

8-1-2018

# ANALYZING THE RELATIONSHIP BETWEEN LARGE SCALE CLIMATE VARIABILITY AND STREAMFLOW OF THE CONTINENTAL UNITED STATES

Swastik Bhandari

*Southern Illinois University Carbondale*, [swastikbhandari7@gmail.com](mailto:swastikbhandari7@gmail.com)

Follow this and additional works at: <https://opensiuc.lib.siu.edu/theses>

---

## Recommended Citation

Bhandari, Swastik, "ANALYZING THE RELATIONSHIP BETWEEN LARGE SCALE CLIMATE VARIABILITY AND STREAMFLOW OF THE CONTINENTAL UNITED STATES" (2018). *Theses*. 2362.  
<https://opensiuc.lib.siu.edu/theses/2362>

This Open Access Thesis is brought to you for free and open access by the Theses and Dissertations at OpenSIUC. It has been accepted for inclusion in Theses by an authorized administrator of OpenSIUC. For more information, please contact [opensiuc@lib.siu.edu](mailto:opensiuc@lib.siu.edu).

ANALYZING THE RELATIONSHIP BETWEEN LARGE SCALE CLIMATE VARIABILITY  
AND STREAMFLOW OF THE CONTINENTAL UNITED STATES

by

Swastik Bhandari

B.S., Tribhuvan University, 2013

A Thesis

Submitted in Partial Fulfillment of the Requirements for the  
Master of Science Degree

Department of Civil and Environmental Engineering

in the Graduate School

Southern Illinois University Carbondale

August 2018

THESIS APPROVAL

ANALYZING THE RELATIONSHIP BETWEEN LARGE SCALE CLIMATE VARIABILITY  
AND STREAMFLOW OF THE CONTINENTAL UNITED STATES

by

Swastik Bhandari

A Thesis Submitted in Partial  
Fulfillment of the Requirements  
for the Degree of  
Master of Science  
in the field of Civil Engineering

Approved by:

Dr. Ajay Kalra, Chair

Dr. Justin T. Schoof

Dr. Rolando Bravo

Graduate School

Southern Illinois University Carbondale

April 25, 2018

## **AN ABSTRACT OF THE THESIS OF**

SWASTIK BHANDARI, for the Master of Science degree in CIVIL ENGINEERING, presented on April 25, 2018 at Southern Illinois University Carbondale.

**TITLE: ANALYZING THE RELATIONSHIP BETWEEN LARGE SCALE CLIMATE VARIABILITY AND STREAMFLOW OF THE CONTINENTAL UNITED STATES**

**MAJOR PROFESSOR: Dr. Ajay Kalra**

Over the years there is an increasing evidence of climate change on the available water resources. The interaction of hydrological cycle with climate variability and change may provide information related with several water management issues. The current study analyzes streamflow variability of the United States due to large-scale ocean-atmospheric climate variability. In addition, forecast lead-time is also improved by coupling climate information in a data driven modeling framework. The spatial-temporal correlation between streamflow and oceanic-atmospheric variability represented by sea surface temperature (SST), 500-mbar geopotential height ( $Z_{500}$ ), 500-mbar specific humidity ( $SH_{500}$ ), and 500-mbar east-west wind ( $U_{500}$ ) of the Pacific and the Atlantic Ocean is obtained through singular value decomposition (SVD). For forecasting of streamflow, SVD significant regions are weighted using a non-parametric method and utilized as input in a support vector machine (SVM) framework. The Upper Rio Grande River Basin (URGRB) is selected to test the applicability of the proposed forecasting model for the period of 1965-2014. The April-August streamflow volume is forecasted using previous year climate variability, creating a lagged relationship of 1-13 months. To understand the effect of predefined indices such as El Nino Southern Oscillation (ENSO), Pacific Decadal Oscillation (PDO), and Atlantic Multidecadal Oscillation (AMO) on the regional streamflow, a wavelet analysis is also performed for regions developed by from 2014 National

Climate Assessment (NCA). Moreover, different SVD approach is performed for streamflow of each of the six NCA regions named as Great Plains, Midwest, Northeast, Northwest, Southeast, and Southwest. In regional case, SVD is applied initially with streamflow and SST; and that spatial-temporal correlation is later correlated with  $Z_{500}$ ,  $SH_{500}$ , and  $U_{500}$  separately to evaluate the interconnections between climate variables.

SVD result showed that the streamflow variability of the URGRB was better explained by SST and  $U_{500}$  as compared to  $Z_{500}$  and  $SH_{500}$ . The SVM model showed satisfactory forecasting ability as the observed and forecasted streamflow volume for different selected sites were well correlated. The best results were achieved using a 1-month lead to forecast the following 4-month period. Overall, the SVM results showed excellent predictive ability with average linear correlation coefficient of 0.89 and Nash-Sutcliffe efficiency of 0.79. Whereas regional SVD analysis showed that streamflow variability in the Great Plains, Midwest, and Southwest region is strongly associated with SST of ENSO-like region. However, for Northeast and Southeast region,  $U_{500}$  and  $SH_{500}$  were strongly correlated with streamflow as compared to the SST of the Pacific Ocean. The continuous wavelet analysis of ENSO/PDO/AMO and the regional streamflow patterns revealed different significant timescale bands that affected their variation over the study period. Identification of several teleconnected regions of the climate variables and the association with the streamflow can be helpful to improve long-term prediction of streamflow resulting in better management of water resources in the regional scale.

## ACKNOWLEDGEMENTS

I would like to extend my deepest gratitude to Dr. Ajay Kalra, my thesis advisor, for his guidance, supervision and mentorship throughout my graduate study. The assistance that he provided was pivotal to the success of this research.

I would also like to acknowledge Dr. Justin T. Schoof and Dr. Rolando Bravo for the help and advice they offered not only during my research but also whilst serving on my thesis committee. I owe special thanks to my lab team for their continuous help in resolving the problems whenever I encountered them. I would like to thank the office of Vice Chancellor for Research at SIU Carbondale for providing support for the current study.

I would like to express my sincere thanks to my parents, family, friends and colleagues for their patience, advice and helps whenever I needed them the most. Thanks goes to Balbhadra Thakur, Pratik Pathak, Ranjit Bhandari, Narayan Nyaupane and Ranjan Parajuli for their continuous support. Finally, I would like to express my deepest appreciation to the entire faculty, staff and colleagues of the Southern Illinois University's Civil and Environmental Engineering Department for their help and guidance throughout my graduate study.

I dedicate this thesis to my parents Mr. Laxmi Prasad Bhandari and Mrs. Pushpa Bhandari who's never ending love and affection helped me believe in myself. I could never thank them enough and can only hope to pay forward their never ending love and support.

## TABLE OF CONTENTS

<u>CHAPTER</u>	<u>PAGE</u>
ABSTRACT.....	i
ACKNOWLEDGEMENTS.....	iii
LIST OF TABLES.....	vii
LIST OF FIGURES.....	viii
CHAPTER 1 - INTRODUCTION.....	1
1.1 Research Background.....	1
1.2 Research Motivations.....	2
1.3 Research Objectives.....	4
1.4 Research Outlines.....	4
CHAPTER 2 - STREAMFLOW FORECASTING USING SINGULAR VALUE DECOMPOSITION AND SUPPORT VECTOR MACHINE FOR THE UPPER RIO GRANDE RIVER BASIN .....	5
2.1 Introduction.....	5
2.2 Study Area and Data .....	10
2.2.1 Study area.....	10
2.2.2 Data.....	10
2.3 Methodology .....	13
2.3.1 Establishing correlation between two variables using SVD.....	15
2.3.2 Screening of predictors .....	16
2.3.3 Predicting streamflow using SVM.....	17
2.3.4 Model Evaluation.....	18

2.4	Results and Discussion.....	19
2.4.1	SVD analysis.....	19
2.4.2	Predictor screening analysis.....	27
2.4.3	SVM Analysis.....	29
2.5	Conclusion.....	37
CHAPTER 3 - INTERRELATIONSHIP BETWEEN REGIONAL STREAMFLOW AND		
OCEAN-ATMOSPHERIC CLIMATE VARIABILITY OF THE UNITED STATES: A		
SINGULAR VALUE DECOMPOSITION AND WAVELET APPROACH.....		
		41
3.1	Introduction .....	41
3.2	Study Area and Data .....	45
3.2.1	Study area.....	46
3.2.2	Data.....	46
3.3	Methodology .....	48
3.3.1	Singular value decomposition.....	48
3.3.2	Continuous wavelet transforms and its derivatives .....	49
3.4	Results and Discussion.....	51
3.4.1	SVD analysis.....	51
3.4.2	CWT analysis.....	70
3.5	Conclusions .....	79
CHAPTER 4 - CONTRIBUTIONS AND RECOMMENDATIONS.....		
		81
4.1	Summary .....	81
4.2	Contributions.....	82
4.3	Limitations .....	82

4.4 Recommendations for future work.....	83
REFERENCES.....	84
VITA.....	97

## LIST OF TABLES

<u>TABLE</u>	<u>PAGE</u>
Table 1: SVD results for different lead-time cases.....	20
Table 2: Best streamflow predictor variables for different lead-time scenarios.....	28
Table 3: SVM model performance for different stations for different lead-times.....	31
Table 4: NCA regions and the constituent states .....	45
Table 5: The number of stations, data range, and percent variance.....	51
Table 6: SCF and NSC values for different lead-time cases .....	53

## LIST OF FIGURES

<u>FIGURE</u>	<u>PAGE</u>
Figure 1: Map showing six streamflow stations in the Rio Grande River Basin.....	13
Figure 2: The SVD-SVM model flowchart. ....	15
Figure 3: Heterogeneous correlation map for the Pacific Ocean.....	20
Figure 4: Heterogeneous correlation map for the Atlantic Ocean. ....	21
Figure 5: Map showing examples of continuous exceedance probability forecast .....	29
Figure 6: Model evaluation plots for 1-month lead-time.....	32
Figure 7: Model evaluation plots for 4-month lead-time.....	34
Figure 8: Model evaluation plots for 13-month lead-time.....	36
Figure 9: Map showing NCA regions and streamflow stations.....	46
Figure 10: Heterogeneous correlation map developed for Great Plains.....	54
Figure 11: Heterogeneous correlation map developed for Midwest.....	58
Figure 12: Heterogeneous correlation map developed for Northeast. ....	61
Figure 13: Heterogeneous correlation map developed for Northwest.....	63
Figure 14: Heterogeneous correlation map developed for Southeast. ....	66
Figure 15: Heterogeneous correlation map developed for Southwest.....	70
Figure 16: Continuous wavelet transforms along with ENSO and PDO.....	71
Figure 17: Continuous wavelet transforms along with NCA regions.....	72
Figure 18: Wavelet coherency spectrums between ENSO and the regional streamflow. ....	73
Figure 19: Wavelet coherency spectrums between PDO and the regional streamflow.....	75
Figure 20: Wavelet coherency spectrums between AMO and the regional streamflow.....	78

# CHAPTER 1

## INTRODUCTION

### 1.1 Research Background

The Intergovernmental Panel on Climate Change (IPCC, 2014) has stated that the human and natural systems have been affected by recent climate change. Alteration of regional hydrological cycle and subsequent effect on quality and quantity of water resources are one of the major consequences of global climate change (Gleick, 1989). Water has become a major natural commodity in the United States, where limited water availability has been exacerbated by past frequent droughts. Extreme hydrologic events such as floods and droughts are associated with hydro-climatic variability; improved knowledge of that variability in response to climatic fluctuations is crucial to mitigating social and economic impacts (Redmond and Koch, 1991). Several studies (e.g., Christensen et al., 2004; Stewart et al., 2004; Nijssen et al., 2001) have shown that climate change can result in uncertainty of water availability ranging from the watershed to global scale. As the impacts of climate change to the hydrologic characteristics of a basin are realized, streamflow forecasting can become difficult for hydrologists and climatologists as past hydrologic conditions are no longer representative of future conditions (Thakali et al., 2016; Pathak et al., 2016a; Pathak et al., 2016b; Pathak et al., 2016c; Pathak et al., 2017; Jobe et al., 2017; Tamaddun et al., 2017; Thakali et al., 2017; Thakali et al., 2018; Bhandari et al., 2018)). It is important to understand the relationship between climate variability and the hydrologic response of a basin such that sustainable and efficient management of water related systems can be implemented (Middelkoop et al., 2001; Pahl-Wostl, 2007; Kundzewicz et al., 2009, Maheshwari et al., 2016; Nyaupane et al., 2017; Nyaupane et al., 2018; Bhandari et al., 2018).

