total for the next-year and refers it as a 1-year lead time. Thus on January 1st of the current year, the cumulative precipitation total for the entire current year can be known using the proposed modeling approach. In their previous work, the authors have successfully increased the streamflow lead time up to three years for the UCRB by using an SVM-based model that incorporated oceanic-atmospheric oscillations (Kalra and Ahmad, 2009). In addition, the authors have applied the SVM modeling approach to successfully capture the spatial and temporal variability in soil moisture in the Lower Colorado River Basin by using remote sensing data (Ahmad et al., 2010). In this current research, the authors propose a moving period SVM model coupled with oceanic-oscillations to increase the precipitation estimate lead time up to one year. A moving period SVM model is proposed in order to better capture the non-stationary data in the precipitation within Colorado River Basin. There have been attempts to increase the precipitation lead time; however, whatever modest skills a climatologist may have at predicting a 3-9 month lead time arises from ENSO and its effect (Colorado River Basin Climate, 2005; Kim et al., 2006 and 2008). Along with extending the forecast lead time by using four oceanic-atmospheric oscillations; this current research performs a rigorous

sensitivity analysis to determine the coupled and also the individual impact of each oscillation mode with respect to annual precipitation. Additionally, the current research evaluates long-term changes (Trend and Step) in annual precipitation by using nonparametric statistical tests. Two statistical tests, Mann-Kendall and Spearman's Rho, are used to evaluate the trends in the precipitation data; the median-based Rank Sum test is used to evaluate the step change. The parametric linear regression test is used to evaluate correlation between the oceanic-oscillations and precipitation that is persistent over time. In order to show that the moving period SVM approach being proposed better captures the interaction between oceanic-oscillations and precipitation, SVM precipitation estimates are compared with a feed-forward back propagation ANN model as well as a Multivariate Linear Regression (MLR) model. All three models are evaluated using root mean squared error (RMSE), mean absolute error (MAE), RMSE-observation standard deviation ratio (RSR), correlation coefficient (R), and Nash Sutcliffe coefficient of efficiency (NSE). It should be noted that an exhaustive comparison between different methods for estimating precipitation under different circumstances is not our goal. We compared the results of SVM model to ANN and MLR to explore and compare how well the SVM model was able to estimate precipitation within CRB by using oceanic-atmospheric oscillations.

The paper is organized as follows. Section 2.2 presents the description of the study region. The precipitation data for the climate divisions is described in Section 2.3. The SVM modeling approach for estimating precipitation as well as a brief description of the statistical tests used in this study is described in Section 2.4. Section 2.5 summarizes the long-term changes in precipitation. Section 2.6 and 2.7 provides a discussion of SVM

modeling results and highlights the comparison of the SVM precipitation estimates with the ones obtained using ANN and MLR models. Section 2.8 summarizes and concludes the paper.

2.2 Study Region

The Colorado River is the most regulated river in the United States, and is governed by the "Law of River" (Sax et al., 2000). It encompasses seven states, and is a major source of water to the southwestern United States (Figure 1a). It provides industrial and municipal water to nearly 25 million people by means of existing reservoirs, Lake Powell and Lake Mead. It provides agriculture water for nearly 3-million acres, and produces 11.5 billion kilowatt-hours of hydroelectric power. The Colorado River Basin is composed of the Upper Basin (Wyoming, Colorado, Utah, and New Mexico) and the Lower Basin (California, Nevada, and Arizona). The flow demand between the upper and the lower basin is established by the flow at Lee's Ferry (depicted by a triangle in Figure 1a), which acts as the hydrologic divide. The majority of the flow (nearly 90%) is generated in the Upper Basin from the spring-summer snowmelt. Based on the flow contribution, the Upper Basin is subdivided further into eight sub-basins (Figure 1b). As evident in Figure 1b, majority of the flow is generated from the Colorado headwaters, with minimum contribution obtained from the Dolores sub-basin (Prairie and Callejo, 2005). The United States is divided into 344 climate divisions, based on the climatic boundaries. The Colorado River Basin encompasses 29 climate divisions. Out of these 29 climate divisions, 17 divisions have greater than 30% of their area within the Colorado River Basin (Figure 1a). For the purpose of this study, the climate divisions have been sorted according to different states, and have been numbered from 1-17. Table 1 shows

the nomenclature used to identify each climate division within a particular state. Divisions 1-7 and 10-12 are within the Lower Basin, and divisions 8-9 and 13-17 encompass the Upper Basin. An area-weighted method was employed in the Geographic Information System to compute the flow contribution from each of the Upper Basin climate divisions. The climate divisions were merged with the sub-basins of the Upper Basin (Figure 1b) to compute the percentage contribution of each sub-basin within that respective division. This resulted in a maximum flow generated by climate division 8 (~57%), followed by climate division 17 (~14%), and 16 (~ 11%), as shown in Figure 1c. The remaining divisions in the Upper Basin generate less than 6% flow, individually. It should be noted that portions of the San Juan and Dirty Devil sub-basins do not intersect with any of the Upper Basin climate divisions, and therefore were not included in the calculations. This accounts for approximately 5% of the flow. Therefore, the total flow percentage in Figure 1c adds up to less than 100%. Additionally, the annual percentage of streamflow that occurs in the winter and spring is about 70% in the northwestern and central parts, and about 35% in the southern part of LCRB (Thomas, 2007).

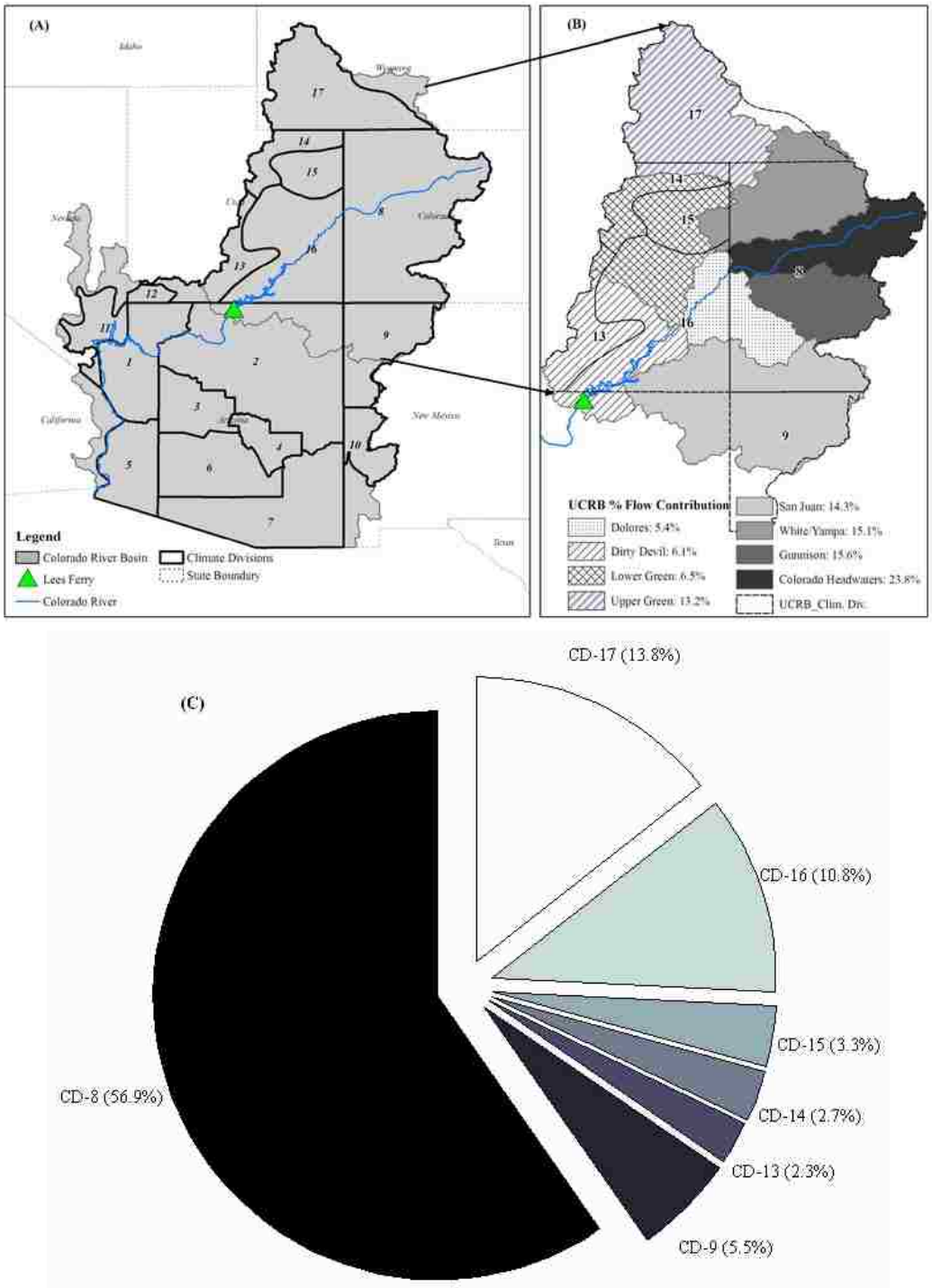


Figure 1: Map showing the location of (a) the Colorado River Basin and the 17 climate divisions, (b) percent flow contribution from UCRB to Colorado River, and (c) flow generated from each climate division in the Upper Basin. (Note: A portion of San Juan and Dirty Devil sub-basins (~5%) are not taken into account, because they do not intersect with UCRB climate divisions). The location of Lee's Ferry is indicated by a triangle.

Table 1: List of climate divisions used in the study.

Climate Division	NAME	State	Region
1	NORTHWEST	AZ	Lower Basin
2	NORTHEAST	AZ	Lower Basin
3	NORTH CENTRAL	AZ	Lower Basin
4	EAST CENTRAL	AZ	Lower Basin
5	SOUTHWEST	AZ	Lower Basin
6	SOUTH CENTRAL	AZ	Lower Basin
7	SOUTHEAST	AZ	Lower Basin
8	CO DRAINAGE BASIN	CO	Upper Basin
9	NORTHWESTERN PLATEAU	NM	Upper Basin
10	SOUTHWESTERN MOUNTAINS	NM	Lower Basin
11	EXTREME SOUTHERN	NV	Lower Basin
12	DIXIE	UT	Lower Basin
13	SOUTH CENTRAL	UT	Upper Basin
14	NORTHERN MOUNTAINS	UT	Upper Basin
15	UINTA BASIN	UT	Upper Basin
16	SOUTHEAST	UT	Upper Basin
17	GREEN AND BEAR DRAINAGE	WY	Upper Basin

2.3 Data

The data sets used to estimate annual precipitation are the oceanic-atmospheric modes of the Pacific and the Atlantic Oceans as well as the precipitation time series. A brief description of the precipitation data is provided below. The oscillation data used in the current analysis is similar to the author's previous work; a detailed description can be found in Kalra and Ahmad (2009).

2.3.1 Precipitation Data

The precipitation data used in this study is the average monthly time series (inch) data for 17 climate divisions, covering a period from 1901-2008. The monthly data set is added to obtain the annual precipitation time series for each of the climate division. This data is obtained from the National Climate Data Center (NCDC) http://www.esrl.noaa.gov/psd/cgi-bin/data/timeseries/timeseries1.pl. The NCDC prepares the data over each climate division by taking an average of temperature and precipitation

from stations, reported by the National Weather Service (NWS) Cooperative Observer Program (COOP), within a division. The dataset is corrected for time bias by adjusting for the variations in average and monthly mean temperatures described in Karl et al. (1986). The count and distribution of the stations within COOP have changed over time and may not be representative of topographical impacts of climate within a division. This may be considered a limitation in the dataset, but the data corresponds well both spatially and temporally to large-scale historic climate anomalies, such as drought (Guttman and Quayle, 1996).

Since its availability, the data has been subject to changes and revisions. The latest significant change occurred in late 1960s. It should be noted that the CRB is composed of highly varied elevations and climate regimes, and it is difficult to integrate all precipitation contributions into a single time series that is representative of the entire-basin precipitation. Therefore, the climate division data used in the analysis is helpful in representing the temporal and spatial variation of precipitation within CRB. The annual spread of the input data for each climate division is shown in vertical box plots in Figure 2. The horizontal line inside the box shows the median value. The box represents the 25^{th} and 75^{th} percentile (interquartile range) values, and the whiskers extend from 5^{th} to 95^{th} percentile values. The dot inside the box shows the historic mean of the input data. The box plots show that the annual precipitation within the CRB exhibits a higher degree of variability, as indicated by wider box plots for the majority of the climate divisions. Estimating this variability by using oceanic-atmospheric oscillations is a challenging task.

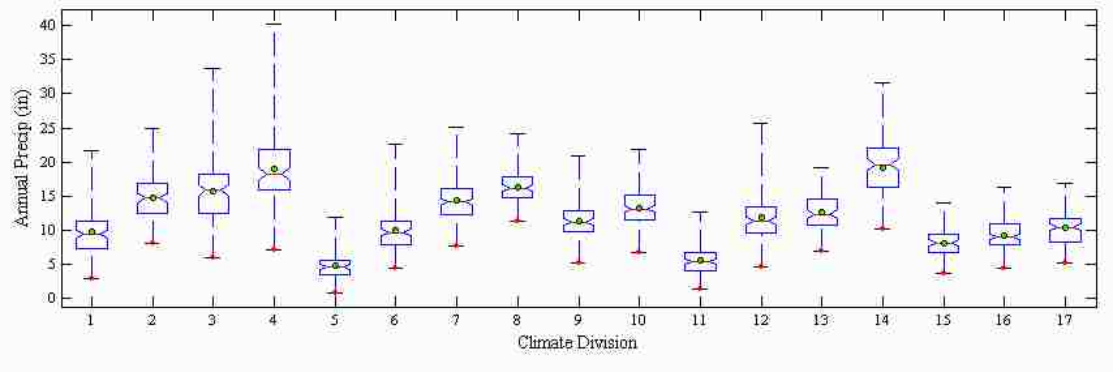


Figure 2: Box plots depicting annual precipitation data from 1901-2008 for 17 climate divisions encompassing the Colorado River Basin. The horizontal line inside the box shows the median value. The box represents the 25th and 75th percentile values (an interquartile range), and the whiskers extend from 5th to 95th percentile values. The circular dot inside the box represents the long–term mean of annual precipitation.

2.4 Methods

In this section, statistical tests are discussed that are used to detect the changes (trend and step) in annual precipitation from 1901-2008 for the 17 climate divisions encompassing the CRB. Also in this section is a description of the SVM modeling framework used to estimate annual precipitation from oceanic-atmospheric oscillations for the 17 climate divisions.

2.4.1 Statistical Tests

The current study used nonparametric tests to evaluate the changes (trend and step) in annual precipitation from 1901-2008 for 17 climate divisions encompassing the Colorado River Basin. The trend in annual precipitation is evaluated using the Mann-Kendall (Mann, 1945; Kendall, 1975) and Spearman's Rho (Lehmann, 1975) tests; the median-based Rank-Sum test is used to evaluate the step change (Chiew and Siriwardena, 2005). In step change studies, the data is divided in two portions in order to evaluate the presence of a step in the hydrologic time series. Based on the literature, 1977 was used as

the year showing the step change (Kerr, 1992; Beamish et al., 1997; Holbrook et al., 1997; Mantua and Hare, 2002; Kalra et al., 2008). The use of nonparametric tests is preferred over parametric tests because nonparametric tests are thought to be more suitable for non-normally distributed data and censored data, which are frequently encountered in hydro-meteorological time series. The serial independence of a time series is still required in nonparametric tests. Moreover, the tests are rank-based procedures and are not influenced by the use of skewed variables (Maidment, 1993). Examples of use of these tests for detecting trends in hydrological and hydro-meteorological time series have been well documented in the available literature (Bunting et al., 1976); Frei and Schar, 2000; Haylock and Nicholls, 2000; Hennessy et al., 1999; Karl and Knight, 1997; Gonzalez-Hidalgo et al., 2001; Luis et al., 2000; Timbal, 2004; Kalra et al., 2008; Miller and Piechota, 2008; Kalra and Ahmad, 2011). Each climate division is evaluated independently to detect any change in the data. Agreement of the Mann-Kendall and Spearman's Rho tests is used to identify the trend in the time series. For step change, the Rank-Sum test has to be significant to show any change in the data. If for a particular climate division both the trend and step changes are significant, the change is referred to be occurring due to the step change in the data.

The tests are evaluated for confidence levels of 95% ($p \leq 0.05$). Trend software by Chiew and Siriwardena (2005) is used for detecting the changes in the annual precipitation. This program is designed to facilitate statistical testing for trend, change, and randomness in hydrological data as well as other time series data. Interested readers are referred to Chiew and Siriwardena (2005) for detailed explanation of these statistical tests. Beside nonparametric tests, the parametric linear regression test is used to compute

the Pearson correlation coefficient between the oscillations modes and lag 1 precipitation. This helps to evaluate the persistence over time among the different oscillations and precipitation for the Colorado River Basin.

2.4.2 SVM Modeling

The following is a brief explanation of the underlying principles of SVM. A more detailed discussion on the subject within the general framework of statistical learning theory can be found in Vapnik (1995 and1998). The basic concept of Support Vector Regression (SVR) is to map, nonlinearly, the original data *x* into a higher dimensional feature space such that

$$f = w\Phi(x_i)+b \tag{1}$$

where $\Phi(x_i)$ is a input feature and both *w* and *b* are coefficients. The coefficients are estimated by minimizing the regularized risk function

$$R(f) = C\frac{1}{N}\sum_{i=1}^{N}(\xi_i + \xi_i^*) + \frac{1}{2}\| w \|^2 \tag{2}$$

$$\text{Subject to } \begin{cases} y_i - \sum_{j=1}^{K}\sum_{i=1}^{L} w_j x_{ji} - b \le \varepsilon + \xi_i \\ \sum_{j=1}^{K}\sum_{i=1}^{L} w_j x_{ji} + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases} \tag{3}$$

where *C* is the cost that determines the trade-off between the complexity of the function *f* and the tolerance for error in the prediction of the function, $y \in \boldsymbol{R}$, taken from a set L of *independent and identically distributed* (*i.i.d.*) observations; *K* is the number of support vectors; and ε is called the Vapnik's insensitive loss function. The Vapnik's ε-insensitive loss functions acts as a threshold, in the sense that errors less than ε are not

considered. Additionally, ξ_i and ξ_i^* are the slack variables that determine the degree to which samples will be penalized with errors larger than ε. Interested readers are referred to Kalra and Ahmad (2009) for the working mechanism as well as an example of the SVM modeling approach.

A typical modeling framework of any AI model consists of the following four steps: 1) preparation of training and testing data suitable for model, 2) training the model using trained data set, 3) testing the trained model using the testing data set, and 4) cross-validating the model using the entire data set. Step 1 is essential in every data-driven modeling application. Steps 2 and 3 are applied collectively in nonparametric stochastic modeling applications, for example, KNN; these two steps have never been merged in SVM approaches, based on information from prior studies and the authors' best knowledge. Step 4 is applied at the end to show the robustness of the model: it tests to see if different training and testing datasets do not yield different results that lead to different conclusions. To make the SVM modeling framework more robust and to improve the efficiency of the SVM approach as a better forecasting tool, the current research proposes a modified SVM modeling framework that combines Steps 2 to 4. The proposed modeling approach is called the Moving Period Support Vector Machine technique and is described in the ensuing section.

2.4.2.1 Moving Period SVM Approach

The Moving Period SVM approach is a special case of the k-fold cross validation technique (Geisser, 1975; Stone, 1974). Each fold is held out, in turn; the model is trained on the remaining dataset and tested on the held-out fold. In this case, k = N, where N is the number of observations in the dataset. The proposed approach is robust, not being

period-specific, and can effectively relate the output to the relative input. The modeling approach for a single climate division is performed as follows.

Let the number of observations be denoted by N. Let i represent the current instance, which is also representative of the testing instance (i = 1). For each instance i, the training set will be comprised of [1,i) U (i,N]. This will train the model on all instances, except for instance i, and will test the model on instance i only. This process is repeated for all instances and stops when i = N+1. A brief description of the step-by-step algorithm is described below:

Step 1. Let [Xi] represent the data matrix comprising of all observations used in the study of length N, where i is the featured instance and varies from 1:N.

Step 2. Partition matrix [Xi] into two sub-matrices, [Ai] and [Bi], such that [Ai] is of length N-i and [Bi] is of length i.

Step 3. Train the SVM model on [Ai] and test the model on [Bi].

Step 4. Repeat steps 1-3 for all the featured instances.

Step 5. Evaluate model performance for all instances (pooled) of [Bi].

Step 6. Apply steps 1-5 for other climate divisions.

The performance of SVM depends on the choice of kernel as being a kernel's parameterization problem. In this study, a radial basis kernel is used in the SVM model; this has performed better when compared with other kernels, such as linear, polynomial, sigmoid or spline, as evident in the past studies (Schölkopf et al., 1997; Smola et al., 1998; Dibike et al., 2001; Yu and Liong, 2007; Khalil et al., 2006; Gill et al., 2006; Asefa et al., 2006; Twarakavi et al., 2006 and 2009).

Based on the modeling approach described, four SVM models are developed using annual oceanic-oscillations PDO, NAO, AMO, and ENSO for time step 't' in order to estimate precipitation at 't+1' (t is in year) for 17 climate divisions encompassing the Colorado River Basin. Each climate division is considered independent, and separate SVM models are developed for each division. The SVM models are developed by using the software package included in the '*R'* software (*http://www.r-project.org/*). Model I is termed as the "base case," and uses all four oceanic modes to estimate annual precipitation. To estimate precipitation, models II and III uses a combination of 3 and 2 oscillation modes, respectively. Model IV uses a single oscillation mode to estimate precipitation. The major reason for developing models II - IV is to evaluate the role of individual and coupled oceanic-oscillations in estimating precipitation within the basin.

The performance of SVM regression depends on the good selection of so-called hyper parameters: cost (C), insensitivity value (ε), and the radial basis kernel width (γ). Previous studies have used the following three procedures to estimate hyper parameters in any SVR formulation: (1) based on prior knowledge and user expertise, (2) using grid based search, and (3) using an analytical approach based on the statistical properties of the training data. In the current study, a grid-based search was adopted to compute the hyper parameters. The objective of a grid search is to obtain hyper parameters by estimating the error prediction for a training data set for every possible combination within a feasible hyper parameter space. The hyper parameters, which result in minimal error, are selected as optimal values. A number of previous studies have used similar approaches (Asefa et al., 2004 and 2005; Gill et al., 2006; Kalra and Ahmad, 2009; Twarakavi et al., 2006 and 2009; Ahmad et al., 2010;)

Furthermore, the performance of the moving period SVM model is compared with the two most widely used hydrologic time series modeling approaches, the Artificial Neural Network (ANN) and the Multivariate Linear Regression (MLR). Many ANN structures have been tested since the 1950s. The most widely employed structure includes the feed-forward neural network with back propagation momentum learning algorithm. The ability to learn is one of the most important characteristics of the ANN model. Despite differences in detailed structures and the number of layers among ANN models, the ANN model performs two basic functions: (1) a learning process that classifies the inputs to hidden layers, and (2) an optimization process that maps the classified inputs to outputs. This current study employed a feed-forward back propagation ANN model, which is composed of one input layer, one hidden layer, and one output layer containing a single node. A tan sigmoid function was employed in the hidden layer neurons, and a linear transfer function was used at the output node. Several combinations of ANN model parameters, such as the number of hidden layers, the type of activation function, and the number of training functions, were tried before selecting an optimal parameter set that gave sufficient convergence. Additionally, this ANN-type model has been used in earlier modeling studies involving different hydroclimatic variables (Raman and Sunilkumar, 1995; Kuligowski and Barros, 1998; Tokar and Johnson, 1999; Hsu et al., 1995, Ahmad and Simonovic, 2005; Melesse et al., 2011). A more detailed description on the theoretical aspects of ANN is available in ASCE Task Committee (2000a). The other type of model developed is the parametric Multivariate Linear Regression model, which consists of oceanic-oscillations as the predictors and annual precipitation as the

predictand. The framework (training and testing data set) used to develop the corresponding ANN model and MLR model is similar to the SVM modeling approach.

2.4.2.2 Model Performance Evaluation

Several techniques have been published in the literature to evaluate the performance of hydrological time series forecasting models. The current study uses five performance measures: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), RMSE-Observation Standard Deviation Ratio (RSR), Correlation Coefficient (R), and Nash Sutcliffe Coefficient of Efficiency (NSE). RMSE, MAE, and R are the most commonly used error index statistics. Lower RMSE and MAE represent better model performance (Singh et al., 2005). R is an index of the degree of linear relationship between observed and predicted values. R varies from -1 to 1; this variable is oversensitive to high extreme values (outliers) and insensitive to additive and proportional difference between model predictions and measured data (Legates and McCabe, 1999). RSR standardizes RMSE by using the standard deviation of observations (Moriasi et al., 2007); it is calculated as the ratio of RMSE to the standard deviation of observed data. RSR can vary between 0 and a large positive value, where 0 indicates zero RMSE or residual variation and, therefore, a perfect model fit. The lower the RSR, the lower the RMSE and better the model performance. RSR is calculated using the following formula:

$$RSR = \frac{RMSE}{STDEV_{Obs}} \tag{4}$$

NSE is used to access the predictive power of hydrological models. It is the normalized statistic that determines the relative magnitude of the residual variance ("noise") compared to the measured data variance, and indicates how well the plot of observed versus predicted data fits the 1:1 line (Moriasi et al., 2007). It is defined as

$$NSE = 1 - \frac{\sum_{i=1}^{n}(y_i - x_i)^2}{\sum_{i=1}^{n}(x_i - \overline{x_i})^2} \quad (5)$$

where y_i is the predicted precipitation value, x_i is the observed value, $\overline{x}$ is the long-term mean of observed data, and N is the total number of years, for example, 108. NSE ranges from -∞ to 1. NSE equal to 1 corresponds to perfect match between the observed and predicted values, NSE equal to 0 shows that model predictions are same as the mean of the observed data, and NSE less than 0 occurs when the mean of the observed data is better predictor than the model.

Table 2, which is abstracted from Moriasi et al. (2007), shows the performance measures for RSR, NSE, and R for predicting precipitation at monthly time step. In general, model performance can be judged satisfactory if RSR ≤ 0.70 and NSE > 0.5 (Moriasi et al., 2007). It should be noted that Moriasi et al. (2007) recommended a value of R > ±0.5 for a model performance to be considered satisfactory; however, the current study uses a stricter criteria, and considers the model performance as satisfactory when R was > 0.7. According to U.S. EPA (2002), the process used to accept, reject, or qualify model results should be based on strict guidelines and documented before evaluating the model. Therefore, based on the above mentioned performance measures, the efficiency of all the three models (SVM, ANN, and MLR) are evaluated in the current research.

Table 2: Recommended performance measures at monthly time steps. Performance measures for RSR and NSE are taken directly from Moriasi et al., 2007; the performance measure for R is modified by the authors.

Performace Rating	RSR	R*	NSE
Very Good	0.0 - 0.50	0.85 - 1.0	0.75 - 1.0
Good	0.51 - 0.60	0.81 - 0.85	0.65 - 0.75
Satisfactory	0.61 - 0.70	0.71 - 0.80	0.51 - 0.65
Unsatisfactory	> 0.70	≤ 0.70	≤ 0.50

2.5 Statistical Properties of Annual Precipitation and its relation with Oscillation Modes

Figure 3 shows the spatial profile of (a) trends and (b) step changes in annual recipitation for the 17 climate divisions encompassing the Colorado River Basin. An increasing trend is noted for climate division 10, and decreasing trends are noted for Divisions 1 and 17 (Figure 3a). Annual precipitation remained relatively unchanged for the remaining climate divisions. Contrary to trend change, an increasing step change in annual precipitation is noted for 9 climate divisions within the Colorado River Basin (Figure 3b). Six of these divisions are in the Lower Basin and the remaining three are in the Upper Basin. Similar to trend change, Division 10 shows an increasing step change and Division 17 shows a decreasing step change in annual precipitation, indicating that the changes for these climate divisions are attributed to an abrupt step change and are not due to a gradual trend change.

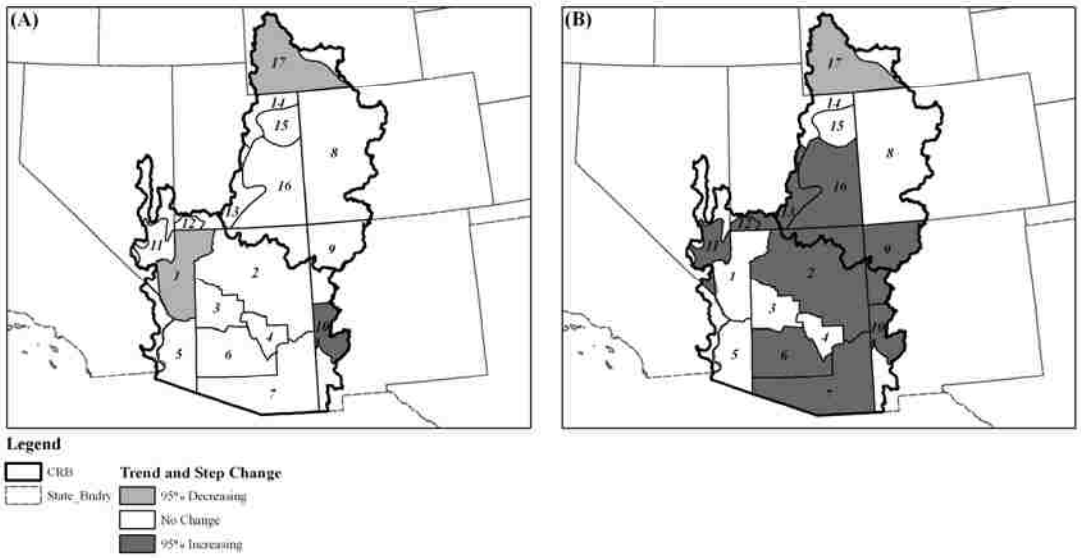


Figure 3: Spatial maps showing the (a) trend change, and (b) step change in annual precipitation for 17 climate divisions encompassing the Colorado River Basin.

The visual identification of step change for selected climate divisions is shown in Figure 4. The plot indicates an apparent jump in the mean value, either an increase or a decrease for the selected climate divisions around the years 1976 to 1977. This jump coincides with the 'regime shift' of mid-1970s in the North Pacific Ocean Sea (Hare and Mantua, 2000). Following the regime shift, a deeper Alteutian low was associated with advection of warmer and moist air to the North American West Coast and colder air over the North Pacific, producing changes in atmospheric thermodynamic and moisture carrying mechanisms (Ingraham et al., 1991). Additionally, these changes manifested as a southward shift of normal tracks, changes in the physical environment, and anomalous precipitation patterns (Trenberth and Hurrell, 1994). This shift has been well documented in the literature (Hollowed and Wooster, 1992, Trenberth and Hurrell, 1994; Mantua et al., 1997; Hare and Mantua, 2000). It should be noted that, universally, there is no common definition of step change. Majority of the times, changes in natural phenomenon over time can be characterized by step change. The shift in hydrologic time series can

also be due to the changes in climate resulting from increased CO_2 emissions or natural variability. Detecting these shifts is of paramount scientific and practical significance for water resource systems that are designed and operated on the assumption of stationarity (Milly et al., 2008). If the assumption is incorrect then the water systems need to be re-designed to adequately serve the purpose.

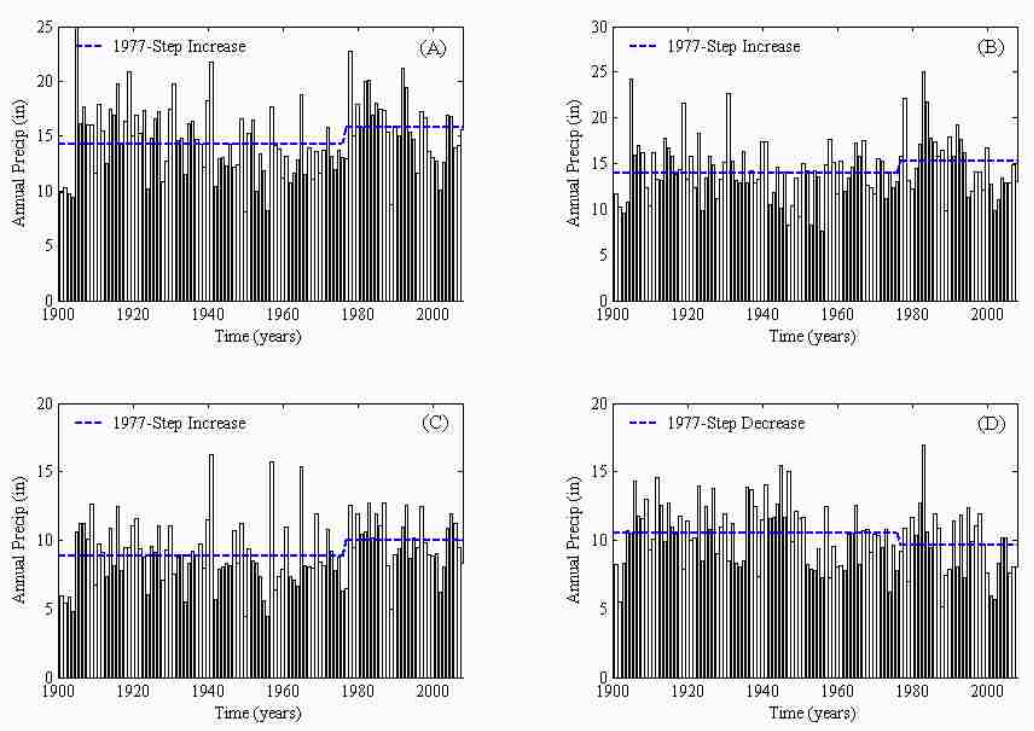


Figure 4: Bar plots depicting step changes (increase/decrease) in annual precipitation for (a) Climate Division 2, b) Climate Division 7, c) Climate Division 16, and (d) Climate Division 17. The dotted line shows the pre- and post-1977 mean values.

The change results, both in trend and step, indicate that the majority of the climate divisions encompassing the northeastern and southern region of the Lower Basin have witnessed increase in annual precipitation due to an abrupt step change, compared to the changes in the Upper Basin precipitation. The mountainous northern region (Climate Division 17) has decreasing precipitation change due to the step change. This region has

precipitation primarily in the form of snow, which is generated by the frontal systems originating in the North Pacific Ocean. Precipitation in this region replenishes the mountain storage and is a source of runoff in the critical spring season, which generates approximately 14% (Figure 1c) of the flow in the Colorado River. Decreasing precipitation changes for this region will affect the runoff generated in the Upper Basin. The trends indicated in the current analysis are in agreement with the findings of some other studies. For example, Regonda et al. (2005) showed statistically significant increases in winter precipitation in the majority of the Lower Basin region, and decline in the monthly snow water equivalent (SWE) over half of the western U.S., including Upper Basin. Coincident with these trends, Knowles et al. (2007) reported a general decline in the SWE for the western United States and the upper Colorado region during a period from 1948 to 2001. In addition to these, Hamlet et al. (2005) and Stewart et al. (2005) have documented decreasing snowpack and earlier runoff in the Colorado region. Trends in annual historic precipitation data compiled by NOAA indicated an average of a 6.1 percent increase in annual precipitation over the conterminous United States since 1900, although there was considerable regional variability (IPCC, 2007). The greatest increases were noted for the eastern, northern, and central U.S. climate regions and also the southern region, which included the Lower Colorado River Basin. Hawaii was the only region to show a decrease in annual precipitation. According to the Intergovernmental Panel on Climate Change (IPCC) report (2007), there will be a global increase in the mean surface air temperatures between 1.1^{o}C to 6.4^{o}C by 2100. This will accelerate the evaporation rate, which may result in more precipitation falling as rain instead of snow in snowmelt driven watersheds, such as the Colorado River Basin. The majority of the

precipitation in the CRB (particularly LCRB) is bi-seasonal with moisture contributions from frontal systems, tropical cyclones, and the summer monsoon. The main source of runoff is the precipitation during the winter and spring season; changes in precipitation can lead to the shift in the timing of the spring-summer peak runoff. This study shows that the increases in annual precipitation within majority of the CRB are due to the step change, and not due to a trend in the data. This is because if a climate division indicates both trend and step change, the final change is attributed to the abrupt step and not to a gradual trend in the data. Therefore, it becomes necessary to distinguish between a gradual trend change and an abrupt step change for climate studies.

Additionally, the pattern of the trend change can be linear and continuous, whereas step changes are nonlinear, can occur abruptly, and may reoccur in the future (McCabe and Wolock, 2002). Furthermore, the changes in annual precipitation are in agreement with the authors' recent work that indicates that the Lower Basin is getting wetter, compared to the Upper Basin, due to an abrupt step change (Kalra and Ahmad, 2011). The authors indicated that the step increases in Lower Basin precipitation can be linked with the climate variability associated with ENSO events. Several studies have documented that ENSO has a pronounced effect on the hydroclimatology in the Lower Basin, compared to the Upper Basin. Furthermore, knowing these shifts in advance can potentially assist water managers in getting prepared for extremes in climate (Regonda et al., 2005). Moreover, finding these changes are significant for management of regional water resources and reservoir operations because precipitation in the form of snow has traditionally played a central role in determining the seasonality of natural runoff within CRB. In many regions of Upper Basin, the precipitation is stored as snow during winter;

this accounts for a significant portion of spring and summer inflow to lower elevation reservoirs (Cayan et al., 1998; Regonda et al., 2005; Stewart et al., 2005). Overall, the change results indicate that the annual precipitation data within CRB exhibits variability; it also brings into question the assumption of stationarity. As mentioned earlier that water resources systems are designed on the assumption of stationarity, and questioning this assumption has implications for the future design of water resource systems. Therefore, it becomes necessary to verify the non-stationarity properties of precipitation in order to better represent the temporal and spatial characteristic of precipitation, such as the seasonal effect and climate variability (Segond et al., 2006). Therefore it becomes necessary to use the climate information when estimating precipitation. Due to this reason, current research used an AI-type, data-driven SVM model coupled with oceanic-oscillations to improve precipitation estimate within the CRB.

Table 3 shows a parametric linear correlation coefficient of lag 1 between the oscillation modes and annual precipitation for 17 climate divisions encompassing the CRB. As reported by previous researchers, the correlation coefficients generally are computed to investigate the link between large-scale oceanic-atmospheric variability and the hydroclimatic variable, in this case annual precipitation, to examine the potential predictors (Grantz et al., 2005; Singhrattna et al., 2005). The significant correlations for each of the combinations are highlighted in bold, and the minimum and maximum correlation values for a particular set are circled.

When correlating PDO index with the 17 climate divisions for a one-year lead time, the correlation values resulted in 11 climate divisions exceeding 90% significance, with Climate Division 10 having the maximum correlation and Climate Division 14 having the

minimum correlation. Five of these divisions are in the Upper Basin, and the remaining six in the Lower Basin. In case of NAO, 8 climate divisions exceeded the 90% significance with Division 17 having the maximum correlation and Division 14 having the minimum correlation. Seven of these divisions are in the Lower Basin, and the remaining division is in the Upper Basin. When relating AMO and ENSO, 12 and 7 climate divisions exceeded the 90% significance, respectively. In case of AMO, Climate Division 7 has the maximum correlation, and Climate Division 13 has the minimum correlation. In case of ENSO, Climate Division 12 has the maximum correlation, and Climate Division 1 has the minimum correlation. PDO shows a stronger association with Upper Basin precipitation, whereas NAO and AMO show a comparable stronger association with Lower Basin precipitation. On the other hand, ENSO shows a pronounced effect in the Upper Basin, as evident by a significant correlation for the majority of the climate divisions (Table 3). Additionally, among the predictors, only the combinations of PDO-ENSO (± 0.50) and NAO-AMO (± 0.41) had significant correlations at 90% significant level. However, to form the basis for a skillful forecast, coefficient values alone cannot be considered. Correlations analysis is a first step to verify the potential predictors for each climate division that show significant relationships with annual precipitation in the CRB. Furthermore, each climate index individually and in combination is analyzed by means of the SVM forecast model, and the results are discussed in the ensuing section.

Table 3: Correlation coefficient between oscillation modes and annual precipitation for 17 climate divisions at p≤0.1 confidence level. The significant correlations are shown in bold. The minimum and maximum correlation values for each subset are circled.

Climate Division	Region	PDO	NAO	AMO	ENSO
1	Lower Basin	0.15	**0.19**	-0.12	-0.06
2	Lower Basin	**0.24**	**0.18**	**-0.33**	-0.16
3	Lower Basin	0.12	**0.18**	**-0.33**	-0.10
4	Lower Basin	0.08	**0.19**	**-0.32**	-0.11
5	Lower Basin	**0.23**	0.16	**-0.24**	-0.13
6	Lower Basin	0.13	**0.17**	**-0.30**	**-0.18**
7	Lower Basin	**0.18**	**0.18**	**-0.34**	-0.14
8	Upper Basin	**0.18**	0.10	**-0.23**	-0.12
9	Upper Basin	**0.28**	0.11	**-0.24**	-0.13
10	Lower Basin	**0.32**	0.08	**-0.19**	-0.12
11	Lower Basin	**0.22**	**0.17**	**-0.21**	**-0.22**
12	Lower Basin	**0.29**	0.05	-0.14	**-0.29**
13	Upper Basin	**0.27**	0.04	-0.06	**-0.24**
14	Upper Basin	0.07	-0.01	-0.07	**-0.18**
15	Upper Basin	**0.20**	0.02	-0.15	**-0.16**
16	Upper Basin	**0.16**	0.05	**-0.19**	-0.15
17	Upper Basin	0.15	**0.22**	**-0.18**	**-0.19**
	90% Significant	***11***	***8***	***12***	***7***

2.6 Results and Discussion

Using the method described earlier, first the SVM Model I (base case) 1-year lead time precipitation estimates for the 17 climate divisions encompassing the CRB are discussed. Next, the coupled and individual effects of oscillations on annual precipitation (Models II-IV) are analyzed. Lastly, the annual precipitation estimates obtained using the SVM models are compared with the ANN and MLR model estimates.

2.6.1 Model I

In Model I, all four oscillations modes are used to estimate annual precipitation with a lead time of one year. This results in one model run for each climate division. Figure 5 shows the scatter plots between the measured and estimated annual precipitation for the 17 climate divisions for the period of record (1901-2008). A good match is obtained between the measured and estimated annual precipitation for majority of the climate

divisions. Correlation values for R exceeding 0.70 are obtained for all the climate divisions, indicating that the model performs reasonably well in capturing the extreme values (high and low). Better predictions are obtained for climate divisions in the Upper Basin compared to the Lower Basin. This is evident by the sample points following the 45^0 bisector line for majority of the Upper Basin Climate Divisions, indicating a good model fit. A point lying far above the bisector line indicates higher estimates, whereas a point far below the line shows lower estimates. For Climate Divisions 1, 5, 6, 11, and 12 of the Lower Basin, the model does fairly well at the low values; however, a few of the high values are scattered away from the bisector, indicating that the model was not able to capture them satisfactorily.

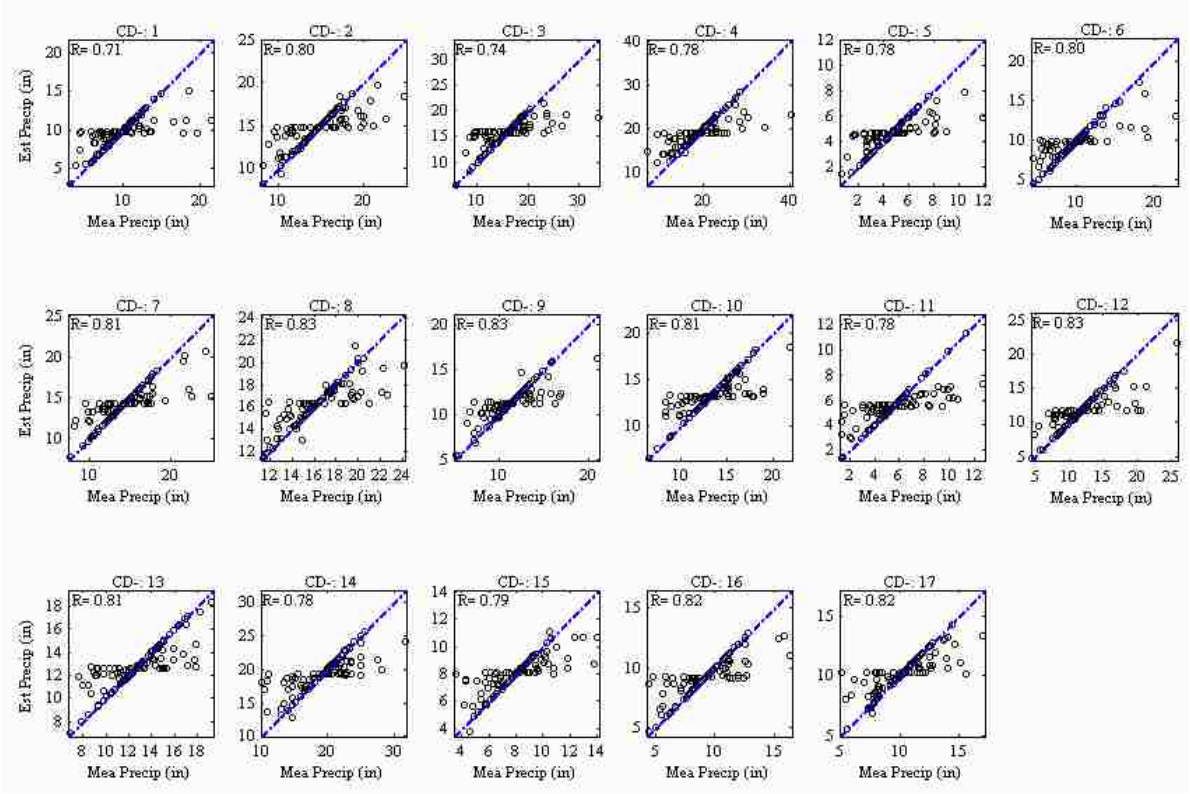


Figure 5: Scatter plot between measured and SVM estimated precipitation for 17 climate divisions for Model I. Dashed line is the 45^0 bisector.

