12-2011

Hydroclimatic forecasting in the western United States using paleoclimate reconstructions and data-driven models

Christopher Allen Carrier
University of Nevada, Las Vegas

HYDROCLIMATIC FORECASTING IN THE WESTERN UNITED STATES USING PALEOCLIMATE RECONSTRUCTIONS AND DATA-DRIVEN MODELS

By

Christopher Allen Carrier

A thesis submitted in partial fulfillment

of the requirements for the

Master of Science in Engineering

Department of Civil and Environmental Engineering

Howard R. Hughes College of Engineering

The Graduate College

University of Nevada, Las Vegas

December 2011

THE GRADUATE COLLEGE

We recommend the thesis prepared under our supervision by

Christopher Allen Carrier

entitled

Hydroclimatic Forecasting in the Western United States Using Paleoclimate Reconstructions and Data-Driven Models

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

Master of Science in Engineering
Department of Civil and Environmental Engineering

Sajjad Ahmad, Committee Chair

Jacimaria Batista, Committee Member

Thomas Piechota, Committee Member

Ashok Singh, Graduate College Representative

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

December 2011

ABSTRACT

Hydroclimatic Forecasting in the Western United States Using Paleoclimate Reconstructions and Data-Driven Models

By Christopher Carrier

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

This thesis investigated climate variability and their associated hydrologic responses in the western United States. The western United States faces the problem of water scarcity, where the management and mitigation of available water supplies are further complicated by climate variability. Climate variability associated with the phases of oceanic-atmospheric oscillations has been shown to influence streamflow and precipitation, where predictive relationships have led to the possibility of producing long-range forecasts. Based on literature review, four oceanic-atmospheric oscillation indices were identified in having the most prominent influence over the western United States including the El Niño – Southern Oscillation (ENSO), Pacific Decadal Oscillation (PDO), Atlantic Multidecadal Oscillation (AMO), and North Atlantic Oscillation (NAO). However, these hydroclimatic processes are not fully understood and are difficult to describe in physically-based models. A viable alternative to generating forecasts is through data-driven models, which extract relationships in a dataset of oscillation inputs and hydrologic outputs to build a structured forecasting model.

One of the limitations to using oceanic-atmospheric oscillations in a data-driven model is a short instrumental record from which the model can train on. Data-driven

models often perform better when they are subjected to a larger training dataset. Reconstructions have the potential to extend the period of record by several centuries, which may aid in identifying important hydroclimatological relationships and improving the quality of forecasts.

With this motivation, this study focused on increasing the forecast lead time through the use of reconstructions of oceanic-atmospheric oscillations in the western United States. First, reconstructions of oscillations were investigated to increase the forecast lead time of four streamflow gages in the Upper Colorado River Basin (UCRB) by using the KStar and M5P data-driven models. Secondly, an expanded spatial examination was performed over the western United States for 21 streamflow gages to increase the forecast lead time using the KStar model. Thirdly, different combinations of oceanic-atmospheric oscillations were tested for precipitation forecasts for 20 climate divisions throughout the western United States. Finally, a support vector machine (SVM) was used to increase the streamflow forecast lead time for 21 gages in the western United States.

In order to accomplish this task, a collection of annual time series, processing techniques, testing procedures, and performance measures were used. Reconstructions were available for oscillation indices, streamflow volumes, and climate division precipitation was developed with a common timeframe available as far back as 1658. The instrumental records used ranged from 1900 to 2007 Noise was removed from the dataset using a 3-year, 5-year, and 10-year moving average filter. A 10-fold cross-validation technique was used as opposed to splitting the dataset into training and testing periods so that the entire dataset could be tested and to better capture the non-stationarity of the

dataset. The performance of the models were evaluated through a series of independent measures which include the root mean squared error (RMSE), mean absolute error (MAE), RMSE-standard deviation ratio (RSR), Pearson's correlation coefficient (R), Nash-Sutcliffe coefficient of efficiency (NSE), and linear error in probability space (LEPS) skill score (SK). In addition, all of the models were compared with a multiple linear regression (MLR) model.

The results indicated that the lead time for streamflow forecasts in the Upper Colorado River Basin were increased up to 5 years with the KStar model. In addition, 1-year and 2-year lead-time forecasts with the KStar model were achieved for 21 streamflow gages in the western United States. A 1-year precipitation forecast was also made for 20 climate divisions with the KStar model throughout the western United States and found that the forecasts deteriorated when any of the four oscillations were dropped as predictors. Finally, the SVM model produced streamflow forecasts in the western United States using the raw data at the 1-year and 5-year lead time. In addition, the results indicated that the use of all four oceanic-atmospheric oscillation indices (i.e. ENSO, PDO, AMO, and NAO) provided the best forecasts, and dropping any of the indices yielded inferior results. It was also found that noise removal increased the performance of the model, by aiding in the identification of the oscillation phases.

The contributions made from this research include an extension of the lead-time for streamflow and precipitation forecasts and a better understanding of the effects of climate variability. This study was the first to use reconstructions in a data-driven forecasting model for streamflow and precipitation. Other studies have incorporated reconstructions for use in determining hydroclimatic behaviors and relationships in

comparison to the observed record; however, there have been no previous attempts to use reconstructions with data-driven techniques for forecasting purposes. Overall, this research provided a better understanding of climate variability and their hydrologic responses in the western United States. The forecasting models produced through this research are expected to aid water managers in the long-term planning and management of water resources in the western United States.

ACKNOWLEDGEMENTS

The funding for this work was provided by National Science Foundation (NSF) Award CMMI 0846952.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 Research Problem

The foundation for life on Earth and the growth of human civilization is reliant upon the deliverance of abundant fresh water. Planning and management is essential to provide an assured water supply for sustaining ecological habitats, maintaining public health, driving agriculture and industry, providing hydro-electric power, preserving community stability, and upholding economic growth (Anderson & Woosley, 2005; Reclamation, 2011). Historically, water planning and management has been based on the assumption of stationarity (Milly et al., 2008), where the observed mean and variances in the annual water supply would not change over time. This left water management practices vulnerable to changes in the water supply brought about by climate variability (Kalra & Ahmad, 2009; Milly et al., 2008; Redmond & Koch, 1991; Tootle et al., 2005). Climate variability directly impacts the available water supply from year to year, which places complexity on water management to mitigate water resources. Climate variability influences the hydrologic cycle and is strongly correlated with changes in the magnitude of streamflow and precipitation, (Cayan et al., 1998; Cayan et al., 1999; Hamlet & Lettenmaier, 1999; Hidalgo & Dracup, 2003; Lins & Slack, 1999; Tootle et al., 2005). In addition, climate variability modulates hydrological extremes, such as flood risks and droughts, which cause large-scale natural and socioeconomic destruction (Cayan et al., 1998; Hamlet & Lettenmaier, 1999; Reclamation, 2011).

Water resources management is particularly more important in the western United States, which has become stressed by an explosive population growth, the emerging need of water for environmental and recreational uses, and the production of food and fiber

from western farms and ranches (Anderson & Woosley, 2005). This arid region faces a real possibility of water scarcity as a result of increased water demands and climate variability (Vorosmarty et al., 2000). Securing water for the western United States is challenging because local sources have already been allocated to prior uses, depleted by overuse, or diminished by drought stress (Anderson & Woosley, 2005). Traditionally, new water supplies are secured to meet the growing demands, but this is an unsustainable solution as water resources are limited (Anderson & Woosley, 2005). The result is an increased pressure on existing water supplies, which presents challenges for water resource managers to provide water for the increasing population, irrigated agriculture, power generation, recreation, scenic value, and fish and wildlife habitat (Cody & Hughes, 2003). With the heightened state of water stress in the western United States, there is a serious concern about the future water availability in this region and long-range forecasts become a necessity in order to plan for water allocations and mitigate extreme events.

Several indices relating to oceanic or atmospheric fluctuations have been developed to identify climate variations. There is much interest in these indices, due to their cyclic nature and their teleconnection with the magnitude and frequency of streamflow and precipitation across the globe (Enfield et al., 2001; Hamlet and Lettenmaier, 1999; Tootle et al., 2005). In addition to their spatial influence, there is often a delay of several months to over a year on streamflow or precipitation responses to these oscillations (Hamlet & Lettenmaier, 1999; Kalra & Ahmad, 2009; Piechota et al., 1997). In the western United States, the four most influential oceanic-atmospheric oscillations are the El Niño – Southern Oscillation (ENSO), Pacific Decadal Oscillation (PDO), Atlantic Multidecadal Oscillation (AMO), and North Atlantic Oscillation (NAO).

ENSO has been linked with streamflow and precipitation in the western United States, where a warm El Niño event generally brings below normal moisture in the Pacific Northwest and above normal moisture in the southwestern United States (Cayan et al., 1998; Piechota et al., 1997, Redmond & Koch, 1991). The PDO has been identified with negative precipitation correlations in the Pacific Northwest and positive precipitation correlations in the southwestern United States (Manuta et al., 1997). Hidalgo & Dracup (2003) revealed that both the ENSO and PDO play a significant influence on streamflow and precipitation in the western United States, specifically the Upper Colorado River Basin (UCRB). Enfield et al. (2001) examined the relationship between precipitation variability across the United States in relation to the AMO and found that the AMO has a strong positive correlation in the Pacific Northwest and a strong negative correlation in the southwestern United States. The indices of ENSO, PDO, and AMO have similar regional influences, which suggests that precipitation and streamflow variability is influenced by all three oceanic-atmospheric oscillations. This was further explored by Tootle et al. (2005), which evaluated the individual and coupled effects of ENSO, PDO, AMO, and NAO in relation to streamflow across the United States and found that all four indices influence the streamflow variability. Hunter (2006) also found a significant relationship between the individual and coupled effects of ENSO, PDO, AMO, and NAO on snowfall throughout the western United States. Furthermore, a study by Timilsena et al. (2009) found that the individual and coupled effects of ENSO, PDO, and AMO are prominent in the Colorado River Basin. Based upon the documented literature, oceanic-atmospheric oscillations have a connection with hydrologic variables i.e., streamflow and precipitation. They also have the potential to be used to extend the lead time in

hydrologic forecasting, where a forecast of streamflow or precipitation may be made a year or more in advance from the state of the current oscillations. However, due to the complex nature of these hydroclimatic processes, they are difficult to encapsulate into a physically-based model (Lin et al., 2009).

1.2 Data-Driven Modeling

Alternative to physically-based models are data-driven models, which use artificial intelligence to automatically discover patterns in a dataset in order to form a predictive model (Witten et al., 2011). This approach is useful for hydrologic forecasting, in which underlying relationships can be extracted between a training dataset comprised of the input oscillation indices and the output streamflow or precipitation. The relationships are stored in the structure of the model through an array of mathematical equations. The resulting relationships extracted through the data-driven models can then be used for predictive purposes by entering a new series of inputs into the model.

Data-driven models encompass artificial neural networks (ANNs), support vector machines (SVMs), instance-based learners, and decision trees (Witten et al., 2011). ANNs and SVMs are complex black box models, which develop relationships that are difficult to extract and interpret in the model (Solomatine & Dulal, 2003; Solomatine & Xue, 2004; Solomatine et al., 2008). Instance-based learners and decision trees are much simpler data-driven models, which can produce comparable and often more accurate predictions than complex ANNs (Solomatine & Dulal, 2003; Solomatine & Xue, 2004; Solomatine et al., 2008).In addition, it is much easier to extract relationships from their more transparent algorithmic structures. A drawback of using data-driven models is that

they are often data-hungry, which requires an ample amount of data in order to build a more accurate and robust model (Witten et al., 2011).

1.3 Research Motivation

There is an abundant amount of research that has studied relationships between oscillation indices and their hydrologic responses. However, the results of these studies are based upon an instrumental record that spans 50-100 years. For data-driven modeling, this shorter time series serves as an important limitation in the training of the model (Lin et al., 2009). A possible solution to this limitation is the incorporation of paleoclimate reconstructions. Reconstructions have the potential to extend the period of record by several hundreds of years through the use of proxies, such as tree rings, coral, and ice cores (Jones & Mann, 2004).These proxies provide fixed annual time series that can be correlated with several climatological and hydrological variables including precipitation, streamflow, air temperature, sea-surface temperature, and sea-level pressures (Jones & Mann, 2004).The resulting time series is filtered in order to attenuate the short-range fluctuations and extract the long-range climatic variations (Hidalgo, 2004; McCabe et al., 2004; Probst & Tardy, 1987). With an extended dataset available for these variables, a more flexible data-driven model may be obtained.

An extended period of record was expected to aid data-driven models through several aspects. The most apparent is the increased number of instances available for the data-driven model to train on. This allows for the model to make critical decisions in the forecasts with a higher degree of certainty. In addition, the extended period of record allows the model to further examine the relationships between the oscillation indices and streamflow or precipitation, which may not be as prevalent in the instrumental record.

The use of reconstructions could be useful for hydrologic forecasting using data-driven models.

With this motivation, this thesis used a data-driven approach to increase the streamflow and precipitation forecast lead time by incorporating reconstructions of oceanic-atmospheric oscillation indices. The data-driven models used were M5P, a decision-tree; KStar, an instance-based learner; and a kernel-based multimodel SVM model. The western United States is considered in the analysis because it exhibits the need for long-range forecasts due to hydrologic sensitivity and water stress. There is also ample reconstruction data available for both streamflow and precipitation in the area. The four oceanic-atmospheric oscillations of ENSO, PDO, AMO, and NAO were used in the research because they are commonly associated with hydroclimatology in the western United States and reconstructed data is available for each index. The models were evaluated with several performance measures including the root mean squared error (RMSE), mean absolute error (MAE), RMSE-standard deviation ratio (RSR), Pearson's correlation coefficient (R), Nash-Sutcliffe coefficient of efficiency (NSE), and linear error in probability space (LEPS) skill score (SK). In addition, visual inspections of the models were made through scatter plots, box plots, non-exceedance probability plots, bar charts, and spatial maps.

1.4 Research Objectives

The objective of this research was to increase the forecast lead time for streamflow and precipitation. In order to increase the lead time, a greater understanding of the relationships between oceanic-oscillation indices and their hydrologic

teleconnections was developed. The resulting forecasts are expected to be useful for water managers to aid in the planning and management of water resources. In order to achieve these objectives, the following questions and their corresponding hypotheses were addressed in this research.

Research Question #1: Can the oceanic-atmospheric oscillations be used to improve the streamflow forecast lead time in the Upper Colorado River Basin and can noise removal improve the model?

Hypothesis #1: There is a strong relationship between oceanic-atmospheric oscillations and streamflow variability within the Upper Colorado River Basin and the forecast lead time can be improved as a result. The removal of noise improves the ability of the model to produce accurate forecasts.

Research Question #2: Do oceanic-atmospheric oscillations play an important role in streamflow variability in the western United States, and can the oscillations be used to increase the forecast lead time for gages located in the headwaters and for downstream gages?

Hypothesis #2: The influence of oceanic-atmospheric oscillations is prominent at all gage locations and can be used to forecast streamflow regardless of the variable streamflow volumes observed in the headwater or downstream locations.

Research Question #3: Can the lead time of precipitation forecasts be improved by using oceanic-atmospheric oscillations and do all indices play an equally important role in the forecast model?

Hypothesis #3: Precipitation variability across the western United States is considerably affected by oceanic-atmospheric oscillations and the forecast lead time can be improved. The forecast model improves when a non-significant oscillation index is removed as a predictor.

Research Question #4: Can oceanic-atmospheric oscillations be used in an SVM model to increase the streamflow forecast lead time throughout the western United States without noise removal?

Hypothesis #4: The SVM model is successful in increasing the streamflow forecast lead time from the raw data throughout the western United States.

1.5Research Tasks

The research is presented in a manuscript format. The current chapter comprises of the introduction and formulates the problem statement for this study. Chapter 2 is a manuscript titled “Extending Streamflow Forecast Lead Time Using Proxy Reconstructions for the Upper Colorado River Basin” and addresses Research Question #1. Reconstructions and two data-driven models, M5P and KStar, were used to forecast annual streamflow for 4 gages in the Upper Colorado River Basin at a 1-5 year lead time. The relationships between streamflow and oceanic-atmospheric oscillations in the Upper Colorado River Basin were also explored, as well as the use of smoothing filters to remove noise and their effect on the model performance. Chapter 3 is a manuscript titled “Streamflow Forecasts using Paleoclimate Reconstructions in the Western United States” and addresses Research Question #2. Reconstructions and the KStar data-driven model were used to forecast annual streamflow for 21 gages in the western United States at a 1-

5 year lead time. The relationships between streamflow and oceanic-atmospheric oscillations were further explored on a larger spatial scale and included streamflow gages located in headwater areas. Chapter 4 is a manuscript titled “Long-Range Precipitation Forecasts Using Paleoclimate Reconstructions in the Western United States” and addresses Research Question #3. This chapter uses reconstructions and the KStar data-driven model to forecast annual precipitation for 20 climate divisions in the western United States at a 1-year lead time. This chapter examines the relationships between precipitation and oceanic-atmospheric oscillations and their spatial characteristics throughout the western United States. It also examines the impact of dropping one or more of the oceanic-atmospheric oscillations on the model performance. Chapter 5 is titled “Long-Range Streamflow Forecasts using Support Vector Machines in the Western United States” and explores Research Question #4. This chapter uses reconstructions and an alternative data-driven model, an SVM model, to forecast annual streamflow for 21 gages in the western United States at 1-year and 5-year lead times. Chapter 6 summarizes and concludes this thesis and provides recommendations for future research.

CHAPTER 2: EXTENDING STREAMFLOW FORECAST LEAD TIME USING PROXY RECONSTRUCTIONS FOR THE UPPER COLORADO RIVER BASIN

2.1 Introduction

Streamflow is an important component of the hydrologic cycle that influences water supplies. Long-term streamflow variability dictates the water supply availability in basins with large amounts of storage available, such as the Colorado River Basin (Cayan et al., 1998, Dettinger, 1998).Streamflow variability also affects extreme events, such as floods and droughts. In order to plan for the allocation of water and mitigation of extreme events, the development of a reliable streamflow forecast becomes an important task for hydrologists, meteorologists, water resource engineers, and water managers (Chang & Chen, 2001; Kahya & Dracup, 1993). A reliable streamflow forecast with a long lead time -- on the order of a year or more -- would allow water resource managers to better plan and allocate available water supplies for the forthcoming water year.

One of the most promising routes to developing a long lead time streamflow forecast is through using oceanic-atmospheric oscillations. There is an inherent lag between the observance of oceanic-atmospheric oscillations and their effects on streamflow, which researchers may utilize to provide a forecast lead time of several years (e.g. Hamlet & Lettenmaier, 1999; Kahya & Dracup, 1994; Kalra & Ahmad, 2009). There is ample evidence showing that bimodal oceanic-atmospheric phenomenon, such as El Niño-Southern Oscillation (ENSO), Pacific Decadal Oscillation (PDO), Atlantic Multi-decadal Oscillation (AMO), and North Atlantic Oscillation (NAO), correlate to global streamflow fluctuations and particularly in the western United States (e.g. Beebee & Manga, 2004; Cayan & Webb, 1992; Dettinger et al., 1998; Enfield et al., 2001;

Goodrich, 2007; Hamlet & Lettenmaier, 1999; Hidalgo & Dracup, 2003; Kahya & Dracup, 1993; Knight et al., 2006;Mantua, 1999; Ropelewski & Halpert, 1986; Tootle et al., 2005). These oscillations indicate atmospheric conditions over the Atlantic and Pacific Oceans. Changes in seas surface temperatures (SSTs) and atmospheric pressures identify warm or cool phases, which influence the climate. A number of studies have identified streamflow responses to oceanic-atmospheric oscillations in the Colorado River Basin (e.g. Hidalgo & Dracup, 2003; Piechota et al., 1997; Tootle et al., 2005). Notably, Kalra and Ahmad (2009) incorporated ENSO, PDO, AMO, and NAO in the development of a Support Vector Machine (SVM) model for a three-year lead time streamflow forecast in the Upper Colorado River Basin (UCRB). Other oceanic-atmospheric oscillations, including Arctic Oscillation (AO), East Atlantic pattern (EA), West Pacific Oscillation (WPO), Tropical/Northern Hemisphere pattern (TNH), and Pacific/North American pattern (PNA), are available. However, on the basis of results from previous streamflow studies and documented literature, the oceanic-atmospheric oscillations of ENSO, PDO, AMO, and NAO show the most influence on streamflow patterns within the United States (e.g. Cayan & Webb, 1992; Dettinger et al., 1998; Enfield et al., 2001; Hamlet & Lettenmaier, 1999; Kahya & Dracup, 1993; Knight et al., 2006; Tootle et al., 2005).

From these studies, it is evident that oceanic-atmospheric oscillations do influence streamflow.In fact, there have been attempts to use oscillations as predictors to estimate streamflow (Asefa et al., 2006; Coulibaly et al., 2000; Hamlet & Lettenmaier, 1999; Kalra & Ahmad, 2009; Wood et al., 2002). The development of a forecasting model for long lead time streamflow is difficult due to the challenge in capturing complex

interactions between oceanic-atmospheric oscillations and streamflow (Coulibaly et al., 2000; Gutierrez & Dracup, 2001). In order to capture the dynamics of oceanic-atmospheric oscillations in a long lead time streamflow forecast, robust non-linear data-driven approaches may be used to identify these relationships (Coulibaly et al., 2000). Data-driven techniques encompass artificial neural networks (ANN), SVMs, regression functions, decision trees, and instance-based learners (Witten et al., 2011). Data-driven techniques operate through the automatic or semiautomatic process of discovering patterns within datasets, encapsulating them in a model and using the model to predict what will happen under new situations (Witten et al., 2011). A data-driven model, which is trained with a larger dataset, generally creates a more robust model (Witten et al., 2011). This is because more data allows the model to learn from more examples that helps in improving the model forecasts (Ahmad & Simonovic, 2005; Melesse et al., 2011).

The instrumental record is generally available for less than 100 years, which is an important limitation for data-driven models for use in long lead time streamflow forecasting (Dettinger et al., 1998; Hidalgo & Dracup, 2003). Several studies suggest addressing this limitation with reconstructions using high resolution paleoclimatic proxy indicators, especially tree-ring chronologies (e.g. Brito-Castillo et al., 2003; Dettinger et al., 1998; Hidalgo & Dracup, 2003; Hunter et al., 2006; Prairie et al., 2008). Tree-rings provide an opportunity to extend the period of record as they provide a fixed annual resolution and absolutely dated time series, where moisture and temperature correlate with annual tree-ring widths (Cook, 1992, Jones & Mann, 2004; Stockton & Jacoby, 1976). Dettinger et al. (1998) reveals that proxy indicators can reflect the observed spatial

and temporal characteristics of climatic variability, which allows for the potential to extend streamflow into pre-instrumental periods. Through tree-ring chronologies, several studies produce reconstructed streamflows that extend several hundred years into the past and show strong influences from the oceanic-atmospheric oscillations (e.g. Brito-Castillo, 2003; Cook & Jacoby, 1983; Gou et al., 2007; Graumlich et al., 2003; Gray et al., 2004a; Hidalgo, 2004;Lara et al., 2007; Meko et al., 2007; Smith & Stockton, 1981; Stockton & Jacoby, 1976; Timilsena et al., 2009; Woodhouse et al., 2006). The reconstructions allow for a full examination of the effects of multi-decadal oceanic-atmospheric oscillations such as the PDO and AMO, where a single phase may persist for 20-40 years, which only allows for the observation of one to two full cycles in the instrumental record (Timilsena et al., 2009). Since both the reconstructions and the instrumental record indicate that oceanic-atmospheric oscillations influence streamflow (Gray et al., 2004a; Meko et al., 2007; Woodhouse et al., 2006), it is possible to develop a streamflow forecast model based upon reconstructed oceanic-atmospheric oscillations.

There are no studies that have used paleoclimate proxy reconstructions in a data-driven model. Since reconstructions for oceanic-atmospheric oscillations and streamflow are available for several hundred years, and are continually improving in quality and quantity, it is expected that an improved streamflow forecast model can be developed compared to models trained on a limited period of instrumental record.

This study focused on the UCRB, which exhibits a need for long-term water resources planning and management. This area was chosen because prior research has identified correlations between oceanic-atmospheric oscillations and streamflow within the basin and abundant proxy reconstruction data is available for oceanic-atmospheric

oscillations and streamflow. Streamflow forecasts are of significant importance in the Colorado River Basin, in which water stress is beginning to develop due to growing populations and agriculture industries (Piechota et al., 2004). Therefore, it is critical for water managers to obtain a reliable streamflow forecast to plan for future water allocations, especially in the snowmelt-driven UCRB, which supplies 90% of the annual streamflow (Christensen et al., 2004).

The goal of this research was to increase the lead time for streamflow forecast by using a data-driven model that incorporates both paleoclimate proxy reconstructions and instrumental record. Forecasting models were developed through the KStar and M5Pdata-driven techniques. KStar is a nearest neighbor algorithm with an entropy-based distance function. M5P is a decision tree with the possibility of linear regression at the leaves. High resolution proxy reconstructions encompassing data for several hundred years have not yet been incorporated in a long lead time streamflow forecast model. Reconstructions are available for oceanic-atmospheric oscillations, including ENSO, PDO, AMO, and NAO, and for water year streamflows for four gages in the UCRB. The model was trained and tested using a 10-fold cross validation technique on a dataset containing reconstructions and instrumental records for both the oscillations and streamflow. To filter anomalies and aid in the identification of oscillation phases, a moving average filter was applied to the dataset. The model was set to produce a 1-5 year lead time forecast of the water year streamflow from as early as January 1st of the input year. A forecast evaluation was performed through the mean absolute error (MAE), root mean squared error (RMSE), RMSE-observations standard deviation ratio (RSR), correlation coefficient (R), Nash-Sutcliffe coefficient of efficiency (NSE), and linear error in

probability space skill score (LEPS SK). The results of the KStar and M5P models were compared with a multivariate linear regression (MLR) model. This proposed modeling technique can be potentially useful for long lead time water resources management within the UCRB.

This paper is organized as follows. The study area and the data used for this study are described in Sections 2.2 and 2.3, respectively. Section 2.4 describes the KStar and M5Pdata-driven techniques as well as the modeling framework for this study. Section 2.5 describes the methods for performing a forecast evaluation. The results from the data-driven models are presented in Section 2.6. Finally, Section 2.7 provides a discussion of the findings and concludes the paper.

2.2 Study Area

The Colorado River is a major source of water for much of the arid southwest United States. However, due to increasing population and agricultural activity, planning for future water allocations becomes a necessity in order to meet the escalating water demands. The Colorado River Basin encompasses a total area of 637,000 km^2 and services nearly 30 million people (CRWUA, 2007). It provides water for municipal and industrial purposes, electricity generation, fish and wildlife, recreation, and irrigation of over 7,000 km^2 of agricultural land (CRWUA, 2007). Under the "Law of the River," water is allocated to seven states within the basin including Colorado, Wyoming, Utah, New Mexico, California, Arizona, and Nevada as well as to Mexico (USBR, 2008) (Figure 1). The Colorado River Basin is often viewed as a two-basin system with the gage at Lees Ferry, Arizona, serving as the division between the basins. The basin upstream from Lees Ferry is defined as the UCRB and serves Colorado, Wyoming, Utah,

and New Mexico. The downstream basin is defined as the Lower Colorado River Basin (LCRB) and serves California, Arizona, Nevada, and Mexico. This designation is important because the annual UCRB snowmelt runoff accounts for 90% of the Colorado River flow. However, under the "Law of the River" the LCRB states are apportioned 7.5 MAF (9.25 km^3) per year for beneficial consumptive use, in which the flow at Lees Ferry is not to drop below an aggregate of 75 MAF (92.5 km^3) over a 10 year period regardless of the annual and decadal fluctuations of streamflow in the UCRB (USBR, 2008). Several basin states are experiencing a rising water demand due to a growing population and agriculture industry (Piechota et al., 2004). As a result, the need for a long lead time streamflow forecast becomes critical to better manage the stressed storage in the Colorado River Basin. The ability to produce long lead time streamflow forecasts of several years would provide better water management for a region that is facing water scarcity.

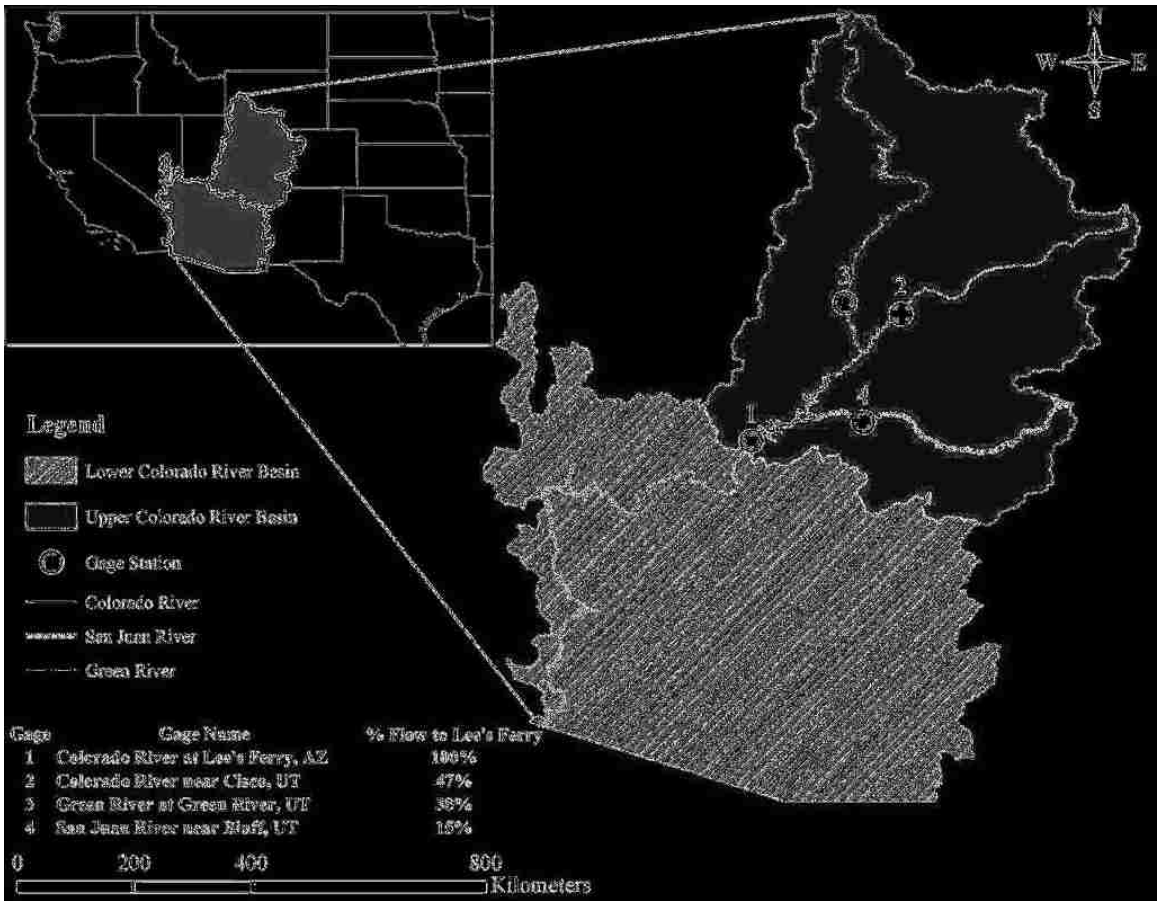

Figure 1: Map showing location of study area and streamflow gages

2.3 Data

The datasets used in the development of long lead time streamflow forecasts included oceanic-atmospheric oscillations and naturalized streamflow data. Paleoclimate proxy reconstructions and instrumental records were available for these datasets. All reconstructions were available through the NOAA Paleoclimatology Program at (http://www.ncdc.noaa.gov/paleo/recons.html). The instrumental records were obtained from several sources, which are outlined in Table 1. There were often multiple reconstruction available, and this study selected those with longer lengths.

Table 1: List of data sources for the streamflow gages and oceanic-atmospheric oscillations used in this study.

Data Sources			
Gage or Oscillation	Source	Period Available	Period in Study
Colorado River at Lees Ferry, AZ; Colorado River near Cisco, UT; Green River at Green River, UT; San Juan River near Bluff, UT	Woodhouse et al. (2006)	1490-2000	1661-1905
	USBR (2009)	1906-2007	1906-2007
ENSO (as SOI)	Jones & Mann (2004)	1650-1980	1661-1905
	Australian Government Bureau of Meteorology (2010)	1876-2010	1906-2007
PDO	Shen et al. (2006)	1470-1998	1661-1905
	JISAO (2010)	1900-2010	1906-2007
NAO	Luterbacher et al. (2001)	1661-2001	1661-1905
	Hurrell (2010)	1865-2010	1906-2007
AMO	Gray et al. (2004b)	1567-1985	1661-1905
	ESRL (2010)	1856-2010	1906-2007

Four streamflow gages located in the Upper Colorado River Basin were used for this study as shown in Figure 1. These gages included (1) Colorado River at Lees Ferry, Arizona (Lees Ferry); (2) Colorado River near Cisco, Utah (Cisco); (3) Green River at Green River, Utah (Green River); and (4) San Juan River near Bluff, Utah (San Juan River). This study made use of tree-ring proxy reconstructions of water year (October – September) streamflow for the four gages in the UCRB from Woodhouse et al. (2006) as shown in Table 1.

The standard chronologies in stepwise regression dataset (Lees-B) was used for the Lees Ferry gage and the full pool of predictor chronologies with stepwise regression was used for the remaining gages in this study. In addition to the paleoclimate reconstructions for streamflow, naturalized observed records for the gages were obtained

from the U. S. Bureau of Reclamation. The update from September 16, 2009 was used in this study.

Yearly oceanic-atmospheric oscillations based upon sea-surface temperature (SST) and sea-level pressure (SLP) were available for ENSO, PDO, NAO, and AMO. The datasets used for these oscillations are described in Table 1. ENSO is defined as the prolonged (1-2 years) warming or cooling of at least 0.5 °C averaged over the east-central tropical Pacific Ocean; this occurs approximately every 4 years, but may vary from 2 – 7 years (Ahrens, 2007; Beebee & Manga, 2004; Kahya & Dracup, 1993). El Niño is the warm phase, which is identified with above-normal streamflow in the southwestern United States. Similarly, La Niña is the cool phase, which is identified with below-normal streamflow in the southwestern United States (Mann et al., 2000; McCabe & Dettinger, 1999). Currently, there is no single measure of ENSO that is universally accepted, but this study makes use of a SLP index known as the Southern Oscillation Index (SOI), which is defined as the difference between the standardized SLP anomalies at Tahiti and Darwin, Australia (Beebee & Manga, 2004; Kahya & Dracup, 1993). For this study, the winter (October to March) SOI was used because it was a better measure of ENSO as opposed to using the entire year (Mann et al., 2000; McCabe & Dettinger, 1999).

The PDO is derived from the leading principal component of monthly sea-surface temperature (SST) anomalies in the North Pacific Ocean, pole ward of 20°N, which exhibits decadal-scale oscillations that typically last between 20 to 30 years (Mantua et al., 1997). Warm phases of the PDO are associated with above-normal precipitation, and

the cool phases are associated with below-normal streamflow in the southwestern United States (Mantua et al., 1997).

The NAO is based on the normalized SLP difference between Ponta Delgada, Azores and Stykkisholmur/Reykjavik, Iceland and oscillates on a large decadal time scale, which may vary annually or may remain in a single phase for several years (Hurrell, 1995). NAO is more predominant over the eastern United States and Europe; however, NAO may be linked to climate variability in the southwestern United States (Hunter et al., 2006).

AMO is an index of SST in the Atlantic Ocean that is calculated from a 10-year running mean of detrended SST anomalies between 0 – 70°N in the Atlantic Ocean (Enfield et al., 2001; Gray et al., 2004b). The AMO exhibits an oscillation that may last 65 – 80 years, with phases that may persist for 20 – 40 years (Enfield et al., 2001; Gray et al., 2004b; Kerr, 2000). Warm phases of the AMO have been linked to below-normal streamflow and drought in the southwestern United States, while the cool phases bring about above-normal streamflow (Gray et al., 2003; McCabe et al., 2004).

Although the inclusion of other oscillations may be beneficial for the development of a streamflow forecasting model, paleoclimate reconstructions are limited to these four oceanic-atmospheric oscillations.

2.4 Methods

This section describes the KStar and M5P data-driven models used in this study as well as the framework to implement the models in the development of streamflow forecasts.

2.4.1 Model Description

This study focused on two algorithms, KStar and M5P. They were featured in the Weka software, which is a data-mining software developed by the University of Waikato in New Zealand (Witten et al., 2011). These algorithms have been successfully utilized in several hydrologic applications, including evaporation estimation, soil moisture content, rainfall-runoff modeling, flood forecasting, and short-term streamflow forecasting (e.g. Elshorbagy et al., 2010; Solomatine & Dulal, 2003; Solomatine & Xue, 2004; Solomatine et al., 2007; Terzi, 2007). Simple instance-based learners and decision trees have been shown to produce equal or better results than complex algorithms, such as artificial neural networks (e.g. Solomatine & Dulal, 2003; Solomatine & Xue, 2004; Solomatine et al., 2007). The KStar and M5P algorithms are described in the following sections.

2.4.2 KStar

KStar is an instance-based learner designed to classify instances through an entropy-based distance measure (Cleary & Trigg, 1995). Like other instance-based learners, the algorithm will compare an instance to pre-classified examples and classify it based on the most similar example. This is performed through a distance function, which determines the similarity between the two instances. In the case of KStar, the distance function is entropy-based. Instead of simply taking the shortest distance between two instances and ignoring all other possible paths, the KStar algorithm determines the distance based upon the probability over all possible paths. This has the advantage of being able to quickly generate relationships and forecasts from the oscillations within seconds, whereas physically-based models often require numerous inputs and tedious calibrations before they can run. Instance-based learners have been successfully utilized

in several hydrologic applications, including evaporation estimation (e.g. Elshorbagy et al., 2010; Terzi, 2007), soil moisture content (e.g. Elshorbagy et al., 2010), rainfall-runoff modeling (e.g. Elshorbagy et al., 2010; Solomatine et al., 2007). Simple instance-based learners have been shown to produce equal or better results than complex algorithms, such as artificial neural networks (e.g. Solomatine et al., 2007).

The detailed discussion of KStar is available in Cleary and Trigg (1995). A brief description of the KStar equations abstracted from Cleary and Trigg (1995) is provided here.

Let **I** be a (possibly infinite) set of instances and **T** a finite set of transformations on **I**. Each t ∈ **T** maps instances to instances: t:**I**→**I**. **T** contains a distinguished memberσ (the stop symbol) which for completeness maps instances to themselves(σ (a) = a). Let **P** be the set of all prefix codes from **T*** which are terminated by σ. Members of **T*** (and so of **P**) uniquely define a transformation on **I**:

$$\bar{t}(a) = t_n\left(t_{n-1}(\dots t_1(a)\dots)\right) \text{ where } \bar{t} = t_1, \dots t_n \quad (1)$$

A probability function p is defined on **T***. It satisfies the following properties:

$$0 \le \frac{p(\bar{t}u)}{p(\bar{t})} \le 1 \quad (2)$$

$$\sum_u p(\bar{t}u) = p(\bar{t}) \quad (3)$$

$$p(\Lambda) = 1 \quad (4)$$

As a consequence it satisfies the following:

$$\sum_{\bar{t}\in P} p(\bar{t}) = 1 \quad (5)$$

The probability function P* is defined as the probability of all paths from instance a to b:

$$P^*(b|a) = \sum_{\bar{t}\in P:\bar{t}(a)=b} p(\bar{t}) \quad (6)$$

It is easily proven that P* satisfies the following properties:

$$\sum_b P^*(b|a) = 1 \quad (7)$$

$$0 \leq P^*(b|a) \leq 1 \quad (8)$$

The K* function is then defined as:

$$K^*(b|a) = -\log_2 P^*(b|a) \quad (9)$$

Figure 2 displays the theoretical partitioning of instance space. In this figure, the circles represent the pre-classified examples. When the instance of interest is compared to the examples, the distance formula will determine which examples are most similar to the instance of interest. The light gray circles are disregarded as they will not affect the result. The remaining black circles are the few prototypical examples that are saved and used for training. The dark filled circles are the most similar instance through the distance function and are the only examples that actually get used in the decision making process.

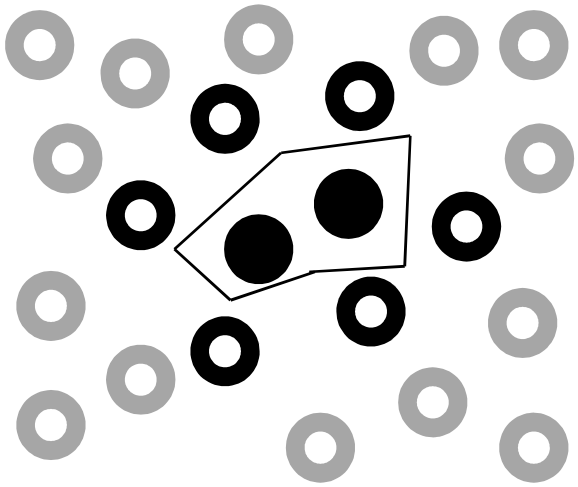

Figure 2: Partitions of an instance space (adopted from Witten et al., 2011)

2.4.3 M5P

M5P is a model tree that incorporates a decision tree approach with the possibility of linear regression functions at the leaves (Wang & Witten, 1996). The M5 model tree was originally developed by Quinlan (1992) and was further improved upon by Wang and Witten (1996) to develop the M5 Prime (M5P) model tree. The M5P model can categorize similar instances and provides a linear regression model for each set of instances (Witten et al., 2011).

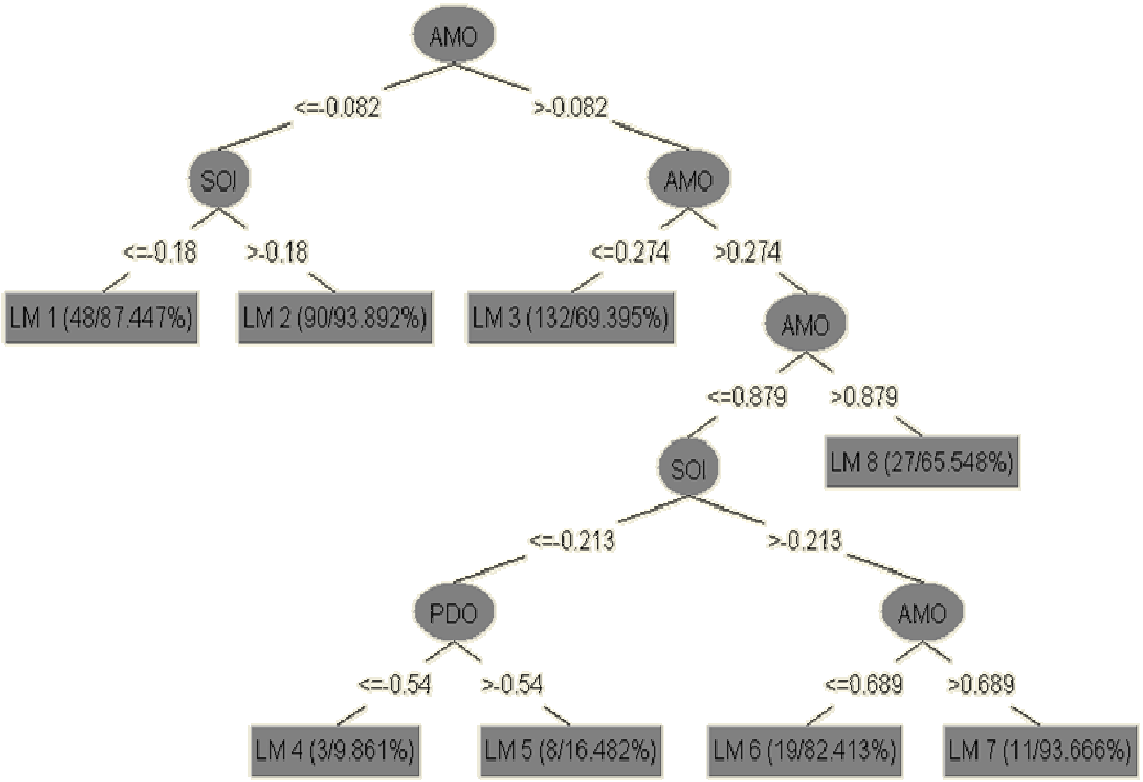

Figure 3: Example of a M5P decision tree. LM 1-8 represent linear regression models.

The M5P decision tree is developed through a series of decision and pruning processes. Figure 3 provides an example of a decision tree. The decision tree uses splitting criterion based upon the standard deviation. This minimizes the variation in the

set of classified values below each node instead of maximizing the information gain at each node. The splitting procedure in M5P stops when the classified values of all instances that reach a node vary only very slightly or if only a few instances remain. The tree then is pruned back from each leaf by turning an inner node into a leaf. The main difference between M5 and M5P is that the leaves in the M5 model are all constants, while the M5P model allows the leaves to be fit with a regression plane (Wang & Witten, 1996). Finally, the tree is smoothed to avoid sharp discontinuities between the sub-trees by combining the linear model predictions at each node along the path back to the root. Interested readers can refer to Wang and Witten (1996) for a detailed description of the M5P model. A brief overview of the M5P model is presented here.

The splitting criterion is based on treating the standard deviation as a measure of error for instances that reach a node, which determines the maximum expected error reduction. The instance that maximizes the expected error reduction is chosen as the splitting node. The standard deviation reduction (SDR) is calculated, as follows.

$$\mathrm{SDR} = \frac{\mathrm{m}}{|\mathrm{T}|} \times \beta(\mathrm{i}) \times \left[\mathrm{sd}(\mathrm{T}) - \sum_{\mathrm{j}\in\{\mathrm{L,R}\}} \frac{|\mathrm{T_j}|}{|\mathrm{T}|} \times \mathrm{sd}(\mathrm{T_j})\right] \quad (10)$$

$$\beta = \mathrm{e}^{7\times\frac{2-\mathrm{k}}{\mathrm{n}}} \quad (11)$$

where m is the number of instances without missing values and T is the set of instances that reach the node. TL and TR are the sets that result from splitting the node according to the chosen instance. β is a correction factor for the M5P model, which improves the model's ability to choose a splitting criterion, where k is the number of values of the

original enumerated attribute and n is the total number of instances. The splitting process only occurs if the produced leaves have at least two examples.

The pruning process performs linear regression on the instances below a node and then drops terms if doing so improves the error estimate. The error function is described, as follows.

$$\frac{(n+v)}{(n-v)} \times \frac{\sum_{examples} |\text{deviation from predicted class value}|}{n} \quad (12)$$

Finally, a smoothing process is applied to the model to smooth each node. The smoothing function is described below.

$$p' = \frac{np+kq}{n+k} \quad (13)$$

where p' is the prediction passed up to the next higher node, p is the prediction passed from the node below, q is the predicted value at the current node, n is the number of instances that reach the node, and k is a constant. The smoothing process substantially increases the accuracy of predictions.

2.4.4 Modeling Framework

A description of the data-driven modeling approach is presented in this section. The oceanic-atmospheric oscillation indices of ENSO, PDO, AMO, and NAO were used as inputs to obtain water year streamflow volumes 1-5 years in advance at select gages in the UCRB. The reconstructed data used in the analysis ranged from 1661 to 1905 and observed data ranged from 1906 to 2007. Both the datasets were combined into one continuous time series with no overlap so that there was no bias introduced into the

dataset. This provided more robust approach as opposed to analyzing the reconstructed and observed dataset individually. The extended time series (i.e. reconstructed and observed) allows the model to learn from a larger training set, where a trend in the reconstructed dataset may or may not be evident in the observed dataset or vice versa. The studies that developed the proxy reconstructions have shown that the overlapping period between the proxy reconstructions and instrumental record dataset were well correlated. The dataset was prepared using a 3-year, 5-year, and 10-year moving average in order to remove the high degree of noise in the dataset. A lead time of 1-5 years was applied to the four streamflow gages in the UCRB to simulate a forecast. The following steps describe the modeling framework.

Step 1. Let i represent the streamflow gage that is the current gage of interest; that is, 1 = Lees Ferry, 2 = Cisco, 3 = Green River, and 4 = San Juan River (i = 1-4). Let $[X_i]$ represent the data matrix that is on the order of T x M, where T represents the number of years and M represents the four oceanic-atmospheric oscillation indices and a streamflow gage. The model used the four oceanic-atmospheric oscillation columns as inputs and the streamflow column as the output. Matrix $[X_i]$ comprises of the complete period of record for this study, including the reconstructions from 1661-1905 and the observed record from 1906-2007 (T = 347). Data was available for the four oceanic-atmospheric oscillations and a streamflow gage in the UCRB (M = 5).

Step 2. Partition the data matrix $[X_i]$ into two sub-matrices, $[A_i]$ and $[B_i]$, so that matrix $[A_i]$ is on the order of N x M and matrix $[B_i]$ is on the order of (T - N) x M, where N is the length of the reconstructed period from 1661-1905 (N = 245).

Step 3. Let j represent the forecast lead time; for example, j = 1 corresponded to a 1-year lead forecast (j = 1-5). The forecast lead time was applied to the column containing the streamflow only. The streamflow column was shifted by j rows; for instance, for year "t" the streamflow column will now correspond to year "t + j". Apply this method to construct sub-matrices $[A_{ij}]$ and $[B_{ij}]$.

