ABSTRACT

PATSKOSKI, JASON. Predicting Streamflow in the Southeastern United States using SST and Tree Ring Chronologies. (Under the direction of Sankarasubramanian Arumugam.)

Long time series of annual streamflow are required for the planning and management of water resources. However, observed streamflow records within the continental United States typically start from 1930. Hence, studies have considered paleo-climatic data and tree rings for extending the observed streamflow records over a longer period of time. The most common approach to reconstruct annual streamflow is to develop a statistical regression relationship between the principal components of the tree rings and the observed annual flow values and then extend the relationship to reconstruct annual streamflow values over the period for which tree ring data is available. However, this approach has limited skill in reconstructing high flow values since tree ring growth reaches its potential limit during wet years. We propose an alternate approach to overcome this limitation by combining information from Sea Surface Temperature (SST) and from tree ring chronologies. Given the role of El Nino Southern Oscillation (ENSO) in influencing hydroclimatology over the Southeastern US, we predict the periodic component of streamflow using Nino 3.4 – an index representing ENSO – and the non-periodic component of streamflow using the non-periodic interannual variability in the tree rings. We employ Multi-Channel Singular Spectrum Analysis (MSSA) on tree rings for streamflow reconstruction, since MSSA can extract low-dimensional components as well as the periodic components associated with a hydroclimatic data. The proposed MSSA based approach is tested with the traditional principal component regression (PCR) approach based on Leave-Five-Percent-Out-Cross-Validation. Results from the study show that MSSA approach reduces the error in reconstructing streamflow and the hybrid approach (SSTs and tree rings) perform better during high flow years. Combination

of the annual streamflow estimates from these three models – PCR, MSSA and the hybrid approach– results in improved reconstruction of annual streamflow.

Predicting Streamflow in the Southeastern United States using SST and Tree Ring Chronologies

by
Jason Patskoski

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Master of Science

Civil Engineering

Raleigh, North Carolina

2012

APPROVED BY:

_____	_____
Sankarasubramanian Arumugam Committee Chair	Emily Zechman
_____	
E. Downey Brill	

DEDICATION

I give God all the glory.

BIOGRAPHY

Jason Patskoski was born on November 18, 1987 in Cocoa Beach, Florida. He and his family moved to the once prosperous mill town of Greer, South Carolina when he was four years old. In Greer, he lived in house that backed up to the woods and spent most of his childhood climbing trees, making forts and swinging on vines. He lived in Greer until he graduated from high school. Greer is the place where he learned many of important life skills such as how to read, ride a bike, start a fire with a magnifying glass, balance a checkbook, shoot a gun, throw a curve ball, patch drywall, drive a stick shift, and post up a defender. In June of 2006 he graduated from Greer High School. During his time there he had the ability to dunk a basketball and run a quarter mile in fifty-one seconds.

After high school graduation, Jason attended Clemson University. While at Clemson, he attended many football and basketball home games, often camping out days for tickets. He still refers to his blood type as "orange." Jason had an amazing experience at Clemson in every aspect. While there, he learned the art of computer programming, a skill to which he attributes much of his success. Other things he learned at Clemson include fluid mechanics, hydrology, statistics, how to play rugby and coincidentally how to use crutches. In May of 2010, he received a Bachelor of Science in Civil Engineering with a concentration in Applied Fluid Mechanics.

He then moved to Raleigh, North Carolina to pursue his Masters of Science in Civil Engineering at North Carolina State University.

ACKNOWLEDGMENTS

Most importantly, I would like to thank God who makes everything possible. Without his strength, I would not have been able to complete this work. I am truly blessed. As it says in Proverbs 16:9, “In his heart, a man plans his course, but the Lord determines his steps.” I would also like to thank my advisor, Dr. Sankar Arumugam, for the opportunity to work with him. Not many people that grew up where I did get the chance to go to graduate school, and I would like to express my appreciation to Dr. Arumugam for believing in me. I would also like to thank him for his guidance and patience with me over the past two years. I truly do not know how I could have accomplished this task without his consistent teaching and direction throughout the process. I cannot thank him enough. I want to express my gratitude to the other committee members, Dr. Zechman and Dr. Brill, for their support during my time here at N.C. State. I would like to thank Dr. Kaye at Clemson University for his guidance and support of my future and opportunity for undergraduate research during my undergraduate career.

I wish to give my most sincere thanks to my parents, Mark and Mary Patskoski, for everything they have done for me throughout the years. They have sacrificed a lot in order to do what they felt was best for my wellbeing and future. They were always there to give me support, encouragement, and unconditional love. I would like to especially thank my dad for always being willing to do things like play catch with me or take me fishing with me when I was a kid. I realize how lucky I am to have a dad like that. I also appreciate him and my mom teaching me the importance and value of work ethic. I definitely could not have achieved what I have without that value. Thanks mom and dad for raising me the right way.

I would also like to thank all my friends I have met while in graduate school, especially those in “The Domination Station” and on the “Honey Badgers” intramural team and its affiliates. You guys were always there to celebrate with me in the good times, willing to grab a beer and help me out in the not so good times and consistently went to eat General Tso’s just about every week. I hope I can repay you in the same fashion.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Introduction

Annual streamflow is essential for water supply planning and management. For instance, most of the design of water supply systems depends on annual streamflow information to estimate the safe yield of the system. But, in most sites over the Southeastern United States, observed streamflow data is not available before 1930's. This implies that considerable uncertainty could arise in the estimates of 90% reliable yield from a reservoir obtained from 80-year annual streamflow record. Hence, studies have focus on extending the streamflow record beyond the observed realm [Cook et al. 1983; Woodhouse et al. 2001, 2006; Gangopadhyay et al. 2009]. Studies have shown that incorporating paleo-climatic flood records with observed flood records resulted in improved design flood estimates corresponding to rare return period events [Hosking et al. 1986; Stedinger et al. 1986]. One approach to extend the streamflow beyond the observed period is by reconstructing streamflow from other proxy climatic data such as tree rings. Creating reconstructed streamflow records can give insight into past variability and could augment the instrumental records. This study focuses on the relationship between annual streamflow, tree ring chronologies, and sea surface temperatures in order to develop low dimensional annual streamflow reconstruction models.

This thesis is organized as follows: Chapter 1 presents the background, literature, and research objectives. Chapter 2 provides the hydroclimatic, annual streamflow, and water quality databases along with the tree ring chronologies used in this study. Chapter 3 develops and evaluates the skill of three annual streamflow reconstruction models and their

combination. Chapter 4 provides a summary of the key findings from the study and discusses future research directions.

1.1 Annual Streamflow Reconstruction

Each year, trees grow a new tree ring in the center of the trunk, where the width is dependent upon the growth of that tree during that particular year. The growth of trees is dependent on many factors: amount of sunlight, air temperature, root depth, and most importantly for this study, moisture availability. Due to these relationships along with the long life of existing trees, tree ring chronologies can be used for the reconstruction of many hydroclimatic factors [Fritts, 1976].

It is well established that annual streamflow is a hydroclimatic factor that can be reconstructed with tree ring chronologies at a high level of skill [Fritts 1976, 1991; Stockton et al. 1976; Hidalgo et al. 2000; Woodhouse et al. 2006]. The basic idea is to develop a statistical relationship between the observed annual streamflow records (predictand) and the tree ring chronologies (predictor) for the corresponding period, so that the statistical relationship could be employed for estimating the annual streamflow beyond the observed period based on the past tree ring chronologies. The annual streamflows within Upper Colorado River Basin were reconstructed using the Principal Components (PCs) of tree ring chronologies based on forward stepwise regression [Stockton et al. 1976; Fritts 1991; Hidalgo et al. 2000; Woodhouse et al. 2006]. Past studies indicate that this approach has become the most popular technique in annual streamflow reconstruction [Fritts et al. 1990, Brockway et al. 1995].

1.2 Limitations in using Tree Ring Chronologies

One limitation in using tree ring chronologies for streamflow reconstruction is the underestimation of high flow values since tree ring growth is often limited by soil properties as well as by the metabolic growth rate of the tree [Meko et al. 1995]. In addition, high flow values are commonly due to high precipitation. High precipitation conditions are accompanied by a high amount of cloud cover which causes lower temperatures and a reduction in sunlight, both of which are limiting factors in tree growth [Fritts 1976]. Recent studies on annual streamflow reconstruction show tree ring chronologies yield conservative reconstructions of extreme events [Margolis et al. 2011].

Another limitation in the use of tree ring PCs in annual streamflow reconstruction stems from process of obtaining the relationships of the PCs. Since PCA determines only the relationship between variables at the same time step, tree ring PCs have a limited ability to contain the impact of climactic factors from previous time periods. It is well documented that groundwater recharge is a significant component of annual streamflow. However, it is possible for the residence time of groundwater in areas with large aquifer volumes to be long enough to result in the time between a change in available moisture indicated in tree growth and change in annual streamflow to stretch over a year. Furthermore, tree ring chronologies have a high autocorrelation due to biological carryover decreasing the signal of available moisture in tree ring chronologies [Hidalgo et al. 2000]. One approach that is commonly employed to address this is by the inclusion of lagged tree ring chronologies in the regression model [Fritts 1976, 1991; Meko 1997; Hidalgo et al. 2000].

The growth season of a tree is also a limitation in the skill of tree ring chronologies in annual streamflow reconstruction. Tree growth occurs during spring and summer months which imply that there is limited skill in the tree ring chronologies' ability to capture the available moisture due to precipitation in the autumn and winter months. In previous annual streamflow reconstruction studies, such as the Upper Colorado River Basin, this limitation is reduced due to both autumn and winter precipitation occurring as snow. Snowmelt does not occur until the spring months, making the precipitation from the autumn and winter months observable in the tree ring chronologies from these areas [Fritts 1976]. However, in the southeast, little winter precipitation results in snow accumulation. Furthermore, melting of that accumulation usually develops a few days after the snowfall, leaving tree ring chronologies in the southeast little chance of observing autumn and winter precipitation.

A final limitation of tree ring chronologies is the small portion of variance in a chronology reflecting past hydroclimatic variations [Fritts 1991]. The effect of non-hydroclimatic variance, noise, often is much larger than the signal from hydroclimatic conditions. Tree ring data collection procedure includes steps to reduce this limitation such as removing trends due to age and cross-dating the samples. It is important to note that these procedures do not ensure the signal from the tree rings will be primarily due to a particular hydroclimatic attribute such as precipitation or soil moisture.