Streamflow forecasting is critical for optimal allocation of water to users along with better understanding of future flood and drought in the region. The prediction of streamflow for low flow is even more significant considering the limited water resources. Streamflow prediction on monthly or seasonal basis is desired for long-term planning of water resources (Kalra et al., 2011; Kalra et al., 2012; Kalra et al., 2008; Sharma et al., 2015; Kalra et al., 2017). The vast geographical extent of the water system with limited data collection network result in uncertainty in hydrological prediction and the non-linearity correlation between input and output variables further impede forecasting process (Zealand et al., 1999). Increased climatic fluctuation makes streamflow prediction even more challenging (Grantz et al., 2005).

## **1.2 Research Motivations**

Various modeling techniques have been employed for streamflow prediction. The conventional forecasting models that are common in practice are conceptual models and time series models. Multiple Linear Regression, Auto Regressive Integrated Moving Average are some of the conventional model extensively used for prediction of hydrological time series. However, these models do not represent the non-linear processes involved in precipitation-streamflow transformation (Zealand et al., 1999). These time series models utilize the concept of data stationarity and hence provide little applicability when dealing with non-stationary data. Artificial Neural Network (ANN) has emerged as a dynamic, self-learning model capable of utilizing noisy, non-linear data in predicting hydrological time series without knowing the physical relationship between input and output data (Nourani et al., 2009). ANNs have been applied and performed well in non-linear processes involved in multivariable conditions. Recently, support vector machines (SVM) have received growing attention as a novel regression technique. SVM uses a statistical machine learning approach in which available data are trained

to predict series of data (Liong and Sivapragasam, 2002). It can minimize prediction error and reduce model complexity (Vapnik, 1995, 1998). SVMs evolve incorporating the noise and non-linearity in the training data without assuming the stationarity proving it ideal while analyzing hydrologic parameters affected by climate change. Several of the previous data driven modeling studies using climate information to improve streamflow forecasts have focused on pre-defined oceanic indices rather than entire SST regions that do not introduce spatial bias. To overcome this limitation, this research proposes a novel-modeling framework that would couple a large-scale climate variability into a data driven model and eliminates the spatial bias at a regional scale.

For preprocessing of geophysical data and obtaining the coupled modes of variability between time series of two field, Bretherton et al. (1992) compared four different methods and concluded that SVD was the simplest, easiest to implement, and found SVD to be superior to other methods. SVD is a widely used reliable statistical tool used to find the relationship between two spatial and temporal field previously used by Uvo et al. (1998), Rajagopalan et al. (2000), Wang and Ting, (2000), Sagarika et al. (2015) in the field of hydro-climatology. Popular predefined indices such as ENSO, PDO, and AMO are conventionally used for teleconnection between specific geographical regions and several hydrological variables such as precipitation and streamflow. However, these predefined indices are capable of explaining relationship for limited region only, therefore, such dependency on the predefined indices lead to spatial biases. Utilization of SVD subsides the use of these predefined indices by obtaining unique spatiotemporal correlation pertinent to the considered study area. Previous studies on streamflow forecasting have primarily focused on SST influence while the current research includes  $Z_{500}$ ,  $SH_{500}$ , and  $U_{500}$  data for the analysis. These climate variables are one of the important climate

variables that affect the hydrological cycle. Including these additional variables broadens the scope of the forecasting ability presented here and identifies significant SH<sub>500</sub> and U<sub>500</sub> regions in Pacific and Atlantic Ocean.

### **1.3 Research Objectives**

The objective of the current research was to understand the relationship between regional streamflow of the United States and ocean-atmospheric climate variability of the Pacific and the Atlantic Ocean. Furthermore, improvement of long lead-time streamflow forecasting of a river basin was also investigated. To achieve the objectives of the research, the following questions were addressed in the research.

**Research Question #1:** How does the proposed modeling framework improve the lead-time of the streamflow forecast?

**Research Question #2:** How is the streamflow within a basin associated with ocean-atmospheric climate variability?

### **1.4 Research Outlines**

The thesis is presented in a manuscript format. The first chapter is introduction whereas the second chapter is titled as “*Streamflow Forecasting Using Singular Value Decomposition and Support Vector Machine for the Upper Rio Grande River Basin*”, investigates the first research question. The third title titled as “*Interrelationship between Regional Streamflow and Ocean-Atmospheric Climate Variability of the United States: A Singular Value Decomposition and Wavelet Approach*” addresses the second research question. The fourth chapter presents summary and future recommendations.

## CHAPTER 2

# STREAMFLOW FORECASTING USING SINGULAR VALUE DECOMPOSITION AND SUPPORT VECTOR MACHINE FOR THE UPPER RIO GRANDE RIVER BASIN

### 2.1 Introduction

Water has become a major natural commodity in the Western United States, where limited water availability has been exacerbated by past frequent droughts. Extreme hydrologic events such as floods and droughts are associated with hydro-climatic variability; improved knowledge of that variability in response to climatic fluctuations is crucial to mitigating social and economic impacts (Redmond and Koch, 1991). Several studies (e.g., Christensen et al., 2004; Stewart et al., 2004; Nijssen et al., 2001) have shown that climate change can result in uncertainty of water availability ranging from the watershed to global scale. In 2016, the United States Army Corps of Engineers issued Engineering and Construction Bulletin No. 2016-25 (ECB 2016-25) incorporated that climate change should be considered for all federally funded projects in planning stages. ECB 2016-25 provisioned qualitative analysis of historical climate trends, as well as assessment of future projections. As the impacts of climate change to the hydrologic characteristics of a basin are realized, streamflow forecasting can become difficult for hydrologists and climatologists as past hydrologic conditions are no longer representative of future conditions (Ghimire et al., 2016; Thakali et al., 2017; Jobe et al., 2018; Pathak et al., 2018, ). It is important to understand the relationship between climate variability and the hydrologic response of a basin such that sustainable and efficient management of water related systems can be implemented (Thakur et al., 2017a; Thakur et al., 2017b; Thakur et al., 2018; Bhandari et al., 2018; Kandissounon et al., 2018).

The dominant drivers of climatic variability affecting the hydrologic cycle all over the world and primarily in the U.S. include the El Niño Southern Oscillation (ENSO), Pacific Decadal Oscillation (PDO), Atlantic Multi-decadal Oscillation (AMO), Madden-Julian Oscillation, North Atlantic Oscillation, Arctic Oscillation, and Pacific-North America Pattern. Throughout the U. S., these teleconnection patterns are significant predictors of hydrologic response (Dettinger et al., 1998; McCabe et al., 2004). Sea surface temperature (SST), atmospheric pressure, humidity, and wind are the major ocean-atmospheric variables that have wide influence in explaining the hydrologic variability of a region. SST variability has been utilized to find teleconnections between streamflow, precipitation, and snowpack. Traditional predefined indices have shown consistent results in specific areas such as El Niño phase influences on the southwest, southeast, and northwest US regions (Kahya and Dracup, 1993). Although the identification of predefined SST regions in the Pacific and Atlantic aid in forecasting streamflow in a certain basin, it may not influence hydrology over all basins (Tootle and Piechota, 2006). Consideration of the entire Pacific and Atlantic Ocean SST avoids regional biases and may lead to improved streamflow estimates (Tootle and Piechota 2006). Studies have associated 500-mbar geopotential height ( $Z_{500}$ ) anomalies with climate change (Wallace and Gutzler, 1981).  $Z_{500}$  is the elevation above mean sea level at which atmospheric pressure is 500-mbar.  $Z_{500}$  has been used as a significant predictor in climate forecasting models and has performed well (Grantz et al., 2005; Soukup et al., 2009; Sagarika et al., 2015). Precipitation is related to ocean evaporation and the movement of clouds; these components of the hydrological cycle are primarily impacted by humidity, wind speed, and air temperature. In order to fully address these components, two additional climate data included in this analysis are: zonal wind stress ( $U_{500}$ ) (i.e., east-west wind force per unit area parallel to the surface of water bodies

corresponding to 500-mbar atmospheric pressure) and specific humidity ( $SH_{500}$ ), corresponding to 500-mbar pressure of both the Pacific and Atlantic Ocean.

Principal component analysis, singular value decomposition (SVD), canonical correlation analysis, and combined principal correlation analysis are some of the techniques commonly used to find interrelationship between two spatial and temporal fields (Wallace et al., 1992).

Bretherton et al. (1992) applied afore-mentioned statistical methods to find the coupled relationship between two spatial-temporal variables and opted for SVD for its simplicity and robustness. Wallace et al. (1992) also concluded that SVD extracts the most significant modes of variability in comparison to other tools. Several studies (Wallace et al., 1992; Uvo et al., 1998; Tootle and Piechota, 2006; Soukop et al., 2009) have been conducted to find the linkage and forecasting ability between large scale climate data and streamflow, snowpack or precipitation using SVD technique. Popular predefined indices such as ENSO, PDO, and AMO are conventionally used as predictors of streamflow while these predefined indices are the source of spatial biases. Utilization of SVD subsides the use of these predefined indices by obtaining unique spatiotemporal correlation pertinent to the considered study area. In order to improve the forecasting ability of a model, several data preprocessing techniques are available. In conjunction with data-driven modeling, singular spectrum analysis (SSA) and discrete wavelet transform (DWT) are most common preprocessing tools and these are efficient in eliminating discontinuity of data and reducing forecasting errors (Marques et al., 2006; Nourani et al., 2009). However, recent research by Du et al. (2017) presented the incorrect usage of SSA and DWT in developing hybrid models and showed that those models may cause significant forecasting errors.