Step 4. Let k represent the number of years of a moving average. A 3-year, 5-year, and 10-year moving average is used in this study to smooth the dataset and better capture relations between the indices and streamflow (k = 3, 5, 10). Apply the moving average to construct sub-matrices $[A_{ijk}]$ and $[B_{ijk}]$.

Step 5. Rejoin sub-matrices $[A_{ijk}]$ and $[B_{ijk}]$ to reform matrix $[X_{ijk}]$. This dataset will contain both observed and reconstructed data with the appropriate lead times and moving averages.

Step 6. Apply a 10-fold cross-validation technique to matrix $[X_{ijk}]$. The 10-fold cross-validation technique randomly divides the matrix $[X_{ijk}]$ dataset into 10 further sub-matrices of equal proportions on the order of (T / 10) x M. Under this technique, 9 of these sub-matrices are used for training and the remaining sub-matrix is used for testing. A forecast is made for each instance in the fold held out for testing. This procedure is executed a total of 10 times with each sub-matrix held out for testing in turn. The 10 errors are averaged to yield an overall error. The 10-fold cross-validation technique is a standard practice in data-driven models (Witten et al., 2011).

Step 7. Apply the KStar data-driven model to the set of sub-matrices created in the previous step by using the 10-fold cross-validation technique described in the previous step. This approach allows for an examination of the performance of the data-

driven model's ability to predict future streamflows. The model will output the streamflow predictions, which are represented by matrix $[C_{ijk}]$ on the order of T x 1. Matrix $[C_{ijk}]$ can be compared with the streamflow column of matrix $[X_{ijk}]$ to provide a performance evaluation.

In addition, the modeling framework was applied to MLR. In this manner, a direct comparison between the results of the KStar model can be made with the MLR model. Both KStar and MLR models were subjected to the forecast evaluation discussed in the following section.

The raw data was filtered in order to attenuate the short-range fluctuations and to extract the long-range climatic variations (Hidalgo, 2004; McCabe et al., 2004; Probst & Tardy, 1987). This technique is not new and has been incorporated into studies pertaining to streamflow and climate fluctuations (e.g. Currie, 1996; Hidalgo, 2004; McCabe et al., 2004; Pekarova et al., 2003; Probst & Tardy, 1987; Riehl & Meitin, 1979; Riehl et al., 1979). The current study used a basic 3-year, 5-year, and 10-year moving average filter to aid in the detection of long-range climatic variations. Other data filtering techniques were available; however Probst and Tardy (1987) used three complementary filtering techniques (cumulative deviation method, moving average, and weighted moving average) and indicated that all the three filters produce similar results, although a difference of one or two years for the localization of minima and maxima was sometimes observed.

2.5 Forecast Evaluation

To assess the quality of the streamflow forecasting models, several performance evaluation measures were applied to the model output (Moriasi et al., 2007; Chowdhury & Sharma, 2009). The measures of performance evaluation were the MAE, RMSE,RSR, R, NSE, and LEPS SK. In addition, a visual inspection of the model performance in comparison to the measured data was performed through scatter plots, box plots, and non-exceedance plots.

The performance evaluation focused on determining the accuracy, error, and skill in predicting streamflows compared to the actual streamflows. The streamflow predictions generated by matrix $[C_{ijk}]$ may be represented by p_i, which is the predicted value for the i^{th} instance. The actual streamflows, as provided in the streamflow column in matrix $[X_{ijk}]$, are represented by a_i, which is the actual value for the i^{th} instance. Let n represent the total number of instances. The following measures of performance evaluation are described by Witten et al. (2005) and Moriasi et al. (2007).

The MAE and RMSE are common measures of model performance that indicate error in the same dimensionality of the measured variable, where a value of 0 indicates a perfect fit (Moriasi et al., 2007; Witten et al., 2011). The RSR standardizes the RMSE by using the standard deviation of the observed dataset. This allows the RSR to be used as an error index that can be compared with other results (Moriasi et al., 2007). This is useful for comparing the model performance at the four gages in the UCRB, where streamflow greatly varies. RSR is calculated as the ratio of the RMSE and standard deviation of the measured data as follows:

$$RSR = \frac{RMSE}{STDEV_{meas}} = \frac{\sqrt{\sum_{i=1}^{n}(X_i - Y_i)^2}}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}} \quad (14)$$

RSR ranges from 0 to a large positive value, with 0 indicating a perfect model (Moriasi et al., 2007). Table 2 shows the performance levels for RSR from Moriasi et al. (2007), where RSR is categorized as 'very good' if it is between 0.00 and 0.50, 'good' if between 0.50 and 0.60, 'satisfactory' if between 0.60 and 0.70, and 'unsatisfactory' if greater than 0.70.

Table 2: Performance Rating for RSR, R, and NSE.

Performance Rating	RSR[a]	R	NSE[a]
Very Good	$0.00 \leq RSR \leq 0.50$	$0.85 \leq R \leq 1.0$	$0.75 \leq NSE \leq 1.0$
Good	$0.50 < RSR \leq 0.60$	$0.80 \leq R < 0.85$	$0.65 \leq NSE < 0.75$
Satisfactory	$0.60 < RSR \leq 0.70$	$0.70 \leq R < 0.80$	$0.50 \leq NSE < 0.65$
Unsatisfactory	$0.70 < RSR \leq 1.0$	$0.00 \leq R < 0.70$	$0.00 \leq NSE < 0.50$

[a]RSR and NSE obtained from Moriasi et al. (2007)

R is a measure of the degree of linear relationship between the actual and predicted values (Moriasi et al., 2007). The measurement ranges from 1 (perfect positive correlation), through 0 (no correlation), to -1 (perfect negative correlation). This study defines R as 'very good' if it is greater than 0.85, 'good' if between 0.80 0.85, 'satisfactory' if between 0.70 and 0.80, and 'unsatisfactory' if less than 0.70 (Table 2).

NSE is a normalized measure of the residual variance ("noise") compared to the measured data variance ("information") (Moriasi et al., 2007). NSE indicates how well the predicted and measured data follow a 1:1 relationship (Legates & McCabe, 1999; Moriasi et al., 2007). The NSE is computed as follows:

$$NSE = 1 - \frac{\sum_{i=1}^{n}(X_i - Y_i)^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \quad (15)$$

NSE ranges from 1 to negative infinity, where a NSE greater than 0 is generally an acceptable level of performance because it indicates that the model is a better predictor than the observed mean of the dataset (Legates & McCabe, 1999; Moriasi et al., 2007). Table 2 shows that the performance level for NSE is categorized as 'very good' if it is between 0.75 and 1.00, 'good' if between 0.65 and 0.75, 'satisfactory' if between 0.50 and 0.65, and 'unsatisfactory' if below 0.50 (Moriasi et al.,2007).

To evaluate the quality of forecast LEPS SK is used, which measures the ability of the model to produce accurate predictions that are weighted more for values that are further from the mean (Potts et al., 1996). The LEPS score is defined as

$$S" = 3 * (1 - |Pf - Po| = Pf2 - Pf + Po2 - Po) - 1 \quad (16)$$

where Pf and Po are the forecasted and observed probabilities, respectively. The average skill (SK) is defined as

$$SK = \frac{\sum 100S"}{\sum S"_m} \quad (17)$$

where the summation S" is for all years of record. If S" is positive, $S"_m$ is the sum of the best possible forecast (i.e. Pf = Po) for all years of record. If S" is negative, then $S"_m$ is the sum of the worst possible forecast (i.e. Pf = 1 or 0) for all years of record. SK ranges from -100 to 100, where a SK of 0 represents the climatological score or equivalently, random data (Casey, 1998). A SK score is considered 'good' if it is greater than 10, 'satisfactory' if it is greater than 5, 'poor' if it is below -5, and 'bad' if it is below -10 (Casey, 1998).

2.6 Results

The results are discussed in the following sections. Section 2.6.1 examines the performance of the KStar model when the dataset was subjected to different moving averages and different lead times. Section 2.6.2 describes the M5P model results. Section 2.6.3 provides a comparison of the KStar model and the M5P model as well as a comparison to the MLR model. The results for the Lees Ferry gage are described in full detail and the remaining three gages are discussed in comparison.

2.6.1 KStar Model

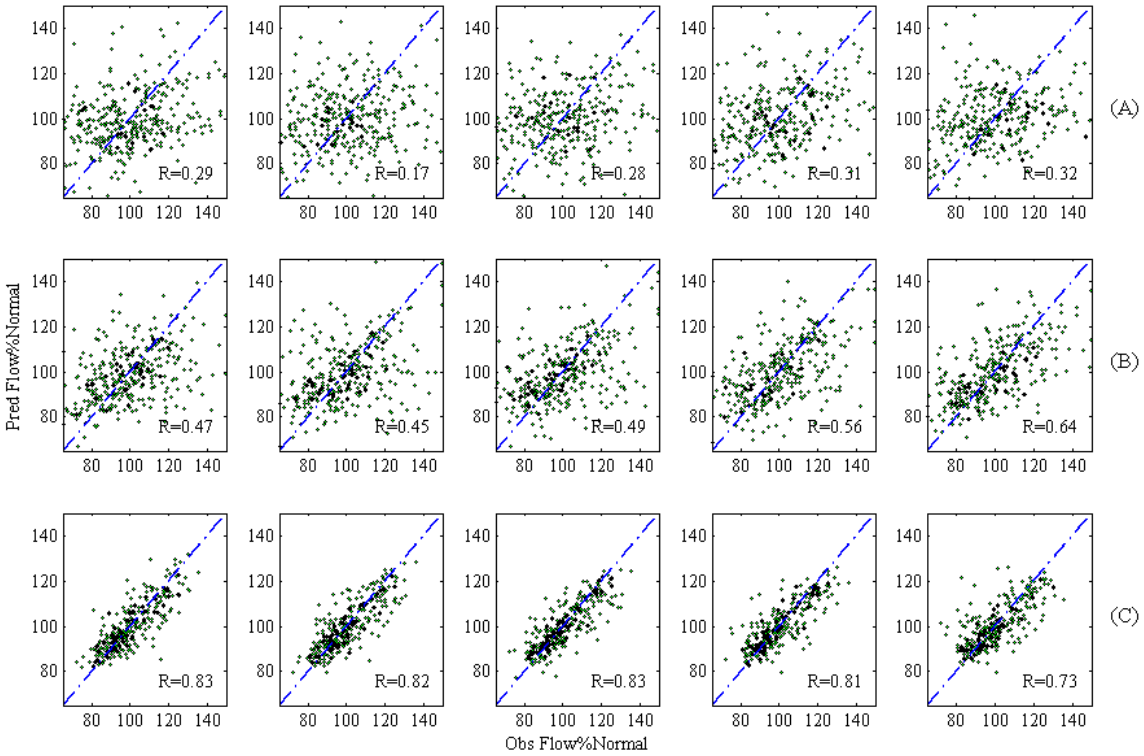

Figure 4: Scatter plots between measured and predicted streamflow for KStar for Lees Ferry under a (A) 3-year moving average, (B) 5-year moving average, and (C) 10-year moving average. The diagonal line is the 45° bisector line. Plots are shown with a 1-year lag on the left and ascending to a 5-year lag on the right.

A visual inspection was performed through scatter plots between the actual and predicted streamflow volumes for the Lees Ferry gage (Figure 4). Figure 4A shows that the streamflow predictions using 3-year moving average filter were widely scattered from the 1:1 bisector for all the lead times. This indicates that the model had difficulty in identifying a relationship between the oscillations and streamflow. This is likely due to the presence of noise, since the moving average window did not sufficiently identify the oscillation signals. Figure 4B shows that the results improved with a 5-year moving average applied to the dataset. The model yielded satisfactory results, with most predictions saturated around the bisector line for all the lead times, indicating a good fit between the measured and predicted streamflow volumes. Still, the 5-year moving average was unable to capture some of the extreme values. A significant improvement in results was noticed when applying a 10-year moving average window to the dataset (Figure 4C). A close match between the measured and predicted streamflow was noticed using the 10-year filter for all the lead times. The sample points were saturated around the bisector line, indicating a good model fit. Additionally, the model did reasonably well in capturing the extreme values, compared to the 3-year (Figure 4A) and 5-year (Figure 4B) results for all the lead times. This was due to the fact that the 10-year moving average filter adequately removed the noise associated with the dataset, making the model efficient at establishing relationships between the large-scale climate oscillations and streamflow. The visual inspection of the model through scatter plots indicates that the 10-year moving average window yielded the best forecasts, while the 5-year moving average and 3-year moving average yielded satisfactory results for 1-5 year lead times.

Table 3: Forecast evaluation for Lees Ferry, using KStar for a 3-year, 5-year, and 10-year moving average.

MA	Lead Time	RMSE	MAE	RSR	R	NSE	LEPS
3	1	3.23	2.52	1.00	0.29	-0.01	14.57
	2	3.46	2.69	1.07	0.17	-0.16	8.86
	3	3.26	2.57	1.01	0.28	-0.02	12.34
	4	3.27	2.51	1.01	0.30	-0.02	16.83
	5	3.22	2.50	0.99	0.32	0.01	15.13
5	1	2.40	1.82	0.91	0.47	0.16	26.54
	2	2.48	1.81	0.94	0.44	0.11	27.64
	3	2.36	1.74	0.89	0.49	0.20	30.46
	4	2.23	1.67	0.85	0.56	0.28	34.46
	5	2.03	1.55	0.77	0.64	0.40	39.82
10	1	1.03	0.77	0.56	0.83	0.69	58.67
	2	1.04	0.78	0.57	0.82	0.68	57.06
	3	1.01	0.74	0.55	0.83	0.69	58.73
	4	1.06	0.78	0.58	0.81	0.66	54.92
	5	1.24	0.89	0.68	0.73	0.53	48.81

[a]In MAF (1 ac-ft = 1233.5 m^3)

Table 3displays the forecast evaluations at different lead times associated with the KStar model results for the Lees Ferry gage. The 3-year moving average produced unsatisfactory results for R ($R < 0.70$), RSR ($RSR > 0.70$), and NSE ($NSE < 0.50$). In addition, the MAE and RMSE were rather high with the MAE ranging from 2.50 – 2.69 MAF and the RMSE ranging from 3.22 – 3.45 MAF. Errors of this magnitude were relatively large for streamflow prediction, where the mean annual flow in the observed record was 15 MAF. However, the LEPS SK show that the model produced 'satisfactory' forecasts ($SK > 5$) for the 2 year lead time, as well as 'good' forecasts ($SK > 10$) for the 1, 3, 4, and 5 year lead times.

The forecast performance improved when a 5-year moving average was applied. Although the model continued to produce unsatisfactory results for R, RSR, and NSE, there was a notable improvement in these values when compared using a 3-year moving

average (Table 3). The 5-year moving average reduces the errors associated with the 3-year moving average, with the MAE ranging from 1.55 – 1.82 MAF and the RMSE ranging from 2.03 – 2.48 MAF. The errors showed improvement compared to the 3-year results but were still on the higher side in the case of streamflow forecasting. The LEPS SK showed that the model produced 'good' forecasts (SK >10) for all lead times, with the 5-year lead time producing the best forecasts when a 5-year moving average was used.

A significant improvement in the forecast performance was experienced when a 10-year moving average was applied compared to the 3-year and 5-year moving averages (Table 3). With the 10-year moving average applied to the dataset, the model produced a satisfactory correlation ($0.70 \leq R < 0.80$) between the predicted and measured data for the 5-year lead time. However, good correlations ($0.80 \leq R < 0.85$) were produced for the remaining lead times. The RSR agrees with the R in which the predictions were satisfactory ($0.60 < RSR \leq 0.70$) for the 5-year lead time and good ($0.50 < RSR \leq 0.60$) for the 1-year to 4-year lead times. The NSE is also in agreement with the R and RSR results in which the 5-year lead time produced satisfactory ($0.50 \leq NSE < 0.65$) model results and the remaining lead times produced good ($0.65 \leq NSE < 0.75$) modeling results. The MAE ranged from 0.74 – 0.89 MAF and the RMSE ranged from 1.01 – 1.24 MAF. These low errors show that model was capable of providing an accurate streamflow forecast. The performance evaluation showed that the noise associated with the dataset is removed and the model performance increased when a larger moving average window was applied. Overall, the KStar model provided good forecasts and confirms the visual analysis (Figure 4) that the 10-year moving average yielded the best streamflow predictions for all lead times.

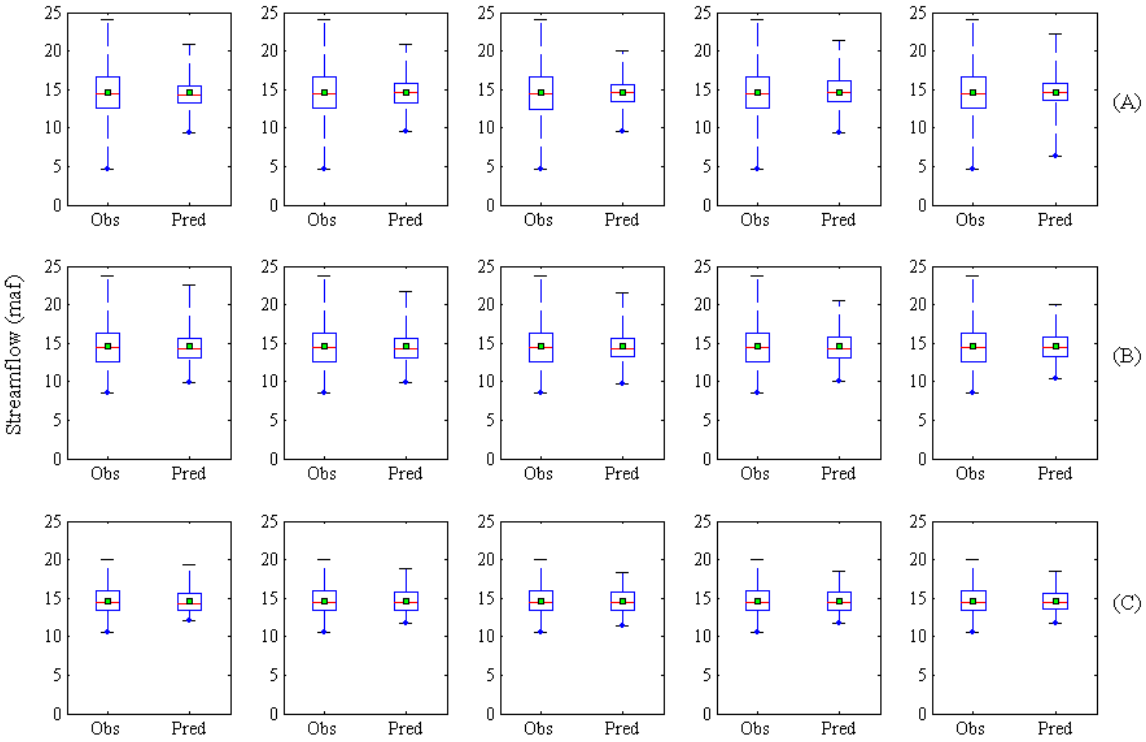

Figure 5: Box plots for KStar for Lees Ferry under a (A) 3-year moving average, (B) 5-year moving average, and (C) 10-year moving average. The box shows the interquartile range (25th-75th percentile). The whiskers extend from the 5th to 95th percentile values. The solid line inside the box shows the median value (50th percentile), and the solid dot represents the mean of the value. Plots are shown with a 1-year lag on the left and ascending to a 5-year lag on the right.

The box plots in Figure 5 examined the ability of the KStar model to capture streamflow variability. Figure 5A shows that the long-term mean of the streamflow was efficiently captured for all the lead times with a 3-year moving average filter. The 3-year moving average filter was able to capture some of the larger variations in streamflow, although it did not do well in capturing the extreme events. Figure 5B shows an improvement in capturing variability as well as extreme events when a 5-year moving average was applied to the dataset as compared to the 3-year moving average window. This demonstrated that a larger moving average window increased the range of the streamflow prediction and aided in better capturing the streamflow variability in relation

to oceanic-atmospheric oscillations. A further improvement was noticed in the ability of the model to capture streamflow variability by applying the 10-year moving average (Figure 5C). The range of predictions within the 25th to 75th percentile increased for all the lead times. Also, it was noticed that the model better represented the extreme values in the dataset. The box plots show that the KStar model performed well in predicting streamflow volumes, especially for flows within the 25th to 75th percentiles (Figure 5).

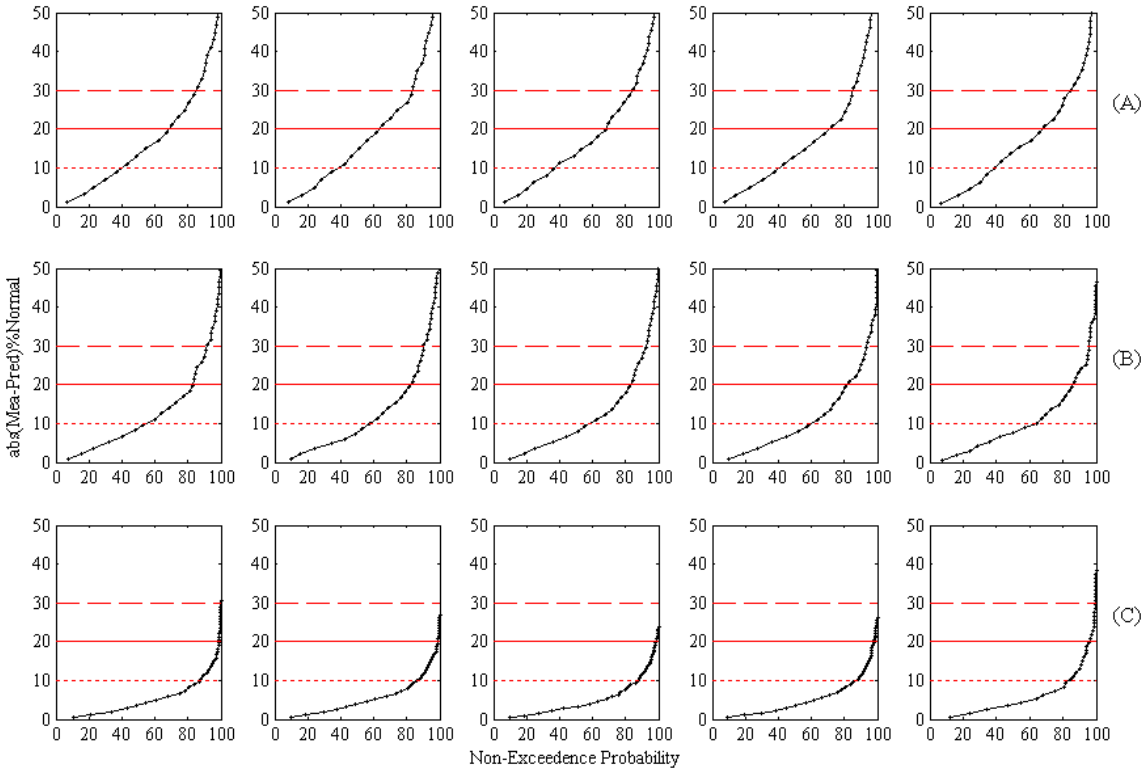

Figure 6: Non-exceedance probability plots for KStar for Lees Ferry under a (A) 3-year moving average, (B) 5-year moving average, and (C) 10-year moving average. The horizontal dotted line shows an error value of 10%, the horizontal solid line shows an error value of 20%, and the horizontal dashed line shows an error value of 30%. Plots are shown with a 1-year lag on the left and ascending to a 5-year lag on the right.

The non-exceedance probability plots in Figure 6displayed the accuracy of the KStar model predictions in terms of their error from the actual streamflow volume.

Figure 6A shows that the 3-year moving average yielded higher errors between the measured and predicted streamflow volumes. The plots show that approximately 35% - 45% of the predictions had an error of 10% or less. The successful predictions increased to 65% - 75% which had an error 20% or less. A further improvement in successful predictions (80% - 85%) was obtained for an error of 30% or less. This implies that the majority of the accurate predictions had higher errors, and may not provide valuable information to water managers. Figure 6B showed an improvement with the 5-year moving average filter, with 50% - 60% of the predictions with an error of 10% or less, 80% - 85% of the predictions with an error 20% or less, and 90% - 95% of the predictions with an error of 30% or less. The predictions under the 5-year moving average were more suitable for a streamflow forecast because it yielded more reliable predictions than the 3-year moving average. Figure 6C showed a considerable improvement in the accuracy of the streamflow forecasts. The plots show that 85% - 90% of the predictions had an error of 10% or less, and very few predictions had errors over 20%. This indicates that the majority of the forecasts had an error of 10% or less; this potentially can help water managers in better planning and management of water resources within the basin. The non-exceedance probability plots highlight that the 10-year moving average was well suited for providing accurate streamflow forecasts, while the 5-year and 3-year moving averages yielded satisfactory forecasts.

Similar to the Lees Ferry gage, other gages (i.e., Cisco, Green River, and San Juan River), were also analyzed with similar moving average and lead times. Since the forecast evaluation for the Lees Ferry gage consistently indicated that the 10-year moving

average provided the best conditions for streamflow forecasting, the results for the other gages were shown only for the 10-year moving average window.

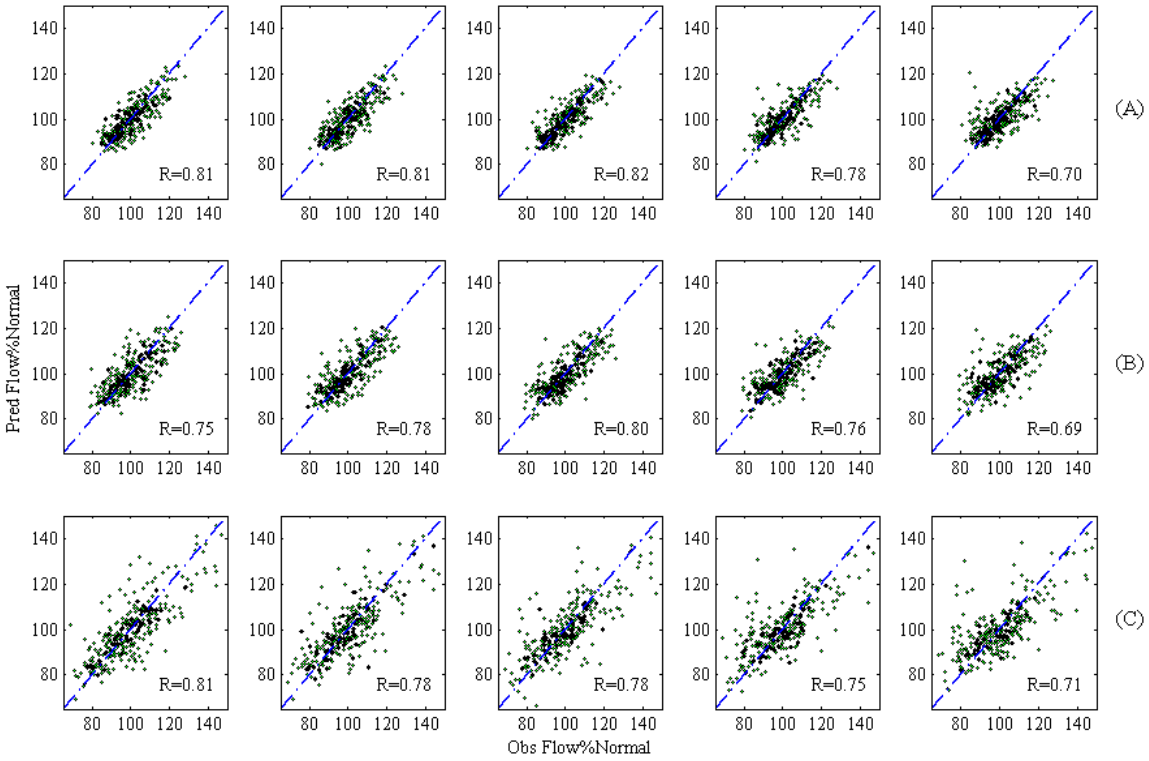

Figure 7: Scatter plots between measured and predicted streamflow for KStar, using a 10-year moving average for (A) Colorado River near Cisco, Utah; (B) Green River at Green River, Utah; and (C) San Juan River near Bluff, Utah. The diagonal line is the 45° bisector line. Plots are shown with a 1-year lag on the left and ascending to a 5-year lag on the right.

Figure 7 displays the scatter plots for the gages at Cisco, Green River, and San Juan River. The plots show that all three gages had a good fit between the predicted and observed streamflow for all lead times, which were similar to the forecasts obtained for the Lees Ferry (Figure 4C). The San Juan River gage in Figure 7C showed a much larger range in the streamflow volumes in terms of the difference from normal flow. This was likely due to the San Juan River having a smaller annual streamflow than the other gages tested, which resulted in larger relative variability to the normal flow.

Table 4: Forecast evaluation, using KStar for Colorado River near Cisco, Green River, and San Juan River; using KStar under a 10-year moving average.

Gage	Lead Time	RMSE	MAE	RSR	R	NSE	LEPS
CO	1	0.35	0.28	0.58	0.81	0.66	55.79
	2	0.36	0.28	0.59	0.81	0.65	54.96
	3	0.34	0.25	0.57	0.82	0.68	59.30
	4	0.38	0.27	0.63	0.78	0.61	54.24
	5	0.44	0.31	0.72	0.70	0.48	46.39
GR	1	0.35	0.27	0.66	0.75	0.56	51.51
	2	0.33	0.25	0.63	0.78	0.60	54.02
	3	0.32	0.24	0.60	0.80	0.63	54.93
	4	0.34	0.26	0.65	0.76	0.58	50.23
	5	0.38	0.28	0.73	0.69	0.47	46.10
SJ	1	0.18	0.13	0.59	0.81	0.65	57.59
	2	0.19	0.14	0.63	0.78	0.61	53.62
	3	0.19	0.14	0.63	0.78	0.60	54.86
	4	0.20	0.14	0.66	0.75	0.56	49.74
	5	0.21	0.15	0.70	0.71	0.51	46.42

[a]In MAF (1 ac-ft = 1233.5 m^3)

Table 4 displays the forecast evaluation for the 3 gages. For the Cisco gage, the model produced 'good' correlation values for the 1-year, 2-year, and 3-year lead times and 'satisfactory' correlations for the 4-year and 5-year lead times per the performance measures shown in Table 2. The RSR and NSE for the Cisco gage showed a similar forecast evaluation with 'good' performance ratings obtained for the 1-year to 3-year lead times and 'satisfactory' performance ratings obtained for the 4-year and 5-year lead times. The MAE was low, ranging from 0.25 – 0.31 MAF. The RMSE was low as well, and ranged from 0.34 – 0.44 MAF. These errors indicated that the streamflow forecasts at Cisco gage were accurate. This was confirmed through the LEPS SK with all lead times indicating a good modeling result (LEPS SK > 10).

The forecast evaluation for the Green River gage shows 'satisfactory' to 'good' results (Table 4). Following the performance ratings in Table 2, the R, RSR, and NSE

showed that the model produced satisfactory predictions for all of the lead times except for the 3-year lead time, which yielded ‘good’ predictions. The MAE ranged from 0.24–0.28 and the RMSE ranged from 0.32 – 0.38. These errors associated with the Green River gage were lower when compared to the Cisco gage. This was due to the difference in the streamflow volumes, since the Cisco gage had a higher yearly streamflow volume than the Green River gage. The LEPS SK indicated a good model performance for all lead times at the Green River gage. However, the streamflow forecasts at Cisco gage were better for all lead times.

The forecast evaluation for the San Juan River gage also produced a ‘satisfactory’ to ‘good’ model performance as well (Table 4). The R, RSR, and NSE indicated that the model produced ‘satisfactory’ predictions for all of the lead times except for the 1-year lead time, which yields ‘good’ modeling results. The MAE was very small, ranging from 0.13 – 0.15 MAF. The RMSE was small as well and ranges from 0.18 – 0.21 MAF. The MAE and RMSE were much lower than the other gages due to the San Juan River having a much smaller annual streamflow volume than the other gages. The LEPS SK showed that a ‘good’ model performance was obtained at all lead times for the San Juan River gage.

The low errors associated with all of the tested streamflow gages showed the robustness and confidence in the model in providing accurate streamflow forecasts within the UCRB. The performance evaluation showed that the model performed well regardless of the streamflow volume. Of the four gages, the KStar model performed best for the Lees Ferry gage. This was of importance because Lees Ferry lies on the hydrologic divide of the upper basin and lower basin, and the water supply to the lower basin is

governed by the flow at Lees Ferry. Overall, the results showed that the KStar model was successful in capturing the relationship between large-scale climate patterns and streamflow within the UCRB. 'Satisfactory' to 'good' streamflow predictions were obtained for 1-5 year lead times with a 10-year moving average filter applied to the dataset.

2.6.2 M5P

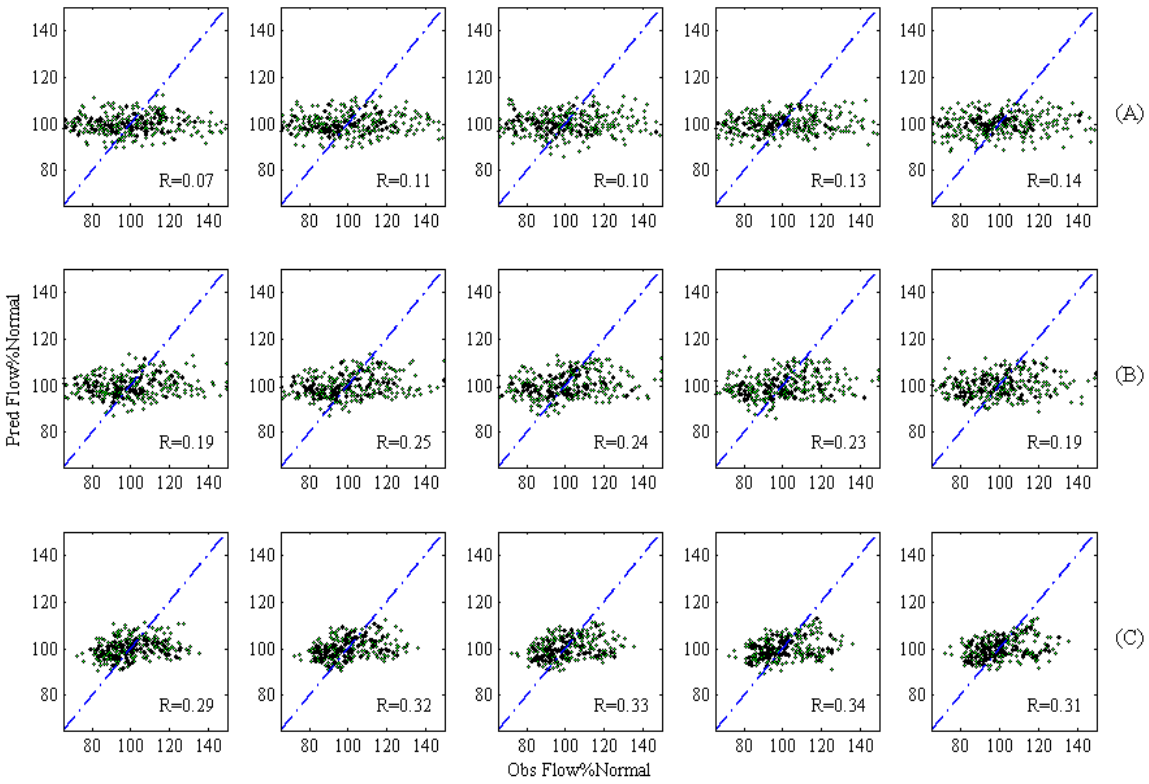

Figure 8: Scatter plots between measured and predicted streamflow for M5P for Lees Ferry under a (A) 3-year moving average, (B) 5-year moving average, and (C) 10-year moving average. The diagonal line is the 45° bisector line. Plots are shown with a 1-year lag on the left and ascending to a 5-year lag on the right.

Figure 8 displays the scatter plots between the actual and predicted streamflow volumes for M5P. The forecast evaluation section explains that the scatter plots show the agreement between predicted and observed values with the 45° bisector line representing

a perfect fit. Figure 8A shows that the streamflow predictions using the 3-year moving average filter were saturated around the predicted 100% normal flow line for all lead times. This indicates that the model was unable to identify a relationship between the oceanic-atmospheric oscillations and streamflow, and returned the mean streamflow as the prediction. The presence of noise and outliers may have hindered the model from identifying hidden relationships. Figure 8B shows a similar result with the streamflow predictions using the 5-year moving average, where the predictions were saturated around the mean streamflow. The 5-year moving average was still unable to identify a relationship between the oceanic-atmospheric oscillations and streamflow. However, the 5-year moving average did perform better than the 3-year moving average. Figure 8C shows a significant improvement in the forecast when a 10-year moving average window was applied to the dataset. The predictions were more saturated around the bisector line, indicating a good model fit. With the removal of the noise with the 10-year moving average filter, the M5P model was capable of identifying critical points for nodes in the development of the decision tree. Without the moving average filter, M5P was unable to categorize similar instances and yielded unsatisfactory predictions because there was too much noise in the dataset. The visual inspection of the model through the scatter plots in Figure 8 indicated that the 10-year moving average window yielded better forecasts, while the 5-year moving average and 3-year moving average were unable to provide satisfactory results for 1-5 year lead times.

Table 5: Forecast evaluation for Lees Ferry, using M5P for a 3-year, 5-year, and 10-year moving average.

MA	Lead Time	RMSE	MAE	RSR	R	NSE	LEPS
3	1	3.27	2.53	1.02	0.08	-0.03	3.29
	2	3.21	2.52	0.99	0.13	0.01	2.67
	3	3.20	2.49	0.99	0.19	0.02	5.63
	4	3.21	2.50	0.99	0.15	0.01	4.02
	5	3.23	2.50	1.00	0.12	0.00	3.56
5	1	2.57	2.04	0.98	0.23	0.05	7.28
	2	2.56	2.01	0.97	0.24	0.05	8.14
	3	2.52	1.96	0.96	0.29	0.08	10.17
	4	2.52	1.98	0.96	0.28	0.08	10.38
	5	2.58	2.01	0.98	0.20	0.03	7.51
10	1	1.56	1.24	0.85	0.53	0.28	24.22
	2	1.51	1.17	0.82	0.57	0.33	27.05
	3	1.43	1.11	0.78	0.63	0.39	30.95
	4	1.48	1.15	0.81	0.58	0.34	27.48
	5	1.55	1.20	0.85	0.52	0.27	22.83

[a]In MAF (1 ac-ft = 1233.5 m^3)

The M5P forecast evaluation for the Lees Ferry gage is shown in Table 5. The results indicate an 'unsatisfactory' performance through R ($R < 0.70$), RSR ($RSR > 0.70$), and NSE ($NSE < 0.50$) for all combinations of lead times and moving average filters. This indicates that the M5P model was unable to produce accurate predictions for use in streamflow forecasting. Although the model produced 'unsatisfactory' forecasts, the model did offer some important insight. The results were shown to steadily improve when larger moving average filters were applied to the dataset. The LEPS SK indicates that a 'satisfactory' model (LEPS SK > 5) was produced at the 3-year and 4-year lead times when a 3-year moving average was applied to the dataset. When a 5-year moving average was applied, the LEPS SK showed that all of the models were 'satisfactory' and a 'good' model (LEPS SK > 10) was achieved at the 2-year lead time. Furthermore, the

10-year moving average produced much better models in comparison with the 3-year and 5-year moving averages. This also indicates that M5P was better than climatology..

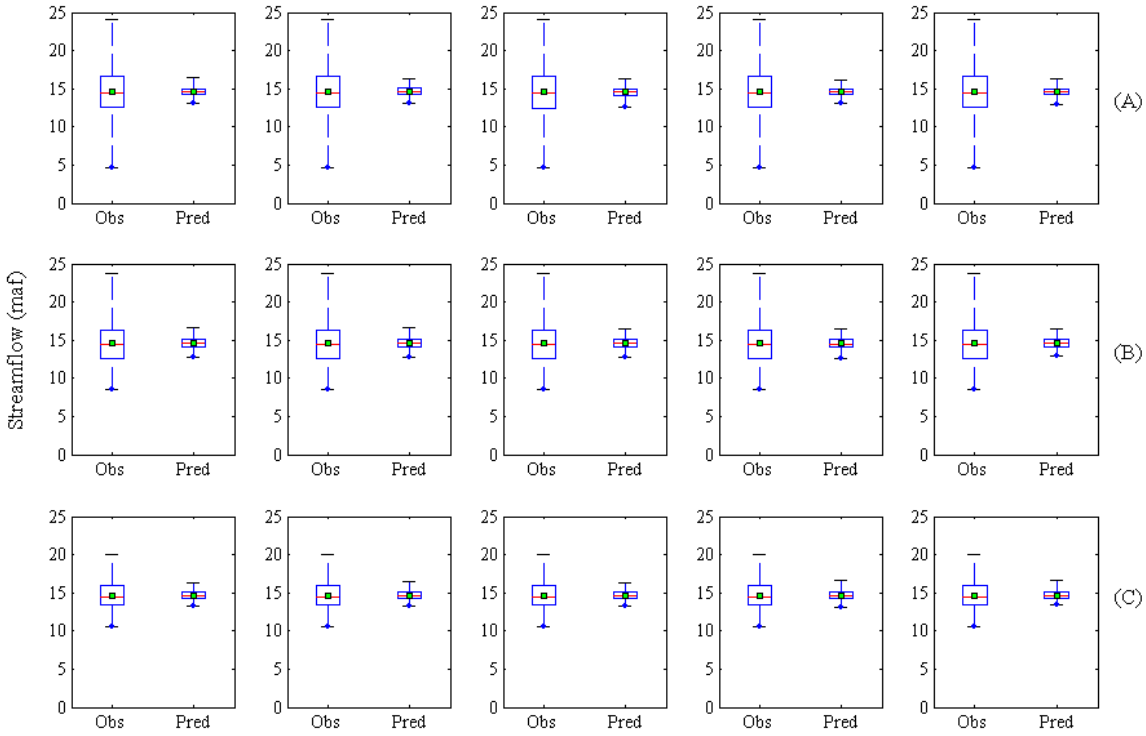

Figure 9: Box plots for M5P for Lees Ferry under a (A) 3-year moving average, (B) 5-year moving average, and (C) 10-year moving average. The box shows the interquartile range (25th-75th percentile). The whiskers extend from the 5th to 95th percentile values. The solid line inside the box shows the median value (50th percentile), and the solid dot represents the mean of the value. Plots are shown with a 1-year lag on the left and ascending to a 5-year lag on the right.

Figure 9shows the box plots, which examined the ability of the M5P model to capture streamflow variability. Figure 9A shows that the 3-year lead time was capable of capturing the long-term mean of the streamflow; however, it was unable to capture much of the variability within the 25th to 75th percentile. This indicates that the M5P model was unable to find viable decision points, leaving the model with few nodes that were crucial to developing flexibility in the M5P model. The extent of the whiskers of the predictions only covered the 25th to 75th percentile of the observed box plot, indicating that the model

was unable to predict extreme events. Figure 9B shows some improvement in capturing the streamflow variability when a 5-year moving average was applied to the dataset, when compared to the 3-year moving average window; however, extreme events were still unaccounted for. The range of predictions within the 25^{th} to 75^{th} percentile increased, but was still indicative that the model was unable to identify critical points for the decision tree. In addition, the model was unable to capture extreme flows under the 5-year moving average. Figure 9C shows that the 10-year moving average window resulted in a considerable improvement. The model improved at capturing the streamflow variability within the 25^{th} to 75^{th} percentile for all lead times, compared to the 3-year and 5-year moving average results. However, extreme events were not captured in the model. The box plots shown in Figure 9indicate that the M5P model accurately predicted streamflow variability for the 10-year moving average, compared to 5-year and 3-year moving average. However, all models experienced difficulty in capturing extreme flows.

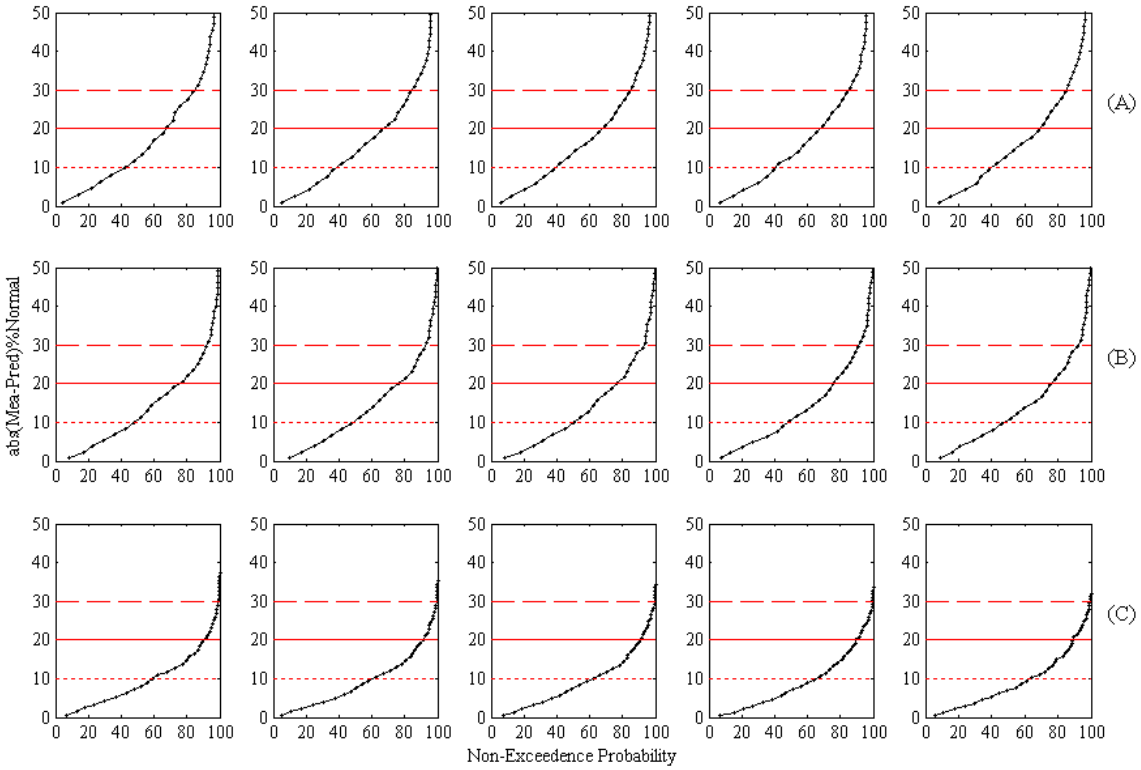

Figure 10: Non-exceedance probability plots for M5P for Lee's Ferry under a (A) 3-year moving average, (B) 5-year moving average, and (C) 10-year moving average. The horizontal dotted line shows an error value of 10%, the horizontal solid line shows an error value of 20%, and the horizontal dashed line shows an error value of 30%. Plots are shown with a 1-year lag on the left and ascending to a 5-year lag on the right.

The non-exceedance probability plots in Figure 10showed the accuracy of the M5P model in predicting the streamflow volume in terms of their error from the actual streamflow volume. Figure 10A shows that the 3-year moving average yielded high errors between the predicted and observed streamflow volumes. The plots show that approximately 38% - 42% of the predictions had an error of 10% or less. Predictions improved to 65% - 70%, having an error 20% or less. A further improvement was observed with 80% - 85% of the predictions, having an error of 30% or less. This implies that the majority of the predictions for the 3-year moving average were insufficient to provide meaningful information to water managers. Figure 10B shows an improvement

when a 5-year moving average was applied with 45% - 50% of the predictions with an error of 10% or less, 75% - 80% of the predictions with an error 20% or less, and 90% - 95% of the predictions with an error of 30% or less. The majority of the predictions under the 5-year moving average had an error of 20% or less, which was not sufficient to provide meaningful information for planning and managing water resources. Figure 10C showed a significant improvement in the accuracy of the streamflow forecasts when a 10-year moving average was applied. The plots show that 65% - 70% of the predictions had an error of 10% or less, and about 95% of the predictions had errors less than 20%. These errors were more acceptable for streamflow forecasting, as the majority of the predictions had an error of 10% or less and therefore can provide useful information to water managers when planning and managing available water resources within the UCRB. Furthermore, the non-exceedance probability plots supported the finding that the 10-year moving average was best suited for providing accurate streamflow forecasts as compared to the 5-year and 3-year moving averages.

Furthermore, the M5P model indicated that the 10-year moving average window consistently provided the best streamflow forecasts for the Lees Ferry gage. Therefore, the results for the remaining gages for M5P were depicted for the 10-year moving average window. Similar to the Lees Ferry gage, the M5P model was applied to Cisco, Green River, and San Juan River gages within the UCRB.