In order to use the model output for tasks ranging from regulation to research, models should be scientifically sound, robust, and defensible (U.S. EPA, 2002). This can be achieved by evaluating the performance of the model in terms of the accuracy of the simulated data compared to the measured data. Based on Table 2, Figure 6 shows the spatial map of the 17 climate divisions, depicting the three performance measures (RSR, R, and NSE). Based on RSR, the model shows good precipitation estimates for the climate divisions in the Upper Basin, and satisfactory estimates for the Lower Basin climate divisions. Based on Figure 1c, climate divisions generating more than 90% of the Upper Basin flow have RSR error statistic in the range of 0.51-0.60 (good estimates per Table 2). All the divisions in the Lower Basin, except Division 1, have RSR in the satisfactory range (0.61-0.70). The RSR measure indicates that the model performs unsatisfactorily in estimating precipitation for climate Division 1 (RSR > 0.7). The correlation statistics R agrees with the results of RSR, indicating that a good correlation (0.81< R < 85) is achieved between the measured and estimated precipitation for the Upper Basin climate divisions and a satisfactory value (R > 0.7) for the climate divisions in the Lower Basin. In case of NSE error statistics, the climate divisions that generate approximately 60% of the Upper Basin flow have good precipitation estimates (NSE > 0.65). The remaining Upper Basin divisions have satisfactory estimates (0.51-0.65), and the Lower Basin's divisions are in the satisfactory range, except for Division 1. Overall, the SVM model is able to provide satisfactory estimates for all the climate divisions, except for Division 1. It should be noted that even in case of Climate Division 1, results are only slightly below or above the satisfactory levels, with RSR, R and NSE values of 0.71, 0.71 and 0.49, respectively. Similar results are noticed for RMSE and MAE error

statistics, as evident in Table 4. Better results – that is, low RMSE and MAE values are obtained for the majority of the climate divisions in the Upper Basin compared to the Lower Basin divisions.

Table 4: Performance measures for SVM Model I output. The RMSE and MAE values are in inches (1 inch =2.54 cms).

Clim Div	*RMSE*	*MAE*	*RSR*	*R*	*NSE*
1	2.38	1.25	0.71	0.71	0.49
2	2.03	1.15	0.63	0.80	0.60
3	3.09	1.68	0.69	0.74	0.52
4	3.50	1.88	0.66	0.78	0.56
5	1.29	0.73	0.66	0.78	0.56
6	2.04	1.03	0.62	0.80	0.61
7	1.98	1.06	0.61	0.81	0.62
8	1.53	0.87	0.57	0.83	0.67
9	1.47	0.81	0.57	0.83	0.67
10	1.64	0.94	0.60	0.81	0.63
11	1.53	0.91	0.66	0.78	0.57
12	2.14	1.10	0.59	0.83	0.65
13	1.57	0.90	0.59	0.81	0.65
14	2.57	1.49	0.64	0.78	0.58
15	1.29	0.75	0.63	0.79	0.59
16	1.38	0.76	0.60	0.82	0.64
17	1.37	0.76	0.59	0.82	0.65

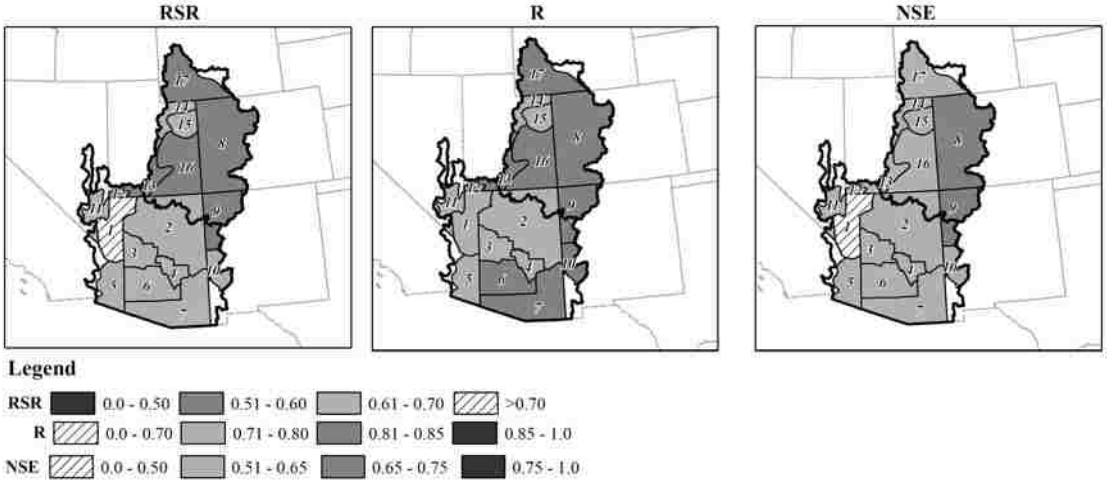


Figure 6: Spatial maps showing the range of performance measures for 17 climate divisions for Model I: (a) RSR, (b) R, and (c) NSE.

Based on Model I results, SVM model performs satisfactorily in capturing the temporal variability in annual precipitation for a one-year lead time by using large-scale climate patterns for the 17 climate divisions within the Colorado River Basin. The scatter plots (Figure 5) and the spatial maps of the performance measures (Figure 6) show that the model produces good precipitation estimates for the Upper Basin divisions, whereas satisfactory estimates (except for Climate Division 1) are obtained for the Lower Basin divisions. This is of importance: an accurate precipitation estimate within the UCRB will be helpful for the water managers in efficient planning and management of water resources within the Colorado River Basin because, on average, 90% of the streamflow is generated in the Upper Basin above Lee's Ferry. The flow to the Lees Ferry gage is controlled by the precipitation in the Upper Basin. This gage is on the hydrologic divide between the Upper and Lower Basins, and is used as the measurement point for the allocation for water between the two basins. Overall, it is justifiable to state that annual precipitation estimates are in the range of satisfactory to good for Model I at 't+1' when using all the four oscillation indices. This shows that all the four oscillations modes have some effect on the hydroclimatology of the Colorado River Basin, and can be satisfactorily captured using an AI-type, data-driven SVM model. The coupled and individual impacts of the oscillations in relation to precipitation can vary within the Upper and Lower Basins. To better understand this variability, Models II-IV are used to analyze the coupled and individual oscillations in relation to annual precipitation within the CRB; these results are described in the ensuing section.

2.6.2 Coupled and Individual Response of Oscillation in relation to Annual Precipitation

To analyze the coupled and individual response of oscillation in relation to annual precipitation, separate SVM models (Models II-IV) are created for each climate division, as described in section 2.6.2.1. Precipitation estimates ranging from satisfactory-to-good are obtained using Model I, and are used as a baseline (or threshold) to compare precipitation estimates obtained from Models II-IV. This will help to better understand the coupled and individual responses of oscillation modes in relation to precipitation within the two regions of CRB. Additionally, understanding the forcing mechanisms that affect hydroclimatic conditions in CRB will provide useful information for water resources planning. Similar to Figure 6, spatial maps depicting the three performance measures RSR, R, and NSE are created for Models II-IV (Figures 7-9).

2.6.2.1 Model II-Model IV

In Model II, oscillations are dropped individually, and the remaining three oscillation modes are used to predict annual precipitation. This results in four model runs for each climate division; the results are shown in Figure 7. By dropping ENSO and using the remaining three oscillations (PDO, NAO, and AMO) as input to the model, an improvement results in RSR (Figure 7a) and R (Figure 7b) error statistics for three climate divisions in the Upper Basin and five climate divisions in the Lower Basin, compared to Model I results. The NSE (Figure 7c) shows improvement in three climate divisions in both the Upper Basin and Lower Basin, compared to Model I results. The divisions showing improvement in the Upper Basin generate approximately 67% of the flow in the Colorado River, based on Figure 1c. The climate divisions showing

improvement in Lower Basin encompass the northern and central portion, generating approximately 70% of the flow (Thomas, 2007) in the Lower Basin. Deterioration in RSR and R error statistics is noted only for Climate Division 13 in the Upper Basin, generating approximately 2% of the flow. Based on the flow contribution, the precipitation estimates are in the range of very good for the Upper Basin and good for the Lower Basin, using a combination of PDO, NAO, and AMO.

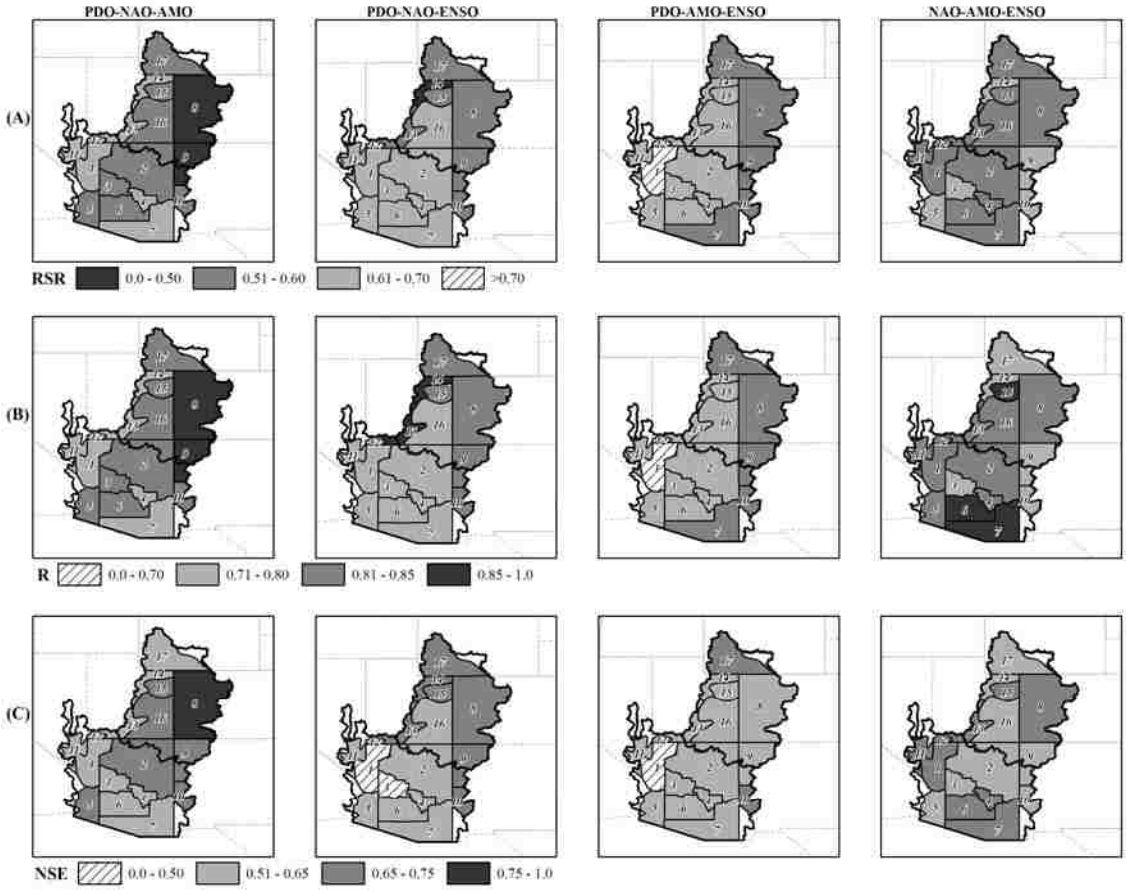


Figure 7: Spatial maps showing the range of performance measures for 17 climate divisions for Model II: (a) RSR, (b) R, and (c) NSE.

By dropping AMO and using the combination of PDO, NAO, and ENSO as inputs in the model, all the three error statistics deteriorate for Climate Division 16, which generates approximately 11% of the flow, compared to Model I results. RSR shows an improvement in Division 14, whereas improved correlation is noted for Divisions 13 and 14 in the Upper Basin. NSE shows an improvement for the divisions of the northwest portion of UCRB, generating approximately 22% of the flow. For the Lower Basin, improvement in precipitation estimates is noted for Divisions 1 and 10 by using RSR statistics; deterioration in the correlation is noted for Divisions 6 and 7, compared to Model I. NSE error statistics show an improvement for Division 10 and deterioration for Division 3, compared to Model I. Overall, the results are comparable with the Model I results for both the Upper Basin and the Lower Basin, and lie in the range of satisfactory to good.

By dropping NAO and using PDO, AMO, and ENSO, deterioration occurs in RSR and R error statistics for 2 climate divisions in the Upper and Lower Basins, compared to Model I results. The NSE statistic indicates deterioration for Climate Divisions 8 and 9, generating approximately 62% of the flow; Division 17, which generates approximately 14% of the flow, shows improvement in the Upper Basin. The NSE error statistics for the Lower Basin are similar to Model I results. Overall, this combination resulted in deterioration of precipitation estimates for the climate division in the Upper Basin; however, no significant improvements are noticed for the climate divisions of the Lower Basin. Dropping PDO and using NAO, AMO, and ENSO as inputs to the model showed similarity with Model I performance measures for majority of the climate divisions in the Upper Basin, and majority of the Lower Basin divisions showed improvement.

Precipitation estimates for the Lower Basin are good, compared to satisfactory Model I estimates. Overall, the best estimates for the Upper Basin are obtained using a combination of PDO, NAO, and AMO. By coupling NAO, AMO, and ENSO results, precipitation estimates improved for the Lower Basin in terms of a one-year lead time. Based on Table 2, very good precipitation estimates are obtained for the Upper Basin, and good estimates are obtained for the Lower Basin climate divisions. These results can be interpreted by indicating that ENSO has a weak association with Upper Basin precipitation, whereas PDO does not have a pronounced effect in the Lower Basin. Also, it should be noted that the worst precipitation estimates for majority of the CRB are obtained when NAO is dropped from the model, indicating that NAO has a strong association with CRB hydroclimatology.

In Model III, oscillations are dropped in pairs, and the remaining two oscillation modes are used to predict annual precipitation. This results in six model runs for each climate division. The three performance measures are shown on a spatial map in Figure 8. Based on the RSR error statistic (Figure 8a), a significant improvement in precipitation estimates is noticed for the Upper Basin is noticed by using a combination of NAO-ENSO, compared to Model I results. Three climate divisions -- 14, 15, and 17 -- show improvements, compared to base case. These divisions account for approximately 21% of the flow in the Colorado River. The estimates are in the range of good to very good. Additionally, coupling NAO and AMO also shows improvement for the Upper Basin, compared to Model I results; good precipitation estimates are obtained based on the RSR error statistic. All other combinations show deterioration in RSR statistics for majority of the climate divisions in the Upper Basin. A stronger correlation (Figure 8b) is obtained

between the measured and estimated precipitation by using a combination of NAO and AMO for climate divisions in the Upper Basin compared to Model I results. Improvement in the R statistic is also noticed by coupling NAO and ENSO for a few climate divisions in the Upper Basin. For the majority of Upper Basin climate divisions, all the other combinations resulted in a lower correlation between the measured and estimated precipitation. The NSE error statistics (Figure 8c) showed improved estimates, using a combination of PDO and NAO as well as NAO and AMO for the Upper Basin divisions, as compared to the Model I results. The improvement in error statistics is noted for three climate divisions that generate approximately 21% of the flow in the Upper Basin.

For a majority of the Upper Basin climate divisions, all other combinations resulted in deterioration in NSE. Model III results agreed with Model II results in indicating that NAO has a stronger presence in the Upper Basin. This was evident by NAO being one of the inputs in the best Model III estimates for the Upper Basin, based on the three error statistics. Overall, none of the combinations resulted in better predictions, compared to Model II best results using PDO-NAO-AMO for the Upper Basin climate divisions. Contrary to this, all the three error statistics were in agreement in indicating that best precipitation estimates for majority of the Lower Basin climate divisions are obtained using a combination of AMO and ENSO, compared to Model I results.

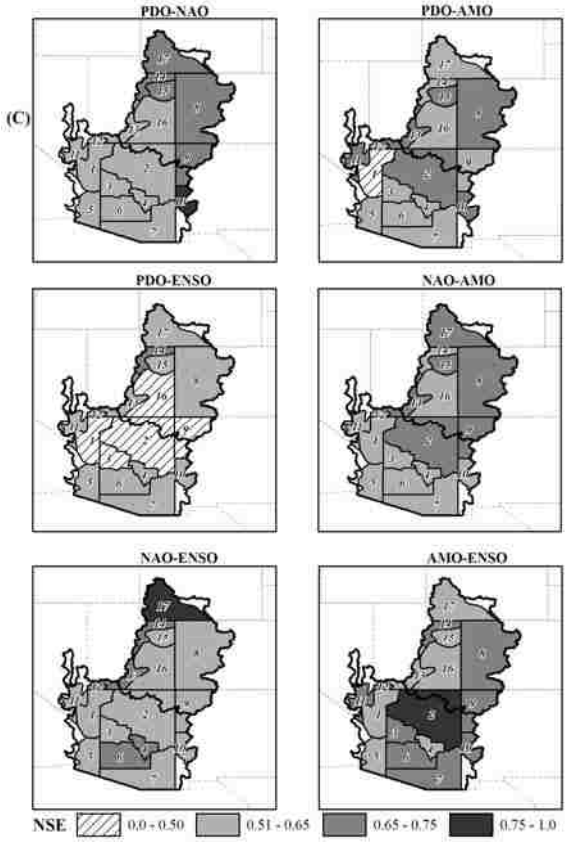


Figure 8: Spatial maps showing the range of performance measures for 17 climate divisions for Model III: (a) RSR, (b) R, and (c) NSE.

All three performance measures showed improvement in the climate divisions covering the north central portion of LCRB. This region generates 70% of the flow in the basin. The estimates for the majority of the climate divisions are in the range of very good compared to satisfactory estimates for Model I and good estimates for the best combination of Model II. This implies that a combination of AMO and ENSO has a stronger influence on precipitation in the LCRB compared to any other combination of indices. Additionally, it was noticed that a combination of PDO and ENSO results in unsatisfactory estimates for the majority of the Lower Basin divisions as compared to the Upper Basin. These climate divisions generate approximately 70% of the flow and cover the northwestern and central portion of the basin. Unsatisfactory precipitation estimates are also obtained for a few climate divisions in Upper Basin by coupling PDO and ENSO, but these divisions account for approximately 15% of the flow in Colorado River. Therefore, combined effect of PDO and ENSO is weakest in the Lower Basin.

In Model IV, each oscillation was used individually to estimate precipitation for each climate division. Figure 9 shows the spatial map representing the three performance measures. It is noticed that none of the oscillations, when used individually, results in improved precipitation estimates for CRB, as compared to Model I results. Although, a satisfactory correlation between measured and estimated precipitation (Figure 9b) is achieved for the majority of the Upper Basin climate divisions, and also for a few Lower Basin divisions using AMO as the sole input in the model. This is expected because AMO had the strongest significant correlation with the majority of the climate divisions within CRB (Table 3). The other two error statistics RSR (Figure 9a) and NSE (Figure 9c) indicate satisfactory predictions for Climate Divisions 17 and 16 by using NAO or

AMO as sole input in the model. Overall, Model IV results indicate that none of the oscillation indices can be used individually to obtain satisfactory annual precipitation estimates for one-year lead time predictions, compared to using coupled oscillations. This is in agreement with the findings from previous studies where researchers showed that a qualitative understanding of the relationship between oscillations and precipitation within the Colorado River Basin can be enhanced by evaluating the coupled response of oscillation indices rather than using an individual oscillation mode (Kim et al., 2006 and 2008; Canon et al., 2007; Hidalgo and Dracup, 2003; McCabe et al., 2007).

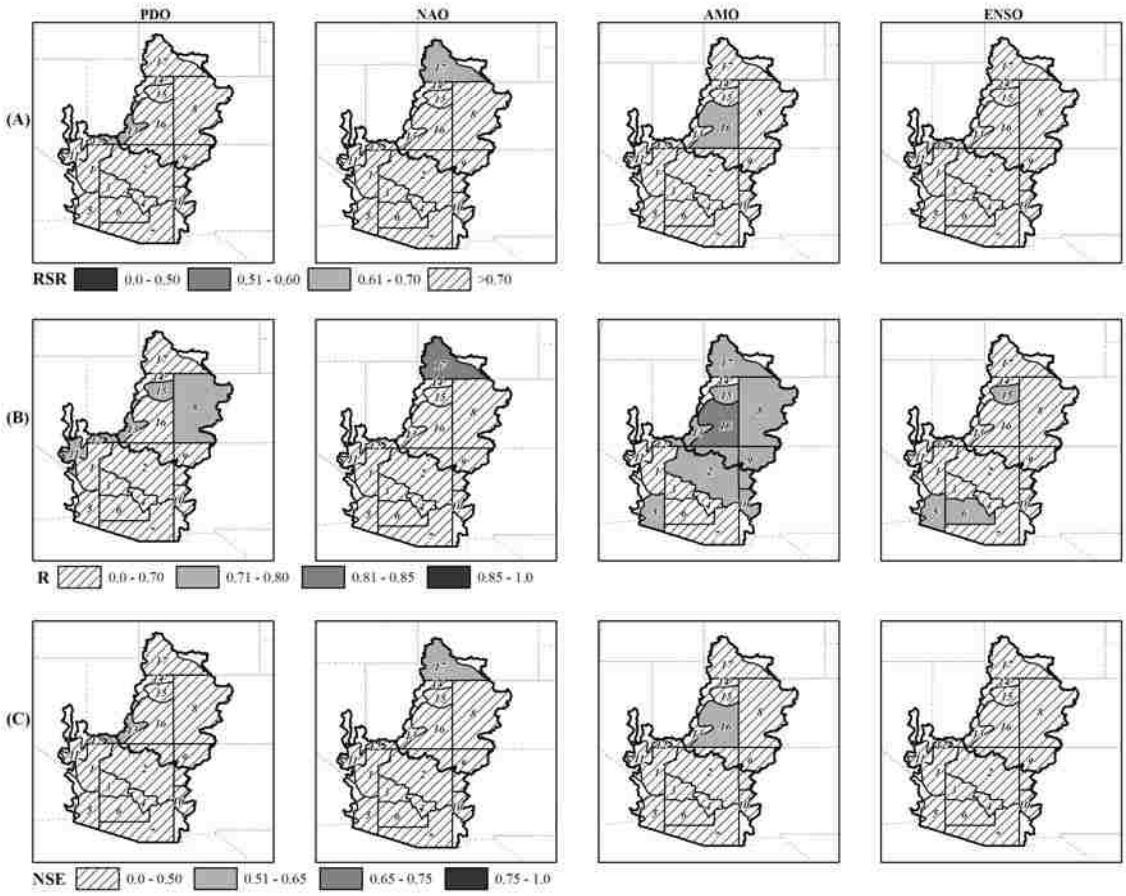


Figure 9: Spatial maps showing the range of performance measures for 17 climate divisions for Model IV: (a) RSR, (b) R, and (c) NSE.

Based on Model II-IV results, it can be inferred that a combination of PDO, NAO, and AMO (Model II) has a stronger association with the annual precipitation for a 1-year lead time in the Upper Colorado River Basin (Figure 7). Interdecadal hydroclimatic variations in the UCRB related to possible PDO influences have been investigated by Hidalgo and Dracup (2003). The study indicated shifts in the mean of variables controlled by UCRB moisture, which are precipitation and streamflow, coincident with PDO shifts suggests a connection between the two processes. Similar to our finding, Hidalgo and Dracup (2003) also concluded that ENSO associations are not always consistent, and may not be linked with the hydrologic fluctuations in UCRB. On the other hand, several studies have established linkages between NAO and precipitation over the Europe and the Mediterranean basin; however, not much emphasis has been given to its association with precipitation in the western United States, particularly to the CRB (Hurrell, 1995; Qian et al., 2000).

Webb et al. (2005) indicated that a combination of AMO and PDO may help to better explain the long-term fluctuations in streamflow within the Colorado River Basin. AMO usually reflects the conditions in the Atlantic Ocean that may affect the climate in North America (Enfield et al., 2001). Colorado River is a snowmelt driven basin with most of the precipitation falling as winter snow in higher elevations of Colorado, Utah, and Wyoming. As a result, the physical mechanism by which the downward trend of the Atlantic Ocean could influence the winter mountain precipitation in UCRB remains a puzzle.

PDO in combination with other oscillations (AO, PNA, and ENSO) besides NAO has been used to evaluate and analyze the hydroclimatology within the CRB by using

different statistical techniques. This work resulted in extending lead times for streamflow and precipitation (Colorado River Basin Climate, 2005; Webb et al., 2005; Kim et al., 2006 and 2008); however, the coupled impact of PDO and NAO for the Upper Basin has not been explored. The findings of the current study are in partial agreement with the other studies that showed that PDO and AMO in combination with other indices can serve as useful predictors, to some extent, for extending lead times of different hydroclimatic variables – in this case, precipitation within the Upper Basin (Hidalgo and Dracup, 2003; Webb et al., 2005; Kim et al., 2006 and 2008; McCabe et al., 2007).

In case of Lower Basin, best estimates are obtained using the combination of AMO and ENSO (Model III). This finding is in agreement with the available literature, indicating that ENSO effects are more pronounced in the Lower Basin than in the Upper Basin (Redmond and Koch, 1991; Webb and Betancourt, 1992; Kahya and Dracup, 1993; Piechota and Dracup, 1996). ENSO has been linked to the occurrence of floods in the Lower Basin (Webb and Betancourt, 1992). Additionally, Thomas (2007) identified that AMO individually and in combination with PDO and ENSO can explain the streamflow variability in the Lower Basin. This is in partial agreement with the current findings. The physical significance of the combined effect of AMO and ENSO on the hydrologic conditions in the Lower Basin is yet to be explored.

Over the time, scientists have learned that no single feature of the system can be used to explain the variability in weather and climate of CRB. It is increasingly evident that various features are interrelated in a complicated fashion (Webb et al., 2005). Their effect on climate depends on the direction and magnitude of their variations, and possibly other factors as yet unidentified can also be linked (Colorado River Basin Climate, 2005; Webb

et al., 2005; McCabe et al., 2007). Though many studies have demonstrated that ENSO and PDO are mainly teleconnnected with precipitation variability in the U.S., it is still necessary to investigate the influence of other climate indices (Ropelewski and Halpert, 1986; Ropelewski and Jones, 1987; Kahya and Dracup, 1993; Dracup and Kahya, 1994; Mantua et al., 1997; Cayan et al., 1999). The results from this current research indicated that there is no single climate system that has a stronger association with the CRB precipitation; however, the coupled impact of these oscillations has a more pronounced effect on the hydroclimatology within the basin. Overall, the moving period SVM modeling approach was successful in establishing the coupled linkage between oscillation indices and annual precipitation. This can be used to increase the lead time for precipitation forecasting within the Colorado River Basin.

2.7 Comparison of SVM with ANN and MLR Models

The SVM application presented in the current research is also compared with ANN and MLR models developed for the 17 climate divisions. The scatter plots between measured and estimated precipitation for ANN and MLR models, using all four oscillations indices, are shown in Figure 10 and Figure 11. The ANN results show that for the majority of the climate divisions, the model estimates the mean of the measured data. The estimates are scattered around the bisector, indicating a poor model fit. Low correlation values are obtained for all climate divisions, indicating the inability of the model to capture the extremes. In general, the model estimates are parallel to the x-axis instead of following the bisector line, showing poor prediction capability.

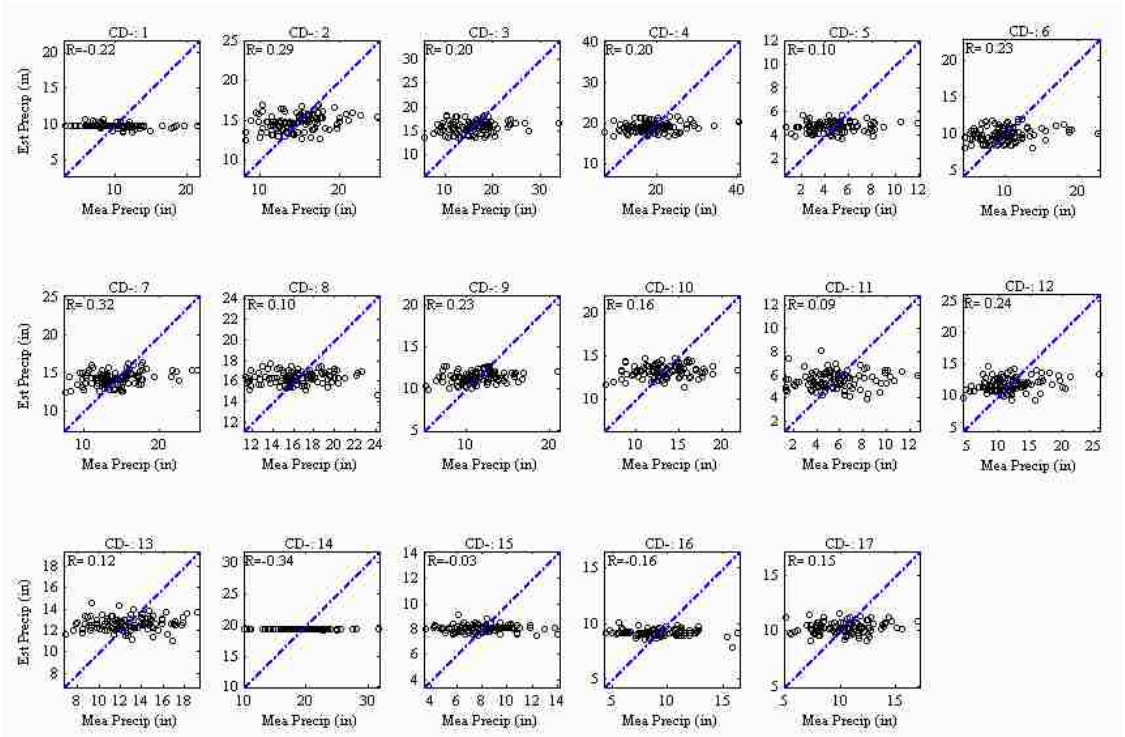


Figure 10: Scatter plot between measured and ANN estimated precipitation for 17 climate divisions for Model I. Dashed line is the 45^0 bisector.

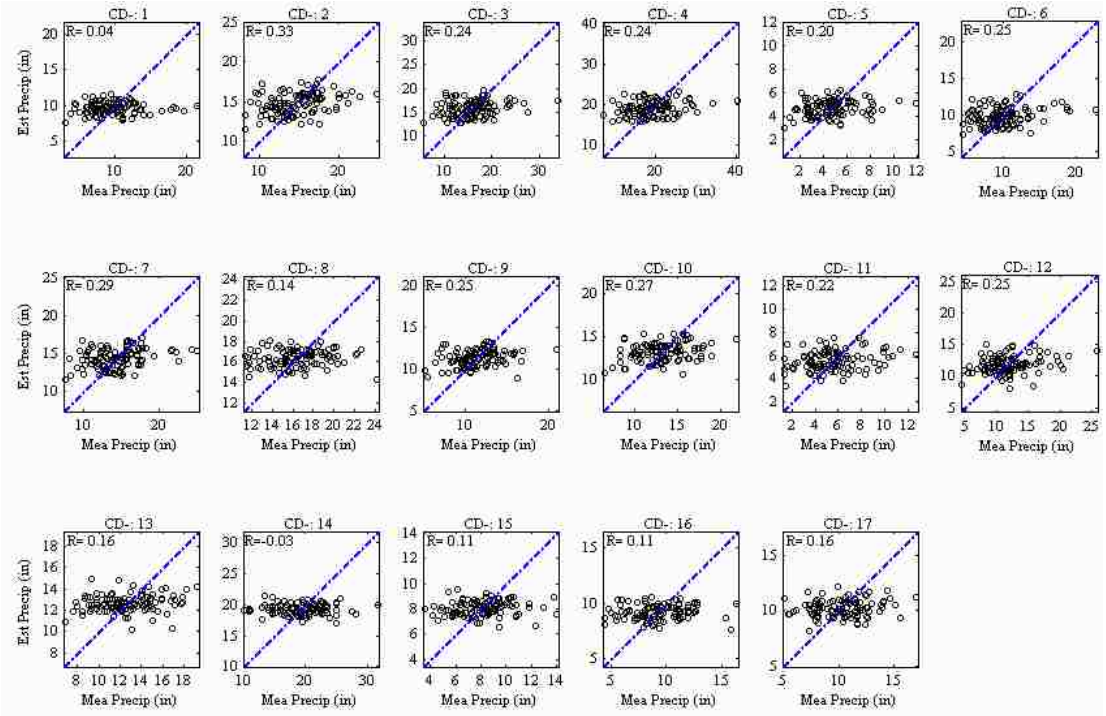


Figure 11: Scatter plot between measured and MLR estimated precipitation for 17 climate divisions for Model I. Dashed line is the 45^0 bisector.

The scatter plots from MLR model (Figure 11) show some improvement, compared to the ANN model estimates. Also, relatively better correlation between measured and estimated precipitation is achieved, compared to ANN model (Figure 10). Additionally, relatively more estimates were around the bisector for the MLR model, as opposed to being parallel to x-axis as for the ANN model.

Compared to the SVM results (Figure 5), both MLR and ANN models performed poorly. Table 5 shows the comparison of different performance measures between the measured and estimated precipitation for ANN and MLR for Model I, using all four oscillations. For all 17 climate divisions, superior performance measures are obtained using the SVM model (Table 4), compared to the ANN and MLR models (Table 5). Based on Table 2, the performance measures obtained from ANN and MLR are in the unsatisfactory range. Although, the SVM model outperforms both the ANN and MLR models, all three models perform comparatively better for the climate divisions within the Upper Basin compared to the Lower Basin divisions. Similar to Model I, Model II-IV (results not shown) were also created, and the results showed a better performance of SVM over both the ANN and MLR models.

There is sufficient evidence from other studies, in different fields of hydrology, that showing the superiority of SVM over the traditional ANN and MLR modeling approaches (Lin et al., 2010 and 2009; Ahmad et al., 2010; Kalra and Ahmad, 2009; Gill et al., 2006; Asefa et al., 2006; Dibike et al., 2001). This is because the SVM model has a better ability to generalize, relating the input to the desired output. In addition, the optimization algorithm used in SVM is more robust than the one used in traditional ANN models.

Table 5: Comparison of performance measures for ANN and MLR outputs for Model I. The RMSE and MAE values are in inches (1 inch =2.54 cms).

	ANN					*MLR*				
Clim Div	**RMSE**	**MAE**	**RSR**	**R**	**NSE**	**RMSE**	**MAE**	**RSR**	**R**	**NSE**
1	3.40	2.62	1.02	-0.22	-0.05	3.39	2.58	1.02	0.04	-0.04
2	3.09	2.38	0.96	0.29	0.08	3.05	2.35	0.94	0.33	0.10
3	4.39	3.28	0.98	0.20	0.03	4.36	3.26	0.97	0.24	0.04
4	5.20	3.95	0.97	0.20	0.04	5.19	3.97	0.97	0.24	0.04
5	1.95	1.50	1.00	0.10	-0.01	1.93	1.50	0.99	0.20	0.02
6	3.21	2.38	0.97	0.23	0.05	3.21	2.42	0.97	0.25	0.05
7	3.06	2.31	0.94	0.32	0.10	3.10	2.33	0.96	0.29	0.07
8	2.67	2.11	1.00	0.10	0.00	2.68	2.11	1.00	0.14	-0.01
9	2.49	1.89	0.97	0.23	0.05	2.49	1.87	0.97	0.25	0.05
10	2.69	2.13	0.99	0.16	0.01	2.62	2.09	0.96	0.27	0.06
11	2.37	1.84	1.01	0.09	-0.04	2.29	1.80	0.98	0.22	0.03
12	3.52	2.73	0.97	0.24	0.05	3.53	2.77	0.97	0.25	0.05
13	2.66	2.18	1.00	0.12	0.00	2.66	2.18	1.00	0.16	0.00
14	3.97	3.14	1.00	-0.34	0.00	4.08	3.27	1.02	-0.03	-0.05
15	2.06	1.60	1.01	-0.03	-0.03	2.05	1.59	1.00	0.11	-0.02
16	2.37	1.81	1.02	-0.16	-0.05	2.33	1.78	1.01	0.11	-0.02
17	2.28	1.86	0.99	0.15	0.02	2.29	1.84	0.99	0.16	0.00

In case of SVM, the determination of model architecture is expressed in terms of quadratic optimization, and the weights are guaranteed to be unique and globally optimal. Contrary to this, different optimal weights could be obtained even when ANN is trained with the same training data; this is because the weights are determined by iterative process. Moreover, different initial weights in ANN lead to different forecasting performances and do not yield a global solution. MLR models are based on the assumption of normality, and can be used efficiently to relate simple processes. In case of hydrological processes where the data does not follow the usual normal distribution, MLR models fail to capture the variability. Even so, the current study MLR model was able to capture the low values evident in scatter plots; however, the values were in the unsatisfactory range and cannot be used for water resources planning and management.

The feed-forward back-propagation algorithm used for ANN in the current analysis is simple and widely used. There are other ANN architectures and activation functions (see Dibike et al., 1999) that may be able to better capture the relationship between precipitation and ocean-atmospheric indices. An exhaustive comparison of methods was not the focus of our work.

2.8 Conclusion

In this study we evaluated the long-term trend and step changes in precipitation in the Colorado River Basin. We also explored the association between individual and coupled oceanic-atmospheric indices and precipitation in the Colorado River Basin. We used an AI-type model to capture the relationship between oceanic-atmospheric indices and precipitation, and used this model to extend the lead time for precipitation estimation up to one year. For this purpose, a moving period data-driven model, using Support Vector Machine and incorporating oceanic-atmospheric oscillations, was constructed for the 17 climate divisions encompassing the Colorado River Basin. The oceanic-atmospheric oscillations used in this study were PDO, NAO, AMO, and ENSO.

Annual precipitation within CRB is variable both at the temporal and spatial scales. It is difficult to construct a single precipitation time series that is representative of the entire basin. For this reason, monthly time series data for the climate divisions were used that extend more than a century in record and spatially cover the entire basin. Currently, Climate Prediction Center issues 3-month forecast for lead times of 0.5 to 12.5 months with modest skill for 3-9 month lead time based on ENSO and its indices. Forecasts are termed ‘skillful’ if they show improvement over the long-term averages of the precipitation record used in the analysis. In general, they have no skill for summer

precipitation during ENSO years and no skill for winter precipitation during non-ENSO years (Colorado River Basin Climate, 2005; Regonda et al., 2005). During ENSO years, the precipitation forecast is higher in the southern part (LCRB) of the basin and has no skill in the headwaters that generate majority of the runoff in the Colorado River (Redmond and Koch, 1991). Therefore, obtaining accurate estimates of precipitation within CRB is a formidable challenge. However, advances along several scientific fronts have opened doors for statistical forecast possibilities. In an attempt to address this challenge, we evaluated the link between individual and coupled oceanic-atmospheric indices and temporal variability in precipitation, and developed a data-driven model to estimate annual precipitation with a lead time of one year.

Four SVM models incorporating individual and coupled oceanic-atmospheric oscillations were developed to estimate precipitation one year in advance. The model outputs were evaluated by using performance measure guidelines presented in Moriasi et al. (2007). The performance measures classified the estimates into four categories that ranged from unsatisfactory to very good. Model estimates for each climate divisions were considered acceptable if they met the requirements of the satisfactory category (NSE $\geq$ 0.5 and RSR $\leq$ 0.70; R is > 0.70) or above. Moriasi et al. (2007) indicated that lower performance measures will suffice if the model is used for basic exploratory research. This current research followed the stricter guidelines to show the robustness of the proposed modeling approach.

Model I indicated that good precipitation estimates for the majority of the Upper Basin climate divisions and satisfactory estimates for the Lower Basin divisions can be obtained using all four oscillation modes. By dropping each oscillation at a time in Model

II, a stronger association of coupled PDO, NAO, and AMO was indicated with the Upper Basin precipitation. This combination resulted in very good precipitation estimates for climate divisions that contribute approximately 67% (two-thirds) of the flow in Colorado River. The best estimates for the Lower Basin were obtained by dropping PDO and using NAO, AMO, and ENSO as input to the model. The estimates were in the range of "good" for majority of the Lower Basin divisions. This model resulted in improved estimates, as compared to Model I results. Also, this analysis indicated that the worst predictions are obtained when NAO is dropped, indicating that NAO has a pronounced effect on hydroclimatology within CRB.

Model III results did not provide any definitive conclusion for the Upper Basin, as different combinations resulted in improved estimates based on the performance measures. Even so, all the combinations showed improvement over Model I results, but were inferior to best results of Model II. A stronger coupled effect of AMO and ENSO was noticed for the majority of the Lower Basin divisions. The estimates for the majority of the Lower Basin divisions were in the range of "very good;" these divisions account for generating approximately 70% of flow in the Lower Basin. Additionally, it was noticed that combination of PDO and ENSO resulted in unsatisfactory estimates for the majority of the Lower Basin divisions, compared to estimates for the Upper Basin. Model IV results indicate that by using individual oscillation modes, the model can identify the strength of the relationship between individual oscillations and precipitation; however, these results cannot be used to extend lead time for precipitation forecasts up to one year. Overall, the best annual precipitation estimates for the Upper Basin were obtained using Model II, where a combination of three oscillations (PDO-NAO-AMO) is used. The best

Lower Basin estimates were obtained using Model III, where a combination of two oscillations (AMO-ENSO) is used as an input in the model.

The precipitation estimates from the SVM model were compared with ANN and MLR modeling approaches. The results showed that SVM approach performs better in capturing the interaction of oscillation indices and precipitation within CRB for a one-year lead time. Along with estimating precipitation, the long-term changes in annual precipitation within CRB were also evaluated. The result showed that the majority of the Lower Basin is getting wetter, compared to the Upper Basin. The changes in the Lower Basin are attributed to an abrupt step change and not to the gradual trend in the data. The step change coincides with the climate regime "shift" of mid 1970s in the North Pacific Ocean. Therefore it is challenging to obtain successful estimates for a non-stationary time series.

The application of SVM as a simple and efficient statistical tool to estimate precipitation has been shown by means of its implementation in the Colorado River Basin. Hydroclimatic variability within the CRB, using PDO and ENSO as climate indices, has been reported extensively (Kahya and Dracup, 1993; Piechota and Dracup, 1996; Mantua et al., 1997; Cayan et al., 1999; Pulwarty and Melis, 2001). Attempts also have been made to study the relationship between AMO and to the hydrologic conditions within the Colorado River Basin. Contrary to this, NAO primarily has been studied in relation to changes in mean sea level pressures (SLP) over the Arctic Ocean (Walsh et al., 1996), to trends in surface wave heights over the North Atlantic (Kushnir et al., 1997), to predicting storm activity and shifts in storm tracks in the Atlantic Ocean (Hurrell, 1995). Lesser attention has been given to the changes in precipitation in relation to NAO within

the Colorado River Basin. Our results indicate that NAO, coupled with other indices, can improve the precipitation estimates in the UCRB.

The major contributions of this research are as follows. First, there is no single oscillation, for example, PDO, NAO, AMO, or ENSO that can be used to adequately explain the climate variability within CRB. It is evident that various oscillations are interrelated and can be used in combination to improve annual precipitation forecasting with a one-year lead time within CRB. Second, NAO coupled with other indices can improve the precipitation estimates in the UCRB. This has been evident in the author's prior work on streamflow (Kalra and Ahmad, 2009); therefore, the effect of NAO on the UCRB hydroclimatology requires in-depth analysis, using more sophisticated techniques. Third, the moving period SVM approach is robust for the entire time series, which confirms that the model is stable. Conclusions drawn from this analysis are not specific to a particular period, and are based on stricter performance measures. Fourth, for the period of record, the majority of the Lower Basin is getting wetter, compared to the Upper Basin. These changes result from an abrupt step instead of a gradual trend in the data.