Various modeling techniques have been employed for streamflow prediction. The conventional forecasting models that are common in practice are conceptual models and time

series models. Multiple Linear Regression, Auto Regressive Integrated Moving Average are some of the conventional model extensively used for prediction of hydrological time series. However, these models do not represent the non-linear processes involved in precipitation-streamflow transformation (Zealand et al., 1999). These time series models utilize the concept of data stationarity and hence provide little applicability when dealing with non-stationary data. Artificial Neural Network (ANN) has emerged as a dynamic, self-learning model capable of utilizing noisy, non-linear data in predicting hydrological time series without knowing the physical relationship between input and output data (Nourani et al., 2009). ANNs have been applied and performed well in non-linear processes involved in multivariable conditions. Recently, support vector machines (SVM) have received growing attention as a novel regression technique. SVM uses a statistical machine learning approach in which available data are trained to predict series of data (Liong and Sivapragasam, 2002). It can minimize prediction error and reduce model complexity (Vapnik, 1995, 1998). SVMs evolve incorporating the noise and non-linearity in the training data without assuming the stationarity proving it ideal while analyzing hydrologic parameters affected by climate change. SVM uses the principal of structural risk minimization unlike the empirical risk minimization principle used by ANNs. SVMs have been extensively applied in various hydrological forecasting problems and have outperformed ANNs approach (Dibike, 2000; Babovic et al., 2000; Cimen and Kisi, 2009). SVM has shown superior generalization ability and it is successful in reducing the overfitting problem compared to ANN (Cimen and Kisi, 2009). Astuti et al. (2014) used SVD for preprocessing and feature extraction and the extracted data were used to forecast location, time, and magnitude of earthquakes using SVM approach and concluded that the proposed methods were relatively better than the other hybrid forecasting models.

Several of the previous data driven modeling studies using climate information to improve streamflow forecasts have focused on pre-defined oceanic indices rather than entire SST regions that do not introduce spatial bias. To overcome this limitation, this research proposes a novel-modeling framework that would couple a large-scale climate variability into a data driven model and eliminates the spatial bias at a regional scale. First, SVD is used to determine a lagged spatial-temporal correlation between April-August streamflow and oceanic-atmospheric variabilities represented by SST,  $Z_{500}$ ,  $SH_{500}$ , and  $U_{500}$  of the Pacific and the Atlantic Oceans. SVD significant regions are weighted using non-parametric approach formulated by Piechota et al. (2001) and utilized as input in SVM framework. The study is conducted in the Upper Rio Grande River Basin (URGRB) for the period of 1965-2014 and the lagged relationship is computed for 1-13 months.

This study is expected to investigate the time-lagged relationship of the URGRB streamflow variability with the ocean-atmospheric variability of the Pacific and the Atlantic Ocean. This research further aims to address the following research questions: (1) How is streamflow within the URGRB associated with ocean-atmospheric variables? (2) What are the dominant predictors among oceanic-atmospheric variables that best describe the streamflow variability of the basin? and (3) How does the proposed modeling framework improve the lead-time of the streamflow forecast? Previous studies on streamflow forecasting in the URGRB have primarily focused on SST influence while the current research includes  $Z_{500}$ ,  $SH_{500}$ , and  $U_{500}$  data for the analysis. Including these additional variables broadens the scope of the forecasting ability presented here and identifies significant  $SH_{500}$  and  $U_{500}$  regions in Pacific and Atlantic Ocean.2.

## **2.2 Study Area and Data**

### **2.2.1 Study area**

The Rio Grande River is one of the major rivers in the United States, which originates in southwestern Colorado, flows through New Mexico and Texas in a southeasterly direction, and discharges into the Gulf of Mexico. The Rio Grande River, which is approximately 3,051 kilometers in length with a catchment area of 472,000 square kilometers, is a major source of water in southern states. More than three million people, agriculture, industries, and wildlife in Colorado, New Mexico, and Texas have been supported by the Rio Grande water supply (Michelsen and Wood, 2003; Booker et al., 2005). During drought conditions, the water allocation conflict among the users is considered among the most intense in the United States (US Department of Interior, 2003). Increased demand, over-allocation of water, and vulnerability to drought and climate change have created and added complexity in active water regulation and allocation in the URGRB region (Booker et al., 2005). The socio-economic importance of the river motivates the need for improved streamflow prediction several months in advance.

### **2.2.2 Data**

The primary datasets used in analysis are streamflow data for six unimpaired gages in the URGRB and oceanic-atmospheric climate data represented by SST, Z<sub>500</sub>, U<sub>500</sub>, and SH<sub>500</sub>. United States Geological Survey (USGS) Hydro-Climatic Data Network 2009 (HCDN-2009) provides the list of streamflow stations which have minimal impact from human activities such as construction of diversion, artificial dams or any activities which can affect the natural flow of streams. The streamflow data from these stations are suitable for the analysis of hydrologic variations and trends for the present climatic context (Lins, 2012). Slack and Landwehr (1992) identified 1659 unimpaired streamflow stations in the United States (Lins, 2012). However, for

the RGRB, our analysis showed only six streamflow stations have minimal impact from human activities which are located in the upper region of the Rio Grande River Basin. These six stations from the upper region of the basin are the reason for selection of the Upper Rio Grande River Basin. These six stations from the upper region of the basin are the reason behind the selection of Upper Rio Grande River Basin. The mean monthly streamflow values from those streamflow stations are extracted from USGS website (<http://www.usgs.gov/>) for 1965 to 2014. Monthly streamflow volumes from April through August are summed to develop seasonal streamflow volumes for the analysis. Figure 1 illustrates the location of six unimpaired streamflow stations. It is commonly observed that the daily streamflow has high uncertainty and it is difficult to find a time lagged relationship between oceanic-atmospheric data and daily streamflow data. To have higher accuracy in the prediction and to have a better estimate of the seasonal streamflow volume, April-August streamflow volume is used since seasonal variation of streamflow is typical in snow-fed rivers of the United States. Further, spring summer streamflow accounts for the major volume of flow of the year and can help water managers to create balance between future annual water demand and annual water availability. Additionally, seasonal analysis of streamflow with climate variability is preferred to water-year analysis because the water-year analysis does not effectively capture the seasonal interaction of streamflow and climatic variables (Sagarika et al., 2015). The analysis, therefore, aims to capture the seasonal relationship of streamflow and climate variability adequately.

National Oceanic and Atmospheric Administration (NOAA) Physical Sciences Division (<http://www.esrl.noaa.gov/psd/>) is the source of SST data for both the Pacific and Atlantic Oceans. The Reanalysis dataset were obtained from Kalnay et al. (1996). The mean monthly SST data is extracted from 2° by 2° grid cells and the spatial extent of SST data in the Pacific Ocean is

100°E to 80°W longitude and 30°S to 70°N latitude. The extent for the Atlantic Ocean is 80°W to 20°W longitude and 30°S to 70°N latitude. The mean monthly SST data was divided into three periods: December to February of the previous year, September to November of the previous year, and December to February of the current year covering a period of 50 years (1964-2013). For example, if streamflow is predicted for April-August of 2010, monthly average SST data for December 2008 to February 2009, September to November of 2009, and December 2009 to February 2010 are considered in the analysis for the three periods. The lead-time in the analysis is defined as the time lag from the last month of SST period to the first month of streamflow period. 1-month lead-time i.e., February to April, 4-month lead-time i.e., November to April, and 13-month lead-time i.e., previous year's February to current year's April are considered as the three forecast lead-times in this study.

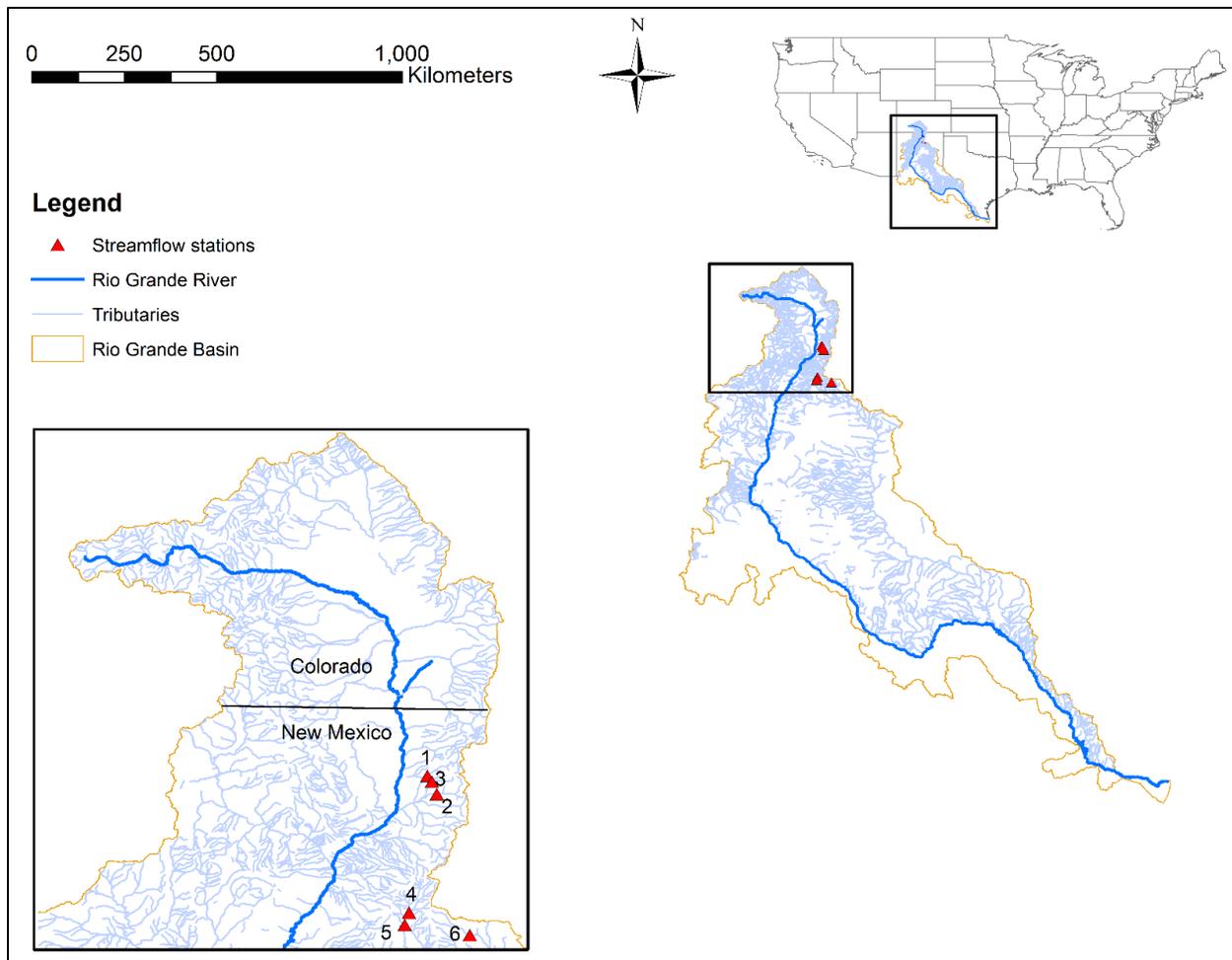


Figure 1: Map showing six unimpaired streamflow stations in the Rio Grande River Basin

In addition to SST, other parameters representing the ocean-atmospheric variability are  $Z_{500}$ ,  $U_{500}$ , and  $SH_{500}$ . NOAA Physical Science Center (<http://www.esrl.noaa.gov/psd/>) provided the mean monthly  $Z_{500}$ ,  $U_{500}$ , and  $SH_{500}$  data from 1964 to 2013. These data are obtained from  $2.5^\circ$  by  $2.5^\circ$  grid cell for both oceans and the spatial extent and division of data is kept the same as that of SST data.