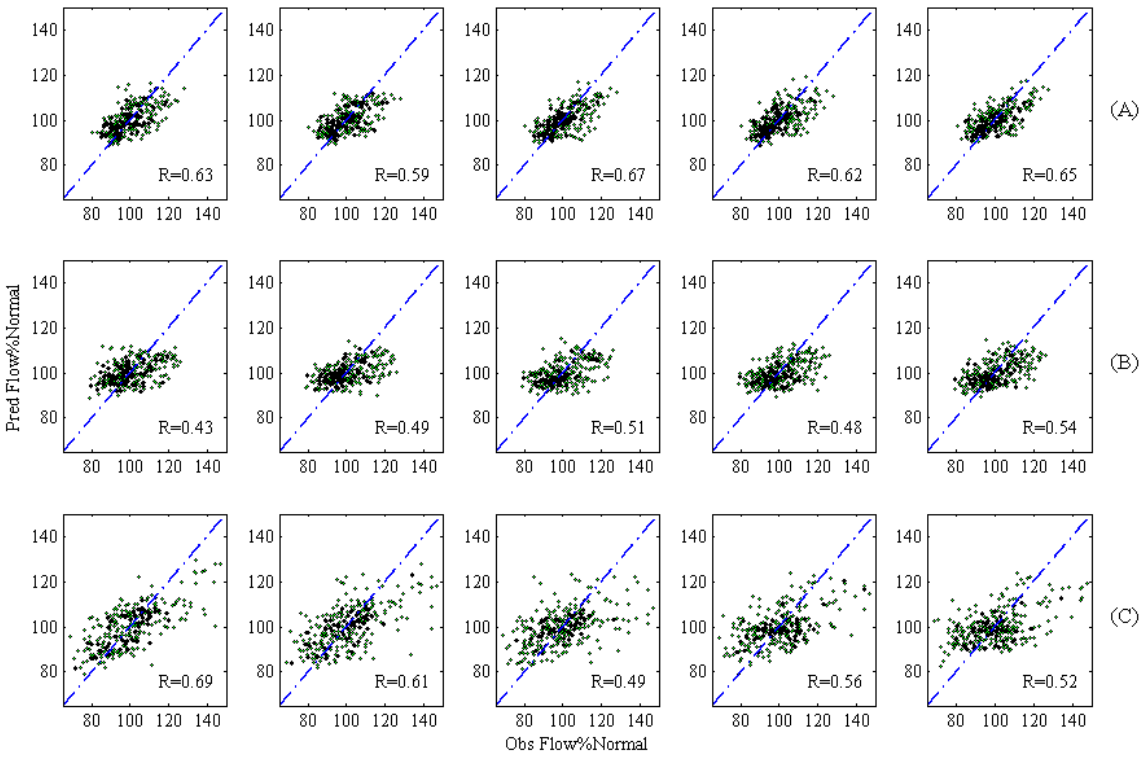

Figure 11: Scatter plots between measured and predicted streamflow for M5P using a 10-year moving average for (A) Colorado River near Cisco, Utah; (B) Green River at Green River, Utah; and (C) San Juan River near Bluff, Utah. The diagonal line is the 45° bisector line. Plots are shown with a 1-year lag on the left and ascending to a 5-year lag on the right.

Figure 11 shows the scatter plots for the three gages. Figure 11A shows that the streamflow predictions for Cisco were a good fit to the observed streamflow for all lead times. The Green River gage in Figure 11B showed satisfactory results for all lead times. The San Juan River gage in Figure 11C showed a much larger range in the streamflow volumes in terms of the difference from normal flow. The lower streamflow volumes observed at the San Juan River gage caused the variations to seem large when compared to the average annual streamflow. The forecasts for the San Juan River gage showed a good fit for all lead times. The scatter plots for Cisco, Green River, and San Juan River

gages were in agreement with forecasts for Lees Ferry, where the majority of predictions followed the bisector line, indicating a good model fit.

Table 6: Forecast evaluation, using M5P for Cisco, Green River, and San Juan River; using KStar under a 10-year moving average.

MA	Lead Time	RMSE	MAE	RSR	R	NSE	LEPS
CO	1	0.47	0.39	0.78	0.63	0.39	31.43
	2	0.49	0.39	0.80	0.60	0.36	30.02
	3	0.45	0.35	0.74	0.67	0.45	36.43
	4	0.47	0.37	0.78	0.62	0.39	32.99
	5	0.46	0.37	0.77	0.65	0.41	31.14
GR	1	0.48	0.39	0.90	0.43	0.18	18.21
	2	0.46	0.37	0.87	0.49	0.23	19.02
	3	0.45	0.35	0.86	0.52	0.26	24.10
	4	0.46	0.36	0.88	0.48	0.23	20.15
	5	0.44	0.35	0.84	0.54	0.29	25.46
SJ	1	0.22	0.17	0.72	0.69	0.47	38.17
	2	0.24	0.19	0.79	0.61	0.37	32.80
	3	0.26	0.20	0.87	0.49	0.24	25.11
	4	0.25	0.20	0.83	0.56	0.31	23.77
	5	0.25	0.20	0.85	0.52	0.27	21.87

[a]In MAF (1 ac-ft = 1233.5 m^3)

Table 6 displays the forecast evaluation obtained at the Cisco, Green River, and San Juan gages when a 10-year moving average was applied. The RSR indicates that all of the models produced 'unsatisfactory' results (RSR > 0.70). The R and NSE agreed with the RSR, but also found that the 1-year lead time for the San Juan gage produced 'satisfactory' results ($0.70 \leq R < 0.80$; $0.50 \leq NSE < 0.65$). These gages further indicated that the M5P model yielded predictions that were insufficient for streamflow forecasting. However, the LEPS SK indicated that 'good' forecasting models were produced and performed better than climatology.

Of the four gages, the M5P model performed best for the Cisco and San Juan River gages, although 'satisfactory' forecasts were made at all gages. Overall, the results showed that the M5P model was successful in capturing the relationship between large-scale climate patterns and streamflow within the UCRB, and forecasted streamflow with 1-5 year lead time using a 10-year moving average filter.

2.6.3 Comparison of MLR with KStar and M5P

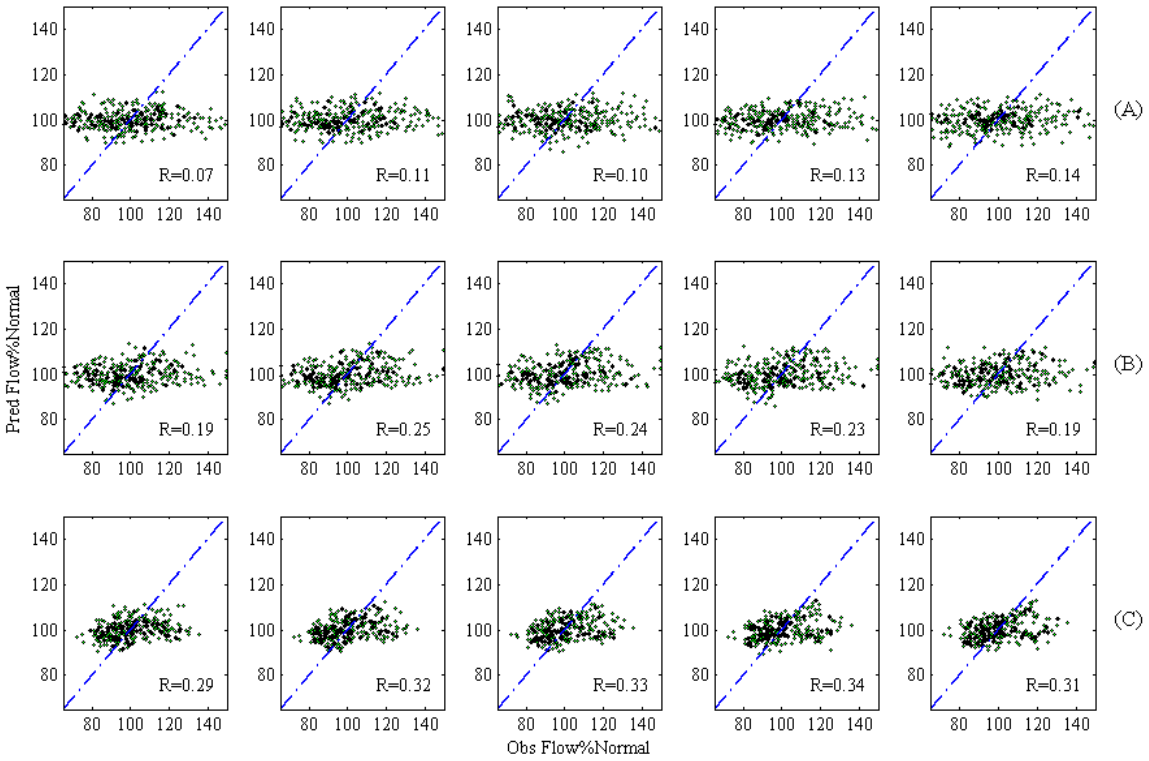

Figure 12: Scatter plots between measured and predicted streamflow for MLR for Lees Ferry under a (A) 3-year moving average, (B) 5-year moving average, and (C) 10-year moving average. The diagonal line is the 45° bisector line. Plots are shown with a 1-year lag on the left and ascending to a 5-year lag on the right.

Similar to KStar and M5P, annual streamflow volumes for 1-5 year lead time were estimated using the MLR model and evaluated using similar performance measures at the Lees Ferry gage. Figure 12shows the scatter plots between measured and predicted

streamflow volumes for the five lead times for MLR. Results for all 3 filters indicated that the predictions lied along long-term mean of the observed data, and the model was unable to capture the streamflow variability. Though there was some improvement when 5 and 10 year average were used, the predictions remained close to the predicted 100% normal flow. In comparison, the KStar model (Figure 4C) performed exceptionally well and provided a very good fit with the majority of predictions saturating the bisector line. The M5P model (Figure 8) behaved similarly to the MLR model at the 3-year and 5-year moving averages, but performed similarly to the KStar model when a 10-year moving average filter was applied. This shows that the M5P model was unable to identify nodes in the decision tree and resulted in only a few linear regression models. When a 10-year moving average filter was applied, the M5P model was able to find numerous nodes to create a well-diversified decision tree. This shows that the KStar and M5P models were better than MLR in identifying significant relationships between the oceanic-atmospheric oscillations and streamflow.

Table 7: Forecast evaluation for Lees Ferry, using MLR for a 3-year, 5-year, and 10-year moving average.

MA	Lead Time	MAE	RSR	R	NSE	LEPS
3	1	2.52	1.00	0.07	-0.01	1.88
	2	2.51	1.00	0.11	0.01	2.46
	3	2.52	1.00	0.10	0.00	2.51
	4	2.47	0.99	0.13	0.01	3.59
	5	2.47	0.99	0.15	0.02	4.04
5	1	2.03	0.98	0.19	0.03	5.98
	2	1.99	0.97	0.24	0.06	8.19
	3	1.98	0.97	0.24	0.06	8.81
	4	1.99	0.97	0.23	0.05	7.92
	5	2.02	0.98	0.19	0.03	6.12
10	1	1.42	0.96	0.29	0.08	10.27
	2	1.39	0.94	0.32	0.10	11.31
	3	1.36	0.94	0.33	0.11	11.82
	4	1.33	0.94	0.34	0.11	12.73
	5	1.36	0.95	0.31	0.09	10.47

[a]In MAF (1 ac-ft = 1233.5 m^3)

Table 7 displays the forecast evaluations for lead times of 1-5 years using the MLR model at the Lees Ferry gage. All of the MLR results produced 'unsatisfactory' results for R ($R < 0.70$), RSR ($RSR > 0.70$), and NSE ($NSE < 0.50$). This indicates that the MLR model was unable to produce significant forecasts, even when the 10-year moving average was applied. In comparison, KStar was shown to produce 'satisfactory' to 'good' predictions using the 10-year moving average (Table 3). However, the M5P produced 'unsatisfactory' predictions (Table 5), but was able to outperform the MLR model. The LEPS SK show that the MLR model exhibited no skill ($SK < 5$) when a 3-year moving average filter was applied, 'satisfactory' forecasts ($SK > 5$) at the 5-year moving average filter, and 'good' forecasts ($SK > 10$) when a 10-year moving average was applied (Table 7). In comparison, the KStar model produced 'good' forecasts with each of the filters (Table 3) and the M5P model produced 'satisfactory' forecasts at the 5-

year moving average and 'good' forecasts with a 10-year moving average (Table 5). When larger moving averages were applied, the KStar model produced much higher LEPS SK compared to the M5P and MLR model.

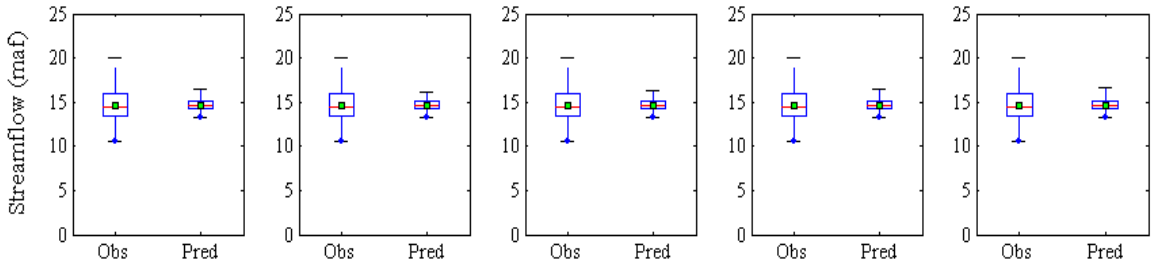

Figure 13: Output from MLR model for Lees Ferry under a 10-year moving average shown as box plots. The box shows the interquartile range (25th - 75th percentile). The whiskers extend from the 5th to 95th percentile values. The solid line inside the box shows the median value (50th percentile), and the solid dot represents the mean of the value. Plots are shown with a 1-year lag on the left and ascending to a 5-year lag on the right.

The MLR models were also evaluated through box plots with a 10-year moving average filter (Figure 13). The MLR model was unable to capture much variability or any extreme events. This indicates that the model was unable to identify relationships between the oscillations and streamflow, returning only the mean streamflow as the prediction. The box plots in Figure 5C showed that the KStar model was able to capture nearly all of the variability associated with the observed dataset as well as a most of the extreme events. Figure 9C indicates that the M5P model was capable of capturing most of the variability, but had difficulty in capturing extreme events.

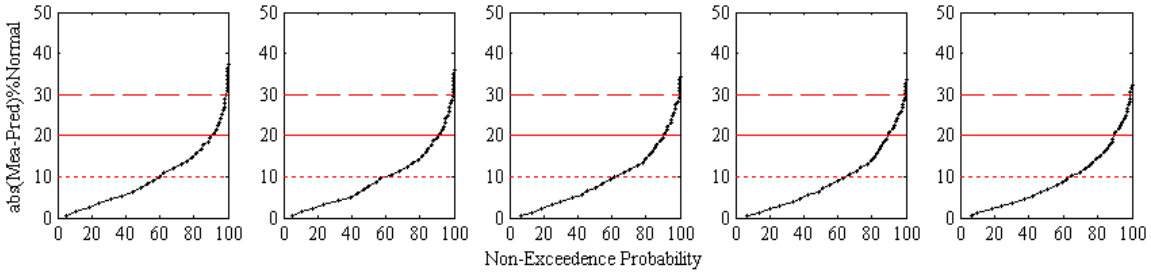

Figure 14: Output from MLR model for Lees Ferry under a 10-year moving average shown as non-exceedance probability plots. The non-exceedance probability plot display the horizontal dotted line which represents an error value of 10%, the horizontal solid line which represents an error value of 20%, and the horizontal dashed line which represents an error value of 30%.Plots are shown with a 1-year lag on the left and ascending to a 5-year lag on the right.

Similarly, non-exceedance probability plots (Figure 14) were used to evaluate the MLR model. Unsatisfactory predictions were made for the MLR model, in which about 60% of the predictions had an error of 10% or less. This was relatively high error for streamflow forecasting. In comparison, the plots for the KStar model in Figure 6C showed that 85% - 90% of the predictions had an error of 10%.This indicated that the KStar model was robust and flexible: it produced a wide range of predictions and was better suited for providing reliable information to water managers. Figure 10C showed that the M5P model was also proficient at producing accurate results as 65% – 75% predictions had an error of 10% or less and can also provide water managers with reliable forecasts.

Overall, the results indicated that that the MLR model was outperformed by both the KStar and M5P models, with the KStar model yielding the best overall forecasts.

2.7 Discussion and Conclusion

The current research presented the KStar and M5P data-driven forecasting models, which incorporated reconstructed and observed annual oceanic-atmospheric

oscillations to forecast annual water year streamflow volumes up to 5 years (t+5) into the future for four gages in the Upper Colorado River Basin. Where other studies were limited to a short period of instrumental record of less than 100 years, the incorporation of proxy reconstructions allowed for the period of record to be extended to over 300 years.

A forecast produced with a 10-year moving average provided the best conditions for streamflow forecasting. A moving average filter was used in this study to smooth out the highly fluctuating data, which had to be filtered in order to attenuate the short-range fluctuations and to extract and clarify the long-range climatic variations (Hidalgo, 2004; McCabe et al., 2004; Probst & Tardy, 1987). The shortcoming of this approach was the inherent loss in streamflow variability, which limited the model's ability to forecast extreme events. However, a moving average filter did not compromise the signals associated with the oceanic-atmospheric oscillations, and aided the model in identifying the phases of the oceanic-atmospheric oscillations (e.g. Hidalgo, 2004). The premise of this study was that the oceanic-atmospheric oscillations can provide predictive information, which can be used to extend the forecast lead time. In addition, it was often difficult for data-driven models to identify phases due to the variability of the signal within a phase, which attributes to noise. According to Cleary and Trigg (1995), a smooth surface was required for predictor attributes so that the model did not make decisions to its detriment. Therefore it was necessary to use a filter to clearly define the phases and remove the short-term fluctuations within each phase. Other hydrologic studies have identified the need to smooth the data with a moving average filter for use in their analysis (e.g. Currie, 1996; Hidalgo, 2004; McCabe et al., 2004; Pekarova et al., 2003;

Probst & Tardy, 1987; Riehl & Meitin, 1979; Riehl et al., 1979). The incorporation of a moving average filter allowed the KStar model to produce more accurate streamflow forecasts up to 5 years into the future.

In comparison with MLR, both the KStar and M5P models yielded better forecasts, with the KStar model providing the best overall results. This was in agreement with previous studies pertaining to hydrologic forecasting (e.g. Gangopadhyay et al., 2009; Lall & Sharma, 1996; Piechota & Dracup, 1999), in which a flexible data-driven approach outperformed standard regression models. The MLR model lacked the ability to capture the non-linear relationships between the oceanic-atmospheric oscillations and streamflow, because it used simple linear regression to make predictions. Since a large majority of the dataset was close to average streamflow conditions, the MLR model was weighted to the average streamflow. KStar directly compared the similarity of each instance to remove mean-weighted biases. M5P categorized similar instances together to give a more transparent structure to the model. These characteristics allowed for the KStar and M5Pmodels to better capture the natural variability that was difficult to capture for the linear models.

The data-driven modeling technique showed that accurate streamflow forecasts could be achieved for the majority of the dataset. The KStar model had success in capturing extreme streamflows, because it compared the current instance directly with extreme streamflow examples by means of the entropy function. This allowed the KStar model to maintain flexibility. Similarly, M5P achieved success by grouping together similar instances based upon the inputs and outputs. However, both models experienced difficulties when similar oscillation inputs had dissimilar streamflow volumes in the

training set. The moving average filter helped mitigate this by establishing the phases of the oscillations as opposed to the raw dataset that contains numerous jumps and discrepancies.

It is beneficial to determine which oscillations show the most influence within the UCRB because there is a possibility that some oscillations do not provide a signal for streamflow or may hinder the signals of other oscillations. The current study performed a rigorous sensitivity analysis to test the significance of each of the oceanic-atmospheric oscillations on streamflow (results not shown). The analysis indicates that ENSO, PDO, AMO, and NAO used together show a stronger association with UCRB streamflow using the paleoclimate record. The removal of one or more of the oscillations results in a deterioration of the forecast. Additionally, the current study indicates that the lead time can be extended up to 5 years by using reconstructed data. This improves upon the 3-year lead time achieved by Kalra and Ahmad (2009) using a data-driven SVM model to predict streamflow volumes for the UCRB using the instrumental record only.

The incorporation of paleoclimatic data into this study aided in understanding the relationships between oceanic-atmospheric oscillations and streamflow response. The use of paleoclimatic data has the tremendous benefit of extending the period of record by several hundred years beyond the instrumental record. The reconstructions allow researchers to examine long-term physical processes and provide a better understanding of a dynamic climate system, which may not be fully studied in the short instrumental record. These reconstructions are particularly useful for data-driven models because they increase the amount of examples used in the training set, possibly creating a more robust and reliable model. There are limitations and uncertainties when using reconstructions

because there are multiple reconstructions available that are often not in good agreement. This is because reconstructions currently are capable of explaining 50% - 80% of the variance observed in the instrumental record (Woodhouse et al., 2011).

Overall, the results from this study contributed to a better understanding of the impacts of long-term physical processes on streamflow in the Colorado River Basin by using paleoclimate reconstructions to extend the period of record back to 1661. In order to identify oscillation phases, a moving average filter was applied to the dataset. Although a decrease in the magnitude of the streamflow variability was experienced, the model results contributed to the increase of the lead time by 1-5 years for quantitative streamflow forecasts in the Colorado River Basin. This was an improvement over the seasonal (Apr-July) streamflow forecasts for the western United States provided by the Colorado Basin River Forecast Center. In addition, using all four oscillation indices (i.e. ENSO, PDO, AMO, and NAO) as predictors resulted in the best streamflow predictions. This was in agreement with other studies indicating that no single climate system can be used to explain the hydroclimatology within Colorado River Basin. Coupled or grouped response of oscillations had stronger association with streamflow compared to their individual effects. The KStar model results provided 'satisfactory' to 'good' streamflow predictions, which can allow water managers to gain important insights about the upcoming water year as early as January 1stand can be useful for long-term water management decisions in the UCRB.

Chapter 3: Streamflow Forecasts using Paleoclimate Reconstructions in the Western United States

3.1 Introduction

Climate fluctuations have the potential to cause large impacts on water supply, electrical power, floods and droughts, and ecological habitat as the water resources in the western United States become stressed (Cayan et al., 2003; Hamlet & Lettenmaier, 1999; Forsee & Ahmad, 2011). The population in the western United States has been estimated to increase by 14.3% from 2000 to 2010 (United States Census, 2010), which places stress upon a limited renewable resource (Qaiser et al., 2011). In addition, a 2011 report from the Bureau of Reclamation has indicated that streamflow in the interior western United States is expected to decline, and these declines are likely to become greater in magnitude over the course of the century (Reclamation, 2011). Furthermore, recent trends in the western United States have shown that high variability in streamflow and persistent high or low flows are synchronous across major basins, resulting in further stress to water resources (Jain et al., 2005; Pagano & Garen, 2005). Water stress or excess within one region can also affect conditions in other regions, because water is often traded or transferred across state and watershed boundaries (Cayan et al., 2003; Ahmad & Prashar, 2010; Vedwan et al., 2008). Changes in climate often result in floods and droughts. The impact of these events on agriculture, soil moisture, water resources, water systems, human settlements and the environment has been studied by many researchers (e.g., Simonovic & Ahmad, 2005; Mosquera-Machado & Ahmad, 2007; Stephen et al., 2010; Puri et al., 2011; Ahmad & Simonovic, 2000, 2001, 2006). The increasing water stress adds complexity to planning and managing water resources in an already challenging environment of changing demographics, changing water quality and competing interests

(Pagano & Garen, 2005; Shrestha et al., 2011; Venkatesan et al., 2011). A reliable long-range streamflow forecast greater than a year can aid in the long-term planning and management of water resources.

Oceanic-atmospheric oscillations provide an opportunity to provide long-range streamflow forecasts. There is an established body of research that demonstrates a predictable relationship between climatic variations and streamflow in the western United States over a wide range of temporal and spatial scales (e.g. Dettinger et al., 1998; Enfield et al., 2001; Gershunov & Barnett, 1998; Hamlet and Lettenmaier, 1999; Hidalgo & Dracup, 2003; Mantua, 1999; McCabe et al., 2007; Sheppard et al., 2002; Tootle et al., 2005). Climatic variations associated with El Niño – Southern Oscillation (ENSO), Pacific Decadal Oscillation (PDO), Atlantic Multi-decadal Oscillation (AMO), and North Atlantic Oscillation (NAO) show the most influence over streamflow variability in the western United States (e.g. Dettinger et al., 1998; Enfield et al., 2001; Gutzler et al., 2002; Hamlet & Lettenmaier, 1999; Hidalgo & Dracup, 2003; Hunter et al., 2006; Kalra & Ahmad, 2009; McCabe et al., 2007; Redmond & Koch, 1991; Sheppard et al., 2002; Tootle et al., 2005). These oscillations have warm and cool phases, which are associated with variations in the climate and in weather patterns. These oscillations may be used for long-range streamflow forecasting, because their effects on streamflow have lags that are greater than a year (Gray et al., 2003; Kalra & Ahmad, 2009). However, because they are not yet fully understood due to their complexity, it is difficult to incorporate oscillations into a physically-based forecasting model (Kalra & Ahmad, 2009).

Data-driven modeling serves as an alternative to physically-based modeling (Lin et al., 2009). Data-driven models use a training dataset comprised of previously known

examples to extract relationships between inputs and outputs (Witten et al., 2011). Data-driven models include artificial neural networks (ANNs), support vector machines (SVMs), decision trees, and instance-based learners. This study uses the KStar model, which is an instance-based learner that uses a generalized distance function based on transforms (Cleary & Trigg, 1995; Witten et al., 2011). This is a relatively simple model, but has been shown to outperform more complex models, such as ANNs (Ahmad & Simonovic, 2005; Solomatine et al., 2008; Wang et al., 2006). Instance-based learners have been used in several hydrologic applications including evaporation estimation (Elshorbagy et al., 2010; Terzi, 2007), rainfall-runoff modeling (Elshorbagy et al., 2010; Solomatine et al., 2008), and soil moisture content (Elshorbagy et al., 2010). Although such data-driven models as KStar are simple in comparison to physically-based models, they may become more robust when trained on a larger dataset because of the greater number of instances (Melesse et al., 2010; Witten et al., 2011). A robust data-driven model would serve to complement existing physically-based models.

The observed instrumental record typically extends from 50-100 years, and is useful in examining historical trends and relationships as well as managing reservoir operations on a daily, weekly, or monthly basis (Franz et al., 2003; McEnery et al., 2005). However, an extended record can potentially provide more information for long-term management. The record can be extended using paleoclimate reconstructions (Dettinger et al., 1998; Gray et al., 2003). This study used reconstructions to extend the oscillation and streamflow record beyond 300 years. Reconstructions are generated from proxies that are well correlated with temperature and moisture (Hidalgo 2004; Jones and Mann 2004; Meko et al., 1995; Stockton & Jacoby, 1976). In particular, tree rings provide an annual

time series where the ring width can be correlated with a streamflow volume or oscillation phases (Hidalgo, 2004; Timilsena et al., 2009; Woodhouse & Lukas, 2006). Longer records using reconstructions are useful to better understand the climate variability associated with oceanic-atmospheric oscillations (Dettinger et al., 1998; Hidalgo, 2004; Timilsena et al., 2009; Woodhouse & Lukas, 2006). Although there is a high level of uncertainty associated with reconstructions, they are able to capture the climate variability (Timilsena et al., 2009). With an extended period of record provided by reconstructions, a robust data-driven model can be potentially developed for streamflow forecasting.

There are no previous studies that have used paleoclimate reconstructions of oceanic-atmospheric oscillations in a data-driven model for streamflow forecasting. With this in mind, the current study used the KStar model to provide a 1-5 year lead time forecast for 21 gage stations in the western United States using reconstructions. The model was trained and tested using a 10-fold cross-validation technique. Forecasts were evaluated by using root mean squared error observations, standard deviation ratio (RSR), Pearson's correlation coefficient (R), Nash-Sutcliffe coefficient of efficiency (NSE), and linear error in probability space skill score (LEPS SK). This modeling technique is intended to aid in the planning and management of water resources in the western United States.

The layout of the paper is as follows. Section 3.2 describes the study area and the data used in this study. A description of the methodology, including the KStar model and model performance measures are provided in Section 3.3. The results of the forecasts are

presented in Section 3.4. Finally, Section 3.5 provides a discussion and concludes the study.

3.2 Study Area and Data

Annual streamflow data was obtained for 40 gages in the western United States. These gages were identified with a period of record from 1658 to 2007, a total of 350years. Reconstructions were used from 1658 to 1899, and the instrumental record was used from 1900 to 2007. Of these 40 gages, 21 gages were selected based upon data availability and compatibility between instrumental records and reconstructions. Figure 15 and Table 8 identify the 21 streamflow gages used in the study. Table 9 describes the observed and reconstructed datasets used for each streamflow gage.

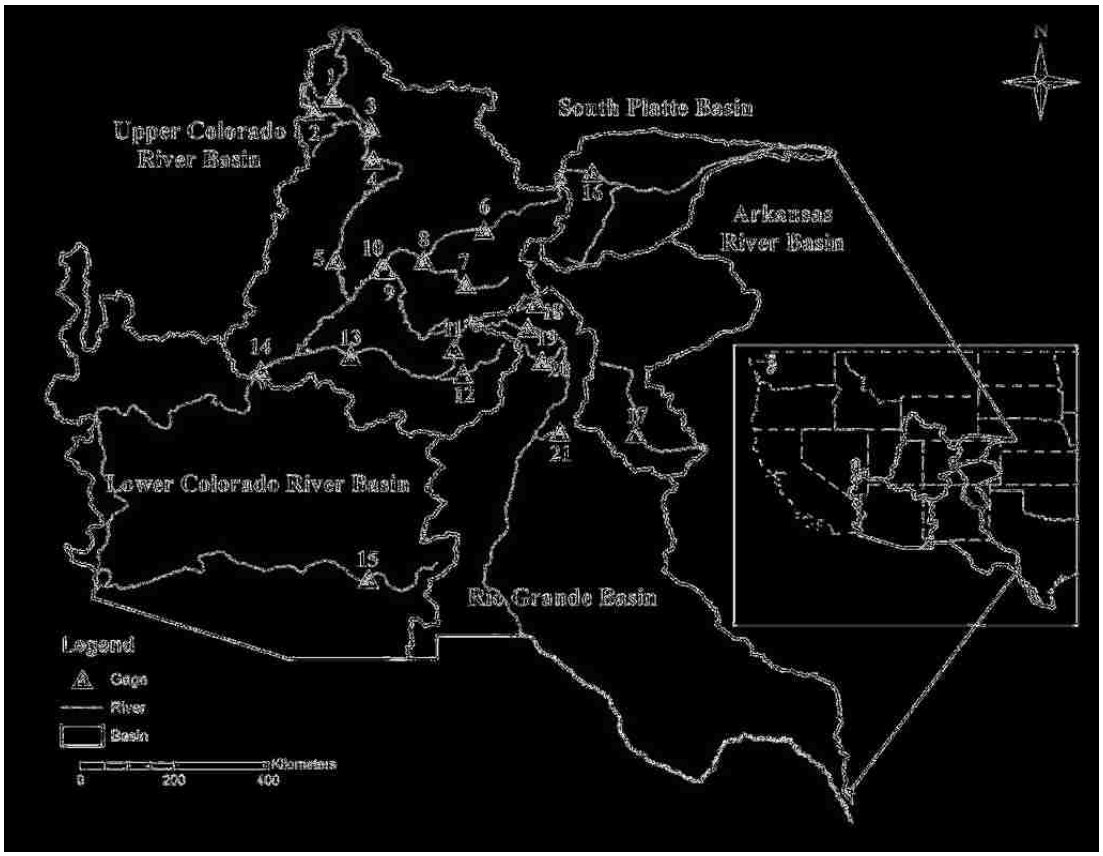

Figure 15: Study area depicting the location of the 21 streamflow gages used in the current study.

Table 8: Streamflow gage locations

Gage Number	Region/Basin	Gage Location
1	UCRB	Fontenelle Creek near Fontenelle, WY
2	UCRB	Hams Fork near Frontier, WY
3	UCRB	Green river at Green River, WY
4	UCRB	Green River near Greendale, UT
5	UCRB	Green River at Green River, UT
6	UCRB	Colorado River above Glenwood Springs, CO
7	UCRB	Gunnison at Crystal Reservoir, CO
8	UCRB	Gunnison River near Grand Junction, CO
9	UCRB	Dolores River near Cisco, UT
10	UCRB	Colorado River near Cisco, UT
11	UCRB	Animas River at Durango, CO
12	UCRB	San Juan River at Archuleta, NM
13	UCRB	San Juan River near Bluff, UT
14	UCRB	Colorado River at Lees Ferry, AZ
15	LCRB	Gila River near Solomon, AZ
16	SPB	Cache La Poudre River near Ft Collins, CO
17	ARK	Canadian River near Sanchez, NM
18	RGB	Saguache Creek near Saguache, CO
19	RGB	Rio Grande near Del Norte, CO
20	RGB	Conejos River near Mogote, CO
21	RGB	Santa Fe River near Santa Fe, NM

Table 9: Data sources for streamflow gages and oscillation indices.

Reconstructions			Instrumental Record	
Gage Number	Source	Data Available*	Source	Data Available*
1, 2, 4	Barnett et al., 2007	1615-1999	United States Geological Survey, 2011	1952-2010
3, 5	Woodhouse et al., 2006	1615-1998	United States Bureau of Reclamation, 2011	1906-2007
6	Woodhouse et al., 2006	1525-1997	United States Geological Survey, 2011	1906-2007
7, 8, 9, 10, 12, 13	Woodhouse et al., 2006	1569-1996	United States Bureau of Reclamation, 2011	1906-2007
11	Woodhouse and Lukas, 2006a	1470-2002	United States Geological Survey, 2011	1928-2010
14	Meko et al., 2007	762-2005	United States Bureau of Reclamation, 2011	1906-2007
15	Meko and Hirschboeck, 2008	1332-2005	United States Geological Survey, 2011	1921-2010
16	Woodhouse and Lukas, 2006a	1615-1999	United States Geological Survey, 2011	1911-2007
17	Meko et al., 2007	1604-1997	United States Geological Survey, 2011	1937-2010
18	Woodhouse and Lukas, 2006b	1520-2000	United States Geological Survey, 2011	1915-2007
19	Woodhouse and Lukas, 2006b	1508-2002	United States Geological Survey, 2011	1891-2010
20	Woodhouse and Lukas, 2006b	1508-2002	United States Geological Survey, 2011	1913-2010
21	Margolis, 2011	1592-2007	United States Geological Survey, 2011	1914-2010
ENSO	Jones and Mann (2004)	1650-1980	Australian Gov. Bureau of Meteorology (2011)	1876-2010
PDO	Shen et al. (2006)	1470-1998	JISAO (2011)	1900-2010
NAO	Luterbacher et al. (2001)	1658-2001	Hurrell (2011)	1865-2010
AMO	Gray et al. (2004)	1567-1985	ESRL (2011)	1856-2010

*Reconstructed data is used from 1685-1952 and instrumental data is used from 1953-2007

Fifteen gages are located in the Upper Colorado River Basin (UCRB). The remaining gages are located in the Lower Colorado River Basin (LCRB), South Platte Basin (SPB), Arkansas River Basin (ARK), and Rio Grande Basin (RGB). Observed flows are provided by the United States Geological Survey (USGS), and naturalized flows are provided by the United States Bureau of Reclamation (USBR). Naturalized flows are required for some gages in order to remove anthropogenic affects, such as consumptive losses and reservoir operations. Reconstructions are archived by TreeFlow (http://treeflow.info).

In addition, this study used four oceanic-atmospheric oscillations indices as predictors, based on anomalies of the sea-surface temperature (SST) or sea-level pressure (SLP). The observed and reconstructed datasets used for each index are described in Table 9. ENSO is a well-known oscillation index with a characteristic return frequency of 4-6 years; it originates over the tropical Pacific Ocean (Hamlet & Lettenmaier, 1999). For this study, an SLP index, known as the Southern Oscillation Index (SOI), was used as an indicator of ENSO, which measures the difference in SLP between Tahiti and Darwin, Australia (Redmond & Koch, 1991). The winter (October to March) SOI was used because it is a better indicator of ENSO than using the entire year (Mann et al., 2000; McCabe & Dettinger, 1999). The warm phase known as El Niño is associated with above-normal streamflow in the southwestern United States and below-normal streamflow in the Pacific Northwest (Mann et al., 2000; McCabe & Dettinger, 1999). The cool phase known as La Niña is identified with the inverse relationships.

The PDO is a bimodal SST climate pattern in the North Pacific Ocean with oscillation cycles on the order of 50 years (Hamlet & Lettenmaier, 1999; Mantua et al.,

1997). Warm phases of the PDO are associated with above-normal streamflow in the southwestern United States and below-normal streamflow in the Pacific Northwest (Mantua et al., 1997). The cool phases are identified with the opposite effects on streamflow.

The AMO is a low-frequency North Atlantic SST index that has a 65-80 year cycle with a 0.4 °C range (Enfield et al., 2001; Gray et al., 2004). Warm phases of the AMO are identified with below-normal streamflow in the southwestern United States, while cool phases are identified with above-normal streamflow in the southwestern United States (Gray et al., 2004; McCabe et al., 2004).

The NAO is measured by the differential SLP between Iceland and the Azores; it oscillates at a decadal time scale (Hurrell & Van Loon, 1997). The phases of NAO shifts the jet stream north during a warm phase and south during a cool phase, which may influence climate variability over the Pacific Northwest and northern Rocky Mountains (Hunter et al., 2006).

3.3 Methodology

This study uses a forecast model that is generated from the KStar algorithm, which is featured in the Weka data-mining software (Witten et al., 2011). KStar is an instance-based learner that uses a generalized distance formula based on transforms to find a measure between two instances (Cleary & Trigg, 1995). The KStar algorithm is motivated by information theory, where the distance between instances is defined as the complexity of transforming one instance into another, so that all possible transforms are considered (Cleary & Trigg, 1995). Other instance-based learners simply use the

Euclidean distance between two instances, which is equivalent to only one transform (Witten et al., 2011). This makes them very sensitive to small changes in the instance space and requires a very smooth surface. The KStar algorithm tries to deal with this problem of smoothness by summing over all possible transformations between two instances (Cleary & Trigg, 1995). A detailed discussion of the KStar model is provided by Cleary and Trigg (1995).

In addition, a standard multiple linear regression (MLR) model is used as a forecast model. The MLR model is subjected to the same methodology and performance analysis as the KStar model. This provides the KStar model with a fair comparison to accepted regression techniques.

The forecast model used all four oceanic-atmospheric oscillation indices (i.e., ENSO, PDO, AMO, and NAO) as predictors, and used one streamflow gage as the predictand. The entire dataset is used from 1658 to 2007. A lead time approach is used so that the indices of a given year are used to predict the streamflow volume 1-5 years into the future. A total of 210 model runs are performed (KStar and MLR for 1-5 year lead times at each of the 21 stations). The forecast models are performed with a stratified 10-fold cross-validation technique, which is the standard procedure of predicting the error rate of data-driven techniques (Witten et al., 2011). This technique randomly partitions the dataset into 10 folds of equal proportions. The model is trained on 9 folds and tested on the remaining fold. This is executed a total of 10 times, where each fold is held out in turn. The error estimates are averaged together to yield an overall result. These results are analyzed using a set of performance analysis measures, which are described in the following section.

When dealing with instance-based learners such as KStar, it is necessary to use a smooth dataset so that there are no large jumps or discontinuities, which may cause the model to make decisions to its detriment (Cleary & Trigg, 1995). The datasets used in this study are smoothed with a 10-year moving average filter in order to attenuate the short-range fluctuations, or noise, in the dataset and also to extract long-range climate signals (Hidalgo, 2004; McCabe et al., 2004; Pekarova et al., 2002; Probst & Tardy, 1987). More complex filtering techniques are available, for example, the cumulative deviation method, the weighted moving average, and the exponentially weighted moving average. Each technique yields a similar filtered time series; however, a difference in the localization of minima and maxima is sometimes observed in the filtered time series (Hidalgo, 2004; Pekarova et al., 2002; Probst & Tardy, 1987). The filtered dataset maintains the signatures of the oscillation indices (Hidalgo, 2004), which helps the model to determine the correct relationships between oscillation phases and streamflow variability.

3.3.1 Performance measures

The forecast results were analyzed using several performance measures, including the RSR, R, NSE, and LEPS SK. In particular, the RSR, R, and NSE performance measures were subjected to strict performance ratings, adopted from Moriasi et al. (2007), and were categorized as ‘unsatisfactory,’ ‘satisfactory,’ ‘good,’ and ‘very good’ (Table 10).

Table 10: Performance Rating for RSR, R, and NSE.

Performance Rating	RSR*	R	NSE*
Very Good	$0.00 \leq RSR \leq 0.50$	$0.85 \leq R \leq 1.0$	$0.75 \leq NSE \leq 1.0$
Good	$0.50 < RSR \leq 0.60$	$0.80 \leq R < 0.85$	$0.65 \leq NSE < 0.75$
Satisfactory	$0.60 < RSR \leq 0.70$	$0.70 \leq R < 0.80$	$0.50 \leq NSE < 0.65$
Unsatisfactory	$0.70 < RSR \leq 1.0$	$0.00 \leq R < 0.70$	$0.00 \leq NSE < 0.50$

*RSR and NSE adopted from Moriasi et al., [2007]

The MAE and RMSE are two commonly used performance measures that provide errors in the same units as the tested data; for both measures, a value of 0 indicates a perfect forecast (Moriasi et al., 2007; Witten et al., 2011). The RSR is defined as the RMSE divided by the standard deviation of the observed dataset; this standardizes the RMSE to allow it to be used as an error index (Moriasi et al., 2007). It also provides a better comparison of the model performance than the RMSE because streamflow volume varies greatly from station to station.

R measures the linear relationship between the observed and predicted values, which typically ranges from 1 (perfect positive correlation), 0 (no correlation), and -1 (perfect negative correlation). Moriasi et al. (2007) indicates that acceptable models are obtained when R is greater than 0.50. However, the current study develops stricter criteria in order to complement the RSR and NSE performance ratings adopted from Moriasi et al. (2007) (Table 10).

The NSE measures the residual variance, or noise, in comparison with the measured data variance, or information, and is defined in Moriasi et al. (2007). NSE indicates how well the observed and predicted data follows a 1:1 relationship (Legates & McCabe, 1999; Moriasi et al., 2007). NSE ranges from 1 to $-\infty$, where a positive NSE is acceptable because it indicates the model is a better predictor than the observed mean (Legates & McCabe 1999; Moriasi et al., 2007).

The LEPS SK is designed to measure the accuracy of forecast predictions, where higher weights are given to values that are further from the mean of the dataset (Potts et al., 1996). LEPS SK includes climatology in its calculation, and serves as a better indicator than climatology (Casey, 1998; Potts et al., 1996). LEPS SK ranges from -100 to 100; it is considered 'bad' if the measure is below -10, 'poor' if it is below -5, 'satisfactory' if it is greater than 5, and 'good' if it is greater than 10 (Casey, 1998).

Besides developing performance measures, a visual inspection of the forecast was performed by means of scatter plots, box plots, and non-exceedance plots. Both the KStar and MLR models were subject to this set of performance measures.

3.4 Results

3.4.1 KStar model

This section describes the results obtained when ENSO, PDO, AMO, and NAO were combined in the KStar model to forecast streamflow volumes for each of the selected gages, as described in the methodology section.

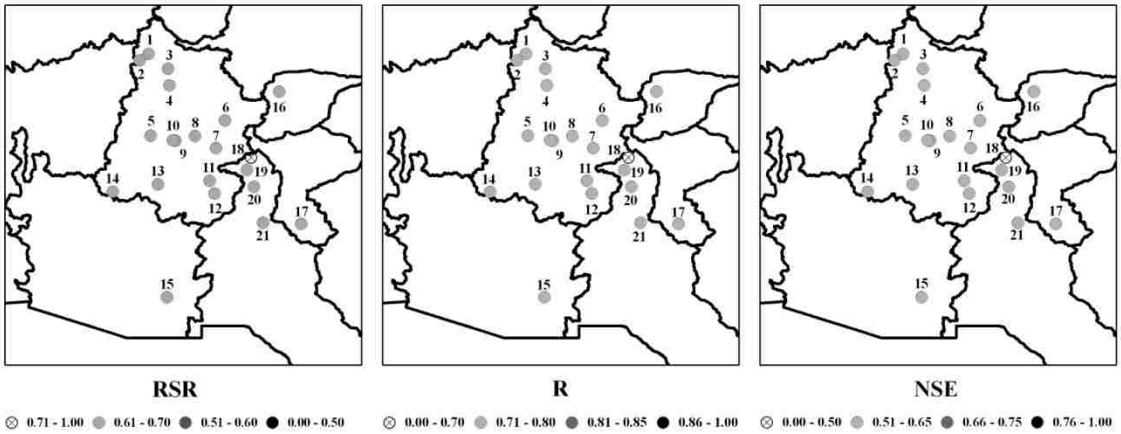

Figure 16: Results displayed in a spatial map using KStar at a 1-year lead time. Performance measures shown are the RSR, R, and NSE.

Figure 16displays a spatial representation of the RSR, R, and NSE results obtained for the 21 streamflow gages using KStar with a 1-year lead time. A lighter color indicates that the gage produced poor results, while a darker color represents a strong forecast was made at the gage. The RSR indicates that the model produced ‘satisfactory’ forecasts for the majority of the gages with one ‘unsatisfactory’ forecast made at Gage 18. Similar results were obtained for R, which showed that the majority of the gages had ‘satisfactory’ correlations between the measured and predicted values. The NSE also agreed with the RSR and R, where the majority of the gages produced ‘satisfactory’ forecasts.

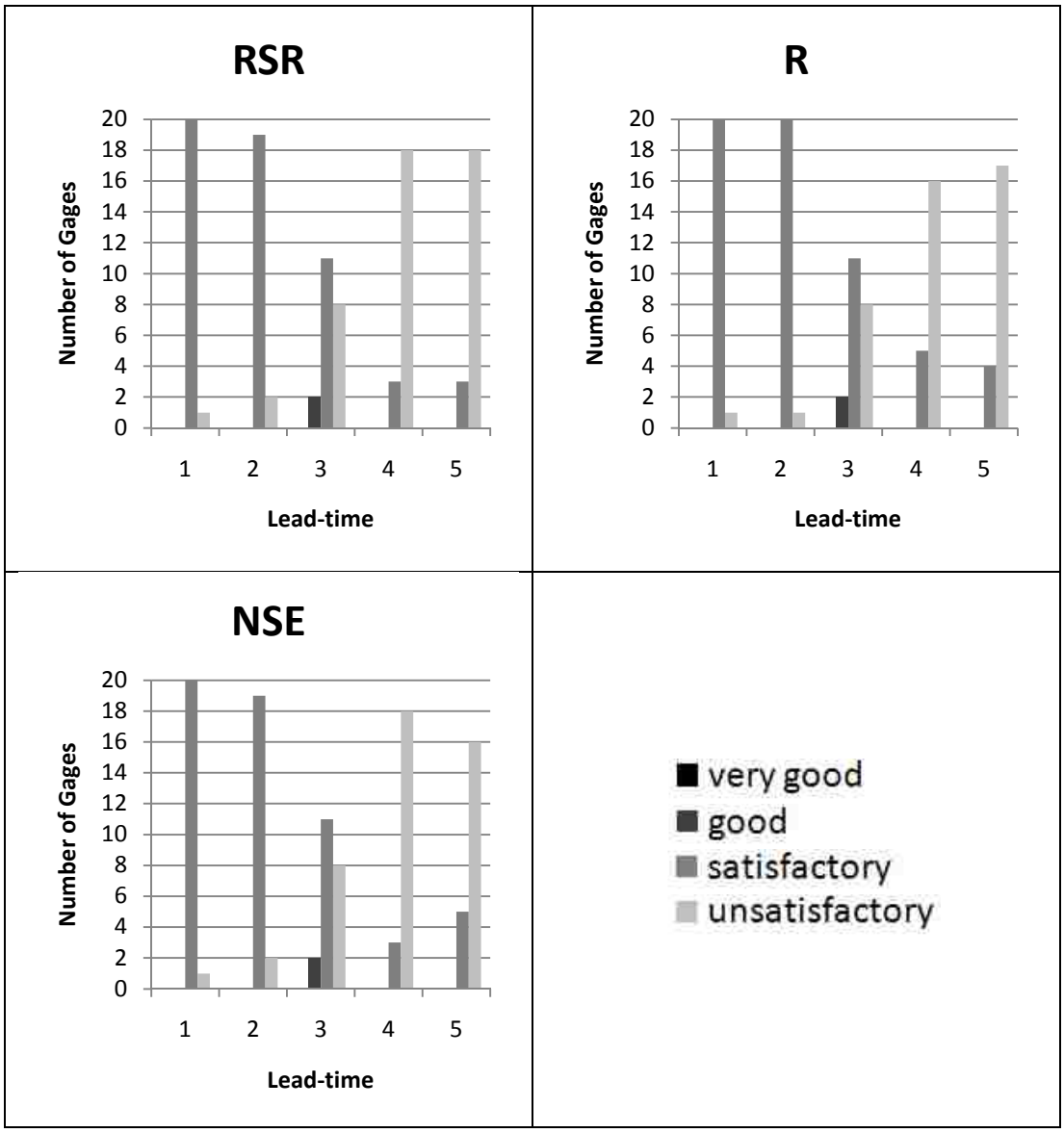

Figure 17: Bar graphs showing the results of the KStar model obtained at 1-5 year lead times. The performance ratings corresponding to the 'very good,' 'good,' 'satisfactory,' and 'unsatisfactory' categories are provided in Table 10.

Figure 17summarizes the performance measures as bar charts for the 21 gage stations using a 1-5 year lead time. The bars represent the number of gages that were 'very good,' 'good,' 'satisfactory,' or 'unsatisfactory,' based upon the performance ratings provided in Table 10. The RSR indicated that the 1-year forecast yielded the best performance. It produced 20 'satisfactory' forecasts and 1 'unsatisfactory' forecast made for Gage 18 in the RGB. The RSR for the 2-year forecast indicated 'satisfactory' results

for 19 of the 21 gages. Gages 18and 21in the RGB were the only gages to show ‘unsatisfactory’ results. Forecasts made with a 3-year lead time began to deteriorate, since only 11 of the gages produced ‘satisfactory’ results. 8 of the gages produced ‘unsatisfactory’ forecasts for RSR. However, Gages 1 and 2 in the UCRB produced ‘good’ forecasts. Forecasts continued to deteriorate with the 4-year and 5-year lead time forecasts; the number of ‘satisfactory’ gages decreased and the number of ‘unsatisfactory’ gages increased.