1.3 Research Objectives

The overall goal of this study is to improve the skill of annual streamflow reconstructions in the southeast. This will be achieved by developing models that reduce the

limitations of tree ring chronologies. The limitation of tree rings' inability to capture the impact of climactic factors from previous time periods can be improved by using Multi-Channel Singular Spectrum Analysis (MSSA) to obtain predictors for annual streamflow reconstruction. MSSA identifies unknown and partially known components of a time series using lagged covariance resulting in Reconstructed Components (RCs) containing information from a current time period and previous time periods. Previously, this limitation has been addressed by performing including lagged chronologies among variables in the PCA [Fritts 1976, Hidalgo et al. 2000]. However, MSSA identifies and reconstructs trends in the data through the covariance of many lags in the data resulting in corresponding RCs. The RCs should perform better in regression as they are representative of the patterns of the raw data as the traditional approach only uses the lags of the raw measurements. Furthermore, since MSSA separates the data into many different trends and there is a long length of tree chronologies available, it is likely that the RCs will represent trends due to different hydroclimatic factors, precipitation, temperature, sunlight, etc. In annual streamflow reconstruction, the highest correlated hydroclimatic factor should be precipitation, so the regression model should choose those corresponding RCs, leading to an improvement in reconstruction.

Another approach that will reduce the limitations of annual streamflow reconstruction using tree ring chronologies is considering Sea Surface Temperatures (SSTs) in addition to the tree ring chronologies as model predictors. The relationship between SSTs and annual streamflow has been well established using many SST anomaly indices. El Nino Southern Oscillation (ENSO) is an SST anomaly in the Pacific Ocean highly correlated to annual

streamflow values in parts the Southeast, and many studies have related the effect of the ENSO state on streamflow sites over the region [Kahya and Dracup 1993]. ENSO is a periodic SST anomalous condition in the tropical Pacific having a frequency of three to seven years. Since ENSO is the primary mechanism by which wet and dry years are modulated, the inclusion of ENSO in streamflow reconstruction should improve the prediction during high flow years. Unlike tree ring chronologies, the ENSO index has no upper limit due to biological processes making it a better predictor for high annual flow values since annual streamflow is also not limited by non-physical processes. The use of ENSO will also reduce the growth season limitation of tree ring chronologies as the annual ENSO state is a representation of the entire water year while tree ring chronologies only observe the available moisture during the spring and summer months. This will be beneficial for streamflow sites in the southeast due to the lack of snow accumulation and especially beneficial for streamflow sites with a strong seasonality in the autumn and winter months.

The goal of this study is to improve the skill of annual streamflow reconstructions over the Southeast by using Multichannel Singular Spectrum Analysis (MSSA) on tree ring chronologies and by considering Sea Surface Temperature (SST) for improving the prediction during wet years. Most of the studies have used either tree ring chronologies or SST for reconstruction of annual flow attribute [Sankarasubramanian et al. 2003]. We propose to improve the Principal Component Regression (PCR) analysis by utilizing both SST and tree ring chronologies for reconstructing annual streamflow estimates.

Chapter 2: Data

This chapter summarizes the tree ring chronologies as well as the annual streamflow, sea surface temperature used in this study.

2.1: Tree Ring Database

The primary predictor of streamflow reconstructions is tree ring chronologies as discussed in Chapter 1. The tree ring chronologies for the study were obtained from the National Atmospheric and Oceanic Administration (NOAA) International Tree Ring Data Bank (ITRDB) [available online at http://www.ncdc.noaa.gov/paleo/treering.html]. Tree Ring chronologies are standardized by removing the growth trend in the tree ring measurements due to the natural aging process of a tree. The selection of these chronologies for streamflow reconstruction is often limited to chronologies located within the boundaries of a reconstructed site's basin [Hidalgo et al 2000; Woodhouse et al 2006]. However, in the Southeast, there is a limited amount of tree ring chronologies, thereby limiting the watersheds for which annual streamflow could be reconstructed. Hence, the selection of chronologies for streamflow reconstruction in the study was extended to chronologies within 200 km of the streamflow site.

2.2: Annual Streamflow Database

Given the purpose of the study is to reconstruct annual streamflow records using SST and tree ring chronologies, we consider only undeveloped basins from the Hydro-Climatic Data Network (HCDN) [Slack et al. 1993]. Streamflow records in the HCDN are nearly void of upstream storage and groundwater pumping. USGS has identified these basins as virgin

watersheds. The locations of the selected streamflow sites can be seen in Figure 1 along with the locations of tree ring chronologies used for reconstruction. The characteristics for each site are summarized in Table 1.

Table 1: List of USGS Streamgage Stations and the corresponding tree ring chronologies used for annual streamflow reconstruction

Site Number	USGS Site Number	Average Annual Streamflow [cfs]	Drainage Area [mi^2]	Tree Chronologies Used [ITRDB Number]
1	02132000	1009	1030	SC004, SC006, NC008
2	02136000	937	1252	SC004, SC006, NC008
3	02198000	603	646	SC004, SC006, GA003
4	02232500	1298	1539	FL007, FL008
5	02236000	3055	3066	FL007, FL008
6	02246000	188	177	FL005, FL007, FL008
7	02321500	410	575	FL005, FL007, FL008
8	02322500	1500	1017	FL005, FL007, FL008

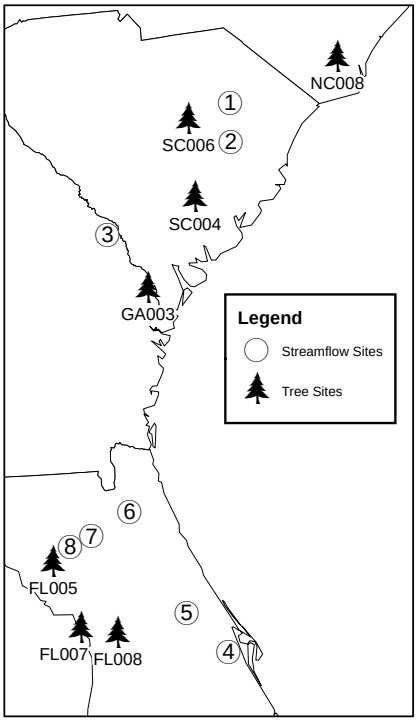

Figure 1: Locations of the streamflow and tree ring chronology sites considered for this study

2.3: SST Database

To improve the prediction of high flow values in streamflow reconstruction, SSTs will be added to the reconstruction predictors. As described in Chapter 1, river basins in the Southeast are greatly influenced by ENSO. Many previous studies have used the Nino-3.4 index to denote the ENSO state [Trenberth and Stepaniak 2001; Devineni and Sankarasubramanian 2010]. The Nino-3.4 index is the anomalous SST condition over the area of 5°S–5°N and 170°–120°W in the tropical Pacific. In the study, average annual Nino-3.4 calculated using Kaplan's Analyses of global sea surface temperatures [Kaplan et al 1998] will be used [available online at http://iridl.ldeo.columbia.edu/SOURCES/.KAPLAN/.Indices/.NINO34/].

Chapter 3: Annual Streamflow Reconstruction

This chapter summarizes the methods, models, validation techniques and performance metrics used in annual streamflow reconstruction. We also discuss results from the validation techniques as well as on the reconstructed streamflow.

3.1 Reconstruction Models

Four different models were used in annual streamflow reconstruction. The traditional approach described in Chapter 1 is included and is denoted as TR-PCA. We also consider the lagged version of tree ring chronologies, which is denoted as TR-PCA Lagged. To improve upon the limitations of the TR-PCA model, two new models were developed, one using MSSA (TR-MSSA) and one combining tree rings and sea surface temperatures (SST-TR) as described in Chapter 1. Furthermore, these models were combined using a nearest neighbor approach, so that the strengths in each model could be maximized to develop an improved reconstructed streamflow time series.

3.1.1 Tree Rings based on PCA Model (TR-PCA and TR-PCA Lagged)

The proximity of tree ring chronologies cause them to be inter-correlated and the most common approach to address the linear codependency in streamflow reconstruction is PCA [Hidalgo et al 2000]. Linear codependency is a common problem among predictors which limits its usage in linear regression. When a multiple linear regression model contains predictors that are highly inter-correlated, regression coefficients estimates can be inaccurate and unstable [Weisberg 1985]. PCA creates a new set of variables called PCs which are independent (orthogonal) of one another by transforming the original data set into linear

combinations of the original variables. Eigen mode analysis from a covariance matrix of the original data set extracts the PCs from the data set. PCA will yield the same amount of PCs as original data variables, each explaining a percentage of the variance from the original data set. In the study, PCA is performed on the tree ring chronologies, and stepwise regression is used to develop a relationship between observed annual streamflow values and the PCs (TR-PCA model). The implementation of stepwise regression greatly reduces the possibility of model "over fitting" by selecting predictors one at a time based on statistical significance. Previous studies have employed stepwise regression for streamflow reconstructions in the Upper Colorado River Basin [Hidalgo et al. 2000; Woodhouse et al. 2006].

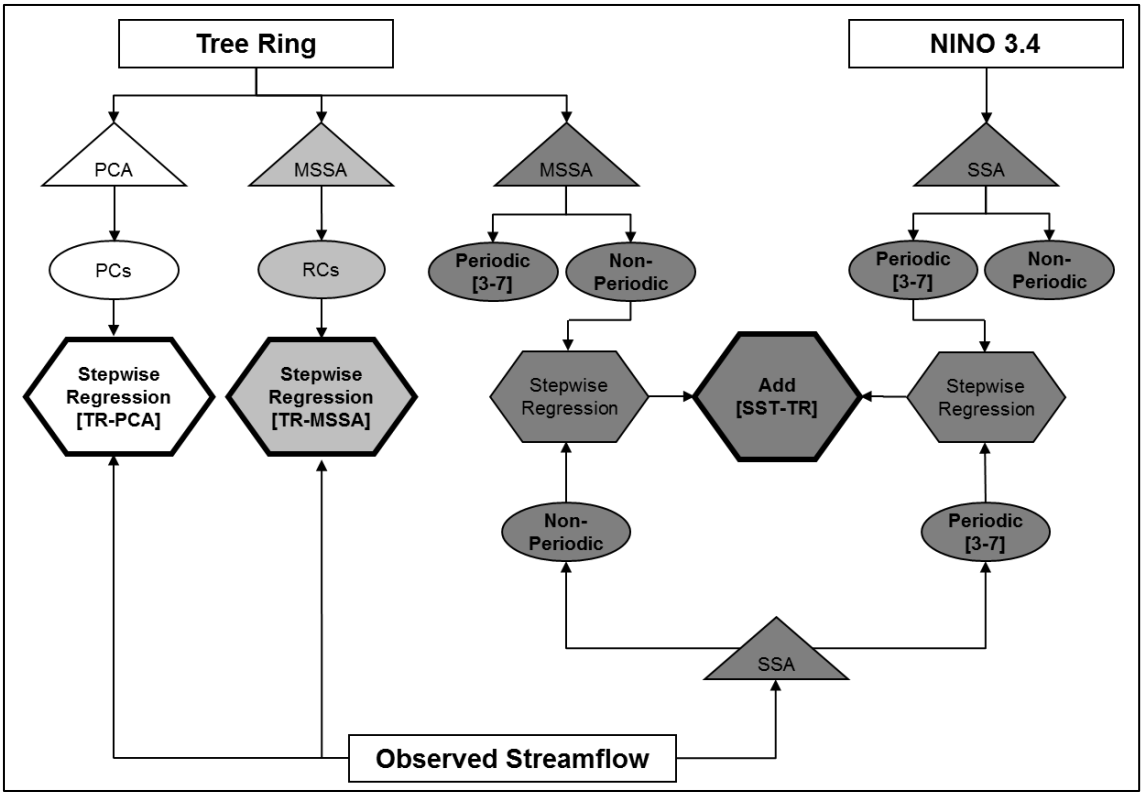

Figure 2: Schematic diagram of the three methodologies employed for annual streamflow reconstruction