### 2.3 Methodology

The methods used here are divided into four steps:

1. Establishing correlation between two variables using SVD

2. Screening of predictors
3. Predicting streamflow using SVM
4. Model evaluation

The flowchart in Figure 2 summarizes the model algorithm to forecast the streamflow from the ocean-atmospheric variables with different lead times. In first step, SVD is applied to find the spatiotemporal correlations between the streamflow data and the ocean-atmospheric variables that results in the temporal expansion series (TES) of significant modes explained later. These TES are screened in the second step further. The screened predictors are used as the input for the SVM model of each stations and the predictors are screened for each streamflow stations independently. After the streamflow are forecasted they are evaluated in the fourth step by comparing estimated and observed streamflow statistical and graphical aspects. A brief description of all the methods abstracted from several sources is provided in the ensuing sections. Interested readers are referred to original references for detailed descriptions (Bretherton et al., 1992; Piechota et al., 2001; Vapnik, 1995).

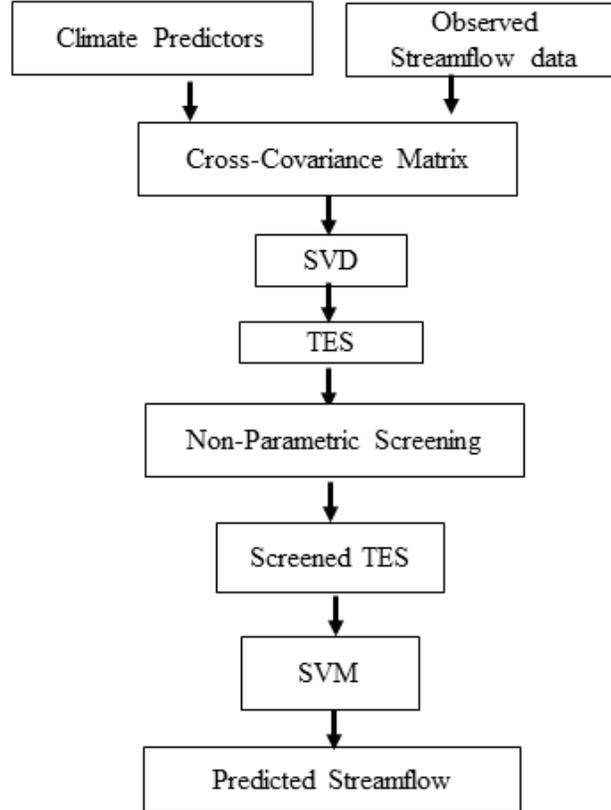


Figure 2: The SVD-SVM model flowchart showing the steps involved in predicting streamflow with the oceanic-atmospheric variables.

### 2.3.1 Establishing correlation between two variables using SVD

SVD is a simple and robust statistical technique primarily useful for differentiating major modes of variability out of extensive series of data. SVD evaluates a cross-covariance matrix between two fields and identifies the correlation between these fields (Bretherton et al., 1992). SVD decomposes an arbitrary matrix  $A$  with  $r$  rows and  $c$  columns into three matrices shown in equation (1):

$$A_{r \times c} = U_{r \times r} S_{r \times c} V_{c \times c}^T \quad (1)$$

Where  $U^T U = I$  and  $V^T V = I$  meaning  $U$  and  $V$  are orthogonal and normalized matrices.  $S$  is a diagonal matrix with non-negative values. A left singular vector and right singular vector are

derived from the columns of those orthogonal and normalized matrices. The first columns and rows of these orthogonal matrices explain more of the correlation between variables compared to subsequent rows/columns. The diagonal matrix provides the singular value of the parent matrix in non-increasing order and these values provide information about the properties of a matrix (Hoecker and Kartvelishvili, 1996. SVD approach to data unfolding. <https://arxiv.org/pdf/hep-ph/9509307.pdf>. Accessed 25 September 1995).

### **2.3.2 Screening of predictors**

The temporal expansion series of eight different ocean-climatic variables obtained from SVD analysis are the possible predictors of streamflow in the Rio Grande. For each predictor, a continuous exceedance probability is calculated by incorporating a kernel density function formulated by Piechota et al. (2001). The forecast developed through the exceedance probability curve of observed streamflow is termed as climatology forecast or no-skill forecast. To determine the skill or reliability of the forecast, Linear Error in Probability Space (LEPS) score approach introduced by Ward and Folland (1991) is used. LEPS score determines the distance between the forecasted and observed value over the cumulative probability function. LEPS score is then calculated for each predictor by incorporating all possible combination of forecasted and measured cumulative probability values and average LEPS skill score (SK) is determined. The average skill (SK) of LEPS determines the forecast performance; a skillful forecast has a 10% or higher LEPS SK score (Potts et al., 1996). Finally, a combined exceedance probability forecast is developed by coupling individual forecast from each predictors with different weights. The weight given for each forecast is such that the summation is equal to one, with higher weights given to those forecasts with higher skill when compared to poor forecasts. The combined forecast is selected based on highest LEPS score for each combination of predictors. The

predictors with highest LEPS values are then defined as the best predictors. This process leads to identification of two best predictors for each streamflow station.

### 2.3.3 Predicting streamflow using SVM

The best combinations of predictors selected are then taken as input for SVM modeling. Unlike traditional learning methods that use an empirical risk minimization principle, SVM uses a machine-learning approach, and this formulation involves a structural risk minimization principle. The application of support vector regression (SVR) is briefly described here. The descriptions and equations are abstracted from Ahmad et al. (2010).

Suppose a training data set with input and output variable represented as,  $\{x_i, y_i\}^N$  where  $x_i \in R^p$  represents independent input variable, and  $y_i \in R$  represents dependent output variable. We need to find a function  $y = f(x)$  that provides the dependency relationship of these two variables. The function can be written as in equation (2):

$$y = f(x) = (w, x) + b, (w + x) \quad (2)$$

In addition, the optimization problem and equality constraints are formulated and shown below in equation (3).

$$\begin{aligned} &\text{Minimize } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \\ &\text{Subject to } \begin{cases} y_i - \sum_{j=1}^K \sum_{i=1}^N w_j x_{ji} - b \leq \varepsilon + \xi_i \\ \sum_{j=1}^K \sum_{i=1}^N w_j x_{ji} + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad (3) \end{aligned}$$

Where  $f(x)$  is the dot product of a weighting vector and input vector represented as  $w$  and  $x$  respectively. The bias is represented by  $b$  while  $\varepsilon$  represents insensitive loss function.  $C$  as the capacity parameter cost,  $\xi_i$  and  $\xi_i^*$  as slack variables, and  $K$  as the number of support vectors are represented in the formulation. The goal is to determine optimal parameters, which minimizes the forecasting error for the SVR model. The optimization of SVM is based on the selection of a kernel function that utilizes non-linear mapping in the feature space (Dibike et al., 2001). Radial bias function kernel is used in the current framework which shows superior efficiency by minimizing test error (Scholkopf et al., 1997). For the detail description of support vector machine, interested readers are advised to go through Vapnik (1995, 1998).

The performance of the model is tested by training the data and validating to the remaining data sets. The training phase intends to find the optimal values of the parameters and attain the best possible generalization conditions. This research utilizes the leave-one-out cross validation approach commonly known as special case of  $k$ -fold cross validation that overcomes the data splitting problem when limited data sets are available for training and testing (Kalra et al., 2012). In this technique one data point is selected for testing by validating the remaining data points and the process is repeated for all the data sets.

#### **2.3.4 Model Evaluation**

SVM performance is evaluated based on various statistical and graphical measures. Time-series plots are used to depict the trend of observed and forecasted streamflow over the years while scatter plots demonstrate the correlation between observed and predicted streamflow values. Similarly, box plots show the statistical variation of streamflow values and non-exceedance probability plots are used to visualize the estimation error at different probability

scenarios. The statistical measures utilize correlation coefficient (R), Nash Sutcliffe model efficiency (NSE), percent bias (PBIAS), and LEPS SK values to evaluate model efficiency.

## **2.4 Results and Discussion**

The results and discussion are described in three different sections. Each section is further divided into 3 sub-sections to describe the results for three different lead-times. The first section discusses the SVD spatial-temporal correlation of streamflow with oceanic-atmospheric variables. Next, the kernel model results are presented followed by the SVM analysis.

### **2.4.1 SVD analysis**

#### **2.4.1.1 1-month lead-time.**

The SVD analysis of the Rio Grande streamflow with Pacific/Atlantic SST,  $Z_{500}$ ,  $SHU_{500}$  and  $U_{500}$  resulted in the identification of significantly correlated regions. The modes in SVD that have a squared covariance fraction (SCF) value greater than 10% are considered for each predictor. SCF value shows the degree of variability explained by SVD analysis. Similarly, normalized squared covariance (NSC) indicates the correlation between two fields averaged over all the grid points and the NSC value has maximum value of 1 if there is perfect correlation between two variables at every grid point (Wallace et al. 1992). Most of the variability of the streamflow in the URGRB was explained by the first mode of SVD and therefore, only the first mode SVD results are reported throughout the section. Table 1 presents the SCF and NSC values obtained for different lead-time scenarios.

Table 1 SVD results for different lead-time cases

Climate variability	Lead-time Months	SST		Z <sub>500</sub>		SH <sub>500</sub>		U <sub>500</sub>	
		SCF (%)	NSC (%)	SCF (%)	NSC (%)	SCF (%)	NSC (%)	SCF (%)	NSC (%)
		Pacific Ocean	1	97.3	6.9	96.4	4.0	96.3	4.6
	4	97.4	5.1	91.4	2.5	95.6	4.3	95.4	3.9
	13	92.4	2.2	89.1	1.6	88.6	2.1	90.5	2.4
Atlantic Ocean	1	96.9	4.5	96.0	2.5	92.5	3.7	90.5	3.0
	4	96.2	4.4	95.4	3.1	95.1	3.8	94.5	3.4
	13	94.6	1.9	92.0	1.5	87.1	2.2	89.5	1.7

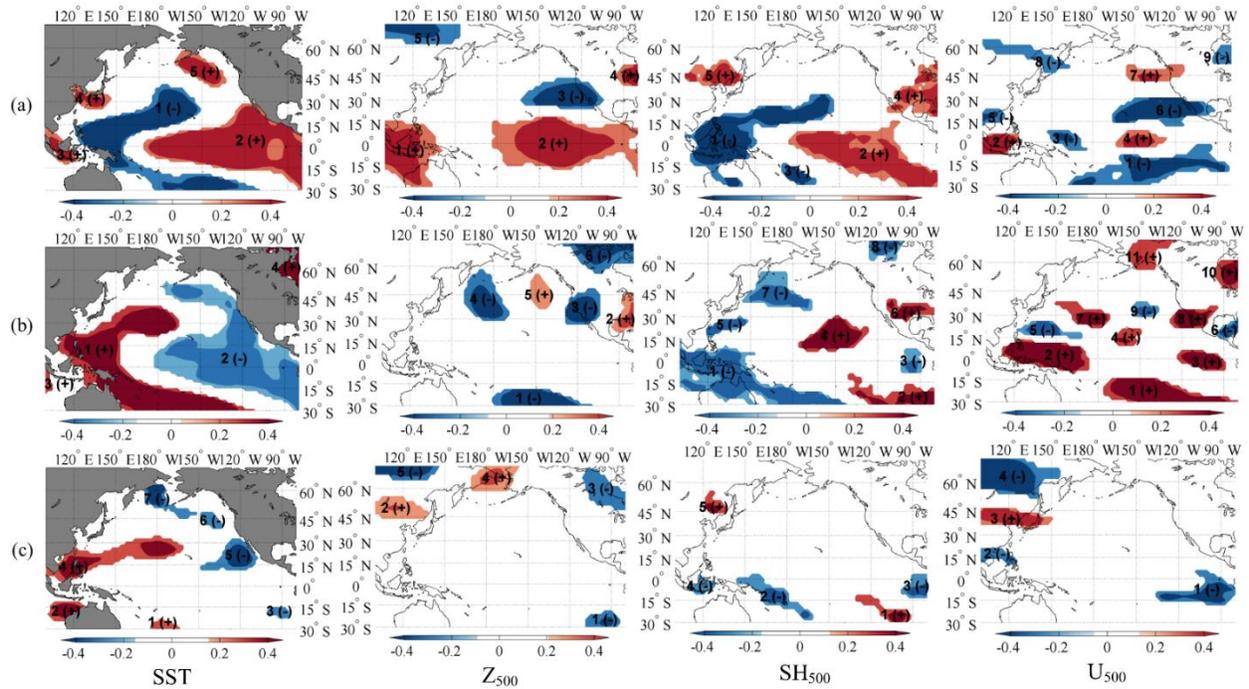


Figure 3: Heterogeneous correlation map for Pacific Ocean (a) 1-month lead-time (b) 4-month lead-time (c) 13-month lead-time SST, Z<sub>500</sub>, SH<sub>500</sub>, and U<sub>500</sub> with April-August streamflow. Significant regions with Positive (negative) correlations are represented by red (blue).

