Reconstruction of hydrologic and climatic variability in the Colorado River Basin

Janak Timilsena

University of Nevada, Las Vegas

Follow this and additional works at: https://digitalscholarship.unlv.edu/rtds

Repository Citation
https://digitalscholarship.unlv.edu/rtds/2752

This Dissertation is brought to you for free and open access by Digital Scholarship@UNLV. It has been accepted for inclusion in UNLV Retrospective Theses & Dissertations by an authorized administrator of Digital Scholarship@UNLV. For more information, please contact
digitalscholarship@unlv.edu.
RECONSTRUCTION OF HYDROLOGIC AND CLIMATIC VARIABILITY IN THE
COLORADO RIVER BASIN

by

Janak Timilsena

Bachelor of Civil Engineering
University of Mysore, India
1992

Master of Engineering
Asian Institute of Technology, Thailand
1997

A dissertation submitted in partial fulfillment
of the requirements for the

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

Graduate College
University of Nevada, Las Vegas
August 2007

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
The Dissertation prepared by

JANAK TIMILSENA

Entitled

RECONSTRUCTION OF HYDROLOGIC AND CLIMATIC VARIABILITY IN THE
COLORADO RIVER BASIN

is approved in partial fulfillment of the requirements for the degree of

DOCTORATE OF PHILOSOPHY IN CIVIL AND ENVIRONMENTAL ENGINEERING

Examination Committee Chair

Dean of the Graduate College
ABSTRACT

Reconstruction of Hydrologic and Climatic Variability in the Colorado River Basin

by

Janak Timilsena

Dr. Thomas C. Piechota, Examination Committee Chair
Associate Professor, Department of Civil and Environmental Engineering
University of Nevada, Las Vegas

This dissertation investigated the reconstruction of hydrologic and climate variability in the Colorado River Basin for the past 500 years. Unimpaired streamflow stations and regional April 1 snow water equivalent of the basin were reconstructed using tree-ring chronologies as predictors in Partial Least Squares Regression (PLSR). Regional April 1 snow water equivalent time series was developed with principal component analysis and cluster analysis on snow course stations in the basin. For the reconstruction of unimpaired streamflow and regional April 1 snow water equivalent, all available standard tree-ring chronologies inside the Colorado River Basin were screened based on the correlation criterion (>95% significance level). Then, PLSR was run using the cross validation approach (i.e. removing the least correlated screened tree-ring chronologies) to obtain an optimal result. Further, a PLSR reconstruction was compared with the different reconstruction procedures (Stepwise Linear Regression and Stepwise Principal Component Regression) in order to find the preferred method for climate variable...
reconstruction. Finally, the non-parametric rank sum test was performed to evaluate the individual and coupled impact of interannual and interdecadal ocean climate phenomenon on climate variables (streamflow and snowpack) of the basin using extended records.

The major contributions of this research are divided into three major categories. First, unimpaired streamflow stations in the Colorado River basin were reconstructed using the Partial Least Square Regression (PLSR) technique. The current practice of reconstructing streamflows by using multiple linear regression and principal component regression were compared with the performance of Partial Least Square Regression technique (PLSR) based on cross validation standard error. The spatial and temporal variability of drought was evaluated for all the unimpaired streamflow stations and the different centuries in the record. Second, snowpack was regionalized using principal component and cluster analysis which helps to identify coherent regions in the basin. Further, regionalized snowpack was reconstructed for the past 500 years using the PLSR in order to understand the long-term regional spatial and temporal variability of drought in the basin. Third, the teleconnection study between interannual and interdecadal ocean climate phenomenon and climate variables (streamflow and snowpack) in the region was performed using reconstructed data considering different lag years (0, +1, +2 and +3).

The outcome of this research identified an improved reconstruction approach to extend the climate variables (unimpaired streamflow and snowpack), regionalized climate variables, and determined the relationship between reconstructed climate variables and reconstructed large-scale ocean atmospheric patterns. These outcomes will help to improve the current practice of drought management planning and drought forecasting in the western United States.
<table>
<thead>
<tr>
<th>TABLE OF CONTENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT: .......................................................................................................................... iii</td>
</tr>
<tr>
<td>TABLE OF CONTENTS ......................................................................................................... v</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS ........................................................................................................ xi</td>
</tr>
<tr>
<td>CHAPTER 1 INTRODUCTION ................................................................................................. 1</td>
</tr>
<tr>
<td>1.1 Research Problem ........................................................................................................ 1</td>
</tr>
<tr>
<td>1.1.1 Reconstruction of Streamflow Using Tree-Ring Data ........................................ 1</td>
</tr>
<tr>
<td>1.1.2 Regionalization and Reconstruction of Snowpack Using Tree Ring Data ........... 3</td>
</tr>
<tr>
<td>1.1.3 Linkages of Streamflow/Snowpack with Large-Scale Ocean/Atmospheric Phenomenon ................................................................. 4</td>
</tr>
<tr>
<td>1.1.4 The Study Area ........................................................................................................ 6</td>
</tr>
<tr>
<td>1.2 Research Questions and Hypothesis ........................................................................ 7</td>
</tr>
<tr>
<td>1.3 Presentation of this Research .................................................................................. 8</td>
</tr>
<tr>
<td>CHAPTER 2 STATE OF KNOWLEDGE .................................................................................. 10</td>
</tr>
<tr>
<td>2.1 Concept of Dendrochronology .................................................................................. 10</td>
</tr>
<tr>
<td>2.1.1 Fundamentals of Tree-ring Growth ...................................................................... 11</td>
</tr>
<tr>
<td>2.1.2 Tree-ring Sampling Procedure ........................................................................... 12</td>
</tr>
<tr>
<td>2.1.3 Cross-dating of Tree-ring Chronologies .............................................................. 12</td>
</tr>
<tr>
<td>2.1.4 Standardization ...................................................................................................... 15</td>
</tr>
<tr>
<td>2.2 Tree-ring Reconstruction of Climate Variables ....................................................... 16</td>
</tr>
<tr>
<td>2.2.1 Tree-ring Screening for Climate Reconstruction ................................................ 17</td>
</tr>
<tr>
<td>2.2.2 Climate Reconstruction Methods .......................................................................... 20</td>
</tr>
<tr>
<td>2.3 Regionalization of Climate Variables ...................................................................... 29</td>
</tr>
<tr>
<td>2.4 Ocean/Atmosphere Influences on U.S. Climate ......................................................... 31</td>
</tr>
<tr>
<td>2.4.1 El Niño-Southern Oscillation (ENSO) ................................................................ 32</td>
</tr>
<tr>
<td>2.4.2 Pacific Decadal Oscillation (PDO) ..................................................................... 33</td>
</tr>
<tr>
<td>2.4.3 Atlantic Multidecadal Oscillation (AMO) ............................................................ 35</td>
</tr>
<tr>
<td>2.4.4 North Atlantic Oscillation (NAO) ........................................................................ 36</td>
</tr>
<tr>
<td>CHAPTER 3 PARTIAL LEAST SQUARE REGRESSION FOR IMPROVED STREAMFLOW RECONSTRUCTION ................................................. 38</td>
</tr>
<tr>
<td>3.1 Introduction ................................................................................................................ 38</td>
</tr>
<tr>
<td>3.2 Data Sources ................................................................................................................ 41</td>
</tr>
<tr>
<td>3.2.1 Unimpaired Streamflow Data ............................................................................... 41</td>
</tr>
<tr>
<td>3.2.2 Tree-Ring Data ...................................................................................................... 42</td>
</tr>
</tbody>
</table>
3.3 PLSR Reconstruction and Comparison ......................................... 44
  3.3.1 Pre-screening of Predictors ................................................. 44
  3.3.2 Partial Least Square Regression Procedure ....................... 47
  3.3.3 Selection of Best Predictors ................................................ 49
  3.3.4 Comparison of PLSR with the Stepwise Principal
Component Regression and the Stepwise Linear Regression ........................................ 50
  3.3.5 Reconstruction of Unimpaired Streamflow and Drought 51

3.4 Results ............................................................................................... 54
  3.4.1 Performance Comparisons of PLSR, STPCR and STLR 54
  3.4.2 PLSR Reconstruction of 17 Streamflow Stations 57
  3.4.3 Spatial and Temporal Variability of Drought .................... 60

3.5 Conclusions ...................................................................................... 65

CHAPTER 4 REGIONALIZATION AND RECONSTRUCTION OF SNOW WATER EQUIVALENT IN THE UPPER COLORADO RIVER BASIN ................................................................. 67

4.1 Introduction ...................................................................................... 67

4.2 Data Sources .................................................................................... 69
  4.2.1 Snow Course Data ................................................................. 69
  4.2.2 Tree-Ring Data ...................................................................... 71
  4.2.3 Unimpaired Streamflow Data ............................................. 74

4.3 Methods ............................................................................................. 74
  4.3.1 Regionalization .................................................................... 74
  4.3.2 April 1 Snow Water Equivalent Reconstruction ............... 76
  4.3.3 Spatial and Temporal Variability of Drought .................... 81

4.4 Results ............................................................................................... 82
  4.4.1 Regionalization .................................................................... 82
  4.4.2 April 1 Snow Water Equivalent Reconstruction ............... 84
  4.4.3 Regional and Temporal Variability of Drought ................ 88

4.5 Conclusions ....................................................................................... 97

CHAPTER 5 INDIVIDUAL AND COUPLED IMPACT OF OCEANIC CLIMATEPHENOMENON ON COLORADO RIVER STREAMFLOW AND SNOWPACK USING EXTENDED PERIOD OF RECORD ................................................................. 99

5.1 Introduction ...................................................................................... 99

5.2 Data Sources ................................................................................. 102
  5.2.1 Reconstructed Streamflow .................................................. 102
  5.2.2 Reconstructed Regional Snow Water Equivalent (SWE) .............. 104
  5.2.3 Reconstructed Interdecadal and Decadal Oceanic Data 104

5.3 Methodology .................................................................................... 107
  5.3.1 Teleconnection Study of Individual Impact of ENSO, PDO, AMO and NAO Using Reconstructed Longer Period of Record 107
LIST OF FIGURES

Figure 1-1 The Study Area – Upper and Lower Colorado River Basin ....................... 6
Figure 2-1 A schematic view of a skeleton plot (from Nash, 2002).......................... 13
Figure 3-1 Location map displaying 17 unimpaired streamflow stations and 42 standard tree-ring chronologies ............................................................. 42
Figure 3-2 Calibration comparison of PLSR with STPCR and STLR for water year streamflow (100*Km$^3$) volume for three different stations ..................... 56
Figure 3-3 PLSR reconstructed water year streamflow (100*Km$^3$) of (a) Lake Fork at Gateview (b) Virgin River at Little field and (c) Salt River near Roosevelt White River ................................................................. 59
Figure 3-4 Number of drought years for each 100 year period (1500-1599, 1600-1699, 1700-1799, 1800-1899 and 1900-1999) for all 17 unimpaired streamflow stations in the basin ............................................................. 61
Figure 3-5 Average drought deficit (100*Km$^3$/year) for each 100 year period (1500-1599, 1600-1699, 1700-1799, 1800-1899 and 1900-1999) for all 17 unimpaired streamflow stations in the Colorado River basin.............. 62
Figure 4-1 Location map displaying 39 snow course stations and 17 standard tree-ring chronologies ................................................................................... 71
Figure 4-2 Regionalization of snow course stations based on S-mode PCA and average linkage cluster analysis ................................................................. 83
Figure 4-3 Calibration plot for PLSR reconstructed regional composite April 1 SWE for Region 1, 2 and 3 .............................................................................. 86
Figure 4-4 The 3-year, 5-year and 10-year moving average of PLSR reconstructed regional composite April 1 SWE for Region 1, Region 2 and Region 3. ......................................................................................................... 90
Figure 4-5 Drought duration identified based on 3-year moving average for three different regions (R1, R2 and R3) and three different streamflow (S1, S2 and S3). R1, R2 and R3 indicate reconstructed regional composite April 1 SWE of region 1, 2 and 3 respectively. .................................................................................. 93
Figure 4-6 Drought duration identified by the 5-year moving average for three different regions (R1, R2 and R3) and three different streamflow (S1, S2 and S3). R1, R2 and R3 indicate reconstructed regional composite April 1 SWE of region 1, 2 and 3 respectively. .................................................................................... 94
Figure 4-7 Drought duration identified by the 10-year moving average for three different regions (R1, R2 and R3) and three different streamflow (S1, S2 and S3). .......................................................................................................... 96
Figure 5-1 Unimpaired streamflow stations and snowpack region in the Colorado River basin. ......................................................................................................... 103
Figure 5-2  Significant (95%) difference in streamflow medians for (a) ENSO (cold) - ENSO (warm) (b) PDO (cold) - PDO (warm) (c) AMO (cold) -AMO (warm). ................................................................. 115

Figure 5-3  Significant (95%) difference in streamflow medians for (a) El Nino/PDO cold - El Nino/PDO warm (b) La Nina/ PDO cold - La Nina/PDO warm. ................................................................. 121

Figure 5-4  Box and Whisker plots that shows mean, median, and percentile (5th, 25th, 75th and 95th) increase/decrease (in terms of percentage) in streamflow volume (considering all 17 stations) due to individual and coupling effect of oceanic climate phenomenon. ........................................ 127

Figure 5-5  Box and Whisker plots that shows mean, median, and percentile (5th, 25th, 75th and 95th) increase/decrease (in terms of percentage) in streamflow volume (considering all 17 stations) due to individual effect of ENSO, PDO and AMO using historical and reconstructed data respectively ................................................................. 127

Figure 5-6  Box and Whisker plots that shows mean, median, and percentile (5th, 25th, 75th and 95th) increase/decrease (in terms of percentage) in streamflow volume (considering all 12 stations of upper basin only) due to individual and coupling effect of oceanic climate phenomenon .... 130

Figure 5-7  Box shows mean, median, and percentile (lowest and highest) increase/decrease (in terms of percentage) in streamflow volume (considering all 5 stations of lower basin only) due to individual and coupling effect of oceanic climate phenomenon. ........................................ 131
LIST OF TABLES

Table 2-1  Summary of Literature on Drought Variables Reconstruction summary of Literature on Drought Variables Reconstruction ...................................... 21
Table 3-1  Summary of correlation between water year streamflow (predictands) and standard tree-ring chronologies (predictors), pool of predictors ...... 45
Table 3-2  Performance comparison of PLSR with STPCR and STLR for 17 unpaired streamflow stations in the Colorado River basin .................. 55
Table 3-3  Reconstruction performance of 17 unimpaired streamflow stations using PLSR ............................................................................................................ 58
Table 3-4  The most severe drought years based on the 3-year, 5-year and 10-year moving average of last 500 years of water year streamflow for 17 reconstructed unimpaired streamflow stations in the Colorado River basin ............................................................................................................. 64
Table 4-1  Summary of the standard tree-ring chronologies utilized for the regional April 1 SWE reconstruction ................................................................. 73
Table 4-2  Summary of PLSR reconstruction performance statistics .......................... 85
Table 4-3  Drought duration ranking based on Regional Composite April 1 SWE and water year streamflow volume using three different moving average scenarios ....................................................................................................... 89
Table 4-4  Drought characteristics based on different moving average (3-year, 5-year and, 10-year) based on three regional composite April 1 SWE (R1, R2 and R3) and three different water year streamflow volume (S1, S2 and S3) ............................................................................................................. 92
Table 5-1  Individual impact of ocean climate phenomenon on streamflow ...... 113
Table 5-2  Individual impact of ocean climate phenomenon on regional snowpack ............................................................................................................. 114
Table 5-3  Coupling effect of ENSO with PDO, AMO and NAO on streamflow 118
Table 5-4  Coupling effect of ENSO with PDO, AMO and NAO on regional snowpack ............................................................................................................. 120
Table 5-5  Coupling effect of PDO, AMO and NAO on streamflow .................... 123
Table 5-6  Coupling effect of PDO, AMO and NAO on regional snowpack ...... 124
ACKNOWLEDGEMENTS

The author wishes to express his deep appreciation and sincere gratitude to his advisor Dr. Thomas C. Piechota for his sincere guidance and timely suggestions throughout the study period. The author really benefited from weekly progress meeting and follow up suggestion by his advisor and his fellow colleagues of B107. The author would also like to thank to his committee Dr. Jacimaria Batista, Dr. David E. James, Dr. Ashok K. Singh, and graduate faculty representative Dr. Matthew S. Lachniet for their invaluable guidance, support and suggestions through out the study period.

Finally, the help, sacrifice and motivation provided by his parents, wife Ramila and son Sayujya are highly appreciated and author would like to sincerely thank them.
CHAPTER 1

INTRODUCTION

1.1 Research Problem

1.1.1 Reconstruction of Streamflow Using Tree-Ring Data

Drought is one of the most devastating natural hazards faced by the United States (Woodhouse and Overpeck, 1998). Out of the 46 U.S. weather-related disasters between 1980 and 1999 causing in excess of $1 billion in damage each, eight of the events were droughts. This includes the most costly national disaster, the 1988 drought, which alone was responsible for $40 billion of losses (Ross and Lott, 2000).

Currently, drought planning and management is primarily based upon 20th century available historical records of climate and streamflow data. Longer records using paleoclimatic data (e.g., tree rings) are needed to evaluate drought from a long-term perspective and for developing better management practices during drought. Dendrohydrological reconstructions (from tree-rings) can be a useful tool for augmenting existing observed data. Tree-rings are effective proxies for streamflow because trees in selected locations respond to a set of climate-related factors including precipitation and evapotranspiration that also influences streamflow variability (Meko et al., 1995).

The reconstruction of streamflow using tree-ring data has been widely applied to regions around the world, and some of the results are available at the NOAA World Data Center for Paleoclimatology including the stations in U.S. (e.g., Stockton and Jacoby
1976; Cleaveland 2000; Woodhouse 2000; Meko et al., 2001; Woodhouse 2001; Graumlich et al., 2003). The period of record for the reconstructed streamflow data in the U.S. goes as far as back 869 A.D. for the Sacramento River.

There are a variety of methods available to reconstruct hydroclimatic variables from tree-ring data. Stahle et al. (1998), Touchan et al. (1999), Diaz et al. (2001), Woodhouse (2001), and Gray et al. (2004) used Simple/Multiple Linear Regressions and Stepwise Linear Regression to reconstruct hydroclimate variables such as precipitation, streamflow and other drought variables. Similarly, Stockton and Jacoby (1976), Cook and Jacoby (1983), Garen (1992), Meko et al. (2001), Hidalgo et al. (2000) Woodhouse (2001), Brito-Castillo et al. (2003) and Woodhouse et al. (2006) used Principal Component Regression (Simple/Multiple Principal Component Regression, Stepwise Principal Component Regression and Alternate Principal Component Regression) for streamflow reconstruction. Cook et al. (1999) reconstructed a gridded drought index from 1700-1978 using the Point by Point Regression (PPR) technique based on Principal Component Regression.

In Principal Component Regression (PCR), the X-scores (scores of independent variable) are chosen to explain as much of the factor variation as possible; therefore this approach yields informative directions in the factor space, but they may not be associated with the shape of predicted surface.

Partial Least Squares Regression (PLSR) is a method for constructing predictive models when there are many factors that are highly collinear. PLSR uses the principal components of both the predictor (X-scores) and predictand (Y-scores). The X-scores and Y-scores are chosen so that the relationship between successive pairs of scores is as
strong as possible. In principle, this is a robust form of redundancy analysis, seeking
directions in the factor space that are associated with the high variations in the responses.

This research reconstructs the streamflow using the Partial Least Square
Regression (PLSR) technique. Additionally, the current practice of reconstructing
streamflows by using Multiple Linear Regression and Principal Component Regression
are compared with the performance of PLSR.

1.1.2 Regionalization and Reconstruction of Snowpack Using Tree Ring Data

Mountains in the western U.S. hold a vast amount of snowpack which provide
50%-80% of the water supply in the region (NRCS, 2006). Although snowpack is limited
to only 15% of the land area of the Colorado basin (Doesken and Stanton, 1991), most of
the streamflows in the Colorado River are derived from the melting of the snowpack from
April through July (Kuhn, 2005). Snowpack trend and characteristics of this basin are
crucial since it is the major source of water in the region. Existing dense and relatively
short records of snowpack are inadequate for assessing the long-term variations of
drought; hence a longer record of snowpack is required to characterize and to obtain the
drought variability.

Eigen techniques (Principal component, Factor analysis) have been extensively
used in the meteorological community for data reduction, grouping of variables and
identification of coherent modes in atmospheric fields (e.g. Richman, 1986). Piechota et
al. (1997) used the S-mode PCA and cluster analysis to isolate the group of stations that
used rotated principal component analysis to define the snowpack region in the western
Colorado River basin.
The research presented here regionalizes the snowpack using principal component and cluster analysis which helps to identify coherent regions in the basin.

Tree-rings have been successfully used to reconstruct hydroclimatic variables such as precipitation, streamflow and other drought variables (e.g. Haston and Michealson 1997; Cook et al. 1999, Cleaveland 2000; Hidalgo et al. 2000, Meko et al. 2001, Woodhouse 2001, Graumlich et al. 2003 and Gray et al. 2004).

Few studies have reconstructed snowpack data and no studies have utilized PLSR to reconstruct snowpack, the research presented here reconstructs the regionalized snowpack using PLSR utilizing the tree-ring chronologies as the predictors.

1.1.3 Linkages of Streamflow/Snowpack with Large-Scale Ocean/Atmospheric Phenomenon

Natural variability of streamflow with respect to short-term and long-term fluctuations as well as long-term trends is crucial for assessing the vulnerability of water resources to current or future climate changes (Graumlich et al., 2003). A major source of rivers in the western U.S. is snowmelt and variability of snow accumulation is influenced by large-scale atmospheric circulation patterns (Graumlich et al., 2003).

It is also important that ocean-atmospheric variability occurs on interannual, decadal and interdecadal timescales (Tootle et al., 2005). El Niño-Southern Oscillation (ENSO) defines the relationship of the periodic large-scale warming or cooling of the central eastern equatorial Pacific Ocean with the Southern Oscillation, with a periodicity of 2-7 years (Hanley et al., 2003). The Pacific Decadal Oscillation (PDO) is a long-lived ENSO-like pattern of Pacific climate variability (Shen et al., 2006). PDO has significant influence on climate-sensitive natural resources in the Pacific and North America,
including the water supplies and snowpack in selected North American regions (Mantua and Hare, 2002) with a periodicity of about 50 years. The Atlantic Multidecadal Oscillation (AMO) is termed as the leading mode of low-frequency, North Atlantic Ocean (0 to 70 degrees) sea surface temperature variability with a periodicity of 65-80 years (Kerr, 2000; Gray et al. 2004). Similar to PDO, AMO has two phases: positive and negative. During the positive phase, most of the U.S. experiences below-normal rainfall (Enfield et al., 2001) and has increased probability of drought in the western U.S. (McCabe et al., 2004).

The coupled influences of interannual ENSO phenomenon with the decadal phenomenon, PDO, AMO, NAO (North Atlantic Oscillation Index) and SSTs (Sea Surface Temperature) using instrumental climate data of length 50-100 years have been performed successfully (e.g. Gershunov and Barnett, 1998; Barlow et al., 2001; Beebee and Manga, 2004; Sutton and Hudson, 2005; and Tootle et al., 2005). However, the investigation of interdecadal variability requires a longer record since phenomena such as the PDO, AMO and NAO cover only one to two phases of the interdecadal phenomena during the instrumental period; which may not be sufficient length to interpret 20-80 years of periodic phenomena.

This research identifies the individual and coupled impacts of ENSO with PDO, AMO and NAO to streamflow and regional snowpack of the Colorado River basin considering time lags of 0, +1, +2 and +3 years of an extended period of record obtained from reconstructed streamflow and snowpack.
1.1.4 The Study Area

The study area of this research is the Upper and Lower Colorado River basin (Figure 1-1) which is a major source of water supply in the south western U.S. that includes several major cities: Las Vegas; Phoenix; Tucson; San Diego; and Los Angeles. The watershed area of the Colorado River basin (Upper and Lower) is approximately 637,100 km$^2$. It serves about 25 million people with water supply and electricity, and provides irrigation for 14,164 km$^2$ of farmland. The present practice of water management in the Colorado River basin is principally based upon the last 100 years (approximately) of instrumental records (Woodhouse, 2002). Therefore, reconstructed past records of unimpaired streamflow stations and regional snowpack in the region using tree-ring chronologies are important for improved drought management.

Figure 1-1: The Study Area – Upper and Lower Colorado River Basin
1.2 Research Questions and Hypothesis

The overall goal of the research is to develop an improved reconstruction approach to extend the climate variables (unimpaired streamflow and snowpack) by using tree-ring chronologies and to determine the relationship between reconstructed climate variables and reconstructed large-scale ocean atmospheric patterns. This will lead to a better understanding of high and low frequency oceanic-climatic phenomenon and its coupled effect on streamflow and snowpack in the Colorado River basin.

The research questions and hypothesis addressed in this dissertation are as follows.

Research Question #1 - Is it possible to reconstruct the streamflow stations using PLSR and does it improve the result as compared to the current reconstruction approach of principal component regression and multiple linear regressions?

Hypothesis #1 – Hydrologic variable (streamflow) can be reconstructed using PLSR based on cross validation, and it improves the performance as compared to the contemporary methods (multiple linear regression and principal component regression) since it generalizes and combines the features from principal component analysis and multiple linear regressions

Research Question #2 - Can snowpack stations be regionalized by using principal component analysis and cluster analysis, and reconstructed to interpret the hydrologic variability of the Colorado River basin?

Hypothesis #2 – Regionalization and reconstruction of April 1 snow water equivalent of the basin facilitate a better understanding of spatial and temporal distribution of snowpack regimes from a long-term perspective.
Research Question #3 - What are the individual impacts of ENSO, PDO, AMO and NAO on streamflow and snowpack of the Colorado River basin based on the reconstructed longer period of record? Also, what are the coupled impacts of ENSO, PDO, AMO and NAO based on the extended period of record?

Hypothesis #3 – The teleconnection study between oceanic atmospheric phenomenon (ENSO, PDO, AMO, and NAO) and climate variables (unimpaired streamflow and regional snowpack) of the basin based on extended period of record will lead to improved understanding since multiple phases of each phenomena will be captured.

1.3 Presentation of this Research

This dissertation is presented in six chapters. Chapter 2 presents the background information on dendrochronology, tree-ring reconstruction of climate variables, regionalization of climate variables and oceanic/atmosphere influences on U.S. climate. Chapter 3 describes the PLSR methodology of streamflow reconstruction and provides the comparison of PLSR performances with the most common methods (principal component regression and multiple linear regressions) of reconstruction. In addition, drought analysis is performed based on each 100 year epoch for the last 500 years. In this chapter, drought years were defined as the years that have less than 10 percentile of water year flow volume for the last 500 years. Chapter 4 describes the procedures and results of regionalization of snow course stations and PLSR reconstruction of regional April 1 snow water equivalent (SWE) for the Upper Colorado River Basin. In addition to the regionalization and reconstruction, drought analysis based on its duration is also
presented. In this part of the research, a drought event was defined as a year that has a negative anomaly or consecutive years with negative anomalies of climate variables (snowpack and streamflow). Similarly, the year elapsed between the beginnings and the ending year of negative anomalies for each drought event was the drought duration for that particular drought event. Chapter 5 summarizes the impacts on Colorado streamflow and snowpack based on individual and coupled effect of interannual and interdecadal climatic influences (ENSO, PDO, AMO and NAO) using reconstructed data and utilizing lags of 0, +1, +2 and +3 years.
CHAPTER 2

STATE OF KNOWLEDGE

2.1 Concept of Dendrochronology

Dendrochronology is defined as the study of the chronological sequence of annual growth ring properties in trees (Ferguson, 1970). Tree-ring properties include the annual growth size, early wood ring-width, late wood ring-width, early wood density and late wood density; however, the most commonly used feature is annual growth size (ring-width). The physical dimension of annual growth rings depends upon the collective role of all climate variables (precipitation, temperature, soil moisture etc.) during the growing season (Meko, 1995; Woodhouse, 2002). These annual rings have long been recognized as an important source of chronological and climate information (Bradley, 1999) and in certain circumstances, hold information about surrounding climatic conditions (Fritts, 1976).

Dendrochronology includes the set of processes that comprises the understanding of tree-ring growth, sampling techniques, proper assignment of tree-ring properties to a calendar year (cross-dating), and removing the age growth and autocorrelation influences in tree-rings properties between consecutive tree-rings in the tree-ring chronologies (standardization).
2.1.1 Fundamentals of Tree-ring Growth

Tree-rings can be observed in the wood of many species of temperate forest trees throughout the world (Fritts, 1985). In general, trees actively grow during the spring and summer (growth period) and are inactive during the fall and winter (dormant period). In a growth period, a tree produces wood (also defined as xylem). There are two different growth stages during the one full growth period: (a) the early part of the growth tentatively starts in April and ends in July (early wood); and (b) the latter part of the growth starts around July and ends in September (late wood). Early wood grows faster and will have thin cell walls but the late wood will not grow as fast as early wood, creating thicker and denser cell walls. The sum of the two growth stages within a year results in an annual tree-ring width, and these two consecutive stages (early wood and late wood) help to separate one annual tree-ring growth from another.

Generally, softwoods, like Conifers (*Juniperus, Pinus* etc), have a clear contrast in the growth boundary, but in the case of hardwood like Oak (*Quercus*), there is a possibility of a lack of distinct separation between two growth rings (patterns). In the southwestern U.S., the most commonly used softwood tree species for tree-ring analysis are Douglas-fir (*Pseudotsuga menziesii*), Ponderosa pine (*Pinus ponderosa*), Rocky Mountain juniper (*Juniperous scopulorum*) and Giant Sequoia (*Sequoiadendron giganteun*). Even though, Giant Sequoias have a very long record in the southwest and much work have been done using this species, it is not considered a good dendrochronological species (Ferguson, 1970), because complacent (uniform growth) tree-rings are more common in this particular species in this region.
2.1.2 Tree-ring Sampling Procedure

The relationship between tree-ring growth and climate conditions is strongest in areas where trees are climate stressed (Bradley, 1999). According to Ferguson (1970), the main criteria of sampling tree-ring species are: (a) trees should not be too close to each other, as the competition may overshadow the climate response; (b) trees should always be climatically stressed (i.e. no subsurface supply of water); and (c) there should not be any outward appearance of injuries or diseases. Tree-ring growth in a climatically stressed region will have distinctive tree rings, because a small change in climate variables will create a distinct difference in the tree-ring growth pattern. In less climate sensitive sites (e.g. near a perennial water source or water table), the tree will produce complacent rings (uniform growth) that makes it difficult to extract climate signals.

The Swedish Incremental Borer is the most commonly used tool to remove the core from trees. The auger of this borer should be driven in a radial direction perpendicular and towards the pith (vertical axis) of the tree. However, it should not be driven beyond the pith because it makes it difficult to draw back the auger and may damage the sample when drawing back. To extract consistent climate signals, 2-3 sample cores should be taken separately from at least 20 trees from each individual site (Bradley, 1999).