As discussed in Chapter 1, it is common to include lagged tree ring chronologies in the regression model to account for autocorrelation in tree ring growth due to residual moisture supply from previous year. This model, TR-PCA Lagged, is executed the same way as the TR-PCA model with the PCs being obtained on a data set consisting of lagged tree ring chronologies. Since we found that the lag relationship in tree ring chronologies is statistically significant up to two years, we obtained PCs using lag-2 tree ring chronologies.

3.1.2 Tree Rings based on MSSA Model (TR-MSSA)

A better approach to take care of the lad dependency in streamflow and tree ring chronologies is by using MSSA. MSSA is a time series approach that quantifies the lag dependence in streamflow by expressing certain features of a multi-dimensional time series through estimates of lagged cross-covariance [Shun and Duffy 1999; Ghil et al 2002]. Like PCA, MSSA produces orthogonal rotated components (RCs), the variance of the RCs (Eigen values) and the eigenvectors corresponding to each RC. Since the MSSA calculation involves lagging the tree ring chronologies, there is no need to have a "Lagged" approach to this method as was done in the TR-PCA model. In the study, MSSA was performed on the tree ring chronologies, and stepwise regression was used to develop a relationship between the observed annual streamflow values and the RCs (TR-MSSA model) as shown in Figure 4. This approach is similar to the TR-PCA Method, but the predictors are now the RCs from MSSA which contain information from previous time periods as opposed to PCs which only contain information from a single year. A brief description of MSSA follows next.

We first define a L dimensional data set of length N as $\boldsymbol{X_l(t)}$, $1 \leq l \leq L$, $1 \leq t \leq N$ where l is the channel number and t is the time index after we remove the mean of each time series. An embedding dimension, M, is then defined to lag the time series. Each time series will be lagged creating a matrix, $\widetilde{\boldsymbol{X}}_{\boldsymbol{l}}$, of N' $(N-M+1)$ time series of length M from the original time series.

$$\widetilde{\boldsymbol{X}_{\boldsymbol{l}}} = \begin{bmatrix} X_l(1) & X_l(2) & \dots & X_l(M) \\ X_l(2) & X_l(3) & \dots & X_l(M+1) \\ \vdots & \vdots & \dots & \vdots \\ X_l(N') & X_l(N'+1) & \dots & X_l(N) \end{bmatrix}, 1 \leq l \leq L. \quad (1)$$

The lag covariance matrix $\widetilde{\boldsymbol{C}}_{\boldsymbol{X}}$ is then given by

$$\widetilde{\boldsymbol{C}}_{\boldsymbol{X}} = \frac{1}{N'}\widetilde{\boldsymbol{X}}^{\boldsymbol{T}}\widetilde{\boldsymbol{X}} = \begin{bmatrix} C_{1,1} & C_{1,2} & \cdots & C_{1,L} \\ C_{2,1} & C_{2,2} & \dots & C_{2,L} \\ \vdots & \vdots & \ddots & \vdots \\ C_{L,1} & C_{L,2} & \cdots & C_{L,L} \end{bmatrix}. \quad (2)$$

Where the components of $\widetilde{\boldsymbol{C}}_{\boldsymbol{X}}$ are defined by

$$\boldsymbol{C_{l,l'}} = \frac{1}{N'}\widetilde{\boldsymbol{X}}_{\boldsymbol{l}}^{\boldsymbol{T}}\widetilde{\boldsymbol{X}}_{\boldsymbol{l'}}. \quad (3)$$

The eigenvectors $\boldsymbol{E_X}$ and eigenvalue $\boldsymbol{\Lambda_x}$ matrices are found by solving

$$\boldsymbol{E_X^T C_X E_X} = \boldsymbol{\Lambda_x}. \quad (4)$$

Each eigenvector $\boldsymbol{E}^k$ will contain L successive segments of length M denoted by $E_l^k(j)$ where l is the segment and j is the time index of the segment. The associated PCs A^k for each eigenvector are extracted by projecting $\widetilde{\boldsymbol{X}}$ onto the eigenvector as follows,

$$A^k(t) = \sum_{j=1}^{M}\sum_{l=1}^{L} \boldsymbol{X_l}(t+j-1)E_l^k(j), 1 \leq t \leq N'. \quad (5)$$

The orthogonal RCs for MSSA are brought back to the original time space by convolving the corresponding PCs with the eigenvectors. Each eigenvector and PC will have to be reconstructed with each channel giving *L* RCs for each PC as follows

$$R_l^k(t) = \frac{1}{M_t}\sum_{j=L_t}^{U_t} A^k(t-j+1)E_l^k(j)\,, 1 \le t \le N \quad (6)$$

The sum of the RCs recovers the original time series. The M_t, L_t, and U_t(normalizing term, lower index limit, and upper index limit) terms are dependent on the number of PCs a given time index *t* was used to calculate. The values of M_t, L_t, and U_t can be seen below,

$$[M_t, L_t, U_t] = \begin{cases} \left[\frac{1}{t}, 1, t\right], & 1 \le t \le M-1 \\ \left[\frac{1}{M}, 1, M\right], & M \le t \le N' \\ \left[\frac{1}{N-t+1}, t-N+M, M\right], & N'+1 \le t \le N. \end{cases} \quad (7)$$

3.1.3 SST and Tree Rings based Reconstruction (SST-TR)

Although the TR-MSSA model contains more information than the TR-PCA model, the limitation of tree growth properties causing an underestimation of high flow values as described in Chapter 1 is still not addressed. The SST-TR model considers SSTs to the predictors to reduce the underestimation of high flow values. Given that wet and dry periods in the basins from the Southeast US are primarily influenced by El Nino Southern Oscillation [Ropelewski and Halpert, 2007; Devineni and Sankarasubramanian, 2010], we proposed to improve the high flow estimation based on the ENSO conditions. As described in Chapter 2, ENSO is the average SST anomaly over 120 W to 170 W and 5S and 5N having a frequency

of three to seven years. For this study, we consider the average Nino3.4 over the period October-September for reconstruction the annual streamflow during the corresponding water year.

Periodic components of a time series can also be identified based on Signal to Noise (S/N) separation of oscillatory pairs of Eigen elements [Ghil et al. 2002]. Geophysical time series, such as ENSO, often have greater power at lower frequencies, so a red noise "null hypothesis" [Allen, 1992] is often used for periodic component detection [Ghil et al. 2002]. The test for periodic detection of RCs against red noise devised by Allen [1992] is known as Monte Carlo Singular Spectrum Analysis (MC-SSA).

Singular Spectrum Analysis (SSA) is a one dimensional version of MSSA with the dimension *L* being one and is often used along with MC-SSA to identify periodic components of a time series [Shun et al. 1999]. MC-SSA generates simulated red noise data from the original time series and computes a covariance matrix $\boldsymbol{C_R}$. The eignevalues $\boldsymbol{\Lambda_R}$ of the red noise data is calculated by projecting $\boldsymbol{C_R}$ onto the eigenvectors $\boldsymbol{E_X}$ from MSSA or SSA as seen below

$$\boldsymbol{\Lambda_R} = \boldsymbol{E_X^T C_R E_X}. \quad (8)$$

The generation of red noise data and subsequent calculation of $\boldsymbol{\Lambda_R}$ are performed a large number of times giving an ensemble of Eigen values $\boldsymbol{\Lambda_E}$. Red noise effect 95% confidence limits are calculated by finding the interval between the 2.5th and 97.5th percentile of a component of the eigenvalue ensemble $\boldsymbol{\Lambda_E^k}$. If a corresponding eigenvalue from MSSA or SSA $\boldsymbol{\Lambda_X^k}$ is outside the 95% confidence range of $\boldsymbol{\Lambda_E^k}$, then the eigenvalue is not due to the red

noise. Hence, the RCs associated with an eigenvalue outside the confidence range can be said to be "periodic."

The SST-TR model utilizes MC-SSA to separate periodic and non-periodic components of streamflow, tree ring chronologies, and SSTs so that separate regressions for periodic and non-periodic components of streamflow can be created as seen in Figure 2. The predictands for each regression are found through SSA on observed annual streamflow. MC-SSA is then performed to separate the periodic and non-periodic components for their associated regressions. The relationship between the non-periodic annual streamflow and tree rings is found using the RCs from a MSSA of tree ring chronologies similar to the method in the TR-MSSA model. However, since only the non-periodic annual streamflow component is being estimated, MC-SSA is performed on the RCs and only the components identified as non-periodic are used in the regression. Similarly, only the periodic component of ENSO will be used in the regression with the periodic annual streamflow component. SSA is performed on Nino 3.4 and the components identified as periodic through MC-SSA are used in stepwise regression to develop a relationship with the periodic annual streamflow estimate. The estimates from the two regressions are added together for the annual streamflow estimate.

3.1.4 Stepwise Regression

Due to the large amount of available predictors, models described in sections 3.1.1-3.1.3 utilize stepwise regression to avoid the inclusion of predictors unrelated to annual streamflow. Stepwise regression computes an F-test on each predictor to determine which will be chosen for the regression [Hocking 1976]. This study uses Forward Selection where the process starts with no predictors in the regression and adds predictors one at a time until

all variables are included or the improvement in the regression due to the addition of another predictor is no longer statically significant. To determine the predictors available for the Leave-Five-Percent-Out-Cross-Validation described in Section 3.2.1, stepwise regression was performed for each model using the entire observed time period. Only predictors chosen were considered for regression during validation. The percent variance of the predictors chosen can be seen in Table 2 where the number in parenthesis is the number of components chosen. The percent variance of the selected tree ring PCs is much higher than that of the tree ring RCs and non-periodic RCs. This is due to the amount of variance explained by each component. For the sites in this study, there are only two or three PCs from PCA making the variance explained by the first PC around 0.7. By comparison, since MSSA includes the covariance with previous time periods, there are more RCs resulting in the variance explained by the first RC around 0.03. This is one advantage of MSSA because each RC will have a strong signal associated with it, and the signal of the RCs chosen will be highly correlated to annual streamflow.