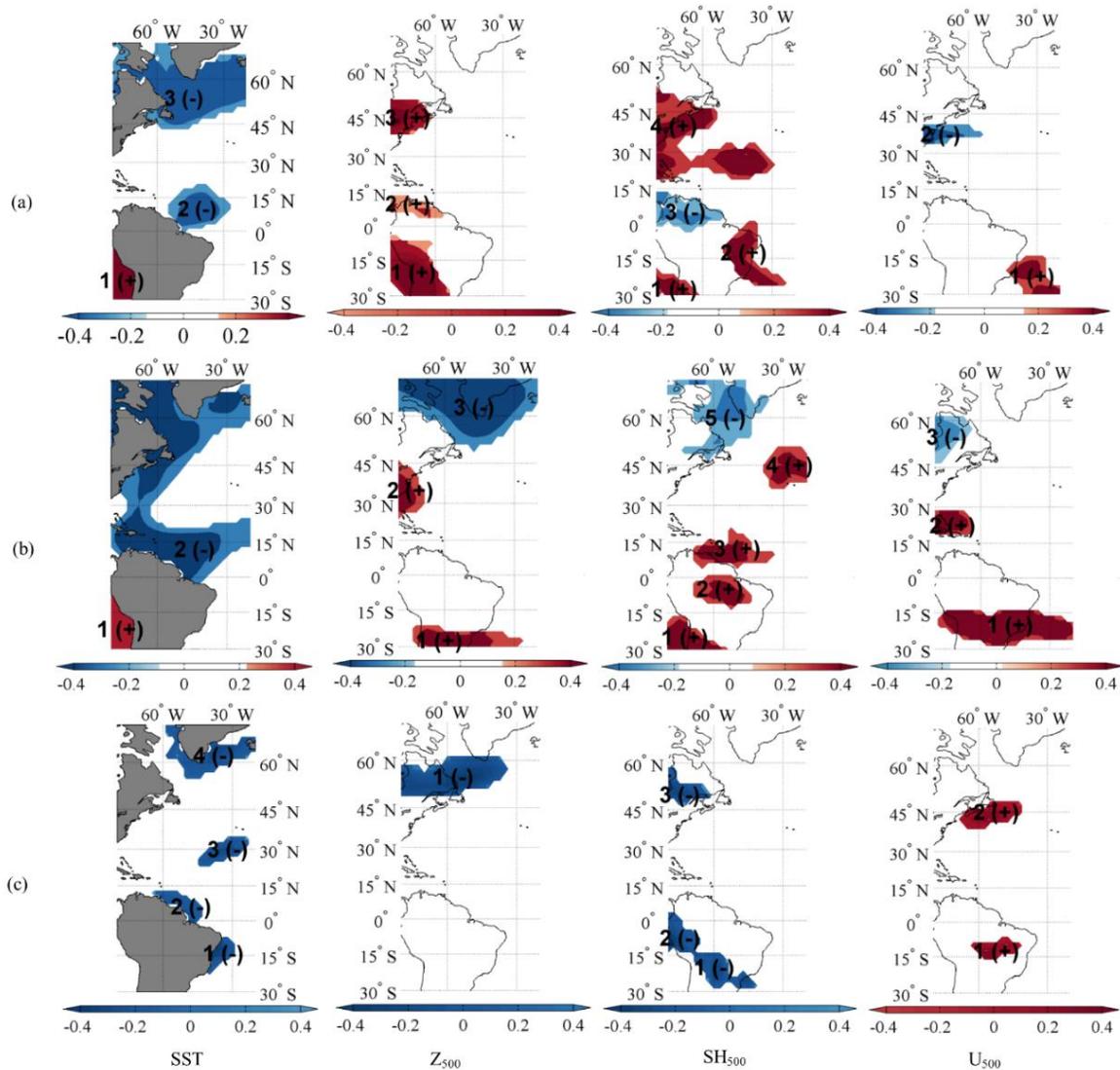


Figure 4: Heterogeneous correlation map for Atlantic Ocean (a) 1-month lead-time (b) 4-month lead-time (c) 13-month lead-time SST, Z<sub>500</sub>, SH<sub>500</sub>, and U<sub>500</sub> with April-August streamflow. Significant regions with Positive (negative) correlations are represented by red (blue).

Figure 3a represents the relationship between Pacific Ocean ocean-atmospheric variability with URGRB streamflow through heterogeneous correlation map at the 90% significance level for the 1-month lead-time period. Resulting significant SST regions are shown in red and blue color. The red and blue color in the map indicates positive or negative correlation of streamflow with climatic variables in the significant regions. Positive or negative correlations

depend upon the station significance. One of the two key significant regions identified for that period are the regions off the coast of Japan, Indonesia and Australia, which has a horseshoe shape, is negatively correlated with the April-August streamflow in URGRB. This identified region is similar to ENSO including the popular Niño 3.4 region previously identified by Trenberth (1997). The identified region also include northwestern US coastal regions representing PDO. Previously, Khedun et al. (2012) has also identified similar results – indicating ENSO and PDO being positively correlated with winter and spring precipitation which is the source of spring summer streamflow. The obtained SST regions-mostly ENSO are affirmed by previous literatures signifying the URGRB being wetter and colder during El Niño years because of the modifications in the mid latitude jet streams. The reason of ENSO being positively correlated with the streamflow can be attributed to the feeding of moisture to the Jet streams moving towards east from the Pacific as a result of above normal SST in ENSO regions during El Niño years. Another dominant region of SST that has a strong positive correlation is the region extending from West to Central Pacific Ocean bounded in between  $90^{\circ}$  W to  $180^{\circ}$  W latitude, and this region shows conformity with Niño Index as demonstrated by other researchers (Rajagopalan et al., 2000). Figure 4a shows a heterogeneous correlation map of Atlantic SST significant regions for the 1-month lead-time period. The identified significant area is separated into two zones, one is near the east coast of Canada and US resembling the AMO region and the other is near the north shore of Brazil. These regions have a negative correlation with streamflow variability. This is also verified by previous literature that cold north Atlantic SST in winter and spring favors the spring summer streamflow. (Trenberth et al., 1998; Pascolini-Campbell et al., 2017).

The second column of Figure 3a and 4a show the heterogeneous correlation map of significant  $Z_{500}$  regions for both oceans. The significant regions for Pacific are more concentrated in equatorial regions of the Pacific Ocean and southeastern Asia. The altitude of  $Z_{500}$  is approximately 18000 feet above the sea level, and this has been associated with diverse weather phenomena (Soukup et al., 2009). The jet stream formation is related to locations where the  $Z_{500}$  contour lines are concentrated.  $Z_{500}$  is found to be more influential during wintertime. The shortwave train as shown in the second column of Figure 3a with red arrow head as a result of warming of SST in ENSO region signifies the fueling of Jetstream with moisture also responsible for the precipitation and streamflow of URGRB region. This physically explains the reliability of obtained teleconnection between  $Z_{500}$  regions and the streamflow of URGRB. For the Atlantic Ocean, the significant regions identified are clustered at Eastern Canada and north to mid-South America. All those identified regions show the positive correlation with the streamflow of the URGRB.

Four significant  $SH_{500}$  regions were prominent in the Pacific Ocean. Two regions with positive correlation are identified at the equatorial region and above the mid-United States while two negatively correlated specific humidity regions were identified at the eastern side of China, Indonesia, and Japan. As mentioned earlier, the warmer than average ENSO region fuels the jet stream with moisture over Pacific moving in east directions responsible for URGRB streamflow. With the extra fuel and pressure the Jetstream shift eastward with higher than normal precipitation in URGRB region. This can be verified in Figure 3a third column, the specific humidity in ENSO region being positively correlated with streamflow of URGRB region. Positive correlated  $SH_{500}$  regions are found near the basin area that may be linked to the direct relationship of distance with humidity influence making further regions less influential in

streamflow variation. For the Atlantic Ocean, the significant regions were identified at eastern Canada and northern South America. These regions have shown the positive correlation with the streamflow of the basin.

Several significant  $U_{500}$  regions are established in the Pacific Ocean as probable predictors of streamflow. The  $U_{500}$  regions signifies below/above normal wind in east west directions which are the key cause of circulation of moisture from the ocean to any watersheds. Here, these complex  $U_{500}$  regions are the result of complex interactions of pressure, geographic features, temperature gradients and other climate variables. One significant region in the Atlantic is found near the northeast coast of Canada, which has shown the negative correlation. The small spatial extent of the region indicates a little influence of Atlantic  $U_{500}$  on the streamflow variability.

#### **2.4.1.2 4-month lead-time.**

The SCF and NSC values are both comparable for 1- and 4-month lead-time scenarios. The first plot of Figure 3b and Figure 4b show the heterogeneous correlation map for September-November SST of Pacific and the Atlantic Ocean respectively. The significant Pacific SST regions are almost the same for 4-month lead-time when compared to 1-month lead-time. However, for the Atlantic, SST regions are continuously extended in a greater area in 4-month lead-time, unlike 1-month lead-time where the significant regions are widely separated in two zones. The SCF and NSC values also dropped slightly in comparison to 1-month lead-time. Significant regions are located in the west to the east region of the northern Pacific Ocean in smaller groups. Furthermore, the spatial extent of significant  $Z_{500}$  regions also decreases in this period compared with a previous period in the Pacific Ocean while the spatial extent slightly increases for the Atlantic Ocean. It can be noted that SCF/NSC values for Pacific are smaller

than that of Atlantic in this period. The identified indices for 4-month lead-time were similar to the identified indices for 1-month lead-time but the signals were weaker with the expense of lead times. As observed in Figure 3b, the indices like ENSO and the climatic phenomenon like short wave terrain and impacts of mid latitude jet stream are still evident like 1-month lead-time.