2.1.3 Cross-dating of Tree-ring Chronologies

Cross-dating is a process of matching the tree-ring patterns of sample chronologies from tree to tree in order to obtain the exact calendar year of a tree-ring and its corresponding dimensions. If the shared pattern of interannual variability in the tree-ring properties is strong enough and if a long enough series of rings are available in each
case, it is possible to match the ring pattern of wood of an unknown age against dated wood and thus attribute each growth ring in the former sample to a calendar year (Hughes, 2006). In the procedure of cross-dating, the sampling cores (cross sections) of trees from all the sites are collected, mounted and sanded to a smooth surface. Then the cores from one sampling location are matched to each other, and the composite chronology can be obtained for the particular location. The cross-dating (matching) can be performed using a visual, analog technique developed in the early part of the century by Andrew Ellicott Douglass of the University of Arizona (Nash, 2002), and this technique produces a graphic illustration of ring growth, which is also known as the “skeleton plot” in which the narrow rings are emphasized. A schematic view of a skeleton plot is illustrated in Figure 2-1.

![Skeleton Plot Diagram](image)

Figure 2-1: A schematic view of a skeleton plot (from Nash, 2002)

In the case of skeleton plots, the ring-width is plotted in the vertical axis in inverse proportion (i.e. narrow rings with longer length and wide rings with shorter length) and the year is plotted in the horizontal axis considering the pith as zero. The
plotted vertical length is not a measurement; it only represents the comparative scale (1-10) representation of the ring-width. Once all the skeleton plots are prepared for all the samples of each site, they are stacked vertically in order to obtain the common growth properties for a common period (year). Once the narrow rings are matched with each other, then those rings are considered as the milestone to date the remaining fairly indistinct tree-rings. The composite chronology that represents the chronology for the particular site is prepared by matching extreme narrow tree-rings based on the majority of samples. Once the composite chronology for the particular area is developed and matched with enough overlap with modern trees (within the sample or outside the sample) whose age is known, the calendar year and corresponding growth pattern of the composite chronology for the particular location can be obtained, and this composite chronology is called the master chronology. Further, with this master chronology, it is possible to obtain a calendar year and its corresponding growth indices (normal width, early density and late density) of each sample (within the site or outside the site) by matching with the master chronology (already cross-dated).

Using the statistical software COFECHA developed by Holmes (1983), it is possible to check the cross-dating and improve the quality of skeleton plot of the chronologies. Dimension (ring-width measurement) and corresponding year of each sample chronology is noted and put into the program (COFECHA) as an input. The program output provides the segmental (varying) cross correlation among all the samples and with the master chronology. Therefore, it is possible to find the missing rings, false rings or extra rings based on the cross correlation results from COFECHA.
In order to obtain better cross-dating results, it is always recommended to use both steps (skeleton plot and COFECHA) to provide robust results. Further, to extend the master chronology back in time, it is matched with the certain archeological specimens such as beams or logs of wood used in Indian pueblos.

2.1.4 Standardization

Standardization is necessary because tree-ring growth is also associated with the age of the tree. Standardization follows two steps (Meko et al., 1993): (a) detrending the measured width to remove the low frequency biological and other non-climatic signals; and (b) averaging the detrended measured ring widths in order to filter non-climatic effects (biological growth) specific to samples. After the cross-dating procedure, each sample’s measured time series width is fitted with the proper trend line (straight line, negative exponential, cubic spline etc.). Fritts (1976) suggested that ring-width from open growth sites with little competition among trees for moisture and light can be fitted with the straight line or modified negative exponential. The cubic spline is preferred if the site condition is unknown depending upon its performance (Meko et al., 1993). The straight line, negative exponential and a polynomial equation curve fitting may lose the low frequency fluctuations (Fritts, 1976) present in the tree-rings, where as the cubic spline may remove the high frequency climate responses (Meko et al., 1993). So fitting the curve is very crucial in extracting the climate signal depending upon the objective of analysis. After fitting the curve, the measured widths are divided by the yearly values of fitted curve. This procedure transforms the ring-width values to the ring-width indices also called standard normal chronologies that have a mean of unity and fairly constant
variance (Fritts, 1976). The standardized normal chronologies of each sample are then averaged for a site to obtain the standard normal chronology.

In addition to variance stabilization (standardization), autocorrelation (due to biological growth and non-climatic effects) in the tree-ring indices can also be removed by fitting a low-order autoregressive model. This process is called the pre-whitening process and the corresponding pre-whitened chronology is called the standard residual chronology.

The ARSTAN software developed by the Dr. Edward R. Cook in 1985 is the commonly used statistical package in the process of standardization and pre-whitening of the tree-ring chronologies (e.g., Woodhouse, 2001; Glueck and Stockton, 2001; Gray et al., 2004). In addition to the option of standardization (fitting the curve) and detrending (different autoregressive order equation), it also performs principal component analysis to separate the different signals contained in the tree-ring chronologies (Cook and Holmes, 1999).

2.2 Tree-ring Reconstruction of Climate Variables

Paleoclimate reconstructions of climate variables at times before the instrumental record, based on proxy indicators sensitive to climate such as tree-rings, ice-cores, pollen, sediments and corals etc., are significant indicators about past climates, natural variability and global climate change (Woodhouse, 2002). Amongst all proxy data, tree-rings are a complex function of climate-related factors that include precipitation, soil and air temperatures, soil moisture conditions, sunshine, evapo-transpiration, and wind (Meko et al, 1995); therefore the tree-ring records can be used as a proxy indicator that
represents the combined climate conditions over a particular period of time. Tree-ring records of past climates are precisely dated, annually resolved, and can be well calibrated and verified (Fritts, 1976). Similar to other climate variables, streamflow repeatedly correlates well with tree-rings because it integrates hydrologic processes including precipitation, soil moisture, and evaporation rates (Knight, 2004).

Furthermore, trees growing in the semi-arid western U.S. have shown good correlation with winter precipitation indicating sensitivity to the cool season precipitation prior to the growing season (Meko et al., 1995), because the winter precipitation recharges the soil moisture for the following tree-ring growth season. The most important seasonal element to water year streamflow in the southwest is a cool-season precipitation or snowfall, which is controlled by local and regional climate variables (Woodhouse, 2002). Therefore, tree-rings are a valuable proxy for reconstructions of water year streamflow and snowpack in this region.

Climate reconstruction includes the screening of tree-ring variables and tree-ring reconstruction using appropriate methods of regression depending upon the predictands (climate variables) and predictors (tree-ring chronologies).

2.2.1 Tree-ring Screening for Climate Reconstruction

The main purpose of prescreening tree-ring chronologies is to find the response of the tree-ring or set of tree-rings (predictors) with respect to climate data (predictands). The tree-rings which are highly correlated, or respond well, are only considered for the prediction (transfer) models for better reconstruction of the climate variables (precipitation, temperature, streamflow etc.).
The correlation criterion is the most common approach of prescreening of the
considering lag +1, +2 and +3 years that were significantly correlated (95%) with the
July-September streamflow of the Potomac River. Hidalgo et al. (2000) used the
correlation criterion (95% significant) to select the tree-ring chronologies considering the
lag -1, 0 and +1 to reconstruct the Colorado River streamflow (Lee’s Ferry). Similarly,
Britto-Castillo et al. (2003) also used the correlation criterion to select the best tree-ring
chronologies for the reconstruction of winter streamflow in the Gulf of California
continental watersheds. The chronologies that had the best correlation with the rainfall
series were considered for the reconstruction of winter streamflow in the region.
Cleaveland (2000) chose three tree-ring chronologies for the reconstruction of summer
streamflow in the White River, Arkansas based on the length and strength of correlation
with the summer streamflow.

Furthermore, Biondi et al. (1999) selected two tree-ring chronologies for July
temperature reconstruction of east-central Idaho based on the correlation, where one
chronology was positively correlated and another was negatively correlated with the July
temperature. These oppositely correlated chronologies were not related to each other;
hence there was no possibility of co-linearity in the regression model. Gray et al. (2004)
used the standard correlation analysis to evaluate the relationship between the tree-ring
chronologies and climate variables (precipitation, temperature, palmer drought severity
index and April 1 snowpack) of northeastern Utah and found that the precipitation in the
region was significantly correlated with the tree-ring growth. Gray et al. (2004) further
used the same procedure for the precipitation reconstruction of the Bighorn Basin, Wyoming.

The regression approach is another way of finding the best predicting variables (tree-ring chronologies). Meko et al. (2001) used the multiple linear regression method to screen the tree ring chronologies for the Sacramento River reconstruction. Water year precipitation was regressed using the tree-ring chronologies in a distributed lag regression model considering lag 0, +1 and +2. Tree-ring chronologies with a regression $R^2$ greater than 0.1, corresponding to marginal significance (alpha = 0.05) of the overall F for the regression, were only considered for further reconstruction. Diaz et al. (2001) used a response function analysis with the chronologies, precipitation and temperature to obtain the best predictors (tree-rings) for the reconstruction of precipitation for Baja California, Mexico.

The correlation with tree-ring data can also be performed by confining the distance of a site to a climate station. The tree-ring chronologies which are within a certain distance from a climate station are only selected for reconstruction. Cook et al. (1999) used the local optimal search radius of 150 km in order to find the best tree-ring sites for a Palmer Drought Severity Index (PDSI) reconstruction. The search radius was increased by every 50 km until at least five candidate tree-ring datasets were obtained. Meko and Woodhouse (2005) screened tree-ring chronologies with the help of the correlation analysis with gridded water year total precipitation in order to find the best predictor for the Sacramento and Blue River streamflow reconstruction. Each possible chronology was correlated with the gridded water year precipitation within a radius of 450 km and accepted for the corresponding streamflow reconstruction as moisture
sensitive if the median correlation with all the grid points in the search radius was positive.

2.2.2 Climate Reconstruction Methods

Climate reconstructions have been successful with different methods, such as Simple Linear Regression (SLR), Multiple Linear Regression (MLR), Stepwise Linear Regression (STLR), Simple Principal Component Regression (SPCR), Multiple Principal Component Regression (MPCR), Stepwise Principal Component Regression (STPCR) and Alternate Principal Component Regression (APCR).

SLR, MLR and STLR consider the original variables (tree-ring chronologies in our case) as the predictors of climate variables (precipitation, temperature, streamflow etc.). But in the case of PCR (SPCR, MPCR, STPCR and APCR), the original variables are transformed into principal components which are orthogonal to each other, and these orthogonal variables are utilized with the regression equation. PCR is strongly recommended if there is multicollinearity among the predictors (i.e. tree-ring data). Multicollinearity in the data sets leads to imprecision of the regression coefficients and incorrect rejection of the variables (Fritts, 1991). A summary of the reconstruction methods by different authors is presented in Table 2-1 and discussed in following sections.
<table>
<thead>
<tr>
<th>Authors/Year</th>
<th>Reconstructed Drought Variable</th>
<th>Reconstructed Methods</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stahle et al. (1998)</td>
<td>Average July PHDI</td>
<td>SLR</td>
<td>Climate divisions of Virginia and north Carolina</td>
</tr>
<tr>
<td>Diaz et al. (2001)</td>
<td>Sept-July precipitation</td>
<td>SLR</td>
<td>Baja California, Mexico</td>
</tr>
<tr>
<td>Pohl and Hadley (2002)</td>
<td>Average monthly PDSI</td>
<td>SLR</td>
<td>Central Oregon</td>
</tr>
<tr>
<td>Cleaveland et al. (2003)</td>
<td>Winter precipitation</td>
<td>SLR</td>
<td>Durango</td>
</tr>
<tr>
<td>Laroque and Smith (2005)</td>
<td>January mean temperature, April 1 snowpack, Previous July mean temperature, July mean temperature</td>
<td>SLR</td>
<td>British Columbia</td>
</tr>
<tr>
<td>Biondi et al. (1999)</td>
<td>July temperature</td>
<td>MLR</td>
<td>East-central Idaho</td>
</tr>
<tr>
<td>Touchan et al. (1999)</td>
<td>October-May precipitation</td>
<td>MLR</td>
<td>Southern Jordan</td>
</tr>
<tr>
<td>Cleaveland (2000)</td>
<td>Summer (June, July and August) streamflow</td>
<td>MLR</td>
<td>White River, Arkansas</td>
</tr>
<tr>
<td>Gray et al. (2004)</td>
<td>Previous year June-current year June precipitation</td>
<td>MLR</td>
<td>Unita Basin</td>
</tr>
<tr>
<td>MacDonald and Case (2005)</td>
<td>PDO</td>
<td>MLR</td>
<td>-</td>
</tr>
<tr>
<td>Woodhouse (2001)</td>
<td>Mean annual streamflow</td>
<td>STLR</td>
<td>Colorado Front Range</td>
</tr>
<tr>
<td>Authors/Year</td>
<td>Reconstructed Drought Variable</td>
<td>Reconstructed Methods</td>
<td>Location</td>
</tr>
<tr>
<td>----------------------------</td>
<td>----------------------------------------------</td>
<td>-----------------------</td>
<td>-----------------------------------------</td>
</tr>
<tr>
<td>Woodhouse (2003)</td>
<td>April 1 snowpack</td>
<td>STLR</td>
<td>Gunnison-Southern River Basin</td>
</tr>
<tr>
<td>Castillo et al. (2003)</td>
<td>Winter streamflow</td>
<td>STLR</td>
<td>Central and southern regions of Gulf of California</td>
</tr>
<tr>
<td>Gray et al. (2004)</td>
<td>Previous year June-June precipitation</td>
<td>SPCR</td>
<td>Bighorn Basin, Wyoming</td>
</tr>
<tr>
<td>Meko et al. (2001)</td>
<td>Annual streamflow</td>
<td>MPCR</td>
<td>Sacramento River</td>
</tr>
<tr>
<td>Cook et al. (1999)</td>
<td>Gridded summer PDSI</td>
<td>MPCR</td>
<td>United States</td>
</tr>
<tr>
<td>Stahle et al. (1998)</td>
<td>Winter Southern Oscillation Index (SOI)</td>
<td>MPCR</td>
<td>-</td>
</tr>
<tr>
<td>D’Arrigo et al. (2001)</td>
<td>PDO</td>
<td>MPCR</td>
<td>-</td>
</tr>
<tr>
<td>Stockton and Jacoby (1976)</td>
<td>Annual streamflow</td>
<td>STPCR</td>
<td>Colorado River Basin</td>
</tr>
<tr>
<td>Touchan et al. (2003)</td>
<td>Spring (May-June) precipitation</td>
<td>STPCR</td>
<td>Southwestern Turkey</td>
</tr>
<tr>
<td>Woodhouse et al. (2006)</td>
<td>Water year streamflow</td>
<td>STPCR</td>
<td>Green, Cisco, San Juan and Lee’s Ferry</td>
</tr>
<tr>
<td>Cook et al. (1998)</td>
<td>NAO</td>
<td>STPCR</td>
<td>-</td>
</tr>
<tr>
<td>Glueck and Stockton (2001)</td>
<td>NAO</td>
<td>STPCR</td>
<td>-</td>
</tr>
<tr>
<td>Gray et al. (2004)</td>
<td>AMO</td>
<td>STPCR</td>
<td>-</td>
</tr>
<tr>
<td>Garen (1992)</td>
<td>Seasonal streamflow</td>
<td>APCR</td>
<td>South Fork Boise River</td>
</tr>
<tr>
<td>Hidalgo et al. (2000)</td>
<td>Water year streamflow</td>
<td>APCR</td>
<td>Lee’s Ferry</td>
</tr>
<tr>
<td>Piechota et al. (2004)</td>
<td>Water year streamflow, Average water year PHDI</td>
<td>APCR</td>
<td>Upper Colorado River basin (Cisco, Green, PHDI)</td>
</tr>
<tr>
<td>Timilsena et al. (2007)</td>
<td>Average water year Palmer Z index</td>
<td>APCR</td>
<td>Upper Colorado River Basin</td>
</tr>
</tbody>
</table>
2.2.2.1 Simple Linear Regression (SLR)

The SLR method contains only one predictor (one tree-ring chronology). In this procedure, the root mean square error (RMSE) and the coefficient of determination ($R^2$) are the important criterion for selecting the best regression equation.

Stahle et al. (1998) utilized SLR using the residual (pre-whitened) regional chronology (residual average of two moisture sensitive tree-ring chronologies) as a predictor to reconstruct the average July Palmer Hydrological Drought Index (PHDI) of climatic divisions in Virginia and north Carolina, for the last 800 years. Similarly, Diaz et al. (2001) reconstructed the September-July precipitation of Baja California, Mexico, using the SLR procedure. Pohl and Hadley (2002) reconstructed the average PDSI of significantly correlated months back to 1455, using SLR for central Oregon, with a single tree-ring chronology. Cleaveland et al. (2003) used a tree-ring chronology to reconstruct the winter precipitation (November-March) in Durango with the same approach. The average of the two tree-ring chronologies was considered as a single predictor. In addition, Laroque and Smith (2005) reconstructed the climate variables (January mean temperature, April 1 snowpack, previous July mean temperature, July mean temperature and summer mean temperature) of the British Columbia Coast Mountains using SLR. Only one highly correlated predictor was selected from the five different tree-ring chronologies for the corresponding climate variable reconstruction.

2.2.2.2 Multiple Linear Regression (MLR)

The MLR method incorporates more than one predictor variables (tree-ring chronologies). As mentioned earlier, the MLR is an appropriate method for reconstruction if the predictors (tree-ring chronologies in our case) do not have any
multicollinearly among each other. Similar to SLR, MLR also utilizes the root mean square error (RMSE) and the coefficient of determination ($R^2$) to obtain the best regression equation.

Biondi et al. (1999) reconstructed July temperature of east-central Idaho using the two correlated tree-ring chronologies (one positively correlated and another negatively correlated) with MLR for the last 858 years. Similarly, Touchan et al. (1999) reconstructed precipitation (October-May) in southern Jordan for the last 396 years. In their study, log-transformed precipitation was used as the predictand in order to incorporate more variability and improve the skill of the regression model. Cleaveland (2000) reconstructed summer (June, July and August) streamflow in the White River (Arkansas) for the last 963 years. Quadratic transformations of three residual tree-ring chronologies were used as predictors in order to improve the performance of the model. Gray et al. (2004) utilized the MLR to reconstruct precipitation (previous year June-current year June) in the Uinta basin back to 1226.

Lastly, MacDonald and Case (2005) reconstructed annual PDO back to 993 using the hydrologically sensitive tree-ring chronologies from southern California and western Canada, and a MLR procedure. These two sites were oppositely correlated to PDO, and thus considered for the reconstruction.

2.2.2.3 Stepwise Linear Regression (STLR)

The STLR method is another approach of climate reconstruction that selects the best predictor (tree-ring chronology) combination among all the feasible predictors. In this procedure, the number of predictors to be selected and the order of entry are decided by statistical criteria (significance tests). At first, the most significant predictor is selected
for the regression equation, and the procedure continues to add a variable (highest F statistic or lowest p-value) until there is no improvement in the performance of the model. This procedure allows skipping of the predictors in order to obtain the better prediction model.

Woodhouse (2001) used the STLR model to reconstruct the Middle Boulder Creek streamflow from 14 tree-ring chronologies in the Colorado Front range as far back as 1703. The square of each of 14 chronologies were used as the independent variable (predictors) to reconstruct the mean annual creek flow (predictand) in order to account for the non-linearity in the climate/tree growth relationships. Woodhouse (2003) also used STLR to reconstruct the snow water equivalent (SWE) of the Gunnison-Southern River basin of the Colorado River. Out of 15 selected tree-ring chronologies, only 4 tree ring chronologies were retained (selected by the stepwise regression procedure) for the best performance of the final regression model. Similarly, Castillo et al. (2003) applied STLR to reconstruct the winter streamflow of the central and southern regions of the Gulf of California continental watershed. In their study, the residual (pre-whitened) chronologies were considered as the predictors for better performance of regression model.

2.2.2.4 Simple Principal Component Regression (SPCR)

The same approach of selecting best predictors for the SLR, as mentioned in Section 2.2.2.1, is followed in the case of SPCR procedure. However, the methods differ in that the predictors of original variables are replaced with principal component scores.

Gray et al. (2004) used the SPCR procedure to reconstruct precipitation (June of previous year -June of this year) back to 1260 in the Bighorn Basin, Wyoming. PC1 (first
principal component) of the tree-ring chronologies was used as the predictor, because PCI was highly correlated with precipitation and PDSI values in the region.

2.2.2.5 Multiple Principal Component Regression (MPCR)

The same approach of selecting best predictors for the MLR Method, as mentioned in Section 2.2.2.2, is followed in the MPCR procedure. But, the methods differ in that the predictors of original variables are replaced with principal component scores.

Meko et al. (2001) used the MPCR procedure to reconstruct the Sacramento River flow back to 869. In their work, the truncated PCs were used as the potential predictors and log-transformed (log,10) streamflow was the predictand. The criterion of allowing PCs into the regression equation was obtained by the significance t-test of the coefficients with an additional test of the PRESS statistic. Cook et al. (1999) developed the Point by Point Regression (PPR) technique, which is defined as the sequential, automated fitting of a single point MPCR model in order to reconstruct U.S. gridded summer PDSI. In their study, candidate tree-rings (predictors) were selected based on a fixed search radius. The criterion of fixing the search radius was to obtain at least 5 chronologies within the radius of 150 km. Otherwise the search radius was increased by increments of 50 km until at least 5 chronologies (predictors) were obtained. Once the tree-ring candidates within the search radius were obtained, the significantly (95%) correlated tree-ring chronologies were only used in the MPCR procedure.

Stahle et al. (1998) used the pre-whitened tree-ring chronologies (early wood and ring width) from Mexico, southwestern United States, southern Great Plains and Java of Indonesia to reconstruct the winter Southern Oscillation Index (SOI). The prediction

26
model was based on MPCR utilizing the first five principal components (predictors) of potential best tree-ring chronologies from the initial screening (correlation considering lag effect). D'Arrigo et al. (2001) used nine highly correlated tree-ring chronologies from the Pacific Northwest, Coastal Alaska and subtropical North America and two reconstructed PDSI grids (Cook et al., 1999) to reconstruct the PDO. The MPCR procedure based reconstruction considering the tree-ring chronologies and reconstructed climate variables (PDSI) incorporated higher variability as compared to reconstruction based on only tree-ring chronologies. Cook et al. (2002) reconstructed winter NAO using multi-proxy data (tree-ring and core-records) of predictors using the MPCR. In Cook et al. (2002), the candidate predictors were selected based on the correlation criterion. Next, the selected candidate predictors (based on correlation criterion) were orthogonalized using PCA. However, the truncated (eigenvalues >1) principal components were only used for the final MPCR procedure.

2.2.2.6 Stepwise Principal Component Regression (STPCR)

The same approach used for the STLR Method, as mentioned in Section 2.2.2.3 is followed in STPCR. But, the procedure differs in that the predictors of original variables are replaced with principal component scores.

Stockton and Jacoby (1976) reconstructed the twelve annual unimpaired streamflow stations in the Colorado River basin using 30 tree-ring chronologies with the help of the STPCR procedure. In this procedure, the principal components of the candidate predictors (original variables and its lag) were calculated first and then they were truncated based on the eigenvalues criterion (greater than unity). Similar to other stepwise regression procedures, the significance tests (F test and p test) were carried out.
before entering the additional components in the final regression model. Broackway and Bradley (1995) reconstructed annual streamflow of the John Day River in Oregon and Yakima River in Washington applying the same approach of STPCR. Touchan et al., (2003) reconstructed spring precipitation (May-June) in southwestern Turkey using STPCR. In addition to other significance tests (F and p test), the PRESS statistic was used to select the best predictors. Touchan et al. (2005) further reconstructed May-July standardized precipitation index back to 1251 based on 6 tree-ring chronologies of Turkey using the same approach. Recently, Woodhouse et al. (2006) reconstructed Lee’s Ferry streamflow stations in the Colorado River basin using the STPCR model considering the predictor as the residual chronologies. The strength of the model was evaluated by the adjusted $R^2$ and F level of regression as done in stepwise regression models.

Cook et al. (1998) reconstructed the NAO using pre-whitened tree-ring chronologies from North America and Europe with the STPCR technique. In their work, truncated principal components based on the eigen cut-off rule were used as the predictors for the STPCR procedure. Glueck and Stockton (2001) used the STPCR procedure to reconstruct the NAO using multi-proxy predictors (22 tree-ring chronologies and Greenland Summit Ice cores). Gray et al. (2004) reconstructed the AMO using the 12 strongly correlated tree-rings records of eastern North America, Europe, Scandinavia and Middle East using the same approach of STPCR.

2.2.2.7 Alternate Principal Component Regression (APCR)

Garen (1992) introduced the Alternate Principal Component Regression (APCR) technique where the t-test and sign test were used as a retaining criteria of principal
components in the regression equation with an additional testing method of cross validation. A PCR with an additional skill of the cross validation finds the best combination of variable that gives optimal or nearly optimal solutions (Garen, 1992), and also demonstrates that independent additional testing techniques can improve the overall skill of the model (Garen, 1992; Hidalgo et al., 2000). In APCR, skipping of the components was not allowed and the lowest cross validation standard error was the criterion of selecting the best predictors.

Hidalgo et al. (2000) reconstructed Lees’ Ferry streamflow using the APCR procedure considering both the rotated and unrotated principal components with a testing method of cross validation. The APCR procedure by Hidalgo et al. (2000) provided a more parsimonious model and had better prediction skill than the conventional STPCR by Stockton and Jacoby (1976). Piechota et al. (2004) reconstructed two streamflow stations (Cisco and Green) and PHDI, back to 1496 of the Upper Colorado River Basin using APCR (Hidalgo et al., 2000) based on 17 tree-ring chronologies. Timilsena et al. (2007) reconstructed the Palmer Z index in the Upper Colorado River basin using tree-ring chronologies by means of the APCR procedure.

2.3 Regionalization of Climate Variables

Climate regionalization is a useful technique that enables generalization of areas based on spatially and temporally varying climate parameters (Comrie and Glenn, 1998). Principal component analysis and cluster analysis are broadly used procedures to obtain the coherent modes of various climate parameters. Adequate knowledge of spatial distribution and regionalization of climate variables (snowpack, streamflow) are
important to evaluate the trend of climate changes in the region, and also allows the study of teleconnections within the region.

In the principal component method of regionalization, the loading or correlation between original variables and principal components are the important criteria of dividing regions. Mallants and Feyen (1990) used the rotational principal component (S-mode PCA) to divide the rainfall (precipitation) variations over the Ijzer watershed of western Belgium and northern France into four coherent sub-regions of consistent rainfall. The station data that were correlated (critical $r = 0.85$) with principal components were considered as the corresponding coherent regions of climate. Similarly, Meko et al. (1993) used the rotated principal component analysis on instrumental gridded data of PDSI to delineate regions of common tree-growth variation in the U.S. and southeastern Canada. In order to delineate regions, the loading map was used and presented. Comrie and Glenn (1998) used the rotational principal component to obtain the coherent regions using the seasonal and monthly data; where divisions of regions were based on the values of rotated principal component loadings. Knapp et al. (2002) also used the rotational principal component in the climate variable (Climate Pointer Year Index) in order to find the spatially homogeneous regions in the interior Pacific northwest region of the United States. Woodhouse (2003) applied rotational principal components (S-mode PCA) to define the snowpack region using April 1 SWE in the western Colorado River basin. In his study, the two coherent regions in the western Colorado basin were identified based on the retained two principal components (eigenvalues criteria >1). Furthermore, Quiring and Papakyriakou (2005) used S-mode unrotated PCA to establish the spatial patterns of June-July moisture (Palmer Z-Index) variability in the Canadian Prairies. The
regionalization was based on the loading values of the first three principal components retained by the scree-test.

Cluster analysis is another technique of climate regionalization that is used to place objects into groups or clusters based on statistical similarities of their properties (Yeh et al., 2000). In cluster analysis, the Euclidean distance method of classification is most commonly used in which the smallest distance between point stations and group of points (cluster) is taken into an account. Yeh et al. (2000) applied cluster analysis to classify the precipitation patterns over northern California. The classified stations were combined into a geographical information system to obtain the regional precipitation indices.

The evaluation of the results of the principal component analysis can be corroborated with the group average linkage cluster analysis (Piechota et al., 1997). Piechota et al. (1997) regionalized 79 streamflow stations in the western U.S using the rotated S-mode PCA and cluster analysis (Group average linkage method). The corresponding coherent streamflow divisions in the region were based on the principal component loading and Euclidian distance between the different clusters. Furthermore, Baeriswyl and Rebetez (1997) used the S-mode PCA and cluster analysis on the two sets of instrumental precipitation data in Switzerland. Cluster analysis was carried out on the leading PCs in order to divide into the different sub-regions.

2.4 Ocean/Atmosphere Influences on U.S. Climate

The oceans and atmospheres are closely linked both in spatial and temporal scales. It has been well documented that the Pacific and Atlantic Ocean climate
variability influences regional hydrologic activity in the United States (e.g. Redmond and Koch, 1991; Piechota and Dracup, 1996; Enfield et al, 2001). The Pacific climate phenomena (ENSO, PDO) and Atlantic climate phenomena (NAO, AMO) act at the scales of interannual to interdecadal (Trenberth and Hurrel, 1994; Hidalgo, 2004). In order to understand the 20-80 years of Pacific periodic climate phenomena such as PDO, AMO and NAO, a long extended climate record is essential because these phenomena only cover only one to two phases during the available instrumental period.

2.4.1 El Niño-Southern Oscillation (ENSO)

ENSO is a natural but largely unpredictable phenomenon that results from complex interaction between clouds and storms, regional winds, oceanic temperatures, and ocean currents along the equatorial Pacific (USGS, 2007). It operates on a timescale of 2-7 years (Hanley et al., 2003). There are two phases of ENSO; warm phase is referred to as El Niño and the cool phase is referred to as La Niña. Although the mechanical process of ENSO is known, the exact origins of the processes that govern its repetitive phenomenon are still not certain (NOAA, 2007). In "normal" conditions of equatorial Pacific, the tropical trade winds blow from east to west, ponding up warm water in the western Pacific and pulling up cold, deep, nutrient-rich waters in the eastern Pacific from the Ecuadorian coast. But in the El Niño situation, the trade winds weakens that causes the upwelling of cool waters in the eastern Pacific and drifting of the pool of warm water towards the central and eastern pacific (South America). As the waters of the central and eastern Pacific warm, the strong tropical Pacific storms begin to form farther east and interact with the jet stream over the North Pacific Ocean. This jet stream collects moisture and storms, and carries them to the southwestern United States and northern
Mexico. Sometimes El Niño events give way to unusually cold sea-surface temperatures and abnormal strong trade winds, a condition now called La Niña. However, La Niña may begin on its own, without preceding an El Niño. The effects of the El Niño and La Niña on global climate are, in part, mirror images of each other (USGS, 2007). For instance, drought is a common occurrence in the southwestern United States during La Niña, in contrast to the wet years associated with El Niño (USGS, 2007).