Table 2: The percent variance of selected predictors from stepwise regression

Site	Percent Variance of Predictors			
	Tree Ring PCs	Tree Ring RC's	Tree Ring Non-Periodic RC's	Periodic SST Components
1	0.715 (1)	0.124 (6)	0.088 (4)	0.257 (5)
2	0.715 (1)	0.114 (7)	0.076 (5)	0.466 (6)
3	0.798 (1)	0.048 (2)	0.032 (1)	0.015 (2)
4	0.719 (1)	0.051 (4)	0.050 (8)	0.011 (1)
5	0.719 (1)	0.052 (4)	0.105 (6)	0.040 (4)
6	0.673 (2)	0.051 (8)	0.112 (10)	0.024 (4)
7	1.000 (3)	0.081 (5)	0.041 (4)	0.294 (2)
8	0.673 (2)	0.114 (7)	0.089 (5)	0.294 (2)

3.1.5 Model Combination

One of the motivations for the development of the SST-TR model was improvement in high flow value reconstruction, whereas the MSSA model was designed for overall improvement. Given that the skill of above three models (TA-PCA, TR-MSSA and SST-TR) differs depending on the flow conditions, we employ model combination conditioned on the flow conditions [Devineni et al., 2008, Li 2011]. The framework can be seen in Figure 3, and the description is given below.

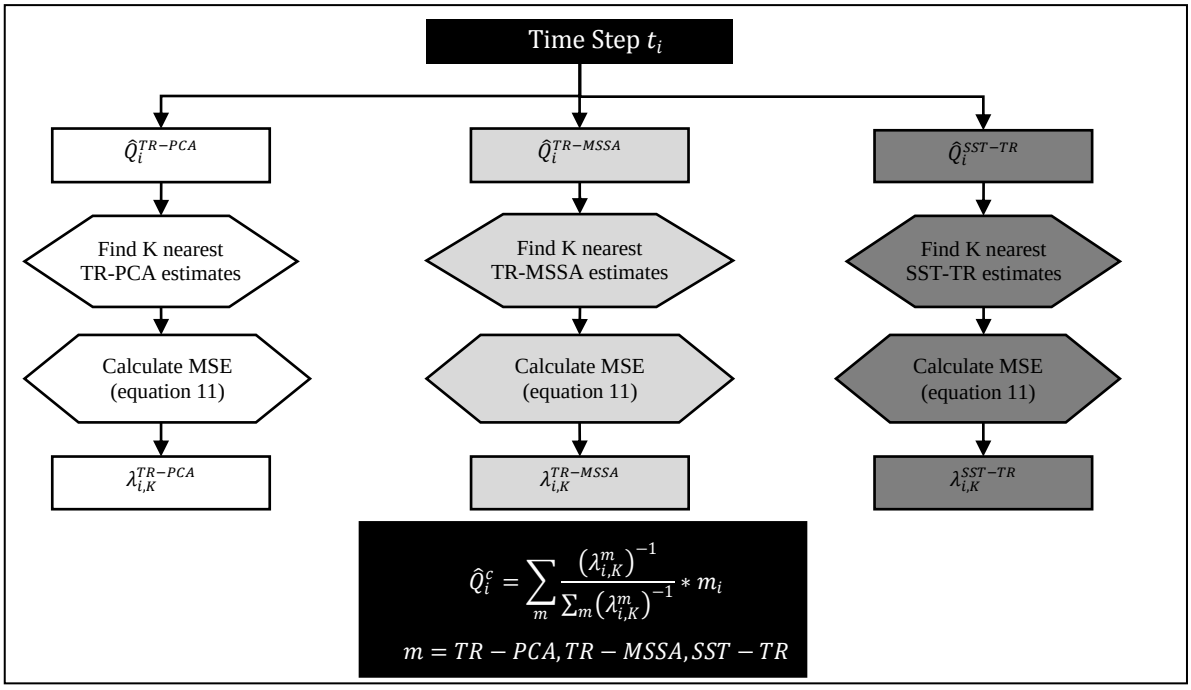

Figure 3: Schematic diagram of the Combination Model employed for annual streamflow reconstruction

For a given time period, t=1,2,3,..,n, there is an observed streamflow Q_t and predicted streamflows, $\hat{Q}_t^m$, where m=1,2,3,..,M denoting a given model. Each model has a corresponding error, ε_t^m, for each time step t. The calculation of ε_t^m can be seen below.

$$\varepsilon_t^m = \left(\hat{Q}_t^m - Q_t\right)^2 \quad (10)$$

For a time outside of the time period, i, given the predicted streamflows, $\hat{Q}_i^m$, we average the error over 'K' neighbors to obtain mean square errors $\lambda_{i,K}^m$ using equation (11).

$$\lambda_{i,K}^m = \frac{1}{K}\sum_{j=1}^{K} \varepsilon_{(j)}^m \quad (11)$$

Weight terms $w_{i,K}^m$ for each model are calculated using the inverses of $\lambda_{i,K}^m$ that gives lower weights for models with higher mean square error over the conditioning neighborhood.

$$w_{i,K}^m = \frac{(\lambda_{i,K}^m)^{-1}}{\sum_{m=1}^{M}(\lambda_{i,K}^m)^{-1}} \quad (12)$$

The weights of each model and the model predictions are then used to obtained the combined model prediction $\hat{Q}_i^c$.

$$\hat{Q}_i^c = \sum_{m=1}^{M} w_{i,K}^m * \hat{Q}_i^m \quad (13)$$

3.2 Validation Techniques and Performance Metrics

Reconstruction models are developed with a limited amount of observed annual streamflow values. The skill in model development increases with the number of observed values used. However, reconstruction models need to be tested against observed values not

used in the regression. For this purpose, we employ leave-five-percent-out cross-validation for evaluating the performance of the three models and the combined model estimates.

3.2.1 Leave-Five-Percent-Out Cross-Validation

To test the skill of the proposed models, a Leave-Five-Percent-Out-Cross-Validation was employed due to its similarity to annual streamflow reconstruction model development and calculations. The framework for this validation can be seen in Figure 4. The predictors available for each model, were determined through the stepwise regressions using the associated predictors from Figure 2 and predictands (annual streamflow for TR-PCA and TR-MSSA, and periodic and non-periodic streamflow for SST-TR). The predictors chosen from these stepwise regressions are the predictors available for the associated model in the validation.

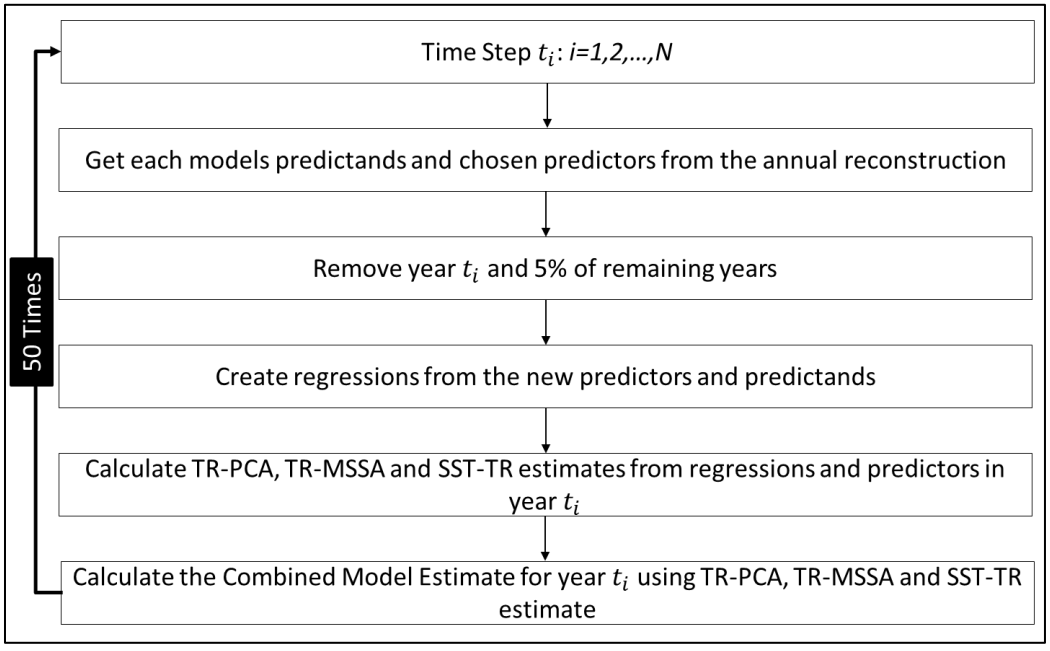

Figure 4: Validation framework employed for the three modeling schemes

For a given year t_i, the predictors and predictands for each model in year t_i and a random five percent of the remaining years in *t*, are removed, and regressions are developed using the remaining predictors and predictands using the framework in Figure 2. The regressions are evaluated using the predictors from year t_i to obtain the annual streamflow estimates for year t_i for the TR-PCA, TR-MSSA, and SST-TR models. These estimates are then used to calculate the annual streamflow estimate for the Combined model in the given year using the procedure outlined in Figure 3. Estimates for all *N* years of observation are calculated to complete the validation. For this study, the validation was performed 50 times.

3.2.2 Performance metrics

A performance measure often used for models is Root Mean Square Error (RMSE) which is a measure of the average deviation between a model and the observed data. A calculation for RMSE can be seen below:

$$RMSE = \sqrt{\frac{\sum_t \left(\hat{Q}(t) - Q(t)\right)^2}{N}}. \quad (14)$$

This study compares the performance of models over several sites, so the Normalized RMSE (NRMSE) will be the performance measure. The calculation can be seen below

$$NRMSE = \frac{RMSE}{\bar{Q}}. \quad (15)$$

NRMSE will allow the performance of models to be compared across many sites without a bias from the average annual streamflow.

Correlation between observed and reconstructed streamflow will also be used to rate the performance the models. Correlation is a measure of the relationship between two variables. It ranges from -1.0 to 1.0 where a value of 0 shows no relationship and values closer to -1.0 and 1.0 show a high relationship. Reconstructed streamflows should be closely related to observed streamflow, so the correlation should be closer to 1.0.