Compared to 1-month lead-time, the SCF and NSC decrease for the Pacific while increases for the Atlantic Ocean in 4-month lead-time. The third plot of Figure 3b and Figure 4b show heterogeneous correlation map for Pacific and the Atlantic Ocean respectively. Similar to SST and  $Z_{500}$ , significant  $SH_{500}$  regions become separated from each other and smaller in areal extent in this period as compared to 1-month lead-time period. Nearer regions to the basin show positive correlation as in the previous 1-month led time case. The SCF and NSC values were higher in 4-month lead-time in comparison to 1-month lead-time. Furthermore, the majority of positively correlated associated regions are prevalent both in Pacific and in Atlantic regions that are clear from Figure 3b and Figure 4b. All other predictors, with the exception of  $U_{500}$ , have shown comparatively better results for 1-month lead-time when compared with 4-month lead-time.

#### **2.4.1.3 13-month lead-time.**

The first plot of Figure 3c and Figure 4c show the heterogeneous correlation map of SST for Pacific and the Atlantic Ocean respectively. From these Figures, it is clear that the significant SST areas become smaller and sparser when compared to smaller lead-time cases. It may be due to the influence of longer lead-time SST period is less effective than that of a shorter lead-time period. A similar drop of SCF and NSC values were obtained for SVD analysis of  $Z_{500}$ . The number of significant regions decreases and these regions move farther away from the ocean as compared to 4-month lead-time as clearly seen in Figure 3c and Figure 4c. For the Atlantic, only

one significant region near the northeast coast of Canada is identified, with negative correlation with streamflow. As seen in Figure 3c and Figure 4c, the spatial extent of significantly teleconnected regions for 13-month lead-time was smaller than other lead times however, these 13-month long lead spatiotemporal associations can also help water managers by providing longer time window for planning and mitigation measures.

The heterogeneous correlation map for Pacific and Atlantic Ocean  $SH_{500}$  are shown in third plot of Figure 3c and Figure 4c. The spatial extent of the significant regions decrease considerably for 13-month lead-time case in comparison to 1- and 4-month lead-time cases; furthermore, the drop in SCF and NSC values also suggest that the 1- and 4-month lead-time period can have better forecasting abilities as compared to 13-month lead-time period. The fourth plot of Figure 3c and Figure 4c show the heterogeneous correlation map for  $U_{500}$  for the Pacific and the Atlantic Ocean, respectively. In this period, number and spatial extent of significant regions decrease considerably similar to SST,  $Z_{500}$ , and  $SH_{500}$ . All the predictor variables have shown better results for smaller lead-time cases than longer lead-time cases.

The SVD results depict the identification of significant regions of SST,  $Z_{500}$ ,  $SH_{500}$ , and  $U_{500}$  in the Pacific and the Atlantic Ocean, which are teleconnected with the streamflow stations in the Rio Grande River. The identification of various significant regions of climatic parameters in this study indicate a dominant influence of these regions on the streamflow variability and can provide better predictive capabilities than other regions. Moreover, these identified regions were found to be similar with results conducted by previous researchers. The lagged SVD analysis clearly showed that smaller lead-time analysis has better forecasting ability as compared to longer lead-time analysis, though the 13-month lead-time result was also useful. The inclusion of entire Pacific and Atlantic Ocean for SST,  $Z_{500}$ ,  $U_{500}$  and  $SH_{500}$  data has eliminated the regional

biases and the dependence on the existing indices to explain the hydrologic variability has reduced.

#### **2.4.2 Predictor screening analysis**

The abovementioned SVD analysis presented all the possible correlations of streamflow with the oceanic-atmospheric climate variables whereas this predictor screening analysis focuses on best correlation for each of the stations leading to screen the best predictor variables out of the eight climate variables. Table 2 shows two best streamflow predictors for each streamflow station for each scenario. The result shows that each variable is one of the best predictors at least once. It is clear from the table that different predictors are dominant at different lead-times. For 1-month lead-time, Atlantic SH<sub>500</sub> was found to be the best overall predictor while Pacific U<sub>500</sub> was found to be the best overall predictor for both 4- and 13-month lead-time scenario. Similarly, Atlantic Z<sub>500</sub> was the least dominant variable followed by Pacific Z<sub>500</sub>. Previous researchers have primarily focused on SST and Z<sub>500</sub> as the important variables in explaining various hydrologic processes. However, this research has included two more climatic variables i.e. U<sub>500</sub> and SH<sub>500</sub> apart from SST and Z<sub>500</sub> for broader scope. These included variables have shown valuable predictive information. For example, for stations 4 and 6, SH<sub>500</sub> of Pacific Ocean could play a key role in forecasting streamflow at those locations, while Atlantic Ocean SH<sub>500</sub> could be an important predictor of streamflow for station 1. The inclusion of these variables can be justified and supported as both U<sub>500</sub> and SH<sub>500</sub> explain the majority of streamflow variability in the URGRB. Results also suggest that the U wind over the Pacific Ocean is one of the major climatic factors that drive the variability of streamflow in the URGRB. Satisfactory performance of U<sub>500</sub> and SH<sub>500</sub> indicate that these climate variables have greater potential in providing finer

results if they are extensively studied and understood. Thus, these variables have possibility of drawing research attention from climatologists and hydrologists in coming days.

Table 2: Best streamflow predictor variables for different lead-time scenarios

Station	Best streamflow predictors		
	1-month lead-time	4-month lead-time	13-month lead-time
1	Atlantic SST	Pacific U <sub>500</sub>	Pacific SST
	Atlantic SH <sub>500</sub>	Atlantic U <sub>500</sub>	Atlantic U <sub>500</sub>
2	Pacific SST	Pacific SST	Pacific SST
	Atlantic Z <sub>500</sub>	Pacific U <sub>500</sub>	Pacific U <sub>500</sub>
3	Atlantic SST	Pacific SST	Atlantic SST
	Atlantic U <sub>500</sub>	Pacific U <sub>500</sub>	Pacific U <sub>500</sub>
4	Atlantic SH <sub>500</sub>	Pacific Z <sub>500</sub>	Pacific SH <sub>500</sub>
	Atlantic U <sub>500</sub>	Pacific U <sub>500</sub>	Pacific U <sub>500</sub>
5	Atlantic SH <sub>500</sub>	Pacific Z <sub>500</sub>	Pacific SH <sub>500</sub>
	Atlantic U <sub>500</sub>	Pacific U <sub>500</sub>	Pacific U <sub>500</sub>
6	Pacific SH <sub>500</sub>	Pacific SST	Pacific SST
	Atlantic SH <sub>500</sub>	Pacific U <sub>500</sub>	Pacific SH <sub>500</sub>

The positive skill shown by continuous exceedance probability forecast depicts the improvement of prediction skill compared against a climatological forecast where the temporal aspect of historical data is taken into account. The continuous exceedance probability forecast labeled as good, fair, and poor compares climatology forecast, modeled forecast, and observed streamflow value and gives an idea about the availability of streamflow at different risk levels in a simple and efficient way. The continuous exceedance probability forecasts labeled as good, fair, and poor forecast are shown in Figure 5. For good forecast, the difference between observed and predicted streamflow value at certain probability is minimum while for poor forecast the difference becomes larger. From Figure 5c, it is clear that when streamflow was predicted for 2014 at 50% exceedance probability, the model predicts 12 million cubic meters (MCM) while the climatology predicts 18 MCM and the observed value is 11 MCM. However, for poor

forecasts certain amount of risk is also present as the forecasted value deviates from the observed value. For majority of years higher number of good forecasts were observed compared to the poor forecasts. In addition, most of the stations showed higher LEPS SK for 1-month lead-time implying greater confidence in forecasting with smaller lead-time.

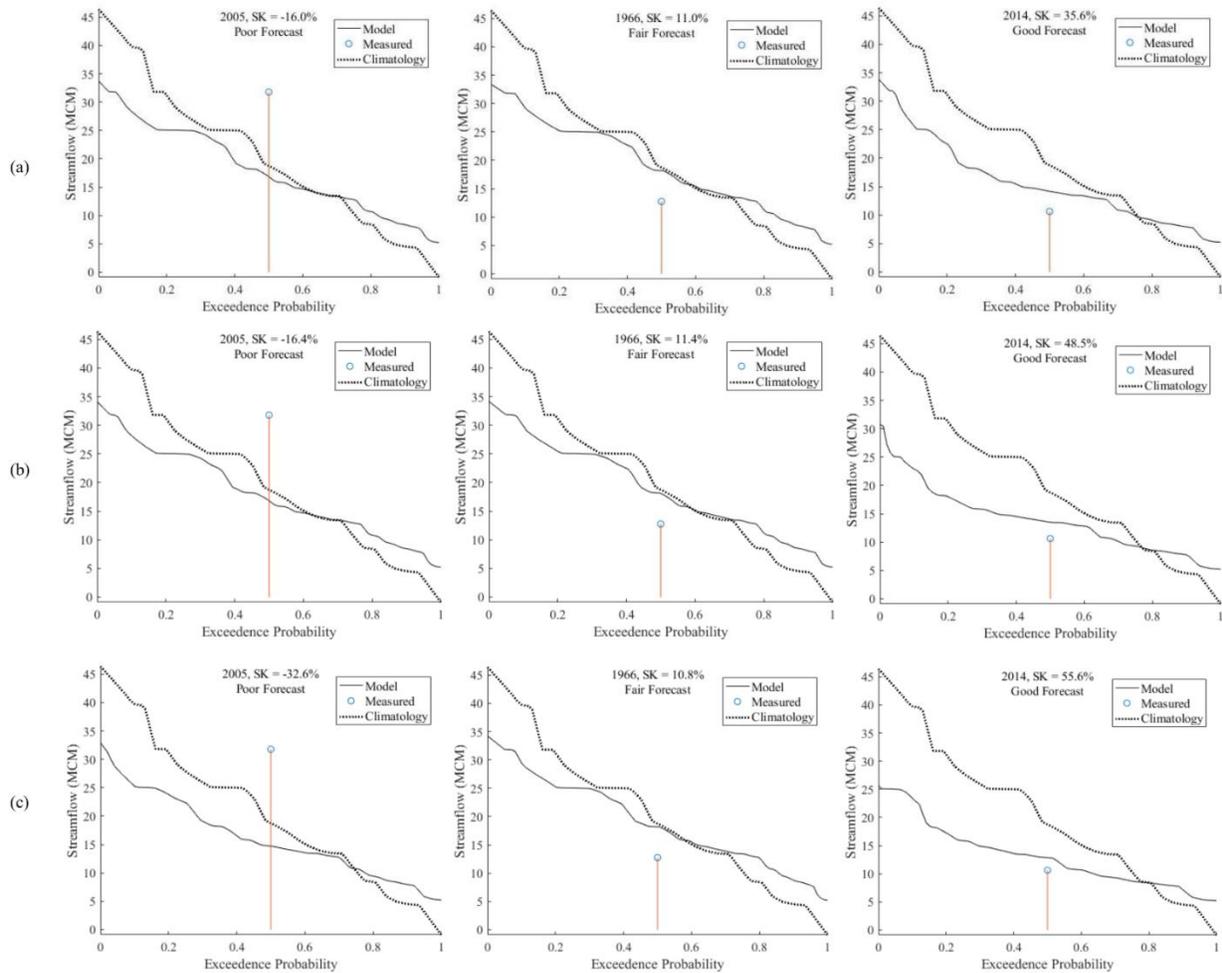


Figure 5: Map showing examples of the poor, fair, and good continuous exceedance probability forecast for (a) 1-month lead-time (b) 4-month lead-time (c) 13-month lead-time

### 2.4.3 SVM Analysis

SVM analysis was used to predict streamflow volumes using input variables from three lead-time cases for all the six stations for a 50-year period. The predicted streamflow is then

compared with observed streamflow, and the performance of the SVM model is described in coming sections.