Teleconnections study between ENSO in the tropical Pacific Ocean and climate variables (precipitation, snow accumulation and streamflow) in U.S. have showed that there is a strong association of ENSO in U.S. climate. Redmond and Koch (1991) found that the El Niño events positively correlate with colder than average temperature and higher than average runoff in the southwestern U.S., however, it was the opposite in the northwestern United States. Piechota and Dracup (1996) found that there were two significant regions (northwest and southwest) with a strong ENSO signal in the United States. A strong relationship was established between ENSO and the southern U.S. in which consistent dry conditions occur during the La Niña years and consistent wet conditions occur during the El Niño years. Cayan et al. (1999) found that high precipitation and high streamflow were more frequent in the southwest U.S. during the El Niño years while there was an opposite pattern during the La Niña years. Clark et al. (2001) also found that the mean of April 1 SWE and annual runoff were wetter than the average during the El Niño years and were opposite during the La Niña years.

2.4.2 Pacific Decadal Oscillation (PDO)

The PDO is a long-lived El Niño-like pattern of Pacific climate variability (Shen et al, 2006). Mantua et al. (1997) defined the PDO as the leading principal
component of north pacific monthly sea surface temperature variability (poleward of
20°N for the 1900-93 periods). The exact mechanism that causes PDO is still not known,
while causes for ENSO are relatively well understood (Mantua, 2002). Similar to ENSO,
PDO has two phases: warm and cold. Each phase persists for about 25-30 years. The
main characteristics that distinguish of PDO from El Niño/Southern Oscillation (ENSO)
is that the PDO cycle persists for 50-60 years, while typical ENSO events persist for 2 to
7 years.

Earlier studies established that the PDO phases (warm or cold) can enhance or
dampen the effect of ENSO teleconnections to climate variables. Gershunov and Barnett
(1998) found that the ENSO signal (El Niño/La Niña) was strongest during the different
phases of PDO (positive and negative); El Niño patterns (wet southwest) were strongest
and consistent during the high phase of PDO. The precipitation patterns during the La
Niña winters (dry winters) were consistent during the low PDO phases. Barlow et al.
(2001) concluded that there is a significant relationship between the Pacific variability
(ENSO, PDO) and U.S. warm season precipitation. The relation of this Pacific variability
has little effect on annual streamflow and drought, but substantially effects the monthly
precipitation. Harshburger et al. (2002) concluded that the combination of ENSO and
PDO phases has the strongest influence on winter precipitation and spring stream
discharge in northern Idaho. The most significant combination was that La Niña with the
negative phase of PDO contributes to wet winters in northern Idaho, which was also
consistent with the study by the Gershunov and Barnett (1998) where La Niña patterns
(wet northwest) were stronger during the negative phase of PDO. Graumlich et al. (2003)
obtained teleconnections between the Upper Yellowstone river flow and ocean climate
variability in the Pacific Ocean using the reconstructed time series of climate variables (reconstructed PDO and reconstructed SOI). It was found that a high flow in the Upper Yellowstone River flow was associated with the strong ENSO and/or PDO climate patterns in the Pacific Ocean. Beebee and Manga (2004) also concluded that the magnitude of the streamflow in central and eastern Oregon was significantly affected by the ENSO and PDO. ENSO was highly correlated with the variability of annual discharge where PDO was highly correlated with the spring snowmelt timing, magnitude and timing of annual floods.

2.4.3 Atlantic Multidecadal Oscillation (AMO)

The Atlantic Multidecadal Oscillation is the leading mode of low-frequency, North Atlantic Ocean (0 - 70°) SST variability with a periodicity of 65-80 years (Kerr, 2000; Gray et al., 2004). Similar to ENSO and PDO, it has two phases: warm and cold each persists for about 30-40 years. Its development is attributed to the thermohaline circulation (Delworth and Mann, 2000). Temperature (thermo) and its salinity (haline) are major elements that control the density of sea water, and difference of density between two places causes the circulation called thermohaline circulation. Due to the density differences between heavier salty water and lighter fresh water, it causes stronger and weaker ocean currents, respectively. This circulation (thermohaline circulation) is believed to be one of the major causes of warm and cold phase of the AMO due to warming and cooling of Atlantic SSTs when it speeds up and slow down respectively. This can impact the climate of different parts of the world.

Sutton and Hodson (2005) concluded that AMO is responsible for changes in the regional atmospheric circulation and associated for anomalies in precipitation and surface
temperature over the United States. Enfield et al. (2001) related the AMO phases with the U.S. rainfall and river flows and found that warm phases of AMO results in below normal rainfall in most parts of the United States. Rogers and Coleman (2003) studied the interactions between AMO, PNA, and warm and cold equatorial pacific Niño 3.4 events producing the Mississippi river discharge variations. In their study, the AMO was significantly linked to the streamflow differences in the Upper Mississippi Valley as well as winter streamflow, but it was not so significant in the Lower Mississippi Valley. The most consistent signal was low (high) streamflow in the Upper Mississippi River basin during the warm (cold) phase of AMO.

2.4.4 North Atlantic Oscillation (NAO)

The NAO is a large-scale fluctuation in atmospheric pressure between the subtropical high pressure system located near the Azores in the Atlantic Ocean and sub-polar low pressure system near Iceland and is quantified as the NAO index. It is a dominant mode of atmospheric and climate variability in the North Atlantic region (Hurrel, 1995; Hurrel and Van Loon, 1997). Similar to PDO and AMO, NAO has both the positive and negative phase. A NAO related impacts on winter climate extend from Florida to Greenland and from northwestern Africa over Europe far into northern Asia (Visveck et al., 2001). The mechanism of NAO that generates low frequency changes in the north Atlantic climate is still not understood (Visveck et al., 2001).

Strong positive phases of the NAO tend to be associated with above-average temperatures in the eastern U.S. and across northern Europe and below-average temperatures in Greenland (NOAA, 2006). Dettinger and Diaz (2000) found that NAO is significantly correlated with streamflow in the eastern United States. During the winter in
the eastern U.S., the negative NAO reduces the moisture transportation and therefore reduces the streamflow in the region. Tootle et al. (2005) used nonparametric testing to evaluate the individual and coupled effect of PDO, AMO, NAO and ENSO using instrumental data of U.S. streamflow and observed data of ocean climate information (PDO, AMO, NAO and ENSO) and found that the warm phase of PDO is associated with increased streamflow in the central and southwest United States, while the warm phase of the AMO is associated with reduced streamflow. The positive phase of the NAO and cold phase of the AMO are associated with increased streamflow in the central United States. Hunter et al. (2006) recently studied the interaction of interannual and interdecadal oceanic-atmospheric influences on April 1 SWE in the western United States. Their study showed that La Niña results in increased SWE in the Pacific northwest and decreased SWE in the southwest region. The coupling of a cold AMO phase increases the effect of La Niña in the northern and central Rocky Mountains resulting in enhanced SWE.
CHAPTER 3

PARTIAL LEAST SQUARE REGRESSION FOR IMPROVED STREAMFLOW RECONSTRUCTION

3.1 Introduction

Reconstructed data of hydrologic variables such as streamflow, precipitation and Palmer Drought Severity Index (PDSI) are important for the study of climate variability (e.g., Cook et al., 1999; Hidalgo et al., 2000; Gray et al., 2004; Woodhouse et al., 2006). Further, longer records using paleoclimatic data (e.g., tree-rings) are needed to evaluate droughts from a long-term perspective and for developing better management practices during droughts. There are a variety of methods available to reconstruct hydroclimatic variables from tree-ring data. Stahle et al. (1998), Touchan et al. (1999), Diaz et al. (2001), Woodhouse (2001), and Gray et al. (2004) used multiple linear regression and stepwise linear regression to reconstruct hydroclimate variables such as precipitation, streamflow and other drought variables. Similarly, Stockton and Jacoby (1976), Cook and Jacoby (1983), Garen (1992), Meko et al. (2001), Hidalgo et al. (2000), and Woodhouse et al. (2006) used principal component regression for streamflow reconstruction.

In multiple linear regressions, the original values of predictors are utilized to explain as much of the dependent variable variation and it does not consider multicollinearity effect on prediction. In principal component regression, the X-scores
(scores of independent variable) are chosen to explain as much of the factor variation as possible. This approach yields informative directions in the factor space, but may not be associated with the shape of the predicted surface.

Another method available for reconstruction of hydroclimate data is partial least square regression (PLSR) which was developed in the late 1960’s (Wold, 1966) and gained importance in the field of chemistry during the 1970’s (Gerlach et al., 1979). PLSR has recently been used for reconstruction of summer temperatures in eastern Norway (Kalela-Brundin, 1999).

PLSR generalizes and combines features from principal component regression and multiple linear regressions (Ablitt et al., 2004). PLSR uses the principal components of both the predictor (X-scores) and predictand (Y-scores). The X-scores and Y-scores are chosen so that the relationship between successive pairs of scores is as strong as possible. In principle, this is a robust form of redundancy analysis, seeking directions in the factor space that are associated with the high variations in the responses. PLSR differs from principal component regression in that the PLSR model is based on the principal components of both the predictor (i.e., tree-ring chronologies) and the predictand (i.e., streamflow). In PLSR, the principal component scores of both tree-ring chronologies and streamflow are used in lieu of the original data to develop the regression model. The ability of PLSR to extract the correlation between predictors and predictands that are itself collinear, allows it to deal with the problems that would be inappropriate for multiple linear regression or principal component regression (Ablitt et al., 2004). This is an attractive feature of PLSR and typically results in improved statistical quality of the reconstruction.
PLSR has similar features to canonical correlation analysis (CCA) which is used to extract and measure the correlation between the two sets of variables (predictors and predictands). The main objective of the CCA is to find the linear combination of the components of the independent vector variable (X variables) that are highly correlated with the linear combination of the dependent variable (Y). It can be understood as an extension of multiple linear regressions, that extracts highly correlated factors (Rao et al., 2006). However, the goal of PLSR is to determine the factors within the independent vector (X variable) that are important for the dependent vector (Y variable) by investigating the modes that explain as much as possible of the covariance between X and Y.

The objectives and contributions of the research presented here were to demonstrate the use of PLSR for improved reconstruction of streamflow in the entire Colorado River basin. A PLSR reconstruction model with tree-ring chronologies was developed using cross validation (drop one approach) as an independent testing and evaluation technique. PLSR with the cross validation technique based on the absolute value of correlation has not been used in dendroclimatology for the reconstruction of hydrologic variables such as streamflow. Further, the reconstruction results and corresponding skill using the PLSR methodology were compared with the performance of two common reconstruction procedures (Stepwise Principal Component Regression (STPCR) and Stepwise Linear Regression (STLR)). Lastly, the reconstruction of streamflow for the entire Colorado River basin builds on prior studies that have focused on single stations in the Upper or Lower basin. A dataset with multiple stations is important for input to system models.
3.2 Data Sources

3.2.1 Unimpaired Streamflow Data

In reconstructing long-term hydroclimate variability, it is important to use unimpaired streamflow data. An unimpaired streamflow station is defined as one that is minimally affected by artificial diversions, storage, or other works of man in or on the natural stream channels or in the watershed. This provides an account of hydrologic responses to fluctuations in climate for watershed (Slack et al., 1993). Unimpaired streamflow stations for the Upper and Lower Colorado River basin were obtained from the United States Geological Survey (USGS), Hydro-Climate Data Network (HCDN) website (http://pubs.usgs.gov/wri/wri934076/1st_page.html). Based on the HCDN website, there were 45 unimpaired streamflow stations in the Upper Colorado River basin and 18 unimpaired streamflow stations in the Lower Colorado River basin (Figure 3-1). In the Upper basin, there were 16 stations which had at least 50 years of regular unimpaired streamflow record. Similarly in the Lower basin, 6 stations had at least a 50 year record of unimpaired regular streamflow. Initially, these 22 (16+6) stations were chosen for the reconstructions since these stations gave a sufficient number of overlapping years with the tree-ring chronologies. Out of these 22 stations, 3 stations (Green River at Warren Bridge station; East Fork River near Big Sandy; and San Pedro River at Charleston) did not have enough predictors (i.e. less than 4 predictors) and were least correlated (<=0.5) with the tree-ring chronologies. Two other stations (Elk River at Clark and Green River at Green River Station) could not incorporate at least 40% of the variability; hence, they were dropped for the PLSR calibration and reconstruction.
procedure. The remaining 17 stations were reconstructed for the last 500 years. Details of these stations are presented in Appendix-A2.

Figure 3-1: Location map displaying 17 unimpaired streamflow stations and 42 standard tree-ring chronologies. Unimpaired streamflow stations are indicated by black dark circle. Standard tree-ring chronologies are indicated by the small dark triangles.

3.2.2 Tree-Ring Data

Tree-ring data were collected and compiled from the International Tree Data Bank website (http://www.ncdc.noaa.gov/paleo/treering.html) maintained by the National Oceanic and Atmospheric Administration (NOAA), World Data Center for
Paleoclimatology. Tree-ring data consisted of information on tree species, standard normal width, standard residual width, and wood density measurement with its location. The standard normal widths of site chronologies (hereinafter referred to as standard tree-ring chronologies) represent the growth indices for each site, and this information was used for the unimpaired streamflow reconstructions. Altogether, the Upper Colorado River basin includes 57 standard tree-ring chronologies; and the Lower Colorado basin contains 54, of which 35 in the Upper and 7 in the Lower have at least the last 500 years of record. These 42 standard tree-ring chronologies are shown in Figure 3-1 and Appendix–A3 were utilized as predictors for the calibration and reconstruction of unimpaired streamflow stations. These 42 standard tree-ring chronologies were selected based on their available data over the last 500 years and sufficient overlapping period with the streamflow stations. These 42 standard tree-ring chronologies were located at different elevation ranging from 1828-3535 meters. Out of these 42 standard tree-species, 19 were Douglas Fir, 11 were Pinyon species (PIED), 4 were Ponderosa pine (PIPO), 3 were Spruce species (PCEN), and 2 were Limber pine (PIFL). The remaining 3 were Great basin B.C. pine (PILO), White pine (PISF) and Rocky Mountain B.C. pine (PIAR).
From the tables Appendix-A2 and Appendix-A3, it is noteworthy that the streamflow has less autocorrelation (lag 1) as compared to standard tree-ring chronologies in the region. This is important to note since one of the overriding assumptions in this research is that streamflow does not have significant autocorrelation (on a year to year basis); however, tree-ring chronologies from previous years may contain useful information.

3.3 PLSR Reconstruction and Comparison

3.3.1 Pre-screening of Predictors

Prescreening of the standard tree-ring chronologies was carried out by correlation analysis for an overlapping period (initial 35 years for comparison) of instrumental water year streamflow and standard tree-ring chronologies. The correlation coefficients between water year streamflow and the selected standard tree-ring chronologies (lag -1, 0 and +1) within the basin were obtained. Chronologies that were positively and negatively correlated with station water year streamflow with less than 95% significance were eliminated from the prescreening, and were not used as predictors for respective streamflow reconstruction, even though these less correlated chronologies may contain important information for a particular season. Inclusion of less correlated tree-ring chronologies may increase the additional noise in the regression model.

The information of correlation with the tree-ring chronologies is presented in Table 3-1. There were 51-126 available standard tree-ring chronologies (including lag -1, 0 and +1 years) prior to prescreening for each streamflow calibration. These chronologies had a correlation value ranging from 0.86-0.00 (in absolute term) with water year
streamflow depending upon the station. Based on the prescreening criterion of correlation
greater than 95% confidence level with the water year streamflow, the number of
standard tree-ring chronologies (predictors) were considerably reduced to 8-40 depending
upon the streamflow station. The highest correlation coefficient with the best correlated
tree-ring ranged from 0.86-0.57 (in absolute value) depending upon the streamflow
station.
Table 3-1: Summary of correlation between water year streamflow (predictands) and standard tree-ring chronologies (predictors), and pool of predictors. Column (1), CAL/VAL PERIOD refers to calibration and validation period. In Column (2), CORREL (CAL/VAL) refers to correlation range (in terms of absolute value) of streamflow stations with the tree-ring chronologies for the overlapping period of calibration and validation. In Column (3), NTREEP refers to the number of available tree-ring chronologies before prescreening including lag (-1, 0 and +1) years. In Column (4), COR RANGE >95% refers to the value of correlation range > 95% confidence level with the tree-ring chronologies (in terms of absolute value). In Column (5), NPOOL refers to available number of tree-ring chronologies in pool of predictors including lag (-1, 0 and +1) years.

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>Station Name</th>
<th>Data (WY)</th>
<th>CAL/VAL PERIOD (1)</th>
<th>CORR. CAL/VAL (2)</th>
<th>N TREEP (3)</th>
<th>COR. &gt; 95% (4)</th>
<th>N POOL (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EAST RIVER AT ALMONT</td>
<td>1935-2005</td>
<td>1936-1970</td>
<td>0.60-0.00</td>
<td>93</td>
<td>0.60-0.34</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>TOMICHI CREEK AT GUNNISON</td>
<td>1938-2005</td>
<td>1939-1973</td>
<td>0.65-0.00</td>
<td>51</td>
<td>0.65-0.40</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>LAKE FORK AT GATEVIEW</td>
<td>1938-2005</td>
<td>1939-1973</td>
<td>0.72-0.00</td>
<td>51</td>
<td>0.72-0.36</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>SMITH FORK NEAR CRAWFORD</td>
<td>1936-1988</td>
<td>1937-1971</td>
<td>0.57-0.01</td>
<td>72</td>
<td>0.57-0.36</td>
<td>17</td>
</tr>
<tr>
<td>5</td>
<td>UNCOMPAGHRE RIVER AT COLONA</td>
<td>1913-2005</td>
<td>1914-1948</td>
<td>0.71-0.00</td>
<td>126</td>
<td>0.71-0.33</td>
<td>33</td>
</tr>
<tr>
<td>6</td>
<td>YAMPA RIVER AT STEAMBOAT SPGS.</td>
<td>1911-1987</td>
<td>1912-1946</td>
<td>0.77-0.00</td>
<td>126</td>
<td>0.77-0.35</td>
<td>39</td>
</tr>
<tr>
<td>7</td>
<td>YAMPA RIVER NEAR MAYBELL</td>
<td>1917-2005</td>
<td>1916-1950</td>
<td>0.75-0.00</td>
<td>123</td>
<td>0.75-0.34</td>
<td>30</td>
</tr>
<tr>
<td>8</td>
<td>WHITE ROCKS RIVER NEAR W.R.</td>
<td>1930-2005</td>
<td>1931-1965</td>
<td>0.68-0.01</td>
<td>96</td>
<td>0.68-0.34</td>
<td>24</td>
</tr>
<tr>
<td>9</td>
<td>WHITE RIVER NEAR MEEKER</td>
<td>1910-2005</td>
<td>1911-1945</td>
<td>0.72-0.00</td>
<td>126</td>
<td>0.72-0.34</td>
<td>30</td>
</tr>
<tr>
<td>10</td>
<td>FISH CREEK ABOVE RESERVOIR NEAR SCOFIELD</td>
<td>1939-2005</td>
<td>1940-1974</td>
<td>0.58-0.00</td>
<td>51</td>
<td>0.58-0.34</td>
<td>11</td>
</tr>
<tr>
<td>11</td>
<td>ANIMAS RIVER AT DURANGO</td>
<td>1928-2005</td>
<td>1929-1963</td>
<td>0.79-0.00</td>
<td>117</td>
<td>0.79-0.34</td>
<td>34</td>
</tr>
<tr>
<td>12</td>
<td>ANIMAS RIVER AT FARMINGTON</td>
<td>1931-2005</td>
<td>1932-1966</td>
<td>0.86-0.00</td>
<td>96</td>
<td>0.86-0.34</td>
<td>29</td>
</tr>
<tr>
<td>13</td>
<td>BRIGHT ANGEL CREEK NEAR GRAND</td>
<td>1924-1973</td>
<td>1925-1959</td>
<td>0.72-0.00</td>
<td>123</td>
<td>0.72-0.34</td>
<td>32</td>
</tr>
<tr>
<td>14</td>
<td>VIRGIN RIVER AT LITTLEFIELD</td>
<td>1930-2005</td>
<td>1931-1965</td>
<td>0.74-0.00</td>
<td>96</td>
<td>0.74-0.37</td>
<td>27</td>
</tr>
<tr>
<td>15</td>
<td>GILA RIVER NEAR GILA</td>
<td>1929-2005</td>
<td>1930-1964</td>
<td>0.66-0.00</td>
<td>99</td>
<td>0.66-0.37</td>
<td>25</td>
</tr>
<tr>
<td>16</td>
<td>SAN FRANCISCO RIVER AT CLIFTON</td>
<td>1936-2005</td>
<td>1937-1971</td>
<td>0.60-0.00</td>
<td>72</td>
<td>0.60-0.37</td>
<td>12</td>
</tr>
<tr>
<td>17</td>
<td>SALT RIVER NEAR ROOSEVELT</td>
<td>1914-2005</td>
<td>1915-1949</td>
<td>0.69-0.00</td>
<td>123</td>
<td>0.69-0.33</td>
<td>40</td>
</tr>
</tbody>
</table>
3.3.2 Partial Least Square Regression Procedure

After the appropriate predictor variables were identified, a PLSR reconstruction model was developed. A brief summary of PLSR follows. As noted earlier, PLSR generalizes and combines features from both principal component analysis (PCA) and multiple linear regressions (MLR) (Abdi, 2003). PLSR is especially useful when there is a need to provide a prediction from a very large set of independent (predictors) variables (Abdi, 2003). In PLSR, the principal component scores of both (X) and (Y) are used in lieu of the original data to develop the regression model. PLSR identifies components from (Y) that are also relevant for (X) (Abdi, 2003). The generalization step results in PLSR searching for a set of components (latent vectors) that explains the maximum covariance between (X) and (Y) which is followed by a regression step where the decomposition of (X) is used to predict (Y) (Abdi, 2003).

The PCA of the matrix (X) (i.e., the matrix of predictors or independent variables) decomposes (X) into a score matrix (T) times a loading matrix (P) and a residual (i.e., error) matrix (E) (Wold et al., 1987).

\[ X = T \ast P' + E \]  \hspace{1cm} (3-1)

Similarly, (Y) is decomposed into a score matrix (U) times a loading matrix (R) and a residual matrix (F).

\[ Y = U \ast R' + F \]  \hspace{1cm} (3-2)

These equations (3-1 and 3-2) are commonly referred to as the outer relations (Geladi and Kowalski, 1986). The objective of the PLSR model is to minimize (F) while maintaining the correlation between (X) and (Y), referred to as the inner relation U (Geladi and Kowalski, 1986).
\[ U = B \cdot T + H \]  

(3-3)

Where \( (H) \) represents the error, \( (B) \) is a diagonal matrix explaining the correlation between \( (X) \) and \( (Y) \). When equation (3-3) is inserted into equation (3-2), a predictive relation for \( (Y) \) is developed where \( (F^*) \) represents the error.

\[ Y = T \cdot R'B + F^* \]  

(3-4)

Equation (3-4) is sometimes referred to as the mixed relation where \( (F^*) \) is to be minimized (Geladi and Kowalski, 1986).

The prediction residual sum of squares (PRESS) statistic is a cross validation calculation that determines the minimum (optimum) number of components required (Geladi and Kowalski, 1986). The cross-validation consists of removing a row (or multiple rows) from the data matrix and then completing the eigen analysis on the reduced matrix. Target testing is then performed on the removed rows using the various levels of the abstract factor space and the difference between the target points and the predicted points is calculated (Malinowski, 2002). This process is repeated until every row has been deleted once and the errors in the target fit for each row are summed (Malinowski, 2002). The PRESS \((j)\) statistic is calculated for each of the \( j \) factor levels using the following equation

\[
PRESS\ (j) = \sum_{i=1}^{r} \sum_{k=1}^{c} (\hat{d}_{ik} - d_{ik}(j))^2
\]  

(3-5)

where \( \hat{d}_{ik}(j) \) and \( d_{ik} \) are the predicted and actual values, respectively, of the deleted rows obtained with \( j \) factors and, \( r \) and \( c \) are the matrix dimensions (Malinowski, 2002). The most popular method to determine the optimal number of latent variables is
the PRESS statistic with the minimum value (Malinowski, 2002), which was applied to the current research.

3.3.3 Selection of Best Predictors

In order to find the best predictors in the case of PLSR, cross validation based on the correlation criterion was followed. First, significantly (95%) positively and significantly (95%) negatively correlated tree-ring chronologies with water year streamflow for the calibration period were ranked based on their absolute value of correlation. Then, PLSR was run for all the significantly correlated tree-ring chronologies. The procedure was followed by eliminating the least significantly correlated variables until the cross validation standard error (CVSE) was minimized (Michealson, 1987).

In this study, independent testing of model performance utilizing the CVSE criterion based on Garen (1992), Hidalgo et al. (2000) and Timilsena et al. (2007) was utilized in order to evaluate the performance of the model. The CVSE is defined as:

\[
CVSE = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n - p}}
\]

where \(y_i\) is the observed streamflow for year \(i\); \(\hat{y}_i\) is the fitted response of the \(i\)th year computed from the fit with the \(i\)th observation removed; \(n\) is the number of years in the data set; and \(p\) is the number of predictors in the regression equation. The best predictors are the remaining tree-ring chronologies after eliminating the least significantly correlated tree-ring chronologies with the minimum CVSE value. Other performance statistics such as Mallows statistics or Akaike information criterion could also be applied in order to select a parsimonious model (Samaniego and Bardossy, 2005).
However, the minimum CVSE criterion was used here to remain consistent with past studies.

In the case of stepwise principal component regression (STPCR) (see section 3.3.4), the value of $p$ is the number of principal components included in the regression equation. Similarly for stepwise linear regression (STLR) (see section 3.3.4), $p$ is the original predictor variables incorporated in the regression equation. In PLSR, $p$ is the optimal number of latent variables that was utilized to develop the regression equation.

3.3.4 Comparison of PLSR with the Stepwise Principal Component Regression and the Stepwise Linear Regression

Various methods are available to perform reconstructions of hydrologic parameters (e.g., streamflow) using tree-ring chronologies (e.g., Hidalgo, 2000; Woodhouse et al., 2006). In this research, the CVSE value from the PLSR reconstruction was compared with the CVSE value from STLR (Woodhouse, 2001) and STPCR (Stockton and Jacoby, 1976). These are two methods that are commonly used to reconstruct hydrologic variables.

STLR is used to sequentially add information (tree-ring chronologies) to a statistical model and evaluate the significance of added parameters. However, this does not always eliminate collinearity between variables. When STLR is used alone with a large number of correlated predictors, the calibration (or test) model results in a good fit for the sample data. However, for new data (verification), the STLR model can result in very large standard deviations of the estimates. In this research, the approach of adding and eliminating variables based on the confidence level (95%) was utilized for the pool of predictors.
As mentioned earlier, STPCR is recommended when there is a high degree of correlation among tree-ring chronologies (predictors). In the case of STPCR, first the matrix (X) (i.e., the matrix of predictors or independent variables) decomposes into a score matrix (T) times a loading matrix (P) and a residual (i.e., error) matrix (E) (Wold et al., 1987). The score and loading from the decomposition of (X) provide information about the systematic structure in (X). PCA is equivalent to Singular Value Decomposition (SVD) and is used to compute the eigenvectors of the covariance matrix (X’X) or the association matrix (XX’). In this research, PC scores (eigen values >1) obtained from the pool of prescreened predictors were only utilized to develop the regression model. Once the significant PC scores were obtained, the stepwise procedure (Stockton and Jacoby, 1976) of selecting best predictors was followed. If the F value was not greater than 3, the value was not used in the calibration equation.

3.3.5 Reconstruction of Unimpaired Streamflow and Drought

3.3.5.1 Reconstruction of Unimpaired Streamflow

Once the performance of PLSR was compared with STPCR and STLR based on a calibration period of the initial 35 years of the overlapping period, all the 17 unimpaired streamflow stations were reconstructed utilizing PLSR. While reconstructing the water year streamflow for the last 500 years, each water year streamflow was calibrated and reconstructed utilizing the tree-ring chronologies ending in 1960, 1970 or 1980 in order to have a longer overlapping period. There were six (6) stations (Uncampahgre River at Colona, Yampa River at Steamboats, Yampa River near Maybell, White River near Meeker, Bright Angel Creek near Grand Canyon and Salt River near Roosevelt) that were calibrated and reconstructed considering all the tree-ring chronologies ending after 1960.
Similarly, six (6) stations (East River at Almont, Whiterocks River near Whiterocks, Animas River at Durango, Animas River at Farmington, Virgin River at Littlefield and Gila River near Gila) that were calibrated and reconstructed utilizing the standard tree-ring chronologies ending after 1970. The remaining 5 stations (Tomichi Creek at Gunnision, Lake Fork at Gateview, Smith Fork near Crawford, Fish Creek near Reservoir and San Francisco River at Clifton) were calibrated and reconstructed utilizing the tree-ring chronologies ending after 1980. Similar to the procedure mentioned earlier, first the tree-ring chronologies were prescreened based on the correlation criterion (> 95% confidence level of positive and negative correlations) during the calibration period. Once the pool of predictors based on the correlation criterion was obtained, predictors were ranked based on their absolute value of correlation. Then PLSR was run eliminating each least correlated tree-ring chronologies until the minimum CVSE was obtained. As it was defined earlier, the best model had the lowest CVSE. Since, streamflow stations did not have significant autocorrelation (see Appendix-A2); the effect of autocorrelation with the CVSE was not taken into an account in this study.

3.3.5.2 Adjustment of Reconstructed Streamflow with Respect to Observed Streamflow Data

The variability between the observed and reconstructed water year streamflows was found to be different. That made it difficult to adjoin the two time series for the drought analysis. In order to have a closer fit of the reconstructions with the observed data, reconstructed water year streamflow were rescaled to the observed variance using the following formula (Timilsena et al., 2007):

$$\hat{x}_i^* = \left(\frac{\hat{x}_i - \hat{x}}{\sigma}\right)\sigma + \hat{x}$$  

(3-7)
where, $\hat{x}_i$ is the original reconstructed water year streamflow, $\hat{x}$ and $\hat{\sigma}$ are the reconstructed mean and standard deviation, $\sigma$ is the historical (observed) standard deviation, and $\hat{x}^*_i$ is the adjusted reconstructed water year streamflow. This adjustment was performed for all the reconstructed 17 water year streamflow data. Though with equation (3-7), the variance of the reconstructed data was improved, there are other biological and physical factors associated with the tree growth such as detrending, autocorrelation and biological growth which have not been addressed in this analysis. After adjoining the observed and reconstructed water year time series, the last 500 years (1500-1999) water year streamflow time series for each unimpaired station were obtained. Further analysis of the drought was performed based on this rescaled (adjusted) reconstructed value plus historical (observed) water year streamflow.