Colorado River Basin is a snowmelt driven watershed with snowpack making up 63% of the annual precipitation within the Upper Basin and 39% of the annual precipitation within the Lower Basin (Serreze et al., 1999). The current study does not differentiate between precipitation as rainfall and snowfall. Furthermore, the trends detected in the current study are dependent upon the period considered for the analysis. Additionally, there are some variations that are still unexplained, and which cannot be addressed using the statistical approach. However, the precipitation estimates for a one-year lead time,

obtained by using the moving period SVM approach proposed in this study, potentially can be useful for water supply and reservoir operations within CRB.

The results from the current research increase our understanding of the association between different oceanic-atmospheric indices and precipitation in the Upper and Lower Colorado River Basin. The results presented in this study are statistical in nature, and the physical mechanisms that drive these relationships are not fully understood at this time. The SVM modeling approach used in the current research was successful in estimating the annual precipitation within the basin. Using the proposed approach, cumulative precipitation totals for the current year can be made available as early as January 1st of that year. The cumulative value can be temporally disaggregated into finer values using stochastic techniques based on the need of end user. Overall, the moving period SVM model used in this study provides very good precipitation estimates that have the potential for better water management within the basin.

Acknowledgments

This work was supported by the National Oceanic and Atmospheric Administration (NOAA-SARP) Award # NA07OAR4310324 and National Science Foundation (NSF) Award # CMMI 0846952.

References

Ahmad, S., A. Kalra, and H. Stephen (2010), Estimating soil moisture using remote sensing data: A machine learning approach, *Advances in Water Resources*, *33* (1), 69-80.

Ahmad, S., and S. P. Simonovic (2005), An artificial neural network model for generating hydrograph from hydro-meteorological parameters, *Journal of Hydrology*, *315* (1-4), 236-251.

Allan, R. J., J. Lindesay, and D. Parker. 1996. El Niño-Southern Oscillation and climate variability. CSIRO Publishing: Collingwood, Victoria.

Angstrom, A. (1935), Teleconnections of climate changes in present time, *Geography Annals*, *17*, 242-258.

ASCE Task Committee (2000a), Artificial neural networks in hydrology I: preliminary concepts, *Journal of Hydrologic Engineering*, *5* (2), 115-123.

ASCE Task Committee (2000b), Artificial neural networks in hydrology II: hydrologic applications, *Journal of Hydrologic Engineering*, *5* (2), 124-137.

Asefa, T., M. Kemblowski, M. McKee, and A. Khalil (2006), Multi-time scale stream flow predictions: The support vector machines approach, *Journal of Hydrology*, *318*, 7-16.

Asefa, T., M. Kemblowski, G. Urroz, and M. McKee (2005), Support vector machines (SVMs) for monitoring network design, *Ground Water*, *43* (3), 413-422.

Asefa, T., M. W. Kemblowski, G. Urroz, M. McKee, and A. Khalil (2004), Support vectors-based groundwater head observation networks design, *Water Resources Research*, *40*, W11509.

Barlow, M., H. Cullen, and B. Lyon (2002), Drought in central and southwest Asia: La Niña, the warm pool, and Indian Ocean precipitation, *Journal of Climate*, *15* (7), 697-700.

Beamish, R. J., C. M. Neville, and A. J. Cass (1997), Production of Fraser River sockeye salmon (Oncorhynchus nerka) in relation to decadal-scale changes in the climate and the ocean, *Canadian Journal of Fisheries and Aquatic Sciences*, *54*, 543-554.

Bjerknes, J. (1966), A possible response of the atmospheric Hadley circulation to equatorial anomalies of ocean temperature, *Tellus*, *18*, 820-829.

Brito-Castillo, L., A. Leyva-Contreras, A. V. Douglas, and D. Lluch-Belda (2002), Pacific decadal oscillation and the filled capacity of dams on the rivers of the Gulf of California continental watershed, *Atmosfera*, *15*, 121-138.

Bunting, A. H., M. D. Dennett, J. Elston, and J. R. Milford (1976), Rainfall trends in the West African Sahel, *Quarterly Journal of the Royal Meteorological Society*, *102*, 59-64.

Cancelliere, A., D. Mauro, B. Bonaccorso, and G. Rossi (2007), Investigating the potential of NAO index to forecast droughts in Sicily, *In* Hydrology Days, Colorado State University, Fort Collins, CO., pp.

Canon, J., J. Gonzalez, and J. Valdes (2007), Precipitation in the Colorado River Basin and its low frequency associations with PDO and ENSO signals, *Journal of Hydrology*, *333*, 252-264.

Cayan, D. R., M. D. Dettinger, H. F. Diaz, and N. E. Graham (1998), Decadal climate variability of precipitation over western North America, *Journal of Climate*, *11* (12), 3148-3166.

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.

Chiew, F., and L. Siriwardena (2005), Trend/change detection software and user guide, CRC for Catchment hydrology, Australia, www.toolkit.net.au/trend.

Colorado River Basin Climate (2005), Special Publication for Association of California Water Agencies and Colorado River Water Users Association Conferences.

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 G. N. Kiladis (1992), Atmospheric teleconnection associated with the extreme phase of the Southern Oscillation *In El Niño: Historical and Paleoclimatic Aspects of the Southern Oscillation*, H. F. D. a. V. Markgraf, editor., Cambridge University Press, 7-28.

Dibike, Y. B., D. Solomatine, and M. B. Abbott (1999), On the encapsulation of numerical hydraulic models in artificial neural network, *Journal of Hydraulic Research*, *37* (2), 147-161.

Dibike, Y. B., S. Velickov, D. Solomatine, and M. B. Abbott (2001), Model induction with support vector machines: Introduction and application, *Journal of Computing in Civil Engineering*, *15* (3), 208-216.

Dickson, R. R., T. J. Osborn, J. W. Hurrell, J. Meincke, J. Blindheim, B. Adlandsvik, T. Vinje, G. Alekseev, and W. Maslowski (2000), The Arctic Ocean response to the North Atlantic oscillation, *Journal of Climate*, *13*, 2671-2696.

Dracup, J. A., and E. Kahya (1994), The relationship between US streamflow and La Niña., *Water Resources Research*, *30* (7), 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 U.S., *Geophysical Research Letters*, *28*, 2077-2080.

Frei, C., and C. Schar (2000), Detection Probability of Trends in Rare Events: Theory and Application to Heavy Precipitation in the Alpine Region, *Journal of Climate*, *14*, 1568-1584.

French, M. N., W. F. Krajewski, and R. R. Cukendall (1992), Rainfall forecasting in space and time using a neural network, *Journal of Hydrology*, *137* (1-4), 1-31.

Geisser, S. (1975), The predictive sample reuse method with applications, *Journal of American Statistical Association*, *70*, 320-328.

Gershunov, A., and T. P. Barnett (1998), Interdecadal modulation of ENSO telecommunications, *Bulletin of the American Meteorological Society*, *79*, 2715-2726.

Giannini, A., M. A. Cane, and Y. Kushnir (2001), Interdecadal changes in the ENSO teleconnection to the Caribbean region and the North Atlantic oscillation, *Journal of Climate*, *14*, 2867-2879.

Gill, M. K., T. Asefa, M. Kemblowski, and M. McKee (2006), Soil moisture prediction using support vector machines, *Journal of American Water Resources Association*, *42* (4), 1033-1046.

Gonzalez-Hidalgo, J. C., M. D. Luis, J. Raventos, and J. R. Sanchez (2001), Spatial distribution of seasonal rainfall trends in a western mediterranean area, *International Journal of Climatology*, *21*, 843-860.

Grantz, K., B. Rajagopalan, M. P. Clark, and E. A. Zagona (2005), A technique for incorporating large-scale climate information in basin-scale ensemble streamflow forecasts, *Water Resources Research*, *41*, W10410, doi:10.1029/2004WR003467.

Gray, S. T., J. L. Betancourt, C. L. Fastie, and S. T. Jackson (2003), Patterns and Sources of Multidecadal Oscillations in Drought-Sensitive Tree-Ring Records from the Central and Southern Rocky Mountains., *Geophysical Research Letters*, *31*, L12205.

Guttman, N. B., and R. G. Quayle (1996), A historical perspective of U.S. climate divisions, *Bulletin of the American Meteorological Society*, *77*, 293-303.

Gutzler, D. S., D. M. Kann, and C. Thornbrugh (2002), Modulation of ENSO-based long-lead outlooks of Southwestern U.S. winter precipitation by the pacific decadal oscillation, *Bulletin of the American Meteorological Society*, *17*, 1163-1172.

Hallowed, A. B., and W. W. Wooster (1992), Variability of winter ocean conditions and strong year classes of northeast Pacific groundfish, *ICES Marine Science Symposia*, *195*, 433-444.

Hamlet, A. F., P. W. Mote, M. P. Clark, and D. P. Lettenmaier (2005), Effects of temperature and precipitation variability on snowpack trends in the western United States, *American Meteorological Society*, *18*, 4545-4561.

Hare, S. R., and N. J. Mantua (2000), Empirical evidence for North Pacific regime shifts in 1977 and 1989, *Progress in Oceanography*, *47* (2-4), 103-145.

Haykin, S. 2003. Neural Networks: A comprehensive foundation, Fourth Indian Reprint edition. Pearson Education, Singapore.

Haylock, M., and N. Nicholls (2000), Trends in extreme rainfall indices for an updated high quality data set for Australia, 1910-1998, *International Journal of Climatology*, *20*, 1533-1541.

Hennessy, K. J., R. Suppiah, and C. M. Page (1999), Australian rainfall changes, 1910-1995, *Australian Meteorological Magazine*, *48*, 1-13.

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., W. A. Leetmaa, Y. Xue, and A. Barnston (2000), Dominant factors influencing the seasonality predictability of U.S. precipitation and surface air temperature, *Journal of Climate*, *13*, 3994-4017.

Holbrook, S. J., R. J. Schmitt, and J. S. Stephens (1997), Changes in an assemblage of temperature reef fishes with climate shift, *Ecological Society of America*, *7* (4), 1299-1310.

Hsu, K. L., X. Gao, S. Sorooshian, and H. V. Gupta (1997), Precipitation estimation from remotely sensed information using artificial neural networks, *Journal of Applied Meteorology*, *36*, 1176-1190.

Hsu, K.-l., H. V. Gupta, and S. Sorooshian (1995), Artificial neural network modeling of the rainfall-runoff process, *Water Resources Research*, *31* (10), 2517-2530.

Hu, Q., and S. Feng (2001), Variations of teleconnection of ENSO and interannual variation in summer rainfall in the central United States, *Journal of Climate*, *14*, 2469-2480.

Hurrell, J. W. (1995), Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation, *Science*, 269 (5224), 676-679.

Ingraham, W. J., R. K. Reed, J. D. Schumacher, and S. A. Macklin (1991), Interannual variability of circulation in the Gulf of Alaska in relation to water properties and fisheries resources, *Eos, Transactions, American Geophysical Union*, *72*, 257-264.

IPCC (2007), Climate Change 2007: The Physical Science Basis., *In* Report of the Intergovernmental Panel on Climate Change, edited by Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor, and H.L. Miller, Cambridge UK, pp.

Kahya, E., and J. A. Dracup (1993), U.S. streamflow patterns in relation to the El Niño/southern oscillation, *Water Resources Research*, *29* (8), 2941-2503.

Kalra, A., and S. Ahmad (2009), Using oceanic-atmospheric oscillations for long lead time streamflow forecasting, *Water Resources Research*, *45*, W03413, doi:10.1029/2008WR006855.

Kalra, A., and S. Ahmad (2011), Evaluating changes and estimating seasonal precipitation for Colorado River Basin using stochastic non-parametric disaggregation technique, *Water Resources Research*, doi:10.1029/2010WR009675.

Kalra, A., T. C. Piechota, R. Davies, and G. A. Tootle (2008), Changes in U.S. streamflow and Western U.S. snowpack, *Journal of Hydrologic Engineering*, *13* (3), 156-163.

Kane, R. P. (1999), El Niño timings and rainfall extremes in India, Southeast Asia and China, *International Journal of Climatology*, *19*, 653-672.

Karl, T. R., and R. W. Knight (1997), Secular trends of precipitation amount, frequency, and intensity in the United States, *Bulletin of the American Meteorological Society*, *79* (2), 231-241.

Karl, T. R., C. N. Williams Jr., and P. J. Young (1986), A model to estimate the time of observation bias associated with monthly mean maximum, minimum, and mean temperatures for United States, *Journal of Climate and Applied Meteorology*, *25*, 145-160.

Kendall, M. G. 1975. Rank Correlation Methods, Griffin, London.

Kerr, R. A. (1992), Unmasking a shifty climate system, *Science*, *255* (5051), 1508-1510.

Khalil, A. F., M. McKee, M. Kemblowski, T. Asefa, and L. Bastidas (2006), Multiobjective analysis of chaotic dynamic systems with sparse learning machines, *Advances in Water Resources*, *29*, 72-88.

Kim, T.-W., J. B. Valdes, B. Nijssen, and D. Roncayolo (2006), Quantification of linkages between large-scale climate patterns and precipitation in the Colorado River Basin, *Journal of Hydrology*, *321*, 173-186.

Kim, T.-W., C. Yao, and J.-H. Ahn (2008), Influence of climate variation on seasonal precipitation in the Colorado River Basin, *Stochastic Environmental Research and Risk Assessment*, *22*, 411-420.

Knowles, N., M. D. Dettinger, and D. R. Cayan (2006), Trends in snowfall versus rainfall in the western United States., *Journal of Climate*, *19*, 4545-4559.

Kuligowski, R. J., and A. P. Barros (1998), Experiments in short-term precipitation forecasting using artificial neural networks, *Weather Forecast*, *13* (4), 1194-1204.

Kumar, M., and C. J. Duffy (2009), Detecting hydroclimatic change using spatio-temporal analysis of time series in Colorado River Basin, *Journal of Hydrology*, *374*, 1-15.

Kushnir, Y., V. J. Cardone, J. G. Greenwood, and M. A. Cane (1997), The recent increase in the North Atlantic wave heights, *Journal of Climate*, *10*, 2107-2113.

Lau, K. M., and H. T. Wu (2001), Principal modes of rainfall-SST variability of the Asian summer monsoon: A reassessment of the Monsoon-ENSO relationship, *Journal of Climate*, *14*, 2480-2895.

Legates, D. R., and G. J. McCabe (1999), Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation, *Water Resources Research*, *35* (1), 233-241.

Lehmann, E. L. 1975. Nonparametrics, Statistical Methods Based on Ranks, Holden-Day Inc, California.

Lin, G.-F., G.-R. Chen, and P.-Y. Huang (2010), Effective typhoon characteristics and their effects on hourly reservoir inflow forecasting, *Advances in Water Resources*, *33*, 887-898.

Lin, G.-F., G.-R. Chen, M.-C. Wu, and Y.-C. Chou (2009), Effective forecasting of hourly typhoon rainfall using support vector machines, *Water Resources Research*, *45*, W08440, doi:10,1029/2009WR007911.

Liong, S.-Y., and C. Sivapragasam (2002), Flood stage forecasting with support vector machines, *Journal of American Water Resources Association*, *38* (1), 173-186.

Luis, M. D., J. Raventos, J. C. Gonzalez-Hidalgo, J. R. Sanchez, and J. Cortina (2000), Spatial analysis of rainfall trends in the region of Valencia (East Spain), *International Journal of Climatology*, *20*, 1451-1469.

Luk, K. C., J. E. Ball, and A. Sharma (2000), A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting, *Journal of Hydrology*, *227* (1-4), 56-65.

Maidment, D. R. 1993. Handbook of hydrology, McGraw-Hill, New York.

Mann, H. B. (1945), Nonparametric tests against trend, *Econometrica*, *13*, 245-259.

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

Mantua, N. J., Y. Z. Hare, 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.

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* (1), 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., M. A. Palecki, and J. L. Betancourt (2004), Pacific and Atlantic Ocean Influences on multidecadal drought frequency in the United States, *Proceedings of the National Academy of Sciences*, *101* (12), 4136-4141.

McCabe, G. J., and D. M. Wolock (2002), A step increase in streamflow in the conterminous United States*, Geophysical Research Letters*, *29*, 2185, doi:10.1029/2002GL015999.

McCulloch, W. S., and W. Pitts (1943), A logic calculus of the ideas immanent in nervous activity, *Bulletin of Mathematical Biophysics*, *5*, 115-133.

Melesse, A. M., S. Ahmad, M. McClain, X. L. Wang, and H. Lim (2011), Suspended sediment load prediction of river systems: An artificial neural network approach, *Agriculture Water Management*, *98* (5), 855-866.

Merideth, R. 2000. A primer on climatic variability and change in the southwest. Udall Center for Studies in Public Policy and the Institute of the Study of Plant Earth, University of Arizona, Tucson, AZ. pp.28.

Miller, W. P., and T. C. Piechota (2008), Regional analysis of trend and step changes observed in hydroclimatic variables around the Colorado River Basin, *Journal of Hydrometeorology*, *9*, 1020-1034.

Milly, P. C. D., J. L. Betancourt, and M. Falkenmark (2008), Stationarity is dead: Whither water management?, *Science*, *319*, 573-574.

Moriasi, D. N., J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith (2007), Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, *Transactions of the American Society of Agricultural and Biological Engineers*, *50* (3), 885-900.

Namias, J. (1963), Interactions of circulation and weather between hemispheres, *Monthly Weather Review*, *91*, 482-486.

Nayak, A., D. G. Chandler, D. Marks, J. P. McNamara, and M. Seyfried (2008), Correction of electronic record for weighing bucket precipitation gauge measurements, *Water Resources Research*, *44*, W00D11, doi:10.1029/2008WR006875.

Nayak, A., D. Marks, D. G. Chandler, and M. Seyfried (2010), Long-term snow, climate, and streamflow trends at the Reynolds Creek Experimental Watershed, Owyhee Mountains, Idaho, United States, *Water Resources Research*, *46*, W06519, doi:10,1029/2008WR007525.

Piechota, T. C., and J. A. Dracup (1996), Drought and regional hydrologic variation in the United States: Association with the El. Nino-Southern Oscillation, *Water Resources Research*, *32* (5), 1359-1373.

Piechota, T. C., J. A. Dracup, and R. G. Fovell (1997), Western U.S. Streamflow and Atmospheric Circulation Patterns El Niño-Southern Oscillation (ENSO), *Journal of Hydrology*, *201* (1-4), 249-271.

Prairie, J., and R. Callejo (2005), Natural flow and salt computation methods, *Rep.* Bureau of Reclamation.

Pulwarty, R. S., and T. S. Melis (2001), Climate extremes and adaptive management on the Colorado River: lessons from the 1997-1998 ENSO event, *Journal of Environmental Management*, *63*, 207-324.

Qian, B., J. Corte-real, and H. Xu (2000), Is the North Atlantic Oscillation the most important atmospheric pattern for precipitation in Europe, *Geophysical Research Letters*, *105*, 11901-11910.

Rajagopalan, B., and U. Lall (1998), Interannual variability in western US precipitation, *Journal of Hydrology*, *210*, 51-67.

Raman, H., and N. Sunilkumar (1995), Multivariate modelling of water resources time series using artificial neural networks., *Journal of Hydrologic Sciences*, *40* (2), 145-163.

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 indices, *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.

Ropelewski, C. F., and M. S. Halpert (1986), North American precipitation and temperature patterns associated with El-Niño-Southern Oscillation (ENSO), *Monthly Weather Review*, *114*, 2165-2352.

Ropelewski, C. F., and P. D. Jones (1987), An Extension of the Tahiti-Darwin Southern Oscillation Index, *American Meteorological Society*, *115* (9), 2161-2165.

Sax, J. L., B. H. Thompson, J. D. Leshy, and R. H. Abrams (2000), Legal control of water resources: Cases and Materials, West Group, pp: 956.

Schölkopf, B., K.-K. Sung, C. J. C. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik (1997), Comparing Support Vector Machines with Gaussian Kernels to Radial Basis Function Classifiers, *IEEE Transactions on Signal Processing*, *45* (11), 2758-2765.

Segond, M.-L., C. Onof, and H. S. Wheater (2006), Spatial-temporal disaggregation of daily rainfall from a generalized linear model*, Journal of Hydrology*, *331* (3-4), 674-689.

Serreze, M. C., M. P. Clark, R. L. Armstrong, and D. A. McGinnis (1999), Characteristics of the western United States snowpack from telemetry (SNOTEL) data, *Water Resources Research*, *35*, 2145-2160.

Singh, J., H. V. Knapp, J. G. Arnold, and M. Demissie (2005), Hydrologic modeling of the Iroquois River watershed using HSPF and SWAT, *Journal of American Water Resources Association*, *41* (2), 361-375.

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

Smola, A. J., B. Schölkopf, and K.-R. Muller (1998), The connection between regularization operators and support vector kernels, *Neural Networks*, *11*, 637-649.

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.

Stockton, C. W., and G. C. Jacoby (1976), Long-term surface water supply and streamflow levels in the Upper Colorado River Basin, *In* Lake Powel Research Bulletin No. 18, Inst. of Geophysics and Planetary Physics, University of California, Los Angeles, pp 70.

Stone, M. (1974), Cross-validatory choice and assessment of statistical predictions, *Journal of the Royal Statistical Society*, *36*, 111-147.

Suykens, J. A. K. (2001), Nonlinear modeling and support vector machines, *In* Proceedings of IEEE Instrumentation and measurement technology conference, Budapest, Hungary, pp 287-294.

Tarboton, D. G. (1994), The source hydrology of severe sustained drought in the southwestern United States, *Journal of Hydrology*, *161*, 31-69.

Thomas, B. E. (2007), Climatic fluctuations and forecasting of streamflow in the Lower Colorado River Basin, *Journal of American Water Resources Association*, *46* (6), 1550-1569.

Timbal, B. (2004), Southwest Australia past and future rainfall trends, *Climate Research*, *26*, 233-249.

Tokar, A. S., and P. A. Johnson (1999), Rainfall-runoff modeling using Artificial Neural Networks, *Journal of Hydrologic Engineering*, *4* (3), 232-239.

Tokar, A. S., and M. Markus (2000), Precipitation-Runoff modeling using artificial neural networks and conceptual models, *Journal of Hydrologic Engineering*, *5* (2), 156-161.

Tootle, G. A., T. C. Piechota, and A. Singh (2005), Coupled oceanic-atmospheric variability and U.S. streamflow, *Water Resources Research*, *41*, W12408, doi:10.1029/2005WR004381.

Trenberth, K. E., and J. W. Hurrell (1994), Decadal atmosphere-ocean variations in the Pacific, *Climate Dynamics*, *9* (6), 303-319.

Tripathi, S., V. V. Srinivas, and R. S. Nanjundiah (2006), Downscaling of precipitation for climate change scenarios: A support vector machine approach, *Journal of Hydrology*, *330*, 621-640.

Twarakavi, N. K. C., D. Mishra, and S. Bandopadhyay (2006), Prediction of arsenic in bedrock derived stream sediments at a gold mine site under conditions of sparse data, *Natural Resources Research*, *15* (1), 15-26.

Twarakavi, N. K. C., J. Simunek, and M. G. Schaap (2009), Development of pedotransfer functions for estimation of soil hydraulic paramters using support vector machines, *Soil Science Society of America Journal*, *73* (5), 1443-1452.

U.S. EPA (2002), Guidance for quality assurance project plans for modeling, EPA QA/G-5M, *Rep.* EPA/240/R-02/007. Washington, D.C.: U.S. EPA, Office of Environmental Information.

Vapnik, V. (1995), The nature of Statistical Learning Theory, Springer, New York.

Vapnik, V. (1998), Statistical Learning Theory, Wiley, New York.

Viles, H. A., and A. S. Goudie (2003), Interannual, decadal and multidecadal scale climate variability and geomorphology, *Earth-Science Reviews*, *61*, 105-131.

Walsh, J. E., W. L. Chapman, and T. Shy, L. (1996), Recent decrease of sea level pressure in the Central Arctic, *Journal of Climate*, *9*, 480-486.

Wang, B., R. G. Wu, and X. H. Fu (2000), Pacific-East Asia teleconnections: How does ENSO affect East Asian climate?, *Journal of Climate*, *13*, 1517-1536.

Wang, X. L., and V. R. Swail (2001), Changes of extreme wave heights in northern hemisphere oceans and related atmospheric circulation regimes, *Journal of Climate*, *14*, 2204-2221.

Webb, R. H., and J. L. Betancourt (1992), Climate variability and flood frequency of the Santa Cruz River, Pima County, Arizona., *In* U.S. Geological Survey Water Supply Paper 2379, pp 40.

Webb, R. H., R. Hereford, and G. J. McCabe (2005), Climatic fluctuations, drought, and flow in the Colorado River Basin, *In*: The state of the Colorado River ecosystem in Grand Canyon. U.S. Geological Survey Circular 1282, pp 57-68.

Wedgbrow, C. S., R. L. Wilby, H. R. Fox, and G. O'Hare (2002), Prospects for seasonal forecasting of summer drought and low river flow anomalies in England and Wales, *International Journal of Climatology*, *22* (2), 219-236.

Wilson, L. (1973), Variations in mean annual sediment yield as a function of mean annual precipitation, *American Journal of Science*, *273*, 335-349.

Xu, Z. X., K. Takeuchi, and H. Ishidaira (2004), Correlation between El Niño-Southern Oscillation (ENSO) and precipitation in South-east Asia and the Pacific region, *Hydrological Processes*, *18*, 107-123.

Yu, X., and S.-Y. Liong (2007), Forecasting of hydrologic time series with ridge regression in feature space, *Journal of Hydrology*, *332*, 290-302.

CHAPTER 3

USING OCEANIC-ATMOSPHERIC OSCILLATIONS FOR LONG LEAD TIME STREAMFLOW FORECASTING (PUBLISHED IN WATER RESOURCES RESEARCH)

Abstract

We present a data-driven model, Support Vector Machine (SVM), for long lead time streamflow forecasting using oceanic-atmospheric oscillations. The SVM is based on statistical learning theory that uses a hypothesis space of linear functions based on Kernel approach, and has been used to predict a quantity forward in time based on training from past data. The strength of SVM lies in minimizing the empirical classification error and maximizing the geometric margin by solving inverse problem. The SVM model is applied to three gages i.e. Cisco, Green River, and Lees Ferry in the Upper Colorado River Basin in the western United States. Annual oceanic-atmospheric indices, comprising of Pacific Decadal Oscillation (PDO), North Atlantic Oscillation (NAO), Atlantic Multidecadal Oscillation (AMO), and El-Niño-Southern Oscillations (ENSO) for a period of 1906–2001 are used to generate annual streamflow volumes with three years lead time. The SVM model is trained with 86 years of data (1906–1991) and tested with 10 years of data (1992-2001). Based on Correlation Coefficient, Root Means Square Error, and Nash Sutcliffe Efficiency Coefficient the model shows satisfactory results, and the predictions are in good agreement with measured streamflow volumes. Sensitivity analysis, performed to evaluate the effect of individual and coupled oscillations, reveals a strong signal for ENSO and NAO indices as compared to PDO and AMO indices for the long lead time streamflow forecast. Streamflow predictions from the SVM model are

found to be better when compared with the predictions obtained from feed-forward back propagation Artificial Neural Network model and linear regression.

Citation: Kalra, A., and S. Ahmad (2009), Using oceanic-atmospheric oscillations for long lead time streamflow forecasting, *Water Resour. Res.*, *45*, W03413, doi:10.1029/2008WR006855.

3.1 Introduction

For decades, streamflow prediction has been regarded a benchmark problem for hydrologists (Chang and Chen, 2001). Water resource managers consider streamflow as one of the most vital surface hydrological variable for predicting water supply and water hazards such as floods and droughts (Chang and Chen, 2001; Grantz et al., 2005; Maier and Dandy, 2000; Zealand et al., 1999; McCabe et al., 2004). Streamflow prediction becomes relatively more important for western United States because its consumption of renewable water supplies (44%) is significantly higher than rest of the United States (4%) (el-Ashry and Gibbons, 1988).

The climate variability has direct impacts, both socially and economically, on the society (Redmond and Koch, 1991). The direct impacts occur through the hydrological cycle, for which climate is the primary driving force, and cause extreme events such as droughts and floods (Grantz et al., 2005; Redmond and Koch, 1991; McCabe and Dettinger, 2002; Dettinger et al., 1998; Hamlet and Lettenmaire, 1999; Regonda et al., 2005). Streamflow prediction becomes even more challenging under the stress of increased climatic variability (Grantz et al., 2005; Gutierrez and Dracup, 2001).

The oceanic-atmospheric modes are linked to climate variability and change, and occur at interdecadal and century time scales (Regonda et al., 2005). The oceanic-

atmospheric modes such as Pacific Decadal Oscillation (PDO), North Atlantic Oscillation (NAO), Atlantic Multidecadal Oscillation (AMO) and, El-Niño-Southern Oscillations (ENSO) Arctic Oscillations (AO), and Sea Surface Temperature (SST) influence streamflow across the globe and particularly in the Western United States (Rogers and Coleman, 2003; Tootle and Piechota, 2006; Kahya and Dracup, 1993; Piechota et al., 1997; Redmond and Koch, 1991; McCabe and Dettinger, 2002; Hamlet and Lettenmaire, 1999; Cayan and Webb, 1992). On one hand, climate variability is a challenge in reliably forecasting long-range streamflow patterns (Kahya and Dracup, 1993) but a correlation between oceanic-atmospheric oscillations and streamflow also provides a forecast opportunity. Researchers have investigated this correlation. Clark et al. (2001) showed the influence of ENSO on streamflow patterns over the United States. Kahya and Dracup (1993) studied the relationship between ENSO and unimpaired streamflow over the conterminous U.S. and indicated a strong ENSO signal in the mid-latitudes of the United States. Tootle et al. (2005) evaluated the streamflow responses to coupled and individual effects of four oceanic-atmospheric modes i.e., PDO, NAO, AMO, and ENSO over the conterminous United States and found a well established ENSO signal along with PDO, NAO, and AMO influencing the streamflow variability. Dettinger et al. (1998) studied multiscale streamflow responses to ENSO phenomena for regions in America, Australia, Northern Europe, and parts of Africa and Asia and indicated that the streamflow changes are associated with the weakening ENSO signals for these regions. McCabe et al. (2004) used the rotated principal component analysis (RPCA) to study the association between PDO and AMO and the multidecadal drought frequency for 344 climate divisions in the United States. The results indicated that the first streamflow component of RPCA was

correlated with PDO and second component with AMO. Hamlet and Lettenmaier (1999) performed streamflow forecasting for the Columbia River Basin using a macroscale hydrologic model and found that the increase in lead time for streamflow forecasting is achieved by using PDO and ENSO modes. Piechota et al. (1997) used the Principal Component Analysis (PCA), Cluster Analysis, and the Jackknifing Analysis to find that spatial and temporal modes of streamflow are associated with ENSO in the western United States. McCabe et al. (2007) studied the decadal to multidecadal sea-surface temperature variability and its association with the Upper Colorado River flow using RPCA. The results show a strong correlation of streamflow with AMO and PDO with first and second modes of RPCA, respectively.

It is evident that streamflow is dependent on climate variability occurring due to oceanic-atmospheric patterns. Based on the results from different previous streamflow studies the main modes of oscillations influencing streamflow patterns across the US comprise of PDO, NAO, AMO, and ENSO (Grantz et al., 2005; Regonda et al., 2005). Although there are other large scale climate indices such as Snow Water Equivalent, Geometric Potential, and Palmer Drought Severity Index for obtaining streamflow predictions but the four teleconnection patterns discussed above by far remain most popular. These teleconnection patterns are dominant on large scale and are important predictors in forecasting streamflow for the western United States (Dettinger et al., 1998). Since the effects of oceanic-atmospheric oscillations have a lag of several years, models based on these oscillations could be developed to increase the forecast lead time.

The common techniques used for modeling hydrological time-series and generating streamflows have been based on conceptual models and time-series models (Hsu et al.,

1995; Kraijenhoff and Moll, 1986; Yapo et al., 1996; Rodriguez-Iturbe and Valdes, 1979; Zealand et al., 1999). Conceptual models are based on mathematically simulating the process and physical mechanisms that contribute to the hydrological cycle (Zealand et al., 1999) and require a great deal of data inputs, which may involve field work and surveying. At times, it becomes challenging to deal with the empirical irregularities and periodicities occurring in the model that are often masked by noise (Zealand et al., 1999).

Time series modeling is a stochastic approach in which the time series models are fitted to the data for the purpose of forecasting and generating sequences used in simulation studies (Gutierrez and Dracup, 2001; Zealand et al., 1999). For modeling the water resources time series, the most commonly used approach in this category is multivariate autoregressive moving average (ARMA) model (Raman and Sunilkumar, 1995; Haltiner and Salas, 1988; Thompstone et al., 1985). The ARMA-type model uses a stationary data (Hipel, 1985) and follows a normal distribution for the data (Irvine and Eberhardt, 1992). The ARMA-type models are best suited for short-term forecasting based on daily or weekly time scales but not for long-term forecasting which involves seasonal or annual time scale (Tang et al., 1991). Although both modeling techniques can produce the long-term mean and variance of streamflow, they do a poor job in predicting the long-term streamflow variability (Dettinger et al., 1998; Gutierrez and Dracup, 2001).

Stochastic disaggregation models are also used to simulate streamflow preserving their temporal and spatial dependencies. These models are based on the nonparametric approaches and do not rely on assumptions that the data are drawn from a given probability distribution. Due to fewer assumptions, their applicability is much wider than the corresponding parametric methods. In hydrology, the most widely used non

parametric approaches of streamflow simulations are based on the traditional kernel nearest neighbor (KNN) time series bootstrap technique developed by Lall and Sharma (1996). The authors show the synthetic streamflow series generation from KNN is better than that from ARMA models. KNN technique is more flexible than the conventional models and is capable of reproducing both linear and nonlinear dependences (Sharma et al., 1997). The KNN method is preferred where the researchers are uncomfortable with the prior assumption about the data (e.g. linear or nonlinear).

Other statistical methods such as artificial neural networks (ANN's) are often considered as the prime choice for modeling hydrologic process. Neural networks are black box models that learn from a training data set mimicking the human-learning ability. They are robust to noisy data and can approximate multivariate non-linear relations among the variables. ANN's have been used for a wide range of different learning-from-data applications and input-output correlations of non-linear processes in water resources, and hydrology (Hsu et al., 1995; Zealand et al., 1999; Chang and Chen, 2001, Tingsanchali and Gautam, 2000; Imrie et al., 2000). A review of ANN applications in hydrology is available in the ASCE Task Committee report (2000b). Ahmad and Simonovic (2005) used feed-forward back-percolation ANN for estimation of runoff hydrograph parameters. Chang and Chen (2001) used ANN model to predict hourly streamflow and showed the superiority of ANN models over the ARMAX models.

Disadvantages associated with using neural networks are that they are 'data hungry', and some training algorithms are susceptible to local minima. Incorrect network definition i.e., number of nodes and number of hidden layers may lead to over fitting of the model, resulting in poor performances during testing.

Recently, another data-driven model i.e., Support Vector Machine (SVM) has gained popularity in many ANN dominated fields and has attracted the attention of many researchers (Liong and Sivapragasam, 2002; Asefa et al., 2006; Yu and Liong, 2007; Khalil et al., 2006; Tripathi et al., 2006). SVMs are trained with learning algorithm derived from optimization theory that uses a hypothesis space of linear functions in a higher dimensional feature space. The learning algorithm is then implemented in a learning bias derived from a statistical learning theory (Cristianini and Shaw-Taylor, 2000). SVMs are considered as kernel based learning systems rooted in the statistical learning theory and structural risk minimization (Haykin, 2003). SVMs have been successfully applied for pattern recognition and regression in different fields such as bio-informatics and artificial intelligence. There are few applications of SVM in hydrology. Liong and Sivapragasam (2002) indicated a superior SVM performance over ANN in forecasting flood stages for the Bangladesh River system. Asefa et al., (2006) applied SVM to forecast flows at seasonal and hourly time scale for the Sevier River Basin. The results indicated a better performance in solving site-specific, (uses local climatological data and requires less inputs than physical models) real-time, water resources problems as compared to the ANN models. Dibike et al. (2001) applied SVM for rainfall/runoff modeling and classification of digital remote sensing image data and compared results with ANN. SVM showed superior performance than the ANN approach. Gill et al. (2006) applied SVM for predicting soil moisture for four and seven days in advance using meteorological variables and compared the results with ANN model. SVMs soil moisture predictions were a good match with the actual soil moisture data and SVM model performed better than ANN model. It is noteworthy that in all the above-mentioned

applications, the SVM modeling results are better than results obtained from ANN models due to the high generalization characteristic of SVM models.

In this paper, a data-driven model, Support Vector Machine (SVM) is presented for predicting streamflow using four oceanic-atmospheric oscillations i.e., PDO, NAO, AMO, and ENSO. Streamflow predictions are made three years in advance for three gages in the Upper Colorado River Basin. Numerous studies have identified Upper Colorado River Basin (UCRB) and other regions in the U.S. showing responses to oceanic- atmospheric oscillations on a seasonal to annual scale but no study has incorporated these oscillations in a SVM model and forecasted streamflow volumes three years in advance. The sensitivity of individual and grouped oscillations in forecasting streamflow is evaluated. Nonparametric statistical tests including Man-Kendall, Spearman's Rho, and Rank Sum, and parametric test including autocorrelation and linear regression are performed to determine the trend/step changes for streamflow and oscillation modes. These tests help in evaluating the trends in the data that are based on the statistical properties such as mean, median, and variance. Moreover, a feed forward-back propagation ANN model is developed to predict streamflow volumes three years in advance. The streamflow volumes obtained using SVM are compared with the volumes obtained using ANN approach. Model performance is evaluated using correlation coefficient, root mean square error and model efficiency.

The paper is organized as follows. Sections 3.2 present a theoretical background on SVM. The study region and the data used are described in sections 3.3 and 3.4, respectively. In section 3.5, the proposed method to forecast streamflow and evaluate the significance of single and grouped oceanic-atmospheric modes for streamflow

predictions is presented. Section 3.6 summarizes the statistical properties of oscillation modes and streamflow using non parametric and parametric testing. Section 3.7 includes the results and discussion of streamflow volumes obtained using SVM for different models and a comparison with the streamflow volumes obtained using ANN. Section 3.8 summarizes and concludes the paper.

3.2 SVM Background

A brief description of theoretical basis of SVMs is provided in this section. A more detailed description on the subject is available in Vapnik (1995, 1998). The idea of learning machines was first proposed by Turing (1950). The trainer of learning machine is ignorant of the processes undergoing inside it, which is considered to be the most important feature of the machine (Turing, 1950). Vapnik (1995) discussed the features of learning machines by Turing (1950) and stated two important factors to control the generalization ability of the learning machine. The first factor is the error-rate on the training data, and the second factor is the capacity of the learning machine measured in terms of Vapnik-Chervonenkis (VC) dimensions (Vapnik and Chervonenkis, 1971). The nonlinearities in the system being modeled were handled by including kernels which act as building blocks for SVMs and are based on the requirements to satisfy Mercer's theorem (Vapnik, 1995; Vapnik, 1998; Cristianini and Shaw-Taylor, 2000). The requirement of kernels in optimization algorithm to achieve global optimum differentiates SVM from other learning machines such as ANNs, that may converge to local optima, and the use of kernels helps in obtaining different "machines".

The ultimate goal of working with statistical learning tools is to find a functional dependency, $f(\mathbf{x})$, between independent variables $\{\mathbf{x_1},\mathbf{x_2},\ldots\mathbf{x_L}\}$, obtained from $\mathbf{R}^K$. The

(dependent) output $\{y_1, y_2, \ldots y_L\}$ is obtained from $y \in \mathbf{R}$ selected from a set L of independent and identically distributed (i.i.d.) observations. The observations are called the regularized functionals, as shown in (Vapnik 1998; Smola et al., 1998) and have the following formulation:

Minimize $\frac{1}{2}\| \mathbf{w} \|^2 + C\sum_{i=1}^{L}\left(\xi_i + \xi_i^*\right)$

Subject to
$$\begin{cases} y_i - \sum_{j=1}^{K}\sum_{i=1}^{L} w_j x_{ji} - b \le \varepsilon + \xi_i \\ \sum_{j=1}^{K}\sum_{i=1}^{L} w_j x_{ji} + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases} \quad (1)$$

where $f(x) = \langle w, x \rangle + b$, $\langle w, x \rangle$ denotes the dot product of w and x**,** x is the input vector, w is the weights vector norm, ε is Vapniks insensitive loss function, C is capacity parameter cost, and b is bias. The first term in the minimizing equation refers to minimizing the VC-dimension of the learning machine, and the second term controls the empirical risk. The trade off between the flatness of *f* and the amount up to which deviations larger than *ε* tolerated are determined by C > 0. This corresponds to Vapniks "ε-insensitive" loss function (shown in Figure 12) and measures the agreement between estimated and actual measurements. An increase in C penalizes large errors and consequently leads to a decrease in approximation error. This is achieved by increasing the weight vector norm, ||**w**||, which does not necessarily guarantee a good generalization performance of the model. Also in equation (1), ξ_i and ξ_i^* are called the slack variables that determine the degree to which sample points are penalized if the error is larger than ε. Hence, for any (absolute) error smaller than ε, $\xi i = \xi i^* = 0$, no data points are required for

the objective function. This implies that not all the variables are used to estimate f(x). The functional dependency $f(\mathbf{x})$ is written as:

$$f(\mathbf{x}) = \sum_{j=1}^{K} w_j x_j + b \tag{2}$$

where, $\langle w, x \rangle$ denotes the dot product of **w** and **x**, K is the number of support vectors, and 'b' is the bias.

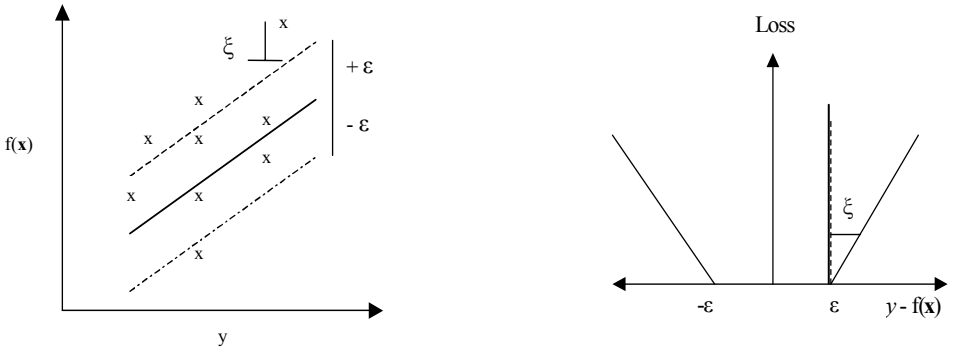


Figure 12: Pre-specified accuracy and slack variable ξ in SVM model.

Another technique of solving the optimization problem subject to constraints in loss function is using the dual formulation. In dual formulation, Lagrange multipliers α^* and α are introduced, and the minimization equation is solved by differentiating with respect to the primary variables, and it results in a maximizing problem.

Maximize

$$W(\alpha^*, \alpha) = -\varepsilon \sum_{i=1}^{L} (\alpha_i + \alpha_i^*) + \sum_{i=1}^{L} (y_i (\alpha_i - \alpha_i^*) - \frac{1}{2} \sum_{i,j=1}^{L} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) k(\mathbf{x}_i, \mathbf{x}_j)$$

subject to constraints $\sum_{i=1}^{L} (\alpha_i^* - \alpha_i) = 0, \qquad 0 \le \alpha_i, \alpha_i^* \le C,$ (3)

where i= 1,….,L is the sample size and the approximating function is

$$f(\mathbf{x}) = \sum_{i=1}^{N} (\alpha_i^* - \alpha_i) k(\mathbf{x}, \mathbf{x}_i) + b \tag{4}$$

In equation (3) and (4), α^*, α are Lagrange multipliers; and $k(\mathbf{x}, \mathbf{x_i})$ is the kernel function that measures non-linear dependence between two input variables. The $\mathbf{x_i}$'s are "support vectors", and N (usually $N << L$) is the number of selected data points or support vectors corresponding to values of the independent variable that are at least ε away from actual observations. The training pattern in the dual can be used to estimate the dot product of two vectors of any dimensions and is regarded as the advantage of the dual formulation (Smola et al., 1998). This advantage in SVM is used to deal with non-linear function approximations. Therefore, the steps involved in SVM modeling are: (1) selecting a suitable kernel function and kernel parameter (kernel width - γ), (2) specifying the 'ε' insensitive parameter, and (3) specifying the capacity parameter cost,'C'.

The working mechanism of SVM is shown in Figure 13. The input vector is transformed into the feature space using a function, Y. The transformation function is not computed explicitly but the dot products that correspond to evaluating kernel functions k at locations k (xi, x) are calculated. These dot products are then summed using weights (that are actually non-zero Lagrange multipliers) and added to the bias "b" that gives the final prediction.