#### **2.4.3.1 1-month lead-time.**

Figure 6a shows time-series graphs showing the trend of simulated and observed streamflow for the 1-month lead-time scenario. Table 3 presents the values of various model performance parameters obtained for different lead-time cases. The dotted line represents observed streamflow values, and the solid line represents predicted streamflow values. It can be seen that the observed and predicted streamflow volume are fitted well, but some small discrepancies are also present. In addition, the predicted and measured streamflow have similar trends. Simulated streamflow is found to be in almost perfect correlation with the measured streamflow for the year 2000 for all the stations. Similarly, Figure 6b shows the scatter plots for the 1-month lead-time scenario. The points lying above the bisector line indicate the prediction is overestimated, while those lying below the line indicate the predictions are underestimated, and points along the line represent perfect predictions. It can be seen that most points lie along the 45° diagonal showing perfect correlation. This indicates that the forecasted and observed streamflow are in good correlation with each other. Better performance at low flows compared to high flows is apparent on the plots. It can be inferred that model applicability is best achieved during low flow events, which may indicate that the model is well suited for drought conditions.

Table 3: SVM model performance for different stations for different lead-times

Streamflow Station	Lead- time Months	Model performance parameter			
		r	PBIAS (%)	NSE	LEPS SK (%)
1	1	0.87	0.79	0.72	64.1
	4	0.95	-0.53	0.87	78.3
	13	0.85	3.84	0.71	69.8
2	1	0.83	2.02	0.67	63.1
	4	0.93	2.32	0.83	75.0
	13	0.85	4.72	0.69	68.4
3	1	0.91	1.8	0.81	78.9
	4	0.93	3.07	0.86	79.6
	13	0.81	1.26	0.63	61.8
4	1	0.94	-1.48	0.87	75.9
	4	0.87	2.05	0.73	68.7
	13	0.86	2.98	0.72	67.9
5	1	0.92	-2.72	0.84	75.8
	4	0.91	0.84	0.80	76.0
	13	0.94	-0.36	0.86	73.6
6	1	0.90	3.87	0.79	75.6
	4	0.80	5.54	0.60	67.3
	13	0.92	4.11	0.84	77.4

The scatter plot also illustrates the PBIAS value and correlation coefficient. According to Moriasi et al. (2007), a perfect model has zero PBIAS value, and for excellent models, the PBIAS value should be below 10%. It is clear from the Figure 6b that each of the stations has a PBIAS value less than 10%. For the 1-month lead-time period, PBIAS value has an average value of 2.11 for all streamflow stations. The average correlation coefficient for observed and simulated streamflow for the stations was 0.89. The higher correlation coefficient further strengthens the forecasting capability of the SVM model. It implies that the model performs well with less error variance. The effectiveness of SVM model is also evaluated through interpretation of the NSE value. Figure 6b shows the NSE values for all the six stations for the 1-month lead-time case. The average NSE value for all the stations was 0.79. These higher values of NSE statistics indicate that the SVM predicted streamflow was satisfactory.

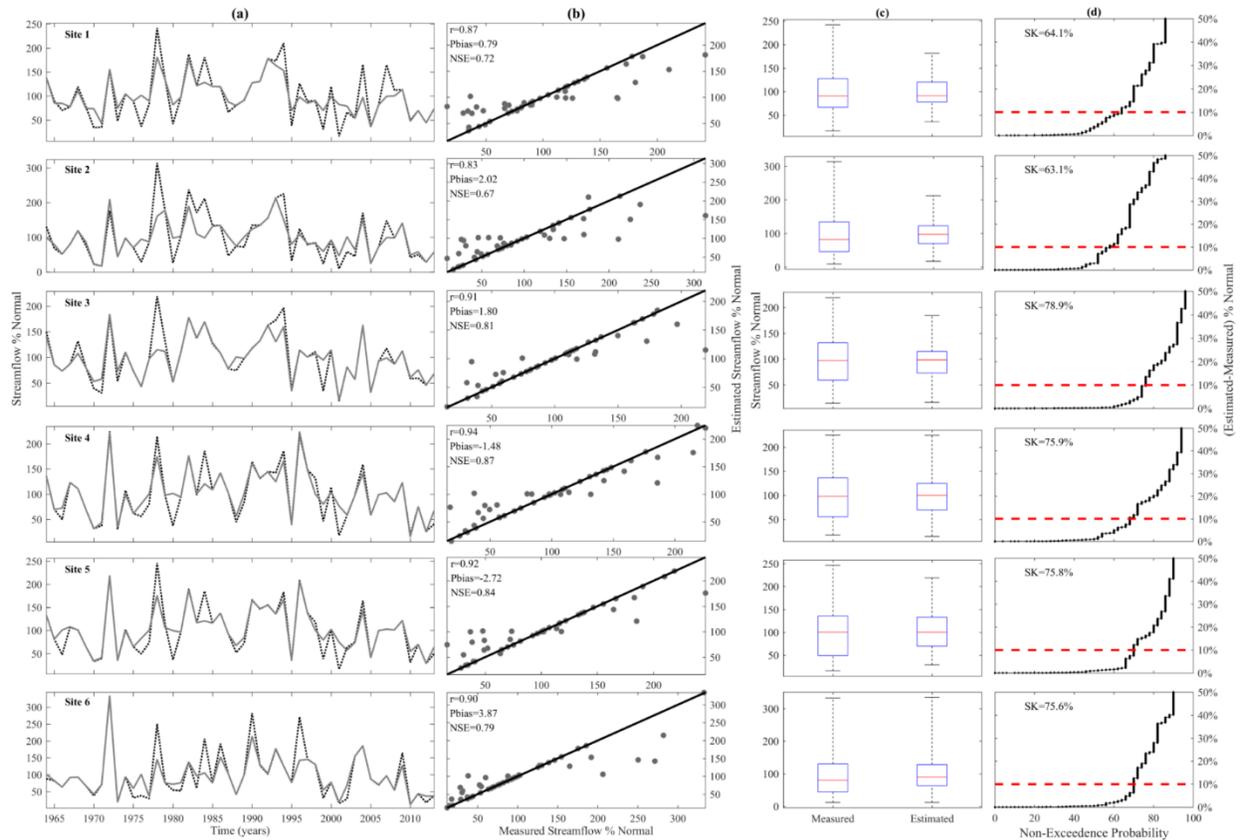


Figure 6: (a) Time series plot (b) Scatter plot (c) Box plot (d) Non-exceedance probability plot for 1-month lead-time depicting the comparison between measured and forecasted streamflow for six streamflow stations.

The box plot of observed and simulated streamflow volume of the model for the December-February period is illustrated in Figure 6c. The horizontal line is the median streamflow while the interquartile range in between 25th percentile and 75th percentile is indicated by the box height. The whiskers in the plot represent extreme 5th and 95th percentiles. The median value for both observed and simulated streamflow values are similar at all sites. Although the interquartile range of estimated streamflow is slightly smaller than observed value, the fifth percentile of both datasets has a closer match. It is clear from the Figures that the interquartile range of measured streamflow is wider as compared to that of predicted streamflow. This illustrates the uncertainty in forecasting ability for high flow range. Furthermore, the model

underestimated the high flow, as most of the predicted high flows are smaller than the observed high flow.

Figure 6d shows the non-exceedance probability plot for the 1-month lead-time scenario. The y-axis represents the percentage of cumulative estimation error and the x-axis is the percentage of predicted data sample which is less than or equal to the value on the x-axis. The dotted line in the plot represents the cumulative modeling error value of 10%. Based on the plot it is clear that at 60% estimate of streamflow, the probabilistic absolute error is around 2% for almost all sites. As per the non-exceedance probability plot, site 3 gives best result, as the absolute error is just 10% at 80% estimate while other sites have more than 10% error for 80% sample estimates. Based on the plot, it is clear that smaller prediction error is achieved at higher estimation percentage, which in turns implies the greater confidence in prediction of streamflow for the water managers. The plot also tells the average skill score to evaluate the performance of the model forecast using LEPS approach. All the stations have LEPS SK more than 60%. The average LEPS SK score value was 72.2% for the streamflow stations. These higher value of LEPS SK score further support good forecasting capability of the model.

### 2.4.3.2 4-month lead-time.

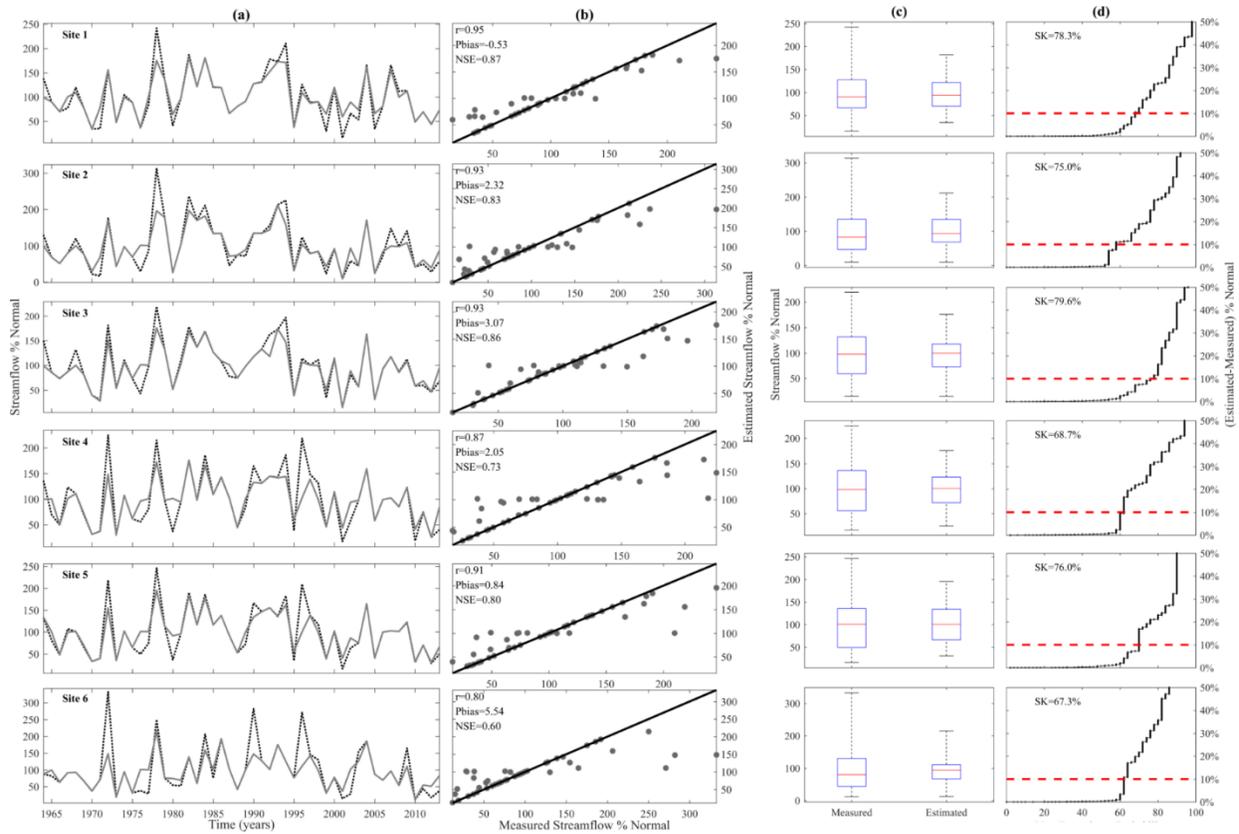


Figure 7: (a) Time series plot (b) Scatter plot (c) Box plot (d) Non-exceedance probability plot for 4-month lead-time depicting the comparison between measured and forecasted streamflow for six streamflow stations.