3.3.5.3 Spatial and Temporal Variability of Drought

The spatial and temporal variability of drought was analyzed by dividing the streamflow data into epochal time periods. In this paper, drought years were defined as the years that have less than 10th percentile water year flow for the last 500 years. Once the drought years were identified for last 500 years, the number of these drought years (<10 percentile water year flow) were separated for each 100 year period. Then, the number of drought years (out of 100) was plotted for each station for each 100 year period (1500-1599, 1600-1699, 1700-1799, 1800-1899 and 1900-1999) respectively. Further, after finding the drought years for each 100 year period, the average drought deficit (sum of the difference between long-term mean and water year streamflow of drought years divided by the number of drought years) for each 100 year period was plotted for each streamflow station. The results of this analysis provided spatial as well as

53

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
temporal information of droughts over the different 100 year periods in the record of the region. In addition, 3-yr, 5-yr and 10-yr moving averages were also found for each streamflow station in order to find the most severe drought year for the last 500 years based on the water year streamflow volume.

3.4 Results

3.4.1 Performance Comparisons of PLSR, STPCR and STLR

Table 3-2 summarizes the reconstruction performance of the three different methods of reconstructions (PLSR, STPCR and STLR) during the calibration period of the initial 35 years of observed water year data. PLSR performed better than STPCR in all 17 streamflow stations based on the minimum CVSE criterion. Similarly, comparing PLSR with STLR, PLSR performed better than STLR in 11 out of 17 stations. Calibration plots based on PLSR, STPCR and STLR for three different streamflow stations (Lakefork at Gateview, Virgin River at Littlefield and Salt River near Roosevelt) are presented in Figure 3-2. In general, the three models in Figure 3-2 produced similar features; however, the performance statistics shown in Table 3-2 suggest that PLSR was able to capture the observed peak (lowest and highest) more effectively.
Table 3-2: Performance comparison of PLSR with STPCR and STLR for 17 unimpaired streamflow stations in the Colorado River basin. CVSE is the cross-validation standard error in 100*km³/water year for streamflow. * indicates the minimum CVSE value for each streamflow station.

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>Station Name</th>
<th>Calibration Period (35 years)</th>
<th>Stepwise Principal Component Regression</th>
<th>Partial Least Square Regression</th>
<th>Stepwise Multiple Linear Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EAST RIVER AT ALMONT</td>
<td>1936-1970</td>
<td>44.07% Var. 6.74 CVSE (100*Km³/year)</td>
<td>51.48% Var. 6.57 CVSE (100*Km³/year)</td>
<td>49.32% Var. 6.43 CVSE (100*Km³/year)</td>
</tr>
<tr>
<td>2</td>
<td>TOMICHI CREEK AT GUNNISON</td>
<td>1939-1973</td>
<td>52.44% Var. 4.92 CVSE (100*Km³/year)</td>
<td>56.56% Var. 4.59 CVSE (100*Km³/year)</td>
<td>62.64% Var. 4.44 CVSE (100*Km³/year)</td>
</tr>
<tr>
<td>3</td>
<td>LAKE FORK AT GATEVIEW</td>
<td>1939-1973</td>
<td>55.85% Var. 4.33 CVSE (100*Km³/year)</td>
<td>59.46% Var. 4.18* CVSE (100*Km³/year)</td>
<td>59.39% Var. 4.35 CVSE (100*Km³/year)</td>
</tr>
<tr>
<td>4</td>
<td>SMITH FORK NEAR CRAWFORD</td>
<td>1937-1971</td>
<td>53.25% Var. 1.10 CVSE (100*Km³/year)</td>
<td>60.38% Var. 1.03* CVSE (100*Km³/year)</td>
<td>62.5% Var. 1.05 CVSE (100*Km³/year)</td>
</tr>
<tr>
<td>5</td>
<td>UNCAMPAHGRE RIVER AT COLONA</td>
<td>1914-1948</td>
<td>54.30% Var. 5.28 CVSE (100*Km³/year)</td>
<td>60.84% Var. 4.74* CVSE (100*Km³/year)</td>
<td>65.55% Var. 4.89 CVSE (100*Km³/year)</td>
</tr>
<tr>
<td>6</td>
<td>YAMPA RIVER AT STEAMBOAT</td>
<td>1912-1946</td>
<td>69.34% Var. 7.98 CVSE (100*Km³/year)</td>
<td>98.25% Var. 6.13* CVSE (100*Km³/year)</td>
<td>67.11% Var. 6.98 CVSE (100*Km³/year)</td>
</tr>
<tr>
<td>7</td>
<td>YAMPA RIVER NEAR MAYBELL</td>
<td>1916-1950</td>
<td>49.99% Var. 29.72 CVSE (100*Km³/year)</td>
<td>59.81% Var. 26.30 CVSE (100*Km³/year)</td>
<td>81.13% Var. 21.61* CVSE (100*Km³/year)</td>
</tr>
<tr>
<td>8</td>
<td>WHITEROCKS RIVER NEAR</td>
<td>1931-1965</td>
<td>52.74% Var. 2.20 CVSE (100*Km³/year)</td>
<td>58.15% Var. 2.03 CVSE (100*Km³/year)</td>
<td>72.67% Var. 1.84* CVSE (100*Km³/year)</td>
</tr>
<tr>
<td>9</td>
<td>WHITE RIVER NEAR MEEKER</td>
<td>1911-1945</td>
<td>68.62% Var. 8.51 CVSE (100*Km³/year)</td>
<td>73.4% Var. 7.75* CVSE (100*Km³/year)</td>
<td>60.58% Var. 8.83 CVSE (100*Km³/year)</td>
</tr>
<tr>
<td>10</td>
<td>FISH CREEK NEAR RESERVOIR</td>
<td>1940-1974</td>
<td>44.03% Var. 1.19 CVSE (100*Km³/year)</td>
<td>55.54% Var. 1.15 CVSE (100*Km³/year)</td>
<td>53.37% Var. 1.12* CVSE (100*Km³/year)</td>
</tr>
<tr>
<td>11</td>
<td>ANIMAS RIVER AT DURANGO</td>
<td>1929-1963</td>
<td>76.49% Var. 13.59 CVSE (100*Km³/year)</td>
<td>80.84% Var. 12.09* CVSE (100*Km³/year)</td>
<td>78.74% Var. 12.32 CVSE (100*Km³/year)</td>
</tr>
<tr>
<td>12</td>
<td>ANIMAS RIVER AT FARMINGTON</td>
<td>1932-1966</td>
<td>76.04% Var. 16.79 CVSE (100*Km³/year)</td>
<td>79.99% Var. 14.75* CVSE (100*Km³/year)</td>
<td>79.92% Var. 15.58 CVSE (100*Km³/year)</td>
</tr>
<tr>
<td>13</td>
<td>BRIGHT ANGEL CREEK NEAR</td>
<td>1925-1959</td>
<td>51.56% Var. 1.11 CVSE (100*Km³/year)</td>
<td>57.62% Var. 1.02* CVSE (100*Km³/year)</td>
<td>59.48% Var. 1.04 CVSE (100*Km³/year)</td>
</tr>
<tr>
<td>14</td>
<td>VIRGIN RIVER AT LITTLEFIELD</td>
<td>1931-1965</td>
<td>50.50% Var. 7.53 CVSE (100*Km³/year)</td>
<td>56.86% Var. 7.14* CVSE (100*Km³/year)</td>
<td>54.27% Var. 7.29 CVSE (100*Km³/year)</td>
</tr>
<tr>
<td>15</td>
<td>GILA RIVER NEAR GILA</td>
<td>1930-1964</td>
<td>53.08% Var. 5.20 CVSE (100*Km³/year)</td>
<td>59.94% Var. 4.69 CVSE (100*Km³/year)</td>
<td>70.84% Var. 4.54* CVSE (100*Km³/year)</td>
</tr>
<tr>
<td>16</td>
<td>SAN FRANCISCO RIVER AT CLIFTON</td>
<td>1937-1971</td>
<td>35.14% Var. 10.15 CVSE (100*Km³/year)</td>
<td>44.26% Var. 9.56* CVSE (100*Km³/year)</td>
<td>44.44% Var. 9.86 CVSE (100*Km³/year)</td>
</tr>
<tr>
<td>17</td>
<td>SALT RIVER NEAR ROOSEVELT</td>
<td>1915-1949</td>
<td>65.22% Var. 46.49 CVSE (100*Km³/year)</td>
<td>82.87% Var. 43.82* CVSE (100*Km³/year)</td>
<td>48.27% Var. 47.09 CVSE (100*Km³/year)</td>
</tr>
</tbody>
</table>
Figure 3-2: Calibration comparison of PLSR with STPCR and STLR for water year streamflow (100*Km³) volume for three different stations (a) Lake Fork at Gateview (b) Virgin River at Littlefield and (c) Salt River near Roosevelt. The solid gray line represents the observed annual water year data. The short-dashed horizontal line with cross symbol indicates the STLR reconstructed water year streamflow volume (100*Km³). The short-dashed line with square symbol indicates the STPCR reconstructed water year streamflow volume (100*Km³). The solid line with circle symbol indicates the PLSR reconstructed water year streamflow (100*Km³).
3.4.2 PLSR Reconstruction of 17 Streamflow Stations

Table 3-3 summarizes the performance statistics of the 17 unimpaired streamflow reconstructions utilizing the PLSR. The annual average and the 10-year moving average of water year streamflow (adjusted reconstructed plus observed water year streamflow) of three different unimpaired streamflow stations (Lake Fork at Gateview, Virgin River at Littlefield and Salt River near Roosevelt) are presented in Figure 3-3. In some of the stations (Figure 3-3(a) and 3-3(b)) the water year streamflow during 1500-1900 was drier than the 1900-1999 period. However, in Figure 3-3(c), the 1900’s do not appear to be wetter than the remaining record. This demonstrates some of the spatial variability in the basin which is further explored in the next section.
Table 3-3: Reconstruction performance of 17 unimpaired streamflow stations using PLSR. CVSE is the cross-validation standard error in 100*km³/water year

<table>
<thead>
<tr>
<th>S. N.</th>
<th>Station</th>
<th>% Var.</th>
<th>CVSE</th>
<th>Cal. Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EAST RIVER AT ALMONT</td>
<td>51.48</td>
<td>6.57</td>
<td>1936-1970</td>
</tr>
<tr>
<td>2</td>
<td>TOMICHI CREEK AT GUNNISON</td>
<td>56.11</td>
<td>4.36</td>
<td>1939-1980</td>
</tr>
<tr>
<td>3</td>
<td>LAKE FORK AT GATEVIEW</td>
<td>65.22</td>
<td>3.91</td>
<td>1939-1980</td>
</tr>
<tr>
<td>4</td>
<td>SMITH FORK NEAR CRAWFORD</td>
<td>61.36</td>
<td>0.93</td>
<td>1937-1980</td>
</tr>
<tr>
<td>5</td>
<td>UNCAMPAGHRE RIVER AT COLONA</td>
<td>72.59</td>
<td>4.82</td>
<td>1914-1960</td>
</tr>
<tr>
<td>6</td>
<td>YAMPA RIVER AT STEAMBOAT SPRINGS</td>
<td>67.86</td>
<td>6.53</td>
<td>1912-1960</td>
</tr>
<tr>
<td>7</td>
<td>YAMPA RIVER NEAR MAYBELL</td>
<td>68.37</td>
<td>24.97</td>
<td>1916-1960</td>
</tr>
<tr>
<td>8</td>
<td>WHITEROCKS RIVER NEAR WHITEROCKS</td>
<td>56.52</td>
<td>2.017</td>
<td>1931-1970</td>
</tr>
<tr>
<td>9</td>
<td>WHITE RIVER NEAR MEEKER</td>
<td>64.04</td>
<td>7.86</td>
<td>1911-1960</td>
</tr>
<tr>
<td>10</td>
<td>FISH CREEK NEAR RESERVOIR</td>
<td>59.30</td>
<td>1.12</td>
<td>1940-1980</td>
</tr>
<tr>
<td>11</td>
<td>ANIMAS RIVER AT DURANGO</td>
<td>79.17</td>
<td>12.55</td>
<td>1929-1970</td>
</tr>
<tr>
<td>12</td>
<td>ANIMAS RIVER AT FARMINGTON</td>
<td>81.07</td>
<td>13.89</td>
<td>1932-1970</td>
</tr>
<tr>
<td>13</td>
<td>BRIGHT ANGEL CREEK NEAR GRAND CANYON</td>
<td>53.77</td>
<td>1.05</td>
<td>1925-1960</td>
</tr>
<tr>
<td>14</td>
<td>VIRGIN RIVER AT LITTLEFIELD</td>
<td>48.52</td>
<td>7.76</td>
<td>1931-1970</td>
</tr>
<tr>
<td>15</td>
<td>GILA RIVER NEAR GILA</td>
<td>48.90</td>
<td>5.428</td>
<td>1930-1970</td>
</tr>
<tr>
<td>16</td>
<td>SAN FRANCISCO RIVER AT CLIFTON</td>
<td>39.40</td>
<td>13.11</td>
<td>1937-1980</td>
</tr>
<tr>
<td>17</td>
<td>SALT RIVER NEAR ROOSEVELT</td>
<td>90.12</td>
<td>39.78</td>
<td>1915-1960</td>
</tr>
</tbody>
</table>
Figure 3-3: PLSR reconstructed water year streamflow (100*km$^3$) of (a) Lake Fork at Gateview (b) Virgin River at Littlefield and (c) Salt River near Roosevelt. White River. Thin short-dashed line indicates the annual water year volume (100*km$^3$). The thick solid line indicates the 10-year moving average. Horizontal medium-dashed line indicates the long-term water year mean.
3.4.3 Spatial and Temporal Variability of Drought

Figure 3-4 summarizes the spatial and temporal variability of the drought based on the 17 unimpaired streamflow stations in the Colorado River basin. The most number of droughts occurred during the 1500’s, 1700’s and 1800’s respectively. Compared to other 100 year periods, the 1900’s had the least number of drought years in the region (Upper and Lower basin).

Similarly, Figure 3-5 also summarizes the spatial and temporal distribution of drought based on the average drought deficit. Out of 17 stations, 13 stations had a maximum average drought deficit during 1500’s, three (3) stations had a maximum average drought deficit during 1900’s, and one (1) station had a maximum average drought deficit during 1800’s. On the other side, 13 stations had minimum average drought deficit during 1700’s, three (3) stations had minimum average drought deficit during 1900’s and one (1) station had minimum average drought deficit during 1600’s. Even though, most of the stations had the maximum drought deficit during the 1500’s and minimum drought deficit during the 1700’s, the magnitude of average drought deficit did not vary as much as the number of drought years.
Figure 3-4: Number of drought years for each 100 year period (1500-1599, 1600-1699, 1700-1799, 1800-1899 and 1900-1999) for all 17 unimpaired streamflow stations in the basin. The black dark circle indicates the unimpaired streamflow stations in the Colorado River basin. Shaded vertical bar indicates the number of drought years in each 100 year period. The symbol ** indicates the 100 year period in which the data were not enough to perform the drought analysis.
Figure 3-5: Average drought deficit (100*Km³/year) for each 100 year period (1500-1599, 1600-1699, 1700-1799, 1800-1899 and 1900-1999) for all 17 unimpaired streamflow stations in the Colorado River basin. The black dark circle indicates the unimpaired streamflow stations. Shaded vertical bar indicates the average drought deficit (100*Km³/year) in each 100 year period. The symbol ** indicates the epoch in which data were not enough to do the drought analysis.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Lastly, the 3-year moving average, 5-year moving average and 10-year moving average of each station water year streamflow volume (adjusted reconstructed plus observed water year streamflow) were obtained, and the year that has the lowest streamflow volume for each moving average in each streamflow station was found. Results are presented in the Table 3-4.
Table 3-4: The most severe drought years based on the 3-year, 5-year and 10-year moving average of last 500 years of water year streamflow for 17 reconstructed unimpaired streamflow stations in the Colorado River basin

<table>
<thead>
<tr>
<th>No.</th>
<th>STATION NAME</th>
<th>3-Year Moving Average</th>
<th>5-Year Moving Average</th>
<th>10-Year Moving Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EAST RIVER AT ALMONT</td>
<td>1585</td>
<td>1883</td>
<td>1671</td>
</tr>
<tr>
<td>2</td>
<td>TOMICHI CREEK AT GUNNISON</td>
<td>1847</td>
<td>1883</td>
<td>1782</td>
</tr>
<tr>
<td>3</td>
<td>LAKE FORK AT GATEVIEW</td>
<td>1847</td>
<td>1883</td>
<td>1782</td>
</tr>
<tr>
<td>4</td>
<td>SMITH FORK NEAR CRAWFORD</td>
<td>1847</td>
<td>1883</td>
<td>1782</td>
</tr>
<tr>
<td>5</td>
<td>UNCOMPAHGRE RIVER AT COLONA</td>
<td>1847</td>
<td>1883</td>
<td>1782</td>
</tr>
<tr>
<td>6</td>
<td>YAMPA RIVER AT STEAMBOAT SPRINGS</td>
<td>1847</td>
<td>1594</td>
<td>1593</td>
</tr>
<tr>
<td>7</td>
<td>YAMPA RIVER NEAR MAYBELL</td>
<td>1847</td>
<td>1585</td>
<td>1592</td>
</tr>
<tr>
<td>8</td>
<td>WHITEROCKS RIVER NEAR WHITEROCKS</td>
<td>1585</td>
<td>1587</td>
<td>1588</td>
</tr>
<tr>
<td>9</td>
<td>WHITE RIVER NEAR MEEKER</td>
<td>1847</td>
<td>1587</td>
<td>1592</td>
</tr>
<tr>
<td>10</td>
<td>FISH CREEK ABOVE RESERVOIR NEAR SCOFIELD</td>
<td>1847</td>
<td>1587</td>
<td>1593</td>
</tr>
<tr>
<td>11</td>
<td>ANIMAS RIVER AT DURANGO</td>
<td>1585</td>
<td>1585</td>
<td>1587</td>
</tr>
<tr>
<td>12</td>
<td>ANIMAS RIVER AT FARMINGTON</td>
<td>1585</td>
<td>1585</td>
<td>1587</td>
</tr>
<tr>
<td>13</td>
<td>BRIGHT ANGEL CREEK NEAR GRAND CANYON</td>
<td>1585</td>
<td>1585</td>
<td>1585</td>
</tr>
<tr>
<td>14</td>
<td>VIRGIN RIVER AT LITTLEFIELD</td>
<td>1585</td>
<td>1585</td>
<td>1587</td>
</tr>
<tr>
<td>15</td>
<td>GILA RIVER NEAR GILA</td>
<td>1585</td>
<td>1670</td>
<td>1585</td>
</tr>
<tr>
<td>16</td>
<td>SAN FRANCISCO RIVER AT CLIFTON</td>
<td>1544</td>
<td>1670</td>
<td>1782</td>
</tr>
<tr>
<td>17</td>
<td>SALT RIVER NEAR ROOSEVELT</td>
<td>1587</td>
<td>1588</td>
<td>1587</td>
</tr>
</tbody>
</table>
Based on the 3-year moving average of water year streamflow volume, 8 stations indicated that the late 1500’s had the lowest water year streamflow volume, but the other 8 stations indicated that the mid 1800’s had the least water year streamflow volume. Further, considering 5-year moving average of water year streamflow volume, 10 stations indicated that the late 1500’s had the lowest streamflow volume, but 5 other stations indicated that the late 1800’s had the lowest streamflow volume. Considering the 10-year moving average, 11 stations indicated the lowest water year volume of streamflow in the late 1500’s, and the other 5 stations indicated the lowest streamflow volume in the mid 1700’s. In summary, considering the 3-year, 5-year and 10-year moving average, most of the stations showed that the drought was most severe at the end of 1500’s, which is consistent to the studies by Hidalgo et al. (2000), Stockton and Jacoby (1976), Piechota et al. (2004) and Timilsena et al. (2007).

3.5 Conclusions

The PLSR methodology provided improved results when compared to STPCR and STLR for the reconstruction of streamflow for the unimpaired streamflow stations in the Colorado River basin. Comparing the results of PLSR with STPCR and STLR, most of the unimpaired streamflow stations demonstrated that PLSR was better in terms of minimum CVSE and improved R². PLSR demonstrated a possible improvement in the statistical methodology used to develop reconstructions; however, the importance of the tree-ring species collected and used in the analysis cannot be understated. Further, the identification of moisture-sensitive trees adjacent to the watershed contributing to the streamflow station being reconstructed could increase reconstruction skill and capture the...
regional scale climate variability. Although primarily applied to chemometrics, PLSR has numerous applications in paleo-hydrology and water resources.

Although, tree-ring chronologies as predictors of streamflow could significantly capture the high-frequency climate variability, additional inclusion of low-frequency proxies (predictors) such as lake sediments would help to integrate the low-frequency climate variability (Moberg et al., 2005). Lastly, there is a possibility of further improving the reconstruction skill in order to capture the low-frequency variability by utilizing the regional curve standardized chronologies as predictors (Esper et al., 2002; Esper et al., 2003) in the reconstruction model.
CHAPTER 4

REGIONALIZATION AND RECONSTRUCTION OF SNOW WATER EQUIVALENT IN THE UPPER COLORADO RIVER BASIN

4.1 Introduction

Mountains in the western U.S. hold a vast amount of snowpack which provide 50%-80% of the water supply in the region (NRCS, 2006). Although snowpack is limited to only 15% of the land area of the Colorado River Basin (Doesken and Stanton, 1991), most of the streamflow in the Colorado River is derived from the melting of the snowpack from April through July (Kuhn, 2005). Trends in the snowpack of the basin are important since it is the major source of water in the basin.

Adequate knowledge of spatial distribution and regionalization of climate variables (snowpack, streamflow) are important to evaluate the trend of climate changes in the region. Eigen techniques (Principal component, Factor analysis) have been extensively used in the meteorological community for data reduction, grouping of variables and identification of coherent modes in atmospheric fields (e.g. Richman, 1986, Woodhouse, 2003). Piechota et al. (1997) used the S-mode principal component analysis and cluster analysis to isolate the group of streamflow stations that have similar El Niño-Southern Oscillation responses in streamflow.
Existing snow records cover much of the basins; however, they are inadequate for assessing long-term variations of drought due to the short periods of available data. A longer record of snowpack is required to characterize and to obtain the drought variability in a long-term perspective. Tree-rings have been extensively used to extend climate variables such as streamflow, temperature and precipitation; (e.g. Cleaveland, 2000; Biondi et al., 1999; Gray et al., 2004 and Touchan et al., 2003) however, few studies (Woodhouse et al., 2003 and Laroque and Smith, 2005) have reconstructed snowpack using tree-ring chronologies as predictors. Common approaches of reconstructing hydrologic variables utilizing tree-rings are multiple linear regression (e.g. Touchan et al., 1999; Diaz et al., 2001; Woodhouse, 2001; and Gray et al. 2004) and principal component regression (Stockton and Jacoby, 1976; Cook and Jacoby, 1983; Garen, 1992; Meko et al., 2001; Hidalgo et al., 2000; and Woodhouse et al., 2006). An alternate approach is Partial Least Square Regression (PLSR) which was developed in the late 1960’s (Wold, 1966) and gained importance in the field of chemistry during the 1970’s (Gerlach et al., 1979). PLSR has been used for temperature reconstruction (Kalela-Brundin, 1999). PLSR incorporates the features of principal component regression and multiple linear regressions (Ablitt et al., 2004) in such a way that the relationship between these two predictors and predictands is as strong as possible.

The goal of this study is to regionalize snow course stations located in the Upper Colorado River Basin (UCRB) and reconstruct the regional snow pack in order to evaluate the regional drought scenarios for the last 475 years. The main contribution of this research is to apply the technique of principal component analysis and cluster analysis to regionalize the snowpack that will help to understand the spatial distribution
of the snowpack regimes in the basin. Identification of regional distribution of the snowpack is important in order to evaluate the spatial distribution of droughts in the region. The second contribution of this research is to reconstruct the regional snowpack with PLSR in order to understand the long-term regional spatial and temporal variability of drought in the basin using a short calibration and verification period. PLSR reconstruction based on cross validation also makes a contribution since a comparatively shorter span of available historical snowpack data is available for the separate calibration and verification. The third contribution of this research is the comparison of the regional reconstruction performance with the individual stations performance. The final contribution of this research is to characterize the droughts in terms of duration for the last 475 years (1500-1975) using three different moving averages (3-year, 5-year and 10-year) of reconstructed regional snowpack, and comparing it with the results based on reconstructed unimpaired water year streamflow.

4.2 Data Sources

4.2.1 Snow Course Data

A snow course is a permanent site where the snow depth and snow water equivalent (SWE) are measured manually by field personnel. SWE is the amount of water contained within the snowpack (NRCS, 2006). Measurements are usually taken around the first of the month during the winter and spring. Based on Natural Resources Conservation Service (NRCS) guidelines, each snow course is approximately 1,000 feet long and protected from the wind in the region of interest. The SWE on or nearest to April 1 is considered as the April 1 SWE in this research. The April 1 SWE is a good
indicator of the water content in the maximum seasonal snowpack and for the water supply for coming months (Woodhouse, 2003). In the study presented here, SWE is regarded as the potential climate variable (indicator). The snow course data were collected from the United States Department of Agriculture (USDA), Natural Resources Conservation Service (NRCS) website (http://www.wcc.nrcs.usda.gov/snowcourse/) and included the information on station ID, station location, and monthly SWE in inches. The snow course stations that had missing values on April 1 SWE were not used in this research.

Altogether, there are 350 snow course stations in the Colorado River basin. Out of these 350 snow stations, only 40 snow course stations in the UCRB have a consistent April 1 SWE readings from 1940 to 1980. This period was selected based on its overlapping period with the tree-ring data (see Section 4.2.2). Out of these 40 snow course stations, one station has several zero values; hence it was eliminated from the analysis.
Figure 4-1: Location map displaying 39 snow course stations and 17 standard tree-ring chronologies. Snow course stations are indicated by small black dark triangle. Standard tree-ring chronologies are indicated by black dark circle.

Therefore these remaining 39 stations were utilized in this research. These 39 stations are illustrated in Figure 4-1 and a detailed description of these stations is in Appendix A-4.

4.2.2 Tree-Ring Data

Tree-ring data were obtained from the International Tree Data Bank website (http://www.ncdc.noaa.gov/paleo/treering.html) maintained by the National Oceanic and Atmospheric Administration (NOAA), World Data Center for Paleoclimatology. In this study, the standard normal widths of site chronologies (hereinafter referred to as standard tree-ring chronologies) were used for the April 1 SWE reconstructions. The Upper Colorado River basin includes 57 standard tree-ring chronologies; and the Lower
Colorado basin contains 54, of which 10 in the Upper and 7 in the Lower have at least 475 years of record before 1975 (Table 4-1).
Table 4-1: Summary of the standard tree-ring chronologies utilized for the regional April 1 SWE reconstruction

<table>
<thead>
<tr>
<th>S. N.</th>
<th>ITRB Site Name</th>
<th>Lat.</th>
<th>Long.</th>
<th>El. (m)</th>
<th>Tree-ring Duration</th>
<th>Auto correl (Lag1)</th>
<th>Tree Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Monarch Lake</td>
<td>40.10</td>
<td>-105.73</td>
<td>2,621</td>
<td>1430-1987</td>
<td>0.58</td>
<td>Ponderosa Pine</td>
</tr>
<tr>
<td>2</td>
<td>Dolores</td>
<td>37.58</td>
<td>-108.55</td>
<td>2,195</td>
<td>1457-1978</td>
<td>0.57</td>
<td>Douglas Fir</td>
</tr>
<tr>
<td>3</td>
<td>Dolores</td>
<td>37.58</td>
<td>-108.55</td>
<td>2,195</td>
<td>1270-1978</td>
<td>0.26</td>
<td>Pinyon Species</td>
</tr>
<tr>
<td>4</td>
<td>Spruce Canyon</td>
<td>37.18</td>
<td>-108.48</td>
<td>2,115</td>
<td>1373-1978</td>
<td>0.40</td>
<td>Douglas Fir</td>
</tr>
<tr>
<td>5</td>
<td>Hogback Rampart Hills</td>
<td>37.38</td>
<td>-108.12</td>
<td>3,230</td>
<td>1484-1994</td>
<td>0.14</td>
<td>Limber Pine</td>
</tr>
<tr>
<td>6</td>
<td>Pumphouse</td>
<td>39.95</td>
<td>-106.52</td>
<td>2,195</td>
<td>1320-1999</td>
<td>0.46</td>
<td>Pinyon Species</td>
</tr>
<tr>
<td>7</td>
<td>Land's End</td>
<td>39.00</td>
<td>-108.15</td>
<td>2,987</td>
<td>1135-2000</td>
<td>0.60</td>
<td>Douglas Fir</td>
</tr>
<tr>
<td>8</td>
<td>Paria Plateau</td>
<td>36.83</td>
<td>-112.05</td>
<td>1,860</td>
<td>1481-1975</td>
<td>0.28</td>
<td>Pinyon Species</td>
</tr>
<tr>
<td>9</td>
<td>Medicine Bow Peak</td>
<td>41.30</td>
<td>-107.70</td>
<td>3,150</td>
<td>1401-1983</td>
<td>0.51</td>
<td>Spruce Species</td>
</tr>
<tr>
<td>10</td>
<td>Wild Horse Ridge</td>
<td>39.42</td>
<td>-111.07</td>
<td>2,805</td>
<td>286-1985</td>
<td>0.41</td>
<td>Gt. basin B.C. Pine</td>
</tr>
<tr>
<td>11</td>
<td>San Francisco Peaks A</td>
<td>35.50</td>
<td>-111.67</td>
<td>3,535</td>
<td>548-1983</td>
<td>0.69</td>
<td>Rocky Mtn. B.C. Pin</td>
</tr>
<tr>
<td>12</td>
<td>Green Mountain</td>
<td>32.38</td>
<td>-110.68</td>
<td>2,194</td>
<td>1460-1986</td>
<td>0.50</td>
<td>Ponderosa Pine</td>
</tr>
<tr>
<td>13</td>
<td>Walnut Canyon</td>
<td>35.17</td>
<td>-111.52</td>
<td>2,057</td>
<td>1420-1987</td>
<td>0.51</td>
<td>Ponderosa Pine</td>
</tr>
<tr>
<td>14</td>
<td>Fort Grant Overlook</td>
<td>32.70</td>
<td>-109.92</td>
<td>2,896</td>
<td>1249-1991</td>
<td>0.56</td>
<td>White Pine</td>
</tr>
<tr>
<td>15</td>
<td>Snow Bowl San Francisco Peak</td>
<td>35.43</td>
<td>-110.20</td>
<td>3,150</td>
<td>1453-1983</td>
<td>0.58</td>
<td>Spruce Species</td>
</tr>
<tr>
<td>16</td>
<td>Reef of Rocks</td>
<td>32.45</td>
<td>-110.78</td>
<td>2,550</td>
<td>1321-1998</td>
<td>0.40</td>
<td>Douglas Fir</td>
</tr>
<tr>
<td>17</td>
<td>Reef of Rocks</td>
<td>32.70</td>
<td>-109.92</td>
<td>3,025</td>
<td>1464-1999</td>
<td>0.42</td>
<td>Douglas Fir</td>
</tr>
</tbody>
</table>
These 17 standard tree-ring chronologies were selected based on their available data over the last 500 years and sufficient overlapping period (35 years) with the snow course stations. These 17 standard tree-ring chronologies are located at different elevations ranging from 1860-3535 meters. Out of these 17 standard tree-species, 5 were Douglas Fir, 3 were Pinyon species (PIED), 3 were Ponderosa pine (PIPO), 2 were Spruce species (PCEN), and remaining 4 were Limber pine (PIFL), Great basin B.C. pine (PILO), White pine (PISF) and Rocky Mountain B.C. pine (PIAR).