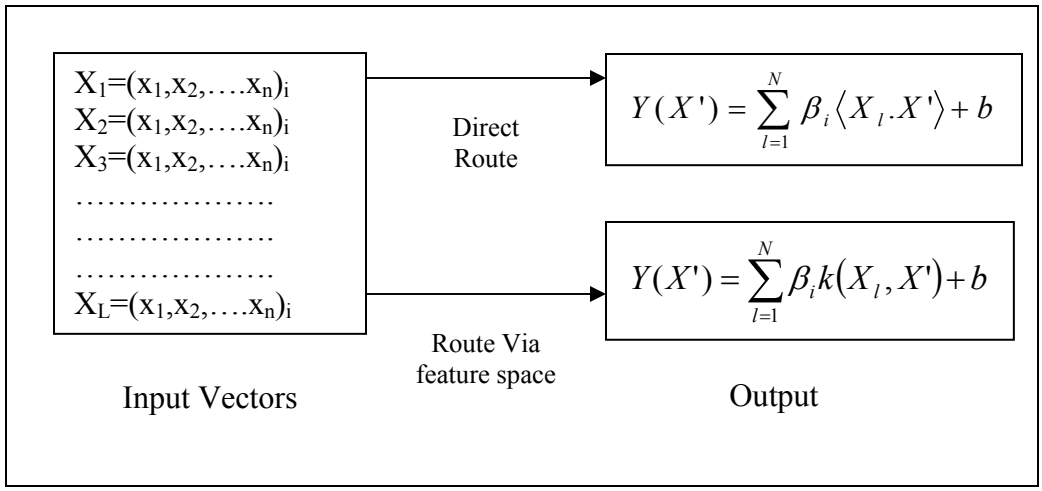


Figure 13: Flow diagram for SVM model.

The above concept is illustrated using the example of a simple 'sinc' function [sinc(x) = (sinx)/x for x = [0 1]. This function is approximated using the SVMs with a radial basis kernel. Figure 14 shows the resulting approximations using Vapniks "ε-insensitive" loss function (ε = 0.01 (a) and ε = 0.1 (b)). Figure 14 shows that few points are needed to capture the behavior of sinc function. The solid line represents the true values, and the dotted lines are the predictions with triangles being the support vectors. The changes in Vapniks "ε-insensitive" loss function result in the change in location and number of support vectors. Increase in Vapniks "ε-insensitive" loss function gives lesser number of support vectors (23 to 7) and results in a slight misfit between the true and predicted values. This demonstrates the ability of SVM to trade between accuracy of approximation and complexity of the approximation given in the objective function (Equation 1).

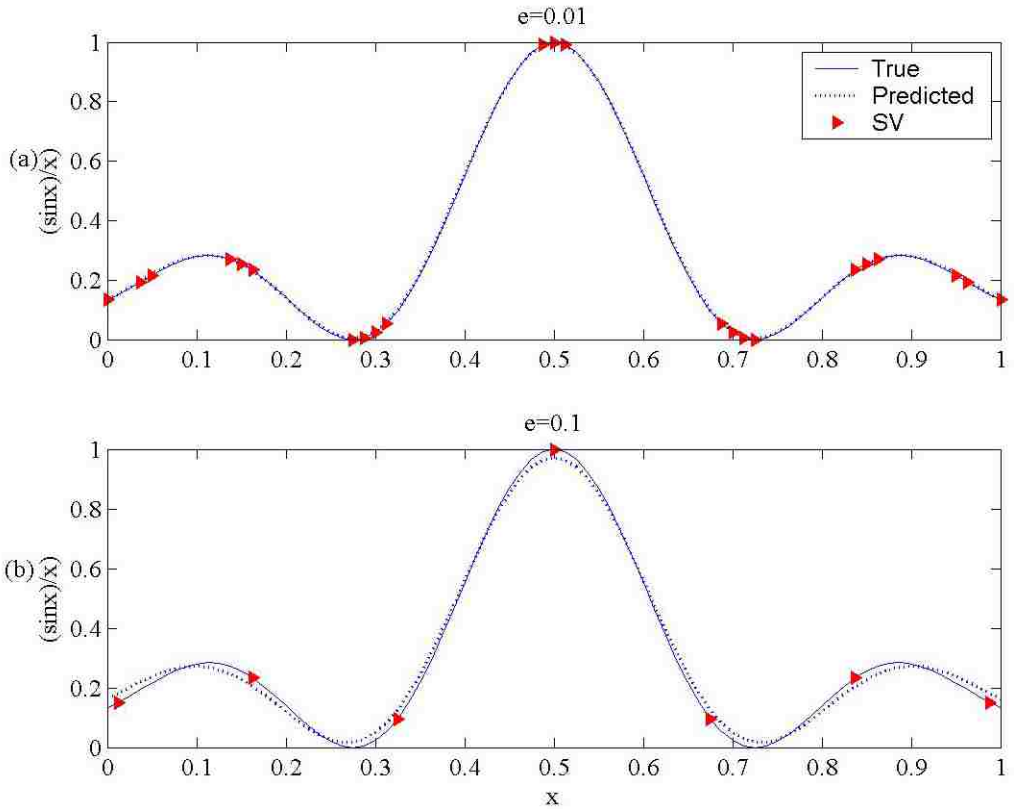


Figure 14: Examples of SVM for 'Sinc' Function for Vapniks "ε-insensitive" loss function (a) ε = 0.01 and (b) ε = 0.1.

3.3 Study Region: Upper Colorado River Basin (UCRB)

The Colorado River is a major source of water supply to the southwestern United States. The water from the Colorado River is allocated to seven states (California, Nevada, Utah, Arizona, Colorado, Wyoming, New Mexico) within the Colorado River basin based on the "Law of River" (Piechota et al., 2004). Due to growing population and agricultural activity, certain states such as California depend on water surpluses from the Colorado River. The Colorado River basin is composed of upper and lower basin. The Upper Colorado River Basin generates 90% of the Colorado River flow from spring-summer runoff due to snowmelt (Figure 15). The UCRB is defined as the part of basin upstream from the gage at Lees Ferry and just downstream of Glen Canyon Dam in

Northern Arizona. It serves Wyoming, Colorado, Utah, and New Mexico. It encompasses a total area of 286,000 km^2 and is comprised of mountains, agricultural, and low-density developments. The streamflow in the UCRB is allocated and regulated on the assumption of negligible changes in the mean and higher moment's statistical distribution of annual and decadal inflow to Lake Powell and Lake Mead. This is because Lake Powell and Lake Mead represent 85% of the storage capacity of the entire Colorado River Basin. The lower basin is downstream of Lees Ferry and serves California, Nevada, and Arizona. The supply to lower basin is governed by the water released from the upper basin.

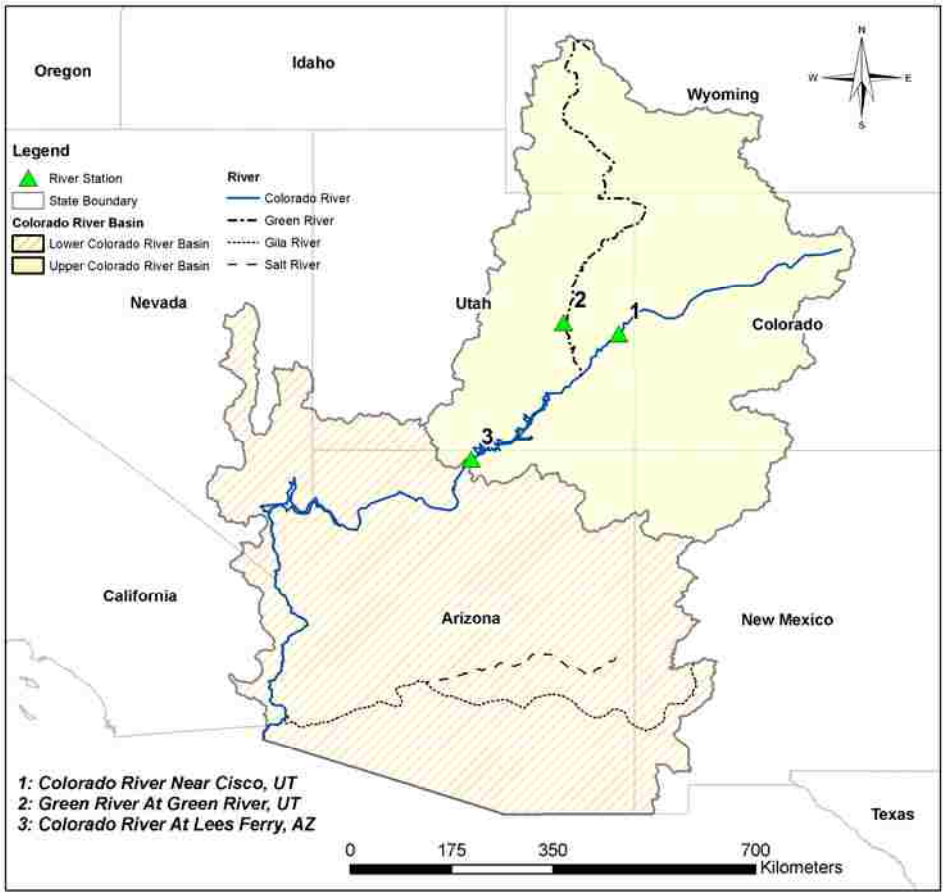


Figure 15: Map showing location of study area and streamflow gaging stations.

Although the water allocations in the UCRB are governed by the "Law of River", it still becomes critical every year to forecast streamflow that would be available for the entire basin (Piechota et al., 2004). This is due to the fact that water supply estimates for the UCRB are released monthly by the collaborative effort of National Weather Service (NWS), Natural Resources Conservation Service, U.S. Bureau of Reclamation, U.S. Geological Survey, local water district managers, and the Colorado River Basin Forecast Center (Tootle and Piechota, 2006). Moreover, water managers face challenges in forecasting streamflow due to the availability of small lead time (Tootle and Piechota, 2006; McCabe et al., 2007). The ability to provide long lead time (2-3 years in advance) forecasting of streamflow volumes for the UCRB could be useful for water managers in managing water resources system which includes the reservoir releases, allocation of water contracts, etc. (Tootle and Piechota, 2006). The focus of this study is on using Pacific and Atlantic Ocean modes i.e., PDO, NAO, AMO, and ENSO as predictor, in a data-driven model, to forecast streamflow three years in advance for the selected gages in UCRB.

3.4 Data

The data sets used to forecast long lead time streamflow are the oceanic-atmospheric modes of Pacific and Atlantic Ocean and the naturalized streamflow data for UCRB. Both the data sets are described in the ensuing sections.

3.4.1 Streamflow Data

Three streamflow gages in the UCRB, shown in Figure 15, are used in this study. These gauges are Colorado River near Cisco, Utah (site 1); Green River at Green River, Utah (site 2); and Colorado River at Lees Ferry, Arizona (site 3). Annual naturalized

streamflows volumes (acre-feet) at these locations are available for the 96 year period spanning 1909 – 2004. These flow volumes have been computed by removing anthropogenic impacts (i.e., reservoir regulation, consumptive water use, etc.) from the recorded historic flows. The natural flow data and additional reports describing these data are available at http://www.usbr.gov/lc/region/g4000/NaturalFlow/index.html. Two out of the three selected gages i.e., Cisco and Green are unimpaired (free from anthropogenic effects) and archived in the Hydro-Climatic Data Network (HCDN) (Slack et al., 1992). The third gage i.e., Lees Ferry is not a part of HCDN and flows at this gage are back calculated accounting for reservoir regulation, consumption and other diversions. However, these back calculated flows do not account for land use changes and are provisional, i.e., subjected to change. Lees Ferry is used in the analysis because of its location; it divides the Colorado River Basin in upper and lower basins.

3.4.2 Oceanic-Atmospheric Data (PDO, NAO, AMO, and ENSO)

Monthly oceanic-atmospheric modes are available for PDO, NAO, AMO, and ENSO. The PDO is an index of decadal-scale Sea Surface Temperature (SST) variability in the North Pacific Ocean (McCabe and Dettinger, 2002) and has been linked to hydro-climatic variability in the western United States (McCabe and Dettinger, 2002; Gershunov and Barnett, 1998). Monthly PDO index values are available from the Joint Institute Study of the Atmosphere and Ocean, University of Washington (http://jisao.washington.edu/pdo). Several studies have indicated two full phases of PDO in the past century (Tootle et al., 2005) with a periodicity of 25 – 50 years (Mantua and Hare, 2002).

The NAO index is the winter climate variability mode in North Atlantic Ocean and is defined as the difference in normalized mean winter (December to March) sea level pressure (SLP) anomalies between the island of Iceland and Portugal (Hurrell, 1995). NAO also has cool (negative index) and warm (positive index) regimes. The NAO index shows annual variability but has the tendency to remain in single phase for intervals lasting several years (Hurrell, 1995; Hurrell and Van Loon, 1995). Monthly NAO values are obtained from the National Center for Atmospheric Research (NCAR) (http://www.cgd.ucar.edu/cas/jhurrell/indices.html). NAO has exhibited interannual variability and long-term persistence in particular phases. Hurrell and Van Loon (1995) defined the NAO cool phase from 1952 – 1972 and again 1977 – 1980 and warm phases from 1950 – 1951, 1973 – 1976, and 1981 – present.

The continuing sequences of long duration changes in the sea surface temperature of the North Atlantic Ocean are termed as AMO (Enfield et al., 2001). AMO indices have been identified as important modes of influencing decadal to multidecadal (D2M) climate variability in the western United States (Enfield et al., 2001; Gray et al., 2003; Rogers and Coleman, 2003; McCabe et al., 2007). Monthly AMO index values comprising of cold (negative index) and warm phases (positive index) are obtained from the National Oceanic and Atmospheric Administration (NOAA) Climate Diagnostics Center http://www.cdc.noaa.gov/ClimateIndices/List/. The cool and warm phases of AMO can last from 20-40 years at a time (Enfield et al., 2001; Gray et al., 2003). Recent studies have indicated that from the mid 1990s, AMO has returned to warm phase (Enfield et al., 2001; Gray et al., 2003; McCabe et al., 2007). Studies have indicated that warm phases of

AMO have led to severe and prolonged droughts in the Midwest and Southwest United States.

The natural coupled cycle in the ocean-atmospheric system over the tropical Pacific is defined as ENSO. ENSO operates on a timescale of 2 – 7 years. Warm (El-Niño, positive index) and cool phases (La-Niña, negative index) of ENSO have been associated with regional and global climate variability and streamflow variability in the western United States (Piechota et al., 1997; Tootle et al., 2005; Regonda et al., 2005). Warm ENSO phases in the eastern coastal tropical Pacific have been used for forecasting streamflows for Columbia (Gutierrez and Dracup, 2001) and have been linked with decrease in fish population due to decreased nutrients (Ahrens, 1994). Currently, there is no single data set universally accepted for measurements of ENSO (Beebee and Manga, 2004). Commonly used ENSO indices include regional SST indices (e.g. Nino-1+2, Nino-3, Nino-4, Nino-3.4 and Japan Meteorological Agency (JMA)) and surface atmospheric-pressure based Southern Oscillation Index (SOI). The SOI is the difference between the de-seasonalized, normalized SLP anomalies over the Tahiti and Darwin used by the Australian Bureau of Meteorology (Ropelewski and Jones, 1987). SOI measures the tendency for easterly winds to blow along the Equatorial Pacific. Positive values of SOI indicate strong easterly winds in the tropics and the tropical Pacific and vice-versa in case of negative SOI values. Initially, SOI data was available since 1950, but has been extended and now available since 1882 from the Climate Prediction Center (CPC). For this study, monthly SOI values are obtained from NOAA-CDC (http://www.cdc.noaa.gov/ENSO/). Annual averages of all indices are computed to obtain the time series from 1906 – 2001. Figure 16 shows the time series plot for the average

annualized oscillation modes. It can be noted that that NAO and ENSO fluctuate every few years, whereas PDO and AMO fluctuate on decadal time scales.

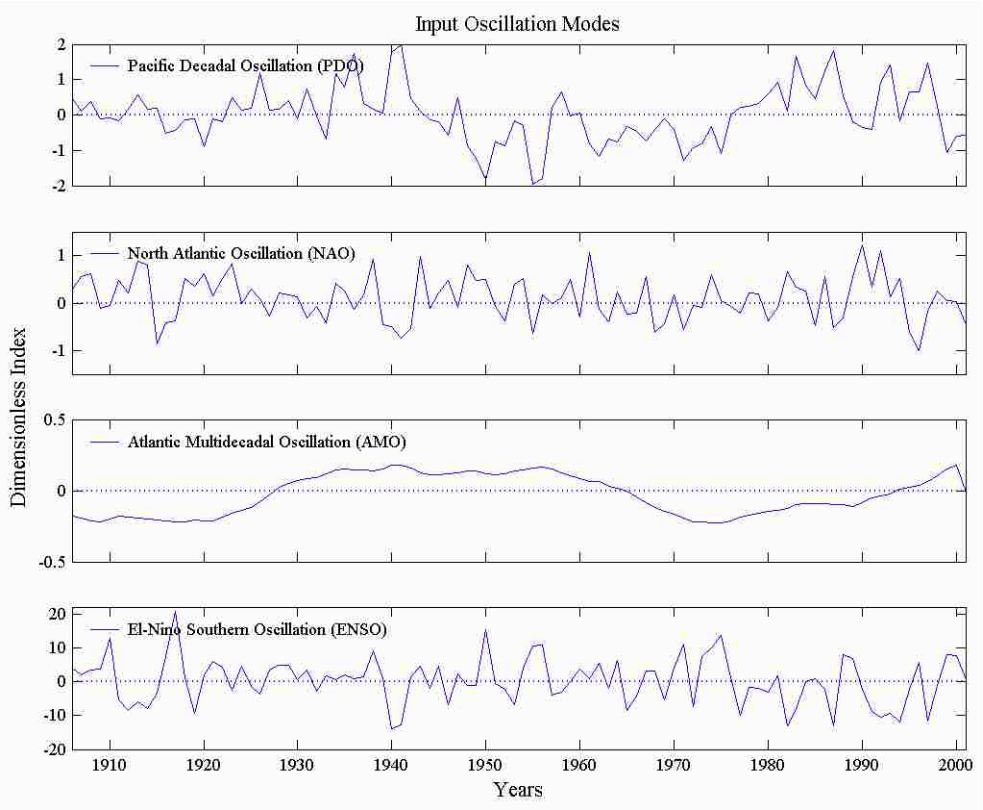


Figure 16: Fluctuations of input oscillation modes during 1906-2001.

3.5 Methods

The annual averaged indices of PDO, NAO, AMO, and ENSO for time step 't' are used to predict annualized streamflow volumes for 't+3' (where 't' is in years) for the three gages in the Upper Colorado River Basin. Four models are developed to predict streamflow volumes. In Model I (base case) all the four oceanic modes are used and that resulted in one model run. The input-output structure of SVM model is shown in Figure 17. The variable 'X' indicates the inputs, which are annualized average PDO, NAO, AMO, and ENSO indices. The variable 'Y' depicts the output, which is the annualized

streamflow volume predictions for 't+3'. The hidden layer takes into account the selection of kernels which is an important component of SVM and satisfies the Mercers Theorem as explained in the SVM Background section. In Model II, each oscillation mode is dropped one at a time and remaining three modes are used to predict streamflow. This resulted in four different model runs. In Model III, oscillation modes are dropped in pairs and then streamflow predictions were obtained using the remaining two modes. This resulted in six different model runs. In Model IV, only one oscillation mode is used (dropping three oscillation modes) to predict streamflow. This resulted in four different model runs. The models II - IV are designed to evaluate the relative significance of oceanic-atmospheric oscillations, individually and in different combination, in predicting streamflow.

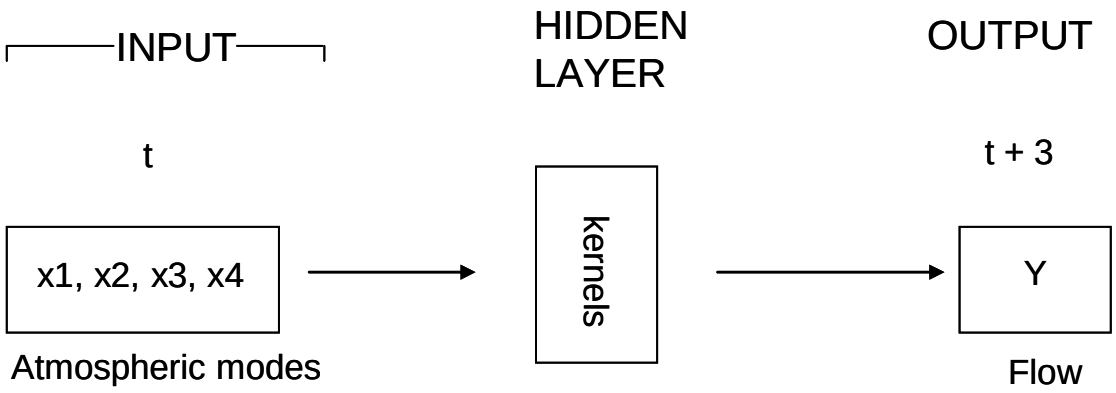
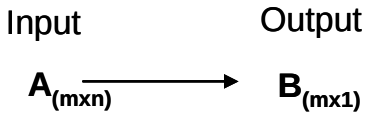


Figure 17: Flow diagram of SVM model structure (Model I).

The SVM model comprises of training and testing phases. The data set is divided in two parts; one is used in training (86 years i.e. 1906 – 1991) the model and other for testing (10 years i.e. 1992 – 2001) the predictions. The training stage aims at finding the optimal estimates of cost, C, insensitivity values, ε, and the kernel width, γ, to achieve the

best generalization. Each streamflow gage is considered independent and separate SVM models are developed for each gage. The matrix in training phase is



where 'A' is of size 'm x n', 'm' is the number of years (86) and 'n' is the total input variables which the model takes into account and equals 4 for model I, 3 for model II, 2 for model III, and 1 for model IV. The output matrix 'B' has a size of 'm x 1' where 'm' is the number of years (86) and the only output variable is streamflow volume. The matrix was replicated for all the selected gages. The SVM software package included in the 'R' software is used in this study (http://www.r-project.org/). The statistical testing criteria used for evaluating the effectiveness of the SVM model during the testing phase are correlations coefficient (R), Root Means Square Error (RMSE), and Nash-Sutcliffe Coefficient (E)

$$E = 1 - \frac{\sum_{i=1}^{n}(y_i - x_i)^2}{\sum_{i=1}^{n}(x_i - \overline{x_i})^2}$$

where, yi are the predicted streamflow volumes during the testing phase, xi are the observed values, $\overline{x}$ is the mean of observed values and n is the number of years in testing phase i.e. 10.

Radial basis kernel is used in SVM model. Schölkopf et al. (1997) concluded that Radial Basis Function (RBF) kernel performs better when compared with other kernels such as linear, polynomial, sigmoid or spline. Dibike et al. (2001) showed the superior efficiency of RBF kernels as compared to other kernels in SVM modeling applications.

Additionally, various other studies have indicated the favorable performances by using RBF kernels in hydrological forecasting problems (Asefa et al., 2006; Yu and Liong, 2007; Khalil et al., 2006; Gill et al., 2006). When RBF kernel is used, the Support Vectors algorithm automatically determines centers, weights and threshold that minimize an upper bound on the expected test error (Schölkopf et al., 1997). Khalil et al. (2006) inferred that the centralized feature of the RBF enables it to model regression process effectively.

In order to assess the relative performance of SVM model, we develop a feed forward-back propagation type ANN model. The feed forward-back propagation is adapted due to its applicability in variety of different problems (Hsu et al., 1995). The structure of ANN model comprised of one input layer, one hidden layer with three nodes, and one output layer with one node. The input layer is the first layer consisting of processing elements (PEs) referred to as nodes that connect the input variables. The input layer passes the input variables onto the subsequent layers of the network. The last layer is the output layer which connects to the output variable(s). The layer between the input and the output layer is called the hidden layer. The main function of the hidden layer is to enhance the networks ability to model complex functions. Details on the theoretical aspects of ANN are available in ASCE Task Committee, 2000a. Four ANN models are developed using the same training and testing data set used for SVM models. The comparison of SVM and ANN model predictions are made using the statistical performance measures of R, RMSE, and E.

Nonparametric Mann-Kendall and Spearman's Rho tests are performed to detect the trends in streamflow. Trends in streamflow (Groisman et al., 2001; Kalra et al., 2008) are

important as they help the water managers in responding to changes in water supply. The Rank Sum test is used to identify the step changes in the data. It is important to clearly differentiate between a gradual trend and a step change for climate change studies because the pattern of the trend change can be linear and continuous, whereas step changes are non linear, occur abruptly, and may reoccur in the future (McCabe and Wolock, 2002; Mantua and Hare, 2002). For analyzing step changes, 1977 was used as the year showing the step change. The "climate regime" shift occurring during the winter of 1977 has been documented by previous researchers (Holbrook et al., 1997; Mantua and Hare, 2002). Pearson correlation coefficient between oscillations modes and streamflow gages is calculated to evaluate the randomness (persistence over time) and correlation among the climate indices and streamflow gages.

The test results are evaluated at significance level of $p \leq 0.05$. The tests are performed using Trend software (www.toolkit.net.au/trend) which is designed to facilitate statistical testing for trend, change and randomness in hydrological and other time series data.

3.6 Statistical Properties of Oscillation Modes and Streamflow

Figure 18 shows the scatter plot for oscillations and streamflow and least square regression line at three gages. The non parametric correlations coefficients for streamflow and parametric correlation coefficients for oscillation modes are shown in Table 6. It can be noted that NAO has the highest correlation for Cisco and Lees Ferry, whereas PDO has the highest correlation for Green River gage. ENSO shows the weakest correlation for Cisco and Green River gage. The streamflow has a decreasing trend as depicted by the negative Mann-Kendall and Spearman's Rho coefficient and it is noteworthy that the

changes are due to trend and not due to any abrupt step change as the rank sum test shows no significance in the median values for the selected gages.

Table 6: Statistical testing of oscillation modes and streamflow gages at p≤0.05 confidence levels.

	Pearson Correlation Coefficient				Trend/Step Change (Streamflow)		
	Oscillation Modes				Man Kendall	Spearman Rho	Rank Sum
Gages	**PDO**	**NAO**	**AMO**	**ENSO**	***Correlation Coefficient***		***p-value***
Cisco	0.25	-0.5	-0.18	-0.14	-0.13	-0.19	0.75
Green River	0.4	0.3	-0.29	0.16	-0.09	-0.15	0.76
Lees Ferry	0.12	-0.32	-0.23	-0.16	-0.1	-0.19	0.76

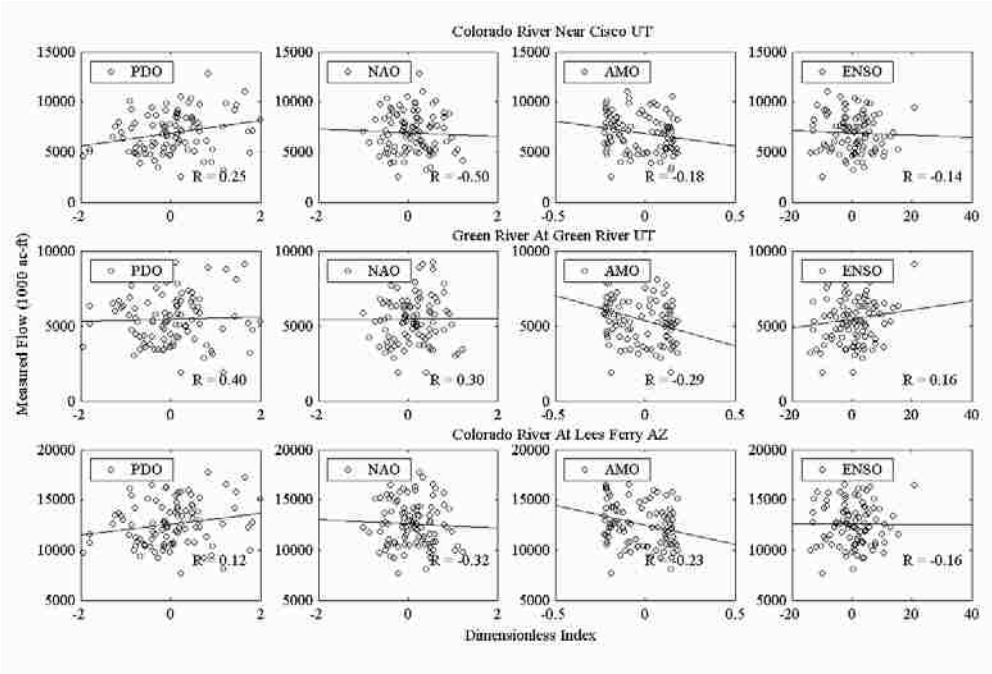


Figure 18: Scatter plots depicting correlation between oscillation modes and streamflow gages.

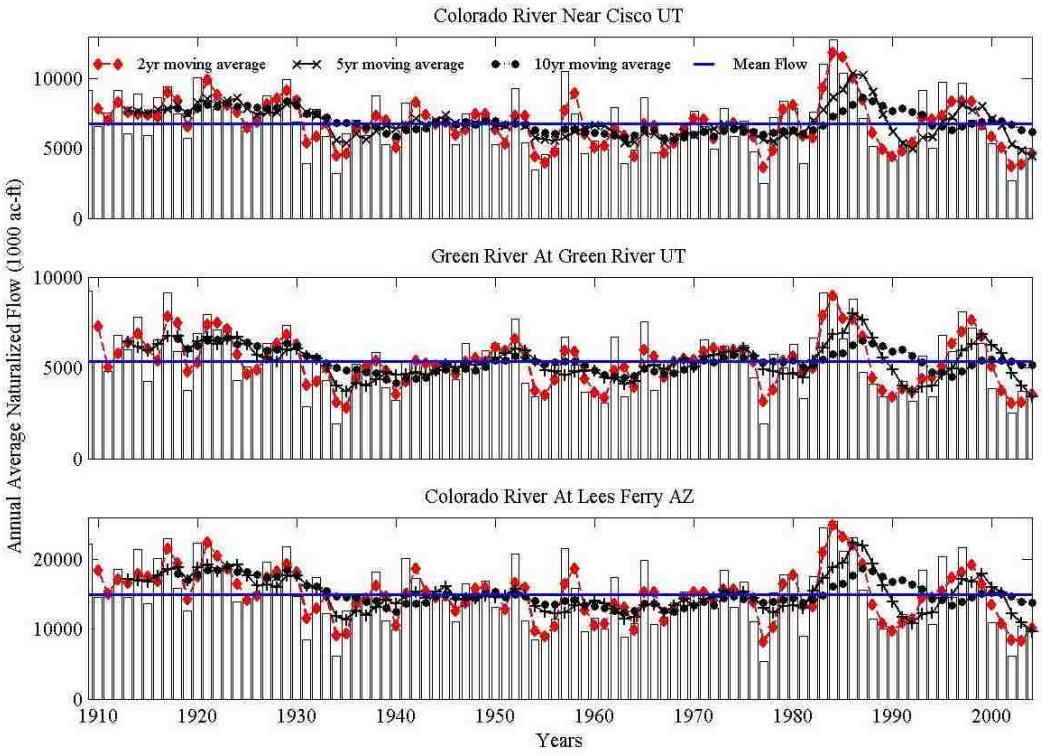


Figure 19: Streamflow variability for the selected gages. Averages using 2 year, 5 year , and 10 year moving windows are shown to depict the trend/step change in the data from 1909-2004. Bars represent the averaged annualized natural flows for the selected gages.

The nonparametric correlation coefficients are significant at $p \leq 0.05$. Figure 19 shows 2 year, 5 year, and 10 year moving averages for the streamflow at the three gages. It can be noticed that there is a decrease in streamflow volumes for the gages in the UCRB and the decreases are more significant around the year 2000, which coincides with the worst drought in the past 80 years for portions of the Upper Colorado River Basin (UCRB) (Piechota et al., 2004).

3.7 Results and Discussion

The results are discussed in two ensuing sections. The first section describes the SVM parameters estimation and modeling results for three gages during the training and testing

phases. The second section presents the ANN modeling results during testing phase and the comparison with the results obtained from SVM models.

3.7.1 SVM Models

The SVM modeling is performed in two stages: (a) training (1906 – 1991) and (b) testing (1992 – 2001). The training stage aims at finding the optimal cost, C, insensitivity value, ε, and radial basis kernel width, γ, to achieve the best generalization. During the testing stage, the ability of the trained SVM to predict final values is evaluated. SVM parameters can be estimated using three procedures: (a) based on a prior knowledge and user expertise, (b) using a thorough grid search approach, and (c) using an analytical estimation based on the statistical properties of the training data set. In this study, we opted for using the grid-based search approach. The optimal hyper-parameters for the SVM are estimated by searching within the feasible parameter space. The feasible parameter space for each hyper-parameter is constructed using the minimum (0) and maximum (100) possible values with 0.1 intervals that are given a priori. This is the most widely used approach and has been well documented (Cherkassky and Ma, 2004; Gill et al., 2006; Tokar and Markus, 2000; Asefa et al., 2006, Tripathi et al., 2006).

In Model I, all four oscillation modes are used for streamflow prediction. This results in one model run for each gage. Figure 20 shows the correlation between measured and predicted streamflows during training (Figure 20 a) and testing phase (Figure 20 b). Based on correlation criterion, the best model predictions are obtained for the Cisco gage with correlation of 0.84 and 0.87 during training and testing phases, respectively. The second best model predictions are for Lees Ferry gage with correlation of 0.72 and 0.81 during training and testing phases, respectively. For Green River gage the correlation

coefficient is 0.63 and 0.72 during training and testing phases, respectively. Similar results are observed based on the performance criteria of model efficiency, where model performance during testing for Lees Ferry (0.47) and Cisco (0.45) gages is better compared to Green river (0.29) gage (Table 7). Based on the performance measures, model I has acceptable predictions but it lacks in capturing the extreme (wet/dry) years in the data. This is evidenced by the predictions lying around the bisector line. Overall, satisfactory streamflow predictions are obtained for model I at 't+3' for the three gages with variations in performance measures.

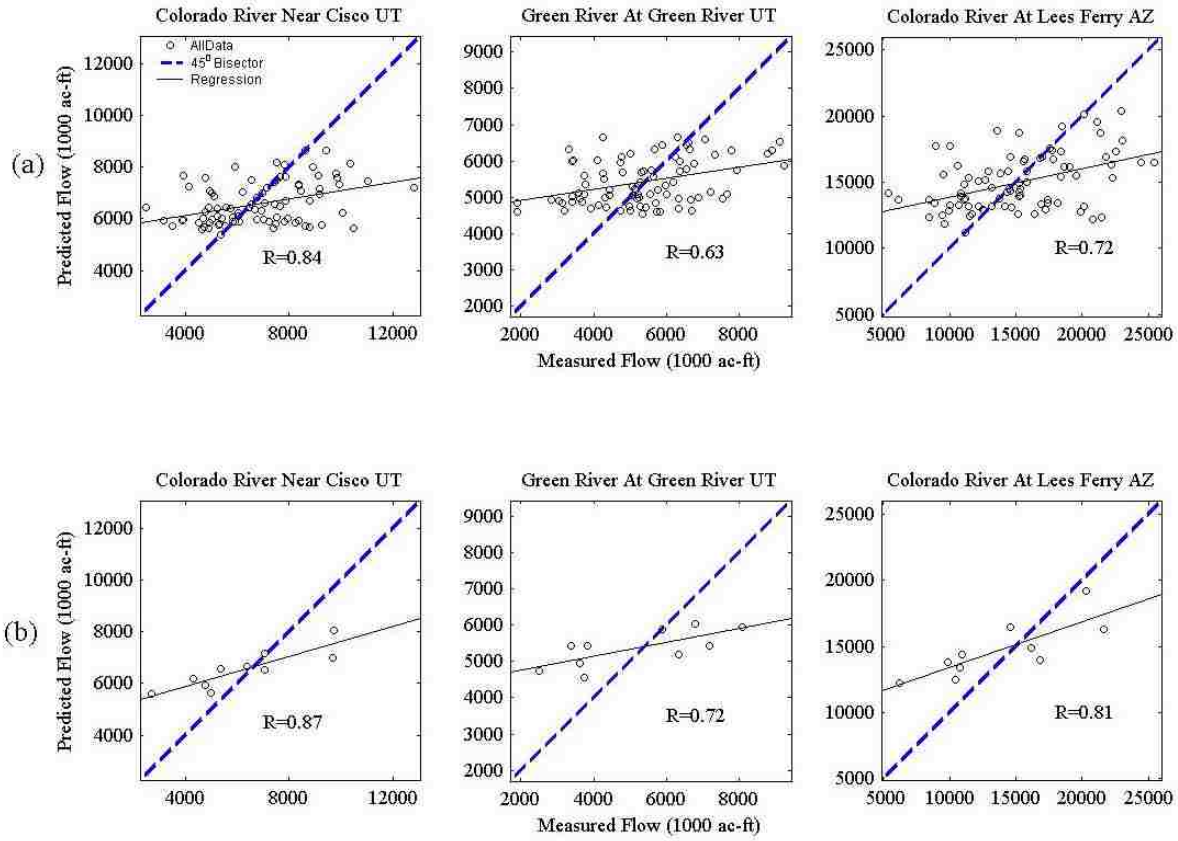


Figure 20: SVM predicted streamflow volumes at 't+3' for model I for (a) training phase and (b) testing phase at the three selected gages. Dashed line is the 45^0 bisector and solid line is true regression line between the measured and predicted streamflow volumes.

Table 7: Comparison of SVM and ANN models during testing phase. Drop 0, 1, 2, 3, and 4 refer to None, PDO, NAO, AMO, and ENSO, respectively. The RMSE values are in 1000 ac-ft. Best model estimates for each model and for each gage are shown in bold.

		SVM				*ANN*		
CISCO	***DROP***	***RMSE***	***R***	***E***		***RMSE***	***R***	***E***
Model I	0	**1592.86**	**0.87**	**0.45**		**1767.95**	**0.53**	**0.29**
Model II	1	**1488.38**	**0.91**	**0.67**		**1742.52**	**0.61**	**0.34**
	2	1713.68	0.77	0.36		1842.49	0.53	0.26
	3	**1453.89**	**0.88**	**0.44**		**1860.37**	**0.58**	**0.31**
	4	1652.48	0.79	0.41		1891.68	0.51	0.22
Model III	1,2	1729.72	0.76	0.35		1951.38	0.72	0.17
	1,3	**1452.38**	**0.88**	**0.54**		**1874.80**	**0.80**	**0.41**
	1,4	1622.62	0.82	0.43		1911.68	0.59	0.20
	2,3	1708.84	0.73	0.36		2253.76	0.57	0.11
	2,4	1823.50	0.68	0.28		2059.54	0.56	0.18
	3,4	1825.87	0.79	0.27		2081.49	0.54	0.16
Model IV	2,3,4	1906.03	0.70	0.19		2159.50	0.60	0.10
	1,3,4	**1806.05**	**0.76**	**0.26**		**2043.63**	**0.69**	**0.17**
	1,2,4	1958.91	0.69	0.16		2198.12	0.58	0.09
	1,2,3	**1671.31**	**0.75**	**0.40**		**2089.67**	**0.63**	**0.14**
GREEN								
Model I	0	**1536.87**	**0.72**	**0.29**		**1939.50**	**0.35**	**0.18**
Model II	1	**1339.96**	**0.81**	**0.46**		**1596.54**	**0.76**	**0.24**
	2	1543.64	0.66	0.29		1890.81	0.13	0.11
	3	**1471.47**	**0.70**	**0.35**		**1469.76**	**0.68**	**0.36**
	4	1603.97	0.61	0.23		2658.30	0.62	0.14
Model III	1,2	1541.00	0.60	0.29		1757.52	0.62	0.17
	1,3	**1341.67**	**0.74**	**0.46**		**1502.87**	**0.62**	**0.33**
	1,4	1459.72	0.72	0.36		1875.96	0.45	0.15
	2,3	1556.68	0.70	0.28		1728.25	0.53	0.13
	2,4	1624.75	0.67	0.21		1831.99	0.45	0.11
	3,4	1579.26	0.56	0.25		1853.94	0.43	0.21
Model IV	2,3,4	1642.33	0.61	0.18		1845.32	0.51	0.10
	1,3,4	**1573.51**	**0.69**	**0.26**		**1837.80**	**0.60**	**0.15**
	1,2,4	1635.08	0.55	0.20		1813.30	0.48	0.11
	1,2,3	**1558.29**	**0.63**	**0.27**		**1758.34**	**0.57**	**0.13**
LEES FERRY								
Model I	0	**3429.31**	**0.81**	**0.47**		**4625.47**	**0.56**	**0.20**
Model II	1	**3260.54**	**0.84**	**0.53**		**3666.08**	**0.65**	**0.28**
	2	3969.84	0.77	0.29		4706.68	0.43	0.18
	3	**3488.44**	**0.81**	**0.45**		**3514.60**	**0.67**	**0.24**
	4	3546.25	0.79	0.35		4353.29	0.50	0.16
Model III	1,2	4015.45	0.69	0.27		4291.79	0.56	0.23
	1,3	**3229.31**	**0.86**	**0.55**		**4156.78**	**0.78**	**0.30**
	1,4	3637.63	0.76	0.40		4378.65	0.50	0.14
	2,3	3876.21	0.79	0.32		4256.65	0.53	0.23
	2,4	4040.62	0.67	0.26		4582.94	0.51	0.17
	3,4	3905.82	0.74	0.31		4777.17	0.52	0.18
Model IV	2,3,4	4028.09	0.62	0.16		4796.22	0.33	0.15
	1,3,4	**3452.71**	**0.79**	**0.31**		**4440.66**	**0.54**	**0.22**
	1,2,4	3902.29	0.66	0.18		4898.89	0.30	0.12
	1,2,3	**3604.23**	**0.69**	**0.27**		**4600.55**	**0.45**	**0.20**

In Model II, oscillations are dropped individually and the remaining three oscillation modes are used to predict streamflow volumes. This results in four model runs for each gage. Results for Model II during the testing phase for the selected gages are shown in Figure 21. The results show significant improvement in all three performance measures and for all three gages compared with results from model I when PDO is dropped (Figure 21 (a), (b), and (c)). When AMO is dropped there is marginal improvement in some performance measures. For example, for the gage at Green (Figure 21 b) correlation slightly reduces compared to Model I but RMSE and E show some improvement (Table 7). When either NAO or ENSO are dropped there is noticeable deterioration in all three performance measures for all gages (Figure 21 (a), (b), and (c)).

In Model III, oscillations are dropped in pairs and the remaining two oscillation modes are used to predict streamflow volumes. This results in six model runs for each gage. It is observed that best performance measures were obtained by dropping PDO and AMO simultaneously for each of the gages (Table 7). An increase in correlation coefficient is observed for all 3 gages compared to Model I. RMSE and E also show improvement pointing to better predictions compared to model I when PDO and AMO are dropped together, compared to the other input combinations for all the gages. Considering all three model performance criteria the worst predictions are obtained when NAO and ENSO are dropped simultaneously.

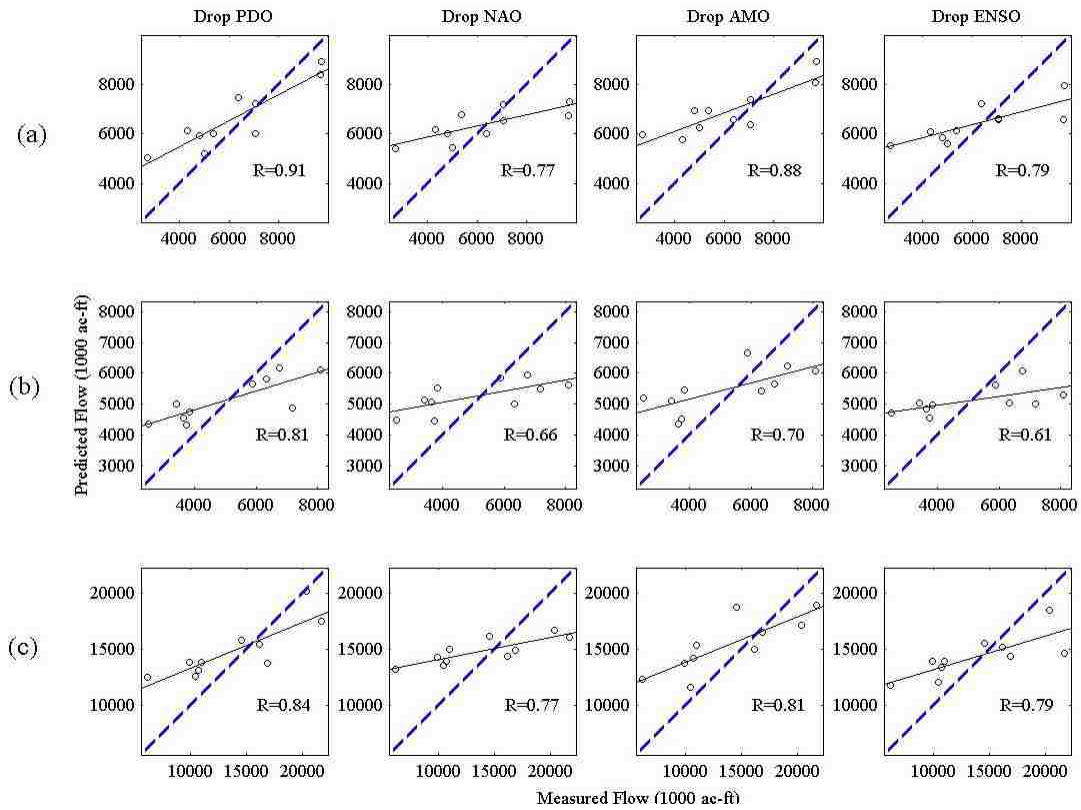


Figure 21: SVM predicted streamflow volumes at 't+3' for model II at (a) Cisco ,(b) Green, and (c) Lees Ferry gage. Dashed line is the 45^0 bisector and solid line is true regression line between the measured and predicted streamflow volumes.