3.3 Results and Discussion

3.3.1 Annual Streamflow Reconstruction

To evaluate the skill of each model, regressions for the TR-PCA, TR-PCA Lagged, TR-MSSA and SST-TR models were developed as described earlier in this chapter using the entire observation period, years when observed SST, streamflow and tree ring chronologies were available. Annual streamflow estimates for these models were then made using the observed predictors, and Combined model estimates were made using the three model estimates as described in Section 3.1.5. The estimates for all models were compared with the observed streamflow at each site. Figure 5 compares the performance of TR-PCA and TR-PCA Lagged models for the observed period. The TR-PCA model has lower NRMSE than the TR-PCA Lagged model in each site except for Sites 1 and 2. However, at these sites the improvement in NRMSE is only slight indicating that the lagging of PCs of tree ring chronologies did not result in significant improvements in reducing the error in reconstruction. Since the TR-PCA Lagged model did not provide additional information in

annual streamflow reconstruction, it will not be considered in the further analysis of this study.

The TR-MSSA, SST-TR and Combined models have a higher correlation than the traditional approach of the TR-PCA model (Figure 6). The increase in correlation illustrates how much more variance the TR-MSSA, SST-TR and Combined models explain. For example, in Site 6 the TR-PCA model has a correlation of 0.526 with the observed streamflow, explaining 27.7% of the observed streamflow variance. In the same site, the Combined model had a correlation of 0.855 which explains 73.1%, 45.2% more than the TR-PCA model, of the observed streamflow variance.

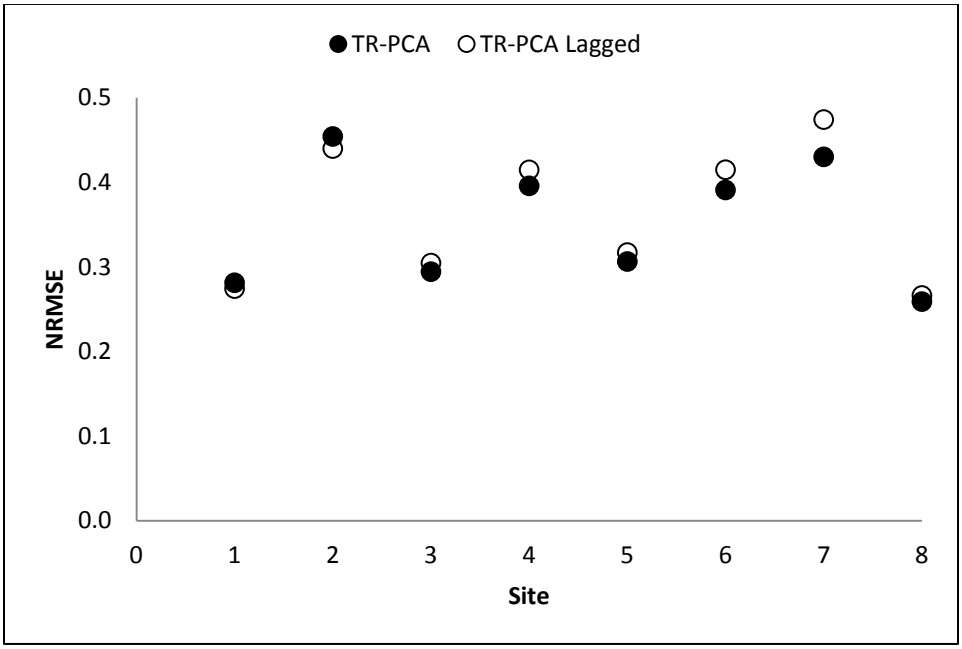

Figure 5: Comparison of the TR-PCA and TR-PCA Lagged models

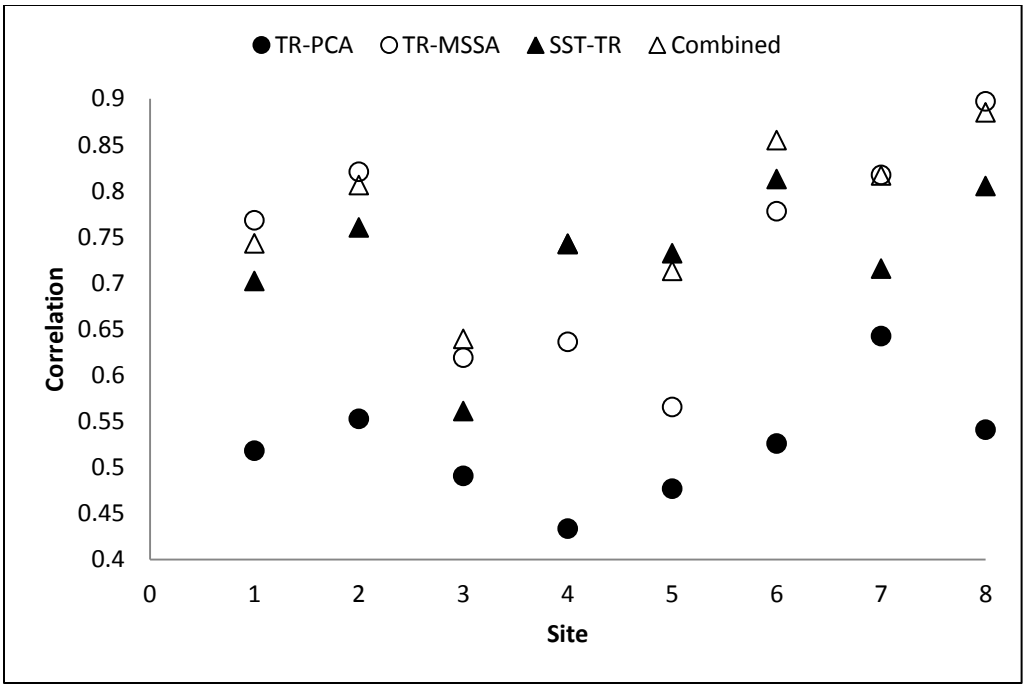

Figure 6: The correlation between observed and reconstructed streamflow for the observed period

The improvement over the traditional approach is shown in terms of NRMSE is seen in Figure 7. In each site, the TR-PCA model has less skill than the other three models. The TR-MSSA model, which developed to capture data from previous climatic events, was has the highest skill in Sites 1, 2, 7 and 8. Figure 8 shows absolute value of the lag 1 correlation of the streamflow, an indication of previous climatic events on annual streamflow, versus the percent improvement of the TR-MSSA model from the TR-PCA model in NRMSE. It can be seen that as the lag 1 correlation of the streamflow increases the % improvement in NRMSE increases, which demonstrates that the RCs from MSSA contain more information on the antecedent basin storage conditions than the PCs and lagged PCs from the TR-PCA model. This is from the ability of MSSA to reduce the dimensions of the lagged covariance matrix. The ability to capture and reconstruct using many positive and negative lagged covariances is

the reason why MSSA had improved skill over TR-PCA lagged model. The improvement of MSSA illustrates the importance of hydroclimatic information from previous years in annual streamflow reconstruction.

The SST-TR model had the highest skill in Sites 4 and 5 (Figure 7). For the source of its skill, the performance of each model is compared with the seasonality of streamflow (Table 3). We discussed in Chapter 1 that tree ring chronologies in the southeast have limited ability to observe precipitation in the autumn and winter months. Table 3 shows the sites having a high seasonality index in the autumn months are the sites where the SST-TR model had the greatest skill in the overall and below normal period. This validates the argument that the use of SSTs in annual streamflow reconstruction will improve upon tree ring chronologies' inability to observe autumn and winter precipitation in the southeast. Sites 1, 2 and 3 have a high seasonality, occurring in the month of February. Since February is near the start of the growing season of trees in the southeast, the infiltration of precipitation in this month will be reachable to the roots when tree growth begins. Thus, the SST-TR model does not improve the overall skill of annual streamflow reconstruction in these sites.

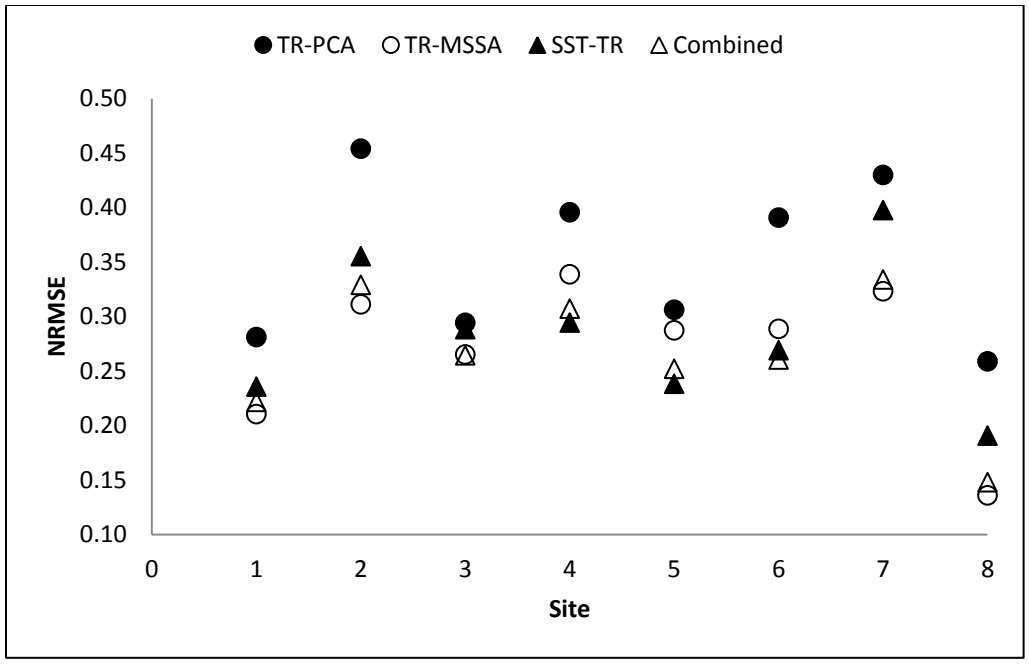

Figure 7: The NRMSE of reconstructed streamflow for the observation period

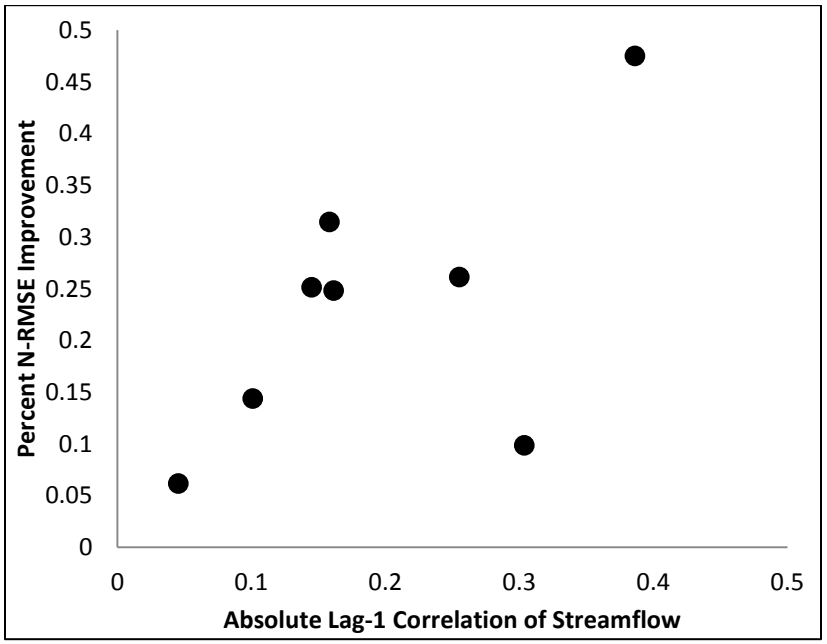

Figure 8: Percent NRMSE Improvement of TR-MSSA model from TR-PCA model based on the absolute value of lag 1 correlation of annual streamflow