4.2.3 Unimpaired Streamflow Data

It is important to utilize unimpaired streamflow data for the study of hydro-climatic variability because it provides an account of hydrologic responses to fluctuations in climate for the watershed (Slack et al., 1993). Unimpaired streamflow data were obtained from the United States Geological Survey (USGS), Hydro-Climate Data Network (HCDN) website (http://pubs.usgs.gov/wri/wri934076/1st_page.html ). Based on the HCDN website, there were 45 unimpaired streamflow stations in the UCRB. Water year streamflow volume of three unimpaired streamflow stations namely: (1) Yampa River near Maybell (2) Whiterocks River near Whiterocks and (3) Uncompahgre River at Colona were utilized in this research.

4.3 Methods

4.3.1 Regionalization

Climate regionalization is a useful technique that enables generalization of areas based on spatially and temporally varying climate parameters (Comrie and Glenn, 1998).
Principal component analysis and cluster analysis are broadly used procedures to obtain the coherent modes of various climate parameters.

4.3.1.1 Principal Component Analysis

Principal component analysis (PCA) is a widely used technique in meteorology and climatology (Baeriswyl and Rebetez, 1997) and sometimes used to reduce the size of climate datasets without losing climate information. Spatial regionalization based on principal components (eigenvectors) of climate variables is called S-mode PCA (Richman, 1986). S-mode PCA is a proven technique for climate regionalization (Knapp et al., 2002). In a procedure of regionalization of climate variables using S-mode PCA, first the eigenvalues and eigenvectors of the correlation or covariance matrix of the time series and loading matrix are computed. The loading matrix represents the correlation of the original variables with the principal components. The loading matrix (C) can be obtained using:

\[ C = A \ast \lambda^{1/2} \quad (4-1) \]

where \( A \) is an orthogonal matrix of the eigenvector of the correlation matrix \( R \) and \( \lambda \) is the diagonal matrix of eigenvalues of the correlation matrix \( R \). Based on the results of the loading matrix, the stations that are highly correlated with the particular principal components can be identified. In this research, truncated and varimax rotated principal component loading was used.

4.3.1.2 Cluster Analysis

Group average linkage cluster analysis is another method of identifying regions with similar climate characteristics (Piechota et al., 1997). The cluster analysis procedure first finds the distance between individual stations, or group of stations, using the
smallest Euclidean distance. The cluster is formed considering the smallest possible
distance, and then proceeds with another cluster considering the distance between the
new station and already formed cluster. The Euclidean distance \( d_{rs} \) is a measure of the
dissimilarity between two pairs of objects (clusters) \( x_r \) and \( x_s \), in an \( n \times p \) data matrix:

\[
d_{rs} = \sqrt{(x_r - x_s)(x_r - x_s)^T}
\]

(4-2)

where, \( n \) is the number of observations and \( p \) is number of variables. Fovell and
Fovell (1993) provided a detailed description of group average linkage cluster analysis.
The result of the principal component can be corroborated with the group average linkage
cluster analysis (Piechota et al, 1997; Baeriswyl and Rebetez, 1997).

4.3.1.3 Regional Standardized Composite of Consensus Snow Course Stations

In this study, snow course stations were regionalized using PCA method, and each
region was mapped based on the principal component loadings. The results from the PCA
loading map were compared with the results of cluster analysis. Once the regions based
on the PCA loading were identified, all the snow course stations within the region were
normalized. These normalized time series of each station within the region were
arithmetically averaged in order to obtain a composite time series for that particular
region. Regional standardized composite April 1 SWE (herein after referred as regional
composite April 1 SWE) series were used to characterize the regional variation of snow
in the UCRB.

4.3.2 April 1 Snow Water Equivalent Reconstruction

Partial Least Square Regression (PLSR) was utilized to reconstruct the regional
composite April 1 SWE. In the procedure, predictors (standard tree-ring chronologies)
including lag -1, 0 and +1 were first screened in order to eliminate the less significant
predictor variables. PLSR based on cross validation was utilized in order to obtain the best prediction model. The detail procedure is explained as follows.

4.3.2.1 Prescreening of Standard Tree-ring Chronologies

Prescreening of the standard tree-ring chronologies was based on the correlation between the regional composite April 1 SWE and standard tree-ring chronologies for the overlapping period (1940-1975). The chronologies (including lag -1, 0 and +1 years) that were positively and negatively correlated with the regional composite April 1 SWE with greater than 95% significance were selected from the prescreening, and were used as a pool of predictors for respective regional April 1 SWE reconstruction.

4.3.2.2 Partial Least Square Regression Procedure

PLSR utilizes the principal component scores of both (X) and (Y) in order to develop the regression model. Once an appropriate number of pools of predictors were obtained after prescreening procedure, a PLSR reconstruction model was developed and run. An important characteristic of PLSR is to identify the components from (Y) that are also related to (X) (Abdi, 2003).

Mathematically, PCA of the matrix (X) (i.e., the matrix of predictors or independent variables) decomposes (X) into a score matrix (T) times a loading matrix (P) and a residual (i.e., error) matrix (E) (Wold et al., 1987).

\[ X = T \cdot P' + E \quad (4-3) \]

Likewise, (Y) is decomposed into a score matrix (U) times a loading matrix (R) and a residual matrix (F).

\[ Y = U \cdot R' + F \quad (4-4) \]

These two equations (4-3 and 4-4) reflect the outer relations (Geladi and

77
Kowalski, 1986). The objective of the PLSR model is to minimize (F) preserving the correlation between (X) and (Y), referred to as the inner relation U (Geladi and Kowalski, 1986).

\[
U = B \times T + H \quad (4-5)
\]

Where (H) represents the error, (B) is a diagonal matrix explaining the correlation between (X) and (Y). When equation (4-5) is inserted into equation (4-4), a predictive relation for (Y) is developed where (F*) represents the error.

\[
Y = T \times R'B + F* \quad (4-6)
\]

Equation (4-6) is sometimes referred to as the mixed relation where (F*) is to be minimized (Geladi and Kowalski, 1986). To perform PLSR, several methods are available including the nonlinear iterative partial least squares (NIPALS) approach. NIPALS is advantageous due to calculation speed and simplicity (Wold et al., 1987), if the model is mainly concerned with the first few principal components. The NIPALS approach was utilized in this research.

Further, in order to determine the optimal number of latent variables, the minimum prediction residual sum of squares (PRESS) statistic approach (Malinowski, 2002) was utilized in this research. PRESS statistic is a cross validation measurement that determines the optimum number of components (latent variables) required in PLSR (Geladi and Kowalski, 1986).

4.3.2.3 Selection of Best Predictors

The significantly (95%) positive and significantly (95%) negative correlated tree-ring chronologies with the regional composite April 1 SWE for the calibration period (35 years of overlapping period) identified in the prescreening procedure were ranked based
on their absolute value (positive and negative) of correlation. PLSR was run for all the
significantly correlated ranked predictors. The PLSR was run repeatedly by eliminating
the least significantly correlated variable at each step until the cross validation standard
error (CVSE) was a minimum. The best predictors were the set of ranked tree-ring
chronologies after eliminating the least significantly correlated tree-ring chronologies
with the minimum CVSE value. In this research, CVSE based on Garen (1992), Hidalgo
et al. (2000) and Timilsena et al. (2007) was utilized in order to evaluate the performance
of the model. The CVSE is defined as:

\[ CVSE = \sqrt{\frac{\sum (y_i - \hat{y}_{i0})^2}{n - p}} \]  

(4-7)

where \( y_i \) is the observed streamflow for year \( i \); \( \hat{y}_i \) is the fitted response of the \( i \)th
year computed from the fit with the \( i \)th observation removed; \( n \) is the number of years in
the data set; and \( p \) is the optimal number of predictors (latent variables) in the regression
equation.

4.3.2.4 Reconstruction of April 1 SWE

Three regional composite April 1 SWE were reconstructed utilizing the PLSR
procedure. Each regional composite April 1 SWE was calibrated using the ranked
prescreened tree-ring chronologies (35 years of overlapping period), and then
reconstructed. The best PLSR model and corresponding best predictors were obtained
based on the minimum CVSE criterion. In addition to the reconstruction of three regional
composite April 1 SWE, a snow course station with the highest loading from each rotated
principal component (i.e. from each region) was also reconstructed in order to compare
the performance with the regional reconstruction.
4.3.2.5 Adjustment of Reconstructed Regional standardized SWE with Respect to Observed Regional standardized SWE

A common problem in using reconstructed data is adjoining time series, one observed and another reconstructed, that have different variability. In order to overcome this problem, the reconstructed regional composites April 1 SWE were rescaled to the observed variance using the following formula (Timilsena et al. 2007):

\[
\hat{x}_i^* = \left( \frac{\hat{x}_i - \hat{x}}{\hat{\sigma}} \right) \sigma + \hat{x}
\]  

(4-8)

where, \( \hat{x}_i \) is the original reconstructed regional composite April 1 SWE, \( \hat{x} \) and \( \hat{\sigma} \) are the reconstructed mean and standard deviation, \( \sigma \) is the historical (observed) standard deviation, and \( \hat{x}_i^* \) is the adjusted reconstructed regional composite April 1 SWE. This adjustment was performed for all three (3) reconstructed regional composite April 1 SWE. Although, the variance of the reconstructed data was improved using this purely statistical approach, there are other parameters associated with the physical factors and biological growth which were not taken in to account.

4.3.2.6 Reconstruction of Streamflow

Three unimpaired water year streamflow stations namely Yampa River near Maybell from Region 1, Whiterocks River near Whiterocks from Region 2, and Uncompahgre River at Colona from Region 3 were also reconstructed using the same approach of PLSR. As it was done in the regional April SWE reconstruction, available standard tree-ring chronologies in the Upper and Lower basin (including lag -1,0 and +1) were prescreened, and best predictors were obtained. Further, all three reconstructed
water year streamflow time series were rescaled to the observed variance using the same approach of adjusting reconstructed regional composite April 1 SWE (Section 4.3.2.5).

Drought analysis (section 4.3.3) was performed using the time series of adjoined observed and adjusted reconstructed data of regional composite April 1 SWE and water year streamflow.

4.3.3 Spatial and Temporal Variability of Drought

Drought was characterized by its duration using three reconstructed (observed plus adjusted reconstructed) regional April 1 SWE (hereafter referred as R1 for region 1, R2 for region 2 and R3 for region 3). The drought duration based on R1, R2 and R3 was also compared with the drought duration based on three reconstructed water year streamflow: (1) Yampa River near Maybell (S1) (2) Whiterocks River near Whiterocks (S2) and (3) Uncompahgre River at Colona (S3).

Drought was defined as an event with a value less than a threshold value in R1, R2, R3, S1, S2 and S3 respectively. The threshold in the case of R1, R2 and R3 was considered to be zero since the reconstruction was based on the standardized value of April 1 SWE data. Similarly in the case of streamflow (S1, S2 and S3), the threshold was considered to be the long-term mean (average of adjoined historical and adjusted reconstructed). The anomalies for R1, R2, R3, S1, S2 and S3 were obtained by subtracting the threshold value from the respective annual April 1 SWE and water year streamflow volume. A drought event was defined as a year that has a negative anomaly or consecutive years with negative anomalies. For example, if there was a one year negative anomaly in between two positive anomalies, it was defined as the drought event of one year duration. Similarly, if there were two consecutive years of negative anomalies in
between two positive anomalies, it was another drought event of two year duration. After obtaining the anomalies of each reconstructed variable (R1, R2, R3, S1, S2 and S3), the beginning of the drought year was identified as the year of the starting negative anomalies. Similarly, the drought ending year was defined when the anomaly returned to a positive value. The year elapsed between the beginnings and ending year of negative anomalies for each drought event was the drought duration for that particular drought event. Furthermore, in order to study the shorter and longer-term effects of droughts in the region, the drought duration was identified using three different moving averages (1-year, 5-year and 10-year) of R1, R2, R3, S1, S2 and S3.

4.4 Results

4.4.1 Regionalization

Figure 4-2 summarizes the regionalization of 39 stations based on the April 1 SWE data from 1940-1980 using two approaches; S-mode PCA and group average cluster analysis. S-mode PCA was performed on the 39*39 station correlation matrix. The truncation of the principal components was based on the eigenvalue criterion (eigenvalue>1). Based on the eigenvalue criterion, four components were retained. Then, varimax rotation on the truncated principal components was carried out. The four truncated rotated principal components incorporated 30.4%, 21.6%, 20.2% and 9.3% respectively. Once the rotated and truncated four principal component loadings for all the stations were plotted, contours (dotted line in Figure 4-2) of the 0.6 loading value for each principal component were plotted. This loading map showed the distinct four regions in the Upper Colorado River basin (1) Region 1 (Eastern part of the basin) (2)
Region 2 (Western Part of the basin) (3) Region 3 (Southern Part of the basin) and (4) Region 4 (Northern part of the basin).

Figure 4-2: Regionalization of snow course stations based on S-mode PCA and average linkage cluster analysis. The dash line indicates the region boundaries based on the 0.6 PC loading map for each truncated and rotated PCs. The symbols at each snow course station correspond to the 6 cluster solution from average cluster analysis.

Similarly, group average linkage cluster analysis was also carried out for the 39 stations. The symbol in Figure 4-2 indicates the different cluster on which each station belongs from the six-cluster solution. The six cluster solution was selected based on the cubic cluster criterion (CCC) and pseudo-$t^2$ statistics (Johnson, 1998).
From Figure 4-2, it was noted that the result of S-mode PCA corroborated the average cluster analysis. The 0.6 PC loading was subjectively selected to isolate the regions and this value of PC loading was able to identify four homogeneous regions in four different spatial locations of the basin. Further, regionalizing based on the 0.6 loading criterion was also able to capture the majority of the classification based on the cluster analysis. It is important to note that the classified regions were principally based on the rotated loading pattern of 0.6 PC loading and the cluster analysis corroborated the results of PC regionalization.

There were 16 stations inside the 0.6 loading contour of PC1 which was identified as Region 1 (eastern part of the basin). Similarly, there were six stations inside the 0.6 loading contours of PC2 which represents Region 2 (western part of the basin). Further, eight stations inside the 0.6 loading of PC3 represent the Region 3 (southern part of the basin). Finally, PC4 or Region 4 (northern part of the basin) included three stations. Six stations did not belong to any regions.

After regionalizing based on the 0.6 PCs loading map and comparing the result with the cluster analysis, each snow course station within the region was normalized and arithmetically averaged in order to develop a regional composite of April 1 SWE. These regional standardized values were considered as a drought indicator for the particular region.

4.4.2 April 1 Snow Water Equivalent Reconstruction

Table 4-2 and Figure 4-3 present a summary of PLSR reconstruction performance statistics of the three regional composite April 1 SWE and three individual snow course stations.
Table 4-2: Summary of PLSR reconstruction performance statistics. CVSE is the cross-validation standard error in value/water year.

<table>
<thead>
<tr>
<th>Snow Station Name</th>
<th>% Variance</th>
<th>CVSE</th>
<th>Calibration Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region 1, Eastern Region</td>
<td>41.28</td>
<td>0.67</td>
<td>1940-1975</td>
</tr>
<tr>
<td>Region 2, Western Region</td>
<td>45.27</td>
<td>0.73</td>
<td>1940-1975</td>
</tr>
<tr>
<td>Region 3, Southern Region</td>
<td>54.48</td>
<td>0.59</td>
<td>1940-1975</td>
</tr>
<tr>
<td>Region 1, Station 05K03</td>
<td>30.12</td>
<td>0.83</td>
<td>1940-1975</td>
</tr>
<tr>
<td>Region 2, Station 09J01</td>
<td>32.55</td>
<td>0.86</td>
<td>1940-1975</td>
</tr>
<tr>
<td>Region 3, Station 07M05</td>
<td>53.07</td>
<td>0.67</td>
<td>1940-1975</td>
</tr>
</tbody>
</table>
Figure 4-3: Calibration plot for PLSR reconstructed regional composite April 1 SWE for Region 1, 2 and 3. The solid gray line represents the observed regional composite April 1 SWE. The solid black line with circle symbol indicates the PLSR reconstructed regional composite April 1 SWE. The horizontal short-dash indicates the mean of historical regional composite April 1 SWE.
For reconstruction of the Region 1, Region 2 and Region 3 SWE, the coefficient of determination ($R^2$) values were 0.41, 0.45 and 0.54 respectively. As mentioned earlier, three individual snow course stations 05K03, 09J01 and 07M05 from each region (Region 1, Region 2 and Region 3) that had the highest loading value of 0.90, 0.85 and 0.84 respectively for each component (PC1, PC2 and PC3) were also standardized and reconstructed. The $R^2$ values for these stations (05K03, 09J01 and 07M05) were 0.30, and 0.32 and 0.53 respectively. Region 4 was eliminated from reconstruction because it had the least correlation with the tree-ring chronologies and none of the chronologies were suitable for use in reconstruction. Based on the summary of statistics shown in Table 4-2, the regional composite April 1 SWE was able to capture more variability as compared to individual stations from each region that had the highest loading.

In addition to the reconstruction of regional April 1 SWE, three unimpaired streamflow stations each from each region (identified by regionalization of snowpack) were also reconstructed. Altogether, there were 41 tree-ring chronologies in the basin (Upper and Lower) that ended after 1960 that were utilized in the case of Yampa River near Maybell (from Region 1) and Uncompahgre River at Colona (from Region 3). Similarly, there were 31 tree-ring chronologies in the basin (Upper and Lower) that ended after 1970 that were considered for the Whiterocks River near Whiterocks (from Region 2) reconstruction. The variability incorporated by the calibration/validation model of Yampa River near Maybell, Whiterocks River near Whiterocks and Uncompahgre River at Colona were 68.37%, 56.52% and 72.59% respectively. The improved skill of streamflow reconstruction was expected since the water year streamflow were better correlated with tree-rings chronologies. Streamflow correlates well with tree-rings
because it integrates hydrologic processes including precipitation, soil moisture, and evaporation rates (Knight, 2004).

4.4.3 Regional and Temporal Variability of Drought

The 3-year moving average, 5-year moving average and 10-year moving average of R1, R2 and R3 were compared with the 3-year, 5-year and 10-year moving average of S1, S2 and S3. Table 4-3, Table 4-4 and Figure 4-4, 4-5, 4-6 and 4-7 summarize the drought characteristics identified by R1, S1, R2, S2, R3 and S3 based on the three different moving averages (3-year, 5-year and 10-year).
Table 4-3: Drought duration ranking based on Regional Composite April 1 SWE and water year streamflow volume using three different moving average scenarios. R1 indicates Region 1, R2 indicates Region 2 and R3 indicates Region 3. S1 indicates Yampa River at Maybell. S2 indicates Whiterocks River near Whiterocks. S3 indicates Uncompahgre River at Colona. This is based on adjusted reconstructed plus historical data of regional composite April 1 SWE and water year streamflow volume respectively.

<table>
<thead>
<tr>
<th>Drought Variables</th>
<th>Moving Average (Years)</th>
<th>Rank #1 Drought Period</th>
<th>Rank #2 Drought Period</th>
<th>Rank #3 Drought Period</th>
<th>Maximum Drought Duration (Years)</th>
<th>Average Drought Duration (Years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>3</td>
<td>1622-1639</td>
<td>1933-1948</td>
<td>1704-1718</td>
<td>18</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1584-1605</td>
<td>1504-1522</td>
<td>1625-1640</td>
<td>22</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>1585-1611</td>
<td>1878-1899</td>
<td>1706-1726</td>
<td>27</td>
<td>10</td>
</tr>
<tr>
<td>S1</td>
<td>3</td>
<td>1581-1604</td>
<td>1773-1784</td>
<td>1502-1512</td>
<td>24</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1501-1685</td>
<td>1623-1640</td>
<td>1773-1786</td>
<td>25</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>1878-1910</td>
<td>1581-1609</td>
<td>1653-1676</td>
<td>33</td>
<td>11</td>
</tr>
<tr>
<td>R2</td>
<td>3</td>
<td>1700-1725</td>
<td>1584-1605</td>
<td>1773-1791</td>
<td>26</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1584-1610</td>
<td>1702-1726</td>
<td>1873-1893</td>
<td>27</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>1704-1733</td>
<td>1585-1611</td>
<td>1876-1894</td>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>S2</td>
<td>3</td>
<td>1573-1595</td>
<td>1774-1791</td>
<td>1872-1885</td>
<td>23</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1575-1596</td>
<td>1532-1549</td>
<td>1774-1791</td>
<td>22</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>1803-1832</td>
<td>1653-1679</td>
<td>1578-1601</td>
<td>30</td>
<td>13</td>
</tr>
<tr>
<td>R3</td>
<td>3</td>
<td>1573-1594</td>
<td>1762-1765</td>
<td>1872-1885</td>
<td>22</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1575-1595</td>
<td>1873-1890</td>
<td>1752-1765</td>
<td>21</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>1576-1609</td>
<td>1801-1832</td>
<td>1653-1677</td>
<td>34</td>
<td>13</td>
</tr>
<tr>
<td>S3</td>
<td>3</td>
<td>1887-1905</td>
<td>1772-1784</td>
<td>1872-1884</td>
<td>19</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1874-1906</td>
<td>1751-1770</td>
<td>1580-1596</td>
<td>33</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>1755-1790</td>
<td>1878-1910</td>
<td>1580-1607</td>
<td>36</td>
<td>10</td>
</tr>
</tbody>
</table>
Figure 4-4: The 3-year, 5-year and 10-year moving average of PLSR reconstructed regional composite April 1 SWE for Region 1, Region 2 and Region 3. The black dotted line indicates the reconstructed regional composite April 1 SWE for Region 1. The solid thick black line indicates the reconstructed regional composite April 1 SWE
for Region 2. Similarly, the thin black line indicates the reconstructed regional composite
April 1 SWE for Region 3.

Figure 4-4 summarizes the 3-year moving average, 5-year moving average and
10-year moving average plot of R1, R2 and R3. The plots showed similar patterns of
snow variability for most of the time in three different regions. It was also noted that the
reconstructed snow indicated the higher variability as compared to the instrumental
record of snow. In addition, the late 1500’s drought was prominently indicated by all the
three different moving average of all the regions.

Table 4-4 and Figure 4-5 summarize the drought identified based on the 3-year
moving average. Considering the 3-year moving average of R1, R2, R3, S1, S2 and S3
from 1500-1975, 42-49 drought events were identified with the average drought duration
ranging 5-6 years. The maximum drought duration indicated by this moving average was
18-26 years depending upon the regional April 1 SWE and water year streamflow. For
the R1 and S1 time series, there were 211 number of negative anamolies years (deficit
years) simultaneously indicated by R1 and S1, which is 81% of total number of deficit
years indicated by R1. There were 186 (77% of the total number of deficit years indicated
by R2) number of deficit years simultaneously indicated by R2 and S2. Similarly, there
were 201 (75% of total number of deficit years indicated by R3) number of deficit years
simultaneously indicated by R3 and S3. There were 119 number of deficit years using the
3-year moving average of R1, R2, R3, S1, S2 and S3 respectively. This demonstrates the
high degree of similarity in April 1 SWE and water year streamflow.

91

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 4-4: Drought characteristics based on three different moving average (3-year, 5-year and, 10-year) utilizing three regional composite April 1 SWE (R1, R2 and R3) and three different water year streamflow volume (S1, S2 and S3). Column (1) indicates the moving average, Column (2) indicates the number of drought events, Column (3) indicates the average drought duration, Column (4) indicates the maximum drought duration, and Column (5) indicates the number of drought years based on all snowpack region and streamflow stations considering three different moving averages. Similarly, Column (6) indicates the percentage of number of deficit years simultaneously indicated by R1 and S1 with respect to total number of deficit years indicated by R1, Column (7) indicates the percentage of number of deficit years simultaneously indicated by R2 and S2 with respect to total number of deficit years indicated by R2, and Column (8) indicates the percentage of number of deficit years simultaneously indicated by R3 and S3 with respect to total number of deficit years indicated by R3.

<table>
<thead>
<tr>
<th>Moving Av. (Years) (1)</th>
<th>Nos. of Drought Events (2)</th>
<th>Av. Drought Duration (years) (3)</th>
<th>Max. Drought Duration (years) (4)</th>
<th>Number of drought years using 3, 5 and 10-year moving average of R1, S1, R2, S2, R3 S3 (5)</th>
<th>Percent of (R1, S1) w.r.t. R1 (6)</th>
<th>Percent of (R2, S2) w.r.t. R2 (7)</th>
<th>Percent of (R3, S3) w.r.t. R3 (8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>42-59</td>
<td>5-6</td>
<td>18-26</td>
<td>119</td>
<td>81%</td>
<td>77%</td>
<td>75%</td>
</tr>
<tr>
<td>5</td>
<td>28-39</td>
<td>7-9</td>
<td>22-33</td>
<td>114</td>
<td>83%</td>
<td>78%</td>
<td>80%</td>
</tr>
<tr>
<td>10</td>
<td>19-27</td>
<td>10-13</td>
<td>27-36</td>
<td>106</td>
<td>80%</td>
<td>70%</td>
<td>68%</td>
</tr>
</tbody>
</table>
Figure 4-5: Histogram of drought duration identified by 3-year moving average for three different regions (R1, R2 and R3) and three different streamflow (S1, S2 and S3). R1, R2 and R3 indicate reconstructed regional composite April 1 SWE of region 1, 2 and 3 respectively. Similarly, S1, S2 and S3 indicate reconstructed water year streamflow volume of Yampa River near Maybell, Whiterocks River near Whiterock and Uncampahgre River at Colona. Horizontal widths of vertical black bars indicate the drought duration.

Table 4-4 and Figure 4-6 summarize the drought characteristics based on the 5-year moving average. The 5-year moving average indicated 28-39 number of drought events with the maximum drought duration ranging from 22-33 years depending upon the region (R1, R2 and R3) and water year streamflow (S1, S2 and S3). The average drought duration indicated by 5-year moving average was 7-9 years. In 475 years (1500-
1975) of R1 and S1 time series, there were 201 (83% of total number of deficit years indicated by R1) number of deficit years simultaneously indicated by R1 and S1. There were 184 (78% of total number of deficit years indicated by R2) number of deficit years simultaneously indicated by R2 and S2. Similarly, 204 (80% of total number of deficit years indicated by R3) number of deficit years were simultaneously indicated by R3 and S3. There were 114 number of deficit years that were identified with the 5-year moving average of R1, R2, R3, S1, S2 and S3 respectively.

Figure 4-6: Histogram of drought duration identified by the 5-year moving average for three different regions (R1, R2 and R3) and three different streamflow (S1, S2 and S3). R1, R2 and R3 indicate reconstructed regional composite April 1 SWE of region 1, 2 and 3 respectively. Similarly, S1, S2 and S3 indicate reconstructed water year streamflow volume of Yampa River near Maybell, Whiterocks River near Whiterock and
Uncampainge River at Colona. Horizontal widths of vertical black bars indicate the
drought durations.

Table 4-4 and Figure 4-7 summarize the drought years based on the 10-year
moving average. The 10-year moving average indicated 19-27 number of drought events
with the maximum drought duration of 27-36 years depending upon the regional
composite April 1 SWE and water year streamflow. This 10-year moving average
showed average drought duration to be 10-13 years. There were 210 (80% of total
number of deficit years indicated by R1) number of deficit years that were
simultaneously indicated by R1 and S1. There were 156 (70% of total number of deficit
indicated by R2) number of deficit years simultaneously indicated by R2 and S2.
Similarly, 199 (68% of total number of deficit years indicated by R3) number of deficit
years were simultaneously pointed out by R3 and S3. There were 106 number of deficit
years that were simultaneously noted using the 5-year moving average of R1, R2, R3, S1,
S2 and S3 respectively.

95

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 4-7: Histogram of drought duration identified by the 10-year moving average for three different regions (R1, R2 and R3) and three different streamflow (S1, S2 and S3). R1, R2 and R3 indicate reconstructed regional composite April 1 SWE of region 1, 2 and 3 respectively. Similarly, S1, S2 and S3 indicate reconstructed water year streamflow volume of Yampa River near Maybell, Whiterocks River near Whiterock and Uncampahgre River at Colona. Vertical black bars indicate the drought duration. Horizontal widths of vertical black bars indicate the drought durations.

In summary, the eastern part of the UCRB (Region 1) had snowfall and streamflow deficit for a period of 50-55% of the last 475 year (1500-1975). Similarly, in the western part of the basin (Region 2), 47-52% of last 475 (1500-1975) year had snowfall and streamflow deficit. Likewise in the southern part of the basin (Region 3), 48-61% of last 475 year had snowfall and streamflow deficit. There were 51 (about 11% of total periods) deficit years that were simultaneously noted using the 3-year, 5-year and
10-year moving average of all three reconstructed regional snowpack and streamflow stations.

4.5 Conclusions

This study provided the temporal and spatial pattern of the regional snowpack for the last 475 years in the Upper Colorado River basin. This research presents the first large-scale reconstruction of SWE for a river basin. A rotated PCA was utilized to regionalize snow course stations and the result was corroborated with the average cluster analysis method of regionalization. The result identified four significant regions (Eastern, Western, Northern and Southern) in the basin. Out of these four significant regions, the northern region of the UCRB did not correlate well with the available tree-ring chronologies in the basin.

Reconstruction of regional snowpack is necessary for the regional discrimination and representation of hydrologic variability with in the UCRB in a long-term perspective. As the snowpack is a crucial for water supply in the region and various studies recognize the trend over the western U.S. using instrumental data (Cayan, 1996; McCabe and Dettinger, 2002; and Hunter et al., 2006), PLSR reconstruction of three regional standardized April 1 SWE of the UCRB using tree-ring chronologies provided regional drought scenarios in an extended horizon. In addition, reconstruction performance of regional SWE reconstruction was also compared to the performance of point station SWE reconstruction. It was concluded that regional SWE achieved better results in terms of variability and CVSE. Regional SWE reconstruction was preferred and reconstructed in order to obtain the regional drought duration for the last 475 years (1500-1975).
The maximum drought duration ranged from 18-36 years and the average drought duration ranged from 6-13 years. The maximum drought duration and average drought duration in the basin were consistent but were shorter for a smaller moving average and longer for a longer moving averages. It was also notable that the 3-year, 5-year and 10-year moving average of R1, R2, R3 and S1, S2, S3 consistently noted that the end of 15th century drought as one of the longest first three droughts in history of the past 475 years. The majority of them have ranked it as the longest one. This result was also found by Stockton and Jacoby (1976), Hidalgo et al. (2000), and Timilsena et al. (2007). This study also concludes that the end of the 15th century was the most severe not only in terms of magnitude or cumulative deficit but also in terms of duration. The collection, selection and inclusion of tree-ring species are important factor in this type of study to increase reconstruction skill and capture the regional scale climate variability. An additional investigation of moisture sensitive trees neighboring the snowpack region and watershed is suggested.
CHAPTER 5

INDIVIDUAL AND COUPLED IMPACT OF OCEANIC CLIMATE PHENOMENON ON COLORADO RIVER STREAMFLOW AND SNOWPACK USING EXTENDED PERIOD OF RECORD

5.1 Introduction

Natural variability of streamflow with respect to short-term and long-term fluctuations and trends are crucial for assessing the vulnerability of water resources to current or future climate changes (Graumlich et al., 2003). A major source of rivers in the western U.S. is snowmelt and variability of snow accumulation is influenced by large-scale atmospheric circulation patterns (Graumlich et al., 2003).