The bar chart for R reveals a similar pattern in the performance of the forecast (Figure 17). The 1-year lead time provided the best set of forecasts with 20‘satisfactory’ results. The 2-year lead time provided acceptable forecasts, with 20 ‘satisfactory’ results as well. At a 3-year lead time, 11gages produced ‘satisfactory’ forecasts, while 8 gages produced ‘unsatisfactory’ forecasts. Forecasts at a 4-year lead time provided only 5 ‘satisfactory’ results. Finally, the 5-year lead time produced 4 ‘satisfactory’ results. This agreed with the results from the RSR in that the forecast deteriorated as a longer lead time was used, resulting in fewer ‘satisfactory’ forecasts and more ‘unsatisfactory’ forecasts.

The bar chart for NSE agrees with the trends in the bar charts for RSR and R (Figure 17). The 1-year lead time proved to have the best set of results, with 20 gages having ‘satisfactory’ forecasts. The 2-year lead time shows that all of the gages produced ‘satisfactory’ forecasts with the exception of Gages 18 and 21 in the RGB, which had ‘unsatisfactory’ forecasts. At the 3-year, 4-year, and 5-year lead times, the number of gages with ‘satisfactory’ results continued to decrease, while the number of gages with ‘unsatisfactory’ results increased; this is in agreement with the RSR and R results.

Box plots showing the measured and predicted streamflow values at a 1-year lead time for the gages within the UCRB are displayed in Figure 18. The box plots indicate that the KStar model produced accurate forecasts. The box representing the 25th to 75th percentiles was nearly identical for the measured and predicted data. This indicated that the model was able to capture the streamflow variability. The whiskers extending to the 5th and 95th percentiles represent extreme events. The model was able to forecast the majority of the extreme events, but was unable to capture the entire extent of the whiskers. This indicated that the model was capable of forecasting very wet or dry years, which is useful for water managers.

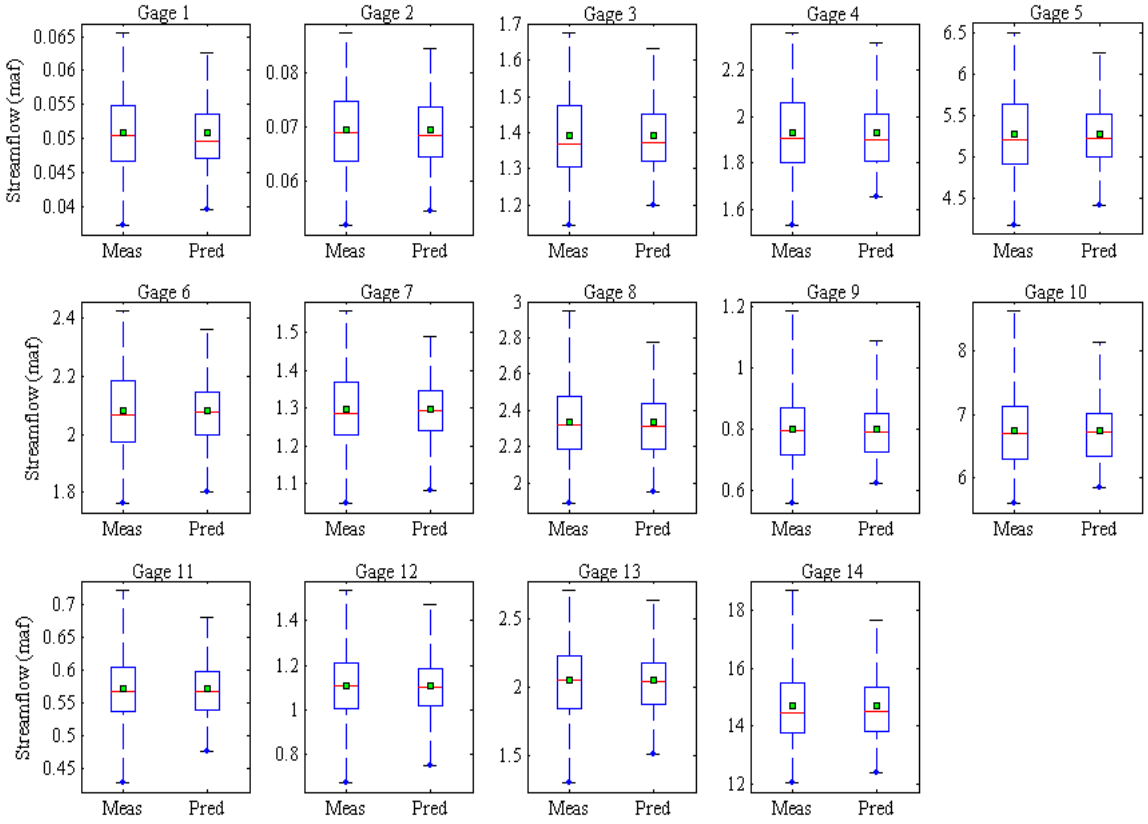

Figure 18: Box plots between measured and predicted streamflow using KStar for the gages within the UCRB. The box represents the interquartile range (25th-75th percentile). The whiskers extend from the 5th to 95th percentile values. The solid line inside the box shows the median value (50th percentile), and the solid dot represents the mean of the data.

Figure 19 displays the non-exceedance probability plots for the gages in the UCRB at a 1-year lead time. The plots indicate that the forecasts have an error of 5% or less for 50-70% of the predictions made at each gage. This indicates that the majority of the predictions had a very low error. The forecasts increased to 80-95% of the predictions that have 10% or fewer errors. This is an indication that nearly all of the predictions resulted in few errors. Once again, the forecasts improve to 90-100% of the predictions that have 15% or fewer errors. This further indicates that the models were able to produce accurate forecasts with few errors at the 1-year lead time.

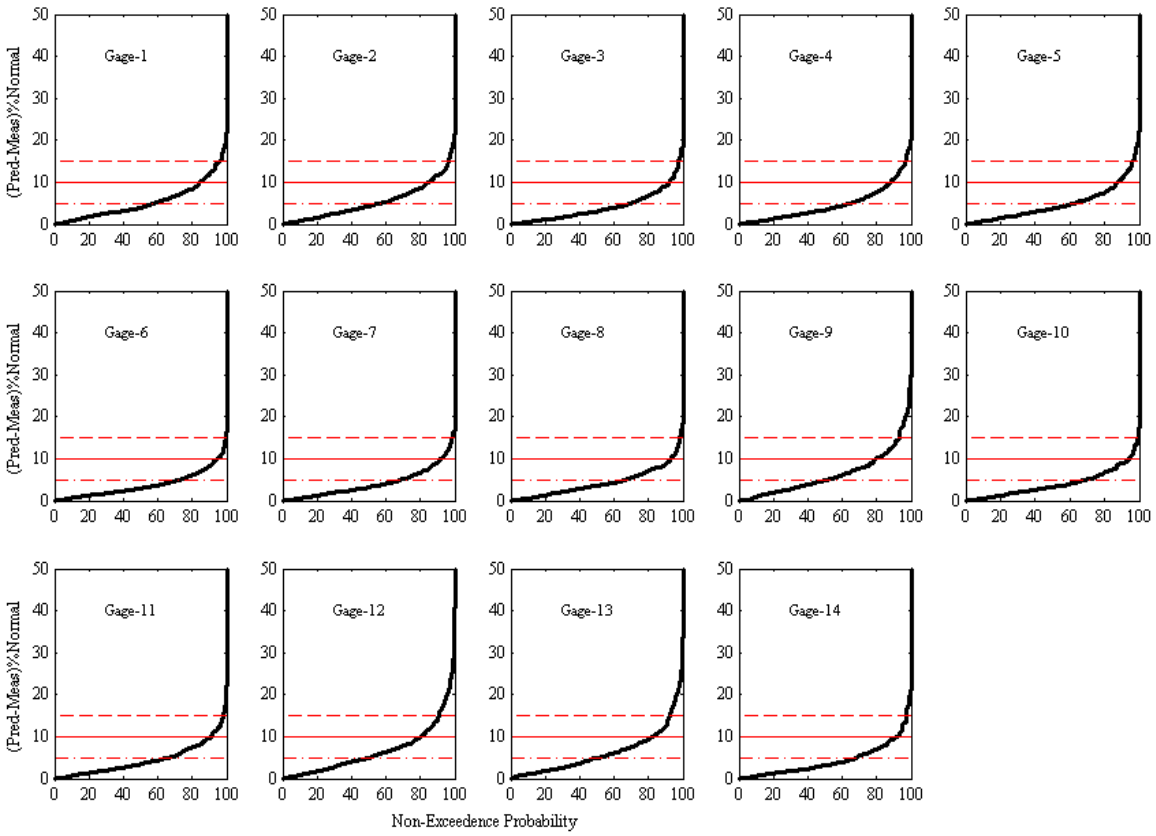

Figure 19: Non-exceedance probability plots using KStar for the gages within the UCRB. The horizontal dash-dot line shows an error value of 5%, the horizontal solid line shows an error value of 10%, and the horizontal dashed line shows an error value of 15%.

Figure 20 shows a bar chart of the LEPS SK for the 21 gages, using a 1-5 year lead time. The LEPS SK incorporates climatology and indicates the position of the predicted value compared to the measured value. All of the lead times for each gage are shown to produce 'good' forecast models (LEPS SK > 10), which indicates that the model is better than climatology. The bar charts indicate that the 1-year lead time produced the best forecast models as compared to the other lead times for the majority of the gage stations. There are a few exceptions, such as at Gage 1, where the 2-year lead time is best, and at Gage 17, where the 3-year lead time is the best. The LEPS SK also tends to agree with the RSR, R, and NSE performance measures, in which the model performance deteriorates as longer lead times are used.

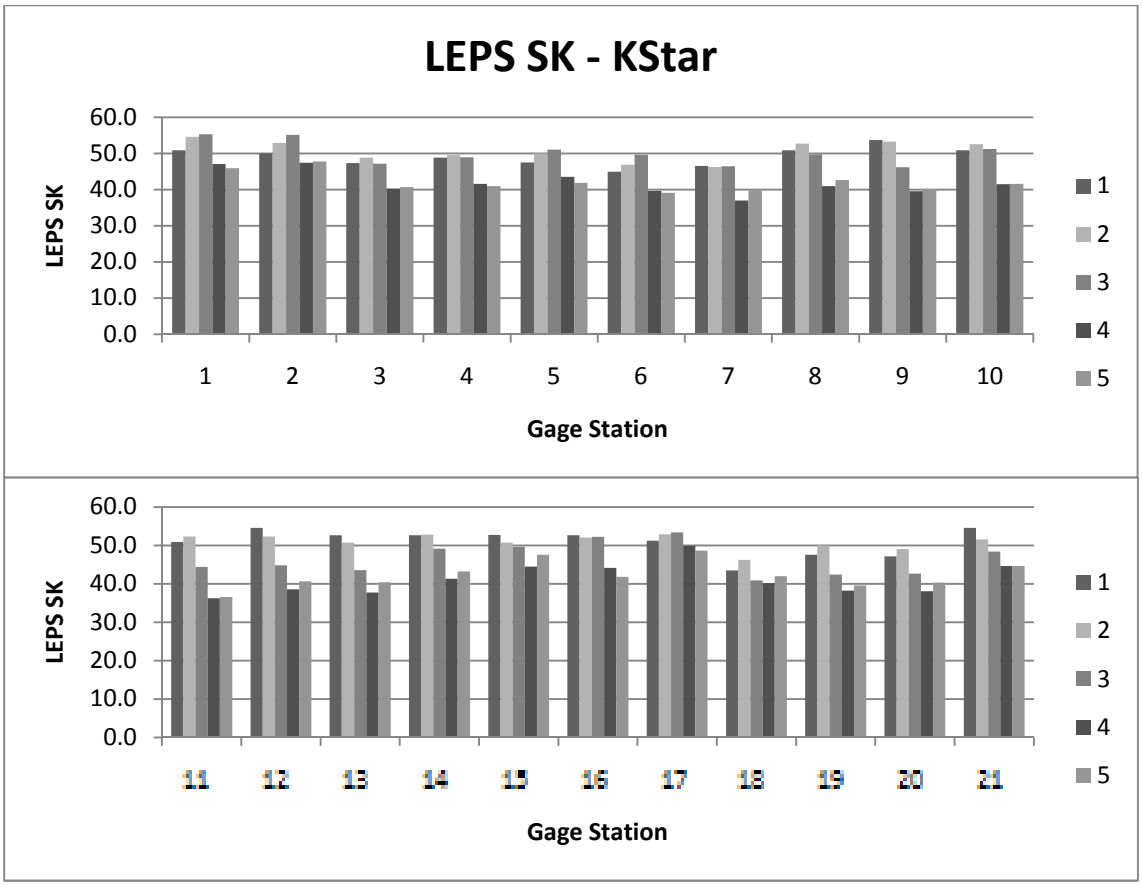

Figure 20: Bar graph displaying the LEPS SK obtained for the 21 streamflow gages using KStar. Each climate division has 5 bars that represent the LEPS SK obtained from 1-5 year lead times (from left to right).

Overall, the results indicate that 'satisfactory' forecasts were made at the 1-year and 2-year lead times. There was a notable decrease in model performance as longer lead times were used in the forecasts. 'Satisfactory' results were still obtainable at the longer lead times; however, in comparison, the 1-year and 2-year lead times produced better forecasts. The box plots (Figure 18) and non-exceedance probability plots (Figure 19) reveal that the model was capable of forecasting extremely wet and dry years with few errors at the 1-year lead time.

In addition, other models were run so that different combinations of oscillation indices could be used to predict streamflow. However, those models produced

'unsatisfactory' results when any of the oscillations were dropped as predictors in the model (results not shown).

3.4.2 MLR model

The KStar model was compared to the MLR modeling approach. As shown in Figure 21, which indicates the spatial representation of the 1-year forecast result for the 21 gages, all of the gages produced 'unsatisfactory' forecasts. In addition, all results obtained for the 1-5 year lead times were 'unsatisfactory' and are not shown. The MAE and RMSE for MLR were much higher than those obtained for the KStar model. However, Figure 22 showed that the LEPS SK was 'satisfactory' (LEPS SK > 5) or 'good' (LEPS SK > 10) for the majority of the forecast models, indicating that the MLR model was better than climatology. However, there were some instances where climatology outperformed the MLR model (LEPS SK < 5).

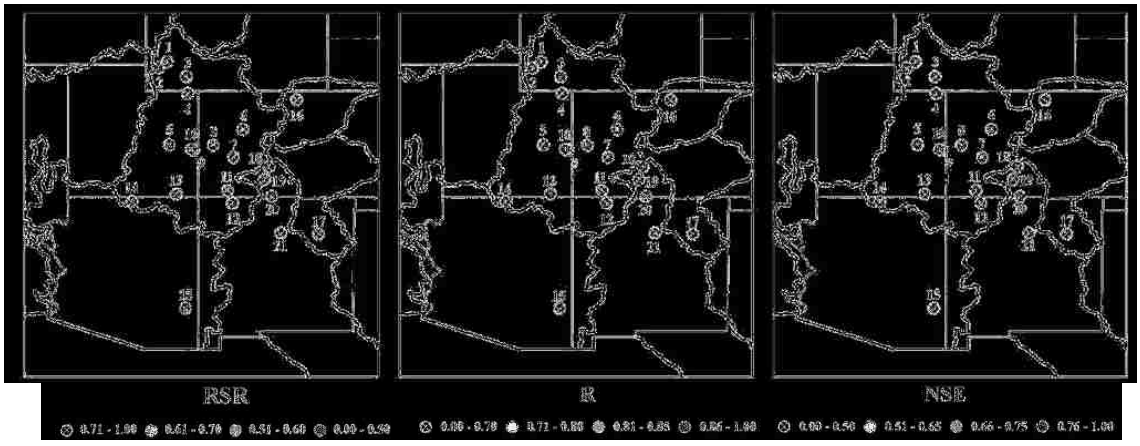

Figure 21: Results displayed in a spatial map using (A) KStar and (B) MLR at a 1-year lead time. Performance measures shown are the RSR, R, and NSE

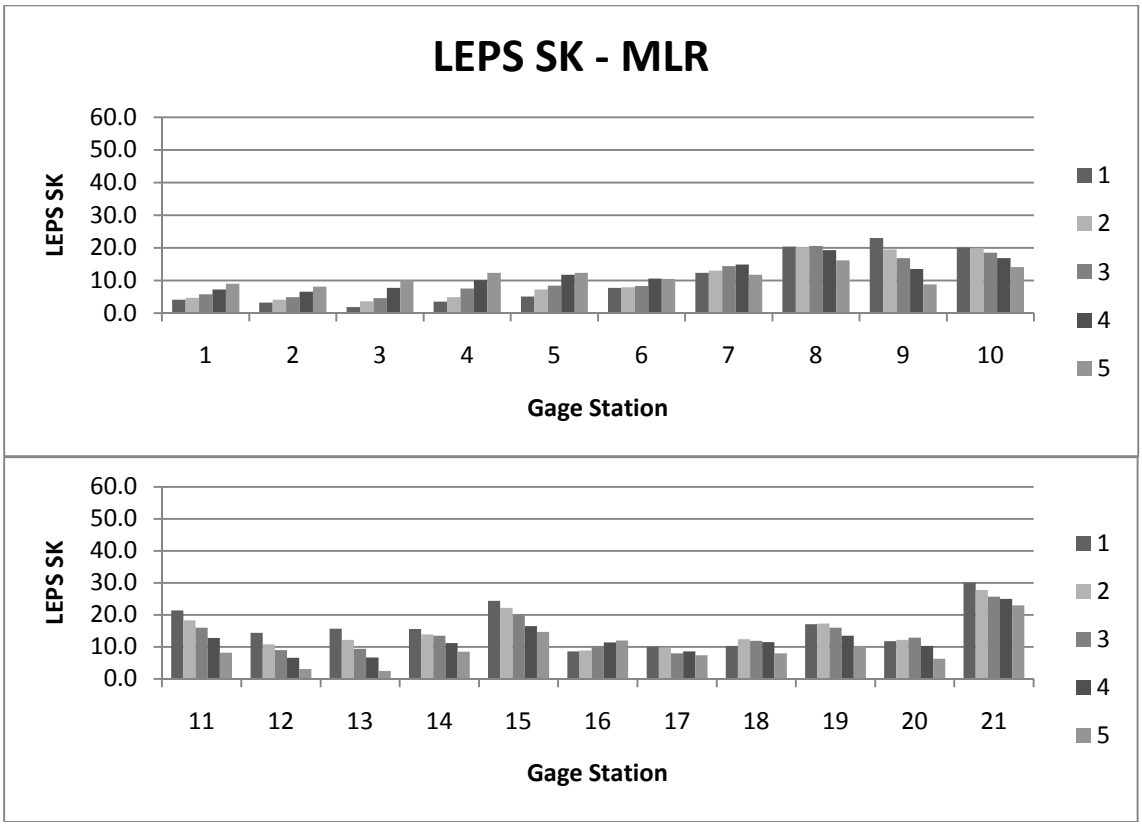

Figure 22: Bar graph displaying the LEPS SK obtained for the 21 streamflow gages using MLR. Each climate division has 5 bars that represent the LEPS SK obtained from 1-5 year lead times (from left to right).

It was unclear as to which lead time provided the best forecasts for some of the gage stations. However, for those stations that had a 'good' forecast model, the MLR model agreed with the KStar model in showing that the 1-year lead time produced the best set of forecasts. The LEPS SK for KStar was much better than those obtained for MLR at all gage stations.

3.5 Discussion and Conclusion

This study used paleoclimate reconstructions to generate a streamflow forecast for 21 gages in the western United States through the use of KStar, a simple yet robust data-driven model. The reconstructions can address the problem of a limited period of record. The instrumental record often limits the use of low-frequency oscillations, such as the

PDO and AMO, because the study of their behavior is often limited to one or two oscillation phases in the instrumental record (Gray et al., 2004; Hidalgo 2004; Timilsena et al., 2009). Paleoclimate reconstructions allow for these low-frequency oscillation indices to be fully studied and incorporated into a forecasting model.

The KStar model, used in this study, was able to produce 'satisfactory' and 'good' streamflow forecasts at a 1-year and 2-year lead time, based upon the performance measures adopted from Moriasi et al. (2007). These yearly forecasts may be disaggregated into finer temporal resolutions through the procedure provided by Kalra and Ahmad (2011).

The spatial map in Figure 16indicates which gage stations yielded 'good' or 'satisfactory' results at a 1-year lead time. It reveals an important interpretation of the model, in that the model may be better suited for forecasting larger, downstream flows. Several of the tested gages are located in the headwaters of the rivers, and the majority of these gages produced 'satisfactory' results. At longer lead times, the gages in the headwaters tended to produce 'unsatisfactory' results. Those gages that are further downstream in the basin tended to have 'good' results, as indicated by Gages 9, 10, 13, 14, and 17 (Figure 16). At longer lead times, the downstream gages proved to be more resilient, with Gage 17 being the only gage that produced 'good' results at all of the lead times for RSR, R, and NSE.

The downstream gages are representative of much of the basin, whereas the gages located in the headwaters are representative of the local area only. Smaller areas may not be affected by one or more of the oscillations, or they may be predominately affected by a single oscillation. Larger basin-scale areas are more likely to be affected by all of the

oscillations, and the streamflow of the downstream gages will reflect their signals. The headwater and downstream gages serve as an important designation because downstream gages often are used by water managers to assess the available water supply. The ability to forecast flows at these downstream gages with a 1-2 year lead time allows for better water planning and management.

Several studies have examined the correlation between oscillation indices and streamflow variability (e.g. Gutzler et al., 2002; Hamlet & Lettenmaier, 1999; Kalra & Ahmad, 2009; McCabe et al., 2007; Redmond & Koch, 1991; Tootle et al., 2005). However, the coupling of oscillation indices, such as the Pacific indices of ENSO and PDO, has been shown to reflect higher streamflow variability as compared to examining the oscillations individually (Gutzler et al., 2002; Timilsena et al., 2009; Tootle et al., 2005). Furthermore, it is clear that oscillations in both the Pacific and Atlantic play a significant role in climate variability across the United States (Gray et al., 2003; McCabe et al., 2004; McCabe et al., 2007; Tootle et al., 2005). In this current study, all four oscillation indices (i.e. ENSO, PDO, AMO, and NAO) were used together in order to represent the interactions of Pacific and Atlantic variability and to extract the maximum information from the indices. Although the results are not shown in this study, the removal of one or more of the oscillation indices resulted in a reduction of the model performance because information becomes lost.

The KStar model is shown to outperform the standard MLR technique. The KStar model develops forecasts based upon the premise that similar inputs yield similar results. KStar, like other instance-based learners, classifies the output from a set of inputs based on similar examples from the training set (Witten et al., 2011). The KStar model takes the

set of oscillation indices of the current year and finds examples that have similar values as the current set of indices to produce a streamflow forecast. This type of data-driven modeling works for streamflow forecasting, because there is extensive research focused on streamflow variability and their correlation with phases of oceanic-atmospheric oscillations (e.g. Gutzler et al., 2002; Hamlet & Lettenmaier, 1999; Kalra & Ahmad, 2009; McCabe et al., 2007; Redmond & Koch, 1991; Tootle et al., 2005). However, the relationship between streamflow and oscillations is inherently non-linear, which makes it difficult for the MLR model to extract relationships using regression methods. In comparison with climatology, MLR has a better forecast skill for the majority of the gages tested, as shown by the LEPS SK (Figure 22); however, KStar outperforms MLR at all of the gages (Figure 20).

The results of this study contributes to a better understanding of the long-term impacts of oceanic-atmospheric oscillations on streamflow variability in the western United States by extending the period of record back to 1658. This is opposed to previous studies, which were limited to a period of record of 50 to 100 years (e.g., Gutzler et al., 2002; Hamlet & Lettenmaier, 1999; Hunter et al., 2006; Kalra & Ahmad, 2009; McCabe et al., 2007; Redmond & Koch, 1991; Tootle et al., 2005). The shorter instrumental record is an important limitation in data-driven modeling. In order to use indices as predictors for streamflow, the raw data must be filtered in order to attenuate short-range fluctuations and extract the long-term oscillations (Hidalgo, 2004; McCabe et al., 2004; Pekarova et al., 2002; Probst & Tardy, 1987). The raw dataset contains a high degree of variance (noise), which repeatedly misclassifies new instances and inevitably lowers the

performance of the model (Witten et al., 2011). A filtered dataset is required for instance-based learners, where a smooth surface that is free of gaps and large jumps is needed so that the model does not make decisions to its detriment (Cleary & Trigg, 1995). The filtered dataset used in this study helped the model to identify the phases of the oscillations and their corresponding correlation to streamflow variability; as a result, the model performance improved significantly.

This current study revealed that the data-driven model, KStar, offers a quantitative forecast that yields 'good' and 'satisfactory' results for up to two years into the future. By using the KStar model, an improvement can be made over current forecasting tools that provide seasonal qualitative forecasts. When all four oceanic-atmospheric oscillations indices become available, as early as January 1st of the current year, the model provides 'good' and 'satisfactory' forecasts for the upcoming water year. The forecasts are better when compared to climatology. The proposed modeling approach provides an alternative to complex, physically-based simulations, and is expected to be useful for long-term water resources management.

Chapter 4: Long-Range Precipitation Forecasts Using Paleoclimate Reconstructions in the Western United States

4.1 Introduction

Fluctuations in precipitation impact the water supply, affect biota, and modulate extreme events, such as floods and droughts (Cayan et al., 1998). Precipitation forecasting poses a greater challenge than other hydrologic variables due to its high temporal and spatial variability (Georgakakos & Hudlow, 1984). For the western United States, where water resource issues are paramount, precipitation forecasts with longer lead times become important (Guttman & Quayle, 1996). A long-range precipitation forecast of a year or more allows water resource managers to allocate water supplies and better plan for reservoir operations.

Oceanic-atmospheric oscillations allow for long-range precipitation forecasts because their effects on precipitation response lag by several months to over a year (Gray et al., 2003; Gutzler et al., 2002). Several oceanic-atmospheric oscillations have been identified that may contribute to climate variability across the western United States. These oscillations are indicators of the atmospheric conditions over the Pacific Ocean and Atlantic Ocean, which are typically indicated by warm or cool phases. These phases are identified with changes in sea surface temperatures (SSTs) and atmospheric flows, which results in variations to weather patterns and climate. Previous studies indicate that the oscillations of El Niño – Southern Oscillation (ENSO), Pacific Decadal Oscillation (PDO), Atlantic Multi-decadal Oscillation (AMO), and North Atlantic Oscillation (NAO) exhibit strong correlations with precipitation throughout the western United States(e.g., Cayan & Webb, 1992; Cayan et al., 2010; Clark et al., 2001; Enfield et al., 2001;Gutzler et al., 2002; Hidalgo & Dracup, 2003; Hunter et al., 2006; Kim et al., 2008; Lee et al.,

2004; Mantua et al., 1997; McCabe et al., 2004; McCabe & Dettinger, 1999; Redmond & Koch, 1991; Ropelewski & Halpert, 1986). Although, these studies indicate that ENSO, PDO, AMO, and NAO show the most influence on precipitation variability within the western United States, it is challenging to incorporate their teleconnections in a physically-based model (Kalra & Ahmad, 2009). The Climate Prediction Center of the National Weather Service (NWS) uses a collection of statistically-based forecast tools to provide qualitative probabilistic precipitation outlooks up to 12.5 months into the future (NWS, 2008).

An alternative to physically-based models are data-driven models, such as artificial neural networks (ANN), support vector machines (SVM), linear regression, decision trees, and instance-based learners. The data-driven approach can be used to examine hidden relationships between oceanic-atmospheric oscillations and precipitation (Partal & Kisi, 2007; Silverman & Dracup, 2000; Zeng et al., 2011). Data-driven models generally perform better with a larger training dataset (Witten et al., 2011). The inclusion of more examples helps the model in identifying critical relationships with a higher degree of certainty. Data-driven models are often compromised by discontinuities and noise in the training dataset, and require a smooth training surface (Cleary & Trigg, 1995). A moving average filter is often used to smooth out short-range fluctuations and extract long-range climatic variations (Pekarova et al., 2003; Probst & Tardy, 1987).

High resolution paleoclimatic proxy reconstructions, especially those derived from tree-ring chronologies, can be used to extend the observed record for use in a data-driven long term precipitation forecast (Prairie et al., 2008). Reconstructions produce absolutely dated time series, which provide an opportunity to extend the climatic and

hydrologic conditions beyond the instrumental record (Cook, 1992; Trenberth & Otto-Bliesner, 2003; Villalba et al., 2011). A study by Dettinger et al. (1998) revealed that the spatial and temporal characteristics of climatic variability that are observable in the instrumental record are also present in tree ring chronologies. Examination of precipitation reconstructions show a strong influence from oceanic-atmospheric oscillations similar to that of the instrumental records (Barlow et al., 2001; Biondi et al., 2001; D'Arrigo et al., 2001; Gedalof & Smith, 2001; Gray et al., 2004a; Hidalgo, 2004; MacDonald & Case, 2005; Villalba et al., 2011). In addition, the longer period of record allows for further examination of factors influencing climate variability, especially from multi-decadal oscillations such as the PDO and AMO (Gray et al., 2003). In the western United States, decadal fluctuations account for 20%–45% of the annual precipitation variance (Cayan et al., 1998). Since the behavior between oceanic-atmospheric oscillations and precipitation are similar in both the reconstructions and the instrumental record, a precipitation forecast model utilizing reconstructions is possible.

The primary objective of this research was to extend the lead time for precipitation forecast by using a data-driven model that incorporates both paleoclimate proxy reconstructions and instrumental record. Reconstructions of oscillations have not yet been incorporated into a precipitation forecast model and are expected to improve the forecast. In this study, the KStar data-driven model, a nearest neighbor algorithm with a generalized distance function based on entropy, was used to develop a forecast model. Reconstructions were available for ENSO, PDO, AMO, and NAO. The model used a dataset containing reconstructions and instrumental records. A moving average filter was applied to the dataset to remove anomalies and aid in the identification of oscillation

phases. A 10-fold cross-validation technique and a lead time approach were used, where the oscillation indices of the current year were used as inputs to forecast precipitation for the following year. In addition, the relative role of individual and coupled oscillations in estimating precipitation was explored.

The results of the KStar model were compared to the multiple linear regression (MLR) modeling approach. Both the models were evaluated using mean absolute error (MAE), root mean squared error (RMSE), RMSE-observations standard deviation ratio (RSR), Pearson's correlation coefficient (R), Nash-Sutcliffe coefficient of efficiency (NSE), and linear error in probability space (LEPS) skill score (SK). A visual examination of the model output was performed using scatter plots, box plots, and non-exceedance probability plots.

This paper is organized as follows. The study region is described in Section 4.2. The data used for this study are described in Section 4.3. Section 4.4 defines the models and performance measures. The results of the data-driven models are presented in Section 4.5. Section 4.6provides a discussion of the findings of the paper and concludes the study.

4.2 Study Area

Water availability is a major concern in the western United States, which is known for low precipitation, aridity, and droughts. Typically, rainfall is less than 50 cm per year west of the 100th Meridian (Anderson & Woosley, 2005). Rapid population growth in the western United States has led to the full allocation of surface water supplies (Anderson & Woosley, 2005). Under these circumstances, it is critical to manage and plan carefully for future water supplies to meet escalating demands. Since climate greatly

varies across the western United States, this study focuses on five specific regions as shown in Figure 23: the Pacific Northwest (PNW), the Great Basin (GB), the Upper Colorado River Basin (UCRB), the Lower Colorado River Basin (LCRB), and the Rio Grande Basin (RBG).

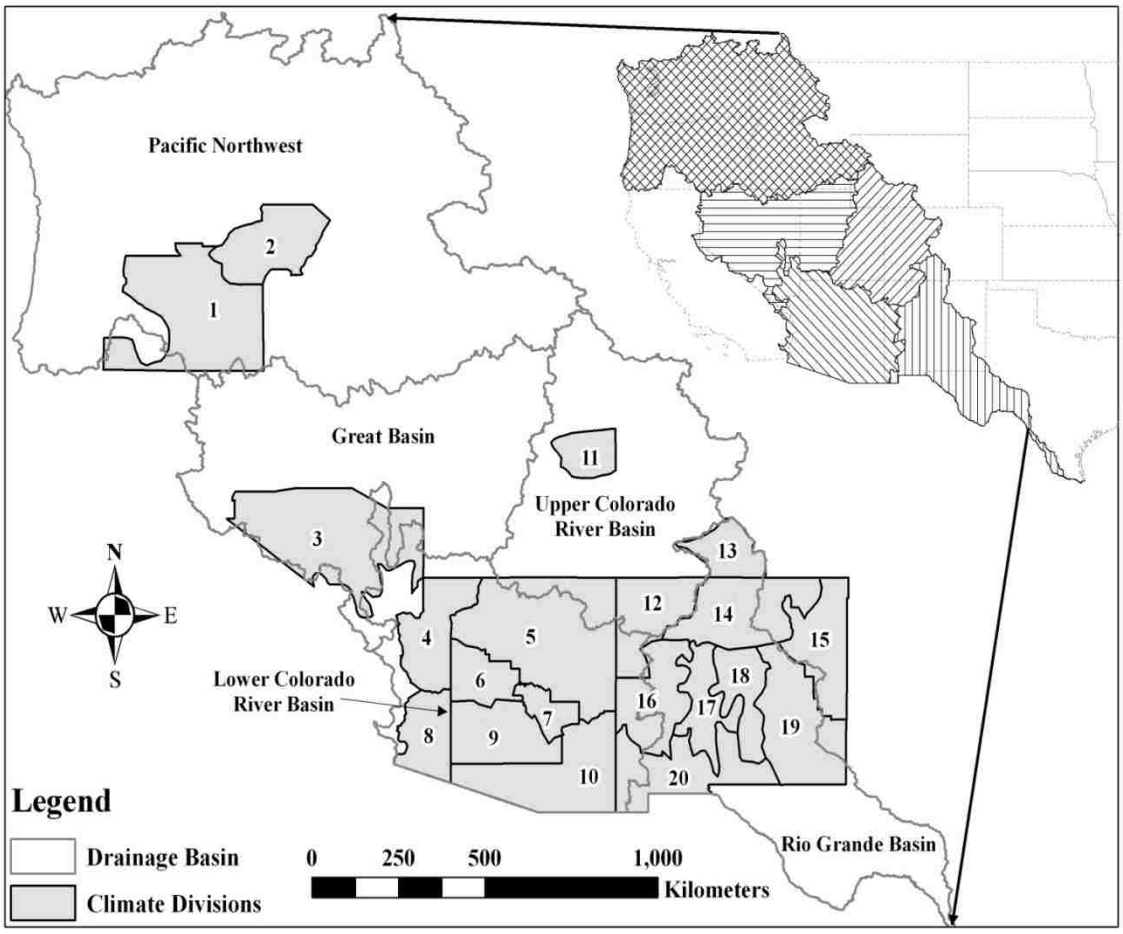

Figure 23: Study area showing the major drainage basins and their representative climate divisions in the western United States.

Since precipitation exhibits large spatial variability, this study simplifies the model in terms of spatial coverage by making use of climate divisions developed by the National Climatic Data Center (NCDC). Climate divisions represent regions that are relatively climatically homogeneous (McCabe & Dettinger, 1999). Climate divisions in the western states are primarily divided by drainage basins (Guttman & Quayle, 1996).

This is important as it allows precipitation to be representative of streamflow, drought, and water supply in the western U.S. The spatial variability is accounted for by testing several climate divisions, which represent different regions within the western U.S. Climate division precipitation also reflects the influences of oceanic-atmospheric oscillations (McCabe & Dettinger, 1999). This allows for climate divisions to sufficiently provide the necessary spatial coverage while maintaining the relationships between oceanic-atmospheric oscillations and precipitation.

Climate divisions are typically identified by state and division number; for example, AZ1 refers to Climate Division 1 in Arizona. For simplicity, climate divisions used in this study were labeled 1-20, as displayed in Figure 23. The actual climate division numbers and locations corresponding to these numbers were reported in Table 11. Precipitation variability associated with each climate division was displayed using box plots in Figure 24.

Table 11: Climate Division Description.

Climate Division	Region	Description
1	PNW	Oregon Climate Division 7: South Central Oregon
2	PNW	Oregon Climate Division 8: Northeast Oregon
3	GB	Nevada Climate Division 3: South-Central Nevada
4	LCRB	Arizona Climate Division 1: Northwest Arizona
5	LCRB	Arizona Climate Division 2: Northeast Arizona
6	LCRB	Arizona Climate Division 3: North-Central Arizona
7	LCRB	Arizona Climate Division 4: East-Central Arizona
8	LCRB	Arizona Climate Division 5: Southwest Arizona
9	LCRB	Arizona Climate Division 6: South-Central Arizona
10	LCRB	Arizona Climate Division 7: Southeast Arizona
11	UCRB	Utah Climate Division 6: Uinta Basin
12	UCRB	New Mexico Climate Division 1: Northwestern Plateau
13	RGB	Colorado Climate Division 5: Rio Grande Drainage
14	RGB	New Mexico Climate Division 2: Northern Mountain
15	RGB	New Mexico Climate Division 3: Northeastern Plains
16	RGB	New Mexico Climate Division 4: Southwestern Mountains
17	RGB	New Mexico Climate Division 5: Central Valley
18	RGB	New Mexico Climate Division 6: Central Highlands
19	RGB	New Mexico Climate Division 7: Southeastern Plains
20	RGB	New Mexico Climate Division 8: Southern Desert

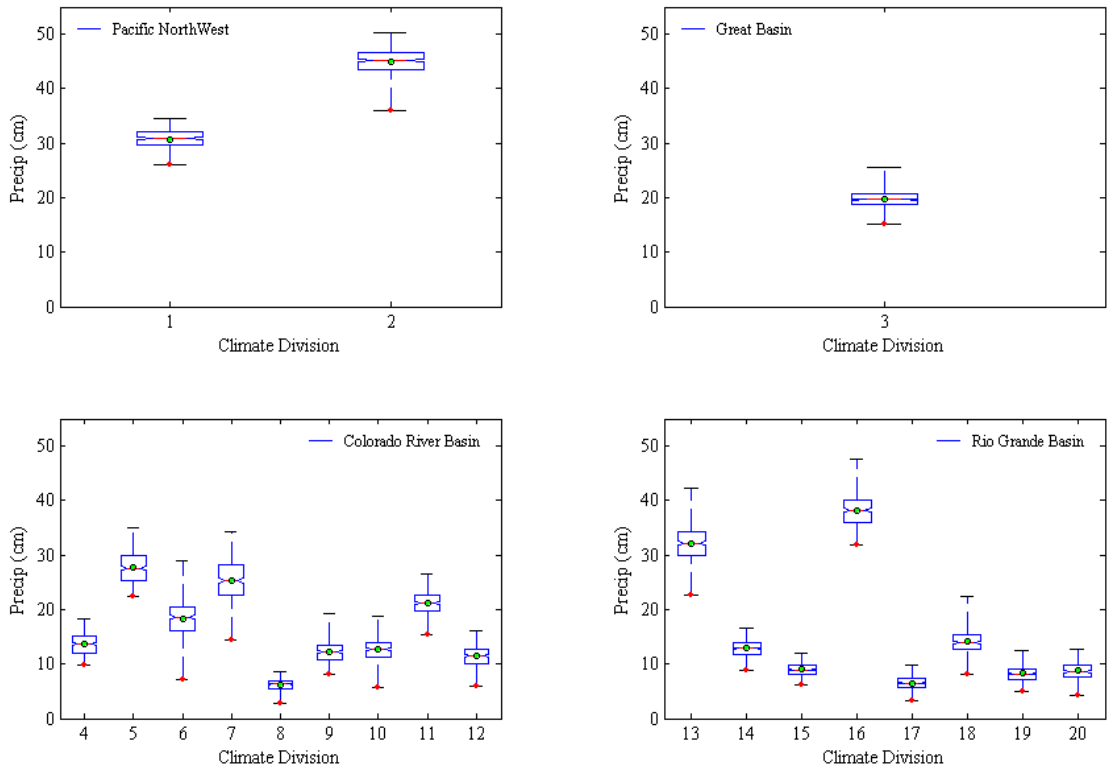

Figure 24: Box plots depicting the annual precipitation variability associated with the 20 climate division using data from 1658-2007. The box shows the interquartile range (25th-75th percentile). The whiskers extend from the 5th to 95th percentile values. The solid line inside the box shows the median value (50th percentile), and the solid dot represents the mean of the value. The climate divisions are divided according to their respective regions in the western United States.

4.2.1 Pacific Northwest

For this study, the PNW was defined as region 17 from the United States Geological Survey's Hydrologic Unit Map (Seaber et al., 1987). This region contains the 668,000 km^2 Columbia River Basin, which drains portions of Wyoming, Montana, Idaho, Washington, Oregon, Alberta, and British Columbia (Figure 23). The PNW is known for its wet and humid climate. However, the PNW exhibits a highly variable climate with coastal areas receiving more than 75 cm of precipitation per year, approximately 250 cm in the Cascades and less than 50 cm in the interior regions (Climate Impact Groups, 2011). From October to March, this region receives approximately two-thirds of its

annual precipitation (Climate Impact Groups, 2011). The PNW is represented by Climate Divisions 1 and 2, located in the interior of the PNW (Figure 23).

4.2.2 Great Basin

The GB region has an area of 477,664 km^2 and covers much of Nevada and western Utah as well as portions of California, Washington, Idaho, and Wyoming (Figure 23). The GB region has an average annual precipitation of about 28 cm per year (Flaschka et al., 1987). Precipitation greatly varies across the basin, with some southern valleys receiving 7 – 13 cm per year, northern valleys receiving 40 – 50 cm, and as much as 150 cm in the high mountain ranges near the eastern and western borders (Flaschka et al., 1987). Climate Division 3 covers the majority of the southern GB (Figure 23).

4.2.3 Colorado River Basin

The Colorado River Basin drains an area of 637,000 km^2 and includes portions of Colorado, Wyoming, Utah, New Mexico, California, Arizona, Nevada, and Mexico; it discharges into the Gulf of California (Figure 23). The average annual precipitation in this basin is about 40 cm per year (Christensen et al., 2004). The Colorado River is often viewed as a two-basin system, with the gage at Lees Ferry, Arizona serving as the division. The downstream basin, defined as the LCRB, serves California, Arizona, Nevada, and Mexico. The LCRB is represented by Climate Divisions 4-10 (Figure 23). The upstream basin, defined as the UCRB, serves Colorado, Wyoming, Utah, and New Mexico. Climate Division 11 represents the northern portion of the UCRB, and the southern portion of the basin is partially represented by Climate Division 12 (Figure 23).

The annual UCRB snowmelt runoff accounts for 90% of the Colorado River flow (Christensen et al., 2004).

4.2.4 Rio Grande Basin

The RGB drains an area of 471,900 km^2 and includes portions of Colorado, New Mexico, Texas, and Mexico; it discharges into the Gulf of Mexico (Figure 23). The average annual precipitation in this basin ranges from about 15 – 20 cm per year in the semi-arid regions to as much as 64 cm in the higher mountains (State of New Mexico Water Quality Control Commission, 2006). The RGB is represented by Climate Divisions 13-20 (Figure 23).

4.2.5 Oceanic-Atmospheric Oscillations

Yearly indices of oceanic-atmospheric oscillations provide valuable information about the current state of each oscillation and their magnitudes. This study used these indices as the predictors for precipitation forecasting, including the ENSO, PDO, AMO, and NAO. Although the inclusion of other oscillations would be beneficial, paleoclimate reconstructions were limited to these four.

ENSO is defined as a warming or cooling of at least 0.5°C in the east-central tropical Pacific Ocean over a period of 1 – 2 years, which occurs approximately every 2 – 7 years (Ahrens, 2007). This study uses the winter (October to March) Southern Oscillation Index (SOI) as a measure of ENSO, which is the difference between the sea-level pressure (SLP) anomalies measured at Tahiti and Darwin, Australia (Mann et al., 2000; McCabe & Dettinger, 1999). El Niño is the warm phase of ENSO that typically enhances warm winter storms and brings above-normal precipitation in the southwestern

United States, while the Pacific Northwest receives below-normal precipitation. Similarly, the cool phase known as La Niña is associated with below-normal precipitation in the southwestern United States and above-normal precipitation in the Pacific Northwest (Mann et al., 2000, McCabe & Dettinger, 1999).

The PDO is derived from the leading principal component of monthly sea-surface temperature (SST) anomalies in the North Pacific Ocean, pole ward of 20°N, which exhibits decadal-scale oscillations that typically last between 20 to 30 years (Mantua et al., 1997). Although the causes of the PDO are currently unknown, warm phases of the PDO are associated with above-normal SSTs and cool phases are associated with below-normal SSTs (Ellis et al., 2010). As a result the PDO exhibits ENSO-like conditions, where a warm PDO phase is linked to above-normal precipitation in the southwestern United States and below-normal precipitation in the Pacific Northwest (Mantua et al., 1997). The cool phase is associated with inverse precipitation patterns.

The AMO index is calculated from a 10-year running mean of de-trended SST anomalies between 0°and 70°N in the northern Atlantic Ocean (Enfield et al., 2001; Gray et al., 2004b). The AMO exhibits a multi-decadal oscillation that may last 65 – 80 years, with phases that may persist from 20 – 40 years (Enfield et al., 2001; Gray et al., 2004b; Kerr, 2000). Cool phases have been identified with wet conditions in the southwestern United States (Enfield & Cid-Serrano, 2006). The warm phases have been linked to an increase in drought frequency over parts of the United States including the Colorado River Basin (Ellis et al., 2010; Gray et al., 2003, 2004b; McCabe et al., 2004). The increase in drought frequency may be due to the weakening of easterly atmospheric flows

from above-normal SSTs in the Atlantic and Gulf of Mexico, which reduces the moisture delivery into the western United States (Shubert et al., 2004).

The NAO index is derived from the normalized SLP difference between Ponta Delgada, Azores and Stykkisholmur/Reykjavik, Iceland and it is shown to vary annually or may remain in a single phase for several years (Hurrell, 1995). Although the NAO predominantly influences the eastern United States and Europe, there are a few studies which have linked NAO to western United States climate variability. Hunter et al. (2006) identified the influence of NAO on snow water equivalent (SWE) in the western United States, although a distinct spatial region was not identified. In addition, the phases of NAO shifts the jet stream north during a warm phase and south during a cool phase, which may influence climate variability over the Pacific Northwest and northern Rocky Mountains (Hunter et al., 2006).

4.3 Datasets

The datasets used in this study included oceanic-atmospheric oscillations indices as predictors and climate division precipitation as predictands. Proxy reconstructions for indices and precipitation were available through the National Climatic Data Center (NCDC, 2011). The instrumental records were obtained from various sources. These datasets are described in Table 12.

It should be noted that there are multiple reconstructions for the oscillation indices available through NCDC; however, reconstructions were selected based on the length of the time series and on their relative statistical properties to the instrumental record. It is also important to note that the annual precipitation values are not measured over the

calendar year, but are measured as the water year or cool season precipitation for the climate division. The period of measure for each water year is provided for each reconstruction (Table 2), in which the appropriate period of measure is applied to the instrumental record.

The model used reconstructions from 1658 to 1899 and observed data from 1900 to 2007 into one continuous time series with no overlap so that no biases were introduced into the dataset. For two climate divisions in the Pacific Northwest, Climate Divisions 1 and 2, reconstructions were used from 1705 to 1899 due to a shorter available reconstruction period. This provided a more robust approach as opposed to analyzing the reconstructions and observed dataset individually. The extended dataset also benefited the data-driven model by allowing it to determine relationships from more available examples resulting in improved model performance (Witten et al., 2011).

Table 12: Data Sources for Climate Divisions and Oscillation Indices

Climate Division / Oscillation Index	Source	Period Available	Period used in Study
1, 2	Garfin and Hughes (1997)	1705-1979	1658-1899
	ESRL (2011)	1895-2010	1900-2007
3	Hughes and Graumlich (1996)	6000BC-1996	1658-1899
	ESRL (2011)	1895-2010	1900-2007
5	Salzer and Kipfmueller (2005)	570-1987	1658-1899
	ESRL (2011)	1895-2010	1900-2007
11	Gray et al. (2004)	1226-2001	1658-1899
	ESRL (2011)	1895-2010	1900-2007
13	Grissino-Mayer and Baisan (1998)	1035-1995	1658-1899
	ESRL (2011)	1895-2010	1900-2007
16	Grissino (1996)	136BC-1992	1658-1899
	ESRL (2011)	1895-2010	1900-2007
4, 6, 7, 8, 9, 10, 12, 14, 15, 17, 18, 19, 20	Ni et al. (2002)	1000-1988	1658-1899
	ESRL (2011)	1895-2010	1900-2007
ENSO	Jones and Mann (2004)	1650-1980	1658-1905
	Australian Government Bureau of Meteorology (2011)	1876-2010	1900-2007
PDO	Shen et al. (2006)	1470-1998	1661-1905
	JISAO (2011)	1900-2010	1906-2007
NAO	Luterbacher et al. (2001)	1658-2001	1661-1905
	Hurrell (2011)	1865-2010	1906-2007
AMO	Gray et al. (2004b)	1567-1985	1661-1905
	ESRL (2011)	1856-2010	1906-2007

4.4 Method

To provide a precipitation forecast, this study used a data-driven modeling technique, KStar, to discover relationships between oceanic-atmospheric oscillation inputs and precipitation outputs. The relationships were applied to give a precipitation forecast based upon the current state of the oscillation indices. To simulate a forecast, a one-year lead time was applied to the climate division precipitation. A 10-fold cross-validation technique was used to test the dataset, which is standard practice in data-driven

modeling (Witten et al., 2011). Under this technique, the dataset was randomly divided into 10 equal folds; 9 of these folds were used to train the model, and the remaining fold was used for testing. Forecasts are made for each instance in the fold that is held out for testing. This procedure was executed 10 times, with each fold held out in turn for testing. The results from the 10 folds were averaged together to yield an overall result for the model. This technique allows for forecasts to be obtained for each year in the dataset (i.e. 1658-2007), which are then averaged together to provide a final forecast.