Table 3: Comparison of model performance with the seasonality of the streamflow

Site	Seasonality Index	Month	NRMSE		
			TR-PCA	TR-MSSA	SST-TR
1	0.294	FEB	0.281	0.211	0.236
2	0.376	FEB	0.454	0.311	0.356
3	0.303	FEB	0.294	0.265	0.288
4	0.329	OCT	0.396	0.339	0.295
5	0.214	OCT	0.306	0.287	0.238
6	0.105	SEP	0.391	0.289	0.269
7	0.025	OCT	0.430	0.323	0.398
8	0.006	SEP	0.259	0.136	0.191

The SST-TR model also has the highest performance in predicting the above-normal inflows (Figure 9). To quantify this, we identify the observed flows categories by grouping them into below-normal, normal and above-normal categories if the observed flow falls within $<33^{rd}$ percentile, 33-67^{th} percentiles and $>67^{th}$ percentiles of the observed flows respectively. It can be seen that the SST-TR model has the lowest NRMSE in five sites out of eight sites in predicting the above-normal inflows. In only sites 2, 7 and 8, the TR-MSSA model has lower NRMSE than the SST-TR model, and there is no instance of the TR-PCA model having lower NRMSE than the SST-TR. This emphasizes the argument made in Chapter 1 that the inclusion of sea surface temperatures in annual streamflow reconstruction would improve the skill in above normal flows over the traditional approach. The inclusion of SSTs did not help the estimation of below normal flows as it did with above normal flows. The models' performance in predicting below normal inflows (Figure 10) is similar to the overall performance. In each site, the model with the lowest overall NRMSE had the lowest NRMSE during below normal years.

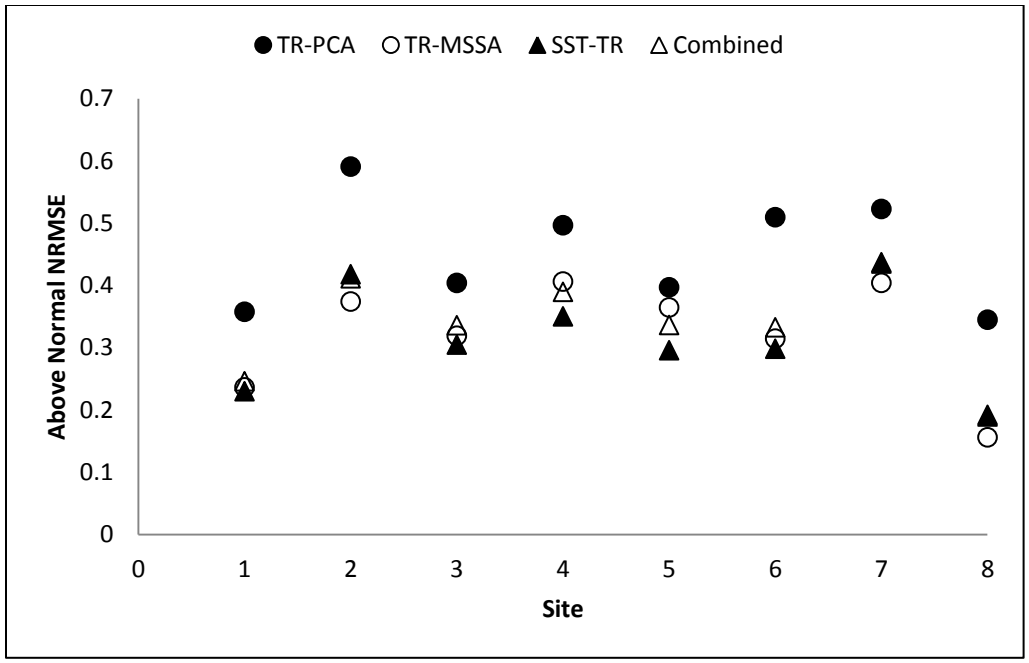

Figure 9: The NRMSE of reconstructed streamflow in above normal flow years during the observed period

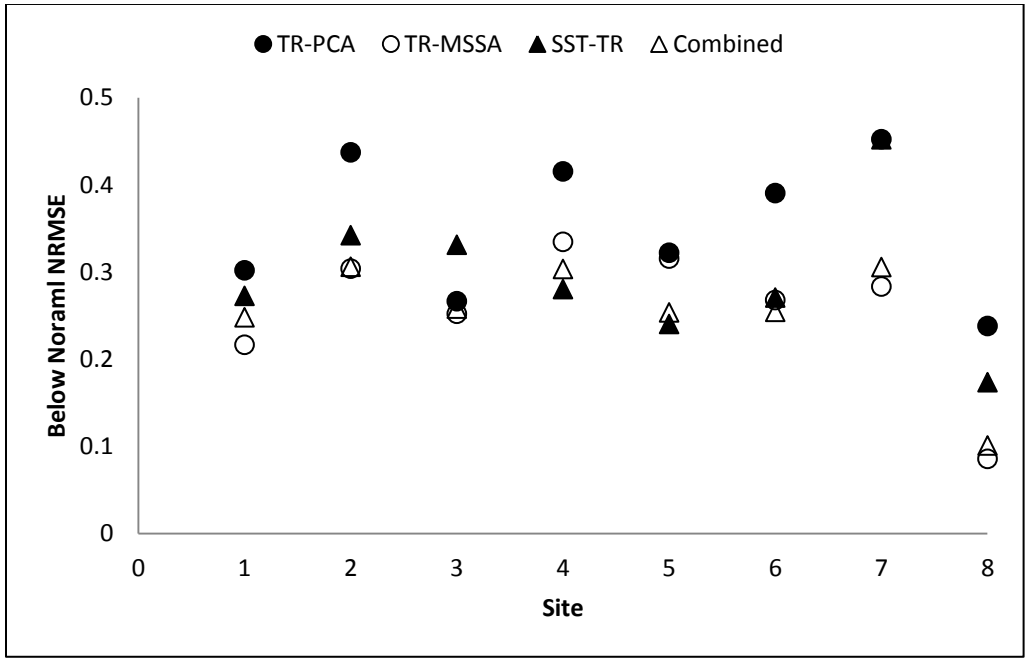

Figure 10: The NRMSE of reconstructed streamflow in below normal flow years during the observed period

3.3.2 Validation Results

For the annual streamflow reconstruction, the entire observed data was used. We expected the reconstructed streamflow to perform well when tested against the observed streamflow, since the same observed annual streamflow was used to develop the regressions for streamflow reconstruction. To test the validity of the findings in annual streamflow reconstruction, Leave-Five-Percent-Out Cross-Validation was employed as described in Section 3.2.1 for the TR-PCA, TR-PCA Lagged, TR-MSSA, SST-TR, and Combined models 50 times. Under each validation, annual streamflow for the Combined model was estimated after the estimates for the other three models had been obtained for the entire time period.

Box plots of correlation and NRMSE from the validation are shown in Figure 11 where the number above each box plot indicates the number of times out of 50 that box plot performed the best (highest correlation or lowest NRMSE) for that site. Like in annual streamflow reconstruction, the TR-MSSA model has higher skill than the TR-PCA model with the exception of Site 5, validating the conclusion that the replacement of PCA with MSSA improves the skill of reconstruction. The SST-TR model also has higher skill than the TR-PCA model in all but one site. This model also had a higher correlation and lower NRMSE than both TR-PCA and TR-MSSA in Sites 4-6, the sites exhibiting high autumn seasonality. It was argued in Chapter 1 that the inclusion of SSTs would improve reconstruction skill in these sites, and these results validate our argument.