In Model IV, individual oscillation modes are used to predict streamflow volumes. This results in four model runs for each gage (Figure 22 (a), (b), and (c)). Based on the performance measures (Table 7) relatively better predictions are obtained using NAO and ENSO as inputs compared to using PDO and AMO.

The results from Model II - IV point that NAO and ENSO individually and in combination have relatively stronger signal than PDO and AMO in three year lead streamflow predictions for the Upper Colorado River Basin. Although ENSO has the weakest correlation with the streamflow (Table 6) and NAO has the strongest correlation, both indices in combination provide the best predictions. The results indicate that the oscillations (NAO, ENSO) with short cycle periodicity (2-7 years) are more useful in

long lead time streamflow predictions as compared to the oscillations (PDO and AMO), which have long cycle periodicity (25 – 40 years).

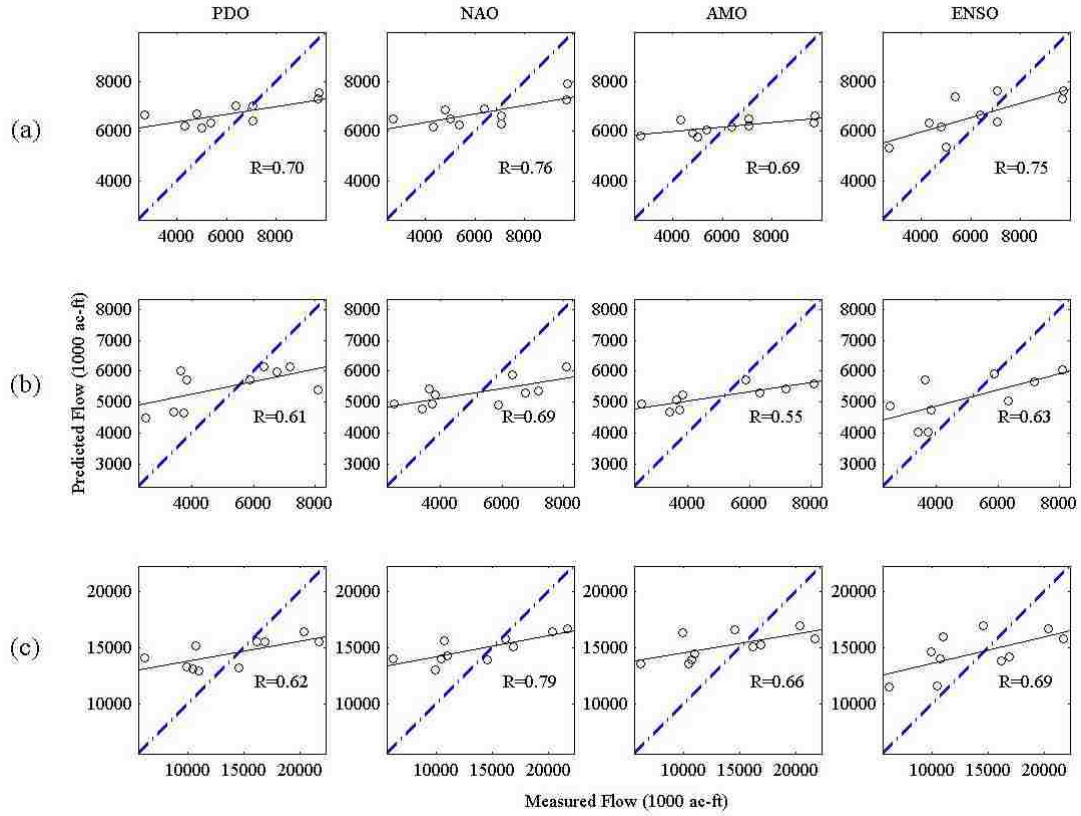


Figure 22: SVM predicted streamflow volumes at 't+3' for model IV at (a) Cisco ,(b) Green, and (c) Lees Ferry gage. Dashed line is the 45^0 bisector and solid line is true regression line between the measured and predicted streamflow volumes.

To test the model performance for different lead time streamflow predictions in the UCRB, R, RMSE, and E were calculated for SVM and ANN models using all four oscillation indices and lead times ranging from one to five years. The correlations between measured and predicted streamflow values during testing using the SVM model are shown in Figure 23 (a) (ANN model results not shown). It is noticed that using all four oscillation indices, correlation coefficient between predicted and measured volumes increases up to three years and then deteriorates. This was counterintuitive as one would

expect a decrease in forecast accuracy with increase in forecast lead time. A rigorous analysis was performed to understand this anomaly; lag 1, lag 2, and lag 3 SVM model streamflow predictions were made using all possible combinations of oscillation indices i.e. models I-IV at all gages. The best predictions obtained for lag 1, lag 2, and lag 3 for Lees Ferry is shown in Figure 23 (b, c, d). An interesting finding was that the best predictions for each lead time were result of a different combination of input indices. For example, the best predictions for lag 1 (Figure 23 (b)) are obtained using all four indices. The best predictions for lag 2 (Figure 23 (c)) are obtained using combination of PDO, NAO and AMO, and the best predictions for lag 3 (Figure 23 (d)) are obtained using combination of NAO and ENSO. Performance measures for Lees Ferry show that lag 2 predictions are better than lag 1 and lag 3 predictions when best possible input combination is used for each lag time, compared to using the same inputs for all lag times. This analysis shows that various combinations of oscillation indices can be used to enhance predictions for different lead times. Moreover, NAO comes across as an important predictor for the UCRB streamflow and can be used to extend lead time up to three years, which is the primary intent of the current research. Same analysis was performed for other gages (results not provided) and resulted in similar findings as reported for Lees Ferry gage.

To evaluate the influence of size of training data set we ran three modeling experiments at all gages with 86, 80 and 76 years of data for training, respectively. Correlation coefficient, RMSE and E were calculated to measure model performance. Figure 23 (e) shows that decreasing the training sample size from 86 years to 80 years and further to 76 years reduces the correlation coefficient between measured and

predicted streamflow volumes for the testing phase significantly. Similar trend was observed in other performance measures and at all three gages.

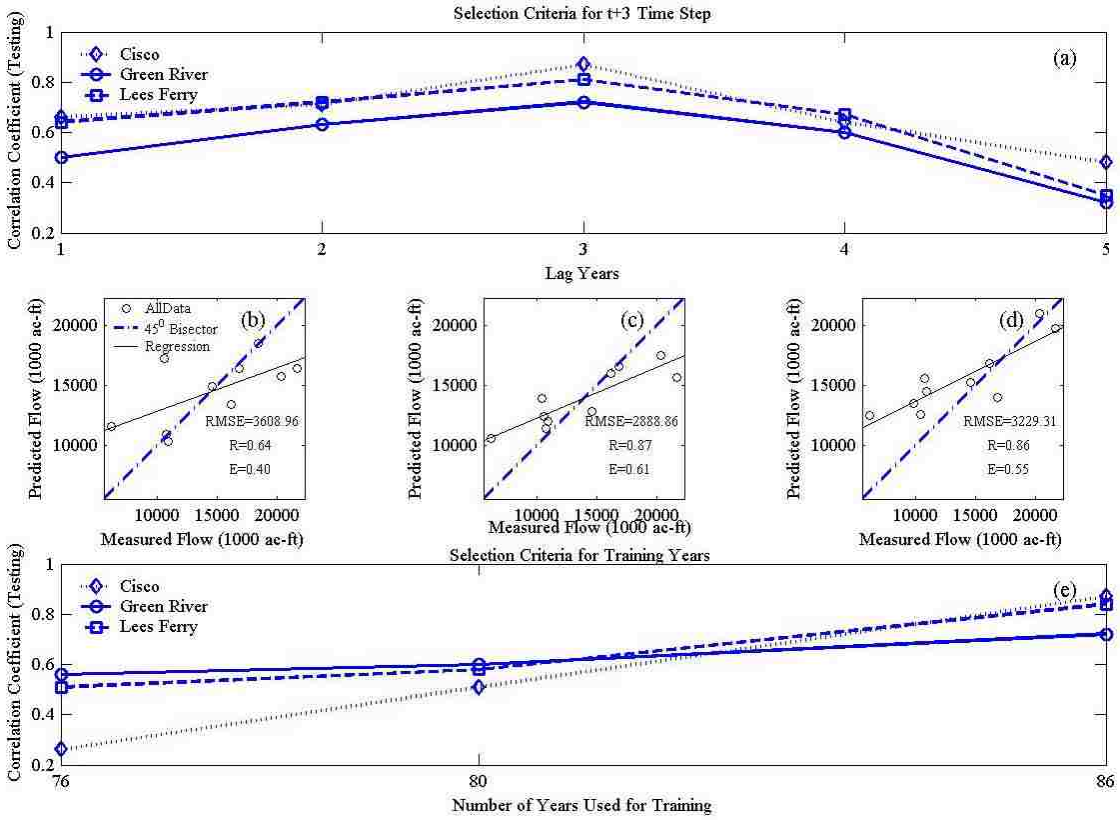


Figure 23: a) Correlation coefficient for test years at three gages for different lead times using all indices (b-d) model performance at Lees Ferry using best input combination for 1, 2 and 3 year lead time forecast (e) selection criteria for training years.

For example, R for the Lees Ferry gage with 86, 80 and 76 year of training data is 0.81, 0.58, and 0.51, respectively. E for the Lees Ferry gage with 86, 80 and 76 year of training data is 0.47, 0.30, and 0.18 respectively. RMSE for the Lees Ferry gage with 86, 80 and 76 year of training data is, 3429, 4192, and 4317, respectively. This analysis led to the basis of dividing the data set only in two parts i.e., training (86 years) and testing (10 years) for all the SVM models. Dividing the data in only two sets i.e., training and testing

is a widely used practice in the SVM modeling and has been adopted in several other studies.

The robustness of the SVM model is verified by cross validation. This is done by dividing the data set into nine 10-year sub periods. The first sub period i.e. 1912-1921 is dropped and the remaining 86 years are used for training the SVM model with NAO and ENSO indices and then tested on the dropped sub period. The process is repeated for other eight sub periods and performance measures are calculated for individual periods and for the pooled values from all sub-periods. Figure 24 shows performance measures for pooled values at three gages. Comparison of performance based on R, RMSE, and E, between SVM model III best prediction (using NAO and ENSO) and pooled results shows that the performance deteriorates for pooled predictions. For example, at Cisco gage R decreases from 0.88 to 0.71, RMSE increases from 1452.38 to 1763.3, and E decreases from 0.54 to 0.38 for pooled predictions. Examination of model performance for different testing sub-periods shows that model performed reasonably well for most sub-periods, the only exception was the period 1972-1981 when performance was poor (sub-period results are not shown). Although, the model performance for pooled predictions at all three gages is lower than the performance for 1992-2001 testing period from the SVM model III (Table 7) but is within the acceptable range. The scatter plots for the pooled values show that the model is able to provide satisfactory results as the predictions are close to the bisector line.

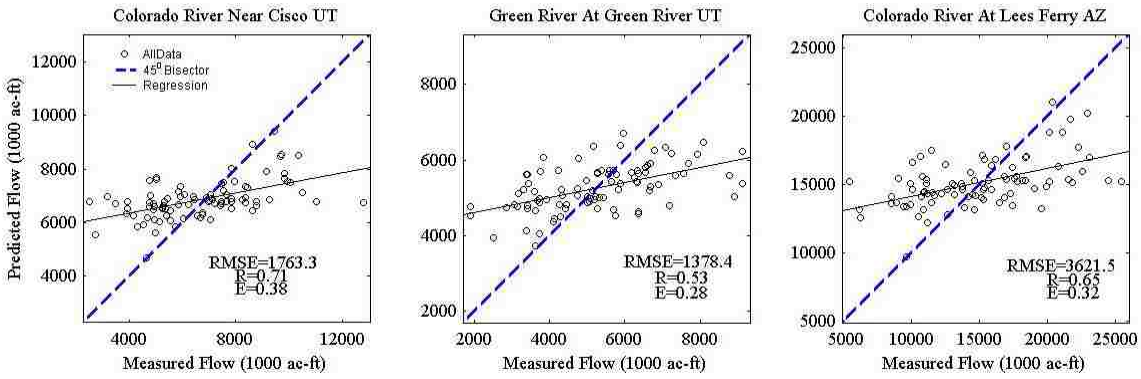


Figure 24: Scatter plot for validation of SVM model using pooled values at the selected gages.

A linear regression model is also developed using NAO and ENSO oscillations modes for 86-years. This model is used to predict streamflow for the testing period for all gages. All three performance measures for linear regression model were weaker than the ones obtained from SVM model III. The R, RMSE and E for Lees Ferry using linear regression model are 0.52, 3876.24, and 0.20, respectively compared to SVM model performance of 0.86, 3229.31, and 0.55. Similar results are obtained for other gages. Linear regression model is not able to capture the association between streamflow and oscillations modes as well as the SVM model does.

3.7.2 Comparison with ANN Models

The results obtained from the SVM models are compared with the traditional machine learning tool used in hydrology known as Artificial Neural Networks (ANNs). ANN model is developed for streamflow predictions at 't+3' time step for the four models discussed earlier. A feed forward-back propagation method with Sigmoid Activation is used in ANN to predict streamflow volumes. NueNet Pro software is used to develop the ANN model (http://www.cormactech.com/neunet/).

Figure 25 shows the correlation coefficient between the training and testing phases for gages using all four inputs. During the training phase (Figure 25 a), the predictions lie far from the bisector resulting in poor predictions in the testing phase (Figure 25 b), also evident by performance measures shown in Table 7. Similar results of lower performance measures are noticed for the other models using ANN approach (Table 7). Comparison of results shows superior performance of SVM model over the ANN model at all three gages. The superiority of SVM over the ANN modeling approach has been well established by Dibike et al. (2001), Asefa et al. (2006), Gill et al. (2006), and Liong and Sivapragasam (2002) in various fields of hydrology. Although SVM model predictions are better compared to ANN model, both modeling approaches reveal stronger signals for NAO and ENSO as compared to PDO and AMO in the Upper Colorado River Basin.

3.8 Conclusions

The application presented in this paper uses the annual averaged oceanic-atmospheric indices to predict annual streamflow volumes three years (t+3) in future for three gages in the Upper Colorado River Basin. Streamflow is used as the hydrological response variable, because streamflow is regarded as the most vital component of the hydrological cycle. We consider hydrologic variability at the regional scale to obtain better streamflow forecast instead of using the entire continental United States. Streamflow volume predictions by Model I at the selected gages, as indicated by the performance measures (Table 7), are satisfactory. The predictions improve for Model II and Model III when PDO and AMO are dropped individually and in pairs for the selected gages (Table 7). An increase in R and E, and decrease in RMSE is noted for Model II and Model III after dropping PDO and AMO. Figure 21 shows that predictions obtained by dropping PDO

and AMO separately are saturated along the 45^0 line as compared to predictions obtained by dropping NAO and ENSO which are scattered. Model IV used single input and identified that better predictions are obtained using NAO and ENSO compared to PDO and AMO for three years lead time which is evident by the three performance measures.

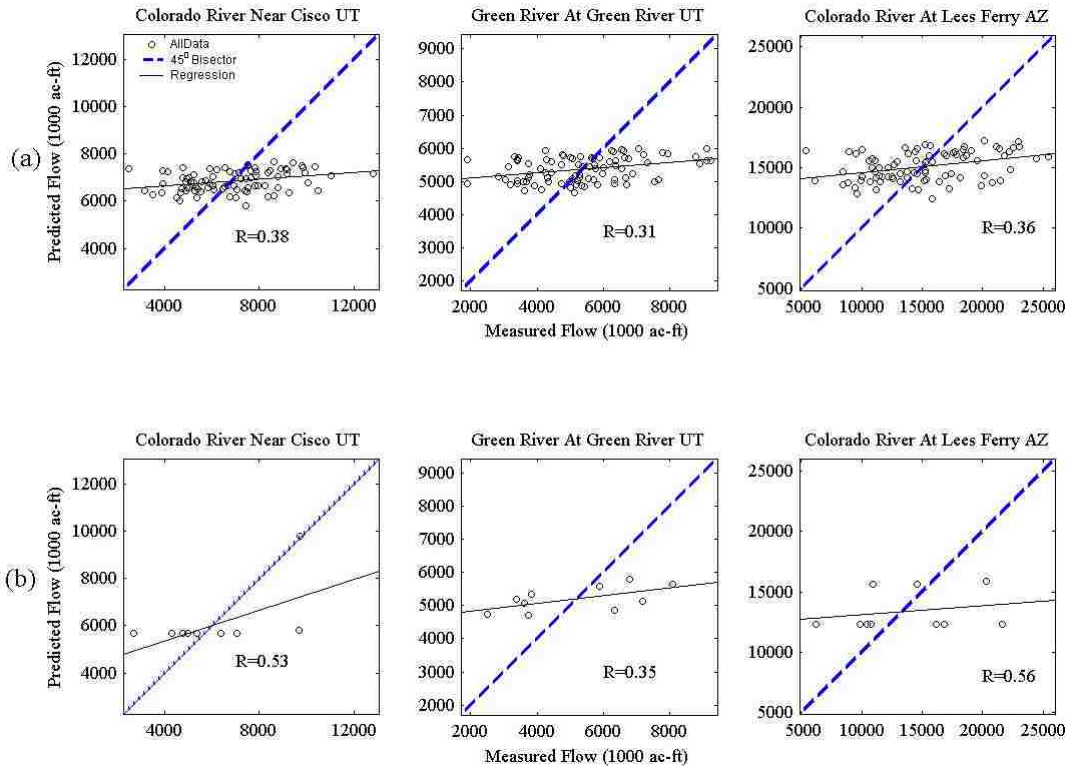


Figure 25: ANN predicted streamflow volumes at 't+3' for model I for (a) training phase and (b) testing phase at the three selected gages. Dashed line is the 45^0 bisector and solid line is true regression line between the measured and predicted streamflow volumes.

The agreement between the results from Models II, III, and IV shows that NAO and ENSO have relatively stronger signal in streamflow predictions as compared to PDO and AMO. The SVM model results are compared with the results obtained from ANN model and linear regression. In general, results from both SVM and ANN models are in agreement in showing stronger signal for NAO and ENSO compared to PDO and AMO

for all the gages. Based on all three performance measures, for all four models, and at all three gages, SVM model outperformed the ANN model and linear regression model.

The application of SVM as a forecasting tool has been shown with its implementation in the Upper Colorado River Basin. The SVM approach comprised of two parts: one part relating to the regularization of the solution; and the other to the ε-insensitive goodness of fit resulting in remarkable generalization capabilities. The SVM models belong to the class of data-driven approaches so it becomes important to determine the dominant model inputs, which helps in reducing the training time and increases the generalization.

The seasonal to annual scale relationship between ENSO and streamflow variability in the UCRB has been reported extensively. NAO has been linked with decreases in mean sea level pressures (SLP) over Artic oceans (Walsh et al., 1996), trends in North Atlantic surface wave heights (Kushnir et al., 1997), changes in storm activity, and the shifts in the Atlantic storm track (Hurrel, 1995). A linkage between NAO and streamflow variability for UCRB has not been conclusively established to date. McCabe et al. (2007) left an open ended question to the significance of AMO in predicting streamflow variability in the UCRB, but identified NAO as influencing streamflow at annual scale. Although Figure 23 shows that a relatively better prediction is possible for a 2-year lead time using a different combination of indices but performance of 3 year lead time is still satisfactory. The present study finds that long-term streamflow predictions i.e. three years for UCRB can be obtained using NAO and ENSO oscillation modes.

The results from the current research contribute to increasing the lead time up to three years for the streamflow forecasting in the UCRB, using NAO and ENSO oceanic-atmospheric indices. During model validation we learned that the SVM model did not

perform equally well for all testing sub-periods. This may be because during some testing sub-periods other oscillations modes, besides NAO and ENSO, may have been dominant. The SVM model also did not adequately capture low and high flows, which points to the fact that the indices used may not fully represent the physical processes linked with streamflow generation. Increasing the size of training data set may also improve the predictions.

Although, the model is unable to successfully capture the extreme events, the long-lead time forecast, developed in this research, would be helpful to the water managers in UCRB in managing water systems in response to inter-decadal climate variability. The research also shows prospects for the use of statistical learning theory (SVM) to predict highly complex process (streamflow) that are difficult to understand and simulate using conceptual models.

Acknowledgments

This material is based upon work supported by the National Oceanic and Atmospheric Administration (NOAA-SARP) under Award # NA07OAR4310324.

References

Ahmad, S., and S. P. Simonovic (2005), An artificial neural network model for generating hydrograph from hydro-meteorological parameters, *Journal of Hydrology*, *315* (1-4), 236-251.

Ahrens, C. D. 1994. Meteorology Today: An Introduction to Weather, Climate, and the Environment. 5th ed. West Publishing.

ASCE Task Committee (2000a), Artificial neural networks in hydrology I: preliminary concepts, *Journal of Hydrologic Engineering*, *5* (2), 115-123.

ASCE Task Committee (2000b), Artificial neural networks in hydrology II: hydrologic applications, *Journal of Hydrologic Engineering*, *5* (2), 124-137.

Asefa, T., M. Kemblowski, M. McKee, and A. Khalil (2006), Multi-time scale stream flow predictions: The support vector machines approach, *Journal of Hydrology*, *318*, 7-16.

Beebee, R. A., and M. Manga (2004), Variation in the relationship between snowmelt runoff in Oregon and ENSO and PDO*, Journal of American Water Resources Association*, *40* (4), 1011-1024.

Cayan, D., and R. Webb (1992), El Niño/Southern Oscillation and streamflow in the western United States Pages 29-68 *In El Niño*, H. F. Diaz and V. Markgraf, editors. Cambridge Univ. Press, New York.

Chang, F. J., and Y. C. Chen (2001), A counterpropagation fuzzy-neural network modeling approach to real time streamflow prediction, *Journal of Hydrology*, *245*, 153-164.

Cherkassky, V., and Y. Ma (2004), Practical selection of SVM parameters and noise estimation for SVM regression, *Neural Networks*, *17*, 113-126.

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

Cristianini, N., and J. Shaw-Taylor (2000), An introduction to support vector machines and other kernel based learning methods, Cambridge University Press, Cambridge, Massachusetts,

Dettinger, M. D., 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*, 3095-4111.

Dibike, Y. B., S. Velickov, D. Solomatine, and M. B. Abbott (2001), Model induction with support vector machines: Introduction and application, *Journal of Computing in Civil Engineering*, *15* (3), 208-216.

El-Ashry, M., and D. Gibbons, (Eds) (1988), Water and Arid Lands of the Western United States. Cambridge University Press, New York.

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 U.S., *Geophysical Research Letters*, *28*, 2077-2080.

Gershunov, A., and T. P. Barnett (1998), Interdecadal modulation of ENSO telecommunications, *Bulletin of the American Meteorological Society*, *79*, 2715-2726.

Gill, M. K., T. Asefa, M. Kemblowski, and M. McKee (2006), Soil moisture prediction using support vector machines, *Journal of American Water Resources Association*, *42* (4), 1033-1046.

Grantz, K., B. Rajagopalan, M. P. Clark, and E. A. Zagona (2005), A technique for incorporating large-scale climate information in basin-scale ensemble streamflow forecasts, *Water Resources Research*, *41*, W10410, doi:10.1029/2004WR003467.

Gray, S. T., J. L. Betancourt, C. L. Fastie, and S. T. Jackson (2003), Patterns and Sources of Multidecadal Oscillations in Drought-Sensitive Tree-Ring Records from the Central and Southern Rocky Mountains., *Geophysical Research Letters*, *31*, L12205.

Groisman, P. Y., R. W. Knight, and T. R. Karl (2001), Heavy precipitation and high streamflow in the contiguous United States: Trends in the twentieth century, *Bulletin of the American Meteorological Society*, *82* (2), 219-246.

Gutierrez, F., and J. A. Dracup (2001), An analysis of the feasibility of long-range streamflow forecasting for Colombia using El Niño-Southern Oscillation indicators, *Journal of Hydrology*, *246*, 181-196.

Haltiner, J. P., and J. D. Salas (1988), Short term forecasting of snowmelt runoff using ARMAX models, *Water Resources Bulletin*, *24* (5), 1083-1089.

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*, *125* (6), 333-341.

Haykin, S. 2003. Neural Networks: A comprehensive foundation, Fourth Indian Reprint edition. Pearson Education, Singapore.

Hipel, K. W. (1985), Time series analysis in perspective, *Water Resources Bulletin*, *21* (4), 609-624.

Holbrook, S. J., R. J. Schmitt, and J. S. Stephens (1997), Changes in an assemblage of temperature reef fishes associated with a climate shift, *Ecological Applications*, *7*, 1299-1310.

Hsu, K.-l., H. V. Gupta, and S. Sorooshian (1995), Artificial neural network modeling of the rainfall-runoff process, *Water Resources Research*, *31* (10), 2517-2530.

Hurrell, J. W. (1995), Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation, *Science*, *269* (5224), 676-679.

Hurrell, J. W., and H. Van Loon (1995), Decadal variations in climate associated with the North Atlantic Oscillation, *Climate Change*, *31*, 301-326.

Imrie, C. E., S. Durucan, and A. Korre (2000), River flow prediction using artificial neural networks: generalization beyond the calibration range, *Journal of Hydrology*, *233*, 138-153.

Irvine, K. N., and A. J. Eberhardt (1992), Multiplicative, seasonal ARIMA models for Lake Erie and Lake Ontario water levels, *Water Resources Bulletin*, *28* (2), 385-396.

Kahya, E., and J. A. Dracup (1993), U.S. streamflow patterns in relation to the El Niño/southern oscillation, *Water Resources Research*, *29* (8), 2941-2503.

Kalra, A., T. C. Piechota, R. Davies, and G. A. Tootle (2008), Changes in U.S. streamflow and Western U.S. snowpack, *Journal of Hydrologic Engineering*, *13* (3), 156-163.

Khalil, A. F., M. McKee, M. Kemblowski, T. Asefa, and L. Bastidas (2006), Multiobjective analysis of chaotic dynamic systems with sparse learning machines, *Advances in Water Resources*, *29*, 72-88.

Kraijenhoff, D. A., and J. R. Moll (1986), River flow modelling and forecasting, Dordrecht, Netherlands.

Kushnir, Y., V. J. Cardone, J. G. Greenwood, and M. A. Cane (1997), The recent increase in the North Atlantic wave heights, *Journal of Climate*, *10*, 2107-2113.

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

Liong, S.-Y., and C. Sivapragasam (2002), Flood stage forecasting with support vector machines, *Journal of American Water Resources Association*, *38* (1), 173-186.

Maier, H. R., and G. C. Dandy (2000), Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications, *Environmental Modelling & Software*, *15*, 101-124.

Mantua, N. J., and S. R. Hare (2002), The Pacific Decadal Oscillation, *Journal of Oceanography*, *59* (1), 35-44.

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* (1), 183-192.

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.

McCabe, G. J., M. A. Palecki, and J. L. Betancourt (2004), Pacific and Atlantic Ocean Influences on multidecadal drought frequency in the United States, *Proceedings of the National Academy of Sciences*, *101* (12), 4136-4141.

McCabe, G. J., and D. M. Wolock (2002), Is streamflow increasing in the conterminous United States?, *In* Proceedings of PACLIM, Pacific Grove, CA, pp 93-98.

Piechota, T. C., J. A. Dracup, and R. G. Fovell (1997), Western U.S. Streamflow and Atmospheric Circulation Patterns El Niño-Southern Oscillation (ENSO), *Journal of Hydrology*, *201* (1-4), 249-271.

Piechota, T. C., J. Timilsena, G. A. Tootle, and H. Hugo (2004), The Western U.S. Drought: How Bad Is It?, *Eos, Transactions, American Geophysical Union*, *85* (32), 301-308.

Raman, H., and N. Sunilkumar (1995), Multivariate modelling of water resources time series using artificial neural networks, *Journal of Hydrological Sciences*, *40* (2), 145-163.

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 indices, *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.

Rodriguez-Iturbe, I., and J. B. Valdes (1979), The geomorphologic structure of hydrologic response, *Water Resources Research*, *15* (6), 1409-1420.

Rogers, J. C., and J. S. M. Coleman (2003), Interactions between the atlantic multidecadal oscillations, El Niño/La Niña, and the PNA in winter Mississippi Valley stream flow, *Geophysical Research Letters*, *30* (10).

Ropelewski, C. F., and P. D. Jones (1987), An Extension of the Tahiti-Darwin Southern Oscillation Index, *American Meteorological Society*, *115* (9), 2161-2165.

Schölkopf, B., K.-K. Sung, C. J. C. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik (1997), Comparing Support Vector Machines with Gaussian Kernels to Radial Basis Function Classifiers, *IEEE Transactions on Signal Processing*, *45* (11), 2758-2765.

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

Slack, J. R., J. M. Landwehr, and A. Lumb (1992), A U.S. Geologic Survey streamflow data set for the United States for the study of climate variations, 1974-1988, *Rep.* 92-129, U.S Geol. Surv., Oakley, Utah.

Smola, A. J., B. Schölkopf, and K.-R. Muller (1998), The connection between regularization operators and support vector kernels, *Neural Networks*, *11*, 637-649.

Tang, Z., C. DeAlmeida, and P. A. Fishwick (1991), Time series forecasting using neural networks vs. box-jenkins methodology, *Simulations*, *57* (5), 303-310.

Thompstone, R. M., J. P. Hipel, and A. I. McLeod (1985), Forecasting quarter-monthly riverflow, *Water Resources Bulletin*, *21* (5), 731-741.

Tingsanchali, T., and M. R. Gautam (2000), Application of tank, NAM, ARMA, and neural network models to flood forecasting, *Hydrological Processes*, *14* (4), 2473-2487.

Tokar, A. S., and M. Markus (2000), Precipitation-Runoff modeling using artificial neural networks and conceptual models, *Journal of Hydrologic Engineering*, *5* (2), 156-161.

Tootle, G. A., and T. C. Piechota (2006), Climate Variability, Water Supply, and Drought in the Upper Colorado River Basin Pages 132-142 *In Climate Variations, Climate Change, and Water Resources Engineering*, J. D. Garbrecht and T. C. Piechota, editors. American Society of Civil Engineers.

Tootle, G. A., T. C. Piechota, and A. Singh (2005), Coupled oceanic-atmospheric variability and U.S. streamflow, *Water Resources Research*, *41*, W12408.

Tripathi, S., V. V. Srinivas, and R. S. Nanjundiah (2006), Downscaling of precipitation for climate change scenarios: A support vector machine approach, *Journal of Hydrology*, *330*, 621-640.

Turing, A. M. (1950), Computing Machinery and Intelligence, *Mind 59*, *236*, 433-460.

Vapnik, V. (1995), The nature of Statistical Learning Theory, Springer, New York.

Vapnik, V. (1998), Statistical Learning Theory, Wiley, New York.

Vapnik, V., and V. Chervonenkis (1971), On the uniform convergence of relative frequencies of events to their probabilities, *Theory of Probability and its Applications*, *16* (2), 264-280.

Walsh, J. E., W. L. Chapman, and T. Shy, L. (1996), Recent decrease of sea level pressure in the Central Arctic, *Journal of Climate*, *9*, 480-486.

Yapo, P., H. V. Gupta, and S. Sorooshian (1996), Calibration of conceptual rainfall-runoff models: sensitivity to calibration data, *Journal of Hydrology*, *181*, 23-28.

Yu, X., and S.-Y. Liong (2007), Forecasting of hydrologic time series with ridge regression in feature space, *Journal of Hydrology*, *332*, 290-302.

Zealand, C. M., D. H. Burn, and S. P. Simonovic (1999), Short term streamflow forecasting using artificial neural networks, *Journal of Hydrology*, *214*, 32-48.

CHAPTER 4

EVALUATING CHANGES AND ESTIMATING SEASONAL PRECIPITATION FOR COLORADO RIVER BASIN USING STOCHASTIC NONPARAMETRIC DISAGGREGATION TECHNIQUE (PUBLISHED IN WATER RESOURCES RESEARCH)

Abstract

Precipitation estimation is an important and challenging task in hydrology due to high variability and changing climate. This research involves i) analyzing changes (trend and step) in seasonal precipitation, and ii) estimating seasonal precipitation by disaggregating water year precipitation using k-nearest neighbor (KNN) nonparametric technique for 29 climate divisions encompassing Colorado River Basin. Water year precipitation data from 1900 to 2008 is subdivided into four seasons' i.e. autumn, winter, spring, and summer. Two statistical tests i.e., Mann-Kendal and Spearman's Rho are used to evaluate trend changes and Rank Sum test is used to identify the step change in seasonal precipitation. The results indicate a decrease in the upper basin and increase in the lower basin winter precipitation resulting due to an abrupt step change. The affect of El Niño-Southern Oscillations (ENSO) in relation to seasonal precipitation is also evaluated by removing the historic El Niño events. Decreasing winter and spring season precipitation trends for the upper basin are not linked with El Niño. Corroborating evidence of changes in seasonal precipitation is established by analyzing the trends in SNOTEL data and streamflow at Lees Ferry gage. KNN disaggregation results indicate satisfactory seasonal precipitation estimates during winter and spring season compared to autumn and summer season and the superiority of KNN results is established when compared with the first

order periodic autoregressive (PAR-1) parametric approach. The analysis of seasonal changes and estimates of precipitation can help water managers in better management of water resources in the Colorado River Basin.

Citation: Kalra, A. and S. Ahmad (2011), Evaluating changes and estimating seasonal precipitation for the Colorado River Basin using a stochastic nonparametric disaggregation technique, *Water Resour. Res.*, doi:10.1029/2010WR009118.

4.1 Introduction

4.1.1 Background

The need for better information about the variability exhibited by precipitation has increased due to the changing climate (New et al., 2001; Karl and Knight, 1997; Regonda et al., 2005; Block and Rajagopalan, 2007; and Hansen et al., 2006). Large scale changes in precipitation due to the changing climate have caused several catastrophic flood and drought events globally. These changes have caused large scale destruction both to the nature and mankind. A few examples of some catastrophic flooding events are the 1993 flooding along the Mississippi River, the 1996 autumn floods in the New England, the winter floods of 1997 in the Pacific Northwest and California, and the Ohio River and the Red River valley floods during the spring of 1997 (Karl and Knight, 1997). Similar to floods, a few notable drought events are the 1995-1996 droughts in the Upper Midwest and the Ohio Valley, the 1991 drought of California (Tarboton, 1994; Lins and Slack, 1999), and the severe sustained drought within the Colorado River Basin (CRB) since 2000 (Piechota et al., 2004). The increased variability in precipitation has forced water managers to develop plans to mitigate these climate extremes. This requires evaluating

the long-term changes in precipitation and estimating precipitation using statistical techniques.

4.1.2 Changes in Precipitation

It would be remiss not to mention some notable work that has emphasized the changes in precipitation (Bunting et al., 1976; Hennessy et al., 1999; Gonzalez-Hidalgo et al., 2001, Luis et al., 2000; Timbal, 2004). Bunting et al (1976) used linear regression to evaluate the trends in the long-term rainfall records for West Africa and indicated that the trends can be used to forecast rainfall one to two seasons ahead. Timbal (2004) used statistical technique based on synoptic situation to study observed rainfall trends in Southwest Australia. He was able to reproduce the trends observed globally during the past 50-years, indicating the sensitivity of the statistical approach to the changes in climate conditions. Hennessy et al. (1999) used Kendall-Tau test to compute trends in seasonal and total rainfall for 379 stations spread over the entire Australia from 1910-1995 and compared the results using linear regression test. They concluded that changes are significant for total rainfall but show non-significant changes during different seasons. Gonzalez-Hidalgo et al. (2001) used spearman's rho and linear regression to evaluate rainfall trends in western Mediterranean area for 97 pluviometric stations for a period from 1960-1990. The results indicated an increase in winter and summer interannual precipitation. Luis et al. (2000) used nonparametric Mann-Kendal and Spearman's Rho tests to evaluate the trends in rainfall for 97 rain-gauge stations in the region of Valencia (East Spain). They observed a decrease in annual rainfall and showed a significant increase in the interannual variability. In Canada and the United States, an increasing trend in the annual totals of precipitation and decreasing trend in lower-

latitude precipitation have been observed during the twentieth century (Groisman and Easterling, 1994; Groisman et al., 2001).

Although, there is no single method of analysis that can comprehensively cover all aspects of changes in precipitation, it is fairly apparent that more consideration should be given to the type of question an analysis can answer. Majority of studies analyzing the changes in precipitation used nonparametric statistical trend tests such as the Mann-Kendall and Spearman's Rho but did not account for an abrupt step change in the precipitation. It is important to clearly differentiate between a gradual trend and a step change for climate change studies. This is necessary because the pattern of the trend change can be linear and continuous, whereas step changes are non linear, occur abruptly, and may reoccur in the future (McCabe and Wolock, 2002, Kalra et al., 2008). It is well documented that rapid climatic changes were noted during the winter of 1976-1977 in the North Pacific region due to the shift in the ocean-atmosphere system (Kerr, 1992; Beamish et al., 1997; Holbrook et al., 1997; Mantua and Hare, 2002). These oceanic changes intensified the weather in the sub-Arctic Pacific, which affected the sea surface temperatures. Variations (i.e., increase and decrease in sea surface temperatures) were noted for the Eastern Pacific and Central Pacific regions (Kerr, 1992; Beamish et al., 1997). This step-like shift in the mean sea level temperatures has been termed as the "climatic regime", following a regime shift in 1977 (Mantua and Hare, 2002). Due to this reason, evaluating a step change in the precipitation becomes important and knowing these shifts in advance can help water managers to improve reservoir operations to meet the competing demands for municipal use, irrigation, environmental need, and power generation (Regonda et al., 2005).

4.1.3 Need for Precipitation Estimation

Precipitation is regarded as a vital governing factor in the temporal and spatial variability of runoff and soil moisture, which in turn control the entire hydrologic regime of a river basin. Accurate estimate of precipitation are often necessary for monitoring the variability in climate extremes and are helpful in understanding the hydrological cycle (Bell, 1987; Hsu et al., 1997; Nayak et al., 2008, 2010). Accessing high resolution temporal precipitation data is of prime importance in a multitude of hydrologic applications (Olsson, 1998; Guntner et al., 2001; Sivakumar, et al., 2001; Srikanthan and McMahon, 2001; Koutsoyiannis, 2003). Increasing climate variability has shifted the focus of different scientific communities i.e. hydrological, meteorological, and agricultural to deal with complex problems, such as pollution transport, rainfall-related pollution effects on plants, soil water infiltration, and soil erosion (Guntner et al., 2001; Sivakumar, et al., 2001). If precipitation data is not available at required spatial and temporal scale, it results in additional uncertainties. A solution to this problem is to collect high-resolution data relevant to the problem but it is costly and time consuming. General Circulation Models (GCMs) are normally used to forecast future weather conditions at global and regional scales. These models generate rainfall data at a very course spatial (in the order of 250 km^2 or greater) resolution (Chiew et al., 2010). In most of the watershed scale modeling work, input data is required at much finer spatial (in the order of 100 m2 grids) resolution and temporal resolution varies upon the need of end user. The only possible alternative that is simple and parsimonious is to transform the available data from one time scale to another. If precipitation data at aggregate scale is available from statistical models or GCMs, disaggregation techniques can be used to

estimate precipitation at finer temporal resolution. Stochastic disaggregation techniques are often necessary for reproducing the right statistical characteristics of the data, at the required time scale, because the disaggregated series is a realization from the original aggregated time series (Mehrotra and Sharma, 2006). The stochastic disaggregation techniques help in establishing long range estimates from the historic data and generate synthetic values not seen in the historic records and also preserve the statistical properties such as mean, median, standard deviation, and skewness.

4.1.4 Disaggregation Applications in Hydrology

The first linear stochastic disaggregation model was developed by Valencia and Schaake (1973) to disaggregate streamflow. The authors aggregated flow using a linear model and then disaggregated it using the linear stochastic model. The model by Valencia and Schaake (1973) was modified and improved by several researches (Mejia and Rousselle, 1976; Lane, 1982; Stedinger and Vogel, 1984). Besides linear stochastic disaggregation techniques, there have been alternate approaches, which allow representation of the non-Gaussian data directly in the disaggregation procedure (Tao and Delleur, 1976; Todini, 1980; Koutsoyiannis, 1992; Koutsoyiannis, 2001). These approaches do not require data transformation and preserve the additive property and higher order statistics of the aggregated and disaggregated data by performing a stepwise disaggregation. The major disadvantage of such techniques is that they assume linearity in the data and are iterative in nature, which makes them computationally intensive.

To overcome the issues in disaggregation procedures, Lall and Sharma (1996) developed a nonparametric bootstrap approach of time series simulations based on the kernel nearest neighbor (KNN). They showed that the synthetic streamflow series

generated from KNN is better than that from ARMA models, and the KNN technique is more flexible and is capable of reproducing both linear and nonlinear dependences. The KNN method is preferred where the researchers are uncomfortable with the prior assumption about the data (e.g. linear or nonlinear). Sharma et al. (1997) used KNN to simulate streamflow at a single site and showed the advantages of using KNN over traditional linear models. Tarboton et al. (1998) developed a kernel based temporal streamflow disaggregation approach for multiple sites. This work was an extension of the single site work by Sharma et al. (1997). Srinivas and Srinivasan (2005) developed a semi parametric disaggregation multisite model called the hybrid moving block bootstrap multisite model (HMM). This method incorporated the merits of both parametric and nonparametric techniques but still required multiple steps, which were computationally intensive. Prairie et al. (2005, 2006, and 2007) modified the KNN disaggregation procedure developed by Lall and Sharma (1996) for streamflow simulations and disaggregated streamflow both spatially and temporally at multiple sites. They were able to generate values not observed in the historic data using a modified bootstrap KNN approach.

Compared to streamflow disaggregation, precipitation disaggregation has greater challenges due to its intermittence characteristics and the lack of gaussianity (Guenni and Bardossy, 2002). Furthermore, the required temporal resolution of precipitation depends on the purpose for which disaggregation will be performed. Precipitation values generated through GCMs may not be directly useable for some application but can be used indirectly in disaggregation schemes. Studies evaluating long-term climate changes, critical crop production decisions and sediment yield within catchments require seasonal

data, whereas estimation of water demand and the simulation of water supply generally needs monthly data (Srikanthan and McMahon, 2001; Hansen et al., 2006). Hydrological applications that control surface and subsurface processes require daily or hourly precipitation data (Rajagopalan et al., 1997; Rupp et al., 2009). Due to the continued need of high-resolution precipitation data several precipitation disaggregation techniques have been used by numerous researchers for transforming the data from one scale to the other (Grace and Eagleson, 1966; Schaake et al., 1972; Woolhiser and Osborn, 1985; Hershenhorn and Woolhiser, 1987; Arnold and Williams, 1989; Econopouly et al., 1990; Bo et al., 1994; Connolly et al., 1998; Olsson, 1998; Olsson and Berndtsson, 1998; Koutsoyiannis, 1988, Koutsoyiannis and Xanthopoulos, 1990; Rajagopalan and Lall, 1999; Harrold et al., 2003; Gangopadhyay et al., 2005; Block and Rajagopalan, 2007). Grace and Eagleston (1966) disaggregated storm depth for shorter duration using a 2-dimensional overland flow model. A Markov chain model for disaggregating monthly rainfall into daily value was proposed by Schaake et al. (1972). A nondimensionalised Markov process for disaggregating storm depth into fractional depths was developed by Woolhiser and Osborn (1985). Hershenhorn and Woolhiser (1987) developed a daily stochastic model to disaggregate rainfall into a number of individual storms in a day. The model did not address the external storm structure and was later modified by Econopouly et al. (1990). A simple stochastic model for generating half-hourly rainfall intensity from daily rainfall totals was proposed by Arnold and Williams (1989). Koutsoyiannis (1988) and Koutsoyiannis and Xanthopoulos (1990) proposed general methods for disaggregating rainfall for time scales finer than monthly. Connolly et al. (1998) proposed a stochastic model for disaggregating daily rainfall totals into multiple storm

events in a day and accounted for the time-varying intensity within each event. Bartlett-Lewis rectangular pulse (BLRP) model developed by Rodriguez-Iturbe et al. (1987) was used by Bo et al. (1994) to disaggregate daily rainfall values into hourly values. Olsson (1988) used the multifractal simulation techniques to model the temporal structure of rainfall. Later, Olsson and Berndtsson (1998) improved on the method and employed a scaling-based cascade model to disaggregate daily seasonal values to 45-minute temporal resolution. Burian et al. (2000) used Artificial Neural Network (ANN) approach to disaggregate hourly rainfall data into shorter time units, based on theory of learning. Sivakumar et al. (2001) used a chaotic approach to disaggregate rainfall of five simulations using the correlation dimension technique to verify the assumption of chaos at the Leaf River Basin in the state of Mississippi. Harrold et al. (2003) used KNN to generate rainfall amounts of wet days, which takes into account the seasonality and reproduces important distributional and dependence properties of observed rainfall. Lall et al. (1996) used a wet/dry KNN model to resample daily precipitation using kernel probability density estimators. Rajagopalan and Lall (1999) used KNN to simulate daily precipitation and other weather variables and compared the results with the multivariate autoregressive model (MAR-1)). The results indicated a better performance of KNN compared to results from the MAR-1 model. Gangopadhyay et al. (2005) used KNN to downscale precipitation and temperature from a weather prediction model for four diverse basins across the conterminous United States. The authors compared the KNN downscaling model with multiple linear regression based downscaling model and showed the superiority of KNN. Tripathi et al. (2006) applied Support Vector Machine (SVM) to downscale GCM output to monthly time scale for operation purposes in India. Block and

Rajagopalan (2007) evaluated the interannual variability and ensemble seasonal forecast of Upper Blue Nile Basin Kiremt in Ethiopia using stochastic approach based on K-nearest neighbor. The authors used the Prairie et al. (2007) disaggregation technique to generate monthly precipitation values from aggregated seasonal precipitation. Interested readers are referred to Koutsoyiannis (2003) for detailed description and applications of different precipitation disaggregation techniques.