In addition, a sensitivity analysis was performed to determine if an improvement in the forecast was made through a particular set of coupled oscillations. In particular, the model performance of using all oscillation indices together as predictors was compared to the model performance of dropping an index so that only 3 of the 4 oscillations were used as predictors.

The raw data was filtered in order to attenuate the short-range fluctuations and to extract the long-range climatic variations (Hidalgo, 2004; McCabe et al., 2004; Probst & Tardy, 1987). It was also required for data-driven models, where a smooth training surface without large jumps and discontinuities was needed so that the model did not make decisions to its detriment (Cleary & Trigg, 1995). Filtering a dataset prior to analysis is not new and has been incorporated into studies that examine hydrologic and climate fluctuations (e.g. Currie, 1996; De Jongh et al., 2006; Hidalgo, 2004; Garbrecht & Rossel, 2002; McCabe et al., 2004; Pekarova et al., 2003; Probst & Tardy, 1987; Riehl & Meitin, 1979; Riehl et al., 1979). This study used a basic 10-year moving average filter to aid in the detection of long-range climatic variations. Studies by Probst and Tardy (1987) and Hidalgo (2004) revealed that different filtering techniques e.g., cumulative

deviation method, moving average, and weighted moving average yield similar filtered time series, but a difference of one or two years for the localization of minima and maxima was sometimes observed.

4.4.1 Model

This study used the KStar model featured in Weka 3.6.2, which is a data-mining software developed by the University of Waikato, New Zealand (Witten et al., 2011). KStar is an instance-based learner that uses an entropy-based distance measure to compare an instance to pre-classified examples (Cleary & Trigg, 1995). This model works under the assumption that a similar set of input oscillation indices will produce a similar output precipitation. The entropy-based distance function measures the complexity of transforming one instance into another (Cleary & Trigg, 1995), which provides a consistent approach to the handling of attributes and allows a comparison to the entire dataset. This is in contrast to simple k-nearest neighbor instance-based learners, where comparisons are limited to the shortest distance to a defined number of neighbors. A study by Solomatine et al. (2008) concludes that instance-based learner methods can be important alternatives to statistical models and non-linear methods such as ANNs, and may play an important role in hydrological forecasting, thus complementing physically-based models. They also have the advantage of being more transparent than ANNs, thus may be more easily accepted by decision-makers (Solomatine et al., 2008). A further strength of instance-based learners is that they are simple models that can be run very quickly. This serves as an advantage over physically-based models, which often require a large amount of input data and calibration. Instance-based learners have been applied in data-driven estimations and forecasts of evaporation, soil moisture content, rainfall-

runoff, and streamflow (Elshorbagy et al., 2010; Solomatine et al., 2008; Terzi, 2007). Cleary and Trigg (1995) provides a detailed discussion of KStar.

This study did not aim to provide an extensive comparison among forecast models, but rather to complement existing forecast models. However, a multiple linear regression (MLR) model was subjected to a similar set of tests to provide base line results.

4.4.2 Performance Measures

The performances of the KStar and MLR forecast models were evaluated with the MAE, RMSE, RSR, R, NSE and LEPS SK. In addition, scatter plots, box plots, and non-exceedance probability plots were used to provide a visual inspection of the model performance in comparison to the measured data. These performance measures were outlined in Moriasi et al. (2007), which provides ratings that were recommended for models being used for high-impact decisions that required much stricter performance requirements. Although much more lenient performance ratings would suffice for research oriented studies (Moriasi et al., 2007), this study maintained the strict performance ratings to evaluate the forecast models.

The MAE and RMSE are commonly used to analyze results because they indicate error in the same dimensionality of the measured variable, with a value of 0 indicating a perfect fit (Moriasi et al., 2007; Witten et al., 2011).The RSR, a standardized version of the RMSE, incorporates the standard deviation of the observed dataset, which allows the RSR to be used as an error index that can be scaled or normalized so that it can be compared with other results (Moriasi et al., 2007). This is useful for climate division

precipitation in the western United States, where the climate shows considerable spatial variability. RSR is calculated as the ratio of the RMSE and standard deviation of the measured data. RSR ranges from 0 to a large positive value, with 0 indicating a perfect model (Moriasi et al., 2007). Table 13shows the performance level used in this study, provided by Moriasi et al. (2007), where RSR is categorized as 'very good' if it is between 0.00 and 0.50, 'good' if between 0.50 and 0.60, 'satisfactory' if between 0.60 and 0.70, and 'unsatisfactory' if greater than 0.70.

Table 13: Performance Rating for RSR, R, and NSE.

Performance Rating	RSR*	R	NSE*
Very Good	$0.00 \leq RSR \leq 0.50$	$0.85 \leq R \leq 1.0$	$0.75 \leq NSE \leq 1.0$
Good	$0.50 < RSR \leq 0.60$	$0.80 \leq R < 0.85$	$0.65 \leq NSE < 0.75$
Satisfactory	$0.60 < RSR \leq 0.70$	$0.70 \leq R < 0.80$	$0.50 \leq NSE < 0.65$
Unsatisfactory	$0.70 < RSR \leq 1.0$	$0.00 \leq R < 0.70$	$0.00 \leq NSE < 0.50$

***RSR and NSE obtained from Moriasi et al., [2007]**

R, an index of the degree of linear relationship between the predicted and measured data (Moriasi et al., 2007), ranges from -1 to 1, where -1 indicates a perfect negative relationship, 0 indicates no relationship, and 1 indicates a perfect positive relationship. Moriasi et al. (2007) indicates that $R > 0.5$ or $R < -0.5$ is considered satisfactory. For the purposes of this study, stricter criteria are generated for R, where R is defined as 'very good' if it is greater than 0.85, 'good' if between 0.80 and 0.85, 'satisfactory' if between 0.70 and 0.80, and 'unsatisfactory' if less than 0.70 (Table 13). R is limited to the linear relationship in the data only, and is more sensitive to outliers than to observations near the mean (Legates & McCabe, 1999; Moriasi et al., 2007).

NSE is a normalized measure of the residual variance ("noise") compared to the measured data variance ("information") (Moriasi et al., 2007). NSE indicates how well the predicted and measured data follow a 1:1 relationship (Legates & McCabe, 1999;

Moriasi et al., 2007). NSE ranges from 1 to negative infinity, where a NSE greater than 0 is generally an acceptable level of performance because it indicates that the model is a better predictor than the observed mean of the dataset (Legates & McCabe, 1999; Moriasi et al., 2007). Table 13 shows that the performance level for NSE is categorized as 'very good' if it is between 0.75 and 1.00, 'good' if between 0.65 and 0.75, 'satisfactory' if between 0.50 and 0.65, and 'unsatisfactory' if below 0.50 (Moriasi et al., 2007).

The LEPS SK measures the ability of the model to produce accurate predictions that are weighted more for values that are further from the mean (Potts et al., 1996). The LEPS SK incorporates climatology in the computation and serves as an indicator for accessing the forecast compared to the climatology. LEPS SK ranges from -100 to 100, where a LEPS SK of 0 represents the climatological score or equivalently, random data (Casey, 1998). LEPS SK is considered 'good' if it is greater than 10, 'satisfactory' if it is greater than 5, 'poor' if it is below -5, and 'bad' if it is below -10 (Casey, 1998).

4.5 Results

The results are discussed in the following sections. Section 4.6.1 evaluates the statistical properties of the relationship between oceanic-atmospheric oscillations and precipitation. Section 4.6.2 examines the results obtained when all four oscillations are used for precipitation forecasting. This is followed by the results when using three oscillations for precipitation forecasting in section 4.6.3. Section 4.6.4 provides a comparison of the KStar model with an MLR model.

4.5.1 Statistical Relationship between Oceanic-Atmospheric Oscillation and Precipitation

Before oceanic-atmospheric oscillations can be used as a potential predictor for precipitation, the relationship between the two must be identified. The relationship between each oscillation and precipitation was examined through the correlation between the oscillation indices and climate division precipitation. Table 14shows the relationship between each index (ENSO, PDO, AMO, and NAO) and precipitation in each climate division (Climate Divisions 1-20), using a typical correlation technique. The correlations are based on the annual water year precipitation for each climate division as described in Table 2. A lead time of 1-year was applied to climate division precipitation to determine the significance of each oscillation as a predictor for a 1-year forecast. There were 330 instances in the final dataset. Under a normal distribution with a mean of 0, and a standard deviation of 1, the correlation between a given oscillation index and climate division precipitation was considered significant at the 95% confidence level when R exceeded ± 0.11. This process was useful for identifying which oscillations play a significant role in explaining precipitation for each region in the western United States.

Table 14: Correlation between each oscillation and climate division. The correlation is significant at the 95% confidence level when R exceeds ± 0.11.

Division	Region	ENSO*	PDO	AMO	NAO
1	PNW	-0.07	0.01	**0.12**	**-0.23**
2	PNW	-0.01	**-0.27**	-0.10	**-0.37**
3	GB	**-0.36**	**0.22**	**-0.26**	-0.07
4	LCRB	**-0.41**	**0.30**	0.06	0.09
5	LCRB	-0.43	**0.33**	-0.10	**0.20**
6	LCRB	**-0.20**	**0.41**	**-0.28**	**0.11**
7	LCRB	**-0.13**	**0.30**	**-0.35**	0.04
8	LCRB	**-0.28**	**0.40**	**-0.27**	0.07
9	LCRB	**-0.31**	**0.56**	-0.08	**0.20**
10	LCRB	**-0.11**	**0.30**	**-0.46**	0.02
11	UCRB	-0.03	**0.28**	-0.01	0.04
12	UCRB	**-0.29**	**0.42**	**-0.20**	0.10
13	RGB	**-0.34**	**0.18**	**-0.28**	**-0.21**
14	RGB	**-0.28**	**0.39**	**-0.22**	0.03
15	RGB	**-0.23**	**0.23**	**-0.33**	-0.04
16	RGB	-0.09	**0.63**	-0.08	**0.21**
17	RGB	**-0.13**	**0.43**	**-0.38**	0.03
18	RGB	**-0.11**	**0.13**	**-0.44**	**-0.12**
19	RGB	**-0.17**	**0.22**	**-0.41**	-0.07
20	RGB	**-0.30**	**0.24**	**-0.36**	-0.07
Total climate divisions significant at the 95% confidence interval		16	19	14	8

*ENSO shown as SOI, where a negative correlation with SOI corresponds to a positive correlation with ENSO.

ENSO was shown to have a significant relationship with 16 of the 20 climate divisions used in this study. Climate divisions in the PNW did not show a significant response to ENSO. However, ENSO did have a significant relationship with precipitation in the GB and in the LCRB. Climate Division 11 in the UCRB did not show a significant relationship between ENSO and precipitation, while Climate Division 12 did show a significant relationship. The majority of climate divisions in the RGB indicated a significant relationship between ENSO and precipitation.

The PDO was shown to have a significant relationship with 19 of the 20 climate divisions, as shown in Table 14. Only Climate Division 1 in the PNW did not show a significant response to the PDO.

The AMO was found to have a significant correlation with precipitation in 14 of the 20 climate divisions (Table 14). In the PNW, Climate Division 1 showed a significant correlation with the AMO. However, Climate Division 2 did not exceed the 95% significance threshold. The GB had a significant correlation with precipitation. In the LCRB, precipitation in the majority of the climate divisions showed a significant response to the AMO. Precipitation in Climate Division 11 in the UCRB did not have a significant correlation with the AMO, but precipitation in Climate Division 12 did show a significant correlation with the AMO. The majority of climate divisions in the RGB showed a significant correlation between the AMO and precipitation.

The NAO showed correlations at the 95% significance level with precipitation in 8 of the 20 climate divisions, as shown in Table 14. In the PNW, precipitation in the climate divisions showed a significant correlation with the NAO. The GB did not show a significant correlation between precipitation and the NAO. In the LCRB, only 3 of the 7 climate divisions showed a significant correlation with the NAO. Precipitation in the UCRB did not have a significant correlation with the NAO. In the RGB, only 3 of the 8 climate divisions had a significant correlation between the NAO and precipitation.

4.5.2 All Oscillations

In this section, ENSO, PDO, AMO, and NAO were used together to predict precipitation in the selected climate divisions using the KStar model as described in the methods section.

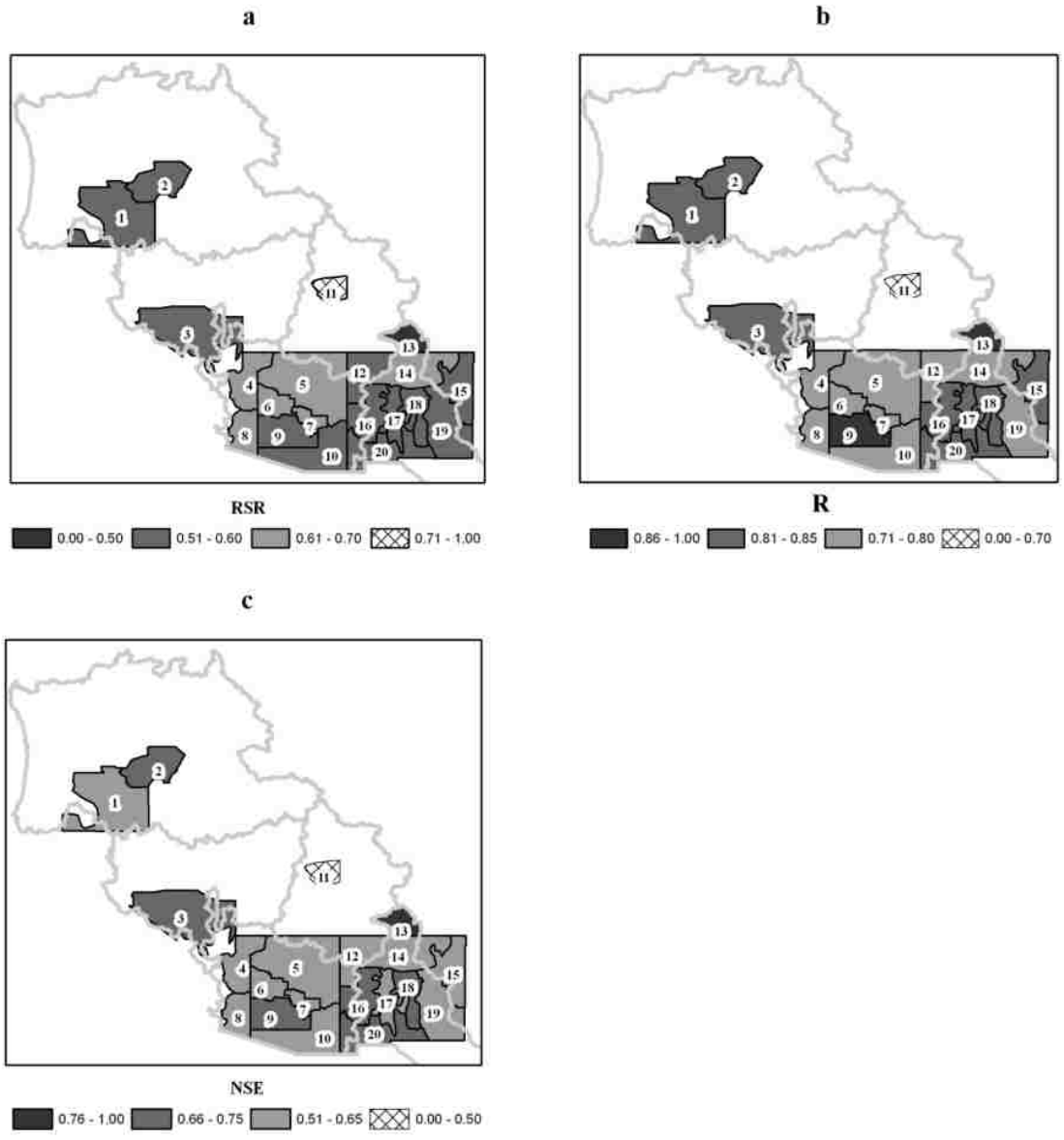

Figure 25: KStar model results obtained when using all four oscillations (ENSO, PDO, AMO, and NAO). Performance measures shown are the (a) RSR, (b) R, and (c) NSE.

Figure **25**adisplays a spatial representation of the RSR values obtained using the KStar model when all oscillations were used to forecast precipitation. Based on RSR, 'satisfactory' ($0.60 \leq RSR < 0.70$) and 'good' predictions ($0.50 \leq RSR < 0.60$) for the majority of the climate divisions. An 'unsatisfactory' predictions ($0.70 < RSR \leq 1.0$) was obtained in the UCRB and a 'very good' prediction was obtained in the RGB. Climate Division1 showed a 'satisfactory' prediction and Climate Division 2 showed a 'good' prediction for precipitation in the PNW. Climate Division 3 in the GB also provided 'good' precipitation predictions. Within the LCRB, 'satisfactory' predictions were made for Climate Divisions 4, 6, 7, and 8. 'Good' predictions were made in the LCRB for Climate Divisions 5, 9 and 10. In the UCRB, Climate Division 11 produced an 'unsatisfactory' prediction, while Climate Division 12 yielded a 'good' prediction. Climate Divisions 15-20 within the RGB showed 'good' precipitation predictions. Climate Division 13 in the northernmost portion of the basin showed 'very good' predictions, while Climate Division 14 showed 'satisfactory' predictions.

Figure 25b shows a spatial map of the R values for the climate divisions in the western United States. The R values generally agree with the RSR values shown in Figure 25a. Climate Division 1 in the in the PNW provided a 'satisfactory' correlation ($0.70 \leq R < 0.80$) and Climate Division 2 provided a 'good' correlation ($0.80 \leq R < 0.85$). Climate Division 3 in the GB had a 'very good' correlation ($0.85 \leq R \leq 1.0$) between the predicted and measured precipitation. A mix of results were obtained in the LCRB. Climate Divisions 4, 6, 7, and 8 showed 'satisfactory' correlations, Climate Divisions 5 and 9 'very good' correlations ($0.85 \leq R \leq 1.0$), and Climate Division 10 showed a 'good' correlation. Climate Division 11 in the UCRB showed an 'unsatisfactory'

correlation ($0.00 \leq R < 0.70$), while Climate Division 12 showed a ‘satisfactory’ correlation. Climate divisions in the RGB provided correlations that range from ‘satisfactory’ to ‘very good’. Climate Division 14 showed a ‘satisfactory’ correlation in the RGB. Climate Divisions 15-20 showed a ‘good’ correlation. Climate Division 13 in the northernmost portion of the basin had a ‘very good’ correlation.

NSE values for the climate division models in the western United States are shown in Figure 25c. NSE generally agreed with the RSR values (Figure 25a) and R values (Figure 25b). In the PNW, Climate Divisions 1 and 2 produced ‘good’ model results ($0.65 \leq NSE < 0.75$). ‘Very Good’ modeling results ($0.75 \leq NSE \leq 0.75$) were obtained for Climate Division 3 in the GB. The climate divisions in the LCRB ranged from ‘satisfactory’ ($0.5\ 0 \leq NSE < 0.65$) to ‘very good’ modeling results. Climate Divisions 4, 6, 8 and 10 showed ‘good’ modeling results. Climate Divisions 5 and 9 provided ‘very good’ modeling results, while Climate Division 7 produced ‘satisfactory’ modeling results. In the UCRB, Climate Division 11 produced ‘unsatisfactory’ modeling results ($0.00 \leq NSE < 0.50$), while Climate Division 12 produced ‘good’ modeling results. Climate divisions in the RGB showed modeling results that were ‘good’ and ‘very good. 'Climate Divisions 14, 15, 17, 18, 19, and 20 showed ‘good’ modeling results in the RGB. Climate Divisions 13 and 16 produced ‘very good’ modeling results.

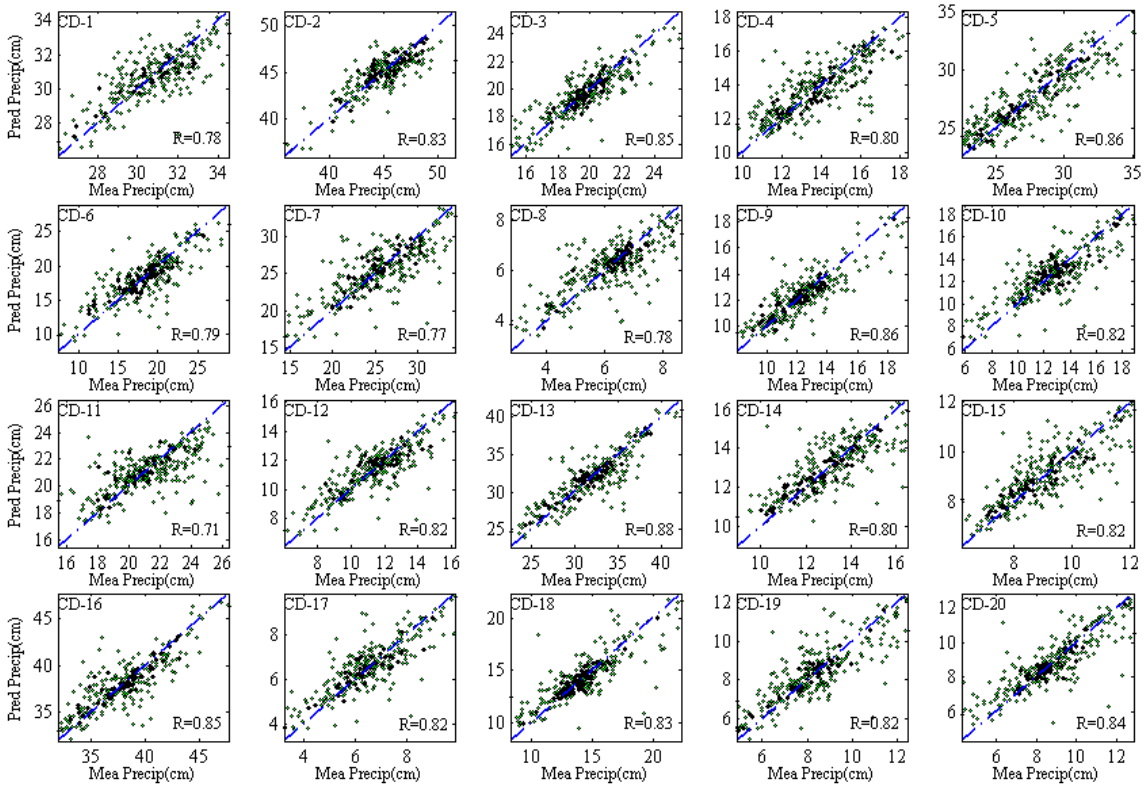

Figure 26: Scatter plots between measured and predicted precipitation using KStar for the 20 climate divisions. The diagonal line is the 45° bisector line.

Figure 26 displays the scatter plots for the climate divisions. The scatter plots showed that the majority of the points saturated the 45° bisector line, indicating that the KStar model was capable of identifying relationships between the oceanic-atmospheric oscillations and precipitation in each of the climate divisions. With the ability to identify these relationships, the KStar model was proficient at providing accurate precipitation predictions for the majority of the dataset.

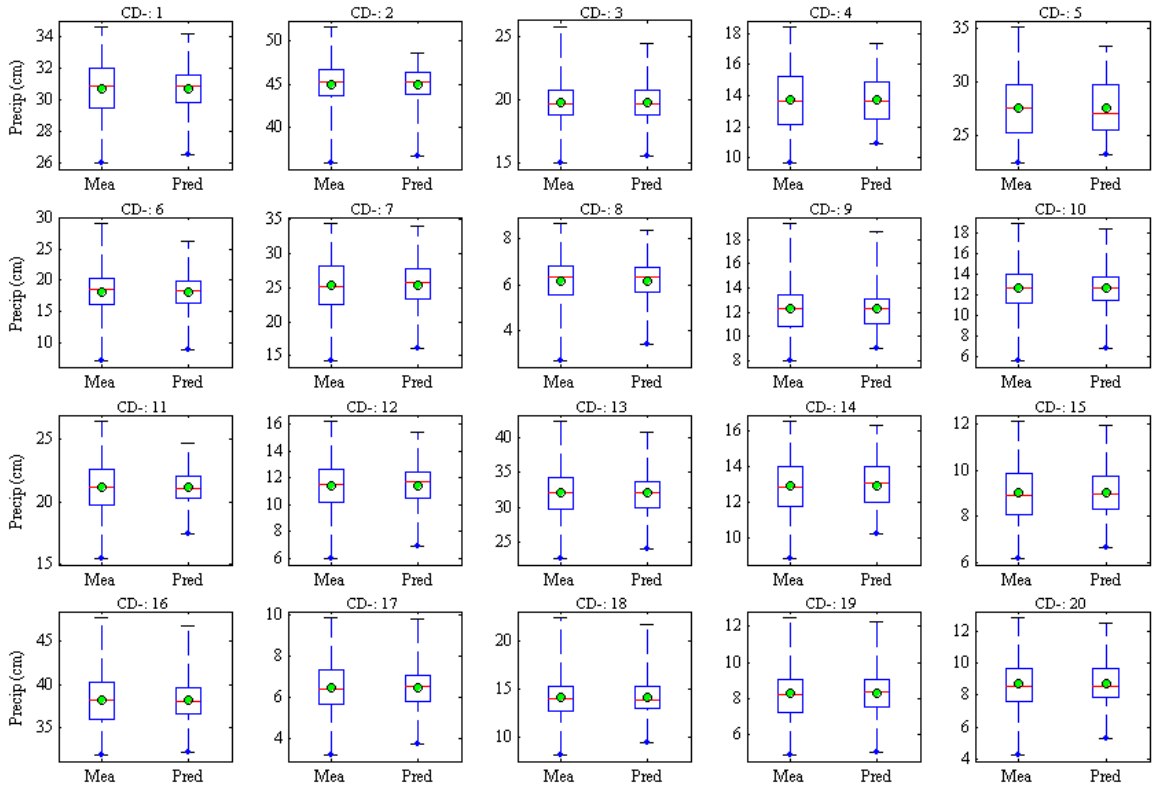

Figure 27: Box plots between measured and predicted precipitation using KStar for the 20 climate divisions. The box shows the interquartile range (25th-75th percentile). The whiskers extend from the 5th to 95th percentile values. The solid line inside the box shows the median value (50th percentile), and the solid dot represents the mean of the value.

Figure 27 shows the box plots for the climate divisions obtained using the KStar model. Two box plots were provided for each climate division to give a comparison between the measured values and the predicted values. The plots illustrate that the KStar model predictions were capable of matching the precipitation variability associated with each climate division. In addition, the box plots show that the model was capable of predicting the majority of extreme precipitation events, which were represented by the whiskers. However, it was unable to capture the very low and very high precipitation events near the 5th and 95th percentiles.

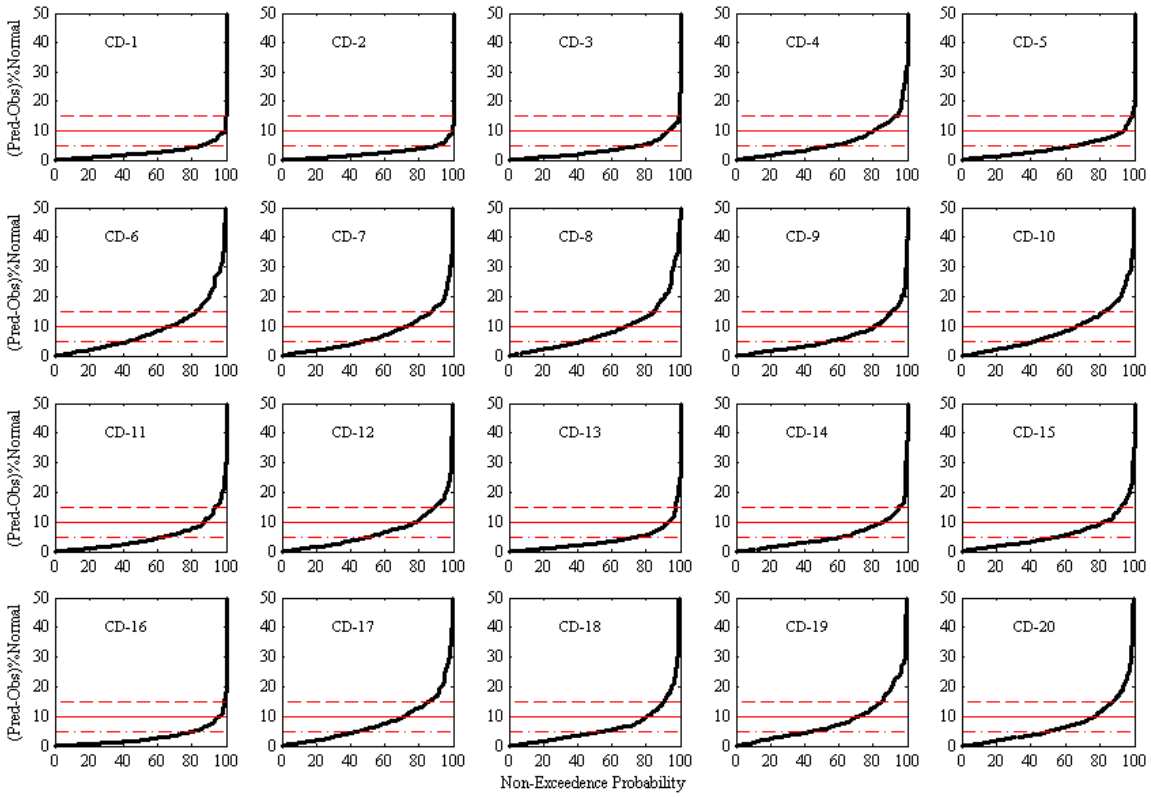

Figure 28: Non-exceedance probability plots for KStar for the 20 climate divisions. The horizontal dash-dot line shows an error value of 5%, the horizontal solid line shows an error value of 10%, and the horizontal dashed line shows an error value of 15%.

Non-exceedance probability plots in Figure 28show the accuracy of the model predictions in terms of their error from the measured values. The plots were interpreted as the percent of predictions that are less than or equal to a given error. About 65% - 100% of the predictions in each climate division had an error of 10% or less. This indicates that the majority of the precipitation predictions in all of the climate divisions had a low error associated with them. Climate Divisions 1 and 2 in the PNW showed that all of the predictions had an error of 10% or less. The number of predictions improved to 85% - 100% of the predictions having an error of 15% or less. Climate Divisions 1, 2, 3, 5, and 16 remained the only climate divisions that had 100% of the predictions with an error of

15% or less. The non-exceedance probability plots show that for the majority of the precipitation values, the KStar model was capable of producing very accurate predictions.

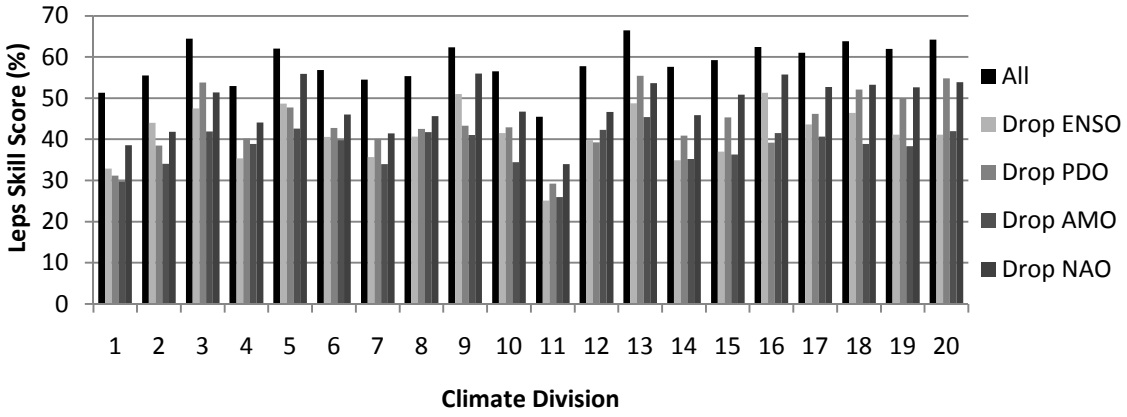

Figure 29: Bar graph displaying the LEPS SK obtained for each of the 20 climate divisions using KStar. Each climate division has 5 bars that represent (from left to right) the LEPS SK obtained using all oscillations, dropping ENSO as a predictor, dropping PDO as a predictor, dropping AMO as a predictor, and dropping NAO as a predictor.

Figure 29 displays a bar graph of the LEPS SK obtained for each climate division. The black bars represent the model performance when using all four oscillations as predictors in the model. The plot indicates that the LEPS SK value when using all oscillations ranged between 45 and 67. As described in the method section, a LEPS SK score greater than 10 indicated a 'good' model and was better than climatology. Therefore the precipitation estimates using the KStar model were much better than the climatology.

Table 15: Model Performance for the KStar Model.

CD	RMSE	MAE	RSR	R	NSE	LEPS
1	1.10	0.85	0.63	0.78	0.60	51.35
2	1.41	1.08	0.56	0.83	0.69	55.51
3	0.97	0.71	0.52	0.85	0.73	64.43
4	1.15	0.83	0.60	0.80	0.64	52.97
5	1.53	1.19	0.52	0.86	0.73	62.04
6	2.25	1.61	0.61	0.79	0.63	56.83
7	2.59	1.88	0.64	0.77	0.60	54.51
8	0.74	0.53	0.62	0.78	0.61	55.32
9	1.04	0.75	0.51	0.86	0.74	62.34
10	1.48	1.08	0.58	0.82	0.67	56.49
11	1.50	1.05	0.71	0.71	0.50	45.45
12	1.09	0.77	0.58	0.82	0.66	57.76
13	1.74	1.23	0.47	0.88	0.77	66.48
14	0.95	0.69	0.61	0.80	0.63	57.61
15	0.74	0.54	0.57	0.82	0.67	59.20
16	1.64	1.15	0.53	0.85	0.72	62.40
17	0.74	0.52	0.57	0.82	0.67	61.00
18	1.40	0.93	0.55	0.83	0.69	63.82
19	0.95	0.68	0.57	0.82	0.68	61.93
20	0.95	0.64	0.54	0.84	0.71	64.19

Table 15 provides a summary of the performance measures obtained for each climate division using the KStar model. The RMSE and MAE for the climate divisions were all low in comparison with the precipitation variability associated with each climate division (Figure 24). The RSR, R, NSE, and LEPS SK values obtained for each climate division also were displayed in Table 15. The values for RSR, R, and NSE corresponded to the values displayed in Figure 25 and were subject to the performance rating as shown in Table 13.

4.5.3 Coupled Oscillations

For this set of KStar models, similar tests were performed; however, instead of all four oceanic-atmospheric oscillation indices used together as predictors for precipitation, one oscillation was dropped so that only three out of the four oscillations were used as predictors in turn. Consequently, each climate division had four model runs using different oscillation index combinations. The results from these coupled oscillations showed a possible loss or gain of information from dropping one of the oscillation indices in comparison to the base case of using all four oscillation indices as predictors. This was useful for identifying the role each oscillation had in precipitation forecasts, especially for those oscillations that were not identified as significant at the 95% confidence level.

4.5.3.1 Drop ENSO

Figure 30ashows a spatial representation of the RSR values obtained under the KStar model when ENSO was dropped so that PDO, AMO, and NAO were used to forecast precipitation. In comparison with the model that used all four oscillations (Figure 25a), there was a noticeable increase in the RSR with only6 of the 20 climate divisions producing ‘satisfactory’ results and the remaining 14 climate divisions had ‘unsatisfactory’ results. The increase in RSR indicated a decrease in the model performance. Figure 30b and Figure 30c agree with the RSR in which the R and NSE were ‘satisfactory’ for the same 6 climate divisions and ‘unsatisfactory’ for the remaining climate divisions. The light gray bars in Figure 29represent the LEPS SK obtained when ENSO was dropped as a predictor from the model. The LEPS SK ranged from 25 to 51, which indicated a ‘good’ forecast and that the model was better than climatology.

However, the LEPS SK was lower for all climate divisions when compared to using all oscillations.

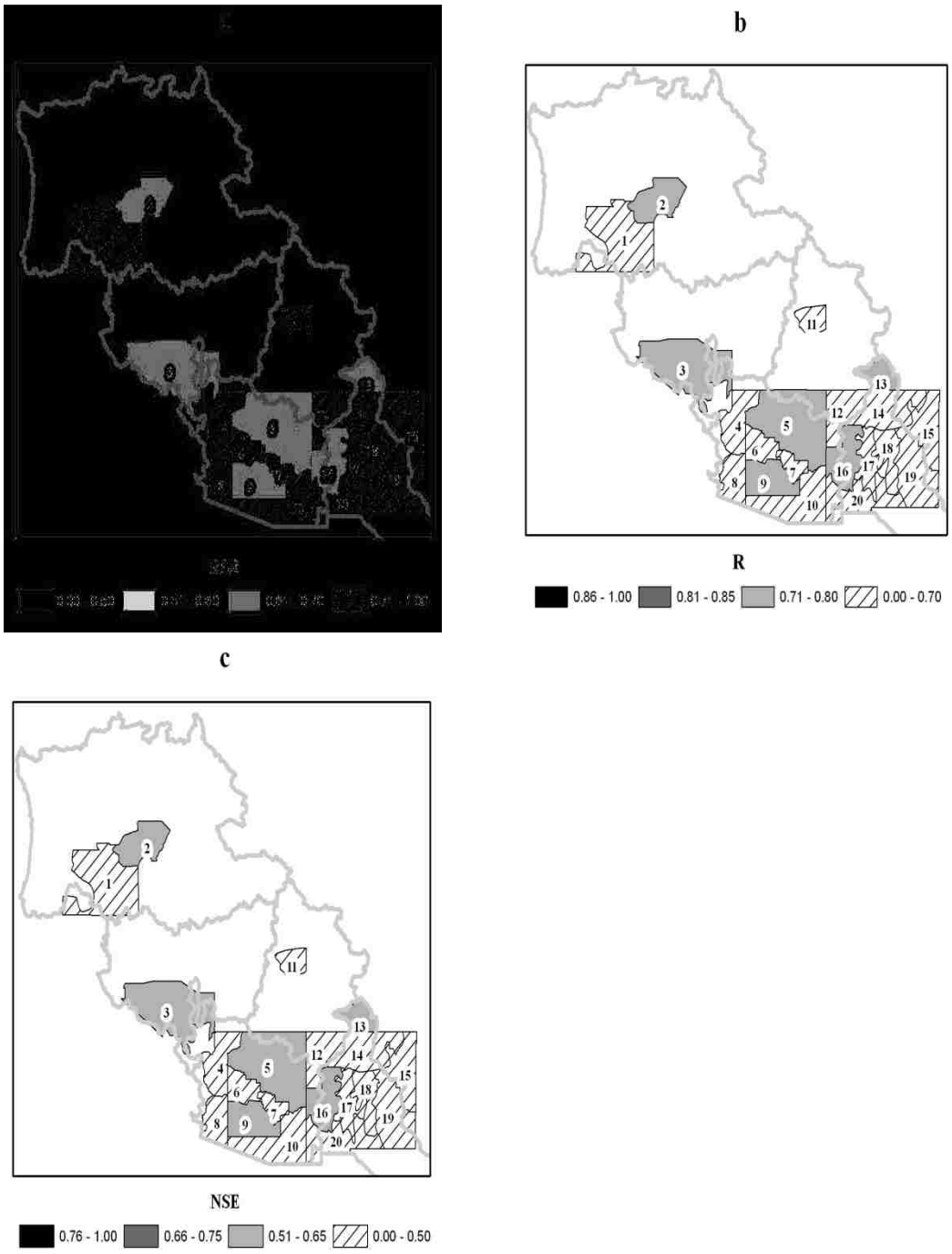

Figure 30: KStar model results obtained when dropping ENSO as a predictor. Performance measures shown are (a) RSR, (b) R, and (c) NSE.

4.5.3.2 Drop PDO

Figure 31a displays the RSR values obtained when PDO was dropped and ENSO, AMO, and NAO were used to forecast precipitation. In comparison with the results obtained when using all four oscillations (Figure 25a), there was an increase in the RSR i.e., model performance deteriorated. This was shown as only Climate Division 13 produced 'good' results, 9 of the 20 climate divisions produced 'satisfactory' results, and the remaining 11 climate divisions produced 'unsatisfactory' results. The climate divisions which performed well were located in the southern portion of the study area, indicating that the PDO had a more prominent influence over the northern portion of the study area. The R values shown in Figure 31b and the NSE values shown in Figure 31c agree with the RSR. Figure 29 shows the LEPS SK as gray bars when PDO was dropped as a predictor. The LEPS SK ranged from 29 to 55, which was lower for all climate divisions in comparison to using all oscillations. However, it still indicated that a 'good' forecasting model was achieved in comparison with climatology.

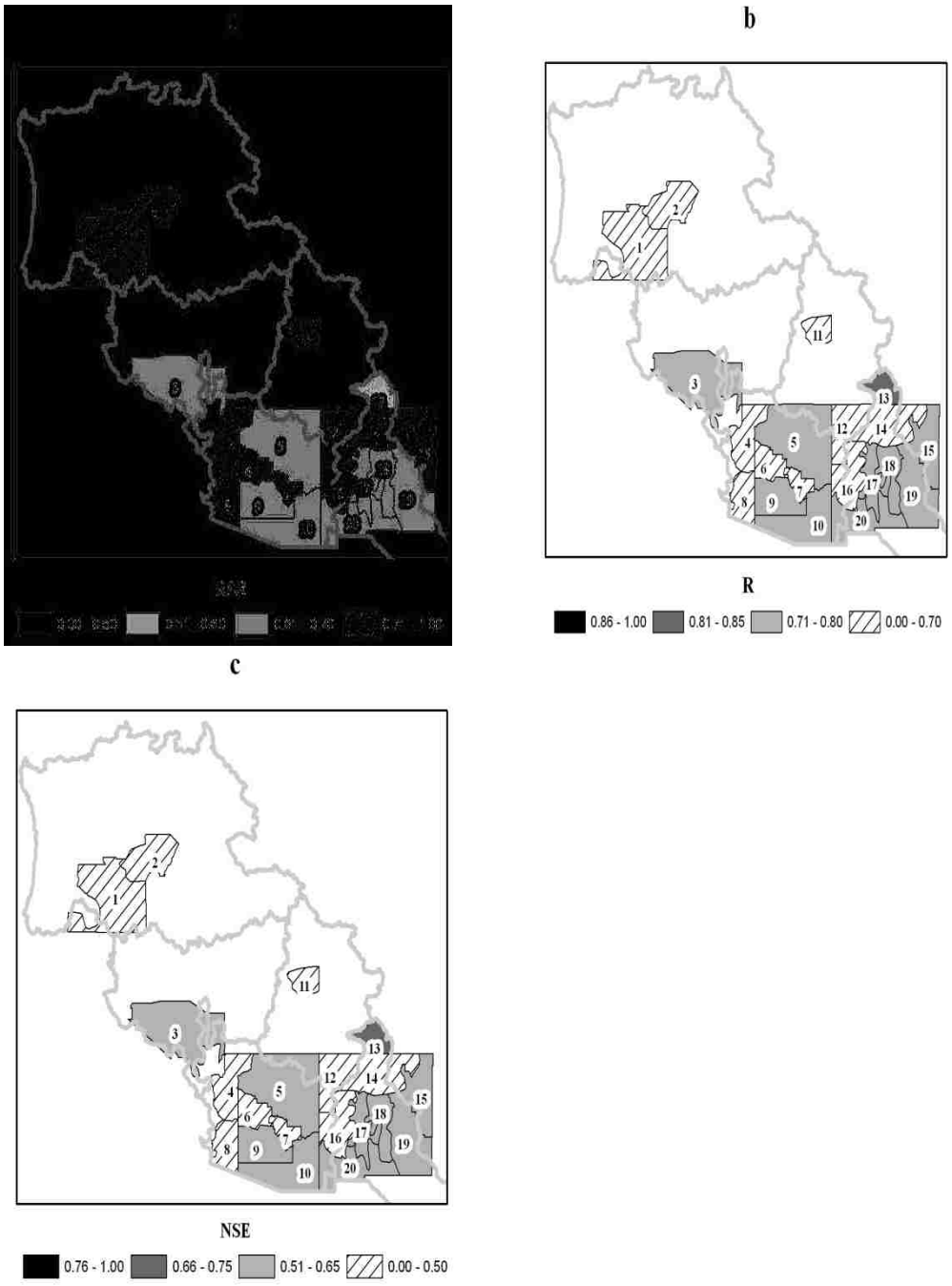

Figure 31: KStar model results obtained when dropping PDO as a predictor. Performance measures shown are (a) RSR, (b) R, and (c) NSE.

4.5.3.3 Drop AMO

Figure 32a displays a spatial representation of the RSR values obtained when AMO was dropped and only ENSO, PDO, and NAO were used to forecast precipitation. In comparison with the results obtained when using all four oscillations (Figure 25a), there was an increase in the RSR, signifying a decrease in the model performance. Only Climate Divisions 5, 13, and 16 produced 'satisfactory' results and the remaining climate divisions produced 'unsatisfactory' results. Figure 32b and Figure 32c showed that the R and NSE values tend to agree with the RSR. However, Climate Divisions 4, 5, 9, 13, and 16 were identified with producing 'satisfactory' results, with the remaining climate divisions producing 'unsatisfactory' results. The dark gray bars in Figure 29 represent the LEPS SK when AMO was dropped as a predictor. The LEPS SK ranged from 25 to 45, which indicated a 'good' model was achieved compared to climatology. Figure 29 also showed a decrease in LEPS SK for all climate divisions in comparison to using all oscillations.

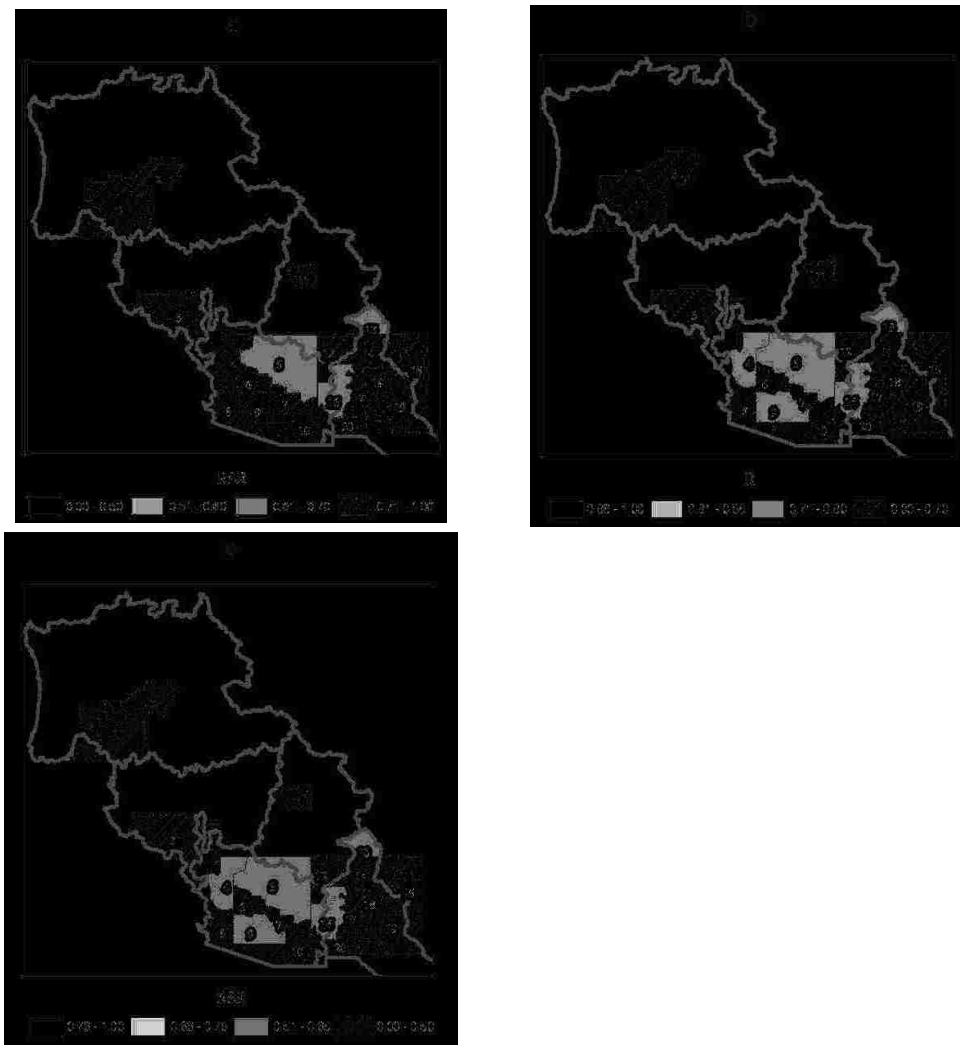

Figure 32: KStar model results obtained when dropping AMO as a predictor. Performance measures shown are (a) RSR, (b) R, and (c) NSE.