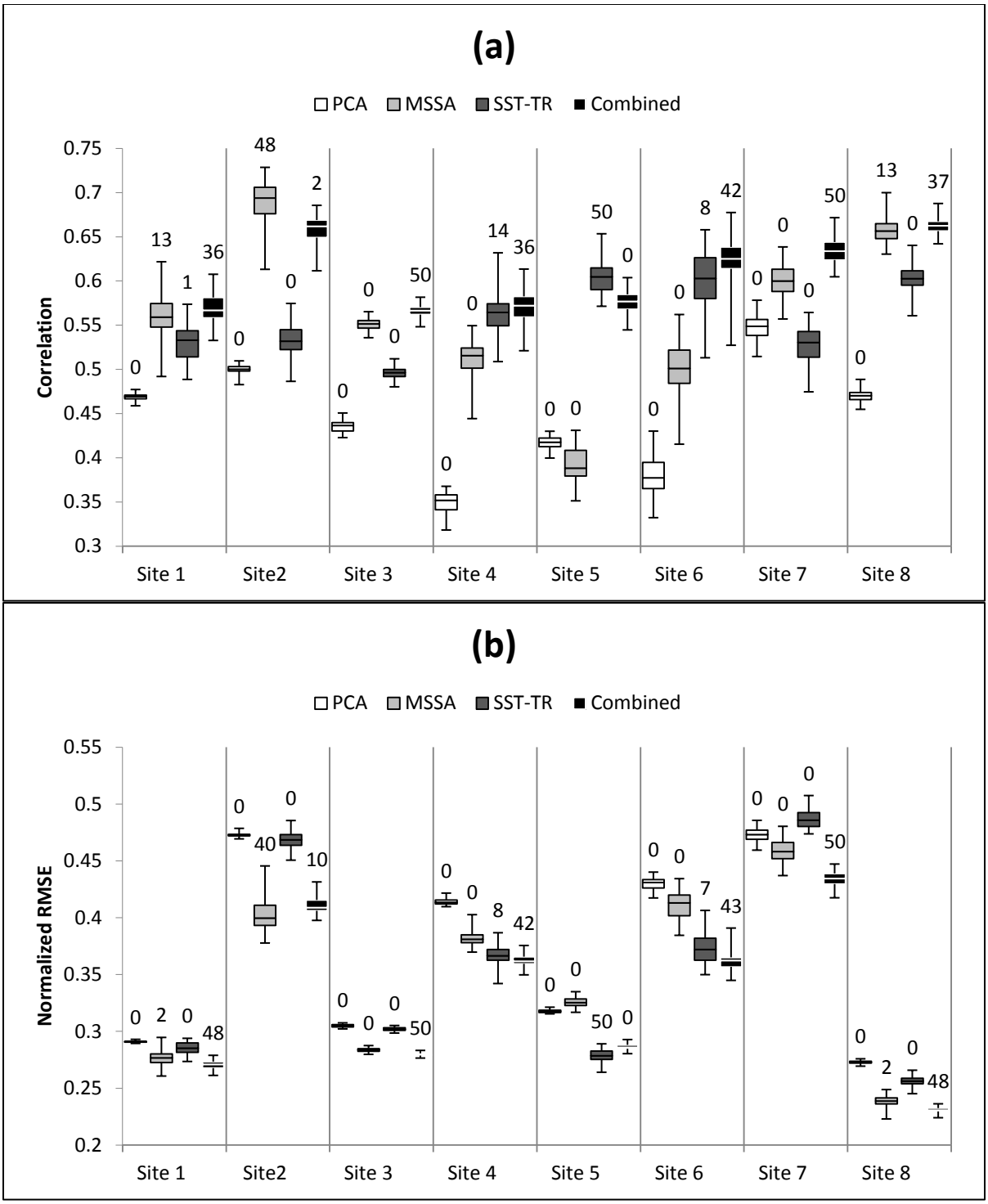

Figure 11: Correlation (a) and NRMSE (b) under Leave-Five-Percent-Out Cross-Validation

As was the case in annual streamflow reconstruction, the SST-TR model has the highest skill in above normal inflow estimation (Figure 12a). In 6 of the 8 sites, the SST-TR model had the lowest NRMSE validating the argument made in Chapter 1 that the inclusion of SSTs as predictors would increase the skill in predicting higher flow values.

The three different models perform differently at each site and for the quantity of streamflow reconstructed, so a combination model was developed to get the overall greatest skill for annual streamflow reconstruction. It can be seen in Figure 11 that the combined model had the highest correlation and lowest NRMSE in six of the eight sites for the validation period. This is different than the results for annual streamflow reconstruction where the Combined model has the highest skill in only 2 sites. The ability to take advantage of the other models different areas of high skill allows the Combined model to have high skill during cross-validation when each model's skill is reduced. The exceptions are Sites 2 and 5 where one model has significantly higher skill than the other two models, reducing the ability for improvement by combining models.

It is also important to note that the combination model is not the best performing model in above and below normal flow values as seen in Figure 12. The Combined model's performance in below normal flows does not follow the trend seen in annual streamflow reconstruction where the model with highest overall skill had the highest below normal skill. However, in most sites, the combination model is the second best in estimating flows during above normal and below normal years.

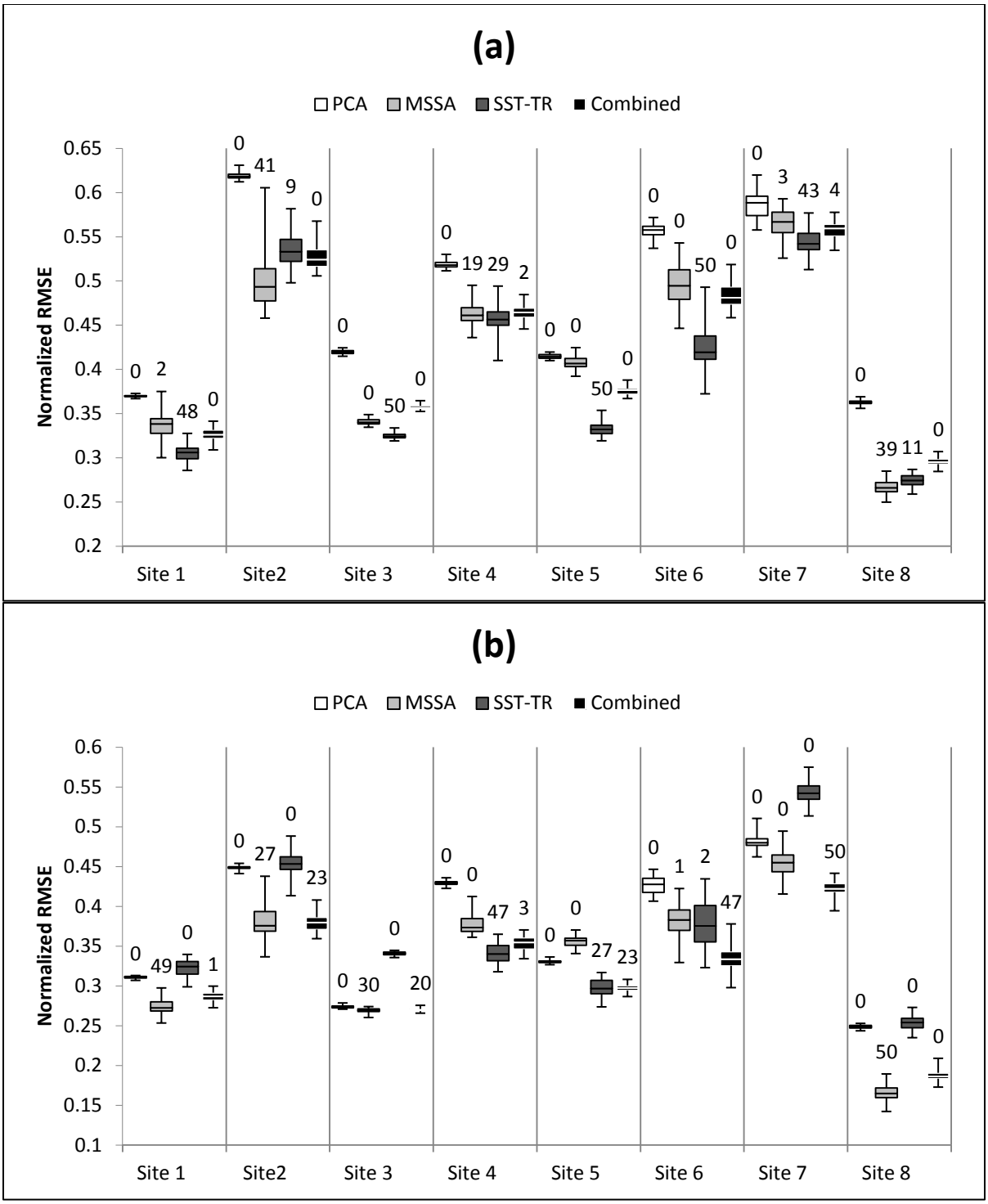

Figure 12: Normalized RMSE under Leave-Five-Percent-Out-Cross-Validation for above normal flow values (a) and below normal flow values (b)

3.3.3 Comparison of Reconstructed Streamflow with observed Precipitation

The relationship between annual precipitation and streamflow is well established, and the correlation between these two attributes is dependent on watershed characteristics such as soil cover, aquifer volume and water table depth. Since the observed precipitation is available even before 1930, we compare the correlation between the reconstructed annual streamflow and the observed precipitation available at the site for the period prior to the observed streamflow is available. The comparison of the observed and reconstructed streamflow correlation with observed precipitation can be seen in Figure 10. The reconstructed streamflows were found by developing step-wise regression between the observed streamflow and the described methodology in Section 3 for each of the model. Only the reconstructed streamflow from these regressions for the observed period were used in the calculations of Figure 10. It can be seen in Figure 10 that the TR-MSSA, SST-TR, and Combined models have correlations similar to the correlation with the observed annual streamflow. The TR-PCA model's correlation is not similar to the observed when the observed correlation is higher, further illustrating the improvement of the TR-MSSA and SST-TR models over the traditional approach.

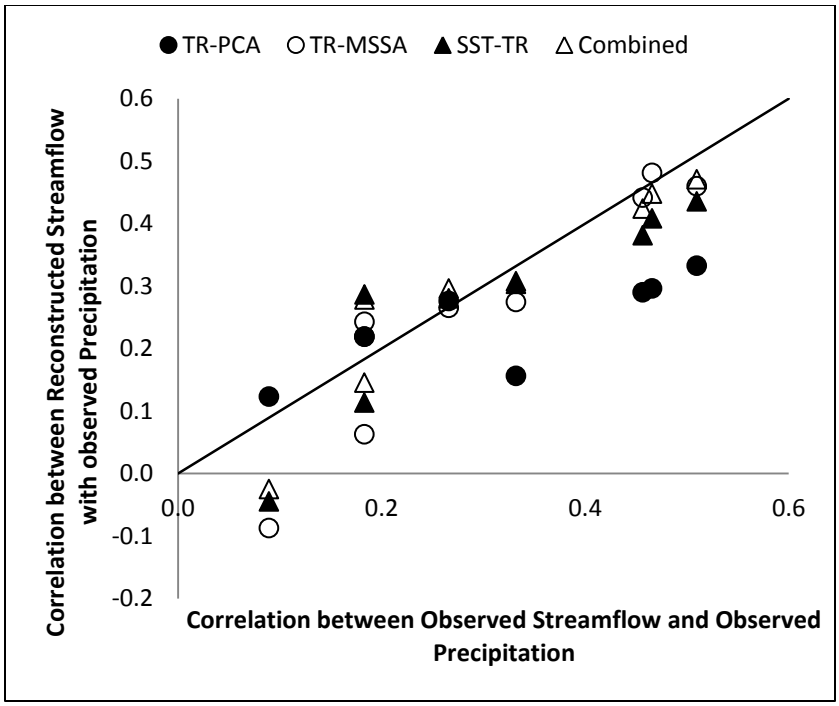

Figure 13: The observed precipitation correlation with observed and reconstructed streamflow for the observed streamflow period