Ocean-atmospheric variability occurs on interannual, decadal and interdecadal timescales (Tootle et al., 2005). El Niño-Southern Oscillation (ENSO) defines the relationship of the periodic large-scale warming or cooling of the central eastern equatorial Pacific Ocean with the Southern Oscillation, with a periodicity of 2-7 years (Hanley et al., 2003). The Pacific Decadal Oscillation (PDO) is a long-lived ENSO-like pattern of Pacific climate variability (Shen et al., 2006). PDO has significant influence on climate-sensitive natural resources in the Pacific and North America, including the water supplies and snowpack in selected North American regions (Mantua and Hare, 2002) with a periodicity of about 50 years. The Atlantic Multidecadal Oscillation (AMO) is termed as the leading mode of low-frequency, North Atlantic Ocean (0 to 70°) sea surface temperature.
temperature variability with a periodicity of 65-80 years (Kerr, 2000; Gray et al. 2004). Similar to PDO, AMO has two phases: positive and negative.

Piechota and Dracup (1996) found that there were two significant regions (northwest and southwest) with a strong ENSO signal in the United States. A strong relationship was established between ENSO and the southern U.S. in which consistent dry conditions occur during the La Niña years and consistent wet conditions occur during the El Niño years. Cayan et al. (1999) found that high precipitation and high streamflow were more frequent in the southwest U.S. during the El Niño years while there was an opposite pattern during the La Niña years. Hunter et al. (2006) recently studied the interaction of interannual and interdecadal oceanic-atmospheric influences on April 1 Snow Water Equivalent (SWE) in the western United States. Their study showed that La Niña results in increased SWE in the Pacific northwest and decreased SWE in the southwest region. Tootle et al. (2005) used nonparametric testing to evaluate the individual and coupled effect of PDO, AMO, NAO and ENSO using instrumental data of U.S. streamflow and observed data of ocean climate information (PDO, AMO, NAO and ENSO) and found that the warm phase of PDO is associated with increased streamflow in the central and southwest United States, while the warm phase of the AMO is associated with reduced streamflow. Sutton and Hodson (2005) concluded that AMO is responsible for changes in the regional atmospheric circulation and associated for anomalies in precipitation and surface temperature over the United States. Enfield et al. (2001) related the AMO phases with the U.S. rainfall and found that warm phases of AMO results in below normal rainfall in most parts of the United States. Dettinger and Diaz (2000) found that NAO is significantly correlated with streamflow in the eastern United States. During
the winter in the eastern U.S., the negative NAO reduces the moisture transportation and therefore reduces the streamflow in the region.

Earlier studies established that the PDO phases (warm or cold) can enhance or dampen the effect of ENSO teleconnections to climate variables. Gershunov and Barnett (1998) found that the ENSO signal (El Niño/La Niña) was strongest during the different phases of PDO (positive and negative); El Niño patterns (wet southwest) were strongest and consistent during the high phase of PDO. The precipitation patterns during the La Niña winters (dry winters) were consistent during the low PDO phases. The coupling of a cold AMO phase increases the effect of La Niña in the northern and central Rocky Mountains resulting in enhanced SWE.

McCabe et al. (2007) concluded that high flow in Upper Colorado River Basin (UCRB) occurred due to coupled impact of AMO cold and PDO warm. Further, the Tootle et al. (2005) study concluded that the warm phase of the NAO and cold phase of the AMO were associated with increased streamflow in the central United States.

The individual and coupled influences of interannual and decadal phenomenon (ENSO, PDO, AMO, NAO) have been evaluated successfully using instrumental data of length 50-100 years length (e.g. Piechota and Dracup (1996), Cayan et al. (1999), Gershunov and Barnett, 1998; Dettinger and Diaz (2000); Enfield et al. (2001); Sutton and Hudson, 2005; Tootle et al., 2005; Hunter et al., 2006 and McCabe et al. (2007)), but no teleconnection studies have been conducted using reconstructed data of hydrologic variables and oceanic climate phenomenon. The limitation of using only available instrumental data of about 100 years is that the phenomena such as the PDO, AMO and NAO cover only one to two phases during the instrumental period; which may not be
sufficient for thorough interpretation of influences of 20-80 years periodic occurrence of oceanic climate phenomenon on hydrologic variables (streamflow and snowpack).

The focus of the research presented in this chapter is to find the relationship between individual impact of ENSO, PDO, AMO, NAO and its combined effect on streamflow and regional snowpack utilizing an extended period of record for the Colorado River basin. While the instrumental data of about 100 years of streamflow and snowpack could not incorporate more than two phases of PDO, AMO and NAO, this study will utilize extended hydrological and climate records in order to sufficiently interpret the 20-80 years of periodic phenomena with adequate number of phases. This will provide comprehensive and valuable information about the impact of high and low frequency ocean climate phenomena in this region. An additional contribution of this research is to achieve the teleconnection study utilizing the different lag years (0, +1, +2 and +3) of streamflow and regional snowpack. Lastly, this research will determine the change in streamflow volume with respect to the long-term mean volume of the basin due to individual and coupled effect of oceanic climate influences. This is important information for water managers to make decisions on water management and allocation issues.

5.2 Data Sources

5.2.1 Reconstructed Streamflow

It is important to use the reconstructed unimpaired streamflow data in order to study long-term hydroclimate variability. An unimpaired streamflow station is defined as one that is minimally affected by artificial diversions, storage, or other works of man in
the watershed. This provides an account of natural hydrologic response to fluctuations in climate for a watershed (Slack et al., 1993). Unimpaired streamflow stations for the Upper and Lower Colorado River basin were obtained from the United States Geological Survey (USGS), Hydro-Climate Data Network (HCDN) website (http://pubs.usgs.gov/wri/wri934076/1st_page.html ). Based on the HCDN website, there were 45 unimpaired streamflow stations in the Upper Colorado River basin and 18 unimpaired streamflow stations in the Lower Colorado River basin. 17 unimpaired reconstructed water year streamflow stations (Refer: Chapter 3, Appendix-A2) were utilized in order to study the teleconnection pattern. These streamflow stations are presented in Figure 5-1.

Figure 5-1: Unimpaired streamflow stations and snowpack region in the Colorado River basin. Dark dots represent the location of unimpaired streamflow stations.
5.2.2 Reconstructed Regional Snow Water Equivalent (SWE)

The April 1 SWE is a good indicator of the water content in the maximum seasonal snowpack and for the water supply for coming months (Woodhouse, 2003). In the study presented here, SWE is regarded as the potential climate variable (indicator). The snow course data were collected from the United States Department of Agriculture (USDA), Natural Resources Conservation Service (NRCS) website (http://www.wcc.nrcs.usda.gov/snowcourse/). Three reconstructed regional composite April 1 SWE (Refer: Chapter 4, Section 4.4.2) obtained from 39 snow course stations were utilized in this research. The regions (R1, R2, and R3) identified in Chapter 4, section 4.4.1 are presented by dashed lines in Figure 5-1.

5.2.3 Reconstructed Interdecadal and Decadal Oceanic Data

5.2.3.1 El Niño-Southern Oscillation (ENSO)

ENSO is a natural coupled cycle in the ocean-atmospheric system over the tropical Pacific that operates on a timescale of 2-7 year (Hanley et al., 2003). It involves large-scale fluctuations in a number of oceanic and atmospheric variables such as sea surface temperature and sea level pressures. The warm phase of ENSO is referred to as the El Niño and the cool phase is defined as the La Niña. The phase and strength of ENSO events are defined by indices; however, there is no single index that best describes the ENSO years and their corresponding strength, timing and duration of events (Hanley et al., 2003). ENSO is represented by the Southern Oscillation Index (SOI), Multivariate
ENSO index (MEI), and Sea Surface temperature data sets (Wright SST, Niño 1 and 2, Niño 3, Niño 3.4 and Niño 4). In this research, the tree-ring reconstructed average SOI for the month of December/January/February by Stahle et al. (1998) was retrieved and utilized from the National Oceanic and Atmospheric Administration (NOAA) Paleoclimatology Program, World Data Center (WDC) for Paleoclimatology website (http://www.ncdc.noaa.gov/paleo/recons.html ). El Niño (warm phase) years were defined as a year which has a value of SOI less than -1.0 and La Niña (cold phase) years were defined as a year which had a value of SOI greater than 1.0 (Pacific ENSO update, 2006). Stahle et al. (1998) utilized 14 tree-ring chronologies from northern Mexico (Durango and Chihuahua), the southwestern U.S.A. (Arizona, Utah and New Mexico) and Java, Indonesia plus the first two factor scores from a network of 9 chronologies in OK and TX as predictors. All chronologies were compiled and detrended to remove biological growth factors.

5.2.3.2 Pacific Decadal Oscillation (PDO)

The Pacific Decadal Oscillation (PDO) is a long-lived ENSO-like pattern of Pacific climate variability (Shen et al., 2006). Similar to ENSO, PDO has two phases: warm and cold. Each phase persists for about 25 years. Mantua et al. (1997) defined the PDO as the leading principal component of North Pacific monthly sea surface temperature variability (poleward of 20°N for the 1900-1993 periods). The tree-ring data reconstructed PDO by Biondi et al. (2001) were retrieved and utilized from the NOAA Paleoclimatology Program and WDC for Paleoclimatology website. Biondi et al. (2001) utilized tree-ring chronologies from the transverse Mountains of Southern California to Sierra San Pedro Martir in northern Baja California situated in a direction roughly
parallel to the coastline. This region was selected after identifying that tree-ring records from this area were better correlated with PDO than with ENSO (Biondi et al. 2001). The correlation value between the tree-ring chronologies utilized by Biondi et al. 2001 and tree-ring chronologies used for the unimpaired streamflow and regional snowpack reconstruction of Colorado River basin in this study were -0.06 to 0.47.

5.2.3.3 Atlantic Multidecadal Oscillation (AMO)

The leading mode of low-frequency, North Atlantic Ocean (0 - 70 °) SST variability with a periodicity of 65-80 years is termed the Atlantic Multidecadal Oscillation (Kerr, 2000; Gray et al, 2004). Similar to ENSO and PDO, it also has two phases: warm and cold, that persists for about 30-40 years. The tree-ring reconstructed time series of AMO by Gray et al. (2004) was utilized in this study. This data is available in the NOAA Paleoclimatology Program and WDC for Paleoclimatology website (http://www.ncdc.noaa.gov/paleo/recons.html). Gray et al. (2004) utilized 12 tree-ring records from eastern North America, Western Europe, Scandinavia, and the Middle East in order to reconstruct AMO from 1567 to 1990 A.D.

5.2.3.4 North Atlantic Oscillation (NAO)

The NAO is a large-scale fluctuation in atmospheric pressure between the subtropical high pressure system located near the Azores in the Atlantic Ocean and subpolar low pressure system near Iceland and is quantified as the NAO index. It is a dominant mode of atmospheric and climate variability in the North Atlantic region (Hurrel, 1995; Hurrel and Van Loon, 1997). Similar to PDO and AMO, NAO has both positive and negative phases. The tree-ring reconstructed winter NAO time series by Cook et al. (1998) was retrieved from the NOAA Paleoclimatology Program and WDC
for Paleoclimatology website (http://www.ncdc.noaa.gov/paleo/recons.html), and was utilized in this research. Cook et al. (1998) utilized ten tree-ring records; six from eastern North America and four from northwestern Europe, in order to develop the reconstruction of the winter North Atlantic Oscillation (NAO).

5.3 Methodology

5.3.1 Teleconnection Study of Individual Impact of ENSO, PDO, AMO and NAO Using Reconstructed Longer Period of Record

This step evaluated the individual impacts of ENSO, PDO, AMO and NAO on individual water year streamflow and regional composite April 1 SWE of the Colorado River basin. The teleconnection study was based on the extended record of ocean climate phenomenon (ENSO, PDO, AMO and NAO) and reconstructed climate variables (unimpaired water year streamflow and regional composite annual April 1 SWE) of the region. The reconstructed data of ocean climate phenomenon and reconstructed climate variables from 1706 to 1970 were utilized in the study because of the availability of ocean climate data and streamflow data. In order to evaluate the lagged response, the impacts of ocean climate phenomenon were evaluated utilizing 0, +1, +2 and +3 lag year time series of streamflow and snowpack. For instance, lag 0 effect was defined as the relation between oceanic climate phenomenon and streamflow/snowpack in the same year. Lag +1 impact is the relation between the current year of oceanic climate phenomenon to the next year data of streamflow and snowpack. Similarly, lag +2 and lag 3 impact studies were carried out using the relation between the current year of oceanic climate phenomenon to the next +2 and +3 year data of streamflow and snowpack. The
study of phases (positive or negative) was also performed for the ENSO, PDO, AMO and NAO for significant (95%) differences in streamflow and SWE medians. Tootle et al. (2005) performed a similar study using 639 streamflow stations in the United States based on the available instrumental data of streamflow and Pacific climate phenomenon from 1951-2002, which covers only one or two cycles of multidecadal oceanic phenomenon. The longer period of record was particularly important to draw a better conclusion on the teleconnection due to multi-decadal oceanic phenomenon.

The Non-parametric rank sum test was used to evaluate the individual impacts of ocean-atmospheric phenomenon on climate variables (unimpaired water year streamflow and regional composite annual April 1 SWE). The rank sum test assumes that the two data sets are identically distributed. There is no assumption of normality (Maidment, 1993). The two data sets should not be paired and can vary in size. This approach does not assume any form of linear relationship as is inherent in correlation analysis. This method compares the two independent data sets and determines if one data set has significantly larger values than the other data set. Applying the rank sum test, it was possible to determine the significant median difference in water year streamflow or snowpack comparing cold and warm phases of ENSO, PDO, AMO and NAO.

Each year of warm and cold phases of ENSO, PDO, AMO and NAO were identified. As mentioned earlier, El Niño (warm phase) years were defined as a year which has a value of SOI less than -1.0 and La Niña (cold phase) years were defined as a year which had a value of SOI greater than 1.0 (Pacific ENSO Update, 2006). In the case of PDO, AMO and NAO, the positive index values was considered as the warm phase where as the negative value was considered as the cold phase. Once the year of warm and
cold phases of each ocean climate phenomenon was determined, the respective streamflow volume time series for each station was obtained. Streamflow for each phase was evaluated considering the significant (greater than 95%) differences in streamflow medians using the rank sum test. As mentioned earlier, the lagged effect of ocean climate phenomenon to the climate variables was also studied in this research utilizing the lag of 0, +1, +2 and +3 years of streamflow and regional snowpack. The streamflow stations, which had the significant difference in medians (cold vs. warm or warm vs. cold), were considered as the impacted streamflow stations or regional snowpack with the respective phase of ocean climate phenomenon.

5.3.2 Teleconnection Study of Coupled Impact of ENSO with PDO, AMO and NAO Using Reconstructed Longer Period of Record

This step included the study of the coupled impact of ENSO with PDO, AMO and NAO on individual unimpaired water year streamflow and regional composite April 1 SWE of the region. The teleconnection study was based on the extended record of ocean climate phenomenon (ENSO, PDO, AMO and NAO) and reconstructed climate variables (unimpaired water year streamflow and regional composite April 1 SWE) of the region. The lagged effect of ocean climate phenomenon to the climate variables was also studied in this research utilizing the 0, +1, +2 and +3 years of water year streamflow and regional snowpack. ENSO was coupled with the PDO, AMO and NAO in order to find whether the interannual high frequency effect is enhanced or dampened due to multidecadal oceanic climate phenomenon. Similar to the previous step, the phases (positive and negative) of ocean-atmospheric phenomenon was utilized for the significant test (95%) in medians.
The initial step of the procedure was the identification of the phases (positive or negative) and corresponding years of ENSO, PDO, AMO and NAO. Once the year of warm and cold phases of each ocean climate phenomenon and corresponding streamflow volume for each station was obtained, it was possible to evaluate the coupling effect of ENSO with PDO, AMO and NAO respectively. As an illustration of this, the coupling effects of ENSO and PDO was determined using streamflow volume or regional composite April 1 SWE that coincides with the year that has PDO cold/El Niño, PDO warm/El Niño, PDO cold/La Niña and PDO warm/La Niña. After that, the coupled climate phenomenon was evaluated considering the significant (greater than 95%) differences in streamflow/snowpack medians of PDO cold/El Niño - PDO warm/El Niño and PDO cold/La Niña - PDO warm/La Niña respectively using the rank sum test. The streamflow station/regionalized April 1 SWE with the significant (greater than 95%) difference in medians using the coupled effect was considered as the impacted stations with the respective coupled phenomenon. This same approach was also utilized for the coupled effect of ENSO with AMO and NAO respectively.

5.3.3 Teleconnection Study of Coupled Impact of PDO, AMO and NAO Using Reconstructed Longer Period of Record

This step included the study of coupled impact of PDO, AMO and NAO on individual unimpaired water year streamflow and regional composite April 1 SWE of the region. The teleconnection study was based on the extended record of ocean climate phenomenon (PDO, AMO and NAO) and reconstructed climate variables (unimpaired water year streamflow and regional composite April 1 SWE) of the region. Coupling of PDO, AMO and NAO was done in order to find the coupling effect of multidecadal
oceanic climate phenomenon. The lagged effect of ocean climate phenomenon to the climate variables was also studied in this research utilizing the lag of 0, +1, +2 and +3 years of water year streamflow and regional snowpack. Similar to the previous step, the phases (positive and negative) of ocean climate phenomenon was utilized for the significant test (95%) in medians. The rank sum test was performed for the analysis.

5.3.4 Mean Difference of Streamflow during the Individual and Coupled Phase of ENSO, PDO, AMO and NAO Using Reconstructed Period of Record

In addition to the rank sum test, the mean differences (in terms of percentage) of water year streamflow of each station for the periods of specific individual/coupled ocean climate phenomenon with respect to the long-term mean were obtained. The mean of the water year volume during specific individual/coupled phenomenon was obtained by taking the average of water year volume during the period of specific individual/coupled ocean climate phenomenon (e.g. PDO cold/AMO warm, PDO cold/AMO cold, PDO warm/AMO cold, PDO warm/AMO cold). Similarly, the long-term mean was obtained by taking an average of water year volume throughout the period (1706-1970). This mean difference study was determined for the lag zero scenarios. The average increase or decrease in percentage flow for all station was obtained in order to obtain the overall basin response for the specific ocean climate phenomenon scenario.
5.4 Results

5.4.1 Individual Impact of ENSO, PDO, AMO and NAO Using Reconstructed Water Year Streamflow and April 1 SWE

Table 5-1, 5-2 and Figure 5-2 summarize the results of the Non-parametric testing of cold and warm phases of ENSO (La Niña vs El Niño), PDO (cold vs warm), AMO (cold vs warm) and NAO (cold vs warm) to water year streamflow and regional snowpack (regional standardized April 1 SWE) using the lag response of 0, +1, +2 and +3 years. For example in Table 5-1 (row 1), there is a significant (95%) negative difference in medians (for 16 out of 17 streamflow stations) for lag 0 year between La Niña and El Niño. This means that the La Niña is associated with decreased streamflow/snowpack in the region and El Niño is associated with the increased streamflow/snowpack. In the Figure 5-2, a black (gray) circle represents a statistically significant (greater than 95%) positive (negative) difference in medians. A spatially distributed black small “dot” represents a station that is not statistically significant.
Table 5-1: Individual impact of ocean climate phenomenon on streamflow stations. The values indicate the number of significant streamflow stations (out of 17 stations) based on the rank sum test under the different scenario of ocean climate phenomenon. P and N indicate the positive and negative significant streamflow stations respectively.

<table>
<thead>
<tr>
<th>Individual Phenomenon</th>
<th>Streamflow Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lag (0)</td>
</tr>
<tr>
<td></td>
<td>Nos</td>
</tr>
<tr>
<td>1 La Niña - El Niño</td>
<td>16</td>
</tr>
<tr>
<td>2 PDO cold-PDO warm</td>
<td>17</td>
</tr>
<tr>
<td>3 AMO cold-AMO warm</td>
<td>4</td>
</tr>
<tr>
<td>4 NAO cold-NAO warm</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 5-2: Individual impact of ocean climate phenomenon on regional snowpack. The values indicate the number of significant snowpack regions (out of 3 regions) based on rank sum test under the different scenario of ocean climate phenomenon. P and N indicate the positive and negative significant regions respectively. R1, R2 and R3 indicate snowpack Region 1, Region 2 and Region 3 respectively.

<table>
<thead>
<tr>
<th>Individual Phenomenon</th>
<th>Snow pack Regions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lag (0)</td>
</tr>
<tr>
<td></td>
<td>Nos</td>
</tr>
<tr>
<td>La Niña – El Niño</td>
<td>2</td>
</tr>
<tr>
<td>PDO cold-PDO warm</td>
<td>2</td>
</tr>
<tr>
<td>AMO cold-AMO warm</td>
<td>0</td>
</tr>
<tr>
<td>NAO cold-NAO warm</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 5-1 and Figure 5-2(a) indicate that 16 out of 17 streamflow stations in the Colorado River basin were associated with the increased streamflow during ENSO warm (El Niño) considering lag 0 years. Further, Table 5-2 indicates that increased April 1 SWE of Region 2 (R2) and Region 3(R3) was associated with the ENSO warm (El Niño) phase. The impact of ENSO to water year streamflow and regional snowpack (regional standardized April 1 SWE) in the same year was significantly noted in the region. The ENSO was active during the lag zero years but did not have any effect during the lag +1, +2 and +3 years. This result of increased streamflow/snowpack due to El Niño and decreased streamflow/snowpack due to La Niña in the southwest is consistent with
previous studies using only historical records by Redmond and Koch (1991), Piechota and Dracup (1996), and Cayan et al. (1999).

(a) ENSO cold - ENSO warm  (b) PDO cold - PDO warm

(c) AMO cold - AMO warm

Figure 5-2: Significant (95%) difference in streamflow medians for (a) ENSO (cold) - ENSO (warm) (b) PDO (cold) - PDO (warm) (c) AMO (cold) –AMO (warm). This result is based on the lag 0 scenario. A larger dark (Gray) dot represents positive
(negative) significant stations. Smaller dark dots indicate the streamflow stations in the basin that are not significant.

The PDO signal was also significantly found in the streamflow and regional snowpack of the Colorado River basin (Figure 5-2 (b), Table 5-1 and Table 5-2). Table 5-1 and Figure 5-2 (b) indicate that the PDO warm/cold phase was associated with the increased/decreased streamflow in all 17 streamflow stations. Similarly to the streamflow, Table 5-2 indicates that the increased/decreased snowpack of Region 2 and 3 were associated with the warm/cold phase of PDO. This result was consistent with the Tootle et al. (2005) study that also concluded the association of increased streamflow in southwest due to the warm phase of PDO. A strong effect of PDO was observed in terms of the number of stations of the basin in this research. Further, Hunter et al. (2006) also concluded that the decreased SWE in Utah and Colorado was associated with the cold phase of PDO. Since, the PDO is a decadal phenomenon, the lag 0 and +1 have similar results, but a decrease in significant stations was observed during the +2 and +3 lag years.

The AMO signal was not so distinct as compared to PDO and ENSO in the Colorado River basin (Figure 5-2(c), Table 5-1, and Table 5-2). There were only 4 (out of 17) streamflow stations that were associated with the increased/decreased streamflow due to the cold/warm phase of AMO. Tootle et al. (2005) and McCabe et al. (2007) also concluded that the low flow generally occurred in the Colorado River basin when the AMO was positive. Even though about 20% of the streamflow stations were only affected, the possibility of an association of a AMO negative causing increased streamflow in the basin can not be neglected at this stage. Further, it was also noted that
the impacted number of stations were slightly (4 to 7 out of 17) increased when the lag
was increased from 0 to +1 years but decreased after the lag of +1 year.

From Table 5-1 and Table 5-2, the NAO signal was not significantly associated
with the streamflow/snowpack in the Colorado River basin in the same year (lag 0 year).
There were no streamflow stations and regional snowpack that were associated with the
increased streamflow/snowpack due to the cold/warm phase of NAO considering the lag
0 effect. This result was consistent with the Visveck et al. (2001), Tootle et al. (2005) and
Hunter et al. (2006) that concluded that there is no association of NAO signal in the
western precipitation, U.S. streamflow and snowpack. Interestingly, Table 5-1 and Table
5-2 clearly indicated the lag +1 effect of NAO warm was noticeably associated with the
increased streamflow/snowpack affecting most of the streamflow stations (15 out of 17)
and all regional snowpack regions (R1, R2 and R3) in the basin. This may be useful for
predicting streamflow.

5.4.2 Coupling of Interdecadal (PDO, AMO and NAO) and ENSO Using
Reconstructed Longer Period of Record

Tables 5-3 and Table 5-4 represent the 6 different combinations of coupling
effects of ENSO with the PDO, AMO and NAO. The coupling effects of ENSO with
PDO were evaluated testing the relationship for PDO cold/ El Niño – PDO warm/ El
Niño and PDO cold/ La Niña – PDO warm/ La Niña (Figure 5-3). Similarly, the coupling
effect of ENSO with AMO was evaluated as AMO cold/ El Niño – AMO warm/ El Niño
and AMO cold/ La Niña – AMO warm/ La Niña. The coupling effect of ENSO with NAO
was evaluated as NAO cold/ El Niño – NAO warm/ El Niño and NAO cold/ La Niña –
NAO warm/ La Niña. Based on the rank sum test, it was concluded that the coupling of
El Niño with the PDO (warm/cold) was associated with a decreased number of impacted streamflow in the basin for lag 0, +1, +2 and +3 years. This result of a decreased number of impacted streamflow stations based on the rank sum was inconsistent with Gershunov and Barnett (1998) who found that the ENSO signal (El Niño/La Niña) was strongest during the different phases of PDO (positive/negative) respectively. Contrary to Tootle et al., (2005) and Hunter et al., (2006), this study identified the coupled effect of PDO and ENSO on UCRB streamflow at the 95% significance level.

There were few stations that were noted in the coupling of AMO and ENSO. It was noteworthy that NAO warm/cold was associated with the increased/decreased streamflow/snowpack when coupled with the El Niño in most of the streamflow stations (16 out of 17) and snowpack regions (2 out of 3) considering the lag +1 year effect in the basin.
Table 5-3: Coupling effect of ENSO with PDO, AMO and NAO on streamflow.

The value indicates the number of significant streamflow stations (out of 17 stations) based on Non-parametric rank sum test under the different scenario of ocean climate phenomenon. P and N indicate the positive and negative significant stations respectively.

<table>
<thead>
<tr>
<th>Coupled Phenomenon</th>
<th>Streamflow Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lag (0)</td>
</tr>
<tr>
<td>El Niño/PDO cold-</td>
<td>14</td>
</tr>
<tr>
<td>El Niño/PDO warm</td>
<td>11</td>
</tr>
<tr>
<td>La Niña/PDO cold-</td>
<td>1</td>
</tr>
<tr>
<td>La Niña/PDO warm</td>
<td>0</td>
</tr>
<tr>
<td>El Niño/AMO cold-</td>
<td>0</td>
</tr>
<tr>
<td>El Niño/AMO warm</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 5-4: Coupling effect of ENSO with PDO, AMO and NAO on regional snowpack. The value indicates the number of significant snowpack regions (out of 3 regions) based on rank sum test under the different scenario of ocean climate phenomenon. P and N indicate the positive and negative significant regions respectively. R1, R2 and R3 indicate snowpack Region 1, Region 2 and Region 3 respectively.

<table>
<thead>
<tr>
<th>Coupled Phenomenon</th>
<th>Snow pack Regions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lag (0)</td>
</tr>
<tr>
<td>1 El Niño/ PDO cold-El Niño/ PDO warm</td>
<td>1(R3)</td>
</tr>
<tr>
<td>2 La Niña/PDO cold-La Niña/PDO warm</td>
<td>1(R3)</td>
</tr>
<tr>
<td>3 El Niño/ AMO cold-El Niño/ AMO warm</td>
<td>0</td>
</tr>
<tr>
<td>4 La Niña/AMO cold-La Niña/AMO warm</td>
<td>0</td>
</tr>
<tr>
<td>5 El Niño/ NAO cold-El Niño/ NAO warm</td>
<td>0</td>
</tr>
<tr>
<td>6 La Niña/NAO cold-La Niña/NAO warm</td>
<td>0</td>
</tr>
</tbody>
</table>
Figure 5-3: Significant (95%) difference in streamflow medians for (a) El Niño/PDO cold - El Niño/PDO warm (b) La Niña/PDO cold - La Niña/PDO warm. The result is based on lag 0 scenarios. Larger Gray dot represents negative significant stations. Smaller dark dots indicate the streamflow stations in the basin that are not significant.
5.4.3 Coupling of Interdecadal (PDO, AMO and NAO) Testing Using Reconstructed Period of Record

The couplings of the interdecadal oscillations (PDO, AMO and NAO) were assessed for 6 different combinations. The 6 different combinations were PDO cold/AMO cold – PDO cold/AMO warm, PDO warm/AMO cold – PDO warm/AMO warm, PDO cold/NAO cold – PDO cold/NAO warm, PDO warm/NAO cold – PDO warm/NAO warm, AMO cold/NAO cold – AMO cold/NAO warm, and AMO warm/NAO cold – AMO warm/NAO warm. Table 5-5 and Table 5-6 summarized the effect based on the different combinations of PDO, AMO and NAO considering the lag 0, +1, +2 and +3 years.
Table 5-5: Coupling effect of PDO, AMO and NAO on streamflow. The value indicates the number of significant streamflow stations (out of 17 stations) based on Non-parametric rank sum test under the different scenario of ocean climate phenomenon. P and N indicate the positive and negative significant stations respectively.