4.5.3.4 Drop NAO

Figure 33a shows a spatial representation of the RSR values obtained using the KStar model when NAO was dropped as a predictor and ENSO, PDO, and AMO were used to forecast precipitation. In comparison with the model that used all four oscillations (Figure 25a), there was an increase in the RSR, which indicated a decrease in the model performance. Climate Divisions 5, 9, 13, and 16 produced 'good' forecasts. Climate Divisions 1, 7, and 11 produced 'unsatisfactory' results. The remaining 13 climate divisions produced 'satisfactory' results. Figure 33b displays the R values, which showed

similar results as the RSR results. Climate Divisions 5, 9, 13, and 16 produced ‘good’ forecasts. Climate Divisions 7 and 11 produced ‘unsatisfactory’ results. The remaining 14 climate divisions produced ‘satisfactory’ results. NSE values for the precipitation forecast models were also shown in Figure 33c, which also showed similar results. Climate Divisions 5, 9, and 16 produced ‘good’ forecasts. Climate Divisions 1, 7, and 11 produced ‘unsatisfactory’ results. The remaining 14 climate divisions produced ‘satisfactory’ results. Figure showed the darker gray bars as the LEPS SK when NAO was dropped as a predictor. The LEPS SK ranged from 34 to 56, which indicates a ‘good’ forecast model was achieved compared to climatology. A decrease in the LEPS SK was observed for all climate divisions in comparison to using all oscillations. However, it was clear that the LEPS SK decreased the least in comparison with the LEPS SK obtained when dropping ENSO, PDO, or AMO.

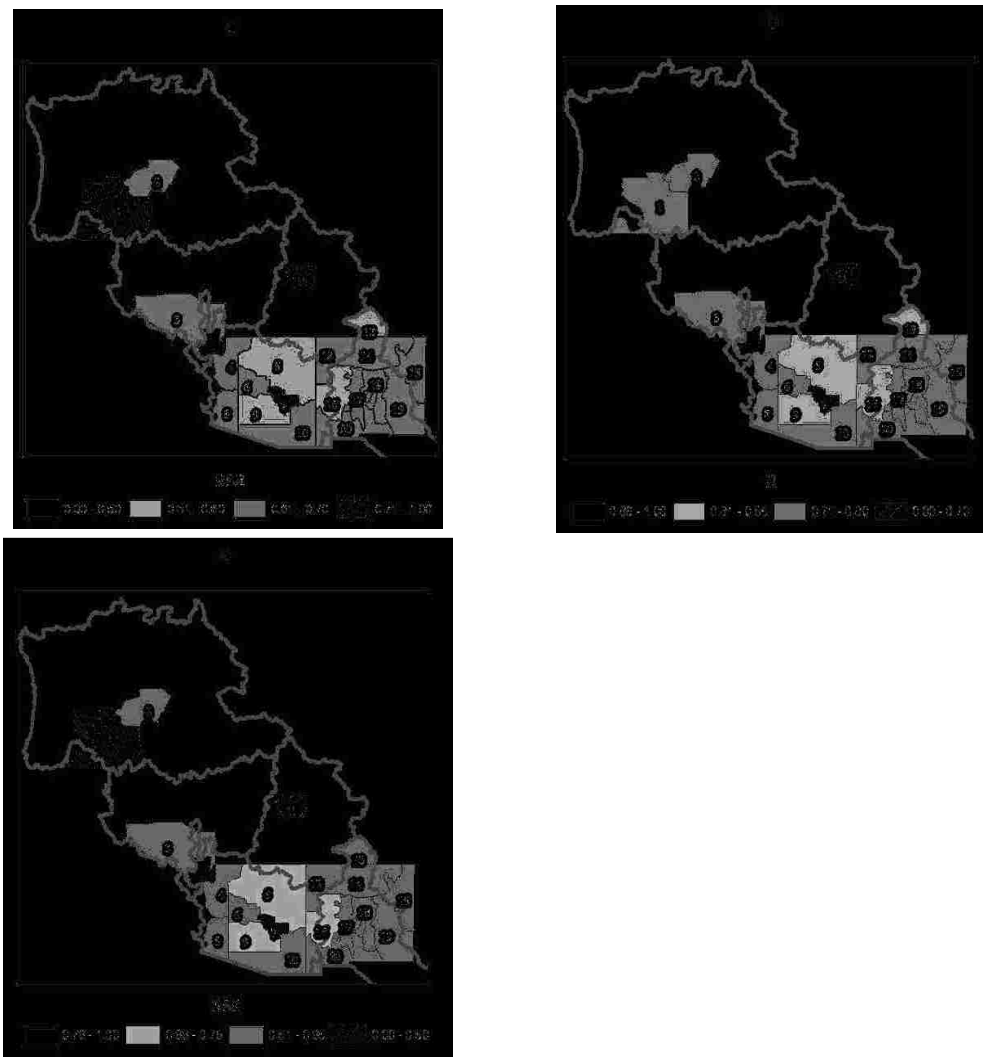

Figure 33: KStar model results obtained when dropping NAO as a predictor. Performance measures shown are (a) RSR, (b) R, and (c) NSE.

4.5.4 Comparison of Coupled Oscillations and All Oscillations

The results for the coupled oscillations analysis showed that dropping an oscillation index as a predictor caused a decline in the model performance in comparison to the performance when all four oscillations were used as predictors. ENSO, PDO, and AMO each had a significant influence over precipitation in the western United States as dropping any one of these resulted in a deterioration of the model performance. Additionally, the NAO did show an influence on precipitation in the western US,

although it was not as apparent as the other oscillation indices. The smallest deterioration in the model performance was obtained when NAO was dropped as a predictor. Therefore, all four oscillation indices together provided the best results for precipitation forecasting.

In addition to the tests performed for all oscillation indices and dropping one oscillation index, similar tests were performed dropping two and three oscillation indices. A total of 6 model runs for each climate division were performed to represent each possible combination of two oscillation indices when they were used as predictors. When single oscillations were used as predictors, four models were used, one for each oscillation index. 'Unsatisfactory' results for RSR ($0.70 < RSR \leq 0.71$), R ($0.00 \leq R < 0.70$), and NSE ($0.00 \leq NSE < 0.50$) were obtained for all 20 climate divisions when two oscillation indices and single oscillation indices were used as predictors. In addition, the LEPS SK continued to deteriorate in comparison with using all oscillations and dropping one oscillation. The results were not shown for these model runs. However, it is clear that the best model results were obtained when all oscillations were used together as predictors.

4.5.5 Comparison with Multiple Linear Regression Approach

The KStar model was compared with the standard MLR modeling approach. Figure 34 shows a spatial representation of the performance measures obtained for each climate division using the MLR model with all four oceanic-atmospheric oscillation indices used as predictors. The RSR (Figure 34a), R (Figure 34b), and NSE (Figure 34c) values for all of the climate divisions using the MLR model were considered 'unsatisfactory' when subjected to the performance ratings, as shown in Table 16. The LEPS SK shows that the MLR model was better than climatology for the majority of the climate divisions. However, the LEPS SK results obtained for KStar were superior to the MLR model. In addition, other combinations of oscillation predictors also produced 'unsatisfactory' results, and the results were not shown.

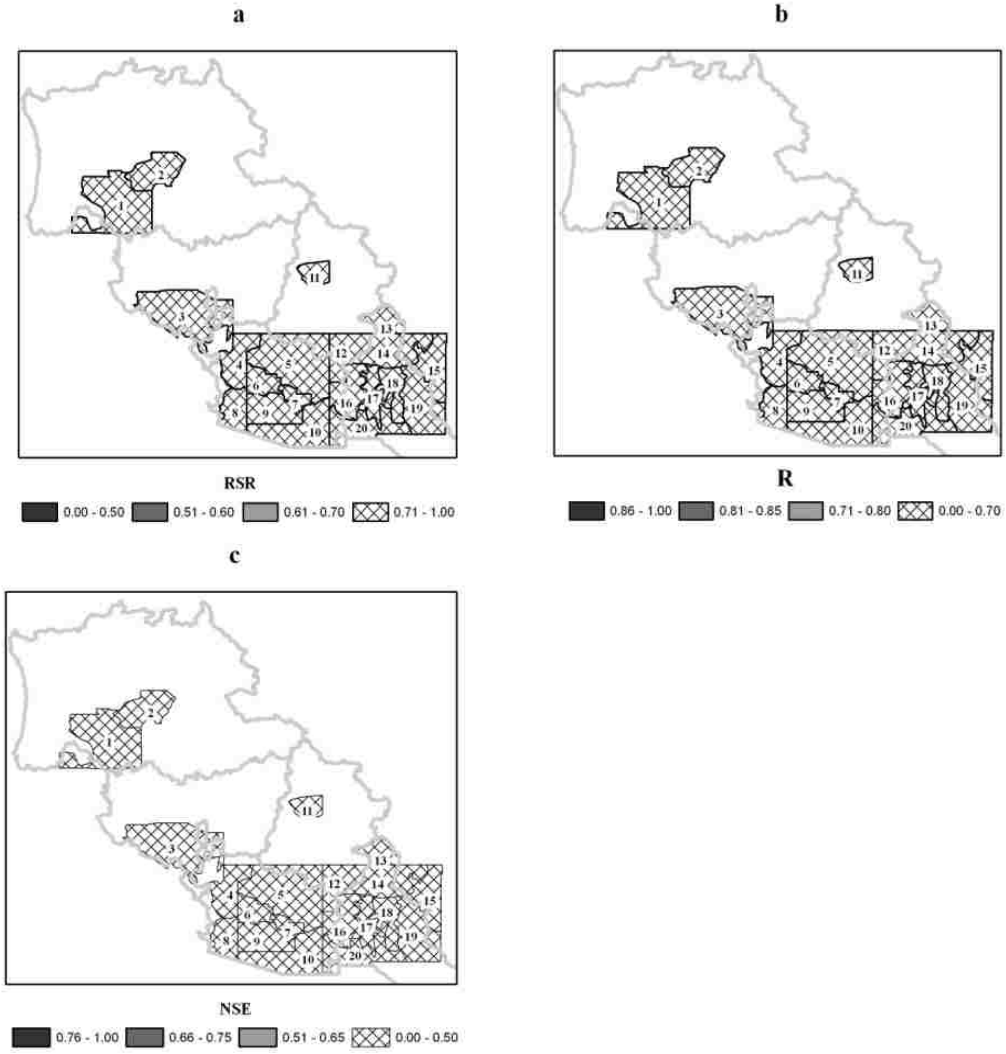

Figure 34: MLR model results obtained when using all four oscillations (ENSO, PDO, AMO, and NAO). Performance measures shown are the (a) RSR, (b) R, and (c) NSE.

Table 16: Model Performance for the MLR Model.

CD	RMSE	MAE	RSR	R	NSE	LEPS
1	1.76	1.45	1.00	0.12	0.00	1.31
2	2.34	1.87	0.92	0.39	0.15	13.70
3	1.61	1.28	0.87	0.49	0.24	23.78
4	1.69	1.37	0.88	0.47	0.22	18.51
5	2.50	2.05	0.84	0.54	0.29	24.00
6	3.18	2.56	0.86	0.51	0.26	23.80
7	3.63	2.93	0.89	0.45	0.21	19.45
8	1.00	0.81	0.84	0.53	0.28	25.58
9	1.64	1.26	0.80	0.60	0.36	30.65
10	2.13	1.73	0.83	0.55	0.31	26.34
11	2.05	1.65	0.97	0.25	0.06	7.78
12	1.62	1.26	0.86	0.52	0.27	24.86
13	3.13	2.58	0.85	0.53	0.28	22.99
14	1.36	1.06	0.87	0.49	0.24	22.70
15	1.15	0.95	0.89	0.45	0.20	15.36
16	2.42	2.00	0.78	0.63	0.39	31.57
17	1.04	0.83	0.81	0.58	0.34	31.09
18	2.18	1.67	0.86	0.50	0.25	23.60
19	1.44	1.14	0.86	0.51	0.26	26.85
20	1.48	1.19	0.84	0.54	0.29	26.05

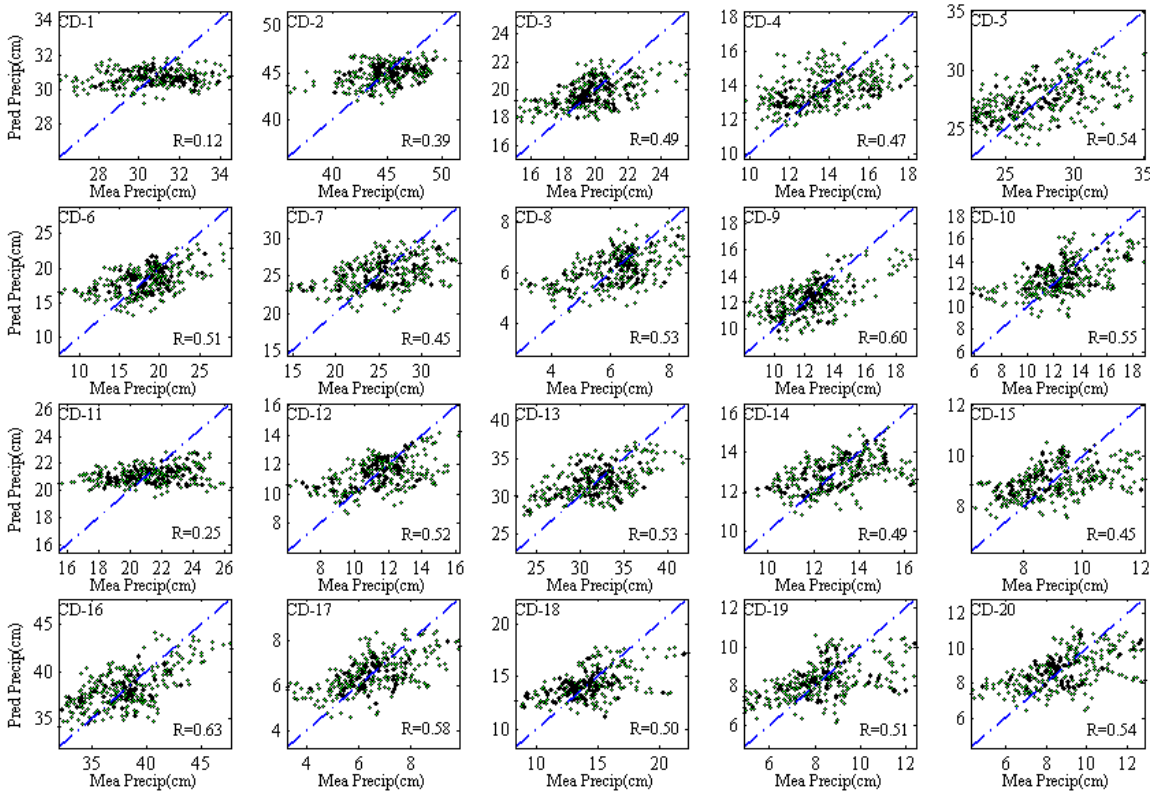

Figure 35: Scatter plots between measured and predicted precipitation using MLR for the 20 climate divisions. The diagonal line is the 45° bisector line.

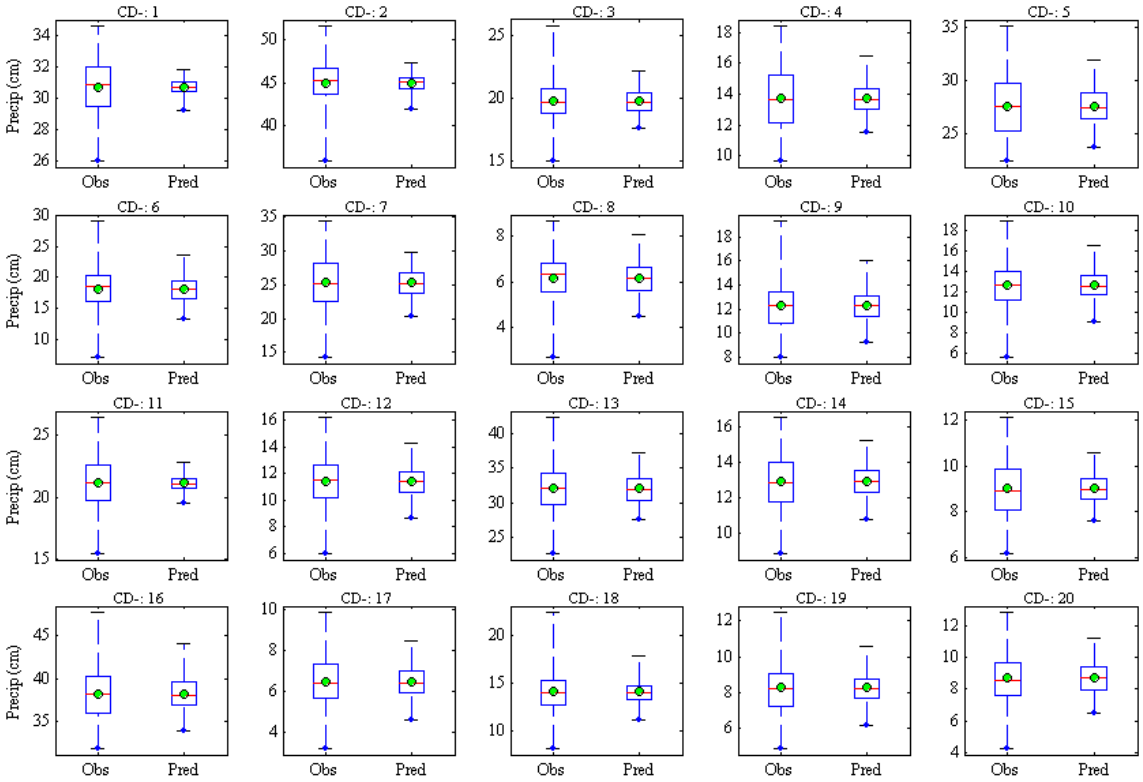

Figure 36: Box plots between measured and predicted precipitation using MLR for the 20 climate divisions. The box shows the interquartile range (25th-75th percentile). The whiskers extend from the 5th to 95th percentile values. The solid line inside the box shows the median value (50th percentile), and the solid dot represents the mean of the value.

4.6 Discussion and Conclusion

This study presented a data-driven forecasting model, KStar, which incorporated paleoclimate reconstructions of annual oceanic-atmospheric oscillation indices to forecast water year precipitation with a 1-year lead time for select climate divisions in the western United States. This study addressed the limitation of using a relatively short period of record, common to current forecasting models (Ahmad et al., 2009; Ahmad et al., 2010). Paleoclimate reconstructions were incorporated to extend the period of record to over 300 years to aid the data-driven model in creating a forecast. This novel approach of incorporating paleoclimate reconstructions allowed for the examination of the influence of long range oceanic-atmospheric oscillations over precipitation, such as PDO and

AMO, which were limited to only 1 or 2 cycles in the instrumental record. The KStar model was able to produce ‘good’ and ‘very good’ forecasts for the majority of the climate divisions based on performance measures by Moriasi et al. (2007). The annual precipitation forecast provided by the KStar model can be disaggregated to a finer temporal resolution depending on the needs of the end user (Kalra & Ahmad, 2011).

The findings showed that the best precipitation forecasts were produced using all four oceanic-atmospheric oscillations indices -- ENSO, PDO, AMO, and NAO -- as predictors in the KStar model. The use of multiple indices as predictors allowed for a data-driven model to grow in complexity. In the case of instance-based learners, such as KStar, the distance function measured the complexity of transforming one instance into another (Cleary & Trigg, 1995). The use of more predictors provided more decision criteria, which enhanced the accuracy of the model. The removal of one or more of the predictor indices resulted in a loss of information, yielding a less accurate model that showed a significant deterioration of the precipitation forecast as shown in Figure 30, Figure 31, Figure 32, and Figure 33. It was shown that dropping ENSO, PDO, or AMO as a predictor resulted in several ‘unsatisfactory’ results. However, NAO was shown to have a smaller impact because there were many ‘satisfactory’ and ‘good’ modeling results when it was dropped as a predictor. This was confirmed in Figure 29, where the LEPS SK showed the smallest deterioration in the results, when NAO was dropped as a predictor, compared to the base case.

To complement the analysis of the coupled effects of the oscillation indices over precipitation, the correlations between the indices and precipitation that were significant at the 95% confidence level revealed the importance of each index in the western United

States. This also identified positive or negative relationships between the indices and precipitation, which were not discernable from the modeling forecasts. When ENSO was compared with climate division precipitation, the correlations were significant for the majority of the southwestern United States (Table 14). This agrees with the spatial representation of the climate divisions when ENSO is dropped as a predictor (Figure 30a), where the majority of the climate divisions experienced a decrease in the model performance. ENSO was shown to have a positive correlation with precipitation in the southwestern United States (Table 14). This finding agrees with previous studies, which have found a positive correlation between ENSO and southwestern United States precipitation over the observed instrumental record (e.g. Dettinger et al., 1998; McCabe & Dettinger, 1999; Redmond & Koch, 1991). However, the previous studies also indicated that precipitation in the PNW is negatively correlated with ENSO (e.g. Dettinger et al., 1998; McCabe & Dettinger, 1999; Redmond & Koch, 1991). The current study could not confirm this relationship with precipitation in the PNW due to the finding that Climate Divisions 1 and 2 were not significant at the 95% confidence level for ENSO.

The correlations in Table 14 indicate that the PDO plays a significant role in the precipitation variability throughout the western United States. This is especially true in the LCRB and RGB where the correlations between PDO and precipitation are relatively strong. Table 14 indicates that Climate Divisions 5, 6, 9, 12, and 16 showed correlations greater than 0.4 for the PDO. In comparison, ENSO correlations did not exceed -0.4. These climate divisions showed a reduction in the model performance when PDO was dropped as a predictor (Figure 31b). Using the instrumental record, previous studies

indicated that the PDO had a positive correlation with precipitation in the southwestern United States and a negative correlation in the northwestern United States (e.g. Brown and Comrie, 2004; Mantua et al., 1997; McCabe & Dettinger, 1999). The findings in Table 14 agree with the previous studies where all of the climate divisions in the GB, UCRB, LCRB, and RGB showed positive correlations with PDO. In the PNW, Climate Division 1 indicated no correlation to PDO. However, precipitation in Climate Division 2 showed a significant negative correlation with the PDO.

According to Enfield et al. (2001), the southwestern United States, including the GB, UCRB, LCRB, and RGB, showed a negative correlation between AMO and precipitation in the observed record, while the PNW showed a positive correlation. The findings in the current study agree with those of Enfield et al. (2001), where a significant number of climate divisions in the southwestern United States showed negative correlations and Climate Division 1 in the PNW showed a positive correlation (Table 14).

In contrast, the NAO signal in the instrumental record was generally not identified in precipitation and streamflow in specific regions in the western United States (Hunter et al., 2006; Tootle et al., 2005; Visbeck et al., 2001). However, Hunter et al. (2006) did identify the NAO signal through the snow water equivalent (SWE) in 40 of 121 Snow Pack Telemetry (SNOTEL) stations that were tested in the western United States; however, they could not identify any distinct spatial regions of NAO influence. The results of the current study agree with the findings of Hunter et al. (2006). Figure 33 indicated that by dropping the NAO as a predictor, a decrease in model performance occurred in some of the climate divisions. The current study found that the NAO was associated with precipitation in 8 of the 20 tested climate divisions in the western United

States (Table 14), but no distinct spatial area was identified, which was also consistent with Hunter et al. (2006). In the current study, a negative correlation between precipitation in Climate Divisions 1 and 2 and the NAO was identified in the PNW. Hunter et al. (2006) also identified a negative correlation between the NAO and SWE in Oregon, but the signal was only evident when coupled with the La Niña phase of ENSO.

A study by Gray et al. (2003) found that moisture variability was significant for multi-decadal periods and may be attributed to the oscillations of PDO and AMO. The study theorized that long-range forecasts greater than a year could be achievable with knowledge about the current phase of the PDO and AMO. The current study confirmed this theory, as the vast majority of the tested climate divisions in the western United States showed a significant reduction in the model performance when the PDO and AMO oscillations are dropped as predictors (Figure 31 and Figure 32). This indicates that climatic variability in the western United States involves complex interactions between the Pacific and Atlantic Oceans, as theorized by Gray et al. (2003). However, this did not agree with the findings of Kalra and Ahmad (2009). They developed an SVM model using the instrumental record to provide a streamflow forecast within the Colorado River Basin using the oceanic-atmospheric oscillations of ENSO, PDO, AMO, and NAO. The study showed that the oscillations of ENSO and NAO serve as the best predictors for streamflow forecasting, whereas the multi-decadal oscillations of PDO and AMO did not provide as strong a signal., This may be attributed to the shorter instrumental record used in the study, where only ENSO and NAO were fully represented.

The ability to capture the relationships between the oceanic-atmospheric oscillations and precipitation was observed in the KStar modeling results, whereas the

MLR modeling results had difficulty in identify the relationships. In comparison with MLR, the KStar model produced better precipitation forecasts. Because MLR uses linear regression to make the precipitation predictions, the model lacked the ability to capture the non-linear relationships between the oceanic-atmospheric oscillations and precipitation. KStar compared the similarity between a tested instance and pre-classified examples to remove mean-weighted biases that may occur in linear regression models (Cleary & Trigg, 1995). The forecasts obtained with KStar are based on the assumption that similar sets of input oscillation indices will produce a similar precipitation forecast. Although the relationships identified through KStar are hidden, a closer examination of their predictors can reveal which combination of oscillation phases have a significant influence over precipitation. This approach can be useful for identifying oscillation combinations that frequently result in extreme events. As a result, this study showed that KStar was better at capturing the natural variability associated with precipitation, which was difficult for the MLR model to identify.

The KStar model served as an improvement over many existing models because of the ability to provide quantitative forecasts. This was as opposed to qualitative forecasts, such as those provided by NWS. With qualitative forecasts water managers had difficulty in making operational decisions. A quantitative forecast provides more meaningful information to water managers for planning and allocation decisions. Since many climate divisions in the western United States were based upon drainage basins (Guttman & Quayle, 1996), the qualitative precipitation forecast could be used to determine the future water supply in the region.

In order for the models to determine the relationships between various oceanic-atmospheric oscillations and precipitation, the data must be filtered. Hydrologic studies that examined long-range fluctuations have used filters to clarify the data for use in an analysis (e.g. Currie, 1996; Garbrecht & Rossel, 2002; De Jongh et al., 2006; Pekarova et al., 2003; Probst & Tardy, 1987; Riehl & Meitin, 1979; Riehl et al., 1979). A moving average filter was used in this study to smooth out the highly fluctuating data, which have to be filtered in order to attenuate the short-range fluctuations and to extract and clarify the long-range climatic variations (Probst & Tardy, 1987).The moving average filter aided in identifying the phases of the oscillations and removed the noise of the index signal within each phase. This ensured that the phases of the oscillations were correctly identified, so that they may be correlated with a precipitation response. In addition, data-driven models required that anomalies, noise, and large jumps within a dataset to be smoothed out so that the model did not make decisions to its detriment (Cleary & Trigg, 1995). This was applicable to the input oscillation indices, where the dataset must be smoothed out so that the model did not make errors in the learning and forecasting phases. The limitation of using a moving average filter was that the natural variability associated with the precipitation became reduced; however, the forecast can be made 1 year into the future.

Other limitations result from using reconstructions for forecasting purposes. Reconstructions are estimates of past climates, which always contain a certain degree of uncertainty. Currently reconstructions are only capable of explaining 50% - 80% of the variance observed in the instrumental record (Woodhouse et al., 2011). This requires the reconstructions to be rescaled so that they are compatible with the instrumental record,

which avoids heteroskedasticity in the dataset. Another source of uncertainty is that there are multiple reconstructions available for the oscillation indices. These reconstructions are often not in very good agreement, which may produce different results for each reconstruction. For purposes of this study, a single reconstruction for each oscillation index was selected based on length and statistical properties. Although there are limitations when using paleoclimate reconstructions, the limitations of using a short period of observed records were addressed. Reconstructions aided in understanding the complex relationships between oceanic-atmospheric oscillations and precipitation variability by extending the period of record by several hundred years beyond the instrumental record. The reconstructions allowed for an examination of long-term processes, which were not fully represented in the short instrumental record. Longer oscillation cycles exhibited by the PDO and AMO show a stronger influence than shorter cycles that were characteristic of ENSO and NAO had a relatively smaller influence over precipitation. For data driven models, more training data helped in improving the model forecasts (Ahmad & Simonovic, 2005; Melesse et al., 2011). Reconstructions aided data-driven models because they increased the amount of examples that the model learns from (Witten et al., 2011).

Overall, the results from this study contributed to a better understanding of the impacts of long-term statistical processes on precipitation in the western United States by using paleoclimate reconstructions to extend the period of record back to 1658. This was as opposed to previous studies, which evaluated the individual and coupled impacts of oceanic-atmospheric oscillations in the observed instrumental record of 50 – 100 years (e.g., Brown & Comrie, 2004; Dettinger et al., 1998; Enfield, 2001; Hunter et al., 2006;

Kalra & Ahmad, 2009; Mantua et al., 1997; McCabe & Dettinger, 1999; Redmond & Koch, 1991). In addition, the proposed model generated quantitative precipitation forecasts with a 1-year lead time in the western United States by using four oscillation indices: ENSO, PDO, AMO, and NAO. This improved upon the current 12.5 month lead time forecasts employed by NWS, which were limited to qualitative precipitation forecasts.

The data-driven model, KStar, provided 'good' forecasts for annual precipitation by using all four oceanic-atmospheric oscillations as predictors. The predictions were better compared to the climatology. When oscillation indices become available, a forecast for the water year is provided as early as January 1st of the current year. This provides important information to water managers about the water availability in the upcoming year. The proposed approach also provides an alternative to complex models that simulate physical processes and is expected to be useful for long-term water resources management.

Chapter 5: Long-Range Streamflow Forecasts using Support Vector Machines in the Western United States

5.1 Introduction

The management of water resources in the western United States is crucial to ensure that appropriations are made for local water supplies, electrical power, and ecological habitat as the water sources become stressed (Cayan et al., 2003; Hamlet & Lettenmaier, 1999). The western United States has experienced a large population growth since 2000, with an estimated average growth of over 14% from 2000 to 2010 (United States Census, 2010). In addition, streamflow in the western United States tends to be highly variable, but persistent high flows and low flows that are synchronous across major basins are not uncommon (Jain et al., 2005; Pagano & Garen, 2005). In particular, periods of sustained droughts can pose a serious problem for water resources in the western United States. Hydrologic conditions within one region can affect conditions in other regions, because water is often traded or transferred across state and watershed boundaries (Cayan et al., 2003). Furthermore, yearly streamflow volumes that serve as the primary source for renewable water resources are expected to decline in the western United States and gradually worsen over the course of the century (Reclamation, 2011). The result of an increasing population, highly variable climate, and declining water supplies is a growing state of water stress in the western United States (Shrestha et al., 2011; Venkatesan et al., 2011 a, b). This adds to the complexity of planning and managing water resources in an already challenging environment of changing demographics and competing interests (Pagano & Garen 2005; Qaiser et al., 2011). In order to properly appropriate water resources in a complex environment, water managers

will need long-range forecasts to aid in planning and management of water resources (Ahmad & Simonovic 2001; Forsee & Ahmad 2011; Vedwan et al., 2008).

The ability to produce a model that yields accurate forecasts has become a central problem in water resources management (Yu & Long, 2007). Better forecasts can not only help in improving the water management but also reduce damages (Ahmad & Simonovic, 2005, 2006; Mosquera-Machado & Ahmad, 2007; Simonovic & Ahmad, 2005). A viable option for improving the accuracy of long range forecasts is to incorporate climate variability in the form of oceanic-atmospheric oscillations. These oscillations are identified with warm and cool phases that have an influence over weather patterns and the climate. Extensive research indicates that predictive relationships can be developed between the phases of an oscillation and their corresponding impacts on streamflow in the western United States (e.g. Beebee & Manga, 2004; Dettinger et al., 1998; Enfield et al., 2001; Hamlet & Lettenmaier, 1999; Hidalgo & Dracup, 2003; Tootle et al., 2005). Oceanic-atmospheric oscillations are identifiable through indices of sea-surface temperatures (SSTs) and sea-level pressures (SLPs) that occur on interannual, decadal and interdecadal timescales (Tootle et al., 2005).

El Niño – Southern Oscillation (ENSO), Pacific Decadal Oscillation (PDO), Atlantic Multi-decadal Oscillation (AMO), and North Atlantic Oscillation (NAO) are the most studied indices that have shown the most prominent influence over streamflow variability in the western United States (e.g. Beebee & Manga, 2004; Cayan et al., 1999; Dettinger et al., 1998; Enfield et al., 2001; Gutzler et al., 2002; Hamlet & Lettenmaier, 1999; Hidalgo & Dracup, 2003; Hunter et al., 2006; Kalra & Ahmad, 2009; Mantua, 1999; McCabe et al., 2007; Redmond & Koch, 1991; Tootle et al., 2005). These

oscillations exhibit two desired qualities for use in streamflow forecasting. The first is that these oscillations have been shown to influence streamflow variability, which can be utilized to improve the accuracy of a forecast (Tootle et al., 2005). The second is that the observed fluctuations in streamflow are often lagged by several months to over a year from when an index identifies a particular phase of an oceanic-atmospheric oscillation (Gray et al., 2003; Kalra & Ahmad, 2009). However, it is difficult to incorporate the complex hydroclimatic relationships into a physically-based forecasting model (Kalra & Ahmad, 2009).

Data-driven modeling provides an alternative to physically-based modeling (Lin et al., 2009). A variety of data-driven models are available including artificial neural networks (ANNs), support vector machines (SVMs), instance-based learners and decision trees. These models work by extracting relationships between inputs and outputs from a training dataset comprised of previously known examples (Witten et al., 2011). However, a frequent problem is that the training data is often noisy and there is no guarantee that a hidden relationship can be correctly captured into a model (Cristianini & Shawe-Taylor, 2000). This study used an SVM model to extract the underlying relationships. SVM is a statistical learning methodology that uses a hypothesis space of linear functions in a higher dimensional feature space (Vapnik, 1995, 1998; Scholkopf & Smola, 2002; Cristianini & Shawe-Taylor, 2000). In order to capture the complex nonlinear relationships associated with hydroclimatology, a kernel-based approach is used. This allowed for nonlinear functions in the hypothesis space that can be transformed into solvable linear functions in the feature space (Scholkopf & Smola, 2002). SVMs are not new in hydrologic modeling and have been successfully used in groundwater monitoring

(Asefa et al., 2004), soil moisture prediction (Gill et al., 2006; Ahmad et al., 2010), precipitation downscaling (Tripathi et al., 2006); rainfall/runoff modeling (Dibike et al., 2001), and streamflow forecasting (Asefa et al., 2006; Kalra & Ahmad, 2009; Yu & Liong, 2007). A robust SVM model would serve to complement existing physically-based models.

The observed instrumental record typically extends from 50-100 years and has been useful for previous studies in examining historical trends and relationships and managing the reservoir operations on daily, weekly, or monthly basis (Franz et al., 2003; McEnery et al., 2005). However, an extended record is available beyond 300 years through paleoclimate reconstructions. Reconstructions provide an annual time series that are well correlated with the instrumental record (Gray et al., 2004; Hidalgo, 2004; Timilsena et al., 2009; Wilson et al., 2007; Woodhouse & Lukas, 2006). Although uncertainty is present when using reconstructions, they are able to capture the climate variability, which is essential for this study (Timilsena et al., 2009). Based on the documented literature and authors' knowledge, there have been no previous studies that have used an extended period of record provided by paleoclimate reconstructions in an SVM model for streamflow forecasting. The availability of the reconstructed record may help in providing useful information and may assist in improving the lead time. With this motivation, the current study used the SVM model to provide a 1-year and 5-year lead time forecast for 21 gage stations in the western United States using the four oceanic-atmospheric oscillations as predictors (i.e. ENSO, PDO, AMO, and NAO). The model was trained and tested using a leave-one-out cross-validation technique as opposed to splitting the dataset into training and testing datasets. The forecasts were evaluated

through the mean absolute error (MAE), root mean squared error (RMSE), RMSE-observations standard deviation ratio (RSR), Pearson's correlation coefficient (R), Nash-Sutcliffe coefficient of efficiency (NSE), and linear error in probability space skill score (LEPS SK). This modeling technique is intended to aid in the planning and management of water resources in the western United States.

The layout of the paper is as follows. Section 5.2 describes the study area and the data used in this study. A description of the methods, which includes the SVM model, is provided in Section 5.3. The model performance measures are provided in Section 5.4. The results of the forecasts are presented in Section 5.5. Finally, Section 5.6provides a discussion and concludes the study.

5.2 Study Area and Data

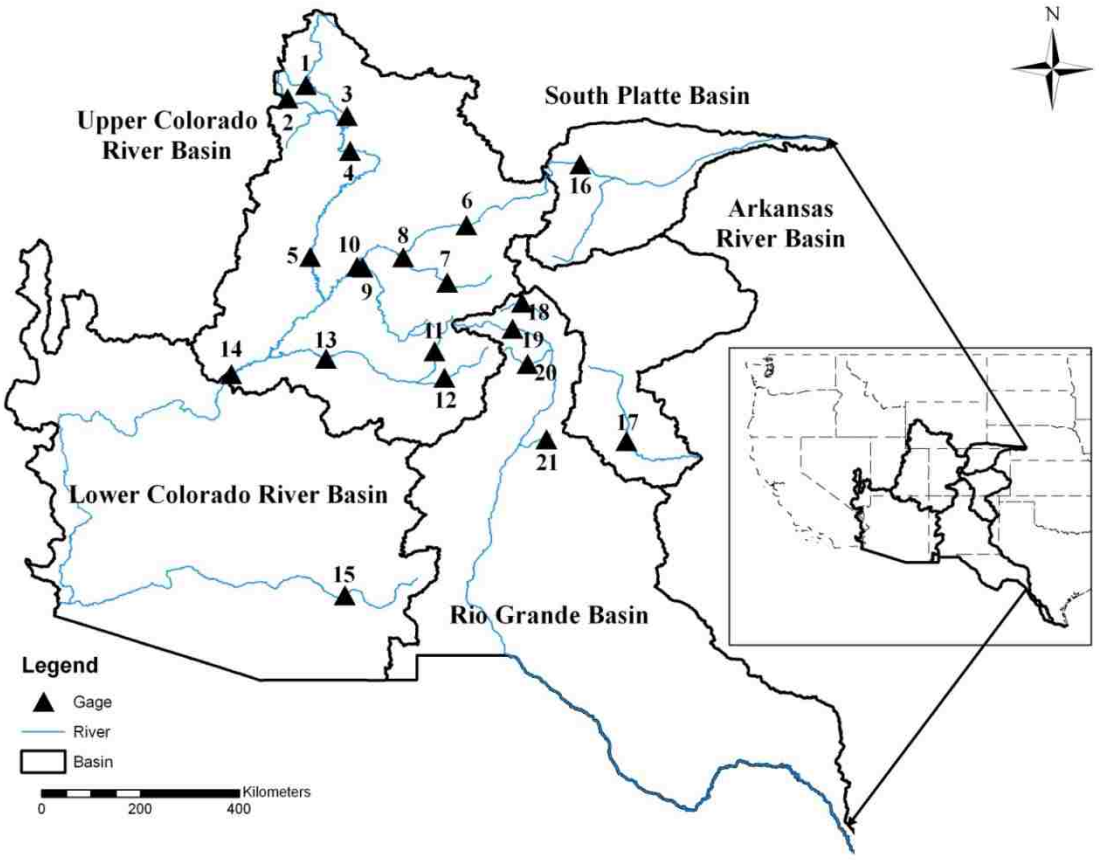

Figure 37: Study area depicting the location of the 21 streamflow gages used in the current study.

Annual observed and reconstructed streamflow data were obtained for 21 selected gages in the western United States. These gages were identified with a period of record from 1658 to 2007 (350 years). Reconstructions were used from 1658 to 1952 and the instrumental record was used from 1953 to 2007.

Table 17 identifies the 21 streamflow gages used in the study and Figure 37 displays their corresponding locations. The observed and reconstructed datasets used for each streamflow gage are provided in Table 18. 14 gages are located in the Upper Colorado River Basin (UCRB). The remaining gages are located in the Lower Colorado River Basin (LCRB), South Platte Basin (SPB), Arkansas River Basin (ARK), and Rio Grande Basin (RGB). Observed flows are provided by the United States Geological Survey (USGS) and the United States Bureau of Reclamation (USBR). Reconstructions are provided by TreeFlow (http://treeflow.info).

In addition, four oceanic-atmospheric oscillations indices were used in this study. Table 18 also describes the datasets used for each index. Reconstructions were limited to ENSO, PDO, AMO, and NAO. In addition, these reconstructions often have multiple reconstructions, which often do not agree with each other. This study selected reconstructions based upon the length of the time series and their statistical properties in relation to the available instrumental record.

The most well-known index is ENSO, which originates over the tropical Pacific Ocean and exhibits a frequency of 4-6 years (Cayan et al., 1999; Hamlet & Lettenmaier, 1999). El Niño is the warm phase of ENSO, which typically brings above-normal streamflow in the southwestern United States and below-normal streamflow in the Pacific Northwest. La Niña is the cool phase of ENSO which brings about below-normal streamflow in the southwestern United States and above-normal streamflow in the Pacific Northwest (Mann et al., 2000; McCabe & Dettinger, 1999). There is no universal indicator for ENSO (Beebee & Manga, 2004); this study used the Southern Oscillation Index (SOI) as a measure for ENSO. The SOI is measured from the difference in SLPs

between Tahiti and Darwin, Australia (Redmond & Koch, 1991). The winter (October to March) SOI was used instead of the annual SOI because it better represents the ENSO signature (Mann et al., 2000; McCabe & Dettinger, 1999).

The PDO is low-frequency climate pattern centered in the North Pacific Ocean with oscillation cycles lasting up to 50 years (Hamlet & Lettenmaier, 1999; Mantua et al., 1997). PDO cycles behave similarly to those of ENSO in which the warm phases bring above-normal streamflow in the southwestern United States and below-normal precipitation in the Pacific Northwest (Mantua et al., 1997). Inverse streamflow patterns are associated with the cool phases of PDO.

The AMO is also a low-frequency climate pattern, but is located in the North Atlantic Ocean that has a 65-80 year cycle (Enfield et al., 2001; Gray et al., 2004). Warm phases of the AMO bring below-normal streamflow over much of the western United States, while cool phases bring above-normal streamflow (Enfield et al., 2001; Gray et al., 2004).

The NAO is measured by the difference in SLPs between Iceland and the Azores and oscillates at a decadal time scale (Hurrell & Van Loon, 1997). The influence of the NAO phases is more dominant over the eastern United States; however, NAO may influence climate over the Pacific Northwest and the northern Rocky Mountains (Hunter et al., 2006).

Table 17: Streamflow gage locations

Gage Number	Region/Basin	Gage Location
1	UCRB	Fontenelle Creek near Fontenelle, WY
2	UCRB	Hams Fork near Frontier, WY
3	UCRB	Green river at Green River, WY
4	UCRB	Green River near Greendale, UT
5	UCRB	Green River at Green River, UT
6	UCRB	Colorado River above Glenwood Springs, CO
7	UCRB	Gunnison at Crystal Reservoir, CO
8	UCRB	Gunnison River near Grand Junction, CO
9	UCRB	Dolores River near Cisco, UT
10	UCRB	Colorado River near Cisco, UT
11	UCRB	Animas River at Durango, CO
12	UCRB	San Juan River at Archuleta, NM
13	UCRB	San Juan River near Bluff, UT
14	UCRB	Colorado River at Lees Ferry, AZ
15	LCRB	Gila River near Solomon, AZ
16	SPB	Cache La Poudre River near Ft Collins, CO
17	ARK	Canadian River near Sanchez, NM
18	RGB	Saguache Creek near Saguache, CO
19	RGB	Rio Grande near Del Norte, CO
20	RGB	Conejos River near Mogote, CO
21	RGB	Santa Fe River near Santa Fe, NM

Table 18: Data sources for streamflow gages and oscillation indices.

Reconstructions			Instrumental Record	
Gage Number	Source	Data Available*	Source	Data Available*
1, 2, 4	Barnett et al., 2007	1615-1999	United States Geological Survey, 2011	1952-2010
3, 5	Woodhouse et al., 2006	1615-1998	United States Bureau of Reclamation, 2011	1906-2007
6	Woodhouse et al., 2006	1525-1997	United States Geological Survey, 2011	1906-2007
7, 8, 9, 10, 12, 13	Woodhouse et al., 2006	1569-1996	United States Bureau of Reclamation, 2011	1906-2007
11	Woodhouse and Lukas, 2006a	1470-2002	United States Geological Survey, 2011	1928-2010
14	Meko et al., 2007	762-2005	United States Bureau of Reclamation, 2011	1906-2007
15	Meko and Hirschboeck, 2008	1332-2005	United States Geological Survey, 2011	1921-2010
16	Woodhouse and Lukas, 2006a	1615-1999	United States Geological Survey, 2011	1911-2007
17	Meko et al., 2007	1604-1997	United States Geological Survey, 2011	1937-2010
18	Woodhouse and Lukas, 2006b	1520-2000	United States Geological Survey, 2011	1915-2007
19	Woodhouse and Lukas, 2006b	1508-2002	United States Geological Survey, 2011	1891-2010
20	Woodhouse and Lukas, 2006b	1508-2002	United States Geological Survey, 2011	1913-2010
21	Margolis, 2011	1592-2007	United States Geological Survey, 2011	1914-2010
ENSO	Jones and Mann, 2004	1650-1980	Australian Government Bureau of Meteorology, 2011	1876-2010
PDO	Shen et al., 2006	1470-1998	JISAO, 2011	1900-2010
NAO	Luterbacher et al., 2001	1658-2001	Hurrell, 2011	1865-2010
AMO	Gray et al., 2004	1567-1985	ESRL, 2011	1856-2010

*Reconstructed data is used from 1685-1952 and instrumental data is used from 1953-2007

5.3 Method

This study used a forecast model that is generated from a multiclass kernel-based SVM model. SVMs are based on statistical learning theory, in which a learning algorithm derived from optimization theory uses a hypothesis space of linear functions in a higher dimensional feature space (Vapnik, 1995, 1998; Scholkopf & Smola, 2002; Cristianini & Shawe-Taylor, 2000). SVMs are designed to minimize the structural risk in order to balance the error rate on the training data and the capacity of the model to learn from the training data (Scholkopf & Smola, 2002). Nonlinearities associated with hydroclimatic processes are handled by incorporating kernels into the SVM model. The kernels aid the model in obtaining a global optimum to produce accurate forecasts, while maintaining generalization within the model so that it can be used to provide successful forecasts with new instances (Scholkopf & Smola, 2002). This serves as an advantage over other data-driven models such as ANNs, where the model may converge upon local optima or over-fit the training data, resulting in poor forecasts. A detailed discussion of the SVM model is provided by Vapnik (1995, 1998) and a brief description of the formulation is given here.

The goal of the SVM model is to formulate a relationship between the oscillation indices as input variables and streamflow volume as output variables given as input vector (x), through the functional dependency f(x). The estimation of f(x) is achieved through an optimization problem from the best subset of instances or number of support vectors (N) as follows:

Minimize

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

Subject to

$$\begin{cases} f(x) - \langle w, x \rangle - b \leq \varepsilon + \xi_i \\ \langle w, x \rangle + b - f(x) \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad (2)$$

To obtain

$$f(x) = \langle w, x \rangle + b \quad (3)$$

where ‹w, x›is the dot product of the support vector weight (w) and the input vector (x), b is the bias, C is the capacity parameter cost, ε is the insensitive loss function (Vapnik, 1995), and ξ_i and ξ_i* are slack variables. ξ_i and ξ_i* are used in Vapnik's (1995) ε-insensitive loss function (Figure 38) to determine the degree to which instances are penalized for absolute errors which are greater than ε. For absolute errors less than ε, the instances have a 0 value in the loss function (i.e. ξ_i=ξ_i* = 0) and are not included in the objective function. This limits the number of instances which are used to estimate f(x). The constant C > 0 determines the trade-off between the complexity of f(x) and the amount up to which deviations greater than ε are tolerated. A larger C places a higher penalty on large errors and will reduce the approximation error, but makes the problem more unconstrained.

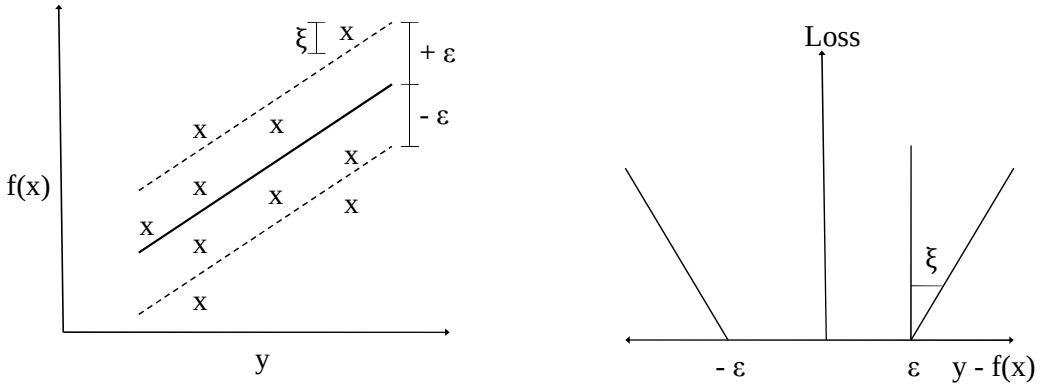

Figure 38: ε-insensitive loss function adopted from Vapnik (1995).