Review of disaggregation literature shows that KNN has been used extensively to disaggregate streamflow, whereas KNN based precipitation disaggregation studies are less common and deal primarily with simulating rainfall at hourly, daily, weekly, and few at monthly time scales. Seasonal estimations reveal the average conditions over a period of time and are not limited to a particular day. Seasonal precipitation disaggregation estimates are necessary for the hydrologic, meteorological, and agricultural communities (Singhrattna et al., 2005; Hansen et al., 2006; Robertson et al., 2007). Estimating seasonal precipitation values is important for resource planning and management e.g. reservoir management, agricultural practices, and flood management (Bindlish and Barros, 2000; Rupp et al., 2009). Seasonal precipitation values are useful in determining the variations in crop productions and can help in adjusting critical decisions (Hansen et al., 2006). Understanding the year to year variability in seasonal precipitation is helpful in mitigating potential disasters (Block and Rajagopalan, 2007). Although, seasonal predictions have been made with generalized circulation models (GCMs), they are often useful for some regions and during particular seasons (Goddard et al., 2003).

Based on the extensive literature review on disaggregation and per author's knowledge, it is noteworthy that none of the precipitation disaggregation studies have

been within the Colorado River Basin (CRB). It will be true to state that there is no river governed more physically and legislatively than the Colorado River (Christensen and Lettenmaier, 2007). Climate change is a major concern in the CRB due to sensitivity of discharge to precipitation and temperature (Christensen and Lettenmaier, 2007, Miller and Piechota, 2008; Kalra and Ahmad, 2009). Regonda et al. (2005) have indicated that climate change may lead to the intensification of different hydrological processes, and may affect the nature of precipitation events within the CRB. Due to this, there has been an increased emphasis on the drought and water availability studies of higher temporal scales within CRB (Hamlet et al., 2005; Mote et al., 2005; Christensen and Lettenmaier, 2007; Pagano and Green, 2005; Easterling et al., 2007).

4.1.5 Motivation of Current Research

To manage available water and analyze drought conditions, there is a need to evaluate the long-term changes in precipitation and provide seasonal precipitation estimates within the Colorado River Basin. With this motivation, the study presented here evaluates both the trend and step changes in seasonal precipitation over 29 Climatic Divisions within the Colorado River Basin over a 109 year time span (1900 – 2008). Nonparametric statistical tests Mann-Kendall and Spearman's Rho are used to evaluate trends in data and nonparametric Rank Sum is used to evaluate the step change. The changes are evaluated for four seasons i.e., autumn, winter, spring, and summer. The seasons were selected in such a manner that the water year can be divided into four seasons and the affect of each season can be analyzed separately. The duration of the seasons are explained in the data section. Miller and Piechota (2008) have evaluated trend and step changes in precipitation within the CRB using the nonparametric tests similar to the current study

using the monthly data from 1951-2005. The current study evaluates seasonal precipitation changes for the entire twentieth century i.e. 1900-2008. Also the effect of El Niño-Southern Oscillations (ENSO) on seasonal precipitation within the CRB is evaluated. In the past, ENSO has been strongly associated with climate fluctuations within the arid and semiarid Southwest, which makes it vital in evaluating changes in precipitation for efficient water resources planning and management within a watershed.

In the western U.S. including the Colorado River Basin, 50%-70% of annual precipitation falls as snow (Seereze et al., 1999), which is largely stored through the winter season. April 1 Snow Water Equivalent (SWE) is usually used to provide estimates and forecast of the annual runoff and is critical in the management of reservoirs and irrigation practices (McCabe and Dettinger, 2002). To verify the findings of the changes in seasonal precipitation in the current study, trends in March 1, April 1, and May 1 SWE data from 1961-2006 (46-years) are also analyzed for snowpack stations within the Colorado River Basin. Moreover, seasonal changes in streamflow for Lees Ferry gage are analyzed from 1922-2009 (88-years) to understand the relationship between upper basin precipitation and streamflow.

Along with evaluating changes in seasonal precipitation, the current research also disaggregates water year precipitation totals into four seasonal values based on nonparametric KNN disaggregation technique. Seasonal precipitation estimates are useful for river basin management and reservoir operation. Seasonal precipitation estimates within the CRB are also vital for paleoclimatic studies, which are used in developing long-duration climate proxies. Tree ring widths are sensitive to precipitation and precipitation that does not impact tree growth is not reflected in the tree ring widths and

cannot be used to reconstruct the past hydrology (Stockton and Jacoby, 1976; Tarboton, 1994). The disaggregation approach used in the current research will be useful in evaluating the seasonal variability exhibited by precipitation within the Colorado River Basin. The results from the KNN approach are compared with the first order periodic autoregressive model (PAR-1) that has been widely used in practice.

The findings of the current research will help in understanding the temporal (109 years) and spatial (location and number of climate divisions) trends in seasonal precipitation that will be useful for water resources planning and management in the Colorado River Basin to meet competing urban, agricultural, environmental, and power generation needs. Secondly, disaggregating water year precipitation into seasonal values will be helpful for paleoclimatic studies for reconstructing the past hydrology (i.e. streamflow). The information available through reconstructions can be used to evaluate the long-term hydrological variability within the Colorado River Basin, which is critical for the effective management of surface water resources.

The paper is organized as follows. The study region and the data used are described in sections 4.2 and 4.3, respectively. In section 4.4, the proposed methods to evaluate changes in the seasonal precipitation along with the KNN disaggregation technique are presented. Section 4.5 covers the trend and step changes and KNN disaggregation results. Section 4.6 summarizes and concludes the paper.

4.2 Study Region

The Colorado River is a major source of water supply to the southwestern United States. The water from the Colorado River is allocated to seven states (California, Nevada, Utah, Arizona, Colorado, Wyoming, New Mexico) within the Colorado River

Basin based on the "Law of River" (Sax et al., 2000). Due to growing population and agricultural activity, certain states such as California depend on water surpluses from the Colorado River. The Colorado River basin is composed of upper and lower basin (Figure 26). The Upper Colorado River Basin generates 90% of the Colorado River flow from spring-summer runoff due to snowmelt. The UCRB is defined as the part of basin upstream of the gage at Lees Ferry (shown as triangle in Figure 26) and just downstream of Glen Canyon Dam in Northern Arizona. It serves Wyoming, Colorado, Utah, and New Mexico. It encompasses a total area of 286,000 km^2 and is comprised of mountains, agricultural, and low-density developments. The streamflow in the UCRB is allocated and regulated on the assumption of negligible changes in the mean and higher moment's statistical distribution of annual and decadal inflow to Lake Powell and Lake Mead (McCabe et al., 2007). This is because Lake Powell and Lake Mead represent 85% of the storage capacity of the entire Colorado River Basin. The lower basin is downstream of Lees Ferry and serves California, Nevada, and Arizona. The supply to lower basin is governed by the water released from the upper basin. In general, the LCRB is a semi-arid region with primarily mixed vegetation and bare soil landcover types.

The conterminous US is divided into 344 climate division out of which 29 climate divisions encompass the Colorado River Basin (Figure 26). The Climate Divisions are structured along county lines, drainage basins, and in some instances reflect the economic and political boundaries defined by the National Climate Data Center (NCDC). The climate divisions are intended to be useful for agricultural, irrigation, transportation, forestry, and engineering communities. For the purpose of this study, the climate

divisions have been sorted according to States and have been numbered from 1-29. Table 8 show the nomenclature used to identify each climate division within a particular state.

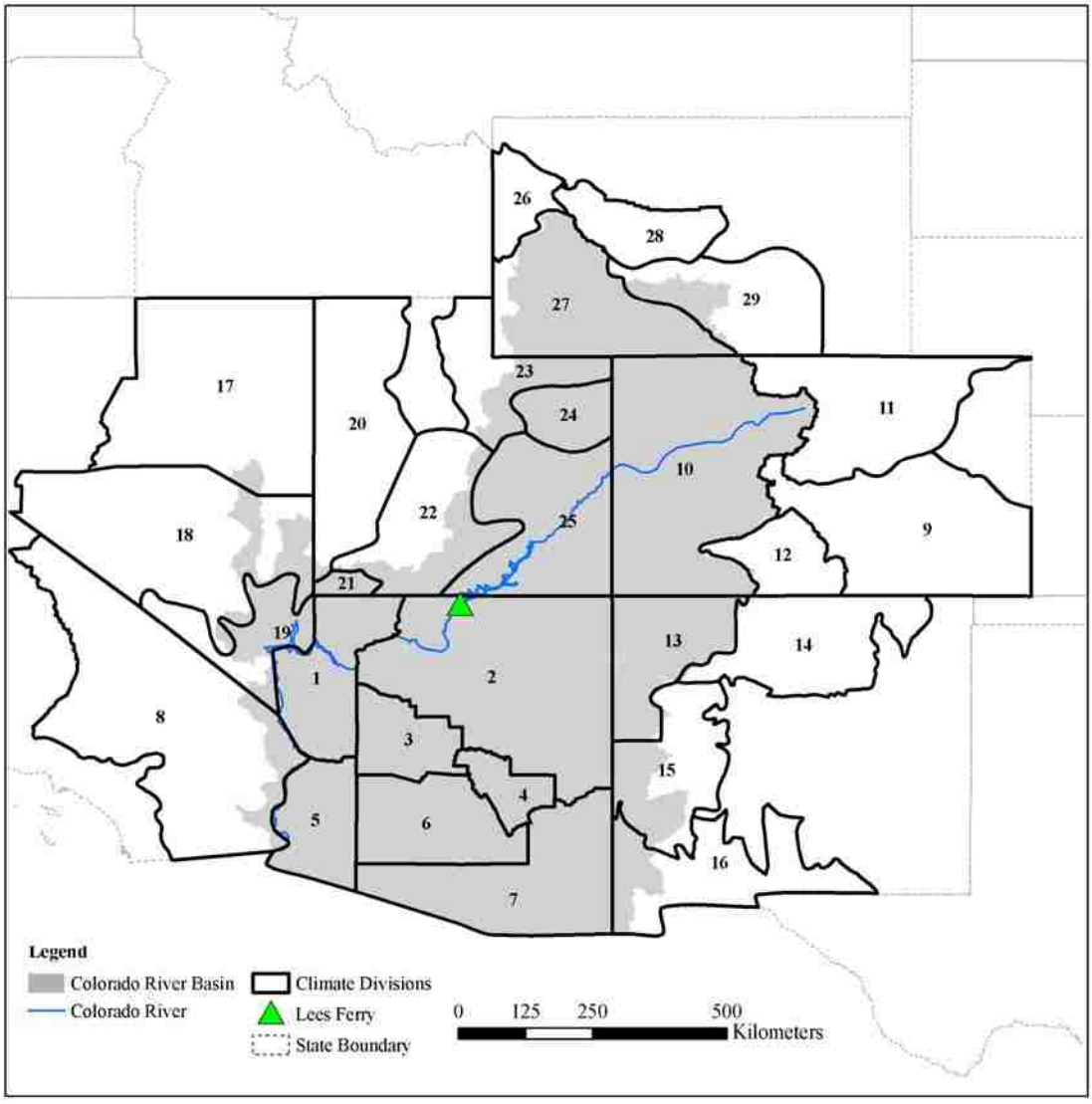


Figure 26: Map showing Colorado River Basin and 29 climate divisions.

Table 8: List of climate divisions used in the study.

Climate Division (CD)	State	NAME
1	AZ	NORTHWEST
2	AZ	NORTHEAST
3	AZ	NORTH CENTRAL
4	AZ	EAST CENTRAL
5	AZ	SOUTHWEST
6	AZ	SOUTH CENTRAL
7	AZ	SOUTHEAST
8	CA	SOUTHEAST DESERT BASINS
9	CO	AR DRAINAGE BASIN
10	CO	CO DRAINAGE BASIN
11	CO	PLATTE DRAINAGE BASIN
12	CO	RIO GRANDE DRNG. BASIN
13	NM	NORTHWESTERN PLATEAU
14	NM	NORTHERN MOUNTAINS
15	NM	SOUTHWESTERN MOUNTAINS
16	NM	SOUTHERN DESERT
17	NV	NORTHEASTERN
18	NV	SOUTH CENTRAL
19	NV	EXTREME SOUTHERN
20	UT	WESTERN
21	UT	DIXIE
22	UT	SOUTH CENTRAL
23	UT	NORTHERN MOUNTAINS
24	UT	UINTA BASIN
25	UT	SOUTHEAST
26	WY	SNAKE DRAINAGE
27	WY	GREEN AND BEAR DRAINAGE
28	WY	WIND RIVER
29	WY	UPPER PLATTE

4.3 Data

The data sets used in the current study comprise of the average monthly precipitation time series (inch), Snow Water Equivalent (inch), and streamflow (acre-ft) data for Lees Ferry. All the data sets are described in the ensuing sections:

4.3.1 Precipitation

The precipitation data used in this study is the averaged monthly time series (inch) data for 29 climate divisions covering a period from 1900-2008 and is obtained from the National Climate Data Center http://www.esrl.noaa.gov/psd/cgi-

bin/data/timeseries/timeseries1.pl. The precipitation data over each climate division is derived by taking an average of each station reported from the National Weather Service (NWS) Cooperative Observer Program (COOP) within that division. The count and distribution of the stations within COOP have changed over time and may not be representative of topographical impacts of climate within a division. This may be considered a limitation in the dataset, but the data corresponds well to large-scale historic climate anomalies such as drought, both spatially and temporally (Guttman and Quayle, 1996). Since the availability of the data, the data has been subject to changes and revisions. The latest significant change occurred in late 1960s.

The monthly precipitations data is summed to obtain precipitation for water year (October of the previous year to September of the current year) and the four seasons i.e. autumn (October of the previous year to December of the previous year), winter (January of the current year to the March of the current year), spring (April of the current year to the June of the current year), and summer (July of the current year to the September of the current year). The periods are selected in such a manner that the water year can be divided into four seasons and the affect of each season can be analyzed separately. Similar to current study, past researchers have also used the same seasonal categorization in different hydrologic studies (Tootle and Piechota, 2004; Regonda et al., 2005; Singhrattna et al., 2005; Tootle and Piechota, 2006; Kalra et al., 2008). The seasonal spread of the input data for each climate division is shown in horizontal box plots (Figure 27 a and 27 b). The horizonatl line inside the box shows the median value. The box represents the 25th and 75th percentile (interquartile range) values, whereas the whiskers extend from 5th to 95th percentile values.

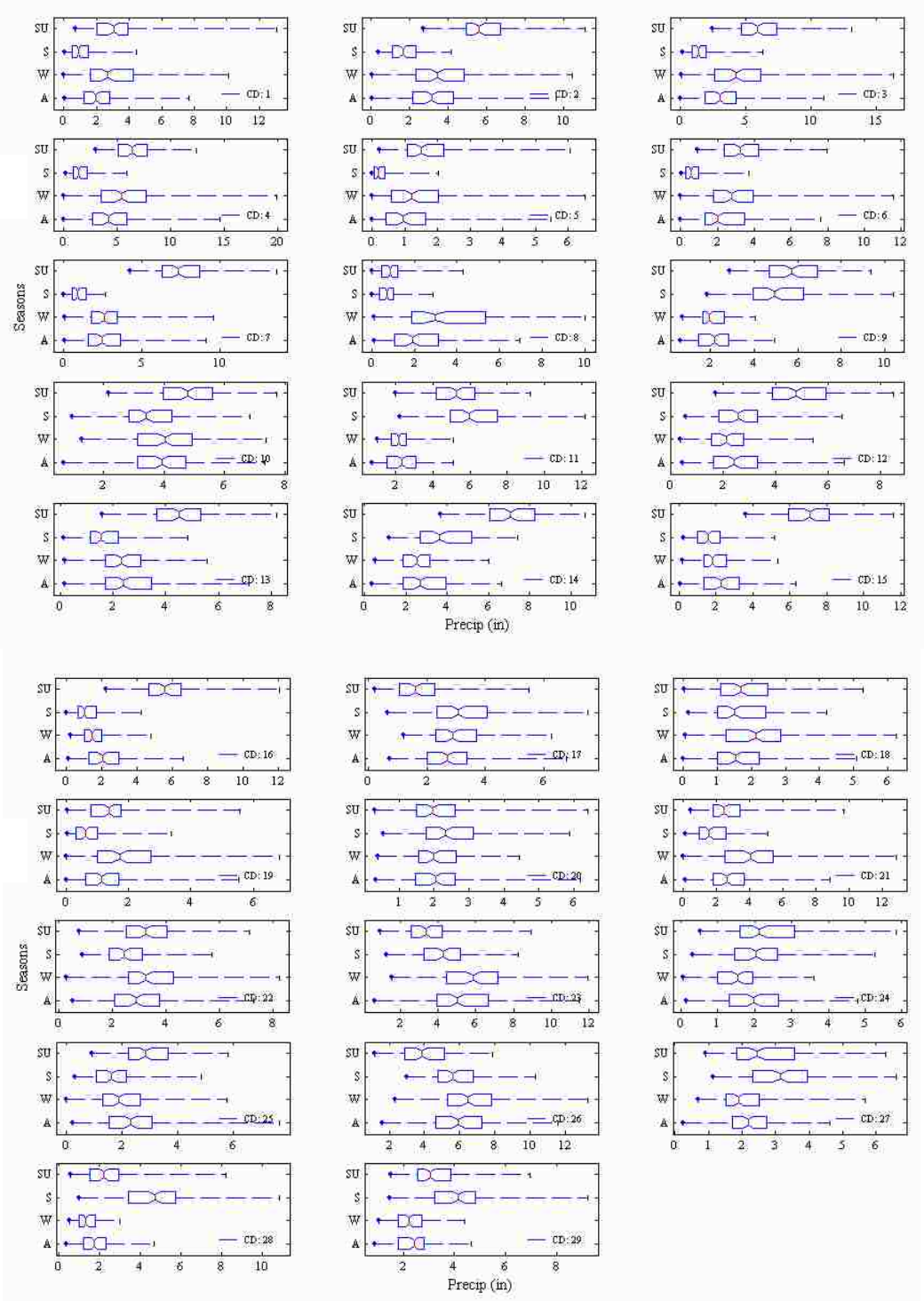


Figure 27: Box plots depicting seasonal precipitation data from 1900-2008 for (a) Climate Divisions 1-15, and (b) Climate Divisions 16-29. The four seasons are Autumn (A), Winter (W), Spring (S), and Summer (SU). The vertical line inside the box shows the median value. The box represents the 25th and 75th percentile (interquartile range) values whereas the whiskers extend from 5th to 95th percentile values.

The box plots show that the seasonal precipitation within the CRB exhibits higher degree of variability and capturing this variability is a challenging task.

4.3.2 Snow Water Equivalent

Historic March 1, April 1, and May 1 SNOTEL data for the Colorado River Basin were used from Aziz et al. (2010). The data ranges from 1961-2006 and is archived in the Natural Resources Conservation Service (NRCS) SNOTEL website (http://www.wcc.nrcs.usda.gov/snotel/). Due to the tradeoff between the length of period and number of stations, 50 snowpack stations within the CRB for March 1 and April 1 and 43 for May 1 are used in the current analysis.

4.3.3 Streamflow

The average monthly streamflow data (cfs) for United States Geologic Survey (USGS) stream gage 09380000 (Colorado River at Lees Ferry, Arizona) is obtained from the USGS_ NWISWeb Data retrieval (http://waterdata.usgs.gov/nwis/). The data ranges from 1922-2009. The monthly streamflow estimates were converted to volume (acre-ft) using appropriate conversion and summed for the water year and the four seasons i.e., autumn, winter, spring and summer as described in section 4.3.1.

4.4 Methods

First, the nonparametric statistical tests used to detect the changes (trend and step) in seasonal precipitation from 1900-2008 for the 29 climate divisions within the CRB are discussed. Next, the modified KNN disaggregation framework used to estimate seasonal precipitation values for the 29 climate divisions is described.

4.4.1 Statistical Tests

Seasonal time series of precipitation ranging from 1900-2008 over each climate division is evaluated independently for detecting change in the data. Trend software by Chiew and Siriwardena (2005) is used for detecting the changes in the seasonal precipitation. The program is designed to facilitate statistical testing for trend, change and randomness in hydrological and other time series data. Two statistical tests i.e., Mann-Kendall and Spearman's Rho are used for trend analysis and Rank-Sum for step change analysis. In what follows is a brief description of the statistical tests. Interested readers are referred to Chiew and Siriwardena, (2005) for detailed explanation of these tests.

The Mann-Kendall test is a nonparametric test in which the rank of the data values within a time series are compared. A test statistic, S, is derived through

$$S = \sum_{i=1}^{n-1}\left[\sum_{j=i+1}^{n}\mathrm{sgn}(R_j - R_i)\right] \quad (1)$$

where R is the rank of the value x within a time series X, n is the number of values, and sgn(x)=1 for x>0, sgn(x)=0 for x=0, and sgn(x)=-1 for x<0. The *z*-statistic i.e. critical values can be derived from a normal probability table

$$z = \frac{|S|}{\sigma^{0.5}}, \text{ where} \quad (2)$$

$$\sigma = \frac{n(n-1)(2n+5)}{18} \quad (3)$$

Spearman's Rho is also a nonparametric rank based test which determines the correlation between variables within a time series. Like the Mann-Kendall test, the n time series values are replaced by their ranks. The *z*-statistic ρ_s is described by following equations and can be obtained from the normal probability tables.

$$\rho_s = \frac{S_{xy}}{(S_x S_y)^{0.5}}, \text{ where} \quad (4)$$

$$S_x = \sum_{i=1}^{n} (x_i - \overline{X})^2, \quad (5)$$

$$S_y = \sum_{i=1}^{n} (y_i - \overline{Y})^2, \text{ and} \quad (6)$$

$$S_{xy} = \sum_{i=1}^{n} (x_i - \overline{X})^2 (y_i - \overline{Y})^2 \quad (7)$$

where x_i is time, y_i is the variable (in this case seasonal precipitation), and $\overline{X}$ and $\overline{Y}$ are their ranks.

The Rank-Sum test is a nonparametric test comparing the medians in two different periods of a dataset. The dataset is divided in two periods based on the step change temporal location. In the current analysis the year 1977 was used as the year showing the step change as it has been documented by other researchers (Kerr, 1992; Beamish et al., 1997; Holbrook et al., 1997; Mantua and Hare, 2002). Values over the entire time series are ranked into a single series without the regard to which sample they are in. The sum *S* of the smaller dataset is computed. In case of equal samples, an average sum is computed. A theoretical mean μ and standard deviation σ are defined as

$$\mu = n(N+1)/2 \text{ and} \quad (9)$$

$$\sigma = \left[nm(N+1)/12\right]^{0.5} \quad (10)$$

where n is the number of values in the small dataset, m is the number of values in the large dataset, and N is the total number of values in the entire time series. The *z*-statistic Z_{rs} is computed as

$$Z_{rs} = (S - 0.5 - \mu) \text{ if } S > \mu \quad (11)$$

$Z_{rs} = 0$ if S= μ and (12)

$Z_{rs} = (S + 0.5 - \mu)$ if S< μ (13)

Z_{rs} can be compared to a normal probability table to derive a level of significance. The tests are rank-based procedures and are not influenced by the use of skewed variables. Also, the tests do not assume any form of linear relationship within the data as is inherent in the correlation analysis. The reliability and efficiency of these tests in evaluating trends in different hydroclimatological variables (i.e. temperature, precipitation, and streamflow) is well established (Bunting et al., 1976; Frei and Schar, 2000; Haylock and Nicholls, 2000; Hennessy et al., 1999; Karl and Knight, 1997; Gonzalez-Hidalgo et al., 2001; Luis et al., 2000; Timbal, 2004; Kalra et al., 2008; Miller and Piechota, 2008). In the current analysis, a trend change for a climate division is termed as increasing or decreasing when both tests i.e. Mann-Kendall and Spearman's Rho are in agreement. For step change, the rank sum test has to be significant to show any change in the data. The tests are evaluated for confidence levels of 90% ($p \leq 0.10$), 95% ($p \leq 0.05$), and 99% ($p \leq 0.01$). Overall confidence levels reported in the study are based on individual tests and the confidence levels between the tests do not have to match. For instance, if for a climate division Mann-Kendall shows significance at $p \leq 0.05$ and Spearman's Rho coefficient shows significance at $p \leq 0.10$, the results are reported at a significance of $p \leq 0.10$. The affect of El Niño-Southern Oscillations (ENSO) in relation to seasonal precipitation is also evaluated. The precipitation values for El Niño years are extracted and the trend analysis is preformed on the remaining data set for each of the 29 climate divisions within the Colorado River Basin. The entire process is than repeated for the SNOTEL data and the streamflow gage to analyze the hydroclimatic variability in

SWE and Lees Ferry streamflow in relation to seasonal precipitation within the Colorado River Basin.

4.4.2 Modified K-Nearest Neighbor Disaggregation Algorithm

The framework used to disaggregate water year precipitation into four seasonal values for the 29 climate divisions within the CRB follows the work of Prairie et al. (2007). Usually, the disaggregation problem amounts to the simulation from the conditional Probability Distribution Function (PDF) f (X/Z) with the constraint that the disaggregated value sum up to the aggregated value. The basic technical details and example of the KNN disaggregation technique are outlined in Prairie et al. (2007). A brief overview of the KNN algorithm abstracted from Prairie et al. (2007) is described below. The algorithm refers to the temporal (water year to seasonal) disaggregation in which the dimensionality d is equal to 4 (i.e. seasons). The conditional PDF can be written as

$$f(X/Z) = f(X,Z) / \int f(X,Z)dx \tag{14}$$

where X is the seasonal precipitation vector and Z is the aggregated water year precipitation. The numerator in equation (14) requires the estimation of d + 1 dimensionality joint density function but due to the additivity requirement, all of the mass of PDF is situated on the d-dimensional hyperplane defined by

$$X_1 + X_2 ..... + X_d = Z \tag{15}$$

The conditional PDF for a particular aggregated value Z is defined by the probability density on a d -1 dimensional hyperplane slice through the d-dimensional density f (X/Z). The disaggregation procedure is considered as a sampling from conditional PDF f (X/Z) with the (additivity) constraint that all X should add up to Z. This is achieved by

orthonormally rotating vector X into a new vector Y whose last coordinate is aligned perpendicular to the hyperplane defined in equation (15). The simulation is performed in the rotated space and back rotated (Tarboton et al., 1998). Gram Schmidt orthonormalization (e.g. Lang, 1970) is used to determine this rotation. The steps involved in modified KNN algorithm for a single climate division are as follows.

Step 1: Orient the historic seasonal data **X** such that the seasonal precipitation values are across rows and time is across column. **X** is then rotated into **Y** (as described above) through a rotation matrix **R** where,

$$\mathbf{Y} = \mathbf{RX} \tag{16}$$

The detailed procedure for obtaining the rotation matrix **R** is outlined in Tarboton et al. (1998). A succinct summary of the procedure is presented herein. The rotation matrix is developed from a standard basis (basis vector aligned with the coordinate axes), which is orthonormal but does not have a basis vector perpendicular to the conditioning plane. A vector perpendicular to the conditioning plane replaces one of the basis vectors. By doing this the basis set of vectors is still non orthonormal. Gram Schmidt orthonormalization procedure is used to obtain an orthonormal vector perpendicular to the conditioning plane. This results in orthonormal perpendicular **R** vector with the property of $\mathbf{R}^{\mathbf{T}}=\mathbf{R}^{-1}$.

Step 2: The next step is to generate the aggregate water year precipitation value z^*. Previous KNN studies generated aggregate value from fitting the data to an appropriate model (Salas, 1985; Lall and Sharma, 1996; Prairie et al., 2005, 2006, and 2007). This approach sometimes resulted in missing the extreme values, which are of particular interest to the water managers and are vital for managing the reservoir operations. Also,

the values generated can be period specific, which is a latent problem in many disaggregation approaches due to fewer samples in the model validation period. To avoid these issues the current analysis uses a weighted moving window of periods to generate the aggregate value. Average monthly values are available for the period of 1900-2008. The available monthly precipitation values are averaged for the water year (as described in section 4.3.1) to obtain the aggregate value z^*. For a particular year '*m*' (where m = 1900:2008, i.e. 109 years) aggregated value, the number of nearest neighbors K of the historic data series are selected by a heuristic scheme $K = \sqrt{N}$ (where N is the sample size). Although there are other methods such as generalized cross validation (GCV) that can also be used to obtain K but the heuristic scheme has performed well in a variety of applications (Lall and Sharma, 1996; Rajagopalan and Lall, 1999; Singhrattna et al., 2005; Prairie et al., 2007). Then the selected neighbors of the generated $z_{sim}' = z^*/\text{sqrt}(d)$ (where d is the dimension) are assigned weights giving more weight to the nearest neighbor and less to the farthest neighbor.

$$W(k) = \frac{1}{k\sum_{i=1}^{K}\frac{1}{i}} \qquad (17)$$

k = 1,2,3….K

Say the seasonal disaggregation for water year 1920 is desired. The years close to 1920 are given higher weight and farther less.

Step 3: Using the weights in equation (17), the neighbors for j^{th} time of the m^{th} historic record are resampled.

Step 4: The rotated matrix Y in equation (16) has its last column $Y_d = z_{sim}' = z^*/\text{sqrt}(d)$ and the first *d* -1 columns of the rotated matrix are denoted by U. Then the rotated matrix

Y becomes a function of U and z_{sim}' such that Y = [U, z_{sim}']. The disaggregated precipitation variable Y^* is generated for the m^{th} year

$$\mathbf{Y}^* = (U_j, z_{sim}') \quad (18)$$

Step 5: The rotation matrix is back rotated to the original space for the m^{th} year

$$\mathbf{X}^* = \mathbf{R}^T\mathbf{Y}^* \quad (19)$$

where X^* is the seasonal vector of the m^{th} year disaggregated water year precipitation and will sum up to z^*. Step 2-5 are repeated for the entire 109-years of data to generate a pool of seasonal disaggregated values. Also, for each m^{th} year aggregated value, 1000 simulations each of 108-years in length are computed to generate ensembles of seasonal values instead of a single trace. The entire process is then repeated for all the climate divisions. The performance measures used for evaluating the effectiveness of the KNN model for the pooled values for each climate division are Root Means Square Error (RMSE), Mean Absolute Error (MAE), and Correlations Coefficient (R). However, the current approach resample the historic data but using weighted moving window of periods helps in better capturing the nonstationarity in the precipitation. The bootstrapping approach helps in generating the seasonal values not seen in the past as opposed to using the simple KNN based approach, which generates values only seen in the historic data. Similar to Prairie et al. (2007) disaggregation approach, the current approach also produces negative values. But using the more robust and not period specific moving window bootstrap approach for generating precipitation simulations the negative values are minimized (less than 0.3% of the simulated values for all climate divisions) and they do not affect the overall results.

In order to assess the relative performance of KNN approach, first order periodic autoregressive model (PAR-1) is also developed to obtain the seasonal precipitation values for the 29 climate divisions within the Colorado River Basin. The AR type models are time series parametric models designed to capture the basic statistical characteristics. Details on the theoretical aspects of AR type models are available in Salas et al. (1980).

4.5 Results and Discussion

The results are discussed in three ensuing sections. Sections 4.5.1.1 - 4.5.1.5 describe the trend and step changes in seasonal precipitation, its relationship to ENSO, trends in SWE, and Lees Ferry streamflow data. Section 4.5.2 highlight the seasonal precipitation disaggregation results using KNN approach for the 29 climate divisions encompassing CRB, and section 4.5.3 provides the KNN disaggregation comparison with the standard parametric alternative.

4.5.1 Trend and Step Changes

The changes in seasonal precipitation and SWE are reported at three confidence levels (described in section 4.4.1) whereas the changes in seasonal flow at Lees Ferry gage are reported at $p \leq 0.05$. It should be noted that the magnitude of trend and step changes for seasonal precipitation and SWE are not computed because the different confidence levels used in the current research represent the quantitative measure of the data.

4.5.1.1 Seasonal Precipitation Trend Changes

The spatial profile of trend changes in seasonal precipitation for the 29 climate divisions encompassing the CRB is shown in Figure 28A. For autumn season (OND) an increasing trend in precipitation is noted for climate divisions 15, 16, and 22 for $p \leq 0.0.5$

and for climate division 25 at p≤0.10. Decreasing trends for climate divisions 8, 27, and 28 are noted at p≤0.05 and climate division 29 show decreasing trend at p≤0.10, while the remaining divisions show no changes during autumn season.

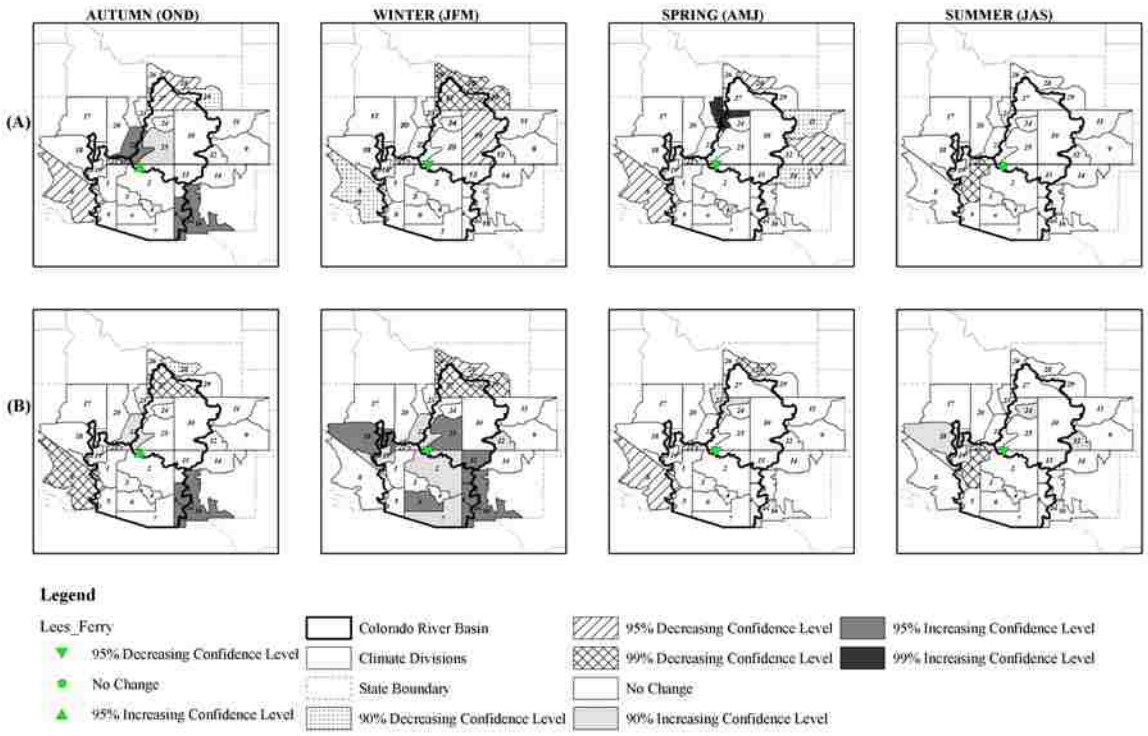


Figure 28: Spatial maps showing the (A) Trend change, and (B) Step change in seasonal precipitation for 29 climate divisions encompassing the Colorado River Basin. The seasonal Trend and Step change in Lees Ferry streamflow are also shown.

For winter season (JFM) climate divisions 8 and 10 show decreasing trends at p≤0.10 and p≤0.05 and climate divisions 26-29 show decreasing trends at p≤0.01. For spring season (AMJ) decreasing trends in seasonal precipitation are dominant with six climate divisions showing decreasing trends and a single climate division showing increasing trends. The decreasing trends are noted for climate divisions 9 and 14 at p≤0.10, climate divisions 4, 8, 9, and 28 at p≤0.05. The increasing trend is noted for climate division 23 at

$p \leq 0.01$ with remaining divisions showing no change. For summer season (JAS) precipitation remained relatively unchanged with single climate division (CD 1) showing decreasing trend at $p \leq 0.01$.

The trend results indicate that overall there is a decrease in seasonal precipitation within the CRB. The decreasing trends in precipitation are dominant during winter and spring seasons; compared to the autumn and summer seasons. The increasing trends are seen for four climate divisions during autumn season and one climate division during the spring season. Majority of the climate divisions that show decreasing precipitation trends during the winter season envelop the northwest mountainous region of the Colorado River Basin. This region primarily has precipitation in the form of snow, which is generated by the frontal systems originating in the North Pacific Ocean. Precipitation in this region replenishes the mountain storage and is a source of snowmelt in the critical spring runoff season. Decreasing precipitation trends for this region indicate the changing character of climate due to variability in the atmospheric circulation patterns and sea surface temperature (SST) of the tropical and North Pacific Oceans, which can affect the frequency and moisture content of frontal systems and alter the long-term trend of winter precipitation within the Colorado River Basin. The winter precipitation within CRB plays a vital role in generating the peak spring-summer streamflow. The decreasing winter precipitation trend coupled with temperature change can lead to the intensification of the hydrological cycle (Huntington, 2006) and cause a shift in the timing of the peak runoff (Cayan et al., 2001). Additionally, increases in surface temperature at higher latitude have been noticed, which have resulted in systematic decrease in snow cover extent and changes in the amount of precipitation falling as rain versus snow during the winter

months (Karl et al., 1993). The decreasing trend of winter precipitation within the UCRB is in agreement with the findings of Christensen et al. (2004). Christensen et al. (2004) assessed the hydrology and water resources of the CRB by comparing the downscaled climate simulations of a Parallel Climate Model (PCM) for three periods (2010-2039, 2040-2069, and 2070-2098) into the future. They indicated decrease in winter precipitation for all three periods, which resulted in large reduction in streamflow within the Colorado River Basin. Decreases in winter precipitation over southern Europe (Brunetti et al., 2001) and Mediterranean, and wet anomalies from Iceland eastward are related to positive phases of North Atlantic Oscillations (NAO) (Hurrell, 1995). Similar climate fluctuations can be related to cause changes in winter precipitation within CRB as streamflow within UCRB is strongly associated with NAO variability (Kalra and Ahmad, 2009). Moreover, results from several GCM runs and scenarios have shown a 10% increase in precipitation above current values in the northwestern U.S. and a 10% decrease below current values for the southwestern United States (Nash and Gleick, 1991; Christensen and Lettenmaier 2007). The increases in autumn precipitation for climate divisions in mid-elevations can be attributed to the increase in temperatures that have lead to more frequent moderate to high intensity non-convective events (Hennessy et al., 1997). Summer precipitation usually has not much role to play within CRB and has remained unchanged during the period of record as indicated by the trend analysis.

4.5.1.2 Seasonal Precipitation Step Changes

The spatial profile of step changes in seasonal precipitation for the 29 climate divisions in the CRB is shown in Figure 28B. An increasing step change in seasonal precipitation is noted for climate division 22 at $p \leq 0.10$ and climate divisions 15 and 16 at

$p \leq 0.05$ for autumn season (OND). Climate division 28 shows decreasing step change at $p \leq 0.10$ and climate divisions 8 and 27 show decreasing step changes at $p \leq 0.01$ for autumn season, while other divisions remained unchanged. Increasing step changes are noted for eleven climate divisions during winter seasons (JFM) but with varying confidence levels. It should be noted that majority of the climate divisions showing increasing step change in winter season precipitation cover the Lower Colorado River Basin. The spring (AMJ) and summer seasons (JAS) remained relatively unchanged with one (spring) and three (summer) climate divisions showing increasing step changes and two (spring) and one (summer) climate divisions showing decreasing step changes at different confidence levels.

The step change results for the seasonal precipitation are similar to the results obtained from trend change analysis for autumn, spring and summer seasons but not for winter season. The step change result show a decrease in winter precipitation for climate divisions enveloping the northwest mountainous region of the CRB (similar to trend results) but also indicated that majority of the LCRB is getting wetter opposed to the trend results. To clearly visualize a step change bar plots for sample climate divisions depicting the abrupt shift of 1976-77 (increase and decrease) in winter precipitation are shown for the lower basin (Figure 29A and B) and upper basin (Figure 29C and D). A clear jump (upward and downward) in the mean value is seen around the year 1976-77 for the selected climate divisions. The precipitation is altered following the step change indicated by the bar plot. This jump may be attributed to the changing climate as a result of increased greenhouse gas concentrations or land use change (urbanization, clearing, aforestation, etc) as documented in previous studies. This jump in the mean value

coincides with the climate 'regime shift' of mid-1970s, which had widespread consequences for the biota of the North Pacific Ocean and Bearing Sea (Hare and Mantua, 2000). This regime was a result of the shift towards warm regime in Californian Current and Gulf of Alaska (Hollowed and Wooster, 1992, Trenberth and Hurrel, 1994; Mantua et al., 1997; Hare and Mantua, 2000). The cooling of central North Pacific Ocean and the warming of the Northeast Pacific Ocean was witnessed following the regime shift (Hare and Mantua, 2000). Increased sea level pressures were witnessed over the western U.S., the central Arctic, northern Africa, and northern Asia. Usually there is no common definition of step change but is characterized by a behavior of a natural phenomenon (e.g., sea level pressure,) over time (Hare and Mantua, 2000). Due to this, it becomes imperative to distinguish between a gradual trend change and an abrupt step change for climate studies.

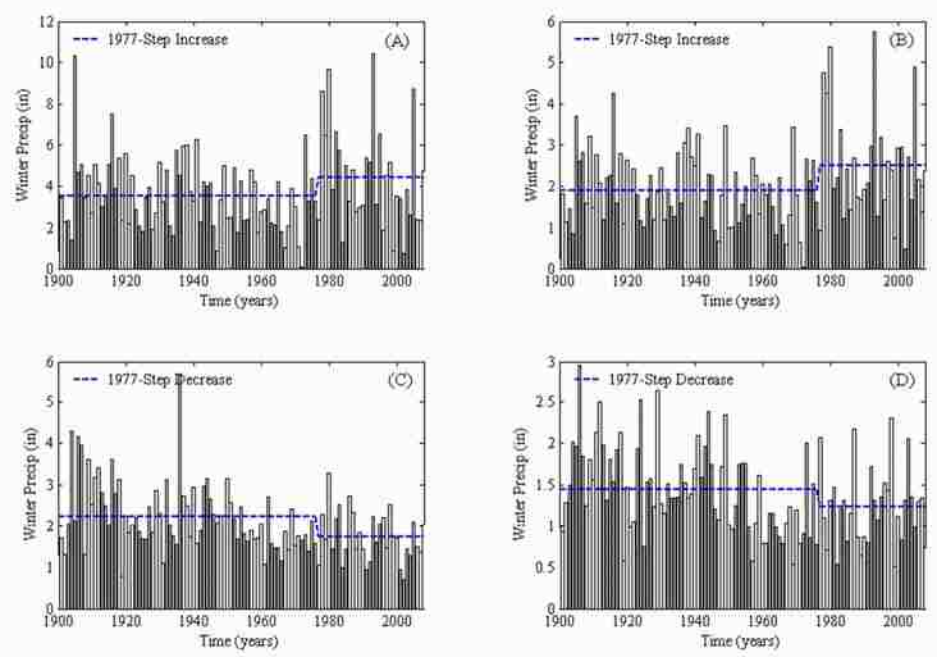


Figure 29: Bar plots depicting step changes (increase/decrease) in winter season precipitation for A) climate division 2, B) climate division 25, C) climate division 27, and D) climate division 28. The dotted line shows the pre and post-1977 mean value.