<table>
<thead>
<tr>
<th>Coupled Phenomenon</th>
<th>Streamflow Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lag (0)</td>
</tr>
<tr>
<td></td>
<td>Nos</td>
</tr>
<tr>
<td>1 PDO cold/AMO cold – PDO cold/AMO warm</td>
<td>0</td>
</tr>
<tr>
<td>2 PDO warm/AMO cold – PDO warm/AMO warm</td>
<td>0</td>
</tr>
<tr>
<td>3 PDO cold/NAO cold – PDO cold/NAO warm</td>
<td>0</td>
</tr>
<tr>
<td>4 PDO warm/NAO cold – PDO warm/NAO warm</td>
<td>0</td>
</tr>
<tr>
<td>5 AMO cold/NAO cold, – AMO cold/NAO warm</td>
<td>0</td>
</tr>
<tr>
<td>6 AMO warm/NAO cold – AMO warm/NAO warm</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 5-6: Coupling effect of PDO, AMO and NAO on regional snowpack. The value indicates the number of significant snowpack regions (out of 3 regions) based on Non-parametric rank sum tests under the different scenario of ocean climate phenomenon. P and N indicate the positive and negative significant stations (p<0.5). R1, R2 and R3 indicate snowpack Region 1, Region 2 and Region 3 respectively.

<table>
<thead>
<tr>
<th>Coupled Phenomenon</th>
<th>Snow pack Regions</th>
<th>Lag (0)</th>
<th>Lag (1)</th>
<th>Lag (2)</th>
<th>Lag (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Nos</td>
<td>Sign</td>
<td>Nos</td>
<td>Sign</td>
</tr>
<tr>
<td>1</td>
<td>PDO cold/AMO cold – PDO cold/AMO warm</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>PDO warm/AMO cold – PDO warm/AMO warm</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>PDO cold/NAO cold – PDO cold/NAO warm</td>
<td>0</td>
<td>-</td>
<td>2 (R2,R3)</td>
<td>N</td>
</tr>
<tr>
<td>4</td>
<td>PDO warm/NAO cold – PDO warm/NAO warm</td>
<td>0</td>
<td>-</td>
<td>1(R3)</td>
<td>N</td>
</tr>
<tr>
<td>5</td>
<td>AMO cold/NAO cold, – AMO cold/NAO warm</td>
<td>0</td>
<td>-</td>
<td>1(R3)</td>
<td>N</td>
</tr>
<tr>
<td>6</td>
<td>AMO warm/NAO cold – AMO warm/NAO warm</td>
<td>0</td>
<td>-</td>
<td>2 (R2,R3)</td>
<td>N</td>
</tr>
</tbody>
</table>

From Table 5-5 and Table 5-6, all the combinations of decadal oscillations (PDO, AMO and NAO) were not clearly associated with the significant increased or decreased streamflow and snowpack in the basin for lag 0 year. The effect could be observed if the significance level of rank sum test is further reduced to 90%.

From Table 5-5 and Table 5-6, it was observed that PDO cold/NAO warm, PDO warm/NAO warm, AMO cold/NAO warm and AMO warm/NAO warm were associated with the increased streamflow and snowpack in the region for lag +1 year scenario. Even though these combinations were associated with the increased streamflow/snowpack for about 20-50% of the stations or regions, the effect of these coupled effects in the region...
should not be neglected. It was also concluded that NAO warm is associated with the increased streamflow and snowpack in the region for lag + 1 year. Contrary to the McCabe et al. (2007) study, this rank sum test did not identify any significant stations that cause the significantly high flow in UCRB due to the coupled impact of AMO cold and PDO warm. This effect could be observed if the significant level of rank sum test is further reduced.

5.4.4 Individual and Coupled Impact of ENSO, PDO, AMO and NAO Using the Mean of Reconstructed Period of Record

The mean differences (in terms of percentage) of water year streamflow of each station for the periods of specific individual/coupled ocean climate phenomenon with respect to the long-term mean were obtained. This was done for 17 streamflow stations under the 32 different scenarios of individual and coupled phenomenon. The result is summarized in Figure 5-4.

In Figure 5-4 (a) and (b), the El Niño, PDO warm, AMO cold, El Niño/PDO warm, El Niño/AMO cold, El Niño/AMO warm, El Niño/NAO cold, El Niño/NAO warm, PDO warm/AMO warm, PDO warm/AMO cold, PDO warm/NAO warm, PDO warm/NAO cold and AMO cold/NAO warm had noticeably increases the streamflow. This result was based on result of the mean, median, 25th, 75th and 5th/95th percentile of the stations that indicated positive percentage of increased mean in all the streamflow stations in the basin. Similarly, La Niña, PDO cold, La Niña/PDO cold, La Niña/AMO cold, La Niña/AMO warm, La Niña/NAO cold, La Niña/NAO warm, PDO cold/ AMO warm, PDO cold/ AMO cold, PDO cold/NAO warm, PDO cold/NAO cold, and AMO warm/NAO warm had decrease in streamflow in the Colorado River basin.
Further, El Niño/ La Niña patterns during the positive/negative phase of PDO had the strongest influence on streamflow and had maximum increased/decreased streamflow of about 22-30%. This result was consistent with the study by Gershunov and Barnett (1998) that found that El Niño was the strongest during the positive phase of PDO, having a wet southwest. This influence was not observed utilizing the rank sum test in terms of number of affected stations, but the mean difference in percentage showed significant increased/decreased in streamflow volume.

From Figure 5-4 (b), it was concluded that the AMO warm was associated with the decreased streamflow in the region as compared to AMO cold when coupled with PDO. AMO warm was dampening the streamflow when coupled with the El Niño, La Niña and PDO warm respectively. The result was similar to the result by Enfield et al. (2001) that concluded warm phases of AMO resulted in below normal rainfall in most parts of the United States. Further, PDO warm coupled with AMO cold enhanced the streamflow volume in the basin and this relation was consistent with McCabe et al. (2007). These results were not observed on the rank sum test of medians considering the significance level of 95% but may be observed if the significant level is further reduced.

In addition, the result of percentage increase or decrease due to individual influence of ENSO, PDO and AMO was compared between the instrumental record (1940-1970) and the reconstructed record (1706-1770) in Figure 5-5. The plot showed that the PDO cold and warm showed higher effect in reconstructed data of the basin as compared to the instrumental shorter period of data. Further, the variability was reduced in reconstructed data of PDO warm and AMO cold, which is rational because a larger number of events were sampled using reconstructed data.
Figure 5-4: Box and Whisker plots that shows mean, median, and percentile (5\textsuperscript{th}, 25\textsuperscript{th}, 75\textsuperscript{th} and 95\textsuperscript{th}) increase/decrease (in terms of percentage) in streamflow volume (considering all 17 stations) due to individual and coupling effect of oceanic climate phenomenon. The box delineates the 25\textsuperscript{th} and 75\textsuperscript{th} percentile. Whiskers are 5\textsuperscript{th} and 95\textsuperscript{th} percentile. The horizontal dash line inside the box indicates mean. The horizontal bold
straight line inside the box indicates the median. The long horizontal dot line indicates the line of zero percentage of increase or decrease.

Figure 5-5: Box and Whisker plots that shows mean, median, and percentile (5th, 25th, 75th and 95th) increase/decrease (in terms of percentage) in streamflow volume (considering all 17 stations) due to individual effect of ENSO, PDO and AMO using historical and reconstructed data respectively. The box delineates the 25th and 75th percentile. Whiskers are 5th and 95th percentile. The horizontal dash line inside the box indicates mean. The horizontal bold straight line inside the box indicates the median. The long horizontal dot line indicates the line of zero percentage of increase or decrease.
Further, the mean difference in percentage was separately calculated and plotted for the streamflow stations located for Upper Colorado River Basin (Figure 5-6) and Lower Colorado River Basin (Figure 5-7) respectively.

In Upper Colorado River basin, there were 12 streamflow stations. The impact due to 32 different scenarios of individual and coupled phenomenon was evaluated utilizing the mean difference in percentage of streamflow stations for all 12 streamflow stations located in upper basin. The result is summarized in Figure 5-6. In the upper basin, the El Niño, PDO warm, AMO cold, El Niño/PDO warm, El Niño/AMO cold, El Niño/AMO warm, El Niño/NAO cold, PDO warm/AMO cold, and PDO warm/NAO cold were associated with increased streamflow. This was based on the result of the mean, median, 25th, 75th and 5th/95th percentile of the stations that indicated a positive percentage of increased mean in all the streamflow stations in the basin. Similarly, La Niña, PDO cold, AMO warm, La Niña/PDO cold, La Niña/AMO cold, La Niña/AMO warm, La Niña/NAO cold, La Niña/NAO warm, PDO cold/AMO warm, PDO cold/AMO cold, PDO cold/NAO warm, PDO cold/NAO cold, and AMO warm/NAO warm were associated with decreased streamflow in the Colorado River basin. Similar to the result of whole basin (upper and lower), El Niño/ La Niña patterns during the positive/negative phase of PDO had the strongest influence on the upper basin streamflow and the maximum increased/decreased streamflow of about 15-25%.
Figure 5-6: Box and Whisker plots that shows mean, median, and percentile (5th, 25th, 75th and 95th) increase/decrease (in terms of percentage) in streamflow volume (considering all 12 stations of upper basin only) due to individual and coupling effect of oceanic climate phenomenon. The box delineates the 25th and 75th percentile. Whiskers are 5th and 95th percentile. The horizontal dash line inside the box indicates
mean. The horizontal bold straight line inside the box indicates the median. The long horizontal dot line indicates the line of zero percentage of increase or decrease.

**Lower Colorado River Basin**

![Box plot showing mean, median, and percentile increase/decrease in streamflow volume](image)

**Figure 5-7:** Box shows mean, median, and percentile (lowest and highest) increase/decrease (in terms of percentage) in streamflow volume (considering all 5 stations of lower basin only) due to individual and coupling effect of oceanic climate

131

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
phenomenon. The horizontal dash line inside the box indicates mean. The horizontal bold straight line inside the box indicates the median. The long horizontal dot line indicates the line of zero percentage of increase or decrease.

In the Lower Colorado River basin, there were only 5 unimpaired streamflow stations that have been utilized in this study. Similar to the upper basin, the impact due to 32 different scenarios of individual and coupled phenomenon was evaluated utilizing the mean difference in percentage of streamflow volume.

The result is summarized in Figure 5-7. The El Niño, PDO warm, AMO cold, El Niño/PDO warm, El Niño/AMO cold, El Niño/AMO warm, El Niño/NAO cold, El Niño/NAO warm, PDO warm/AMO warm, PDO warm/AMO cold, PDO warm/NAO warm, PDO warm/NAO cold and AMO cold/NAO warm had increased streamflow volume on lower basin. This was based on result of the mean, median and percentile (highest and lowest) of the stations that indicated positive percentage of increased mean in all the streamflow stations in the basin. Similarly, La Niña, PDO cold, AMO warm, La Niña/PDO cold, La Niña/PDO warm, La Niña/AMO cold, La Niña/AMO warm, La Niña/NAO cold, La Niña/NAO warm, PDO cold/ AMO warm, PDO cold/ AMO cold, PDO cold/NAO warm, PDO cold/NAO cold, and AMO warm/NAO warm had decreased streamflow. In most of the cases, the individual and coupled impact of ocean climate phenomenon had similar effect on the upper and lower basin but the magnitude of percentage increase or decrease was higher in the lower basin. The El Niño/ La Niña patterns during the positive/negative phase of PDO had the strongest influence on
streamflow and had the maximum increased/decreased streamflow of about 40-50%, which was about the double the magnitude than that of the upper basin.

5.5 Conclusions

The important aspect of this study was to evaluate the hydrologic variability of the Colorado River basin utilizing the reconstructed longer period of record. The previous research (e.g.; Tootle et al., 2005, Hunter et al. 2006, McCabe et al., 2007, and Gershunov and Barnett, 1998) evaluated the individual and coupling effect of ocean climate phenomenon (ENSO, PDO, AMO, and NAO) to the U.S. streamflow and western snowpack in the region using the observed period of about last 50 years, this analysis using the median and mean based on extended period record (1706-1970) strengthened the conclusion.

Overall, it was found that there is an increase in streamflow/snowpack during El Niño and decreased streamflow/snowpack during La Niña. This is consistent with the previous studies using instrumental data (e.g.; Redmond and Koch, 1991, Piechota and Dracup, 1996, and Cayan et al. 1999). Similarly, the result of association of increased streamflow/snowpack in the basin due to the warm phase of PDO was consistent with the result of Tootle et al. (2005) and Hunter et al. (2006) for the southwest. However, this study found notably higher strength of associations in terms of number of significant stations (17 out of 17) as compared to Tootle et al. (2005) and Hunter et al. (2006). Further, low flows were observed in the Colorado River basin when the AMO was positive, and this conclusion was also reached by Tootle et al. (2005) and McCabe et al. (2007).
Based on the rank sum test, the coupling impact of El Niño with the PDO (warm/cold) was associated with the decreased number of impacted streamflow stations in the basin. However, the result based on mean (percentage increase or decrease on mean) streamflow was consistent with Gershunov and Barnett (1998) that concluded the coupling effect of El Niño/ La Niña with the PDO (warm/cold) was associated with increased/decreased streamflow volume and snowpack in the basin.

It was found that combinations of decadal oscillations (PDO, AMO and NAO) were not associated with increased or decreased in streamflow volume and snowpack in the region. It was also concluded that the NAO warm was associated with increased streamflow and snowpack in the region for lag + 1 year. Contrary to the McCabe et al. (2007) study, this study did not identify any significant stations based on rank sum test that cause the significantly high flow in Colorado River basin due to the coupled impact of AMO cold and PDO warm. However, as obtained by McCabe et al. (2007), the result based on mean (percentage increase or decrease) concluded that the coupled effect of AMO cold and PDO warm was associated with high flow in the basin.

In summary, the result based on streamflow and snowpack region are showing similar patterns of impact in the most of the scenarios. For instance, if there were large number of stations that were affected by the specific oceanic climate phenomenon, there were also a large number of snowpack regions affected by the same oceanic climate phenomenon. While comparing the result based on mean (percentage increase or decrease) of streamflow stations located in two basins; upper and lower, it was found that there was not a big difference in results of individual and coupled impact of oceanic climate phenomenon between these two basins. But, the magnitude of impact (in terms of
percentage) was noted significantly higher in the lower basin as compared to the upper basin. Even though, there is uncertainty associated with reconstructed variables due to possibility of losing important climatological information during the tree-ring reconstruction, this research has shown that tree ring analysis using long records can capture multiple phases of ENSO, PDO, AMO and NAO.
CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

This research made several contributions to the field of water resources, more specifically, the use of an improved reconstruction method of hydrologic variables and the study of the impact of ocean climate phenomenon using the longer period of record. The first contribution of this research was to introduce and demonstrate PLSR for improved reconstruction of streamflow in the entire Colorado River basin. The second contribution of this research was to regionalize snow courses stations located in the Upper Colorado River Basin (UCRB) using S-mode PCA and group average cluster analysis. In addition, regional snow pack was reconstructed using PLSR in order to evaluate the regional drought scenarios for the last 475 years (1500-1975). The third contribution of this research was to find the relationship between individual reconstructed ENSO, PDO, AMO, NAO and its combined effect using tree ring reconstructed data of streamflow and regional snowpack of the Colorado River basin.

Chapters 3, 4 and 5 utilized the different statistical method for the reconstruction, regionalization and teleconnection studies. The goal of Chapter 3 was to develop reconstructed data for the last 500 years with the development of an improved reconstruction procedure for all unimpaired streamflow stations located in the Colorado
River basin. Introduction of PLSR for streamflow reconstruction, and comparison with current methods of reconstruction (STPCR and STLR) lead to recommendations of PLSR as an improved prediction model in the field of hydrology. Further, reconstructed streamflow time series for the last 500 years provided important information for water planners that enable them to anticipate the nature of drought in the past so that the changes in water allocations and water conservation measures can be implemented in order to mitigate the negative impacts on the region.

The goal of Chapter 4 was to develop an improved approach of regionalization of snowpack data for the Colorado River basin and reconstruct this regionalized snowpack data for the last 500 years using the improved method of reconstruction (PLSR). In addition to reconstruction, a significant contribution of this part of the research was the development of regional snowpack depth that interprets most of the hydrologic variability in the region. Another important goal of this part of the research was to characterize the droughts in terms of duration for the last 475 years (1500-1975) using three different moving averages (3-year, 5-year and 10-year) of reconstructed regional snowpack, and comparing it with the results based on reconstructed unimpaired water year streamflow. The reconstructed time series of the snowpack provides valuable information for water planners since it represents a storage reservoir in the Colorado River basin.

The goal of the Chapter 5 was to obtain the relationship between individual and coupled effects of reconstructed ENSO, PDO, AMO, and NAO and reconstructed streamflow and regional snowpack considering lag 0, +1, +2 and +3 years. The analysis provided valuable information about the impact of high and low frequency ocean climate phenomena on Colorado River basin. An improved understanding of these coupled
impacts using a longer period of record with the lag approach may lead to improved
drought forecasting.

6.2 Recommendations for Future Research

As a result of the findings of this dissertation study, several future investigations
could be carried out. These include:

1. The identification of further moisture sensitive trees within and adjacent to the
   watershed contributing to the streamflow station being reconstructed (Chapter 3) could
   increase reconstruction skill and capture the regional scale climate variability. In
   addition, developing and considering residual chronologies as predictors instead of
   standard tree ring chronologies may further increase the skill of the regression model.

2. Cook et al. (1999) reconstructed the gridded Palmer Drought Severity Index (PDSI) for
   the entire U.S using tree ring chronologies. But there is a lack of reconstructed
   hydrological variables (streamflow and snowpack) through out the United States. Such
   data sets may be used to evaluate the magnitude, severity, and frequency characteristics
   of hydrologic extreme events in the United States. Further, it may also assist to provide
   important information on spatial and temporal hydrological variability of extreme
   events (floods and droughts) in the United States in a long-term perspective.

3. Currently, reconstructed time series of interdecadal and interannual large-scale patterns
   such as El Niño-Southern Oscillation (ENSO), the Pacific Decadal Oscillation (PDO),
   the Atlantic Multidecadal Oscillation (AMO) and the North Atlantic Oscillation (NAO)
   are available. The study of teleconnection between these reconstructed oceanic climate
phenomenons with the proposed reconstructed streamflow and snow of entire U.S. will allow for a detailed investigation of oceanic climate influences on U.S. hydrology.

4. The extended record of streamflow and snow data can be used as an aid to water resources managers and planners to evaluate the full range of natural water supply variability than is provided by existing instrumental record. Further, the extended hydrologic variability can be used as an input for water budgeting and water allocation models for the reliable and better performance of a water resource system in the Colorado River basin.
APPENDIX A

DATA, SOURCES OF DATA AND SUMMARY
## Appendix - A1  Data, Sources of Data and Summary

<table>
<thead>
<tr>
<th>Data</th>
<th>Sources of Data</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Streamflow</td>
<td>USGS NWISWeb Data retrieval: (<a href="http://waterdata.usgs.gov/nwis/">http://waterdata.usgs.gov/nwis/</a>)</td>
<td>Daily, Monthly and Yearly data of streamflow stations through out the United States</td>
</tr>
<tr>
<td>Snow Course Data of Upper and Lower Colorado River Basin</td>
<td>United States Department of Agriculture (USDA), Natural Resources Conservation Service (NRCS) website (<a href="http://www.wcc.nrcs.usda.gov/snowcourse/">http://www.wcc.nrcs.usda.gov/snowcourse/</a>)</td>
<td>Monthly Snow Water Equivalent (SWE) in inches through out the United states</td>
</tr>
<tr>
<td>Tree ring data</td>
<td>National Climatic Data Center website (<a href="http://www.ncdc.noaa.gov/paleo/treeing.html">http://www.ncdc.noaa.gov/paleo/treeing.html</a>) maintained by the National Oceanic and Atmospheric Administration (NOAA), World Data Center for Paleoclimatology.</td>
<td>Tree-ring data of entire world consists of information on tree species, standard normal width, standard residual width, and wood density measurement with its location.</td>
</tr>
<tr>
<td>Data</td>
<td>Sources of Data</td>
<td>Summary</td>
</tr>
<tr>
<td>------</td>
<td>----------------</td>
<td>---------</td>
</tr>
<tr>
<td>North Atlantic Oscillation (NAO) NAO Index Cook et al. (1998)</td>
<td>National Oceanic and Atmospheric Administration (NOAA) Paleoclimatology Program, World Data Center (WDC) for Paleoclimatology website (ftp://ftp.ncdc.noaa.gov/pub/data/paleo/treering/reconstructions/nao_recon.txt)</td>
<td>The positive numerical value of NAO index was defined as the warm phase NAO while the cold phase has a negative number. Cook et al. (1998) reconstructed Winter NAO index from 1701-1980.</td>
</tr>
</tbody>
</table>
Appendix - A2 Summary of Unimpaired Streamflow Stations in the Colorado River Basin Utilized in this Research

<table>
<thead>
<tr>
<th>Station Number</th>
<th>Station Name</th>
<th>State</th>
<th>Drainage Area (Sq. Km)</th>
<th>Datum El. (m.)</th>
<th>Lat.</th>
<th>Long.</th>
<th>Autocorr -elation (Lag 1)</th>
<th>Data (WY)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9112500</td>
<td>EAST RIVER AT ALMONT</td>
<td>CO</td>
<td>749</td>
<td>2440</td>
<td>38.66</td>
<td>-106.85</td>
<td>0.16</td>
<td>1935-2005</td>
</tr>
<tr>
<td>9119000</td>
<td>TOMICHI CREEK AT GUNNISON</td>
<td>CO</td>
<td>2748</td>
<td>2325</td>
<td>38.52</td>
<td>-106.94</td>
<td>0.26</td>
<td>1938-2005</td>
</tr>
<tr>
<td>9124500</td>
<td>LAKE FORK AT GATEVIEW</td>
<td>CO</td>
<td>865</td>
<td>2386</td>
<td>38.30</td>
<td>-107.23</td>
<td>0.15</td>
<td>1938-2005</td>
</tr>
<tr>
<td>9128500</td>
<td>SMITH FORK NEAR CRAWFORD</td>
<td>CO</td>
<td>111</td>
<td>2161</td>
<td>38.73</td>
<td>-107.51</td>
<td>0.26</td>
<td>1936-1988</td>
</tr>
<tr>
<td>9147500</td>
<td>UNCOMPAHGRE RIVER AT COLONA</td>
<td>CO</td>
<td>1160</td>
<td>1926</td>
<td>38.33</td>
<td>-107.78</td>
<td>0.33</td>
<td>1913-2005</td>
</tr>
<tr>
<td>9239500</td>
<td>YAMPA RIVER AT STEAMBOAT SPRINGS</td>
<td>CO</td>
<td>1564</td>
<td>2041</td>
<td>40.48</td>
<td>-106.83</td>
<td>0.24</td>
<td>1911-2005</td>
</tr>
<tr>
<td>9251000</td>
<td>YAMPA RIVER NEAR MAYBELL</td>
<td>CO</td>
<td>8832</td>
<td>1798</td>
<td>40.50</td>
<td>-108.03</td>
<td>0.29</td>
<td>1917-2005</td>
</tr>
<tr>
<td>9299500</td>
<td>WHITEROCKS RIVER NEAR WHITEROCKS</td>
<td>UT</td>
<td>293</td>
<td>2182</td>
<td>40.59</td>
<td>-109.93</td>
<td>0.27</td>
<td>1930-2005</td>
</tr>
<tr>
<td>9304500</td>
<td>WHITE RIVER NEAR MEEKER</td>
<td>CO</td>
<td>1955</td>
<td>1920</td>
<td>40.03</td>
<td>-107.86</td>
<td>0.31</td>
<td>1910-2005</td>
</tr>
<tr>
<td>9310500</td>
<td>FISH CREEK ABOVE RESERVOIR NEAR SCOFIELD</td>
<td>UT</td>
<td>156</td>
<td>2338</td>
<td>39.77</td>
<td>-111.19</td>
<td>0.21</td>
<td>1939-2005</td>
</tr>
<tr>
<td>9361500</td>
<td>ANIMAS RIVER AT DURANGO</td>
<td>CO</td>
<td>1792</td>
<td>1982</td>
<td>37.28</td>
<td>-107.88</td>
<td>0.09</td>
<td>1928-2005</td>
</tr>
<tr>
<td>9364500</td>
<td>ANIMAS RIVER AT FARMINGTON</td>
<td>NM</td>
<td>3522</td>
<td>1609</td>
<td>36.72</td>
<td>-108.20</td>
<td>-0.06</td>
<td>1931-2005</td>
</tr>
<tr>
<td>9403000</td>
<td>BRIGHT ANGEL CREEK NEAR GRAND CANYON</td>
<td>AZ</td>
<td>262</td>
<td>760</td>
<td>36.10</td>
<td>-112.10</td>
<td>0.11</td>
<td>1924-1973</td>
</tr>
<tr>
<td>9415000</td>
<td>VIRGIN RIVER AT LITTLEFIELD</td>
<td>AZ</td>
<td>13183</td>
<td>538</td>
<td>36.89</td>
<td>-113.92</td>
<td>0.01</td>
<td>1930-2005</td>
</tr>
<tr>
<td>9430500</td>
<td>GILA RIVER NEAR GILA</td>
<td>NM</td>
<td>4828</td>
<td>1419</td>
<td>33.06</td>
<td>-108.54</td>
<td>0.04</td>
<td>1929-2005</td>
</tr>
<tr>
<td>9444500</td>
<td>SAN FRANCISCO RIVER AT CLIFTON</td>
<td>AZ</td>
<td>7164</td>
<td>1047</td>
<td>33.05</td>
<td>-109.30</td>
<td>0.09</td>
<td>1936-2005</td>
</tr>
<tr>
<td>9498500</td>
<td>SALT RIVER NEAR ROOSEVELT</td>
<td>AZ</td>
<td>11153</td>
<td>664</td>
<td>33.62</td>
<td>-110.92</td>
<td>0.16</td>
<td>1914-2005</td>
</tr>
</tbody>
</table>
## Appendix - A3 Summary of Standard Tree-ring Chronologies Utilized in this Research

<table>
<thead>
<tr>
<th>S.</th>
<th>ITRB Site Name</th>
<th>Lat.</th>
<th>Long.</th>
<th>El. (m)</th>
<th>Tree-ring Duration</th>
<th>Auto correl (Lag1)</th>
<th>Tree Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Monarch Lake</td>
<td>40.10</td>
<td>-105.73</td>
<td>2621</td>
<td>1430-1987</td>
<td>0.58</td>
<td>Ponderosa Pine (PIPO)</td>
</tr>
<tr>
<td>2</td>
<td>Schulman Old Tree Number 1 Mesa Verde</td>
<td>37.20</td>
<td>-108.50</td>
<td>2103</td>
<td>1400-1963</td>
<td>0.27</td>
<td>Douglas Fir (PSME)</td>
</tr>
<tr>
<td>3</td>
<td>Eagle</td>
<td>39.65</td>
<td>-106.87</td>
<td>1951</td>
<td>1107-1964</td>
<td>0.60</td>
<td>Douglas Fir (PSME)</td>
</tr>
<tr>
<td>4</td>
<td>Black Canyon of the Gunnison River</td>
<td>38.57</td>
<td>-107.70</td>
<td>2426</td>
<td>1478-1964</td>
<td>0.52</td>
<td>Douglas Fir (PSME)</td>
</tr>
<tr>
<td>5</td>
<td>Upper Gunnison</td>
<td>38.68</td>
<td>-106.87</td>
<td>2530</td>
<td>1322-1964</td>
<td>0.37</td>
<td>Douglas Fir (PSME)</td>
</tr>
<tr>
<td>6</td>
<td>Eagle East (Job 105 Reworked)</td>
<td>39.67</td>
<td>-106.72</td>
<td>2164</td>
<td>1314-1964</td>
<td>0.39</td>
<td>Pinyon Species (PIED)</td>
</tr>
<tr>
<td>7</td>
<td>Dolores</td>
<td>37.58</td>
<td>-108.55</td>
<td>2195</td>
<td>1457-1978</td>
<td>0.57</td>
<td>Douglas Fir (PSME)</td>
</tr>
<tr>
<td>8</td>
<td>Dolores</td>
<td>37.58</td>
<td>-108.55</td>
<td>2195</td>
<td>1270-1978</td>
<td>0.26</td>
<td>Pinyon Species (PIED)</td>
</tr>
<tr>
<td>9</td>
<td>Satan Pass</td>
<td>35.60</td>
<td>-108.13</td>
<td>2286</td>
<td>1381-1972</td>
<td>0.46</td>
<td>Douglas Fir (PSME)</td>
</tr>
<tr>
<td>10</td>
<td>Bobcat Canyon</td>
<td>37.17</td>
<td>-108.52</td>
<td>2042</td>
<td>1390-1971</td>
<td>0.27</td>
<td>Douglas Fir (PSME)</td>
</tr>
<tr>
<td>11</td>
<td>Spruce Canyon</td>
<td>37.18</td>
<td>-108.48</td>
<td>2115</td>
<td>1373-1978</td>
<td>0.40</td>
<td>Douglas Fir (PSME)</td>
</tr>
<tr>
<td>12</td>
<td>Hogback Rampart Hills</td>
<td>37.38</td>
<td>-108.12</td>
<td>3230</td>
<td>1484-1994</td>
<td>0.14</td>
<td>Limber Pine (PIFL)</td>
</tr>
<tr>
<td>13</td>
<td>Pumphouse</td>
<td>39.95</td>
<td>-106.52</td>
<td>2195</td>
<td>1320-1999</td>
<td>0.46</td>
<td>Pinyon Species (PIED)</td>
</tr>
<tr>
<td>14</td>
<td>Land's End</td>
<td>39.00</td>
<td>-108.15</td>
<td>2987</td>
<td>1135-2000</td>
<td>0.60</td>
<td>Douglas Fir (PSME)</td>
</tr>
<tr>
<td>15</td>
<td>Navajo National Monument</td>
<td>36.67</td>
<td>-110.50</td>
<td>2012</td>
<td>1304-1962</td>
<td>0.30</td>
<td>Douglas Fir (PSME)</td>
</tr>
<tr>
<td>16</td>
<td>Defiance West (Defiance/Nazlini)</td>
<td>35.87</td>
<td>-109.43</td>
<td>2134</td>
<td>1474-1965</td>
<td>0.40</td>
<td>Douglas Fir (PSME)</td>
</tr>
<tr>
<td>17</td>
<td>Canyon de Chelly</td>
<td>36.10</td>
<td>-109.38</td>
<td>1828</td>
<td>1376-1972</td>
<td>0.48</td>
<td>Douglas Fir (PSME)</td>
</tr>
<tr>
<td>18</td>
<td>Tseh-Ya-Kin Canyon</td>
<td>36.22</td>
<td>-109.35</td>
<td>1951</td>
<td>1500-1971</td>
<td>0.59</td>
<td>Douglas Fir (PSME)</td>
</tr>
</tbody>
</table>
## Appendix – A3 (Contd.) Summary of Standard Tree-ring Chronologies Utilized in this Research