In order to introduce non-linear dependencies, a dual formulation to the standard SVM (Equations 1-3) is often used. This is accomplished by introducing the Lagrange multipliers α and α* and the kernel function k(x, x_i) to replace the dot products of input instances. The kernels may be represented as the dot products of nonlinear transformation functions. The result is a linear function in the feature space that can be more easily solved (Figure 39). Differentiating the dual form with respect to the primal variables (w, b, ξ_i, ξ_i*) yields the following optimization problem:

Maximize

$$W(\alpha^*, \alpha) = \begin{matrix} -\varepsilon \sum_{i=1}^{N}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{N} f_i * (\alpha_i - \alpha_i^*) \\ -\frac{1}{2}\sum_{i,j=1}^{N}(\alpha_i - \alpha_i^*)\left(\alpha_j - \alpha_j^*\right)k\left(x_i, x_j\right) \end{matrix} \quad (4)$$

Subject to

$$\begin{matrix} \sum_{i=1}^{N}(\alpha_i^* - \alpha_i) = 0, \\ 0 \le \alpha_i, \alpha_i^* \le C \end{matrix} \quad (5)$$

To obtain

$$f(x) = \sum_{i=1}^{n}(\alpha_i^* - \alpha_i)k(x, x_i) + b \quad (6)$$

where the x_is are the selected instances that have an error greater than ε and are called the support vectors, and n (usually n « N) is the number of support vectors. The SVM has the advantage over other models is dealing with nonlinear approximations. In order to successfully develop an SVM model, three parameters must be specified including the kernel function, ε-insensitive function, and the capacity parameter cost C (Vapnik, 1995, 1998).

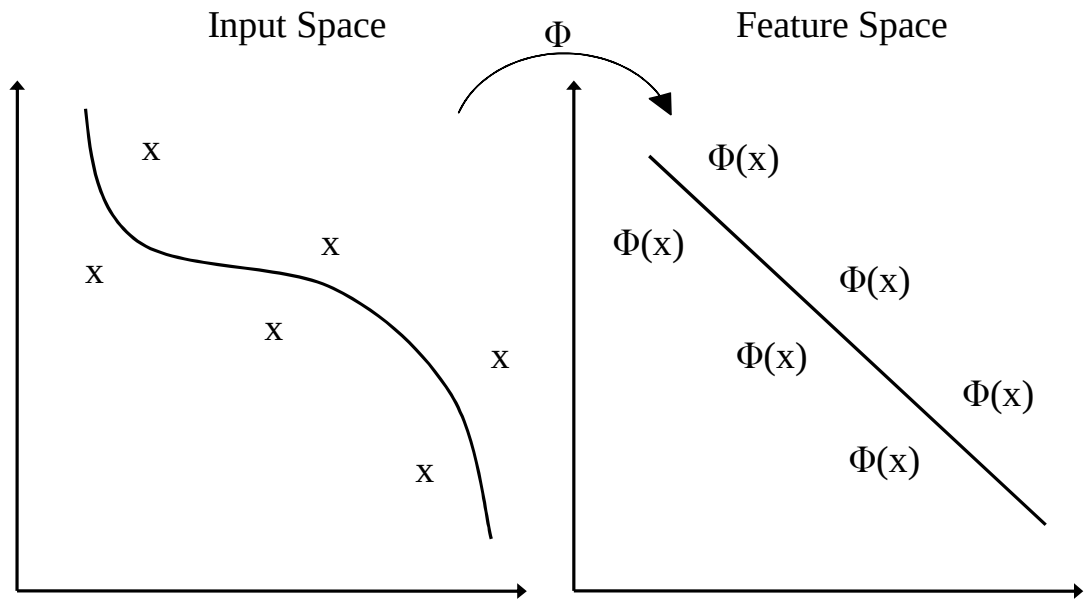

Figure 39: Conceptual representation of Kernel transformation adopted from Asefa et al. (2004). Φ is the transform function.

In addition, a standard multiple linear regression (MLR) model was used as a forecast model to provide a comparison to the SVM model. The MLR model was subjected to the same methodology and performance analysis as the SVM model.

The forecast model used all four oceanic-atmospheric oscillation indices (i.e. ENSO, PDO, AMO, and NAO) as predictors, and used one streamflow gage as the predictand. The entire dataset was used from 1658 to 2007. A 1-year lead time was used to simulate a forecast. The forecast models were performed with a leave-one-out cross-validation technique, where each instance was held out in turn for testing and the model was trained on all of the remaining instances (Witten et al., 2011). The error estimates were averaged together to yield an overall result. This allows the model to examine the

entire dataset and better capture the non-stationarity associated with the dataset as compared to the standard practice of splitting the dataset into training and testing datasets. The forecasts were subjected to a set of performance measures that are described in the following section.

5.4 Performance measures

Several performance measures were used to analyze the forecasts including the MAE, RMSE, RSR, R, NSE, and LEPS SK. These performance measures provided a thorough and independent analysis. Categorical performance ratings were adopted from Moriasi et al. (2007) which defined forecasts as 'unsatisfactory,' 'satisfactory,' 'good,' or 'very good' (Table 19).

Table 19: Performance Rating for RSR, R, and NSE.

Performance Rating	RSR*	R	NSE*
Very Good	$0.00 \leq RSR \leq 0.50$	$0.85 \leq R \leq 1.0$	$0.75 \leq NSE \leq 1.0$
Good	$0.50 < RSR \leq 0.60$	$0.80 \leq R < 0.85$	$0.65 \leq NSE < 0.75$
Satisfactory	$0.60 < RSR \leq 0.70$	$0.70 \leq R < 0.80$	$0.50 \leq NSE < 0.65$
Unsatisfactory	$0.70 < RSR \leq 1.0$	$0.00 \leq R < 0.70$	$0.00 \leq NSE < 0.50$

*RSR andNSE adopted from Moriasi et al., [2007]

The MAE and RMSE are common performance measures because they provide errors in the same units as the tested data, and a value of 0 indicates a perfect forecast (Moriasi et al., 2007; Witten et al., 2011). The RSR is defined as the RMSE divided by the standard deviation of the observed dataset to standardize the RMSE (Moriasi et al., 2007). This provides a better comparison of the model performance than the RMSE because streamflow volume will greatly vary between stations.

R is a measure of the linear relationship between the measured and predicted values, which typically ranges from 1 (perfect positive correlation), 0 (no correlation),

and -1 (perfect negative correlation). Moriasi et al. (2007) indicates that acceptable models are obtained when R is greater than 0.50. However, the current study develops stricter criteria to complement the RSR and NSE performance ratings adopted from Moriasi et al. (2007) which are shown in Table 19.

The NSE measures the residual variance (noise) in comparison with the measured data variance (information) and is defined by Moriasi et al., 2007. It measures the tendency of the measured and predicted data to follow a 1:1 relationship (Legates & McCabe, 1999; Moriasi et al., 2007). NSE ranges from 1 to -∞, where a positive NSE is acceptable because it indicates the model is a better predictor than the observed mean (Legates & McCabe, 1999; Moriasi et al., 2007).

The LEPS SK determines the accuracy of forecast predictions, where higher weights are given to values that are further from the mean of the dataset (Potts et al., 1996). LEPS SK includes climatology in its calculation and serves as a better indicator than climatology (Casey, 1998; Potts et al., 1996). LEPS SK ranges from -100 to 100, where it is considered ‘bad’ if it is below -10, ‘poor’ if it is below -5, ‘satisfactory’ if it is greater than 5, and ‘good’ if it is greater than 10 (Casey, 1998).

In addition, a visual inspection of the forecast was performed through scatter plots, box plots, and non-exceedance plots. The SVM and MLR models were subjected to this set of performance measures.

5.5 Results

This section describes the results of the forecasts obtained for the SVM model and the MLR model. Section 5.6.1 describes the forecasts obtained from the SVM model

when using a 1-year lead time. Section 5.6.2 presents the 5-year forecasts obtained with the SVM model. Finally, Section 5.6.3 provides the results of the MLR model in comparison to the SVM model.

5.5.1 1-Year Lead Time SVM Forecast

This section describes the results obtained when ENSO, PDO, AMO, and NAO were used together to forecast streamflow volumes for each of the selected gages using the SVM model with a 1-year lead time as described in the method section.

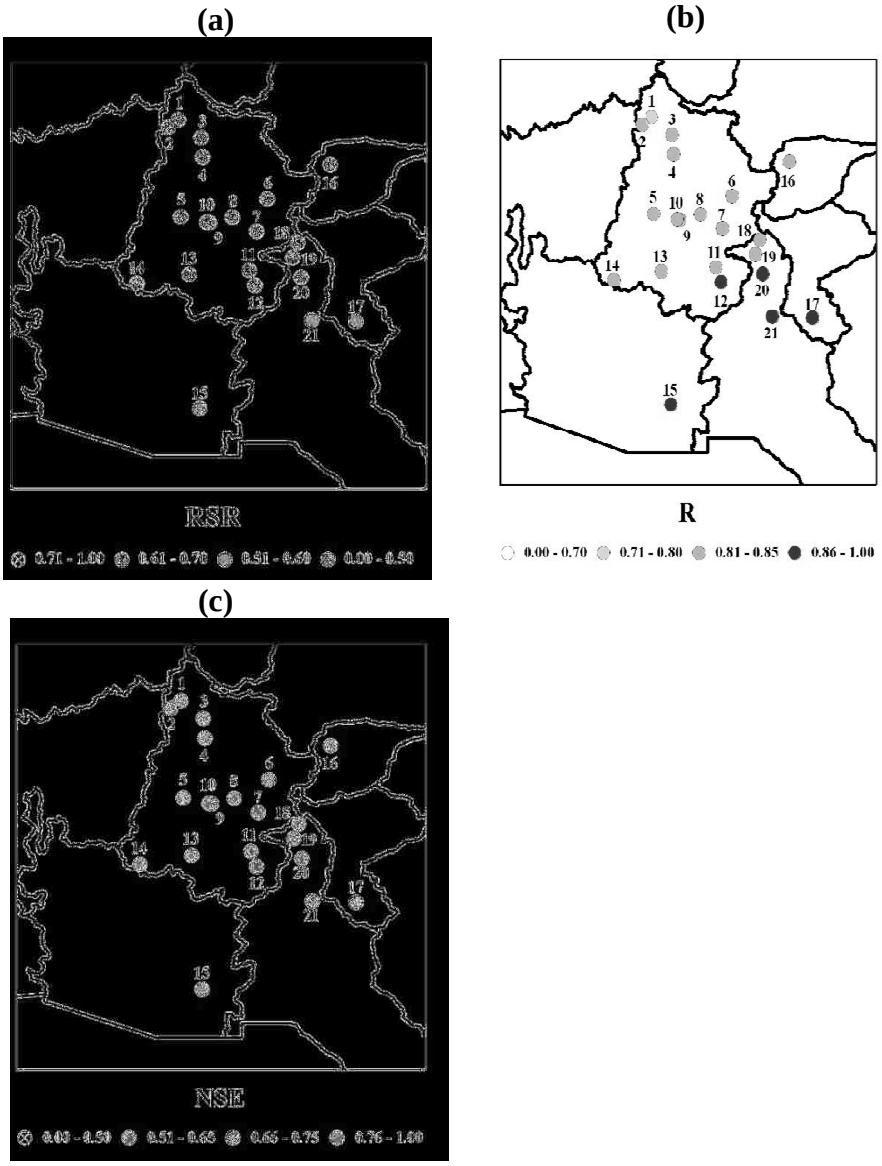

Figure 40: SVM results run with a 1-year lead time displayed as spatial map showing (a) RSR, (b) R, and (c) NSE.

Figure 40a provides a spatial representation of the performance measures obtained for the SVM model run with a 1-year lead time. The results for RSR indicated that 'satisfactory' forecasts were obtained for 18of the 21 gages. These gages were located in the northern portion of the study area including nearly all of the gages in the UCRB, RGB, and SPB. Of the remaining gages, two 'good' forecasts were made in the middle of the study area for Gages 17 and 21. Finally, a 'very good' forecast was made for Gage 15

in the LCRB. The results for NSE agreed with the results for RSR. A ‘very good’ forecast was obtained at Gage 15 and ‘good’ forecasts were obtained at Gages 17 and 21 near the middle of the study area. The remaining gages produced ‘satisfactory’ forecasts in the northern areas of the study area. While the RSR and NSE indicated that the model produced ‘satisfactory’ forecasts, the R indicated that the forecasts were more accurate. Gages 1 in the northern region of the UCRB was the only gage with a ‘satisfactory’ correlation. ‘Good’ correlations were obtained for 15of the 21 gages and included the majority of the gages in the UCRB and SPB. The remaining 5gages were located in the southern portions of the study area and produced ‘very good’ correlations ($0.70 \leq R < 0.80$). These included Gages 12, 15, 17, 20, and 21 with the last three gages located in the RGB. The spatial representation of these performance measures showed that a better model performance was achieved for gages located in the southeastern portion of the study area. Overall, the results indicated that accurate forecasts are produced at all gages with the SVM model.

Table 20 provides the individual results of the performance measures. The MAE and RMSE were relatively low in comparison to the average flow at each gage; however, it was difficult to compare these measures directly between gages. The LEPS SKs for the all of the gages indicated that a ‘good’ forecast model was achieved. In fact the LEPS SKs were all greater than 50, which indicated that very good forecasting models were produced. This shows that the SVM model was able to produce accurate forecasts and that the model performed better than climatology. The RSR, R, and NSE results have been summarized in the spatial representation shown in Figure 40a.

Table 20: Summary of SVM Results run with a 1-year Lead Time

Gage	RMSE*	MAE*	RSR	R	NSE	LEPS SK
1	11.90	6.90	0.69	0.79	0.53	54.9
2	15.97	9.53	0.68	0.80	0.54	53.7
3	233.38	138.15	0.65	0.82	0.58	55.8
4	369.58	218.45	0.64	0.82	0.59	56.5
5	1069.50	613.32	0.65	0.82	0.58	56.5
6	345.16	197.46	0.66	0.81	0.57	56.9
7	241.68	144.12	0.66	0.81	0.57	52.5
8	455.79	269.91	0.65	0.81	0.58	54.1
9	220.19	137.11	0.64	0.83	0.59	52.4
10	1252.80	746.13	0.65	0.81	0.57	52.6
11	108.19	66.88	0.62	0.85	0.62	53.9
12	299.22	185.59	0.60	0.85	0.64	54.0
13	531.69	329.18	0.63	0.83	0.61	52.2
14	2387.70	1414.30	0.62	0.84	0.62	55.3
15	126.48	52.53	0.49	0.91	0.76	67.5
16	61.04	36.55	0.66	0.81	0.56	54.1
17	50.71	19.38	0.54	0.89	0.71	67.7
18	10.82	6.56	0.65	0.81	0.58	54.6
19	119.77	73.56	0.61	0.84	0.62	54.1
20	43.37	26.30	0.61	0.85	0.62	53.8
21	1.27	0.66	0.55	0.88	0.70	64.6

*In 1000 ac-ft (1 ac-ft = 1233.5 m^3)

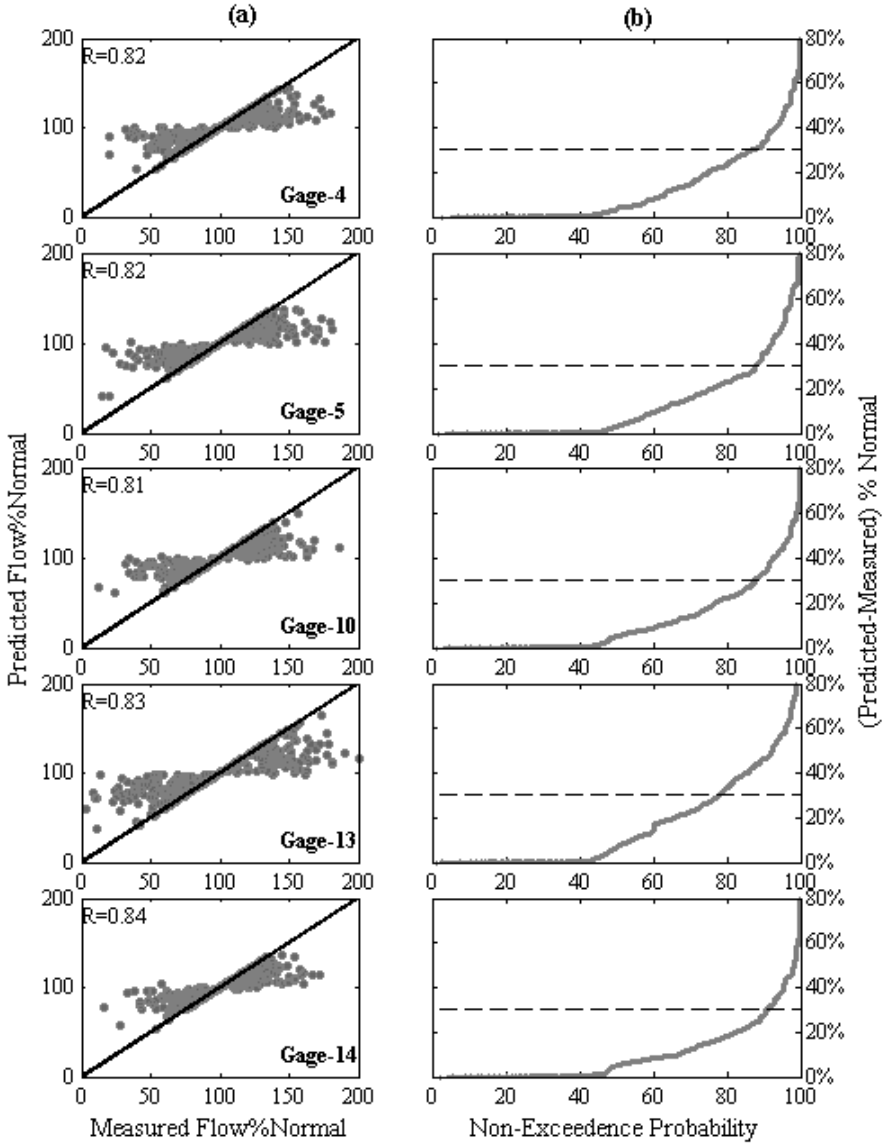

Figure 41: 1-year lead time results for5 selected gages are shown for the SVM model as (a) scatter plots and (b) non-exceedance plots as well as the MLR model as (c) scatter plots and (d) non-exceedance plots. The diagonal line in the scatter plots represents the 1:1 bisector line. The horizontal dashed line in the non-exceedance plots shows an error value of 30%.

Scatter plots and non-exceedance plots in Figure 41 show a visualization of the ability of the model to produce accurate forecasts at the 1-year lead time. The scatter plots (Figure 41a) showed that several of the instances lied on or near the 1:1 bisector line, indicating a perfect forecast was made. However, the plots also showed that the

model was prone to missing extreme events. The SVM model under predicted very high flows and over predicted very low flows. The magnitude of the error was shown to increase for measured flows which are further from the mean of the dataset. The non-exceedance plots (Figure 41b) examined the accuracy of the model in terms of the errors associated with the model performance. About 40% - 45% of the forecasts exhibited no error, which indicated a perfect forecast was made. In addition, approximately 75% - 90% of the forecasts had an error of 30% or less, which indicates that the model was capable of producing accurate predictions for the majority of the forecasts. The instances that had errors greater than 30% correspond to the instances that were further away from the 1:1 bisector line on the scatter plots (Figure 41a). From these results, the SVM model was shown to produce accurate forecasts.

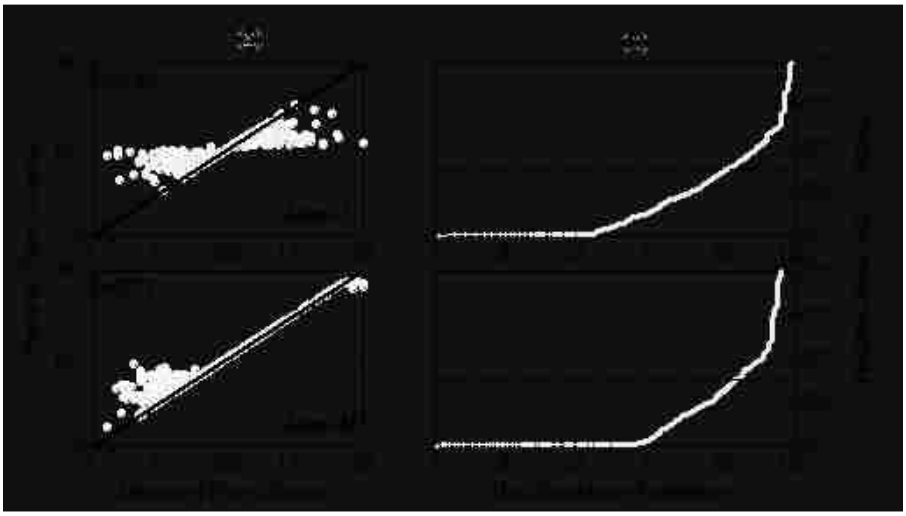

Figure 42: 1-year lead time results for Gage2 and Gage 15 are shown for the SVM model as (a) scatter plots and (b) non-exceedance plots. The diagonal line in the scatter plots represents the 1:1 bisector line. The horizontal dashed line in the non-exceedance plots shows an error value of 30%

A further analysis was performed on Gage 2, which performed the worst, and Gage 15, which performed the best. Scatter plots and non-exceedance plots were used to

show the comparison in Figure 42. Figure 42a shows a scatter plot for Gage 2 that was comparable to the scatter plots for the other gages (Figure 41a). Gage 2 performed worse than the other gages due to the higher number of extreme events that were forecasted as the mean flow. In comparison, Gage 15 performed the best as the majority of the instances lie on the 1:1 bisector line. Figure 42b indicates that Gage 2made perfect forecasts for about 42% of the predictions and approximately 80% of the forecasts had an error of 30% or less. In comparison perfect forecasts were made at Gage 15 for 58% of the forecasts and about 85% of the forecasts had an error of 30% or less. These plots indicate that the SVM model performs exceptionally well as nearly half of the predictions are perfect; however, the errors tend to be high for instances representing extreme events.

5.5.2 5-Year Lead Time SVM Forecast

The SVM model was run with a 5-year lead time using ENSO, PDO, AMO, and NAO to forecast streamflow volumes.

A spatial representation of the performance measures obtained for the SVM model run with a 5-year lead time is shown in Figure 43a. The RSR results indicated that ‘satisfactory’ forecasts were obtained for 16 of the 21 gages, which included the majority of the gages in the UCRB, RGB, and SPB. Out of the remaining gages, 4‘good’ forecasts were made at Gages 1, 14, 17, and 21. With the exception of Gage 1, these gages were located in the middle of the study area. Finally, a ‘very good’ forecast was obtained for Gage 15 in the LCRB. The R indicated that the forecasts were very well correlated with the observations at each gage. ‘Good’ correlations were made for 15 of the 21 gages, including the majority of the gages in the UCRB, RGB, and SPB. Gages 1, 6, 10, 15, 17, and 21 produced ‘very good’ correlations, which did not have a distinct spatial layout

with the 5-year lead time. The results for NSE indicated that a ‘very good’ forecast was obtained at Gage 15 in the LCRB. Gages 17 and 21 in the southeast portion of the study area produced ‘good’ forecasts. Finally, ‘satisfactory’ forecasts were obtained at the remaining gages in the UCRB, SPB, and RGB. Similar to the 1-year lead time, forecasts made with the 5-year lead time indicated that model performance was better for gages located in the southeastern portions of the study area. Overall, the results indicated that accurate forecasts are produced at all gages with the SVM model.

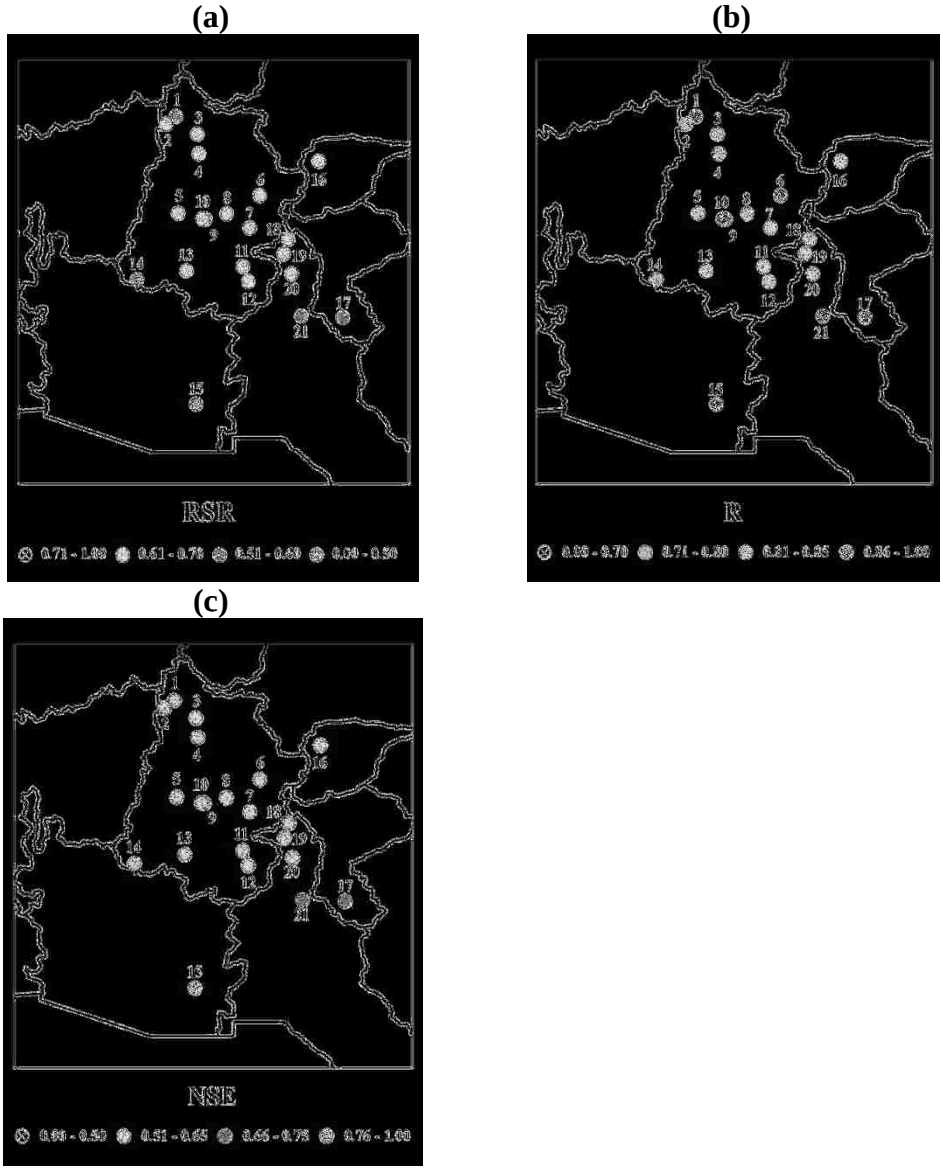

Figure 43: SVM results run with a 5-year lead time displayed as spatial map showing (a) RSR, (b) R, and (c) NSE.

The individual results of the performance measures are displayed in Table 21. The MAE and RMSE were each small in comparison to the mean flow for each gage. A 'good' forecast model was achieved for all of the gages, with each gage having a LEPS SK greater than 50. The SVM model was able to produce accurate forecasts and the model performed better than climatology. Figure 43 summarized the RSR, R, and NSE results according to their performance ratings in a spatial map. The results from the 5-

year lead time are comparable to the results from the 1-year lead time (Table 20), indicating that the forecast lead time can be extended out to 5 years without a degradation of the performance. In fact, an improvement in the forecast was experienced for several of the instances with a 5-year lead time in comparison to the 1-year lead time forecasts.

Table 21: Summary of SVM Results run with a 5-year Lead Time

Gage	RMSE*	MAE*	RSR	R	NSE	LEPS SK
1	10.32	5.95	0.59	0.86	0.65	61.8
2	14.26	8.02	0.60	0.84	0.63	62.0
3	232.50	141.00	0.65	0.84	0.58	54.5
4	354.17	214.09	0.61	0.85	0.63	57.5
5	1017.20	576.70	0.61	0.84	0.62	60.0
6	320.25	185.81	0.61	0.85	0.63	60.5
7	224.23	132.00	0.61	0.85	0.63	57.3
8	438.43	261.23	0.62	0.83	0.61	56.2
9	209.40	127.80	0.60	0.85	0.63	55.9
10	1164.00	698.22	0.61	0.85	0.63	56.6
11	108.56	64.56	0.62	0.83	0.62	55.8
12	306.20	184.02	0.61	0.84	0.62	54.9
13	517.62	308.40	0.61	0.84	0.63	55.8
14	2303.30	1327.00	0.59	0.85	0.65	58.4
15	122.86	52.51	0.47	0.92	0.78	67.8
16	57.11	34.87	0.62	0.84	0.62	56.6
17	49.36	18.54	0.52	0.90	0.73	70.5
18	10.19	6.24	0.61	0.85	0.62	57.0
19	119.92	71.62	0.61	0.84	0.63	55.5
20	42.89	24.37	0.60	0.84	0.63	58.0
21	1.25	0.67	0.54	0.90	0.71	64.9

*In 1000 ac-ft (1 ac-ft = 1233.5 m^3)

Scatter plots and non-exceedance plots are shown in Figure 44 for 5 streamflow gages, which exhibited the highest annual streamflow volume. The scatter plots in Figure 8a indicated that several of the instances lied on or near the 1:1 bisector line, which represents a perfect forecast. However, the high flows were often under predicted and the low flows were over predicted. The magnitudes of the errors increased for instances that

were further from the mean, such as extreme events, as compared to flows that were closer to the mean of the dataset. Figure 44b showed the non-exceedance plots, which examined the accuracy of the model in terms of the errors associated with the model performance. About 40% - 50% of the forecasts exhibited no error, which indicated a perfect forecast was made. In addition, approximately 80% - 90% of the forecasts had an error of 30% or less, which indicates that accurate forecasts were made for the majority of the forecasts. The instances with higher errors were shown to lie further from the 1:1 bisector line on the scatter plots (Figure 8a). From these results, the SVM model was shown to produce accurate forecasts with a 5-year lead time.

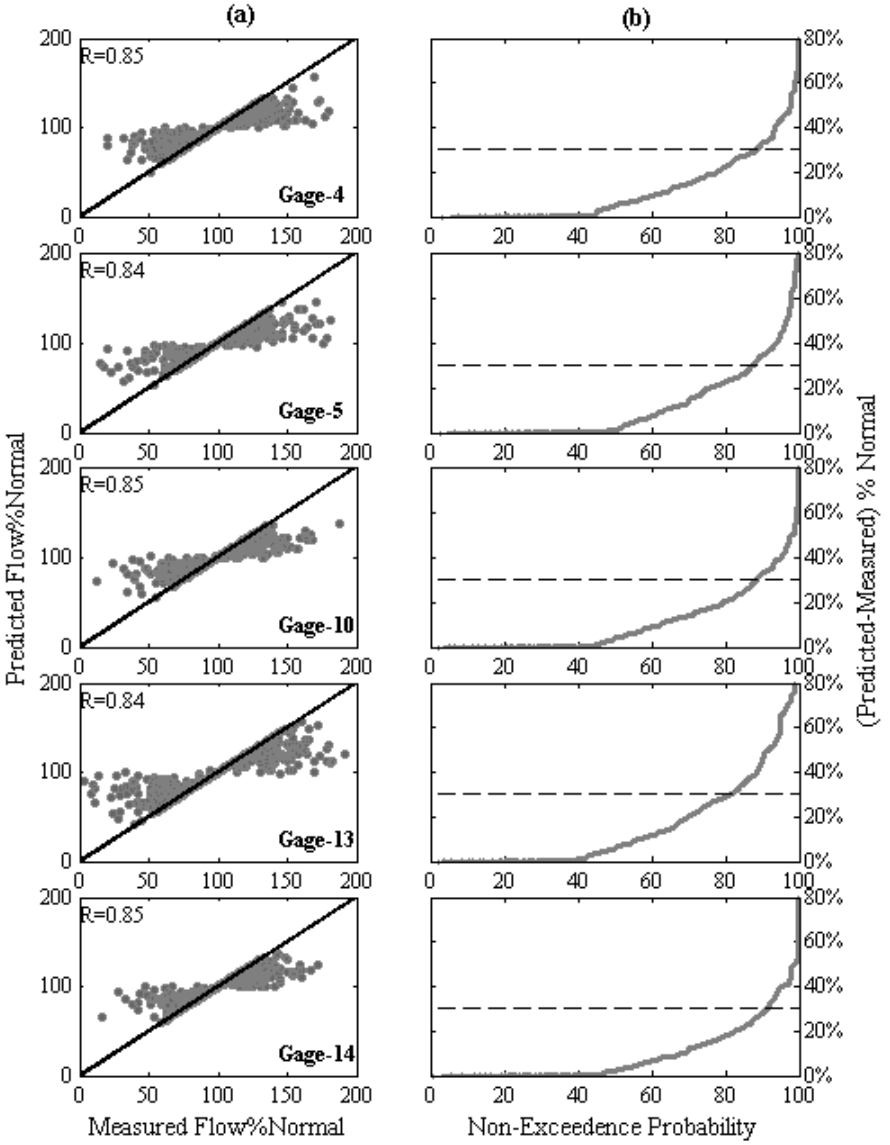

Figure 44: 5-year lead time results for 5 selected gages are shown for the SVM model as (a) scatter plots and (b) non-exceedance plots. The diagonal line in the scatter plots represents the 1:1 bisector line. The horizontal dashed line in the non-exceedance plots shows an error value of 30%.

Of the gages tested, Gage 11 performed the worst and Gage 15 performed the best. A further examination of these gages was provided through scatter plots and non-exceedance plots in Figure 45. Figure 45a shows a scatter plot for Gage 11 and shows that while the majority of the instances lied on or near the 1:1 bisector line, there are several instances that were not correctly predicted in the model. The model tended to

miss extreme events. In contrast, Gage 15 provided the best performance with the vast majority of the instances lying on or near the 1:1 bisector line and much fewer incorrectly forecasted instances. Figure 45b showed that perfect forecasts were made for about 42% of the predictions at Gage 11 and approximately 90% of the forecasts had an error of 30% or less. In comparison perfect forecasts were made at Gage 15 for approximately 55% of the forecasts and about 85% of the forecasts had an error of 30% or less. Although Gage 11 produced fewer forecasts with high errors than Gage 15, the model for Gage 15 produced many more forecasts with little or no error. These results show that the SVM model was accurate for the majority of the forecasts, although the extreme events were difficult to forecast in the model.

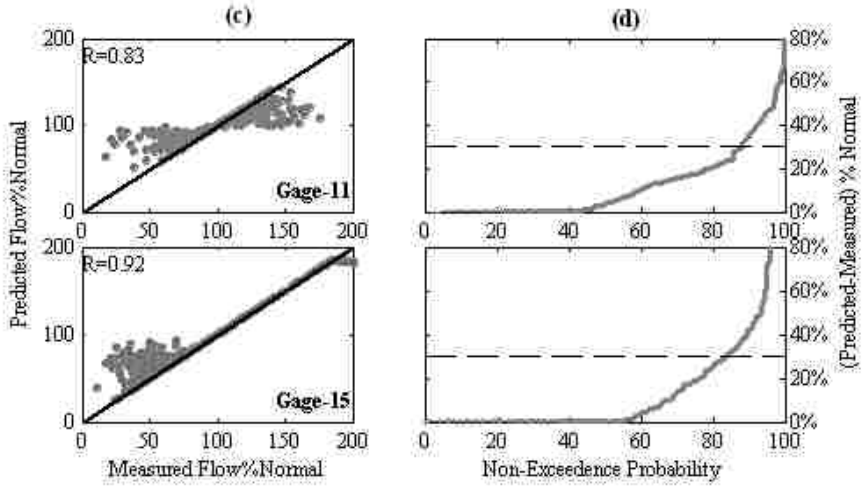

Figure 45: 5-year lead time results for Gage 11 and Gage 15 are shown for the SVM model as (a) scatter plots and (b) non-exceedance plots. The diagonal line in the scatter plots represents the 1:1 bisector line. The horizontal dashed line in the non-exceedance plots shows an error value of 30%.

5.5.3 MLR Forecast

The MLR modeling approach was run in comparison with the SVM models. As shown in Figure 46, all of the gages produced 'unsatisfactory' forecasts for the 1-year lead time in terms of the RSR, R, and NSE. The MAE and RMSE for MLR were much higher than those obtained for the SVM model (Table 22). However, the LEPS SK indicates that the forecast model has no skill with a LEPS SK < 5 for all of the forecast models, indicating that climatology outperforms the MLR model. In addition, the scatter plots in Figure 47 show that the forecasts tended to return the mean of the dataset as the predictions, indicating that the MLR model was unable to determine any relationships in the dataset. In addition, the non-exceedance plots in Figure 47 show that 30% - 75% of the forecasts had an error of 30% or less and no perfect forecasts were made. This indicates that the MLR model produced inaccurate results.

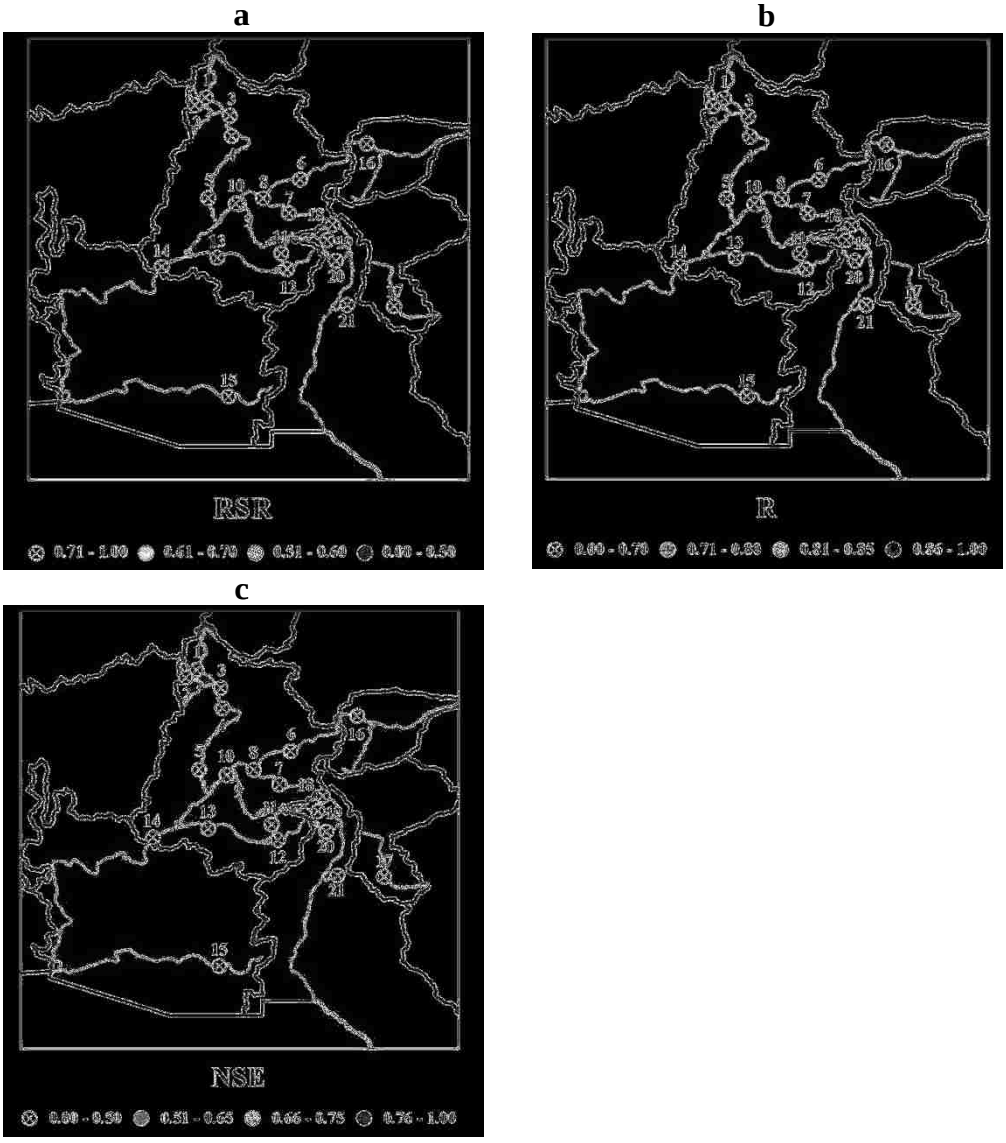

Figure 46: MLR results at a 1-year lead-time displayed in a spatial map for (a) RSR, (b) R, and (c) NSE.

Table 22: Summary of MLR Results run with a 1-year Lead Time

Gage	RMSE*	MAE*	RSR	R	NSE	LEPS SK
1	17.46	13.80	1.01	-0.10	-0.02	1.0
2	23.80	18.97	1.01	-0.15	-0.03	1.0
3	361.47	291.08	1.01	-0.06	-0.02	1.0
4	583.08	471.66	1.01	-0.08	-0.02	1.2
5	1664.20	1317.40	1.01	-0.13	-0.02	1.4
6	529.29	417.71	1.01	-0.11	-0.02	1.1
7	370.21	300.69	1.01	-0.01	-0.01	0.6
8	705.47	574.59	1.00	0.05	-0.01	0.3
9	345.79	284.70	1.00	0.07	0.00	0.8
10	1918.50	1562.90	1.00	0.04	-0.01	0.5
11	175.65	144.24	1.00	0.04	-0.01	0.6
12	500.79	412.79	1.01	0.00	-0.01	0.1
13	852.58	701.42	1.00	0.01	-0.01	0.2
14	3860.90	3173.20	1.00	0.08	0.00	0.5
15	253.28	192.06	0.98	0.21	0.04	3.4
16	92.63	74.51	1.01	-0.04	-0.02	0.8
17	94.71	67.82	1.00	0.04	-0.01	0.0
18	16.78	13.54	1.01	-0.07	-0.02	0.5
19	196.96	161.98	1.01	-0.01	-0.02	0.1
20	71.34	58.18	1.01	-0.08	-0.02	0.4
21	2.28	1.81	0.98	0.18	0.03	2.8

*In 1000 ac-ft (1 ac-ft = 1233.5 m^3)

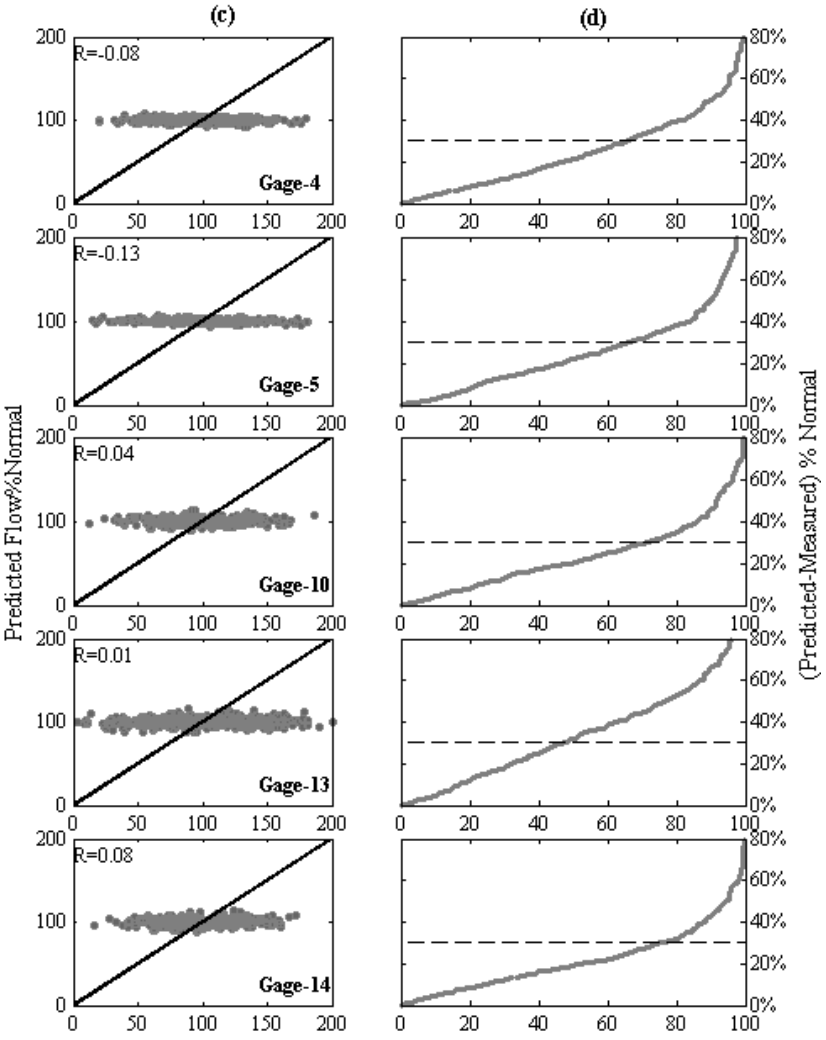

Figure 47: 1-year lead time results using the MLR model for 5 selected gage stations shown as scatter plots and non-exceedance plots. The diagonal line in the scatter plots represents the 1:1 bisector line. The horizontal dashed line in the non-exceedance plots shows an error value of 30%

5.6 Discussion and Conclusion

This study used an extended dataset through the use of reconstructions in order to generate a streamflow forecast for 21 gages in the western United States using an SVM model. The SVM model was able to produce 'satisfactory' to 'very good' streamflow forecasts at a 1-year lead time, based upon the performance measures adopted from Moriasi et al. (2005). These yearly forecasts may also be disaggregated into finer temporal resolutions as described by Kalra and Ahmad (2011). The forecasts generated by the SVM model can complement existing physically-based models for water resources management.

The spatial maps in Figure 40 and Figure 43 reveal that the SVM model performed better for gages located in the southeastern portion of the study area as compared to gages located in the northwestern portion of the study area. The RSR and NSE reveal that 'good' and 'very good' forecasts were made in the southeastern portions of the study area as compared to the 'satisfactory' forecasts obtained for the remainder of the study area. In addition, the spatial map for R indicates that 'very good' correlations were obtained in the southeastern portion of the basin; gages located in the middle of the study area had 'good' correlations; and gages in the northwestern portion of the study area had 'satisfactory' correlations. This is indicative that the oceanic-atmospheric oscillations have a more prominent influence over the southwestern portion of the study area compared to the remainder of the study area. Such oscillations as the ENSO and PDO have been shown to have a stronger influence over streamflow in the LCRB and the RGB compared to the UCRB (Cayan et al., 1999). It is also clear from the spatial maps that the SVM model was capable of producing forecasts regardless of the streamflow

gages being located in the headwaters of the basin or located in the downstream portions of the basin.

The correlation between oscillation indices and streamflow variability has been examined in several studies (e.g. Gutzler et al., 2002; Hamlet & Lettenmaier, 1999; Kalra & Ahmad, 2009; McCabe et al., 2007; Redmond & Koch, 1991; Tootle et al., 2005). However, the coupling of oscillation indices, such as the Pacific indices of ENSO and PDO, has been shown to reflect higher streamflow variability as compared to examining the oscillations individually (Gutzler et al., 2002; Timilsena et al., 2009; Tootle et al., 2005). Furthermore, it is clear that oscillations in both the Pacific and Atlantic play a significant role in climate variability across the United States (Gray et al., 2003; McCabe et al., 2004; McCabe et al., 2007; Tootle et al., 2005). In the current study, all four oscillation indices (i.e. ENSO, PDO, AMO, and NAO) are used together to represent the interactions of Pacific and Atlantic variability and extract the maximum information from the indices. Although the results are not shown in this study, the removal of one or more of the oscillation indices results in a reduction of the model performance because information is lost.

The SVM model was shown to outperform the standard MLR technique. At all gages with a 1-year lead time, the RMSE, MAE, RSR, R, and NSE results for the SVM model (Table 20) were much better than the respective results obtained with the MLR model (Table 22). In addition, the MLR model was unable to outperform climatology as shown by the LEPS SK, but the SVM model was much better than climatology. The SVM model develops forecasts based upon the premise that similar inputs yield similar outputs. The kernel function allows the SVM model to identify instances which have both

similar inputs and outputs. This allows the model to retain the nonlinear relationships between the input oscillations and output streamflow when they are transformed into the feature space and fitted with a regression plane. In comparison, it was difficult for the MLR model to extract the nonlinear relationships using basic linear regression methods and as a result it was only able to capture the mean of the dataset.

The forecasts provided using the SVM model can be further improved by providing it with more information. Witten et al. (2011) explains that data-driven models are often more robust and perform better when trained on a larger training set. This is where paleoclimate reconstructions significantly help the SVM model, by extending the period of record much further than the instrumental record. A study by Kalra and Ahmad (2009) used an SVM model approach to forecast streamflow in the UCRB using the instrumental record only. The model was capable of producing accurate forecasts with an R of 0.72, 0.87, and 0.81 for Green River at Green River (Gage 5), Colorado River near Cisco (Gage 10), and Colorado River at Lees Ferry (Gage 14), respectively. In comparison to the current study, the addition of reconstructions (Table 20) results in an R of 0.83, 0.82, and 0.84 for the same respective gages. From this simple comparison, the addition of paleoclimate reconstructions did not seem to significantly improve the quality of the forecasts nor did it substantially degrade the forecasts.