Apart from statistical results (Figure 29), increases in the LCRB precipitation can be linked to the warm winter storms, which tap the moisture from the Tropical Pacific Ocean and may increase the rainfall events and even their intensities resulting in high runoff and floods on major rivers within the basin. Other possible explanations for an abrupt step change are attributed to anthropogenic global warming caused by increased urbanization resulting in greater CO2 emissions (McCabe and Wolock, 2002). The increasing step change for LCRB climate divisions is in agreement with the works of Miller and Piechota (2008) but for a different period of record. Increasing step changes witnessed in winter precipitation for an arid region like the LCRB with sparse and shrubland like vegetation can affect the soil characteristics within the region (Hansen et al., 2006; Hansen and Ines, 2005; Robertson et al., 2007; and Ahmad et al., 2010). Changes in soil characteristics can be resourceful in evaluating the long-term variations in crop production, which in turn can be used to adjust decisions related to water releases. Furthermore, the decreasing winter precipitation in the northwest mountainous region of the upper basin is attributed to an abrupt step change compared to a gradual trend in data. This is indicated by the spatial profile (Figure 28) and corroborated by the jump (upward/downward) shown in the bar plots (Figure 29). This declining precipitation step change is of concern because approximately 20% of the basins precipitation falls in the highest 10% and roughly 40% of the precipitation falls in the highest 20% of the basin. This winter precipitation is primarily stored and transferred to the dry summers. With the increasing temperature (due to global warming) and declining precipitation, reduced summer streamflow volumes are obtained that significantly affect the water resources planning and management within the basin (Cayan et al., 2001; Stewart et al., 2005).

4.5.1.3 El Niño and Seasonal Precipitation Trend Changes

The increasing impacts of droughts and flood on agriculture, water resources, and environment due to natural climate variability on interdecadal and decadal time scales have captured the attention of scientific community (Mann et al., 1995; Hu and Feng, 2001). Signatures of recent climate trends are seen in several regional and global variables including precipitation. One of the most important characteristics of precipitation within the Southwest U.S. is the high degree of seasonal, interannual, and decadal variability induced by large scale oceanic-atmospheric oscillations. Much attention has been devoted to why precipitation varies in relation to ENSO (Ropelewski and Halpert, 1986; Diaz and Kiladis, 1992). ENSO events typically last from 6 to 18 months and is the single most important factor affecting the interannual variability on global scale and particularly in the western U.S. and the Colorado River Basin (Diaz and Kiladis, 1992; Cayan et al., 2001; Webb et al., 2005). In the past, different phases of El Niño-Southern Oscillations (ENSO) i.e. El Niño (warm) and La Niña (cold) have been used to explain the weather conditions in the Southwest and particularly within the Colorado River Basin. In general, El Niño events have been associated with wetter than average winters in the CRB while La Niña have been linked with dry conditions. The spatial profile of the trend (Figure 28A) and step change (Figure 28B) results along with the bar plots (Figure 29B) indicated that winter season precipitation in the UCRB has a decreasing trend while the LCRB precipitation showed an increasing trend (Figure 28 and Figure 29A). The changes in precipitation are attributed to an abrupt step change compared to the gradual trend indicated by the statistical results. This step shift in winter

precipitation can have serious implication for water resources and need to be analyzed for its linkage with the climate variability at interdecadal or decadal time scales.

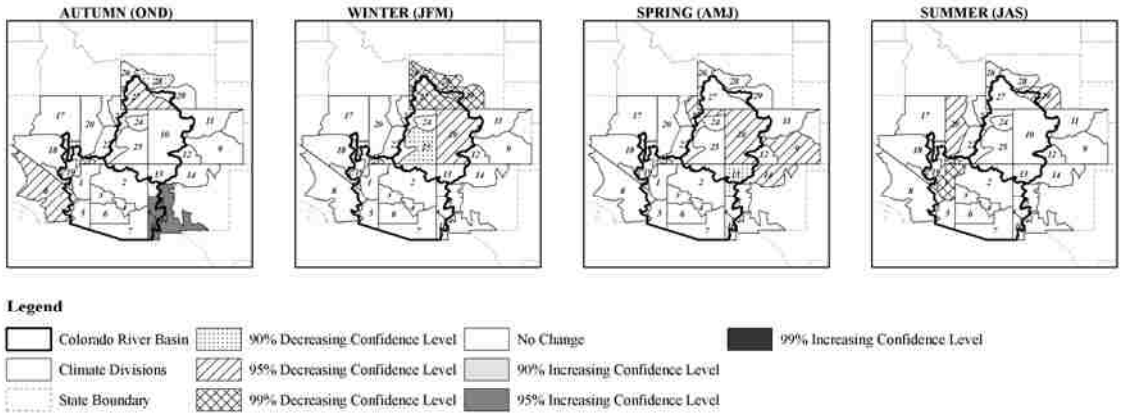


Figure 30: Spatial map showing the Trend change in seasonal precipitation removing the ENSO years for 29 climate divisions encompassing the Colorado River Basin.

For this reason, the impact of ENSO on the regional climatology of the CRB needs to be evaluated. Historic El Niño events were removed from the seasonal precipitation data for all the 29 climate divisions and the trend analysis was performed on the remaining data set. The information about the historic El Niño events was obtained from the National Weather Service (NWS) Climate Prediction Center (CPC) website (http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml). It should be noted that the step change year i.e. 1976-77 used in the current analysis was an El Niño year and was removed from the data so only the trend analysis was performed. Figure 30 shows the spatial profile of trend changes in seasonal precipitation for the 29 climate divisions in the CRB after removing the data for El Niño years. During the autumn season, none of the climate divisions in the UCRB show increasing trends in precipitation. Although increasing trends were noticed for divisions 22 and 25 (Figure

28A) when using the entire data set. This indicates that the increasing trends in autumn season precipitation for divisions 22 and 25 are results of El Niño events caused by shifting of jet streams. These jet streams are pulled south of California and cause storms in the Pacific Ocean. This combination of jet streams and storms results in warmer than average ocean waters, and often is the cause of increase in precipitation. Decreasing trend for the divisions 8, 27, and 28 and increasing precipitation trends for divisions 5 and 16 in the LCRB during autumn season are similar to the results from Figure 28A indicating no linkages with El Niño. The decreasing trend in winter season precipitation for climate divisions in the northwest mountainous region is shown in Figure 30 similar to Figure 28A indicating no linkages with El Niño. This is in agreement with the works of Redmond and Koch (1991) where they highlighted that the precipitation in the mountains of Colorado, Wyoming, Utah are not greatly impacted by ENSO and better forecast of ENSO and its effects are not likely to greatly improve the upper basin mainstem streamflow forecasts. The decreasing precipitation trends for climate divisions in the UCRB for spring and summer season are quite similar to results from Figure 28A indicating no linkages with El Niño. Overall, the results from this analysis indicated that the decreasing seasonal precipitation trends for majority of the climate divisions in the UCRB indicated in Figure 28A are not due to the effect of El Niño. Also, the increasing trends in autumn precipitation in upper basin are linked with El Niño. Another important inference drawn through this analysis is that the increasing winter precipitation step change (Figure 28B) for the climate divisions in the LCRB may be a result of El Niño. This is because ENSO effects are more pronounced in the lower basin than in the upper basin. Piechota and Dracup (1996) showed that ENSO events coincide with major dry

and wet spells in the Lower Colorado River Basin evidenced by the Palmer Drought Severity Index (PDSI). Moreover, when southern regions of the southwest are wet, precipitation in the Upper Colorado is often average or below average (Guido, 2010). This corroborates the findings of current research indicating that winter precipitation in upper basin is decreasing and whereas lower basin has an increasing trend in precipitation over the twentieth century. The summer season precipitation has no signatures of ENSO associated with it. Although, ENSO exerts a strong influence in modulating wet and dry conditions within the CRB, there might be other climate patterns such as Pacific Decadal Oscillation, Atlantic Multidecadal Oscillation, North Pacific Oscillations etc., which individually or coupled with other climate patterns can affect the hydroclimatology within the Colorado River Basin.

4.5.1.4 Streamflow Trend and Step Changes

Lees Ferry gage is a point on the Colorado River located on the hydrologic divide between the upper basin and the lower basin. The water supply to the lower basin is governed by the available water at this gage. Due to this reason evaluating long-term trends become vital at Lees Ferry and can be used by the water managers in efficient planning and management of water resources within the Colorado River Basin. Similar to seasonal precipitation, trends in seasonal Lees Ferry streamflow are evaluated using nonparametric tests. The increasing (decreasing) trend and step change for each season are shown by upward (downward) facing triangles and reported at $p \leq 0.05$ confidence level. Figure 28 shows the seasonal trend (A) and step (B) changes for Lees Ferry streamflow from 1922-2009. An increasing trend and step change is noted for the autumn season streamflow. Decreasing trend and step changes are noted for winter and spring

season flows. The summer season streamflow showed a decreasing step change. The decreasing streamflow trends during winter and spring season are in agreement with the decreasing trends in precipitations for the climate divisions in the upper basin. The winter snowpack generates the peak summer streamflow. Changes in the timing of the snowmelt due to rising winter and spring temperatures results in earlier snowmelt driven streamflow and shifting the peak earlier in season (Cayan et al., 2001) resulting in reduced summer flows. Although the climate divisions in the upper basins showed decreasing trends in autumn season precipitation, the flow at Lees Ferry showed an increasing trend. This could be attributed to the late summer storms, which are caused by the moist air from the Gulf of Mexico, the Gulf of California, and the eastern Pacific Ocean (Webb et al., 2005). These storms can result in late August and September month's high-intensity rainfall at elevations below 7,000 feet and contribute to the autumn peak streamflow. It should be noted that the seasonal trends in Lees Ferry gage are due to an abrupt step change and not due to gradual trend change. This is of importance because unlike trend changes, step changes are non-linear, occurs abruptly, and may reoccur in the future and can lead to extreme events such as floods and droughts caused by the increases and decreases in precipitation.

4.5.1.5 Snow Water Equivalent Trend and Step Changes

In CRB, snowpack is an important source of runoff and water supply, accounting for 50 to 70 percent of the annual precipitation. Majority of flow to the headwaters of Lees Ferry is generated by the winter season snowpack in the mountainous regions of Colorado, Utah, and Wyoming. Evaluating long-term trends in the winter season snowpack can be useful for water managers and forecasters in improving the spring-

summer runoff estimates, which are critical in the management of reservoir operations and agricultural demands. The trend and step changes were evaluated for March 1, April 1, and May 1 Snow Water Equivalent. Figure 31 shows the spatial profile of the trend (A) and step (B) changes for the three SWE temporal periods. The results indicate a decreasing trend and step change in SWE for the three temporal periods within the upper basin. The decreasing trend is more pronounced in May 1 SWE, than April 1 SWE, and least in March 1 SWE data. The decreasing trends in April 1and May1 SWE is dominant in the northwest mountainous regions of the basin, which is a major contributor to the streamflow within the basin. A couple of stations also showed increasing trends in the April 1 SWE value. The reductions in SWE are directly attributable to higher winter temperatures and the resulting decreases in the ratio of precipitation falling as snow versus rain (Mote et al., 2005; Hamlet et al., 2005). It is noteworthy that the changes in SWE are also due to an abrupt step change and not due to a gradual trend. Also, the results are in agreement with the trend and step change analysis in seasonal precipitation (Figure 28), which indicated decreasing step changes for winter precipitation for the climate divisions encompassing the northwest mountainous regions within the Colorado River Basin. The seasonal changes for Lees Ferry (section 4.5.1.4) also agree with April 1 and May 1 SWE results and decreasing spring and summer flows are a results of the decreasing trends of the mountain snow-pack. Decreases in mountain snowpack occur due to winter rainfall events as they do not replenish mountain storage and can come at the expense of snowfall events (Groisman et al., 2001).

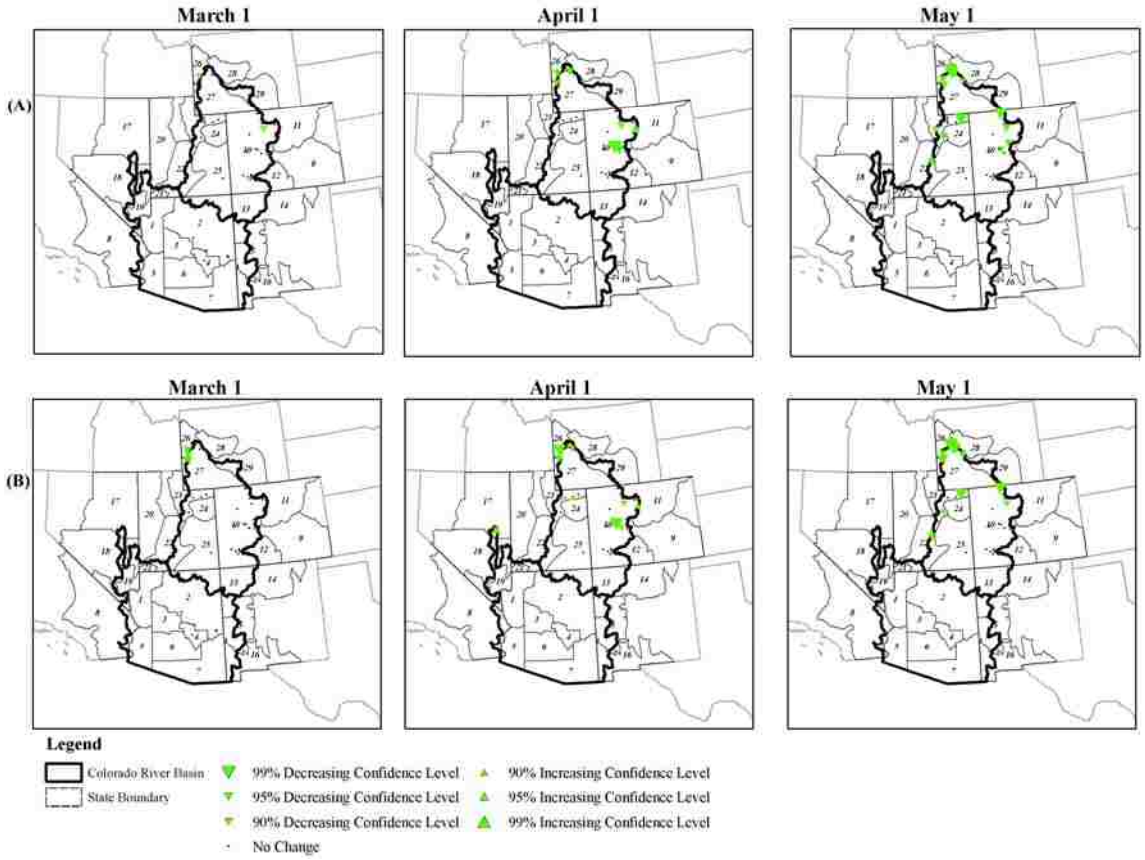


Figure 31: Spatial map showing the (A) Trend change and (B) Step change for March 1, April 1, and May 1 SWE stations within the Colorado River Basin.

Similar to Figure 29, step change (increase/decrease) in April 1 SWE values for selected locations is shown in Figure 32. The increasing step change is depicted for SNOTEL stations in Nevada and Utah, whereas the decreasing step change is shown for stations in Wyoming and Colorado. A clear jump (upward/downward) in the mean of SWE values is observed around the year 1976-77. Following the step change, variation in snow depth is witnessed (Figure 32). This jump in the mean value coincides with the climate 'regime shift' of mid-1970s discussed earlier. Overall, the SWE results are in agreement with the changes in seasonal precipitation and indicate that the precipitation in the upper basin has a decreasing trend.

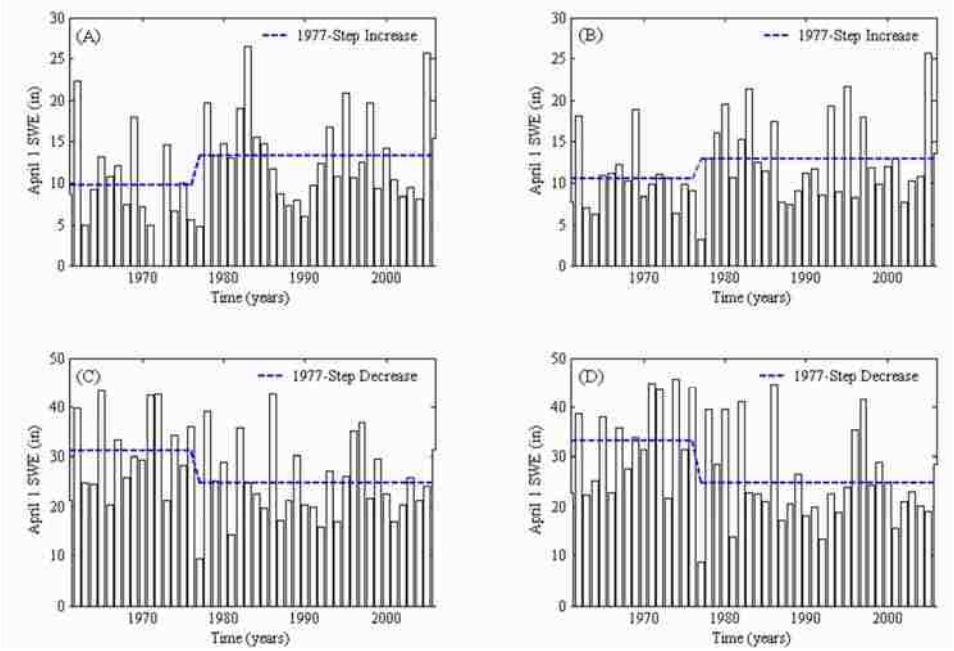


Figure 32: Bar plots depicting step changes (increase/decrease) in April 1 SWE for SNOTEL stations A) NV-14K05S, B) UT-09J05S C) WY-10G20S, and D) WY-10G02S. The nomenclature used for the SNOTEL sites is similar to that of archived in NRCS. The dotted line shows the pre and post-1977 mean value.

The trend and step change analysis indicated that seasonal precipitation within the CRB exhibits higher variability. The seasonal variability of precipitation has tremendous affect on runoff generation in the basin. The value of evaluating long-term changes in precipitation for scientific communities is in getting prepared for climate extremes. These changes serve the qualitative purpose by providing valuable information for developing future climate change scenarios. Furthermore, precipitation is the driving mechanism of several hydrological processes that are assumed to be operating under the assumption of stationarity. Detecting changes in precipitation brings the stationarity assumption in question (Segond et al., 2006; Milly et al., 2008), along with the design of water resource systems that operate under this assumption. Therefore, it becomes necessary to include the non-stationary properties of precipitation to better represent the temporal

characteristic of precipitation such as the seasonal effect and climate variability (Segond et al., 2006). For resource planning and management of the water resources, managers are interested in quantitative aspect of precipitation. GCMs can provide precipitation estimates at a coarser temporal and spatial resolution. Multitude of hydrologic applications and climate change studies often rely on climate information available at finer temporal resolution. In the changing climate it becomes essential to achieve continuous realistic realizations of precipitation through statistical downscaling techniques. For this reason, nonparametric techniques such as KNN become useful, which can be used for temporal disaggregation of precipitation preserving the standard distributional statistics. Additionally, there is renewed interest in disaggregation methods due to their ease of use, simplicity, and robustness as the climate related issues (regional ENSO forecasts or downscaling of climate change scenarios) have come to the fore. With this motivation, the current study developed a weighted moving window KNN scheme to disaggregate water year precipitation into four seasonal values for the 29 climate division encompassing the Colorado River Basin. Disaggregating water year values into seasonal values is useful for river basin management, reservoir operation, agriculture related decisions, and paleoclimatic studies to construct the past hydrology within the Colorado River Basin. The disaggregation results are discussed in the ensuing section.

4.5.2 Seasonal Precipitation Disaggregation

Three performance measures i.e., RMSE, MAE and R are used to analyze the seasonal precipitation disaggregation results for the 29 climate divisions covering the Colorado River Basin and are reported in Table 9. The RMSE values range from 0.72-2.69 (in) during autumn season, 0.60-2.67 (in) during winter season, 0.44-1.61 (in) during

spring season, and 0.86-2.07 (in) during summer season. The MAE ranges from 0.58-1.96 (in) during autumn season, 0.47-2.05 (in) during winter season, 0.30-1.22 (in) during spring season, and 0.63-1.65 (in) during summer season. The correlation coefficient R ranges from 0.10-0.68 during autumn season, 0.12-0.83 during winter season, 0.13-0.68 during spring season, and 0.10-0.58 during summer season. In general it is noticed that majority of climate divisions have smaller RMSE and MAE errors between measured and disaggregated precipitation during the four seasons.

Table 9: Performance statistics for the measured and disaggregated seasonal precipitation for the 29 climate divisions. The RMSE and MAE are in inches (1 inch =2.54 cms).

	Autumn			Winter			Spring			Summer		
Clim_Div	RMSE	MAE	R	RMSE	MAE	R	RMSE	MAE	R	RMSE	MAE	R
1	1.53	1.20	0.18	1.61	1.24	0.63	0.83	0.61	0.30	1.79	1.30	0.20
2	1.43	1.07	0.53	1.63	1.25	0.60	0.92	0.72	0.31	1.65	1.38	0.20
3	2.02	1.54	0.44	2.47	1.84	0.60	1.22	0.89	0.19	2.07	1.65	0.27
4	2.69	1.96	0.49	2.67	2.05	0.64	1.13	0.86	0.30	2.02	1.63	0.29
5	0.80	0.60	0.57	0.92	0.67	0.65	0.44	0.30	0.13	1.02	0.73	0.31
6	1.40	1.08	0.59	1.42	1.11	0.71	0.70	0.52	0.27	1.57	1.24	0.10
7	1.33	1.04	0.68	1.29	0.95	0.57	0.69	0.54	0.17	1.73	1.35	0.37
8	1.42	1.09	0.30	1.40	1.05	0.83	0.67	0.50	0.41	0.92	0.63	0.11
9	0.95	0.77	0.40	0.74	0.58	0.14	1.35	1.06	0.66	1.46	1.21	0.40
10	1.26	1.03	0.35	1.16	0.96	0.44	0.96	0.75	0.52	1.13	0.93	0.49
11	1.10	0.89	0.29	0.78	0.61	0.12	1.61	1.20	0.56	1.52	1.25	0.38
12	1.07	0.80	0.53	0.93	0.72	0.48	0.99	0.82	0.45	1.34	1.04	0.42
13	1.13	0.91	0.49	1.10	0.85	0.33	0.79	0.62	0.46	1.29	0.98	0.39
14	1.26	1.00	0.41	0.91	0.70	0.39	1.24	1.01	0.58	1.66	1.35	0.25
15	1.20	0.94	0.53	0.89	0.66	0.39	0.81	0.63	0.53	1.39	1.15	0.54
16	1.16	0.90	0.57	0.99	0.72	0.17	0.81	0.61	0.26	1.61	1.28	0.45
17	1.02	0.81	0.33	1.00	0.79	0.28	1.17	0.97	0.54	0.93	0.74	0.37
18	1.01	0.81	0.27	1.12	0.90	0.46	0.92	0.73	0.39	1.02	0.81	0.44
19	1.04	0.79	0.10	1.15	0.86	0.71	0.60	0.45	0.23	1.09	0.76	0.13
20	0.79	0.58	0.49	0.79	0.63	0.38	0.85	0.63	0.54	0.90	0.66	0.40
21	1.66	1.28	0.27	2.04	1.59	0.63	1.11	0.84	0.25	1.75	1.22	0.10
22	1.29	0.95	0.42	1.22	0.99	0.54	0.89	0.67	0.49	1.32	1.04	0.24
23	1.91	1.54	0.45	1.62	1.28	0.58	1.34	1.06	0.56	1.35	1.09	0.38
24	0.92	0.75	0.47	0.72	0.55	0.28	0.81	0.64	0.52	0.90	0.73	0.53
25	1.09	0.81	0.47	0.99	0.77	0.45	0.77	0.56	0.46	1.04	0.84	0.34
26	1.64	1.31	0.55	1.77	1.40	0.58	1.56	1.22	0.30	1.41	1.15	0.45
27	0.81	0.65	0.36	0.80	0.57	0.36	0.97	0.76	0.59	0.93	0.73	0.58
28	0.88	0.70	0.30	0.60	0.47	0.12	1.39	1.09	0.68	1.10	0.83	0.55
29	0.72	0.60	0.46	0.72	0.58	0.22	0.96	0.75	0.63	0.86	0.70	0.58

The performance measures are mapped to see the spatial extent of performance measures during the four seasons for 29 climate divisions within the CRB (Figure 33). Figure 33A shows that of the total of 29 climate divisions, 27 divisions have a RMSE value of less than 1.50 (in) during spring season, whereas 23, 22, and 19 divisions have a RMSE value of less than 1.50 (in) during autumn, winter, and summer seasons, respectively. In case of MAE, 23 climate divisions have a MAE value of less than 1.00 (in) during spring season, whereas 18, 19, and 13 climate divisions have MAE value of less than 1 (in) during autumn, winter, and summer seasons (Figure 33B). An acceptable correlation coefficient between the measured and disaggregated precipitation is observed for most of the climate divisions during autumn, winter, and spring seasons but not during summer season (Figure 33C). A correlation value greater than 0.5 between the measured and disaggregated precipitation is obtained for 8 climate divisions during autumn, 13 divisions during winter, and 12 divisions during spring season, respectively. Five climate divisions have an R value greater than 0.5 during summer season.

RMSE (Figure 33A) and MAE (Figure 33B) spatial maps indicate that the best KNN predictions are obtained during spring season and worst during summer season. The KNN model does a better job in disaggregating upper basin precipitation compared to the lower basin for all the seasons as evident by low RMSE values (less than 1 inch) for majority of the climate divisions within the upper basin. Besides spring season, the other three seasons have few climate divisions in the LCRB showing higher RMSE values (greater than 1.5 inch). The MAE spatial maps agree with the RMSE results indicating that better precipitation predictions are obtained for divisions in the upper basin compared to lower basin for all the seasons. The best predictions for divisions in lower

basin are obtained during spring season compared to the other seasons. The correlations maps (Figure 33C) show that the disaggregated precipitation correlates best with winter seasons precipitation values for the climate divisions covering the lower basin compared to the upper basin. Whereas spring season disaggregated precipitation correlates best with divisions in the upper basin compared to the lower basin divisions. Based on the performance measures it is evident that the KNN model does a better job in disaggregating precipitation for divisions within the upper basin compared to the lower basin.

Figure 33: Spatial maps showing the range of performance measures i.e. (A) RMSE, (B) MAE, and (C) R during the four seasons for the 29 climate divisions.

To better examine the temporal variability in seasonal precipitation, bisector plots are created between the measured and estimated precipitations for sample climate divisions in the upper basin and lower basin. The sample divisions within the upper basin (CD 10, 24, 25, and 27) and lower basin (CD 1-7) are selected such that they envelop majority of the CRB and can effectively demonstrate the model performance. Figure 34 shows the scatter plots between the measured and estimated seasonal precipitation for the selected divisions in the upper basin. A good match is obtained between the measured and predicted seasonal precipitation for the climate divisions in the upper basin for the four seasons. The model does a fairly good job in capturing the extremes (low and high values) during winter (Figure 34B) and spring (Figure 34C) seasons compared to autumn (Figure 34A) and summer seasons (Figure 34D). This is evident by the majority of sample points lying close to the 45^0 bisector line indicating a good model fit. A point lying far above the bisector line indicates higher predictions whereas a point far below the line shows lower predictions. During autumn and summer seasons, model does fairly well in capturing the low values but misses the extreme at few locations. Accurate winter and spring season predictions for divisions 10 and 27 show the robustness of the model as decreasing trend change (Figure 28A) was noticed for these divisions and the KNN model was able to capture this variability efficiently. Moreover, the precipitation variability in division 25, having an increasing step change (Figure 28B), is adequately captured by the model. Although few sample points lie far away from the bisector line but majority of the points are saturated around the bisector line.

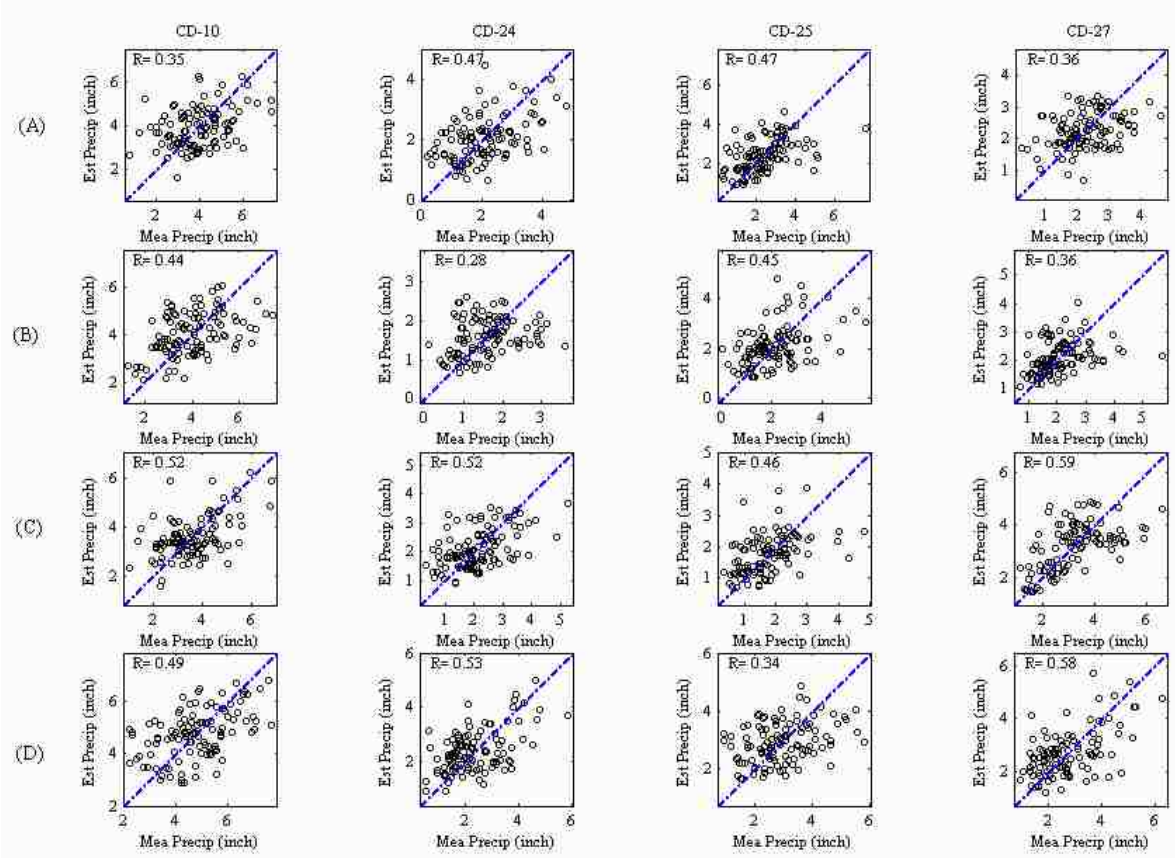


Figure 34: Scatter plots between the measured and disaggregated seasonal precipitation for the selected climate divisions in the upper basin for (A) Autumn, (B) Winter, (C) Spring, and (D) Summer seasons. Dashed line is the 45^0 bisector line.

The scatter plots between the measured and estimated seasonal precipitation for the divisions in the lower basin are shown in Figure 35. Satisfactory predictions are obtained for the selected divisions during autumn season (Figure 35A). The model does a good job in capturing the low values for divisions 1-5 and perfectly captures the extremes for divisions 6 and 7. A very good match is obtained between the measured and predicted winter season precipitation for all the selected divisions (Figure 35B). This is evident by majority of sample points following the bisector line indicating that model does reasonably well in estimating both the low and high precipitation values. For spring season, the model shows acceptable predictions for low precipitation values but fails to

capture the high values (Figure 35C). This is indicated by low values saturated around the bisector line and high values are scattered above the bisector. For summer season, the model does fairly well for low precipitation values for majority of the selected divisions but fails to capture the high values (Figure 35D). The best predictions for the lower basin are obtained during winter season and worst during spring. Moreover, divisions 2, 6, and 7 showed increasing step change (Figure 28B) in winter precipitation, which is efficiently captured by the model and shows the robustness of weighted moving window KNN approach.

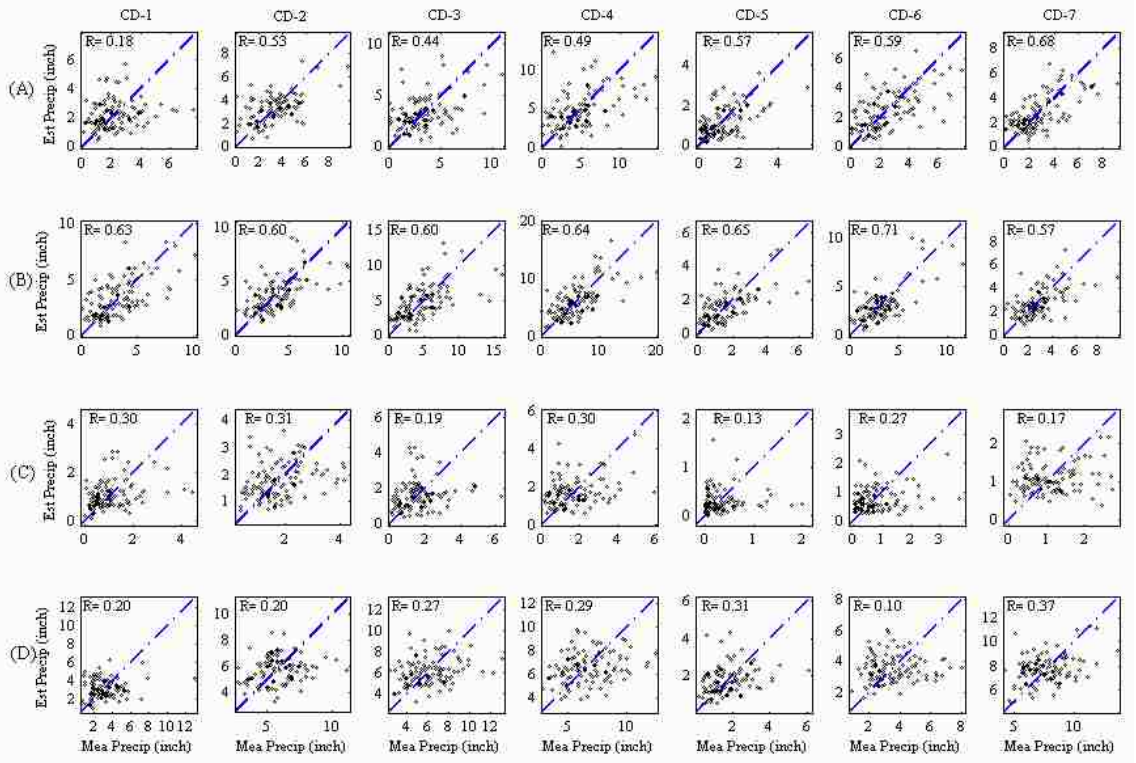


Figure 35: Scatter plots between the measured and disaggregated seasonal precipitation for the selected climate divisions in the lower basin for (A) Autumn, (B) Winter, (C) Spring, and (D) Summer seasons. Dashed line is the 45^0 bisector line.

Box plots depicting the standard statistical properties i.e. mean, standard deviation, and skewness of the simulations for the 29 climate divisions during winter seasons are shown in Figure 36. The box represents the 25th and 75th percentile (interquartile range) values, whereas the whiskers extend from 5th to 95th percentile values. The vertical line inside the box shows the median (50th percentile) value. The statistics of the historic value are represented by the triangle, connected by the dotted line. Historic values lying inside the box is judged as good while increased variability is indicated by a wider box plot. Mean and standard deviations are well reproduced for all the climate divisions as the historic values are captured by the interquartile range. The skew for majority of the climate divisions is preserved but for few climate divisions it is over represented and lies outside the interquartile range. Similar results were obtained for other seasons (results not shown) but the best preservation of the standard distributional statistics was obtained during the winter season.

The results from Figure 34 and Figure 35 show that the model does fairly well in estimating winter and spring season precipitation compared to autumn and summer season. This is of importance because the precipitation exhibits higher degree of variability during winter and spring seasons compared to autumn and summer seasons in the CRB and obtaining satisfactory predictions during these seasons can be useful for water managers. Even the worst estimates obtained during autumn and summer seasons for the upper and lower basin indicate that although the model misses the high precipitation values, it is able to capture the low precipitation values, which can be helpful for water managers for monitoring the low flows and analyzing droughts within the Colorado River Basin. The box plots for the predicted values during winter season

indicate that model performs satisfactorily in preserving the distributional dependence by efficiently capturing the long-term statistical properties.

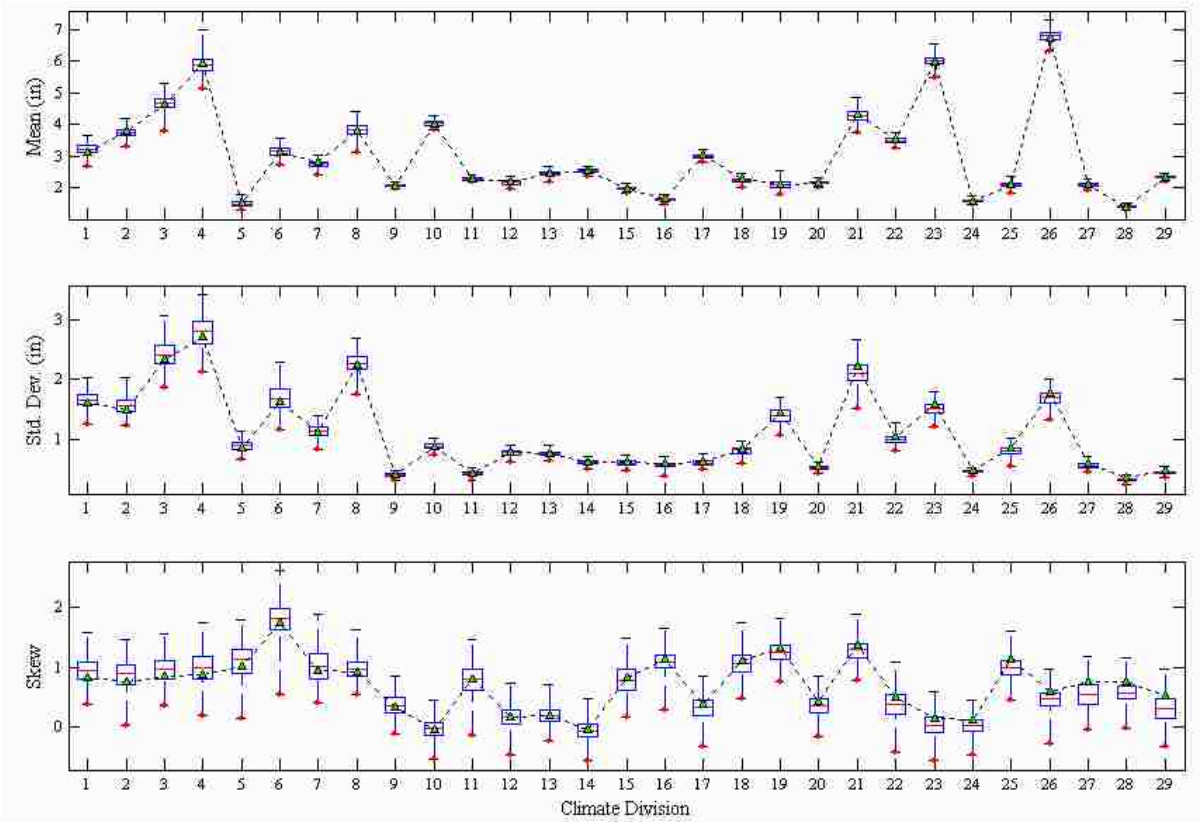


Figure 36: Box plots of seasonal precipitation statistics for the 29 climate divisions during winter season. The box shows the interquartile range (25th – 75th percentile). The whiskers extend from 5th to 95th percentile values. The solid line inside the box shows the median value (50th percentile) and the triangle represents the historic statistic.

4.5.3 Comparison of KNN and Parametric Model (PAR-1)

The simulations of the KNN approach developed in this research are compared with the traditional parametric model (PAR-1). Figure 37 shows the scatter plot between measured and predicted seasonal precipitation values for the selected climate divisions in the upper basin. A poor fit is obtained between the measured and predicted precipitation for autumn (Figure 37A) and summer season (Figure 37D). The model misses both the

low and high values evident by majority of the points scattered around the bisector. During winter (Figure 37B) and spring (Figure 37C) the model performance is satisfactory in estimating low precipitation values for majority of the climate divisions but not the high values. Overall, the range of variability of the predicted values is much lower (in some plots only half) than the range of data; evidenced by the scattered points lying around the lines with angular coefficients much lower than 1 (bisector). Comparing the results obtained using KNN (Figure 34) approach, superiority of KNN approach is noticed over the PAR-1 modeling approach. Similar results were obtained for the divisions in the lower basin (results not shown).

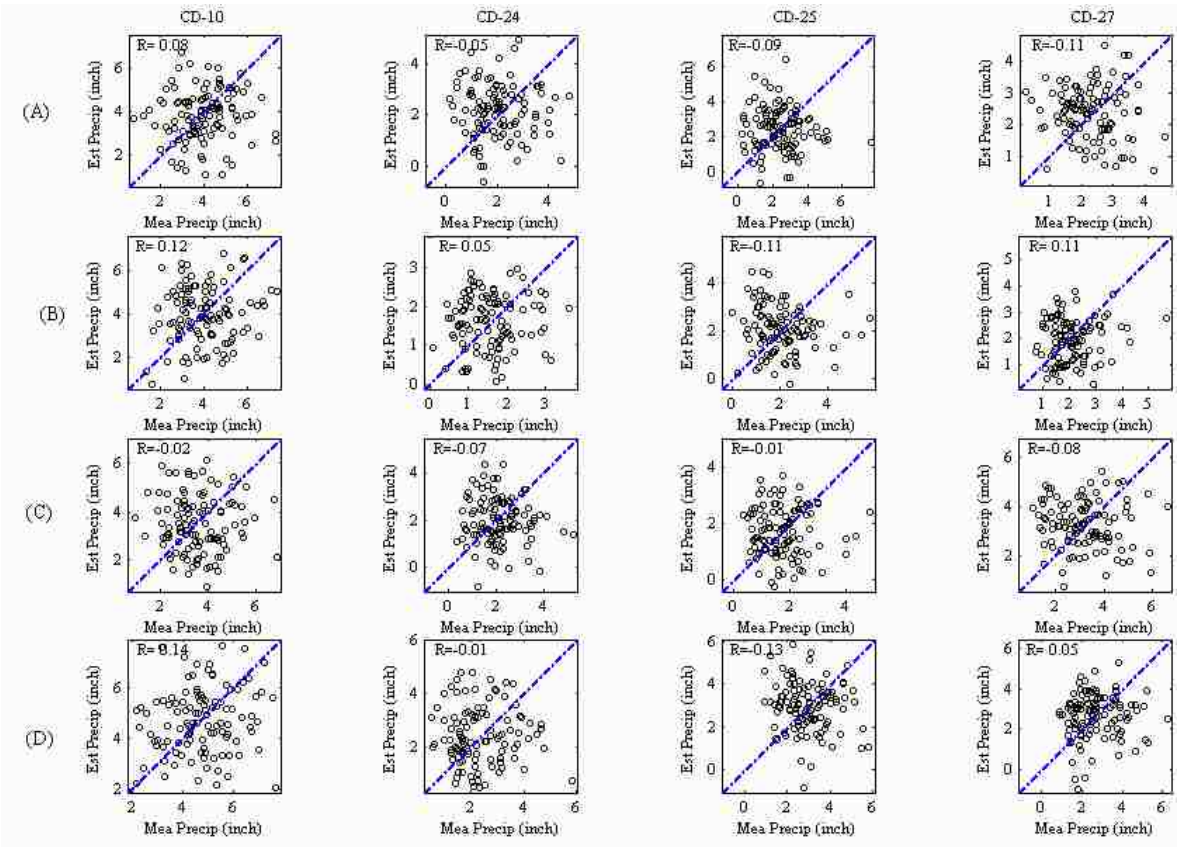


Figure 37: Scatter plot between the measured and disaggregated seasonal precipitation for the selected climate divisions in the upper basin for (A) Autumn, (B) Winter, (C) Spring, and (D) Summer seasons using PAR-1 approach. Dashed line is the 45^0 bisector line.

Previous studies have indicated that parametric models are designed to capture the basic statistical properties but have difficulty in capturing the skewness. Figure 38 shows the box plot depicting the mean, standard deviation, and skewness of the simulations for the 29 climate divisions during winter seasons using PAR-1 model.