However, the reconstructed streamflow's correlation with precipitation must remain consistent to the observed streamflow's correlation beyond the observed period. The correlation between the observed precipitation and observed/reconstructed streamflow over the observed period is presented in Table 4, whereas the correlation between the observed precipitation and reconstructed streamflow from 1900 until the observed period is shown in Table 5. Both tables also provide the number of data points, N, employed in calculating the correlation for each site. It is shown in Table 4 that the Combined model has correlations consistently similar to the observed correlations supporting the argument that the combined model has the highest skill for annual streamflow reconstruction. Since the computed correlations in Table 4 and Table 5 are independent, we test whether the correlations between

the observed period and the reconstructed period are equal. The significance of the difference of two independent time series' correlations is calculated with the Fisher Transformation [Fisher 1915, 1921]. The calculation can be seen below in Equation 16 where *r* denotes the correlation and *N* denotes the size of a given sample (1 or 2).

$$z = \frac{0.5ln\left(\frac{1+r_1}{1-r_1}\right) - 0.5ln\left(\frac{1+r_2}{1-r_2}\right)}{\sqrt{\frac{1}{N_1-3} + \frac{1}{N_2-3}}} \quad (16)$$

Table 4: Correlation between the observed precipitation and observed/reconstructed streamflow for the observed period

Site	Observed Q	TR-PCA	TR-MSSA	SST-TR	Combination	N
1	0.466	0.297	0.481	0.409	0.448	56
2	0.456	0.290	0.442	0.382	0.424	56
3	0.510	0.333	0.460	0.436	0.471	48
4	0.089	0.123	-0.087	-0.044	-0.025	60
5	0.183	0.220	0.063	0.114	0.146	70
6	0.332	0.157	0.275	0.304	0.308	62
7	0.266	0.276	0.265	0.281	0.296	62
8	0.183	0.219	0.243	0.287	0.279	61

The Fisher Transformation is used on the null hypothesis that the two correlations are equal. Therefore if the absolute value of *z* is greater than 1.96, the null hypothesis is rejected at a 95% confidence level, meaning that the two correlations of the independent time series are significantly different.

The z values for each model at each site is presented in Table 6, and the sites for which the correlations are significantly different are shown in bold. It can be seen that TR-PCA model in Sites 1 and 2 has correlations that differ significantly. On the other hand, the

computed correlations are not statistically different between the two periods for any of the sites under TR-MSSA model, SST-TR model, and combined model, showing the constancy of these models.

Table 5: Correlation between the observed precipitation and the reconstructed streamflow from 1900 until the first year of instrumental record

Site	TR-PCA	TR-MSSA	SST-TR	Combination	N
1	0.651	0.427	0.448	0.569	30
2	0.658	0.424	0.572	0.586	30
3	0.560	0.377	0.076	0.430	38
4	0.218	0.152	0.348	0.356	34
5	0.210	-0.255	-0.127	-0.089	34
6	0.332	0.275	0.456	0.382	32
7	0.540	0.421	0.371	0.445	32
8	0.266	0.298	0.329	0.335	33

The comparison of reconstructed streamflow with observed precipitation validates many of the findings from Section 3.3.1. The TR-PCA model did not perform as well as the rest of the three models. Furthermore, the Combined model consistently performed better than the rest of the models.

Table 6: The z values from the Fisher Transformation of the difference of correlations before and during the observed period

Site	TR-PCA	TR-MSSA	SST-TR	Combined
1	**-1.994**	0.291	-0.202	-0.692
2	**-2.073**	0.092	-1.051	-0.924
3	-1.275	0.447	1.737	0.228
4	-0.435	-1.078	-1.827	-1.780
5	0.047	1.490	1.115	1.087
6	-0.826	-0.002	-0.789	-0.370
7	-1.410	-0.778	-0.445	-0.763
8	-0.221	-0.261	-0.205	-0.278

Chapter 4: Conclusions and Future Work

One of the major goals of this study was to improve the skill of annual streamflow reconstruction using the traditional approach of using PCA. A limitation of this approach is the inability of the PCs to capture climactic information from previous time periods on which streamflow is dependent. The use of MSSA instead of PCA reduced this limitation as demonstrated by results in Section 3. Other limitation in using PCA for annual streamflow reconstruction primarily arises from the limited skill in reconstructing high flow values. The addition of SSTs to tree ring chronologies in annual streamflow reconstruction reduced these limitations as indicated by the skill of SST-TR in predicting high flows. It was also demonstrated that this approach reduced the error in sites with high autumn seasonality due to SST being observable during the autumn and winter months. Furthermore, the use of MSSA and addition of SSTs in reconstruction would produce streamflow with a correlation precipitation that is similar to that of observed annual streamflow.

The study also showed that each approach has different skill for different flow values and at different sites, but combining the traditional approach with the use of MSSA and addition of tree rings provides estimates with the highest skill. The increase in skill is due to the combinations ability to use different models in their areas of highest skill. Combination showed the ability to maintain high skill when the skill of each model was reduced during cross-validation.

The next step in this research field is to test this methodology on streamflow sites over the United States as well as on the global streamflow database. Another future goal is to extend the methodology to develop annual streamflow forecasts using both tree rings and

SST forecasts. Once annual streamflow forecasts can be made, they can be paired with reservoir management models to optimize for hydropower supply, water supply efficiency, and flood risk minimization.

REFERENCES

Allen, M., Interactions between the atmosphere and oceans on time scales of weeks to years, Ph.D. thesis, 202 pp., Univ. of Oxford, Oxford, England, 1992.

Brockway, C. G., and A. A. Bradley (1995), Errors in Streamflow Drought Statistics Reconstructed From Tree Ring Data, *Water Resour. Res.*, 31(9), 2279–2293, doi:10.1029/95WR01141.

Cook, E. R., and G. C. Jacoby, 1983: Potomac River streamflow since 1730 as reconstructed by tree rings. *J. Climate Appl. Meteor.*, 22, 1659–1672.

Devineni, N., A.Sankarasubramanian, and S. Ghosh, Multi-model Ensembling of Probabilistic Streamflow Forecasts: Role of Predictor State Space in skill evaluation, *Water Resources Research*,44, W09404, doi:10.1029/2006WR005855,2008.

Devineni, N. and A. Sankarasubramanian, Improving the Prediction of Winter Precipitation and Temperature over the Continental United States: Role of ENSO State in Developing Multimodel Combinations,*Monthly Weather Review*,138(6), 2447-2468, 2010.

Fisher, R. A. (1915), "Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population", Biometrika (Biometrika Trust), 10 (4), pp. 507-521.

Fisher, R. A. (1921), "On the 'probable error' of a coefficient of correlation deduced from a small sample", Metron, 1, pp. 3-32.

Fritts, H. C.: 1976, *Tree Rings and Climate*, Academic Press, London.

Fritts, H. C., J. Guiot, G. A. Gordon, and F.Schweingruber, Methods of calibration, verification, and reconstruction, in *Methods of Dendrochronology*, etided by E.R. Cook and L. A. Kairiukstis, pp 163-217, Kluwer Academic, Norwell, Mass., 1990.

Fritts, H. C., Reconstructing Large-scale Climatic Patterns From Tree-Ring Data: A Diagnostic Analysis, Univ. of Ariz. Press, Tucson, 1991.

Gangopadhyay, S., B. L. Harding, B. Rajagopalan, J. J. Lukas, and T. J. Fulp (2009), A nonparametric approach for paleohydrologic reconstruction of annual streamflow ensembles, Water Resour. Res., 45, W06417, doi:10.1029/2008WR007201.

Ghil, M., et al. (2002), Advanced spectral methods for climatic time series, *Rev. Geophys.*, 40(1), 1003, doi:10.1029/2000RG000092.

Hidalgo, H. G., T. C. Piechota, and J. A. Dracup (2000), Alternative principal components regression procedures for dendrohydrologic reconstructions, *Water Resour. Res.*, 36(11), 3241–3249, doi:10.1029/2000WR900097.

Hosking, J. R. M., and J. R. Wallis (1986), Paleoflood Hydrology and Flood Frequency Analysis, Water Resour. Res., 22(4), 543–550, doi:10.1029/WR022i004p00543.

Hocking, R. R. (1976) "The Analysis and Selection of Variables in Linear Regression," *Biometrics.*

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

Kaplan, A., M. Cane, Y. Kushnir, A. Clement, M. Blumenthal, and B. Rajagopalan, 1998: Analyses of global sea surface temperature 1856–1991. J. Geophys. Res., 103, 18 567–18 589.

Li, Weihua, "Uncertainty Reduction in Hydrologic Modeling and Regional Water Management Utilizing Inter-basin Transfer". Ph.D. dissertation, North Carolina State University, 2011.

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

Meko, D., Stockton, C. W. and Boggess, W. R. (1995), THE TREE-RING RECORD OF SEVERE SUSTAINED DROUGHT. JAWRA Journal of the American Water Resources Association, 31: 789–801. doi: 10.1111/j.1752-1688.1995.tb03401.

Meko D 1997 Dendroclimatic reconstruction with time varying predictor subsets of tree indices J. Clim. 10 687–96.

Sankarasubramanian, A. and U. Lall, Flood Quantiles and Changing Climate: Seasonal forecasts and causal relations, *Water Resources Research*, 39(5), art.no.1134,2003.

Shun, T., and C. J. Duffy (1999), Low-frequency oscillations in precipitation, temperature, and runoff on a west facing mountain front: A hydrogeologic interpretation, *Water Resour. Res.*, 35(1), 191–201, doi:10.1029/98WR02818.

Slack, J.R.; Lumb, A.; Landwehr, J.M. *Hydro-Climatic Data Network (HCDN) Streamflow Data Set, 1874-1988;* U.S. Geological Survey Report, 1993.

Stedinger, J. R., and T. A. Cohn (1986), Flood Frequency Analysis With Historical and Paleoflood Information, *Water Resour. Res.*, 22(5), 785–793, doi:10.1029/WR022i005p00785.

Stockton, C., and G. Jacoby (1976), Long-term surface-water supply and streamflow levels in the Upper Colorado River basin, *Lake Powell Res. Proj. Bull. 18*, Inst. of Geophys. and Planet. Phys., Univ. of Calif., Los Angeles, Calif.

Trenberth, K. E., and D. P. Stepaniak, 2001: Indices of El Niño evolution. *J. Climate*, 14, 1697–1701.

Weisberg,S .,Applied Linear Regression John Wiley, New York, 1985.

Woodhouse, C. A. (2001), A Tree-Ring Reconstruction Of Streamflow For The Colorado Front Range. JAWRA Journal of the American Water Resources Association, 37: 561–569. doi: 10.1111/j.1752-1688.2001.tb05493.x

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