The results of this study contribute to a better understanding of the long-term impacts of oceanic-atmospheric oscillations on streamflow variability in the western United States. In this study, using reconstructions, the period of record was extended back to 1658. This differs from previous studies, which were limited to a period of record of 50-100 years (e.g. Gutzler et al., 2002; Hamlet & Lettenmaier, 1999; Hunter et al.,

2006; Kalra & Ahmad, 2009; McCabe et al., 2007; Redmond & Koch, 1991; Tootle et al., 2005).

The data-driven SVM model provides a quantitative forecast which yields 'satisfactory' to 'very good' results with lead times up to 5 years in the current study. This serves as an improvement over current forecasting tools, which mostly provide qualitative forecasts at seasonal lead-times. Better forecasts are obtained when using all four oceanic-atmospheric oscillations as predictors. The forecasts are better when compared to climatology. Although the SVM model was not able to accurately forecast extreme events, the model results with up to 5 years of lead time can still be very useful for water resource planning and in determining the streamflow response to climate variability. The proposed modeling approach provides an alternative to complex, physically-based models and is expected to complement existing models for use in long-term water resources management.

Chapter 6: Contributions and Recommendations

6.1 Summary

Climate variability has a profound influence over the hydrologic cycle, which complicates water resources planning and management. Climate variability is an especially important variable in the arid western United States, where the scarce water resources are stressed by population growth, agricultural and industrial demands, and the desire to provide water for recreation and wildlife. Oceanic-atmospheric oscillations affect hydrologic variables such as streamflow and precipitation, which are indicators of the available water supply. It becomes necessary to understand these factors contributing to climate variability in the western United States. Since it is difficult for physically-based models to capture the complex dynamics of hydroclimatic variables, an alternate approach to understanding these relationships are provided through the use of data-driven models.

In this study, a data-driven modeling approach was performed to provide long-range forecasts of streamflow and precipitation in the western United States. This was accomplished by discovering relationships between oceanic-atmospheric oscillations and their hydrologic response through streamflow and precipitation. Paleoclimate reconstructions were used to provide a larger dataset for the data-driven models. The study focused on the arid regions of the western United States. An elevated interest was placed on the Colorado River Basin, which exemplifies the highest need for long-range forecasts due to its susceptibility to climate variability and highly strained water resources. Four tasks were performed to address the research questions and their hypotheses posed in this study.

The first task investigated Research Question #1: Can the oceanic-atmospheric oscillations be used to improve the streamflow forecast lead time in the Upper Colorado River Basin and can noise removal improve the model? It was hypothesized that there was a strong relationship between oceanic-atmospheric oscillations and streamflow variability within the Upper Colorado River Basin and the forecast lead time could be improved as a result. It was also hypothesized that the removal of noise would improve the ability of the model to produce accurate forecasts. Research Question #1 was addressed by using two data-driven models, KStar and M5P, to forecast streamflow for four gages in the UCRB. Reconstructions from 1658-1905 and instrumental records from 1906-2007 were obtained for four oceanic-atmospheric oscillation indices i.e., ENSO, PDO, AMO, and NAO to predict annual streamflow volumes at 1-5 year lead-times. A forecast evaluation was performed through the mean absolute error (MAE), root mean squared error (RMSE), RMSE-observations standard deviation ratio (RSR), correlation coefficient (R), Nash-Sutcliffe coefficient of efficiency (NSE), and linear error in probability space (LEPS) skill score (SK). The results of the KStar and M5P models were compared with a multivariate linear regression (MLR) model. The results indicated that the KStar model was capable of producing 'satisfactory' to 'good' forecasts at all lead-times when a 10-year moving average filter is applied to the dataset. The M5P model produced 'unsatisfactory' to 'satisfactory' forecasts. The LEPS SK revealed that both the KStar and M5P models were able to outperform the MLR model and climatology. 'Unsatisfactory' results were obtained when a 5-year or 3-year moving average filter was used for KStar and M5P. This indicated that the model was unable to identify patterns in the dataset because there was too much 'noise' in the dataset. These findings confirmed

the hypothesis that the lead time can be improved by up to 5 years, and that the model performance improves with the removal of noise.

The next task addressed Research Question #2: Do oceanic-atmospheric oscillations play an important role in streamflow variability in the western United States, and can the oscillations be used to increase the forecast lead time for gages located in the headwaters and for downstream gages? It was hypothesized that the influence of oceanic-atmospheric oscillations were prominent at all gage locations and could be used to forecast streamflow regardless of the variable streamflow volumes observed in the headwater or downstream locations. Research Question #2 was examined by using the KStar model to provide 1-5 year lead-time forecasts for 21 streamflow gages in the western United States, using all four of the oceanic-atmospheric oscillations as predictors. Reconstructions were used from 1658-1899 and the instrumental record was used from 1900-2007 were obtained for ENSO, PDO, AMO, and NAO as well as the 21 selected streamflow gages. Forecasts were made at 1-5 year lead-times, with a 10-year moving average applied to the dataset. Forecast evaluations were performed with the RSR, R, NSE, and LEPS SK. An MLR model was run in comparison to the KStar model. The KStar model produced ‘satisfactory’ and ‘good’ forecasts at the 1-year and 2-year lead times at all gages. This indicated that the model is capable of producing accurate forecasts regardless of the location of the streamflow gages or the magnitude of the streamflow volume. Deterioration in the forecasts was observed at longer lead times, although ‘good’ and ‘satisfactory’ forecasts were still obtainable at some streamflow gages. The MLR model produced ‘unsatisfactory’ forecasts for all gages at all lead times. The LEPS SK revealed that the KStar model was able to outperform climatology. These

findings confirmed the hypothesis that the lead time can be improved by up to 2 years, and that the model can be used to forecast streamflow regardless of the streamflow volumes or gage locations.

The third task examined Research Question #3: Can the lead time of precipitation forecasts be improved by using oceanic-atmospheric oscillations and do all indices play an equally important role in the forecast model? It was hypothesized that precipitation variability across the western United States was considerably affected by oceanic-atmospheric oscillations and the forecast lead time could be improved. In addition it was theorized that the forecast model would improve when a non-significant oscillation index was removed as a predictor. In order to address Research Question #3, the KStar data-driven model was used to provide a 1-year forecast for precipitation for 20 climate divisions in the western United States using the four oceanic-atmospheric oscillations. For this model, paleoclimate reconstructions from 1658-1899 and the instrumental record from 1900-2007 were used with a 10-year moving average filter. The model was evaluated using the MAE, RMSE, RSR, R, NSE, and LEPS SK. 'Good' precipitation forecasts were achieved with the KStar model. A thorough sensitivity analysis indicated that using all four oscillations together (i.e. ENSO, PDO, AMO, and NAO) as predictors results in the best overall forecasts. Deterioration in the forecasts was experienced when any of the four indices were dropped as predictors. The forecasts were found to be better than MLR and climatology. The first part of the hypothesis was confirmed by increasing the precipitation forecast lead time by 1 year. However, the hypothesis that dropping an oscillation index in the forecast was proven false.

The final task investigated Research Question #4: Can oceanic-atmospheric oscillations be used in an SVM model to increase the streamflow forecast lead time throughout the western United States without noise removal? It was hypothesized that the SVM model was successful in increasing the streamflow forecast lead time from the raw data throughout the western United States. This study addressed Research Question #4 by using an SVM model to provide 1-year to 5-year lead-time forecasts for 21 streamflow gages throughout the western United States The results indicated that 'satisfactory' to 'very good' forecasts were achieved for the gages at a 1-year and 5-year lead-time. It was discovered that better forecasts were obtained for gages in the southeastern portion of the study area compared to the gages in the northwestern portion of the study area. The forecast results show improvement over standard MLR models and climatology. The hypothesis that the forecast lead time could be increased using the raw data in an SVM model was confirmed.

6.2 Contributions

The results from Research Questions #1, #2, #3, and #4 provided several contributions to the fields of hydroclimatology and machine learning. The results indicated that **using all four oceanic-atmospheric oscillation indices (i.e., ENSO, PDO, AMO, and NAO) as predictors provided the best streamflow and precipitation forecasts**. A reduction in the model performance was experienced when using individual indices or combinations of 2 or 3 oscillation indices. This suggests that hydrology in the western United States cannot be explained by a single process, but is influenced by complex atmospheric interactions originating over both the Atlantic and Pacific Oceans.

This research extended the lead-time of quantitative forecasts for streamflow and precipitation. Quantitative streamflow forecasts were achieved at lead-times up to 5 years, which improves upon the seasonal forecasts of the Colorado River Basin Forecast Center. These long-range forecasts are expected to be useful for the long-term planning of water resources.

Secondly, **this study is the first attempt at using paleoclimate reconstructions in a data-driven model to forecast streamflow and precipitation at lead-times greater than a year**. Several studies suggested that the use of paleoclimate reconstructions have the potential to improve forecasts. However, there are no studies which have attempted to use paleoclimate reconstructions in a data-driven model for forecasting precipitation and streamflow

The removal of noise in the data set improves the quality of the forecasts. It was difficult for data-driven models to identify relationships in the dataset when there is too much noise present. The moving average filter used in this study did not compromise the signals associated with the oscillations, but aided in the identification and extraction of oscillation phases. A larger 10-year moving average filter was found to provide better forecasts compared to the smaller 5-year and 3-year moving average filters.

6.3 Limitations

Many of the watersheds in the western United States are snow-melt driven from accumulated winter snowpack. This study operated at an annual timescale and did not differentiate the precipitation or streamflow that is derived from rainfall or snowfall.

Several limitations of the study were brought about by the use of proxy reconstructions. Proxies such as tree rings are limited to localized areas, where a large number of samples can be cross-dated. Trees must also be free of other influences which may hinder a precipitation signal, such as ground water, location in a flood plain, etc. This makes it difficult to develop streamflow and precipitation reconstructions for all areas across the United States. Reconstructions for oceanic-atmospheric oscillations can be developed from multiple proxies including tree-rings, coral, and ice cores. High-quality and precise reconstructions for the oscillation indices can be developed from coral and ice cores, but are problematic for developing very long multi-century time series (Woodhouse et al., 2011). Tree rings can develop hydrologic reconstructions at millennial scales, but the full series is not included in the current study because a complimentary set of oscillation indices are only available up to 300 years. Many reconstructions are based on conservative regression techniques, which capture more low-frequency climate variability. There is generally a certain degree of uncertainty involved with reconstructions because they only capture 50% - 80% of the variance that is observed in the instrumental record, but they are able to identify events such as the shifts in oscillation phases. It is very difficult to replicate the instrumental record due to other influences that are not accounted for in the reconstruction techniques.

As a result of the uncertainty and noise in the reconstructions, this study used a large moving average filter to identify and extract the long-term climate fluctuations in the dataset. The use of the moving average filter makes the model unsuitable for operational purposes. The model loses the ability to forecast extreme events, floods and droughts at expense of forecasting at longer lead times.

This study operated on the fundamental notion that the past is indicative of the future. Although climate change may be partially explained by climate variability, anthropogenic influences may result in climate changes that have not been observed in the instrumental record or examined in the reconstructed timeline. Furthermore, reconstruction proxies such as tree-rings have shown a change in response to climate variability as a result of the increased CO2 in the atmosphere (Woodhouse et al., 2011), which have to be taken into account in recent reconstructions.

6.4 Recommendations

The data-driven models presented in this study were able to produce 'satisfactory,' 'good,' and 'very good' forecasts based upon relationships between oceanic-atmospheric oscillations and such hydrologic variables as streamflow and precipitation. The quality of these forecasts can be improved as hydroclimatological research expands to implement new strategies and explore new theories in order to further explore the dynamics of climate variability and climate change. As a result of this study, future hydroclimatic explorations can be made:

1) This study used data-driven models as an alternative to physically-based models in order to explore relationships between oceanic-atmospheric oscillations and hydrologic responses in streamflow and precipitation. Future research should explore the physical properties and dynamics of these atmospheric phenomena that drive the changes in the hydrologic cycle. A physically-based model could then be developed based on the properties.
2) Data-driven modeling is a growing field that continues to see improvement over existing models. New methods and models could improve upon current long-

range forecasts. Categorical data-driven models could also be developed to provide qualitative forecasts. Emerging models can potentially find new insights into the effects of climate variability to aid water managers and help in the understanding of climate dynamics.

3) A further exploration is recommended into the development of paleoclimate reconstructions using more advanced methods designed to reduce uncertainty. Most reconstructions are developed through parametric approaches, mostly regression. Nonparametric approaches to developing reconstructions may aid in capturing the variability.
4) This study explores the oceanic-atmospheric oscillation indices of ENSO, PDO, AMO, and NAO. These are the four most commonly studied indices that affect streamflow and precipitation in the western United States. They are also the only indices which have been reconstructed. It is recommended that further research into other oscillation indices, such as the Pacific North American Pattern, West Pacific Pattern, Arctic Oscillation, Madden-Julian Oscillation, etc. be explored for their effects on hydrology in the western United States. Reconstructions of these oscillations should also be considered if they are found to have a significant hydroclimatic influence.

References

Ahmad, M.M., Ghumman, A.R., Ahmad, S., & Hashmi, H.N. (2009). Estimation of Clark's Instantaneous Unit Hydrograph Parameters and Development of Direct Surface Runoff Hydrograph. *Water Resources Management, 23*, 2417-2435.

Ahmad, M.M., Ghumman, A.R., Ahmad, S., &Hashmi, H.N. (2010). Estimation of a Unique Pair of Nash Model Parameters: An Optimization Approach. *Water Resources Management, 24(12)*, 2971-2989.

Ahmad, S., & Prashar, D. (2010). Evaluating Municipal Water Conservation Policies Using a Dynamic Simulation Model. *Water Resources Management, 24(13)*, 3371-3395.

Ahmad, S., & Simonovic, S. P. (2000). Modeling reservoir operations using system dynamics. *Journal of Computing in Civil Engineering, 14(3)*, 190-198.

Ahmad, S., & Simonovic, S. P. (2001). Integration of heuristic knowledge with analytical tools for selection of flood control measures. *Canadian Journal of Civil Engineering, 28(2)*, 208-221.

Ahmad, S., & Simonovic, S. P. (2006). An intelligent decision support system for management of floods. *Water Resources Management, 20(3)*, 391-410.

Ahmad, S., & Simonovic, S.P. (2005). An Artificial Neural Network model for generating hydrograph from hydro-meteorological parameters. *Journal of Hydrology, 315(1-4)*, 236-251.

Ahmad, S., Kalra, A., & Stephen, H. (2010). Estimating Soil Moisture using Remote Sensing Data: A Machine Learning Approach. *Advanced Water Resources, 33(1)*, 69-80.

Ahrens, C. D. (2007). *Meteorology Today: An Introduction to Weather, Climate, and the Environment*. 8th ed. Thomson Brooks/Cole, Canada.

Anderson, M. T., & Woosley, L. H. (2005). Water availability for the western United States - Key scientific challenges. *U.S. Geological Survey Circular 1261*. 85 p.

Asefa, T., Kemblowski, M. W, Urroz, G., McKee, M., & Khalil, A. (2004). Support vectors–based groundwater head observation networks design. *Water Resources Research, 40*, W11509. doi:10.1029/2004WR003304

Asefa, T., Kemblowski, M. W., McKee, M., & Khalil, A. (2006). Multi-timescale stream flow predictions: The support vector machines approach. *Journal of Hydrology, 318*, 7-16. doi:10.1016/j.jhydrol.2005.06.001.

Australian Government Bureau of Meteorology (2010). S.O.I. (Southern Oscillation Index) Archives – 1876 to present. Commonwealth of Australia 2010, Bureau of Meteorology. Retrieved from http://www.bom.gov.au/climate/current/soihtm1.shtml

Barlow, M., Nigam, S., & Berbery E. H. (2001). ENSO, Pacific decadal variability, and U.S. summertime precipitation, drought, and stream flow. *Journal of Climate, 14(9)*, 2105-28.

Barnett, F. A., Gray, S. T. & Tootle, G. A. (2010). Upper Green River Basin (United States) Streamflow Reconstructions. *Journal of Hydrologic Engineering*, 15, 56.

Beebee, R.A., & Manga, M. (2004). Variation in the Relationship between Snowmelt Runoff in Oregon and ENSO and PDO. *Journal of the American Water Resources Association, 40(4),* 1011-1024. doi:10.1111/j.1752-1688.2004.tb01063.x

Biondi, F., Gershunov, A., & Cayan, D. R. (2001). North Pacific Decadal Climate Variability since 1661. *Journal of Climate, 14(1),* 5-10.

Brito-Castillo, L., Di'az-Castro, S., Salinas-Zavala, C. A., & Douglas, A. V., 2003. Reconstruction of long-term winter streamflow in the Gulf of California continental watershed. *Journal of Hydrology, 278(1-4),* 39-50. doi:10.1016/S0022-1694(03)00131-8.

Brown, D. P., & Comrie, A. C. (2004). A winter precipitation 'dipole' in the western United States associated with multidecadal ENSO variability. *Geophysical Research Letters, 31(9),* L09203, 1-4.

Casey, T. M. (1998). Assessment of a seasonal forecast model. *Australian Meteorological Magazine, 47,* 103-111.

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

Cayan, D. R., Dettinger, M. D., Diaz, H. F., & Graham, N. E. (1998). Decadal Variability of Precipitation over Western North America. *Journal of Climate, 11(12),* 3148-3166. doi:10.1175/1520-0442(1998)011<3148:DVOPOW>2.0.CO;2

Cayan, D. R., Dettinger, M. D., Redmond, R. T., McCabe, G. J., Knowles, N., & Peterson, D. H. (2003). The transboundary setting of California's water and hydropower systems, linkages between the Sierra Nevada, Columbia, and Colorado hydroclimates. *Climate and Water: Transboundary Challenges in the Americas,* H. Diaz and B. Morehouse, Eds., Springer, New York, Ch. 10.

Cayan, D. R., Dettinger. M. D., Diaz, H. F., & Graham, N. E. (1998). Decadal Variability of Precipitation over Western North America. *Journal of Climate, 11(12),* 3148-66.

Cayan, D. R., Redmond, K. T., & Riddle, L. G. (1999). ENSO and hydrologic extremes in the western United States. *Journal of Climate, 12(9),* 2881-93.

Cayan, D. R., Redmond, K. T., & Riddle, L. G. (1999). ENSO and Hydrologic Extremes in the Western United States. *Journal of Climate, 12,* 2881-2893. doi:10.1175/1520-0442(1999)012<2881:EAHEIT>2.0.CO;2

Chang, F. J., & Chen, Y. C., 2001. A counterpropagation fuzzy-neural network modeling approach to real time streamflow prediction. *Journal of Hydrology, 245(1-4),* 153-64. doi:10.1016/S0022-1694(01)00350-X

Chowdhury, S., & Sharma, A. (2009). Multisite seasonal forecast of arid river flows using a dynamic model combination approach. *Water Resources Research, 45,* W10428. doi:10.1029/2008WR007510

Christensen, N. S., Wood, A. W., Voisin, N., Lettenmaier, D. P., & Palmer, R. N., (2004). The Effects of Climate Change on the Hydrology and Water Resources of the Colorado River Basin. *Climatic Change, 62(1),* 337-63. doi:10.1023/B:CLIM.0000013684.13621.1f

Clark, M. P., Serreze, M. C., & McCabe, G. J. (2001). Historical effects of El Nino and La Nina events on the seasonal evolution of the mountain snowpack in the Columbia and Colorado River Basins. *Water Resources Research, 37(3)*, 741-57.

Cleary, J. G., & Trigg, L. E. (1995). K*- An instance-based learner using an entropic distance measure. *Proceedings of the 12th International Conference on Machine Learning*. 108-14. doi:10.1.1.51.4098

Climate Impacts Group (2011). About Pacific Northwest Climate. Retrieved from http://cses.washington.edu/cig/pnwc/pnwc.shtml#figure2

Cody, B. A., & Hughes, H. S. (2003). Water resource issues in the 108th Congress: Congressional Research Service Report to Congress, RS20569. Retrieved from http://www.ncseonline.org/nle/crsreports/03Feb/RS20569.pdf

Cook, E. R. (1992). Using tree rings to study past El Niño/Southern Oscillation influences on climate. In: Diaz, H. F & Markgraf, V. (Eds.). *El Niño: Historical and Paleoclimatic Aspects of the Southern Oscillation.* Cambridge University Press, pp. 203-14.

Cook, E. R., & Jacoby, G. C. (1983). Potomac River Streamflow since 1730 as Reconstructed by Tree Rings. Journal of Climate and Applied Meteorology. 22 (10), 1659-72. doi:10.1175/1520-0450(1983)022<1659:PRSSAR>2.0.CO;2

Coulibaly, P., Anctil, F., Arsmussen, P., & Bobée, B. (2000). A recurrent neural networks approach using indices of low-frequency climatic variability to forecast regional annual runoff. *Hydrologic Processes, 14(15)*, 2755-77. doi:10.1002/1099-1085(20001030)14:15<2755::AID-HYP90>3.0.CO;2-9

Cristianini, N., & Shawe-Taylor, J. (2000). An introduction to support vector machines: and other kernel-based learning methods. Cambridge University Press, Cambridge, United Kingdom.

CRWUA (2007). The Colorado River Water Users Association. The Colorado River Water Users Association. Retrieved from http://www.crwua.org

Currie, R. G. (1996). Variance contribution of luni-solar (M_n) and solar cycle (S_c) signals to climate data. *International Journal of Climatology, 16(12)*, 1343-1364.

D'Arrigo, R., Villalba, R., & Wiles, G. (2001). Tree-ring estimates of Pacific decadal climate variability. *Climate Dynamics, 18(3-4)*, 219-24.

De Jongh, I. L. M, Verhoest, N. E. C., & De Troch, F. P. (2006). Ananlysis of a 105-year time series of precipitation observed at Uccle, Belgium. *International Journal of Climatology, 26*, 2023-2039.

Dettinger, M. D., Cayan, D. R., Diaz, H. F., & Meko, D. M. (1998). North-South precipitation patterns in western North America on interannual-to-decadal timescales. *Journal of Climate, 11(12)*, 3095-111.

Dibike, Y. B., Velickov, S., Solomatine, D., & Abbott, M. B. (2001). Model induction with support vector machines: Introduction and application. *Journal of Computing in Civil Engineering, 15(3),* 208-216. doi:10.1061/(ASCE)0887-3801(2001)15:3(208)

Elshorbagy, A., Corzo, G., Srinivasulu, S., & Solomatine, D. P. (2010). Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology - Part 2. *Applications in Hydrologic Earth Systems Science, 14(10),*1943-61.

Enfield, D. B., Mestas-Nunez, A. M., & Trimble, P. J. (2001). The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental U.S. *Geophysical Research Letters, 28(10),* 2077-2080. doi:10.1029/2000GL012745

ESRL (2011). Create a monthly or seasonal time series of climate variables. *Earth System Research Laboratory*. Retrieved from http://www.esrl.noaa.gov/psd/data/

Flaschka, I., Stockton, C. W., & Boggess, W. R. (1987). Climatic variation and surface water resources in the Great Basin Region. *Water Resources Bulletin, 23(1),* 47-57.

Forsee, W., & Ahmad S. (2011). Evaluating Urban Stormwater Infrastructure Design in Response to Projected Climate Change. *Journal of Hydrologic Engineering,* (in press).

Franz, K. J., Hartmann, H. C., Sorooshian, S., & Bales, R. (2003). Verification of National Weather Service Ensemble Streamflow Predictions for Water Supply Forecasting in the Colorado River Basin. *Journal of Hydrometeorology, 4,* 1105–1118. doi:10.1175/1525-7541(2003)004<1105:VONWSE>2.0.CO;2

Gangopadhyay, S., Harding, B. J., Rajagopalan, B., Lukas, J. J., & Fulp, T. J. (2009). A nonparametric approach for paleohydrologic reconstruction of annual streamflow ensembles. *Water Resources Research, 48,* W06417. doi:10.1029/2008WR007201

Garbrecht, J. D., & Rossel, F. E. (2002). Decade-Scale Precipitation Increase in Great Plains at End of 20th Century. *Journal of Hydrologic Engineering, 7(1),* 64-75.

Garfin, G. M., & Hughes, M. K. (1996). Eastern Oregon divisional precipitation and Palmer drought severity index from tree-rings. Final Report to U.S. Forest Service, Cooperative Agreement. PNW-90-174. Laboratory of Tree-Ring Research, University of Arizona. Tucson, Arizona, USA.

Gedalof, Z., & Smith, D. J. (2001). Interdecadal climate variability and regime-scale shifts in Pacific North America. *Geophysical Research Letters, 28(8),* 1515-8.

Georgakakos, K. P., & Hudlow, M. D. (1984). Quantitative Precipitation Forecast Techniques for Use in Hydrologic Forecasting. *Bulletin of the American Meteorological Society, 65(11),* 1186-200.

Gershunov, A., & Barnett, T. P. (1998). Interdecadal modulation of ENSO teleconnections, *Bulletin of the American Meteorological Society, 79(12),* 2715–2725.

Gill, M. K., Asefa, T., Kemblowski, M., & McKee, M. (2006). Soil moisture prediction using support vector machines. *Journal of the American Water Resources Association, 42(4),* 1033–1046. doi:10.1111/j.1752-1688.2006.tb04512.x

Goodrich, G. B. (2007). Influence of the Pacific Decadal Oscillation on Winter Precipitation and Drought during Years of Neutral ENSO in the Western United States. *Weather Forecasting. 22(1),* 116-124. doi:10.1175/WAF983.1

Gou, X., Chen, F., Cook, E., Jacoby, G., Yang, M., & Li, J. (2007). Streamflow variations of the Yellow River over the past 593 years in western China reconstructed from tree rings. *Water Resources Research, 43,* W06434. doi:10.1029/2006WR005705

Graumlich, L. J., Pisaric, M. F. J., Waggoner, L. A., Littell, J. S., & King, J. C. (2003). Upper Yellowstone River Flow and Teleconnections with Pacific Basin Climate Variability during the Past Three Centuries. *Climatic Change. 59(1-2),* 245-262. doi:10.1023/A:1024474627079

Gray, S. T., Betancourt, J. L., Fastie, C. L., & Jackson, S. T. (2003). Patterns and sources of multidecadal oscillations in drought-sensitive tree-ring records from the central and southern Rocky Mountains. *Geophysical Research Letters, 30(6),* 1316. doi:10.1029/2002GL016154

Gray, S. T., Fastie, C. L., Jackson, S. T., & Betancourt, J. L. (2004a). Tree-Ring-Based Reconstruction of Precipitation in the Bighorn Basin, Wyoming, since 1260 a.d. *Journal of Climate, 17(19),* 3855–3865. doi:10.1175/1520-0442(2004)017<3855:TROPIT>2.0.CO;2

Gray, S. T., Graumlich, L. J., Betancourt, J. L., & Pederson, G. T (2004b). A tree-ring based reconstruction of the Atlantic Multidecadal Oscillation since 1567 A.D. *Geophysical Research Letters, 31,* L12205. doi:10.1029/2004GL019932

Grissino-Mayer, H. (1996). A 2129-year reconstruction of precipitation for northwestern New Mexico, USA. pp. 191-204. In: Dean, J. S., Meko, D. M., & Swetnam, T. W. (Eds.). *Tree Rings, Environment and Humanity: Proceedings of the International Conference, Tucson, AZ, May 17-21, 1994.* Radiocarbon, Tucson, AZ.

Grissino-Mayer, H. D., Baisan, C. H., & Swetnam, T. W. (1998). A Multicentury Reconstruction of Precipitation for Great Sand Dunes National Monument, Southwestern Colorado. Mid-continent Ecological Science Center.

Gutierrez, F., & Dracup, J. A. (2001). An analysis of the feasibility of long-range streamflow forecasting for Columbia using El Nino-Southern Oscillation indicators. *Journal of Hydrology, 246(1-4),* 181-196. doi:10.1016/S0022-1694(01)00373-0

Guttman, N. B. & Quayle, R. G. (1996). A historical perspective of U.S. climate divisions. *Bulletin of the American Meteorological Society, 77(2),* 293-303.

Gutzler, D. S., Kann, D. M., & Thornbrugh, C. (2002). Modulation of ENSO-based long-lead outlooks of Southwestern U.S. winter precipitation by the Pacific Decadal Oscillation. *Weather Forecast,17(6),* 1163-72. doi:10.1175/1520-0434(2002)017<1163:MOEBLL>2.0.CO;2

Hamlet, A. F., & Lettenmaier, D. P. (1999). Columbia River streamflow forecasting based on ENSO and PDO climate signals. *Journal of Water Resources Planning and Management, 125(6),* 333-341. doi:10.1061/(ASCE)0733-9496(1999)125:6(333)

Hidalgo, H. G. (2004). Climate precursors of multidecadal drought variability in the western United States. *Water Resources Research, 40,* W12504, doi:10.1029/2004WR003350

Hidalgo, H. G., & Dracup, J. A. (2003). ENSO and PDO Effects on Hydroclimatic Variations of the Upper Colorado River Basin. *Journal of Hydrometeorology, 4(1),* 5-23. doi:10.1175/1525-7541(2003)004<0005:EAPEOH>2.0.CO;2

Hughes, M. K., & Graumlich, L. J. (1996). Climatic variations and forcing mechanisms of the last 2000 years. In: *Multi-millennial dendroclimatic studies from the western United States*. NATO ASI Series, 141, pp. 109-124.

Hunter, T., Tootle, G., & Piechota, T. (2006). Oceanic-atmospheric variability and western U.S. snowfall. *Geophysical Research Letters, 33,* L13706. doi:10.1029/2006GL026600

Hurrell, J. W. (1995). Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science, 269(5224),* 676-9.

Hurrell, J. W. (2011). Monthly Station Based NAO Index. National Center for Atmospheric Research. Retrieved from: http://www.cgd.ucar.edu/cas/jhurrell/indices.data.html#naostatmon

Hurrell, J. W., & Van Loon, H. (1997). Decadal variations in climate associated with the North Atlantic oscillation. *Climatic Change, 36(3-4),* 301-326. doi:10.1023/A:1005314315270

Jain, S., Hoerling, M., & Eischeid, J. (2005). Decreasing reliability and increasing synchroneity of Western North American streamflow. *Journal of Climate, 18(5),* 613–618.

JISAO (2010). PDO Index. Joint Institute for the Study of the Atmosphere and Ocean. Retrieved from http://jisao.washington.edu/pdo/PDO.latest

Jones, P. D., & Mann, M. E. (2004). Climate over past millennia. *Reviews of Geophysics, 42,* RG2002. doi:10.1029/2003RG000143

Kahya, E., & Dracup, J.A. (1993). U.S. streamflow patterns in relation to the El Niño/Southern Oscillation. *Water Resources Research, 29(8),* 2491–2503. doi:10.1029/93WR00744

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

Kalra, A., Ahmad, S. (2011). Evaluating changes and estimating seasonal precipitation for Colorado River Basin using stochastic non-parametric disaggregation technique. *Water Resources Research, 47,* W05555. doi:10.1029/2010WR009675

Kerr, R. A. (2000). A North Atlantic climate pacemaker for the centuries. *Science, 288(5473),* 1984-6. doi:10.1126/science.288.5473.1984

Kim, T., Yoo, C., & Ahn, J. (2008). Influence of climate variation on seasonal precipitation in the Colorado River Basin. *Stochastic Environmental Resources Risk Assessment, 22(3),* 411-20.

Knight, J. R., Folland, C. K., & Scaife, A. A. (2006). Climate Impacts of the Atlantic Multidecadal Oscillation. *Geophysical Research Letters, 33,* L17706, doi:10.1029/2006GL026242

Lall, U., & Sharma, A. (1996). A Nearest Neighbor Bootstrap for Resampling Hydrologic Time Series. *Water Resources Research, 32(3),* 679-693. doi:10.1029/95WR02966.

Lara, A., Villalba, R., & Urrutia R. (2007). A 400-year tree-ring record of the Puelo River summer-fall streamflow in the Valdivian Rainforest eco-region, Chile. *Climatic Change, 86(3-4),* 331-56. doi:10.1007/s10584-007-9287-7

Lee, S., Klein, A., & Over, T. (2004). Effects of the El Niño-southern oscillation on temperature, precipitation, snow water equivalent and resulting streamflow in the Upper Rio Grande river basin. *Hydrologic Processes, 18(6),* 1053-71.

Legates, D. R., & McCabe, G. J. (1999). Evaluating the use of 'goodness-of-fit' measures in hydrologic and hydroclimatic model validation. *Water Resources Research, 35(1),* 233-41.

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

Lins, H. F., & Slack, J. R. (1999). Streamflow trends in the United States. *Geophysical Research Letters, 26(2),* 227-230. doi:10.1029/1998GL900291

Luterbacher, J., Xoplaki, E., Dietrich, D., Jones, P. D., Davies, T. D., Portis, D., Gonzalez-Rouco, J. F., Von Storch, H., Gyalistras, D., Casty, C., & Wanner, H. (2001). Extending North Atlantic Oscillation reconstructions back to 1500. *Atmospheric Science Letters, 2(1-4),* 129-39.

MacDonald, G. M., & Case, R. A. (2005). Variations in the Pacific Decadal Oscillation over the past millennium. *Geophysical Research Letters, 32(8),*1-4.

Mann, M. E., Gille, E., Bradley, R. S., Hughes, M. K., Overpeck, J. T., Keimig, F. T., & Gross, W. (2000). Global temperature patterns in past centuries: An interactive presentation. *Earth Interactions, 4(4),*1-29.

Mantua, N. J. (1999). The Pacific Decadal Oscillation and Climate Forecasting for North America. Retrieved from http://www.atmos.washington.edu/~mantua/REPORTS/PDO/PDO_cs.htm

Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M., & Francis, R. C. (1997). A Pacific Interdecadal Climate Oscillation with Impacts on Salmon Production. *Bulletin of the American Meteorological Society, 78,* 1069–1079. doi:10.1175/1520-0477(1997)078%3C1069:APICOW%3E2.0.CO;2

Margolis, E. Q., Meko, D. M., & Touchan, R. (2011). A tree-ring reconstruction of streamflow in the Santa Fe River, New Mexico. *Journal of Hydrology, 397(1-2),* 118-127.

McCabe, G. J., & Dettinger, M. D. (1999). Decadal variations in the strength of ENSO teleconnections with precipitation in the western United States. *International Journal of Climatology, 19(13),* 1399-410.

McCabe, G. J., Betancourt, J. L., & Hidalgo, H. G. (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. doi:10.1111⁄j.1752-1688.2007.00015.x

McCabe, G. J., Palecki, M. A., & Betancourt, J. L. (2004). Pacific and Atlantic Ocean influences on multidecadal drought frequency in the United States. *PNAS, 101(12),* 4136-4141. doi:10.1073/pnas.0306738101

McEnery, J., Ingram, J., Duan, Q., Adams, T., & Anderson, L. (2005). NOAA's advanced hydrologic prediction service: building pathways for better science in water forecasting. *Bulletin of the American Meteorological Society, 86(3),* 375-385. doi:10.1175/BAMS-86-3-375

Meko, D. M., & Hirschboeck, K. K. (2008). The Current Drought in Context: A Tree-Ring Based Evaluation of Water Supply Variability for the Salt-Verde River Basin. The University of Arizona, The Laboratory of Tree-Ring Research. LTRR-SRP II.

Retrieved from http://fp.arizona.edu/kkh/SRP/SRP2/SRP-II-Final-Final-Report-08-08-08.pdf

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

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

Melesse, A. M., Ahmad, S., McClain, M., Wang, X., & Lim, H. (2011). Suspended Sediment Load Prediction of River Systems: An Artificial Neural Networks Approach. *Agricultural Water Management, 98(5)*: 855-866.

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., & Stouffer, R. J. (2008). Stationarity is Dead: Whither Water Management? *Science, 319*, 573-574. doi:10.1126/science.1151915

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

Mosquera-Machado, S. C., & Ahmad, S. (2007). Flood hazard assessment of Atrato River in Colombia. *Water Resources Management, 21*, 591-609.

NCDC (2011). World Data Center for Paleoclimatology – Climate Reconstructions. National Climatic Data Center. Retrieved from http://www.ncdc.noaa.gov/paleo/recons.html

Ni, F., Cavazos, T., Hughes, M. K., Comrie, A. C., & Funkhouser, G. (2002). Cool-season precipitation in the southwestern USA since AD 1000: Comparison of linear and nonlinear techniques for reconstruction. *International Journal of Climatology, 22(13)*, 1645-62.

NWS (2008). Long-Lead Forecast Tool Discussion and Analysis. National Weather Service. Retrieved from http://www.cpc.ncep.noaa.gov/products/predictions/long_range/tools.html

Pagano, T., & Garen, D. (2005). A recent increase in western US streamflow variability and persistence. *Journal of Hydrometeorology, 6(2)*, 173– 179.

Partal, T., & Kişi, O. (2007). Wavelet and neuro-fuzzy conjunction model for precipitation forecasting. *Journal of Hydrology, 342(1-2)*, 199-212.

Pekarova, P., Miklanek, P., & Pekar, J. (2003). Spatial and temporal runoff oscillation analysis of the main rivers of the world during the 19th-20th centuries. *Journal of Hydrology, 274(1-4)*, 62-79.

Piechota, T. C., & Dracup, J. A. (1999). Long-Range Streamflow Forecasting Using El Niño-Southern Oscillation Indicators. *Journal of Hydrologic Engineer, 4*, 144. doi:10.1061/(ASCE)1084-0699(1999)4:2(144)

Piechota, T. C., Dracup, J. A., & Fovell, R. G. (1997). Western US streamflow and atmospheric circulation patterns during El Nino-Southern Oscillation. *Journal of Hydrology, 201(1-4)*, 249-271. doi:10.1016/S0022-1694(97)00043-7

Piechota, T. C., Timilsena, J., Tootle, G. A., & Gugo, H. (2004). The western U.S. drought: How bad is it? *Eos Trans. AGU. 85(32)*, 301-308. doi:10.1029/2004EO320001

Potts, J. M., Folland, C. K., Jolliffe, I. T., & Sexton, D. (1996). Revised 'LEPS' Scores for Assessing Climate Model Simulations and Long-Range Forecasts. *Journal of Climate, 9(1),* 34-53.

Powell, J. W. (1878). Lands of the arid regions of the United States. Congress of the United States (3rd Session) in the House of Representatives.

Prairie, J., Nowak, K., Rajagopalan, B., Lall, U., & Fulp, T. (2008). A stochastic nonparametric approach for streamflow generation combining observational and paleoreconstructed data. *Water Resources Research, 44(6),* W06423.

Probst, J. L., & Tardy, Y. (1987). Long range streamflow and world continental runoff fluctuations since the beginning of this century. *Journal of Hydrology, 94(3-4),* 289-311.

Puri, S., Stephen, H., & Ahmad, S. (2011). Relating TRMM Precipitation Radar Land Surface Backscatter Response to Soil Moisture in the Southern United States. *Journal Hydrology, 402,* 115-125.

Qaiser, K., Ahmad, S., Johnson, W., & Batista, J. (2011). Evaluating the impact of water conservation on fate of outdoor water use: A study in an arid region. *Journal of Environmental Management, 92(8),* 2061-2068.

Reclamation (2011). *SECURE Water Act Section 9503(c) – Reclamation Climate Change and Water, Report to Congress.*

Redmond, K. T., & Koch, R. W. (1991). Surface climate and streamflow variability in the western United States and their relationship to large-scale circulation indices. *Water Resources Research, 27(9),* 2381-99.

Riehl, H., & Meitin, J. (1979). Discharge of the Nile River: a barometer of short period climate variation. *Science, 206(4423),* 1178-1179.

Riehl, H., El-Bakry, M., & Meitin, J. (1979). Nile River discharge. *Monthly Weather Review, 107(11),* 1546-1553.

Ropelewski, C. F., & Halpert, M. S. (1986). North American precipitation and temperature patterns associated with the El Nino/Southern Oscillation (ENSO). *Monthly Weather Review, 114(12),* 2352-62. doi:10.1175/1520-0493(1986)114<2352:NAPATP>2.0.CO;2

Salzer, M. W., & Kipfmueller, K. F. (2005). Reconstructed temperature and precipitation on a millennial timescale from tree-rings in the southern Colorado Plateau, U.S.A. *Climate Change, 70(3),* 465-87.

Scholkopf, B., & Smola, A. J. (2002). Learning with Kernels - Support Vector Machines, Regularization, Optimization, and Beyond. Massachusetts Institute of Technology

Seaber, P. R., Kapinos, F. P., Knapp, G. L. (1987). Hydrologic Unit Maps (USA). US Geological Survey Water-Supply Paper. 2294.

Shen, C., Wang, W., Gong, W., & Hao, Z. (2006). A Pacific Decadal Oscillation record since 1470 AD reconstructed from proxy data of summer rainfall over eastern China. *Geophysical Research Letters, 33(3),* L03702. doi:10.1029/2005GL024804

Sheppard, P. R., Comrie, A. C., Packin, G. D., Angersbach, K., and Hughes, M. K (2002). The climate of the Southwest, U.S. *Climate Research, 21(3),* 219–238.

Shrestha, E., Ahmad, S., Johnson, W., Shrestha, P., & Batista, J. R. (2011). Carbon Footprint of Water Conveyance versus Desalination as Alternatives to Expand Water Supply. *Desalination*.doi:10.1016/j.desal.2011.06.062

Silverman, D., & Dracup, J. A. (2000). Artificial neural networks and long-range precipitation prediction in California. *Journal of Applied Meteorology, 39(1),* 57-66.

Simonovic, S. P., & S. Ahmad, (2005). Computer-based model for flood evacuation emergency planning. *Natural Hazards, 34(1),* 25-51.

Smith, L. P., & Stockton, C. W. (1981). Reconstructed Stream Flow for the Salt and Verde Rivers From Tree-Ring Data. *Journal of the American Water Resources Association, 17(6),* 1752-1688. doi:10.1111/j.1752-1688.1981.tb01925.x

Solomatine, D. P., & Dulal, K. N. (2003). Model trees as an alternative to neural networks in rainfall-runoff modeling. *Hydrologic Sciences Journal, 48(3),* 399-411.

Solomatine, D. P., & Xue, Y. (2004). M5 Model Trees and Neural Networks: Application to Flood Forecasting in the Upper Reach of the Huai River in China. *Journal of Hydrologic Engineering, 9(6),* 491-501. doi:10.1061/(ASCE)1084-0699(2004)9:6(491).

Solomatine, D. P., Maskey, M., & Shrestha, D. L. (2008). Instance-based learning compared to data-driven methods in hydrological forecasting. *Hydrologic Processes, 22,* 275-287.

State of New Mexico Water Quality Control Commission (2006). 2004-2006 State of New Mexico Integrated Clean Water Act Section 303(d) / Section 305(b) report. State of New Mexico.

Stephen, H., Ahmad, S, Piechota, T. C., & Tang, C. (2010). Relating Surface Backscatter Response from TRMM Precipitation Radar to Soil Moisture: Results over a Semi-Arid Region. *Hydrologic Earth Systems Science, 14,* 193-204.

Stockton, C. W., & Jacoby, G. C. (1976). Long-Term Surface-Water Supply and Streamflow Trends in the Upper Colorado River Basin Based On Tree-Ring Analyses. *Lake Powell Research Project Bulletin, 18,* 70 pp., Univ. of Ariz., Tucson.

Terzi, Ö. (2007). Data mining approach for estimation evaporation from free water surface. *Journal of Applied Science, 7(4),* 593-6.

Timilsena, J., Piechota, T., Tootle, G., & Singh, A. (2009). Associations of interdecadal/interannual climate variability and long-term Colorado River basin streamflow. *Journal of Hydrology, 365(3-4),* 289-301. doi:10.1016/j.jhydrol.2008.11.035

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

Trenberth, K. E., & Otto-Bliesner, B. L. (2003). Toward integrated reconstruction of past climates. *Science, 300(5619),* 589-91.

Tripathi, S., Srinivas, V. V., & Nanjundiah, R. S. (2006). Downscaling of precipitation for climate change scenarios: A support vector machine approach. *Journal of Hydrology, 330,* 621– 640. doi:10.1016/j.jhydrol.2006.04.030

United States Bureau of Reclamation (2011). Colorado River Basin Natural Flow and Salt Data. Bureau of Reclamation: Lower Colorado Region. Retrieved from http://www.usbr.gov/lc/region/g4000/NaturalFlow/current.html

United States Census Bureau (2011). Population Estimates. U.S. Census Bureau. Retrieved from http://www.census.gov/popest/eval-estimates/eval-est2010.html

United States Geological Survey (2011). USGS Surface-Water Data for the Nation. USGS Water Resources. Retrieved from http://waterdata.usgs.gov/usa/nwis/sw

USBR (2008). The Law of the River. United States Bureau of Reclamation: Lower Colorado Region. Retrieved from http://www.usbr.gov/lc/region/g1000/lawofrvr.html

USBR (2009). Colorado River Basin Natural Flow and Salt Data. United States Bureau of Reclamation: Lower Colorado Region. Retrieved from http://www.usbr.gov/lc/region/g4000/NaturalFlow/current.html

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

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

Vedwan, N., Ahmad, S., Miralles-Wilhelm, F., Broad, K., Letson, D., & Podesta, G. (2008). Institutional evolution in Lake Okeechobee Management in Florida: Characteristics, Impacts, and Limitations. *Water Resources Management, 22*, 699-718.

Venkatesan, A. K., Ahmad, S., Johnson, W., & Batista, J. R. (2011). Salinity Reduction and Energy Conservation in Direct and Indirect Potable Water Reuse. *Desalination, 272(1-3)*, 120-127.

Villalba, R., Luckman, B. H., Boninsegna, J., D'Arrigo, R. D., Lara, A., Villanueva-Diaz, J., Masiokas, M., Argollo, J., Soliz, C., Le Quesne, Stahle, D., Roig, F., Aravena, J. C., Huges, M. K., Wiles, G., Jacoby, G., Hartsough, P., Wilson, R. J. S., Watson, E., Cook, E. R., Cerano-Paredes, J., Therrell, M., Cleaveland, M., Morales, M. S., Graham, N. E., Moya, J., Pacajes, J., Massacchesi, G., Biondi, F., Urrutia, R., & Pastur, G. M. (2011). Dendroclimatology from Regional to Continental Scales: Understanding Regional Processes to Reconstruct Large-Scale Climatic Variations Across the Western Americas. *Developments in Paleoenvironmental Research, 11*, 175-227.

Visbeck, M. H., Hurrell, J. W., Polvani, L., & Cullen, H. M. (2001). The North Atlantic oscillation: Past, present, and future. *Proceedings of the National Academy of Sciences U.S.A., 98(23)*, 12876-7.

Vorosmarty, C. J., Green, P., Salisbury, J., & Lammers, R. B. (2000). Global Water Resources: Vulnerability from Climate Change and Population Growth. *Science, 289*, 284-288. doi:10.1126/science.289.5477.284

Wang, H., Ting, M., & Ji, M. (1999). Prediction of seasonal mean United States precipitation based on El Niño sea surface temperatures. *Geophysical Research Letters, 26(9)*, 1341-4.

Wang, H., Zheng, H., Simpsons, D., & Azuaje, F. (2006). Machine learning approaches to supporting the identification of photoreceptor-enriched genes based on expression data. *Bioinformatics, 7*, 116.

Wilson, R., Wiles, G., D'Arrigo, R., & Zweck, C. (2007). Cycles and shifts: 1,300 years of multi-decadal temperature variability in the Gulf of Alaska. *Climate Dynamics*, *28*, 425-440. doi:10.1007/s00382-006-0194-9

Witten, I. H., Frank, E., & Hall MA (2011). Data Mining: Practical Machine Learning Tools and Techniques. 3rd Ed. Burlington, MA: Elsevier Inc.

Wood, A. W., Maurer, E. P., Kumar, A., & Lettenmaier, D. P. (2002). Long-range experimental hydrologic forecasting for the eastern United States. *Journal of Geophysical Research, 107(D20)*, 4429. doi:10.1029/2001JD000659.

Woodhouse, C. A., & Lucas, J. J. (2006a). Multi-century tree-ring reconstructions of Colorado streamflow for water resource planning. *Climatic Change*, *78*, 293-315. doi:10.1007/s10584-006-9055-0

Woodhouse, C. A., & Lucas, J. J. (2006b). Drought, Tree Rings and Water Resource Management in Colorado. *Canadian Water Resources Journal, 31(4)*, 297-310.

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

Woodhouse, C., Lukas, J., Griffin, D., Hirshboeck, K., Hartmann, H., Lay, E., Littell, J., Meko, D., & Gray, S. (2011). TreeFlow. Retrieved from http://treeflow.info/

Yu, X., & Liong, S. Y. (2007). Forecasting of hydrologic time series with ridge regression in feature space. *Journal of Hydrology*, *332(3-4)*, 290-302. doi:10.1016/j.jhydrol.2006.07.003

Zeng, Z., Hsieh, W. W., Shabbar, A., & Burrows, W. R. (2011). Seasonal prediction of winter extreme precipitation over Canada by support vector regression. *Hydrologic Earth Systems Science, 15(1)*, 65-74.

VITA

Graduate College
University of Nevada, Las Vegas

Christopher Allen Carrier

Degrees:

Bachelor of Science, Civil and Environmental Engineering, 2009
University of Nevada, Las Vegas

Thesis Title:

Hydroclimatic Forecasting in the Western United States Using Paleoclimate Reconstructions and Data-Driven Models

Thesis Examination Committee:

Chairperson, Sajjad Ahmad, Ph. D., P.E.
Committee Member, Thomas Piechota, Ph. D.
Committee Member, Jacimaria Batista, Ph. D.
Graduate Faculty Representative, Ashok Singh, Ph. D.