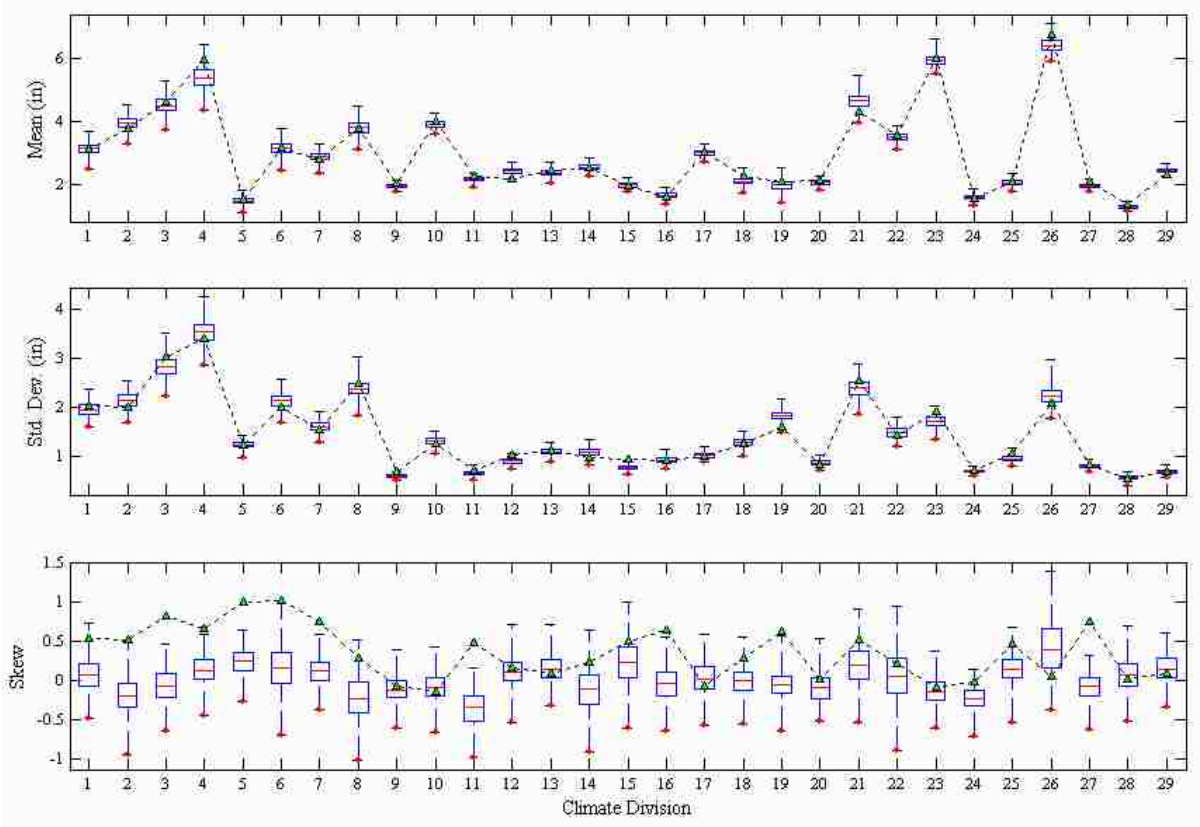


Figure 38: Box plots of seasonal precipitation statistics for the 29 climate divisions during winter season using PAR-1 approach. The box shows the interquartile range (25^{th} – 75^{th} percentile). The whiskers extend from 5^{th} to 95^{th} percentile values. The solid line inside the box shows the median value (50^{th} percentile) and the triangle represents the historic statistic.

The mean and standard deviation are well reproduced for majority of the climate divisions. This is evident by historic statistical values captured within the interquartile range. The skewness coefficient is not well represented by the parametric model indicated by majority of the climate divisions having historic value outside the

interquartile range and for some cases lying outside the whiskers. . Overall, the parametric model results were in agreement with KNN model results as both had best predictions for winter and spring season compared to autumn and summer season. However, the quality of parametric model estimates, based on performance measures, was lower compared to nonparametric model estimates.

Based on the results, it is seen that nonparametric disaggregation techniques such as KNN can be used as an efficient statistical tool for generating seasonal precipitation values within the Colorado River Basin. The spatial maps (Figure 33) and scatter plots (Figure 34 and Figure 35) show that model performs well during winter and spring season compared to autumn and summer seasons. In similar studies, the aggregate value of the hydrological variable is estimated using a model fitted to the data and then the disaggregated values miss the extremes and model cannot be extrapolated to obtain those values (Prairie et al., 2007). The current approach uses a weighted moving window of aggregate values to account for the variability in the data and extreme values, and shows that the model is not period specific and can be used to evaluate the higher degree of variability exhibited by seasonal precipitation.

4.6 Summary and Conclusions

The current study evaluates the trend and step changes in seasonal precipitation over 29 Climate Divisions within the Colorado River Basin over 109 years (1900–2008); and estimates four seasonal precipitation values from water year precipitation using nonparametric KNN disaggregation technique. More than 100 years of precipitation records are examined for evidence of change and to identify the pattern of seasonal precipitation within the CRB. The trend and step changes are evaluated using multiple

statistical tests at three confidence levels of $p \leq 0.10$, $p \leq 0.05$, and $p \leq 0.01$. The trend results indicate that overall there is a decrease in seasonal precipitation in the Upper Colorado River Basin. The main source of flow generation within CRB is the winter precipitation. The winter precipitation falls as snow in the upper elevations of Colorado, Utah, and Wyoming. The decreasing trends in precipitation are pronounced in the northwest mountainous region of the upper basin during the autumn, winter, and spring seasons. The precipitation that falls as snowfall is stored and transferred to relatively dry summers. Temperature increases have lead to more frequent moderate to high intensity non-convective currents causing increase in mid-elevation autumn precipitation within the basin. This is in agreement with results depicting increasing trend in autumn precipitation within the Colorado River Basin. Summer precipitation usually has not much role to play within CRB and has remained unchanged during the period of record as indicated by the trend analysis.

The step change results for seasonal precipitation are similar to the results noted during trend change analysis for autumn, spring and summer seasons but not for winter season. A step increase in winter season precipitation is seen for majority of the climate divisions covering the lower basin due to the warm winter storms. Another possible explanation for an abrupt step change can be the climate variability causing intensification of the hydrologic systems (Huntington, 2006). One notable result from the step change analysis is the decreasing winter precipitation in the northwest mountainous region of the upper basin, which is due to an abrupt step change compared to a gradual change. Figure 29 depicted apparent step changes for selected climate divisions in the upper and lower basin. The plot explicitly indicated the jump (upward/downward) in the

mean value around the year 1976-77. This jump coincides with the warm climate "regime shift" of 1976-77 in the North Pacific Ocean, which had significant biological, ecological, and climatic consequences on the southwestern hydroclimatology. Moreover, the increasing step change for LCRB climate divisions indicated in the current research is in agreement with the previous works of Miller and Piechota (2008). The identification of abrupt step change rather than a gradual trend is important due to significant differences in their implications. Gradual trends are interpreted to continue in the future, whereas abrupt step change signifies that climate system has shifted to a new regime and will likely remain in new regime unless another shift occurs.

With the increasing scientific evidence of climate variability, it becomes necessary to understand fluctuations of long-term changes in associations with climate signatures. ENSO is the most single important factor affecting the interannual variability on global scale and particularly in the western U.S. and the Colorado River Basin. The current study evaluates the impact of ENSO events on the seasonality of the CRB by removing the El Niño events from the historical records and performing the trend analysis on the remaining data set. The increasing trends in autumn precipitation are linked with El Niño, caused by shifting of jet streams in the Pacific Ocean. The summer season precipitation has no signatures of ENSO associated with it. The decreasing trend in winter precipitation within the upper basin is not due to the result of past El Niño events. Contrary to this, the increasing winter season precipitation within the lower basin is linked to the El Niño events. The findings of this analysis are in agreement with other studies that have depicted stronger association of ENSO with the upper basin compared to the lower basin (Piechota and Dracup, 1996, Guido, 2010). The relationship between

upper basin precipitation and seasonal changes in streamflow at Lees Ferry gage is evaluated. An increasing trend is noted for the autumn season streamflow, whereas decreasing trends are noted for winter, spring, and summer season flows. The changes in streamflow are a result of an abrupt step change, not due to gradual trend. To corroborate the seasonal precipitation change results, SNOTEL stations within the CRB are evaluated for trends in March 1, April 1, and May 1 SWE data. The results are in agreement with the trend and step change analysis in seasonal precipitation, which indicated decreasing step changes for winter precipitation for the climate divisions encompassing the northwest mountainous regions within the Colorado River Basin. About 70% of the water originates in this region through winter precipitation and contributes to the flow in the Colorado River. Decreasing precipitation trend in this reason can seriously affect the water resources within the basin. Moreover, the changes in SWE are also due to an abrupt step change and not due to a gradual trend. Similar to Figure 29, the step change (increase/decrease) in April 1 SWE for selected SNOTEL stations is depicted in Figure 31. A clear shift in the regime of snow depth is witnessed around the year 1976-77 for stations indicting step change. Similar to precipitation results, the shift in snow values coincides with the historic regime shift witnessed in the Pacific Ocean. The findings of the analysis excluding ENSO events, with flow at Lees Ferry, and with SWE confirmed that winter season precipitation within the upper basin is decreasing and the decreases are a result of abrupt step change not a gradual trend in the data.

The majority of the water resources systems have been designed and operated on the assumption of stationarity. With the increasing scientific evidence that global climate has changed, is changing and will continue to change the assumption of stationarity may not

be valid (Milly et al., 2008). Detection of long-term change in precipitation trends is important and can assist in evaluating the assumption of stationarity leading to a better planning and management of the water resources within a basin. But for design and operational purposes, water managers are also interested in the precipitation at appropriate temporal scale. Many GCMs models have been developed by numerous researchers for obtaining precipitation estimates but the information available is at a coarser scale and cannot be used for regional scale hydrology. In such cases, statistical disaggregation techniques are an attractive alternative for obtaining continuous realizations of precipitation. Several statistical disaggregation techniques have been used by numerous researchers to transform precipitation from one scale to the other. Precipitation disaggregation in most cases is not an end in itself and provides information to understand and potentially act upon the impacts that are likely to be caused by climate extremes and future climate changes.

The current study used a weighted moving window K-nearest neighbor approach to transform water year precipitation into four seasonal values for 29 climate divisions within the Colorado River Basin. The KNN disaggregation results indicate that the model does a satisfactory job in estimating seasonal precipitation within the Colorado River Basin. The model does a fairly good job in capturing the extremes (low and high values) during winter and spring seasons compared to autumn and summer seasons. The box plots for predicted values during winter season indicate that the model performs satisfactorily in preserving the distributional dependence by efficiently capturing the long-term statistical properties. Overall, best predictions are obtained during winter and spring season compared to autumn and summer season. This is of importance because

majority of the streamflow within the CRB is generated from the winter and late spring precipitation (Mote et al., 2005; Pagano and Green, 2005, Stewart et al., 2005). Better winter and spring precipitation estimates can lead to better streamflow estimates and in turn can help the water managers in planning and management of the water resources within the Colorado River Basin. Superiority of KNN approach is witnessed by comparing the simulations with the traditional parametric approach (PAR-1). The advantages of using the current approach compared to the previous studies are 1) better representation of the temporal variability exhibited by seasonal precipitation; 2) method is not period specific and can be used to generate extreme value.

It should be noted that the distinction between precipitation as rainfall and snowfall is not explicitly addressed in the current study. Snow pack is considered to be dominant in the CRB, making up 63% of the annual precipitation within the upper basin and 39% of the annual precipitation within the lower basin (Serreze et al., 1999). Also, it is important to note that hydrological trends are crucially dependent upon the time period considered in the analysis. Unlike other studies, the aggregate value used in the current study is not generated through a model but historical values are used. The aim of the current study is to show that a stochastic disaggregation technique such as KNN can serve as a useful tool in obtaining satisfactory seasonal precipitation estimates within the CRB, which exhibits high variability. So if an annual aggregate value is obtained from another source such as GCM's, empirical method, or statistical techniques, KNN method can be used to downscale to a temporal resolution useful for regional scale studies. For non stationary time series with a strong seasonal component, many desirable characteristics should be preserved, which can be achieved through reliable and robust disaggregation technique.

Using a simple, efficient and robust technique such as KNN can help in evaluating future climate change scenarios and in efficient water resource planning and management.

The study presented here shows prospects for analyzing the precipitation trends and seasonal variability within the Colorado River Basin, which are of primary importance to the water resource managers in the Southwest and the Bureau of Reclamation. Previous studies (Huntington, 2006 and Miller and Piechota, 2008) have evaluated the trends in precipitation but at different spatial and temporal scales. The current study was successful in capturing the changes (trend and steps) in seasonal precipitation for the entire 20th century within the CRB, which have not been evaluated is the past studies. The current study was also able to provide a distinction between a gradual trend and an abrupt step change not addressed in other studies. The identification of step and gradual trend changes are important for climate change studies, and for hydrological, meteorological and agricultural communities and can help in managing water resources and reservoir operations in the Colorado River Basin. Furthermore, none of the precipitation disaggregation studies have been performed within the Colorado River Basin. Most of the disaggregation studies within CRB have focused on transforming streamflow from one scale to the other. With the increasing and stronger evidence of global warming, changes in precipitation and other climate variables are evident and will be amplified in streamflow (Sankarasubramaniam et al., 2001; Fu et al., 2007). Therefore, developing precipitation disaggregating techniques for CRB are as important as performing streamflow disaggregation. Further research is underway to predict annual precipitation using large scale climate information that can be useful in transforming the aggregate precipitation value into finer resolutions depending on the need of end user.

Acknowledgments

This work was funded by the National Oceanic and Atmospheric Administration (NOAA-SARP) Award # NA07OAR4310324. The authors are thankful to Dr. Glenn Tootle at University of Tennessee for providing the SWE data and to Dr. Navin Twarakavi and Dr. Haroon Stephen for providing useful suggestions during the preparation of this manuscript.

References

Ahmad, S., A. Kalra, and H. Stephen (2010), Estimating soil moisture using remote sensing data: A machine learning approach, *Advances in Water Resources*, *33* (1), 69-80.

Arnold, J. G., and J. R. Williams (1989), Stochastic generation of internal storm structure, *Transactions of the ASAE*, *32* (1), 161-166.

Aziz, O. A., G. A. Tootle, S. T. Gray, and T. C. Piechota (2010), Identification of Pacific Ocean sea surface temperature influences of Upper Colorado River Basin snowpack, *Water Resources Research*, *46*, W07536, doi:10.1029/2009WR008053.

Beamish, R. J., C. M. Neville, and A. J. Cass (1997), Production of Fraser River sockeye salmon (Oncorhynchus nerka) in relation to decadal-scale changes in the climate and the ocean, *Canadian Journal of Fisheries and Aquatic Sciences*, *54*, 543-554.

Bell, T. L. (1987), A space-time stochastic model of rainfall for satellite remote-sensing studies, *Geophysical Research Letters*, *92* (D8), 9631-9643.

Bindlish, R., and A. P. Barros (2000), Disaggregation of rainfall for one-way coupling of atmospheric and hydrological models in regions of complex terrain, *Global and Planetary Change*, *25*, 111-132.

Block, P., and B. Rajagopalan (2007), Interannual variability and ensemble forecast of upper blue Nile Basin Kiremt season precipitation, *Journal of Hydrometeorology*, *8*, 327-343.

Bo, Z., S. Islam, and E. A. B. Eltahir (1994), Aggregation-disaggregation properties of a stochastic rainfall model, *Water Resources Research*, *30* (12), 3423-3435.

Brunetti, M., M. Colacino, and M. Maugeri (2001), Trends in the daily intensity of precipitation in Italy from 1951 to 1996, *International Journal of Climatology*, *21*, 299-316.

Bunting, A. H., M. D. Dennett, J. Elston, and J. R. Milford (1976), Rainfall trends in the West African Sahel, *Quarterly Journal of the Royal Meteorological Society*, *102*, 59-64.

Burian, S. J., S. R. Durrans, S. Tomic, R. L. Pimmel, and C. N. Wai (2000), Rainfall disaggregation using artificial neural networks, *Journal of Hydrologic Engineering*, *5*, 299-307.

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 American Meteorological Society*, *82* (3), 399-415.

Chiew, F., and L. Siriwardena (2005), Trend/change detection software and user guide, CRC for Catchment hydrology, Australia, www.toolkit.net.au/trend.

Chiew, F. H. S., D. G. C. Kirono, D. M. Kent, A. J. Frost, S. P. Charles, B. Timbal, K. C. Nguyen, and G. Fu (2010), Comparison of runoff modelled using rainfall from different downscaling methods for historical and future climates, *Journal of Hydrology*, *387*, 10-23.

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

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

Connolly, R. D., J. Schirmer, and P. K. Dunn (1998), A daily rainfall disaggregation model, *Agricultural and Forest Meteorology*, *92* (2), 105-117.

Diaz, H. F., and G. N. Kiladis (1992), Atmospheric teleconnection associated with the extreme phase of the Southern Oscillation *In* El-Niño: Historical and Paleoclimatic Aspects of the Southern Oscillation, H. F. D. a. V. Markgraf, editor., Cambridge University Press, 7-28.

Easterling, D. R., T. W. R. Wallis, J. H. Lawrimore, and R. R. Heim Jr. (2007), Effects of temperature and precipitation trends on U.S. drought, *Geophysical Research Letters*, *34*, L20707, doi:10.1029/2007GL031541.

Econopouly, T. W., D. R. Davis, and D. A. Woolhiser (1990), Parameter transferability for a daily disaggregation model, *Journal of Hydrology*, *118*, 209-228.

Frei, C., and C. Schar (2000), Detection Probability of Trends in Rare Events: Theory and Application to Heavy Precipitation in the Alpine Region, *Journal of Climate*, *14*, 1568-1584.

Fu, G., S. P. Charles, and F. H. S. Chiew (2007), A two-parameter climate elasticity of streamflow index to assess climate change effects on annual streamflow, *Water Resources Research*, *43*, W11419, doi:10.1029/2007WR005890.

Gangopadhyay, S., M. P. Clark, and B. Rajagopalan (2005), Statistical downscaling using K-nearest neighbors, *Water Resources Research*, *41*, W02024, doi:10.1029/2004WR003444.

Goddard, L., A. G. Barnston, and M. N. Ward (2003), Evaluation of the IRI's "net assessment" seasonal climate forecast: 1997-2001, *Bulletin of the American Meteorological Society*, *84*, 1761-1781.

Gonzalez-Hidalgo, J. C., M. D. Luis, J. Raventos, and J. R. Sanchez (2001), Spatial distribution of seasonal rainfall trends in a western mediterranean area, *International Journal of Climatology*, *21*, 843-860.

Grace, R. A., and P. S. Eagleson (1966), The synthesis of short-time-increment rainfall sequences, Rep. 91, Hydrodynamics Laboratory, Massachusetts Institute of Technology, Cambridge, Mass.

Groisman, P. Y., and D. R. Easterling (1994), Variability and trends of total precipitation and snowfall over the United States and Canada, *Journal of Climate*, *7*, 184-205.

Groisman, P. Y., R. W. Knight, and T. R. Karl (2001), Heavy precipitation and high streamflow in the contiguous United States: Trends in the twentieth century, *Bulletin of the American Meteorological Society*, *82* (2), 219-246.

Guenni, L., and A. Bardossy (2002), A two steps disaggregation method for highly seasonal rainfall, *Stochastic Environmental Research and Risk Assessment*, *16*, 188-206.

Guido, Z. (2010), El Niño-Southern Oscillations: the causes, impacts in the Southwest, and future*, Southwest Climate Outlook*, *9* (1), 1-3.

Guntner, A., J. Olsson, A. Calver, and B. Gannon (2001), Cascade-based disaggregation of continuous rainfall time series: the influence of climate, *Hydrology and Earth System Sciences*, *5* (2), 145-164.

Guttman, N. B., and R. G. Quayle (1996), A historical perspective of U.S. climate divisions, *Bulletin of the American Meteorological Society*, *77*, 293-303.

Hallowed, A. B., and W. W. Wooster (1992), Variability of winter ocean conditions and strong year classes of northeast Pacific groundfish, *ICES Marine Science Symposia*, *195*, 433-444.

Hamlet, A. F., P. W. Mote, M. P. Clark, and D. P. Lettenmaier (2005), Effects of temperature and precipitation variability on snowpack trends in the western United States, *American Meteorological Society*, *18*, 4545-4561.

Hansen, J. W., A. Challinor, A. V. M. Ines, T. Wheeler, and V. Moron (2006), Translating climate forecasts into agricultural terms: advances and challenges, *Climate Research*, *33*, 27-41.

Hansen, J. W., and A. V. M. Ines (2005), Stochastic disaggregation of monthly rainfall data for crop simulation studies, *Agricultural and Forest Meteorology*, *131*, 233-246.

Hare, S. R., and N. J. Mantua (2000), Empirical evidence for North Pacific regime shifts in 1977and 1989*, Progress in Oceanography*, *47* (2-4), 103-145.

Harrold, T. I., A. Sharma, and S. J. Sheather (2003), A nonparametric model for stochastic generation of daily rainfall amounts, *Water Resources Research*, *39* (12), doi:10.1029/2003WR002570.

Haylock, M., and N. Nicholls (2000), Trends in extreme rainfall indices for an updated high quality data set for Australia, 1910-1998, *International Journal of Climatology*, *20*, 1533-1541.

Hennessy, K. J., J. M. Gregory, and J. F. B. Mitchell (1997), Changes in daily precipitation under enhanced greenhouse conditions, *Climate Dynamics*, *13* (9), 667-689.

Hennessy, K. J., R. Suppiah, and C. M. Page (1999), Australian rainfall changes 1910-1995, *Australian Meteorological Magazine*, *48*, 1-13.

Hershenhorn, J., and D. A. Woolhiser (1987), Disaggregation of daily rainfall, *Journal of Hydrology*, *95*, 299-322.

Holbrook, S. J., R. J. Schmitt, and J. S. Stephens (1997), Changes in an assemblage of temperature reef fishes with climate shift, *Ecological Society of America*, *7* (4), 1299-1310.

Hsu, K. L., X. Gao, S. Sorooshian, and H. V. Gupta (1997), Precipitation estimation from remotely sensed information using artificial neural networks, *Journal of Applied Meteorology*, *36*, 1176-1190.

Hu, Q., and S. Feng (2001), Variations of teleconnection of ENSO and interannual variation in summer rainfall in the central United States, *Journal of Climate*, *14*, 2469-2480.

Huntington, T. G. (2006), Evidence for intensification of the global water cycle: Review and synthesis, *Journal of Hydrology*, *319*, 83-95.

Hurrell, J. W. (1995), Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation, *Science*, *269* (5224), 676-679.

Kalra, A., and S. Ahmad (2009), Using oceanic-atmospheric oscillations for long lead time streamflow forecasting, *Water Resources Research*, *45*, W03413, doi:10,1029/2008WR006855.

Kalra, A., T. C. Piechota, R. Davies, and G. A. Tootle (2008), Changes in U.S. streamflow and Western U.S. snowpack, *Journal of Hydrologic Engineering*, *13* (3), 156-163.

Karl, T. R., P. Y. Groisman, R. W. Knight, and R. R. Heim Jr. (1993), Recent variations of snow cover and snowfall in North America and their relation to precipitation and temperature variations, *Journal of Climate*, *6*, 1327-1344.

Karl, T. R., and R. W. Knight (1997), Secular trends of precipitation amount, frequency, and intensity in the United States, *Bulletin of the American Meteorological Society*, *79* (2), 231-241.

Kerr, R. A. (1992), Unmasking a shifty climate system, *Science*, *255* (5051), 1508-1510.

Koutsoyiannis, D. 1988. A disaggregation model of point rainfall. Ph.D. thesis. Pages 310, National Technical University of Athens, Athens.

Koutsoyiannis, D. (1992), A nonlinear disaggregation method with a reduced parameter set for simulation of hydrologic series, *Water Resources Research*, *35* (4), 3175-3191.

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

Koutsoyiannis, D. (2003), Rainfall disaggregation methods: theory and applications, *In* Workshop on Statistical and Mathematical Methods for Hydrological Analysis, Rome, pp.

Koutsoyiannis, D., and T. Xanthopoulos (1990), A dynamic model for short-scale rainfall disaggregation, *Journal of Hydrological Sciences*, *35* (3), 303-321.

Lall, U., B. Rajagopalan, and D. G. Tarboton (1996), A nonparametric wet/dry spell for resampling daily precipitation, *Water Resources Research*, *32* (9), 2803-2823.

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

Lane, W. L. (1982), Corrected parameter estimates for disaggregation schemes, *In* Statistical analysis of rainfall and runoff, edited by V.P.Singh, Water Resour. Publ., Highlands Ranch, pp Colorado.

Lang, S. (1970), Linear Algebra Pages 400 in, Addison-Wesley, Reading, Mass.

Lins, H. F., and J. R. Slack (1999), Streamflow trends in the United States, *Geophysical Research Letters*, *26* (2), 227-230.

Luis, M. D., J. Raventos, J. C. Gonzalez-Hidalgo, J. R. Sanchez, and J. Cortina (2000), Spatial analysis of rainfall trends in the region of Valencia (East Spain), *International Journal of Climatology*, *20*, 1451-1469.

Mann, M. E., U. Lall, and B. Saltzman (1995), Decadal-to-centennial-scale climate variability: Insights into the rise and fall of the Great Salt Lake, *Geophysical Research Letters*, *22*, 937-940.

Mantua, N. J., and S. R. Hare (2002), The pacific decadal oscillation, *Journal of Oceanography*, *5*8, 35-44.

Mantua, N. J., Y. Z. Hare, 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.

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* (1), 183-192.

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* (1), 13-25.

McCabe, G. J., and D. M. Wolock (2002), A step increase in streamflow in the conterminous United States, *Geophysical Research Letters*, *29*, 2185, doi:10.1029/2002GL015999.

Mehrotra, R., and A. Sharma (2006), Conditional resampling of hydrologic time series using predictor variables: A K-nearest neighbour approach, *Advances in Water Resources*, *29* (7), 987-999.

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

Miller, W. P., and T. C. Piechota (2008), Regional analysis of trend and step changes observed in hydroclimatic variables around the Colorado River Basin, *Journal of Hydrometeorology*, *9*, 1020-1034.

Milly, P. C. D., J. L. Betancourt, M. Falkenmark et al., (2008), Stationarity is dead: Whither water management?, *Science*, *319*, 573-574.

Mote, P. W., A. F. Hamlet, M. P. Clark, and D. P. Lettenmaier (2005), Declining mountain snowpack in Western North America, *Bulletin of American Meteorological Society*, *86*, 39-49.

Nash, L. L., and P. H. Gleick (1991), Sensitivity of streamflow in the Colorado Basin to climatic changes, *Journal of Hydrology*, *125*, 221-241.

Nayak, A., D. G. Chandler, D. Marks, J. P. McNamara, and M. Seyfried (2008), Correction of electronic record for weighing bucket precipitation gauge measurements, *Water Resources Research*, *44*, W00D11, doi:10.1029/2008WR006875.

Nayak, A., D. Marks, D. G. Chandler, and M. Seyfried (2010), Long-term snow, climate, and streamflow trends at the Reynolds Creek Experimental Watershed, Owyhee Mountains, Idaho, United States, *Water Resources Research*, *46*, W06519, doi:10,1029/2008WR007525.

New, M., M. Todd, M. Hulme, and P. Jones (2001), Precipitation measurements and trends in the twentieth century, *International Journal of Climatology*, *21*, 1899-1922.

Olsson, J. (1998), Evaluation of a scaling cascade model for temporal rainfall disaggregation, *Hydrology and Earth System Sciences*, *2* (1), 19-30.

Olsson, J., and R. Berndtsson (1998), Temporal rainfall disaggregation based scaling properties, *Water Science and Technology*, *37* (11), 73-79.

Pagano, T., and D. Garen (2005), A recent increase in western U.S. streamflow variability and persistence, *Journal of Hydrometeorology*, *6*, 173-179.

Piechota, T. C., and J. A. Dracup (1996), Drought and regional hydrologic variation in the United States: Association with the El. Nino-Southern Oscillation, *Water Resources Research*, *32* (5), 1359-1373.

Piechota, T. C., J. Timilsena, G. A. Tootle, and H. Hugo (2004), The Western U.S. Drought: How Bad Is It?, *Eos, Transactions, American Geophysical Union*, *85* (32), 301-308.

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

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

Prairie, J. R., B. Rajagopalan, U. Lall, and F. Terrance (2007), A stochastic nonparametric technique for space-time disaggregation of streamflows, *Water Resources Research*, *43*,W03432, doi:10.1029/2005WR004721.

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

Rajagopalan, B., U. Lall, and D. G. Tarboton (1997), Evaluation of kernel density estimation methods for daily precipitation resampling, *Journal of Stochastic Hydrology and Hydraulics*, *11*, 523-547.

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 indices, *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.

Robertson, A. W., A. V. M. Ines, and J. W. Hansen (2007), Downscaling of seasonal precipitation for crop simulation, *Journal of Applied Meteorology and Climatology*, *46*, 677-693.

Rodriguez-Iturbe, I., D. R. Cox, and V. Isham (1987), Some models for rainfall based on stochastic point processes, *In* Proc. R. Soc. Lond., A 410, pp 269-288.

Ropelewski, C. F., and M. S. Halpert (1986), North American precipitation and temperature patterns associated with El-Niño-Southern Oscillation (ENSO), *Monthly Weather Review*, *114*, 2165-2352.

Rupp, D. E., R. F. keim, M. Ossiander, and M. Brugnach (2009), Time scale and intensity dependency in multiplicative cascades for temporal rainfall disaggregation, *Water Resources Research*, *45*, W07409, doi:07410.01029/02008WR007321.

Salas, J. D. (1985), Analysis and modeling of hydrologic time series, *In* Handbook of Hydrology, edited by D.R. Maidment, McGraw-Hill, New York, pp. 19.11-19.72.

Salas, J. D., J. W. Delleur, V. Yevjevich, and W. L. Lane. 1980. Applied modeling of hydrologic time series, 484 pp., Water Resour., Littleton, CO.

Sankarasubramaniam, A., R. M. Vogel, and J. F. Limburner (2001), Climate elasticity of streamflow in the United States, *Water Resources Research*, *37*, 1771-1781.

Sax, J. L., B. H. Thompson, J. D. Leshy, and R. H. Abrams (2000), Legal control of water resources: Cases and Materials, West Group, pp: 956.

Schaake Jr, J. C., M. J. Ganslaw, J. W. Fothergill, and T. E. Harbaugh (1972), Multivariate rainfall generator for annual, seasonal, monthly and daily events, *In* International Symposium on Mathematical Modeling Techniques in Water Resources System, Vol 2, Ottawa, Canada, pp 437-460.

Segond, M.-L., C. Onof, and H. S. Wheater (2006), Spatial-temporal disaggregation of daily rainfall from a generalized linear model, *Journal of Hydrology*, *331* (3-4), 674-689.

Serreze, M. C., M. P. Clark, R. L. Armstrong, and D. A. McGinnis (1999), Characteristics of the western United States snowpack from telemetry (SNOTEL) data, *Water Resources Research*, *35*, 2145-2160.

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

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

Sivakumar, B., S. Sorooshian, H. V. Gupta, and X. Gao (2001), A chaotic approach to rainfall disaggregation, *Water Resources Research*, *37* (1), 61-72.

Srikanthan, R., and T. A. McMahon (2001), Stochastic generation of annual, monthly and daily climate data: A review, *Hydrology and Earth System Sciences*, *5* (4), 653-670.

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

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

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.

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

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

Tarboton, D. G. (1994), The source hydrology of severe sustained drought in the southwestern United States, *Journal of Hydrology*, *161*, 31-69.

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

Timbal, B. (2004), Southwest Australia past and future rainfall trends, *Climate Research*, *26*, 233-249.

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

Tootle, G. A., and T. C. Piechota (2004), Identification of climate teleconnections and forecasting of the upper Tuckee River, *Journal of Nevada Water Resources Association*, *40* (2), 7-19.

Tootle, G. A., and T. C. Piechota (2006), Climate Variability, Water Supply, and Drought in the Upper Colorado River Basin Pages 132-142 *In* Climate Variations, Climate Change, and Water Resources Engineering, J. D. Garbrecht and T. C. Piechota, editors. American Society of Civil Engineers.

Trenberth, K. E., and J. W. Hurrell (1994), Decadal atmosphere-ocean variations in the Pacific, *Climate Dynamics*, *9* (6), 303-319.

Tripathi, S., V. V. Srinivas, and R. S. Nanjundiah (2006), Downscaling of precipitation for climate change scenarios: A support vector machine approach, *Journal of Hydrology*, *330*, 621-640.

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

Webb, R. H., R. Hereford, and G. J. McCabe (2005), Climatic fluctuations, drought, and flow in the Colorado River Basin, *In*: The state of the Colorado River ecosystem in Grand Canyon. U.S. Geological Survey Circular 1282, pp 57-68.

Woolhiser, D. A., and H. B. Osborn (1985), A stochastic model of dimensionless thunderstorm rainfall, *Water Resources Research*, I (4), 511-522.

CHAPTER 5

CONTRIBUTIONS AND RECOMMENDATIONS

5.1 Summary

In an era of changing climate due to natural and anthropogenic factors, it becomes crucial for us to understand the underlying relationships between the variables that describe climatic variability and its effects on the hydrological variables (such as streamflow, precipitation, soil moisture etc.). Climate variability has an enormous effect on the hydrology of our planet. This is especially true in the southwest U.S. where water resources are strained by human demands. The spatial and temporal scale involved as well as the complexity of these relationships makes understanding through physical models a very demanding task. Given this setback, statistical approaches provide a viable alternative to understand these relationships.

In this dissertation, a comprehensive analysis of the applicability of the statistical approaches (such as SVM and KNN) to improve hydrologic predictions in response to climate variability was performed. The area of interest in this dissertation was the Colorado River Basin. Colorado River Basin represents an area with one of the highest population growth rates and as a consequence high strain on water resources. It would be correct to state that no other river in the world is governed more socially, economically, and politically than the Colorado River. Three research questions were addressed through this research.

Research Question # 1: What role do oceanic-atmospheric oscillations play in generating precipitation in the Colorado River Basin, and, can precipitation forecast lead time be increased using oceanic-atmospheric oscillations?

Research Question # 2: What role do oceanic-atmospheric oscillations play in generating streamflow in the Upper Colorado River Basin and can streamflow forecast lead time be increased using oceanic-atmospheric oscillations?

Research Question # 3: How can temporal precipitation disaggregation be achieved when the data possess higher climate variability?

To address first research question, the hypothesis tested was that there is a strong linkage between oceanic-atmospheric oscillations and precipitation within Colorado River Basin and the linkage can be used to improve the annual precipitation forecast lead time. For verifying the hypothesis, a moving period SVM model coupled with oceanic-atmospheric oscillations (PDO, NAO, AMO, and ENSO) from 1900-2007 (108 years) was used to estimate annual precipitation for a 1-year lead time for 17 climate divisions encompassing the Colorado River Basin. The model was evaluated using root mean squared error (RMSE), mean absolute error (MAE), RMSE-observation standard deviation ratio (RSR), correlation coefficient (R), and Nash Sutcliffe coefficient of efficiency (NSE). The results indicated that long-term precipitation predictions for the Upper Colorado River Basin can be successfully obtained using a combination of PDO, NAO, and AMO indices, whereas coupling AMO and ENSO results in improved precipitation predictions for the Lower Colorado River Basin. Furthermore, the results obtained from the analysis were not specific to any period. The precipitation estimates obtained from the SVM model were compared with the ANN and MLR model outputs. The results showed the superiority of SVM approach over the ANN and MLR modeling approaches. Along with estimating precipitation, long-term changes in annual precipitation within CRB were also evaluated. The result showed that the majority of the

Lower Basin is getting wetter, compared to the Upper Basin. The changes in the Lower Basin are attributed to an abrupt step change and not to the gradual trend in the data. The step change coincides with the climate regime "shift" of mid 1970s in the North Pacific Ocean. Overall, the hypothesis tested for research question 1 was true and indicated a strong relationship between oceanic-atmospheric oscillation and annual precipitation for 1-year lead time within Colorado River Basin.

The hypothesis tested for second research question was that there is a strong linkage between oceanic-atmospheric oscillations and streamflow within Upper Colorado River Basin and the linkage can be used to improve the annual streamflow forecast lead time. This resulted in the formulation of a SVM based model to relate the various oceanic-atmospheric oscillations (PDO, NAO, AMO, and ENSO) and annual stream flow for three naturalized gages in the Upper Colorado River Basin. The SVM model was trained with 86 years of data (1906–1991) and tested with 10 years of data (1992–2001). On the basis of correlation coefficient, root means square error, and Nash Sutcliffe Efficiency Coefficient the model showed satisfactory results, and the predictions were in good agreement with measured streamflow volumes. Sensitivity analysis, performed to evaluate the effect of individual and coupled oscillations, revealed a strong signal for ENSO and NAO indices as compared to PDO and AMO indices for the long lead time streamflow forecast. Best streamflow predictions at 3-year lead time were obtained using a combination of NAO and ENSO indices. The results were corroborated by a moving period cross validation analysis. Streamflow predictions from the SVM model were found to be better when compared with the predictions obtained from feed-forward back propagation artificial neural network model and linear regression. Overall, the hypothesis

tested for research question 2 was true and indicated a strong coupled relationship between oceanic-atmospheric oscillation and annual streamflow for 3-year lead time within Upper Colorado River Basin.

The third research question focused on enhancing the temporal resolution of precipitation within Colorado River Basin. The hypothesis tested for this research questions was that the nonparametric trend and step (Mann-Kendall, Spearman's Rho, and Rank Sum) change tests can be used as robust statistical methods to assess the variability in precipitation within the Colorado River Basin. Also, the KNN approach can be used to temporally disaggregate precipitation that exhibits high variability. For this purpose, the current research investigated the long-term changes (Trend and Step) in seasonal precipitation from 1900-2008 (109 years) within CRB and estimated seasonal precipitation using nonparametric K-Nearest Neighbor disaggregation approach. Evaluating the changes in CRB precipitation on seasonal basis is of primary importance to the water resource managers in the Southwest and the Bureau of Reclamation. Several studies have evaluated the trends in CRB precipitation but at different spatial and temporal scales. The current dissertation analyzed the changes in seasonal precipitation for the entire 20^{th} century for 29 climate divisions encompassing the Colorado River Basin. The results indicated a decrease in the upper basin and increase in the lower basin winter precipitation resulting due to an abrupt step change. The effect of El Niño-Southern Oscillations in relation to seasonal precipitation was also evaluated by removing the historic El Niño events. The results indicated that decreasing winter and spring season precipitation trends for the upper basin are not linked with El Niño. The findings of the changes and climate analysis were corroborated by analyzing the trends in

SNOTEL data and streamflow at Lees Ferry gage. A clear identification was made between the step and gradual trend change that are important for climate change studies and for hydrological, meteorological and agricultural communities for managing water resources in the Colorado River Basin. The change results serve the qualitative purpose but for resource planning and management of the water resources, managers are interested in quantitative aspect of precipitation. For this reason a weighted moving window KNN scheme was developed to disaggregate water year precipitation into four seasonal values for the 29 climate division encompassing the Colorado River Basin. The KNN disaggregation results indicate that the model does a fairly good job in capturing the extremes (low and high values) during winter and spring seasons compared to autumn and summer seasons. This is of importance because majority of the streamflow within the CRB is generated from the winter and late spring precipitation. Better winter and spring precipitation estimates can lead to better streamflow estimates and in turn can help the water managers in planning and management of the water resources within the Colorado River Basin. Superiority of the KNN approach is witnessed by comparing the simulations with the traditional parametric approach (PAR-1). The advantages of using the current KNN approach compared to the previous studies are 1) better representation of the temporal variability exhibited by seasonal precipitation; 2) method is not period specific and can be used to generate extreme value. Overall, the hypothesis tested for research question 3 was true and indicated that nonparametric tests can be used to evaluate the long term changes (trend and step) in seasonal precipitation within Colorado River Basin. Additionally, the moving period stochastic KNN disaggregation approach can be used to enhance the temporal resolution of precipitation Colorado River Basin.

5.2 Contributions

The results from research question 1 and 2 indicated **that there is no single climate signal that can be used to explain the hydroclimatology within Colorado River Basin. Coupled response of several oceanic-oscillations in relation to streamflow and precipitation is more pronounced in CRB compared to oscillations individual effects**. Numerous studies have identified Colorado River Basin and other regions in the U.S. showing responses to oceanic-atmospheric oscillations on a seasonal to annual scale. However based on documented literature and prior knowledge, no previous study has attempted to incorporate oceanic-atmospheric oscillations in a SVM model and estimated precipitation and streamflow forward in time. **This research is the first in its kind and resulted in a new data-driven modeling framework that incorporated large scale climate patterns and improved the hydrologic forecast lead time.** The lead time developed in this research would be helpful to water managers in the CRB in managing water systems in response to inter-decadal climate variability. The research also shows prospects for the use of statistical learning theory (SVM) to predict the complex processes (i.e. precipitation and streamflow) that are difficult to understand and simulate using physically based models. **Also, it is noteworthy that no other research has attempted to enhance the temporal resolution of precipitation within Colorado River Basin.** Other studies have focused on disaggregating streamflow from one scale to the other within Colorado River Basin. **This is the first study that used a nonparametric moving window KNN approach for obtaining seasonal precipitation estimates from annual water year value. Furthermore, the results obtained from the analysis are not specific to any period**. The variability exhibited in precipitation is

amplified in streamflow. Therefore, similar to disaggregating streamflow, precipitation disaggregation is also important.

5.3 Limitations

Colorado River Basin is a snowmelt driven watershed with snowpack making up 63% of the annual precipitation within the Upper Basin and 39% of the annual precipitation within the Lower Basin. The current study does not differentiate between precipitation as rainfall and snowfall. Furthermore, the trends detected in the current study are dependent upon the period considered for the analysis. Additionally, the streamflow estimates obtained using the SVM modeling approach is specific to the testing period. Still, there are some variations that remain unexplained in the current dissertation, and which cannot be addressed using the statistical approach.

5.4 Recommendations

Although statistical methods performed satisfactorily in characterizing the relationships between oceanic-atmospheric variables and precipitation and streamflow and improving the temporal resolution of precipitation, a lot is still left to be desired in terms of the accuracy of their predictions. The fields of hydrology and the science of climate change are constantly evolving as researchers and scientists continue to explore new ideas, develop new methods, and make new observations. As a result of this dissertation, several future research investigations could be performed. These include: 1.) This dissertation identified the relationship between coupled oceanic-atmospheric indices and hydrologic variables, which is an important first step. Follow-up studies could focus more on the explanation of physical processes driving this mechanism.

2.) While this dissertation evaluated the annual response of oscillations in relation to hydrologic variables, the phase interaction (i.e. warm and cold) of these oscillations should be considered for future analysis.

3.) Streamflow and precipitation lead times are improved using observed data. Extending the period of record using reconstructed tree ring data may potentially help in improving the forecasts.

4.) This dissertation investigated the coupled and individual impact of most commonly used oscillations i.e. PDO, NAO, AMO, and ENSO. There are other climate indices such as SWE, Geopotential Height Index, PNA, AO, etc. that could be used to represent the hydrologic conditions within CRB.

5.) Several new approaches such as relevance vector machines are emerging and could potentially improve the forecasts. Therefore, the potential of these techniques in improving hydrologic forecast in relation to climate variability could be evaluated.

6.) It is also recommended that similar models be developed for various other locations in U.S. and an effort be made to understand the hydrologic variability within these models as a function of the geography and landscape.

7.) Finally, the ultimate goal would be to move towards a physical based approach that characterizes the relationships between oceanic-atmospheric oscillations and hydrologic variables. It is recommended that the statistical techniques developed here could pave the way for more physically-based models.

VITA

Graduate College
University of Nevada, Las Vegas

Ajay Kalra

Degrees:
Bachelor of Engineering, Civil Engineering, 2002
Punjab Engineering College, India

Master of Science, Civil and Environmental Engineering, 2005
Utah State University, Logan, Utah

Publications:

Kalra, A., and S. Ahmad (2011), Evaluating changes and estimating seasonal precipitation for Colorado River Basin using stochastic nonparametric disaggregation technique*, Water Resources Research,* doi:10.1029/2010WR009118.

Ahmad, S., **A**. **Kalra**, and H. Stephen (2010), Estimating soil moisture using remote sensing data: A machine learning approach, *Advances in Water Resources*, *33* (1), 69-80.

Kalra, A., and S. Ahmad (2009), Using oceanic-atmospheric oscillation for long lead-time streamflow forecasting, *Water Resources Research*, *45*, W03413, doi:10.1029/2008WR006855.

Kalra, A., T. C. Piechota, G. A. Tootle, and R. Davies (2008), Changes in U.S. streamflow and western U.S. snowpack*, Journal of Hydrologic Engineering*, *13* (3), pp. 156-163.

Dissertation Title: Association of Oceanic-Atmospheric Oscillations and Hydroclimatic Variables in the Colorado River Basin

Dissertation Examination Committee:
Chairperson, Sajjad Ahmad, Ph.D., P.E.
Committee Member, David E. James, Ph.D.
Committee Member, Moses Karakouzian, Ph.D., P.E.
Committee Member, Kumud Acharya, Ph.D.
Graduate Faculty Representative, Ashok Singh, Ph.D.