<table>
<thead>
<tr>
<th>S. N.</th>
<th>ITRB Site Name</th>
<th>Lat.</th>
<th>Long.</th>
<th>Altitude (m)</th>
<th>Tree-ring Duration</th>
<th>Autocorrelation (Lag1)</th>
<th>Tree Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>Betatakin Canyon</td>
<td>36.68</td>
<td>-110.53</td>
<td>2042</td>
<td>1263-1972</td>
<td>0.35</td>
<td>Douglas Fir (PSME)</td>
</tr>
<tr>
<td>20</td>
<td>Shonto Plateau</td>
<td>36.83</td>
<td>-110.73</td>
<td>2134</td>
<td>1365-1971</td>
<td>0.16</td>
<td>Pinyon Species (PIED)</td>
</tr>
<tr>
<td>21</td>
<td>Tsegi Point Road</td>
<td>36.68</td>
<td>-110.53</td>
<td>2196</td>
<td>1490-1972</td>
<td>0.25</td>
<td>Pinyon Species (PIED)</td>
</tr>
<tr>
<td>22</td>
<td>Paria Plateau</td>
<td>36.83</td>
<td>-112.05</td>
<td>1860</td>
<td>1481-1975</td>
<td>0.28</td>
<td>Pinyon Species (PIED)</td>
</tr>
<tr>
<td>23</td>
<td>Wind River Mountains Site D</td>
<td>43.08</td>
<td>-110.07</td>
<td>2500</td>
<td>1492-1972</td>
<td>0.51</td>
<td>Limber Pine (PIFL)</td>
</tr>
<tr>
<td>24</td>
<td>Medicine Bow Peak</td>
<td>41.30</td>
<td>-107.70</td>
<td>3150</td>
<td>1401-1983</td>
<td>0.51</td>
<td>Spruce Species (PCEN)</td>
</tr>
<tr>
<td>25</td>
<td>Unita Mountain</td>
<td>40.78</td>
<td>-110.00</td>
<td>3353</td>
<td>1433-1971</td>
<td>0.71</td>
<td>Spruce Species (PCEN)</td>
</tr>
<tr>
<td>26</td>
<td>La Sal Mountains Site A</td>
<td>38.50</td>
<td>-109.25</td>
<td>2323</td>
<td>1489-1972</td>
<td>0.41</td>
<td>Pinyon Species (PIED)</td>
</tr>
<tr>
<td>27</td>
<td>Kane Spring</td>
<td>37.53</td>
<td>-109.90</td>
<td>1966</td>
<td>1445-1971</td>
<td>0.36</td>
<td>Pinyon Species (PIED)</td>
</tr>
<tr>
<td>28</td>
<td>Navajo Mountain</td>
<td>37.02</td>
<td>-110.85</td>
<td>2286</td>
<td>1469-1971</td>
<td>0.22</td>
<td>Pinyon Species (PIED)</td>
</tr>
<tr>
<td>29</td>
<td>White Canyon</td>
<td>37.62</td>
<td>-110.02</td>
<td>1859</td>
<td>1347-1972</td>
<td>0.44</td>
<td>Douglas Fir (PSME)</td>
</tr>
<tr>
<td>30</td>
<td>Milk Ranch Point</td>
<td>37.62</td>
<td>-109.73</td>
<td>2286</td>
<td>1276-1970</td>
<td>0.38</td>
<td>Pinyon Species (PIED)</td>
</tr>
</tbody>
</table>
Appendix – A3 (Contd.) Summary of Standard Tree-ring Chronologies Utilized in this Research

<table>
<thead>
<tr>
<th>S. N.</th>
<th>ITRB Site Name</th>
<th>Lat.</th>
<th>Long.</th>
<th>Altitude (m)</th>
<th>Tree-ring Duration</th>
<th>Autocorrelation (Lag1)</th>
<th>Tree Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>Uinta Mountains Site D</td>
<td>40.62</td>
<td>-109.95</td>
<td>2289</td>
<td>1423-1971</td>
<td>0.47</td>
<td>Pinyon Species (PIED)</td>
</tr>
<tr>
<td>32</td>
<td>Nine Mine Canyon High</td>
<td>39.78</td>
<td>-110.30</td>
<td>1920</td>
<td>1194-1964</td>
<td>0.44</td>
<td>Douglas Fir (PSME)</td>
</tr>
<tr>
<td>33</td>
<td>Nine Mine Canyon</td>
<td>39.78</td>
<td>-110.30</td>
<td>1920</td>
<td>1201-1949</td>
<td>0.32</td>
<td>Douglas Fir (PSME)</td>
</tr>
<tr>
<td>34</td>
<td>Water Canyon Bryce Canyon National Park</td>
<td>37.67</td>
<td>-112.10</td>
<td>2098</td>
<td>1336-1964</td>
<td>0.52</td>
<td>Ponderosa Pine (PIPO)</td>
</tr>
<tr>
<td>35</td>
<td>Wild Horse Ridge</td>
<td>39.42</td>
<td>-111.07</td>
<td>2805</td>
<td>286-1985</td>
<td>0.41</td>
<td>Gt. basin B.C. Pine (PILO)</td>
</tr>
<tr>
<td>36</td>
<td>San Francisco Peaks A</td>
<td>35.50</td>
<td>-111.67</td>
<td>3535</td>
<td>548-1983</td>
<td>0.69</td>
<td>Rocky Mtn. B.C. Pin (PIAR)</td>
</tr>
<tr>
<td>37</td>
<td>Green Mountain</td>
<td>32.38</td>
<td>-110.68</td>
<td>2194</td>
<td>1460-1986</td>
<td>0.50</td>
<td>Ponderosa Pine (PIPO)</td>
</tr>
<tr>
<td>38</td>
<td>Walnut Canyon</td>
<td>35.17</td>
<td>-111.52</td>
<td>2057</td>
<td>1420-1987</td>
<td>0.51</td>
<td>Ponderosa Pine (PIPO)</td>
</tr>
<tr>
<td>39</td>
<td>Fort Grant Overlook Pinaleno Mountains</td>
<td>32.70</td>
<td>-109.92</td>
<td>2896</td>
<td>1249-1991</td>
<td>0.56</td>
<td>White Pine (PISF)</td>
</tr>
<tr>
<td>40</td>
<td>Snow Bowl San Francisco Peak</td>
<td>35.43</td>
<td>-110.20</td>
<td>3150</td>
<td>1453-1983</td>
<td>0.58</td>
<td>Spruce Species (PCEN)</td>
</tr>
<tr>
<td>41</td>
<td>Reef of Rocks</td>
<td>32.45</td>
<td>-110.78</td>
<td>2550</td>
<td>1321-1998</td>
<td>0.40</td>
<td>Douglas Fir (PSME)</td>
</tr>
<tr>
<td>42</td>
<td>Reef of Rocks</td>
<td>32.70</td>
<td>-109.92</td>
<td>3025</td>
<td>1464-1999</td>
<td>0.42</td>
<td>Douglas Fir (PSME)</td>
</tr>
</tbody>
</table>
Appendix – A4  Summary of snow course stations in the Upper Colorado River basin utilized in this study

<table>
<thead>
<tr>
<th>Sl. Nos.</th>
<th>Station</th>
<th>State</th>
<th>Site Name</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Altitude (m)</th>
<th>Period of Record</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>09M01</td>
<td>Utah</td>
<td>Buckboard Flat</td>
<td>37.87</td>
<td>-109.45</td>
<td>2,743</td>
<td>1930-2005</td>
</tr>
<tr>
<td>2</td>
<td>11J23</td>
<td>Utah</td>
<td>Daniels-Strawberry</td>
<td>40.30</td>
<td>-111.25</td>
<td>2,438</td>
<td>1930-1989</td>
</tr>
<tr>
<td>3</td>
<td>11J07</td>
<td>Utah</td>
<td>East Portal</td>
<td>40.17</td>
<td>-111.18</td>
<td>2,304</td>
<td>1934-1985</td>
</tr>
<tr>
<td>4</td>
<td>10K01</td>
<td>Utah</td>
<td>Indian Canyon</td>
<td>39.90</td>
<td>-110.75</td>
<td>3,367</td>
<td>1930-1989</td>
</tr>
<tr>
<td>5</td>
<td>09J01</td>
<td>Utah</td>
<td>King's Cabin-Upper</td>
<td>40.72</td>
<td>-109.55</td>
<td>2,661</td>
<td>1930-1989</td>
</tr>
<tr>
<td>6</td>
<td>10J10</td>
<td>Utah</td>
<td>Lakefork #1</td>
<td>40.60</td>
<td>-110.43</td>
<td>3,078</td>
<td>1931-1989</td>
</tr>
<tr>
<td>7</td>
<td>09L01</td>
<td>Utah</td>
<td>Lasal mountain -Lower</td>
<td>38.48</td>
<td>-109.28</td>
<td>2,682</td>
<td>1931-2005</td>
</tr>
<tr>
<td>8</td>
<td>11K09</td>
<td>Utah</td>
<td>Seeley Creek</td>
<td>39.32</td>
<td>-111.43</td>
<td>3,048</td>
<td>1932-1989</td>
</tr>
<tr>
<td>9</td>
<td>11J08</td>
<td>Utah</td>
<td>Strawberry Divide</td>
<td>40.18</td>
<td>-111.22</td>
<td>2,560</td>
<td>1934-1989</td>
</tr>
<tr>
<td>11</td>
<td>05K06</td>
<td>Colorado</td>
<td>Arrow</td>
<td>39.92</td>
<td>-105.77</td>
<td>2,950</td>
<td>1938-1989</td>
</tr>
<tr>
<td>12</td>
<td>05K03</td>
<td>Colorado</td>
<td>Berthoud Pass</td>
<td>39.83</td>
<td>-105.75</td>
<td>2,957</td>
<td>1936-2006</td>
</tr>
</tbody>
</table>
Appendix – A4 (Contd.) Summary of snow course stations in the Upper Colorado River basin utilized in this study

<table>
<thead>
<tr>
<th>Sl. Nos.</th>
<th>Station</th>
<th>State</th>
<th>Site Name</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Altitude (m)</th>
<th>Period of Record</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>07K02</td>
<td>Colorado</td>
<td>Burro Mountain</td>
<td>39.88</td>
<td>-107.60</td>
<td>2,865</td>
<td>1936-2000</td>
</tr>
<tr>
<td>14</td>
<td>07M05</td>
<td>Colorado</td>
<td>Cascade</td>
<td>37.65</td>
<td>-107.80</td>
<td>2,694</td>
<td>1936-1989</td>
</tr>
<tr>
<td>15</td>
<td>07L01</td>
<td>Colorado</td>
<td>Crested Butte</td>
<td>38.88</td>
<td>-107.00</td>
<td>2,719</td>
<td>1936-2000</td>
</tr>
<tr>
<td>16</td>
<td>06J01</td>
<td>Colorado</td>
<td>Dry Lake</td>
<td>40.53</td>
<td>-106.78</td>
<td>2,560</td>
<td>1936-1989</td>
</tr>
<tr>
<td>17</td>
<td>06J15</td>
<td>Colorado</td>
<td>Elk River #2</td>
<td>40.85</td>
<td>-106.97</td>
<td>2,652</td>
<td>1936-1989</td>
</tr>
<tr>
<td>19</td>
<td>05J10</td>
<td>Colorado</td>
<td>Lake Irene</td>
<td>40.42</td>
<td>-105.82</td>
<td>3,261</td>
<td>1938-2000</td>
</tr>
<tr>
<td>20</td>
<td>05K07</td>
<td>Colorado</td>
<td>Lapland</td>
<td>39.88</td>
<td>-105.88</td>
<td>2,835</td>
<td>1938-2000</td>
</tr>
<tr>
<td>21</td>
<td>07M03</td>
<td>Colorado</td>
<td>Lizard Head</td>
<td>37.80</td>
<td>-107.93</td>
<td>3,109</td>
<td>1936-2006</td>
</tr>
<tr>
<td>22</td>
<td>06J06</td>
<td>Colorado</td>
<td>Lynx Pass</td>
<td>40.08</td>
<td>-106.67</td>
<td>2,707</td>
<td>1936-1989</td>
</tr>
<tr>
<td>23</td>
<td>08K04</td>
<td>Colorado</td>
<td>Mesa Lakes</td>
<td>39.05</td>
<td>-108.08</td>
<td>3,048</td>
<td>1937-2000</td>
</tr>
<tr>
<td>24</td>
<td>06K12</td>
<td>Colorado</td>
<td>Middle Fork Campground</td>
<td>39.78</td>
<td>-106.02</td>
<td>2,743</td>
<td>1936-2006</td>
</tr>
<tr>
<td>25</td>
<td>05J09</td>
<td>Colorado</td>
<td>North Inlet Grand Inlet</td>
<td>40.28</td>
<td>-105.77</td>
<td>2,743</td>
<td>1938-2006</td>
</tr>
<tr>
<td>26</td>
<td>07K01</td>
<td>Colorado</td>
<td>North Lost Train</td>
<td>39.07</td>
<td>-107.15</td>
<td>2,804</td>
<td>1936-1990</td>
</tr>
</tbody>
</table>
Appendix – A4 (Contd.) Summary of snow course stations in the Upper Colorado River basin utilized in this study

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>Station</th>
<th>State</th>
<th>Site Name</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Altitude (m)</th>
<th>Period of Record</th>
</tr>
</thead>
<tbody>
<tr>
<td>27</td>
<td>07K06</td>
<td>Colorado</td>
<td>Park Reservoir</td>
<td>39.05</td>
<td>-107.88</td>
<td>3,036</td>
<td>1940-1990</td>
</tr>
<tr>
<td>28</td>
<td>05J04</td>
<td>Colorado</td>
<td>Phantom Valley</td>
<td>40.40</td>
<td>-105.85</td>
<td>2,752</td>
<td>1936-1990</td>
</tr>
<tr>
<td>29</td>
<td>06L03</td>
<td>Colorado</td>
<td>Porphyry Creek</td>
<td>38.48</td>
<td>-106.33</td>
<td>3,280</td>
<td>1940-2006</td>
</tr>
<tr>
<td>30</td>
<td>08M05</td>
<td>Colorado</td>
<td>Rico</td>
<td>37.67</td>
<td>-108.03</td>
<td>2,652</td>
<td>1936-1992</td>
</tr>
<tr>
<td>31</td>
<td>07J01</td>
<td>Colorado</td>
<td>Rio Blanco</td>
<td>40.03</td>
<td>-107.28</td>
<td>2,591</td>
<td>1936-2000</td>
</tr>
<tr>
<td>32</td>
<td>07M02</td>
<td>Colorado</td>
<td>Telluride</td>
<td>37.93</td>
<td>-107.80</td>
<td>2,682</td>
<td>1936-2006</td>
</tr>
<tr>
<td>33</td>
<td>07K05</td>
<td>Colorado</td>
<td>Trickle Divide</td>
<td>39.13</td>
<td>-107.90</td>
<td>3,048</td>
<td>1940-1985</td>
</tr>
<tr>
<td>34</td>
<td>06M03</td>
<td>Colorado</td>
<td>Upper San Juan</td>
<td>37.48</td>
<td>-106.83</td>
<td>3,109</td>
<td>1936-2000</td>
</tr>
<tr>
<td>35</td>
<td>06J05</td>
<td>Colorado</td>
<td>Willow Creek Pass</td>
<td>40.35</td>
<td>-106.10</td>
<td>2,908</td>
<td>1938-2000</td>
</tr>
<tr>
<td>36</td>
<td>10F15</td>
<td>Wyoming</td>
<td>Kendall R.S.</td>
<td>43.25</td>
<td>-110.02</td>
<td>2,359</td>
<td>1938-1992</td>
</tr>
<tr>
<td>37</td>
<td>10F16</td>
<td>Wyoming</td>
<td>Loomis Park</td>
<td>43.17</td>
<td>-110.13</td>
<td>2,512</td>
<td>1939-1991</td>
</tr>
<tr>
<td>38</td>
<td>06H10</td>
<td>Wyoming</td>
<td>Old Battle</td>
<td>41.15</td>
<td>-107.97</td>
<td>3,024</td>
<td>1936-1990</td>
</tr>
<tr>
<td>39</td>
<td>10G13</td>
<td>Wyoming</td>
<td>Snider Basin</td>
<td>42.47</td>
<td>-110.53</td>
<td>2,457</td>
<td>1937-1992</td>
</tr>
</tbody>
</table>
APPENDIX B

PARTIAL LEAST SQUARE REGRESSION
Partial Least Square Regression

Descriptions

The general theory of Partial Least Square Regression (PLSR) is to extract latent factors, accounting for as much of the variation as possible while modeling the responses. For this reason, the acronym PLS has also been taken to mean "projection to latent structure." The common procedure in PLSR is to search for a set of components (latent vectors) that explains the maximum covariance between (X) and (Y) which is followed by a regression step where the decomposition of (X) is used to predict (Y) (Abdi, 2003). In PLSR, the principal component scores of both (X) and (Y) are used in lieu of the original data to develop the regression model. PLSR determines the components from (Y) that are also relevant for (X) (Abdi, 2003).

In principle, MLR can be used with very many factors. However, if the number of factors gets too large (for example, greater than the number of observations), it is likely to get a model that fits the sampled data perfectly but that will fail to predict new data well. This phenomenon is called over-fitting. In such cases, although there are many remarkable factors, but only a few underlying or latent factors account most of the variation in the response.

In case of PLSR, the PCA of the matrix (X) (i.e., the matrix of predictors or independent variables) decomposes (X) into a score matrix (T) times a loading matrix (P) and a residual (i.e., error) matrix (E) (Wold et al., 1987).

\[ X = T \ast P' + E \]  

(1)
Similarly, \( Y \) is decomposed into a score matrix \( U \) times a loading matrix \( R \) and a residual matrix \( F \).

\[
Y = U \ast R' + F \quad (2)
\]

These equations (1 and 2) are commonly referred to as the outer relations (Geladi and Kowalski, 1986). The objective of the PLSR model is to minimize \( F \) while maintaining the correlation between \( X \) and \( Y \), referred to as the inner relation \( U \) (Geladi and Kowalski, 1986).

\[
U = B \ast T + H \quad (3)
\]

Where \( H \) represents the error, \( B \) is a diagonal matrix explaining the correlation between \( X \) and \( Y \). When equation (3) is inserted into equation (2), a predictive relation for \( Y \) is developed where \( F* \) represents the error.

\[
Y = T \ast R'B + F* \quad (4)
\]

Equation (4) is sometimes referred to as the mixed relation where \( F* \) is to be minimized (Geladi and Kowalski, 1986).
Figure-1 gives outline of the PLSR method (Tobias, 1995). The overall objective is to use the factors (predictors) to predict the responses (predictand). This can be achieved by extracting latent variables T and U from sample factor space of factors and responses respectively. First, factors T also known as X-scores are used to predict Y-scores (U). Then predicted Y-scores (U) are used to predict for the responses.
There are several methods of extracting precise (optimal) number of factors. One approach is to choose by some heuristic approach based on the amount of residual variation. Another approach is to construct the PLS model for a given number of factors on one set of data and then to test it on another, and choosing the number of extracted factors for which the total prediction error is minimized. Further, Van der Voet (1994) recommended choosing the least number of extracted factors whose residuals are not significantly greater than those of the model with minimum error. Another important approach is an approach of cross validation which is strongly recommended if there is lack of enough convenient test data set. In this approach, each observation can be used in turn as a test set; this is known as cross validation (Tobias, 1995).

The prediction residual sum of squares (PRESS) statistic is a cross validation calculation that determines the minimum (optimum) number of components required (Geladi and Kowalski, 1986). The cross-validation consists of removing a row (or multiple rows) from the data matrix and then completing the eigen analysis on the reduced matrix. Target testing is then performed on the removed rows using the various levels of the abstract factor space and the difference between the target points and the predicted points is calculated (Malinowski, 2002). This process is repeated until every row has been deleted once and the errors in the target fit for each row are summed (Malinowski, 2002). Several methods are available to perform PLSR including the nonlinear iterative partial least squares (NIPALS) approach. NIPALS is advantageous due to calculation speed and simplicity (Wold et al., 1987), if the model is primarily concerned with the first few principal components, and this method was utilized in this research.
PROC IMPORT OUT= Work.streamflow_treering
   DATAFILE="C:\Documents and Settings\streamflow_treering.xls"
   DBMS=EXCEL REPLACE;
   SHEET="Sheet1$";
   GETNAMES=YES;
   MIXED=NO;
   SCANTEXT=YES;
   USEDATE=YES;
   SCANTIME=YES;
RUN;

/*The following SAS CODE block sets the parameters to invoke SAS macros*/
% global xvars yvars predname resname xscrname yscrname num_x num_y lv;
% let title1=PLS Reconstruction of Streamflow;
% let xvars=T9 T7 T28 T10 T26 T27 T14 T6 T18 T17; /*These define ranked tree
   ring chronologies*/
% let yvars=Q1; /* This defines streamflow time series (predictands)*/
% let ypred=Qhat1; /* This defines predicting streamflow*/
% let yres=Qres1; /*This defines the residual value of streamflow*/
% let xscrname=xscr; /*This defines x-score*/
% let yscrname=yscr; /*This defines y-score*/
% let num_y=1; /*This defines number of predictand*/
% let num_x=10; /*This defines total number of predictors*/

/*The Following SAS Code block gives the minimum cross validated statistics and
number of latent variable associated for the calibration period. Method ITER represents
the iterative approach (NIPALS) of PLSR*/

data Work.streamflow_treering_cal; set Work.treering ; /*This command picks the data
of streamflow and tree ring for calibration period from original data set*/
if Year GT 429;
run;

PROC PLS data=Work.streamflow_treering_cal method=pls (ITER) cv = split (35)
cvtest (stat=press);
model Q1 = T9 T7 T28 T10 T26 T27 T14 T6 T18 T17;
output out=outpls p=Qhat1 press=press t2=t2; /*This command gives the minimum
press statistics, optimal latent variables and one drop cross validation statistics */
run;

/* The following block develops the calibration equation*/
proc pls data=Work.streamflow_treering_cal
method = pls (ITER) outmodel=est1
lv=3 details; /*This command gives the regression coefficients for calibration period based on the optimal number of latent variables*/
model Q1= T9 T7 T28 T10 T26 T27 T14 T6 T18 T17/solution;
run;

/*The following SAS code develops prediction model based on minimum cross validated statistics and corresponding number of latent variable*/
data Work.streamflow_treering_pred; set Work.Treering;
if Year LT 430 then do; /*This picks the data for prediction period*/
  Q1=.;
end;
proc pls data= Work.streamflow_treering_pred
  method=pls (ITER)
  lv=3;
model &yvars= T9 T7 T28 T10 T26 T27 T14 T6 T18 T17;
output out=outpls2 p=Qhat1 yresidual = yres1;
run;

/* The following SAS code give the output result and merges the data for output*/
data outpls2a; set outpls2 (keep=Qhat1);
  YEAR = _N_;
run;
data Work.streamflow_treering_c; set Work.streamflow_treering (keep=Q1); YEAR = _N_; run;

data Work.predict; merge Work.streamflow_treering_c outpls2a; YEAR = _N_; run;
data Work.predict; set Work.predict;
yresl=Q1-Qhat1; /* This obtains difference between original streamflow and predicted streamflow for the calibration period*/
run;
proc print data=Work.predict;
run;

---------------------------------------------------------------
APPENDIX C

ROTATED PRINCIPAL COMPONENT ANALYSIS AND AVERAGE CLUSTER ANALYSIS
Principal Component Analysis and Cluster Analysis

Descriptions

Principal component analysis and cluster analysis are two important methods to obtain the coherent modes of various climate parameters.

Principal Component Analysis (PCA)

The PCA is widely used technique in meteorology and climatology. If there is very large data set, it is important to reduce its size in order to understand and interpret the structure of data. The first objective of using PCA is to reduce the dimensionality of data without loosing any information, and these reduced numbers of transformed variables can be used in consequent analysis. Second objective of PCA is to identify the meaning underlying variables that can statistically interpret the original data variables. These underlying variables can be used in variety of statistical analysis such as data screening, assumption checking and cluster verifying (Johnson, 1998).

The Spatial regionalization based on principal components (eigenvectors) of climate variables is called S-mode PCA (Richman, 1986). In case of regionalizing of climate variables using S-mode PCA, first the eigen values and eigenvectors of the correlation or covariance matrix of the time series and loading matrix are computed. In case of S-mode analysis, the eigen vectors have a component for each station and represent an orthogonal spatial pattern. The loading matrix represents the correlation of the original variables with the principal components. The loading matrix \( C \) can be obtained using:

\[
C = A\lambda^{1/2}
\]  

(1)
where $A$ is an orthogonal matrix of the eigenvector of the correlation matrix $R$ and $\lambda$ is the diagonal matrix of eigenvalues of the correlation matrix $R$.

Richman (1986) explained different techniques of rotating PCA in the field of climatology. The objective of rotation is to distinguish the clear pattern of loading and, typical rotational methods are varimax and quartimax (Baeriswyl and Rebetez, 1997). The varimax method was utilized to rotate the components and this method allows the maximization of variances.

**Cluster Analysis**

Cluster analysis is another most widely used statistical technique of classification. The cluster analysis procedure first finds the distance between individual stations, or group of stations, using the smallest Euclidean distance. Euclidean distance is the geometric distance in multidimensional space. The cluster is formed considering the smallest possible distance, and then proceeds with another cluster considering the distance between the new station and already formed cluster. The Euclidean distance $(d_{r,s})$ is a measure of the dissimilarity between two pairs of objects (clusters) $x_r$ and $x_s$, in an $n*p$ data matrix:

$$d_{r,s} = \sqrt{(x_r - x_s)^2}$$

(2)

where, $n$ is the number of observations and $p$ is number of variables. Fovell and Fovell (1993) provided a detailed description of group average linkage cluster analysis. The result of the principal component was corroborated with the group average linkage cluster analysis (Piechota et al, 1997; Baeriswyl and Rebetez, 1997).
/*Part A: A sample SAS Code of Running Rotated Principal Component Analysis based on Dallas E. Johnson (1998)*/

TITLE 'Varimax Rotation Principal Component Analysis-SAS procedure';

/*Input data of April 1 Snow water equivalent (SWE) for stations 1-39*/

DATA SWE;

INPUT Year SW1 SW2 SW3 SW4 ..................SW39;

CARDS;

/*data*/

/*The following SAS code conducts Rotated Principal Component Analysis using varimax rotation */

TITLE 'Varimax Rotation Principal Component Analysis';

PROC FACTOR DATA=SWE METHOD=PRIN PRIORS=ONE ROTATE=VARIMAX
EV SCREE FLAG = 0.4;

VAR SW1 -- SW39;

RUN;

--------------------------------------------------------------------------------------------------------------------------
/*Part B: A sample SAS Code of Running Average Cluster Analysis based on Johnson (1998)*/

TITLE 'Research Cluster Analysis- SAS PROCEDURE';

DATA SWE;
INPUT Year SW1 SW2 SW3 SW4 ...SW39;
CARDS;
/*data input*/

PROC CLUSTER DATA=SWE S METHOD=AVERAGE CCC PSEUDO OUTTREE=TREE;
VAR SW1—SW39;
ID Year;
RUN;

/*The following SAS Code gives the statistical information for the average clustering process (1998)*/

PROC PLOT DATA=TREE;
PLOT CCC*_NCL=_NCL_/HAXIS=0 TO 16 BY 2;
RUN;
APPENDIX D

NON-PARAMETRIC RANK SUM TEST
Non-parametric Rank sum test

The Non-parametric rank sum test compares the two independent data sets and determines if one data set has significantly larger or smaller values than the other data sets (Maidment, 1993). The two data sets can vary in size, and the data may not be necessarily normally distributed. It does not assume any form of linear relationship between two data sets as is necessary in correlation analysis. Applying the rank sum test, it is possible to determine the significant median difference in water year streamflow or snow pack with respect to individual and coupling effect of oceanic/atmospheric climate phenomenon.

% A Sample - Matlab Program to run Non-parametric Rank Sum Test

Q = xlsread('1.xls');

% Streamflow Number One
% Independent effect of Oceanic oscillation – data input
ELNINO = [Q(8) Q(14) .... Q(261) Q(263) Q(265)];

% The streamflow water year time series during the El Niño – data input
LANINA = [Q(12) Q(20) ........ Q(241) Q(246) Q(262)];

% The streamflow water year time series during the PDO cold and PDO warm
PDOC = [Q(1) Q(2) ........ Q(263) Q(264) Q(265)];
PDOW = [Q(6) Q(7) ........ Q(239) Q(240) Q(241)];
% Significance level due to Independent effect of Oceanic oscillation

[p1, h11, stats1] = ranksum(LANINA, ELNINO); % this function does the rank sum test between LANINA and ELNINO

[p2, h22, stats2] = ranksum(PDOC, PDOW); % this function does the rank sum test between PDOC and PDOW

m1=median(LANINA)-median(ELNINO); % this function finds median difference in order to find the sign of significance

m2=median(PDOC)-median(PDOW);

% Output

P = [p1 p2];

M = [m1 m2];

xlswrite('results', P, 'sheet1', 'A1');

xlswrite('results', M, 'sheet2', 'A1');

% (1-P)*100 value represents the significance level

% M gives the difference in median and sign of it indicates positive or negative significance
REFERENCES


Garen, D.C., 1992. Improved Techniques in Regression-based Streamflow Volume
Forecasting. Journal of Water Resources Planning and Management, 118:654-670


Chimica Acta, 185:1-17.


Yellowstone River Flow and Teleconnections with Pacific Basin Climate Variability

reconstruction of the Atlantic Multidecadal Oscillation since 1567 A.D.. Geophysical

interannual to decadal scale precipitation variability for northeastern Utah since 1226

reconstruction of precipitation in the Bighorn Basin, Wyoming, since 1260 A.D..
Journal of American Meteorological Society, 17:3855-3865


172


Stahle, D.W., R.D. D'Arrigo, P.J. Krusic, M.K. Cleaveland, E.R. Cook,
R.J. Allan, J.E. Cole, R.B. Dunbar, M.D. Therrell, D.A. Gay, M.D. Moore,
Experimental dendroclimatic reconstruction of the Southern Oscillation.
Bulletin of American Meteorological Society 79:2137-2152

Trends in the Upper Colorado River Basin. Lake Powell Research Project, Bulletin
No. 18, National Science Foundation, 70 Pp.

European Summer Climate. Science, 309:115-118.

Association, in press.

of the Twentieth Annual SAS Users Group International Conference, Cary,


Touchan, R., D.M. Meko and M.K. Huges, 1999. A 396-year reconstruction of
precipitation in southern Jordan. Journal of the American Water Resources


VITA

Graduate College
University of Nevada, Las Vegas

Janak Timilsena

Home Address:
1600 E. University Avenue, #109
Las Vegas, Nevada, 89119

Degrees:
Bachelor of Civil Engineering, 1992
University of Mysore, India

Master of Civil Engineering, 1997
Asian Institute of Technology, Thailand

Dissertation Title: Reconstruction of Hydrologic and Climatic Variability in the Colorado River Basin

Dissertation Examination Committee:
Chairperson: Dr. Thomas C. Piechota, Ph.D.
Committee Member, Dr. David James, Ph.D.
Committee Member, Dr. Ashok Singh, Ph.D.
Committee Member, Dr. Jacimaria Batista, Ph.D.
Graduate Faculty Representative, Dr. Matthew S. Lachniet, Ph.D.