Figure 7a and Figure 7b show the time-series plot and scatter plot of simulated and observed streamflow values for September to November period, respectively. Similar to 1-month lead-time the time series plot for this period show similar trends. Scatter plots for this period are also similar to those shown at the 1-month lead-time period. Most of the points lie over the bisector line while high flow points are below the line implying high flow values are underestimated by the model. In the September-November case, average PBIAS value was 2.29 at all stations. Compared to the 1-month lead-0time, the PBIAS value has slightly increased at the 4-month lead-time. The correlation coefficient for the stations has an average value of 0.89.

The average value of correlation coefficient for 1-month lead-time is equal to that of 4-month lead-time. . Four stations have NSE values higher than 0.8. The average NSE value was 0.78, which is slightly lower than the 4-month lead-time NSE average. In this period, high flows were also underestimated by the model as seen from the box plot of Figure 7c. The interquartile range of forecasted streamflow for station 6 was the smallest among the stations while the ranges were comparable to one another for rest of the stations. The non-exceedance probability plot for this period is shown in Figure 7d. The average LEPS SK score was 74% that is slightly higher than the 1-month lead-time case.

#### **2.4.3.3 13-month lead-time.**

Figure 8a shows the time-series plot of simulated and observed streamflow values for the 13-month lead-time case. This graph also tells the higher prediction skill of the model. Despite the longest lead-time scenario, the 13-month lead-time scenario shows satisfactory forecasting results. Figure 8b is the scatter plot for this period. The average PBAIS value for this period was 2.87. When PBIAS values are compared, the 1-month lead yields the best results followed by 4- and 13-month lead-times respectively. Even though the smallest lead-time forecast demonstrates the best forecasting ability, the 13-month lead-time forecasting ability of the model is still comparatively satisfactory. The correlation coefficient also follows the same trend. The average R-value in this period was 0.87, slightly smaller than 1- and 4-month lead-time scenarios. When the model capability was measured based on NSE value, the 13-month lead-time case gave the smallest average value at 0.74, a NSE value that is considered satisfactory. It was observed that the NSE value improves as the lead-time period decreases, further supporting anticipated better forecasting ability for the lowest lead-time scenario.

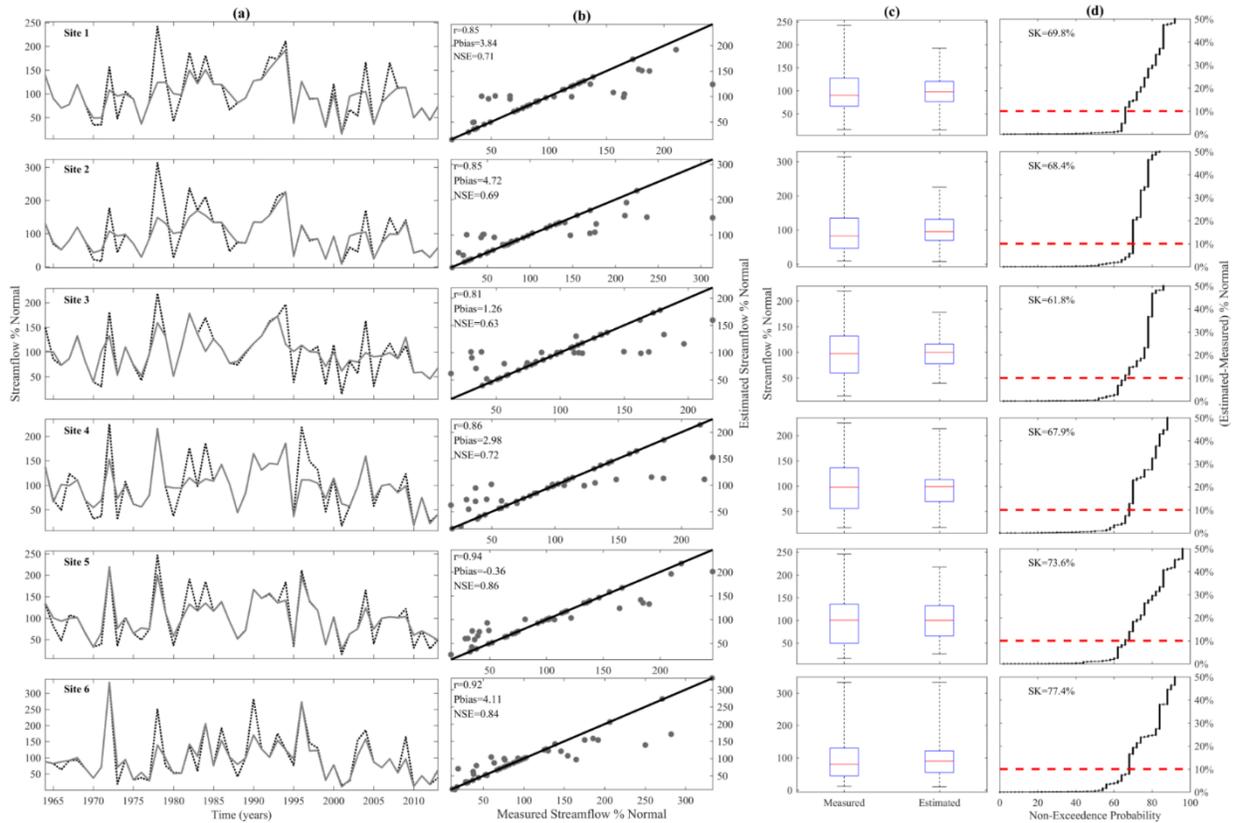


Figure 8: (a) Time series plot (b) Scatter plot (c) Box plot (d) Non-exceedance probability plot for 13-month lead-time depicting the comparison between measured and forecasted streamflow for six streamflow station

Figure 8c shows the box plot at the 13-month lead-time depicting the comparison of modeled and observed streamflow for six streamflow stations. The boxplot for this period showed similar results from previous lead-time periods because, for all three scenarios, the interquartile range of measured streamflow is bigger as compared to that of predicted streamflow. This signifies that all the three scenarios captured both high and low flows, but underestimated high flow, as most of the predicted high flows are smaller than the observed. It can be inferred from the box plot that the model is efficient at predicting the low flow of URGRB as compared to high flow and may be effective for water management during drought seasons. Figure 8d shows the non-exceedance probability plot for the 13-month lead-time case.

The average LEPS SK score value was 69.8%. The average values for this lead-time scenario were the lowest among the three lead-time scenarios. The anticipated higher LEPS SK score for smallest lead-time is not seen here because the 4-month lead-time case has highest average value followed by 1-month lead-time case. The LEPS SK score at all the three lead times indicates the satisfactory forecasting capability of the model.

The SVM model incorporates important oceanic-atmospheric climate variables for improving prediction of URGRB streamflow. The SVM model evaluation by different graphical and statistical analysis resulted in satisfactory results. Scatter plots clearly indicate improved forecast ability. The higher R and NSE values derived between measured and predicted streamflow with consistently smaller PBIAS values further support model forecasting capability. NSE, PBIAS, and R values all indicate that the best forecasting could be achieved for the smallest lead-time while LEPS SK value showed better forecasting ability at the 4-month lead-time followed by 1-month lead-time. Box plots, scatter plot, and time-series plots also suggest the adequate predictability of the model. These plots indicate SVM analysis was able to perform well in capturing low flow as compared to high flow and intermediate flow. This may be due to parameters used in the model that may not adequately represent the physical system governing the generation of streamflow. Possible future research may further investigate the improved forecasting of high flows by considering more physical processes involved in climate-streamflow relation.

## **2.5 Conclusion**

The primary goal of the study was to develop a modeling framework for improving streamflow lead-time in the URGRB using large-scale climate information of the Pacific and Atlantic Oceans. Spatial-temporal relationship of streamflow with each climate variable

represented by SST,  $Z_{500}$ ,  $U_{500}$ , and  $SH_{500}$  was analyzed by SVD approach for three lead-time cases. SVD temporal expansion series for each variable was weighted and screened by a non-parametric approach. These screened variables were used as input in SVM model to predict streamflow at six streamflow stations of the basin. Overall, the research approach is new: the utilization of SVD temporal expansion series of climate variables and employing non-parametric approach for screening the variables used as input for SVM model to forecast streamflow is a novel approach in the field of hydrology.

SVD analysis resulted in new significant  $SH_{500}$  and  $U_{500}$  regions in the Pacific and Atlantic Oceans in addition to SST and  $Z_{500}$  regions. The result presented an extensive idea about all the possible associations between streamflow of the basin and ocean-atmospheric variability. The inclusion of these climate variables led to identify associated significant regions for URGRB and showed equally competent potential for explaining streamflow variability of the basin. These variables have received little attention in previous research efforts. Moreover, the higher correlation of streamflow with  $U_{500}$  and  $SH_{500}$  shows that several other climate variables can be considered together and studied extensively to fully understand the streamflow variability in a basin leading to better water resource management.

The study has shown that SVM model can be a useful method in streamflow forecasting by coupling an extensive range of climate variability with different lead-times. SVM model showed satisfactory forecast results for all the three lead-time cases. The best streamflow forecasting was achieved at the 1-month lead-time followed by 4-month lead-time scenario. The capability to improve long lead-time prediction can be helpful in efficient decision making process and various water management issues when the context of climate change are considered in the basin where snowmelt is the primary source of water. The model showed forecasting

ability over the entire flow range, whereas forecasts at the low flow range were excellent. The Rio Grande River heavily supports domestic use, agriculture, and industry. This river is highly utilized for water supply, and downstream supplies are significantly decreasing over the years (Bartolino and Cole, 2002). The basin has experienced low flows since 2000, and frequent droughts have been reported over the years (Woodhouse et al., 2013). The rainfall pattern and water demand also differ considerably as more precipitation is observed in summer while the peak demand occurs in spring. The ability to capture low flows efficiently aids in water management during drought seasons and below average periods. Better forecasting of low flow events several months ahead may aid in the better allocation of water to competing users during dry periods.

The study doesn't assume stationary climate system and makes the assumption that stationarity is not valid. Relying on the appraisal of past climate by inferring from extreme events or changes in mean is not beneficial as the climate is constantly changing. At this moment, based on the results we obtained, conclusions can be drawn about the magnitude of change in streamflow in the future. While viewing these results, it must be looked at with a range of uncertainties as it is a statistical analysis. But, these uncertainties need to be looked at while making infrastructural investments as these decisions are irreversible. The streamflow predictions that the present study has made in terms of volume should be utilized by the water managers by making decisions so that these infrastructures can effectively respond to conditions that are changing and completely unknown. Detecting changes in past and future is not sufficient to make policy decisions and is the subject that needs more research on. Future work, may explore extended lead-time scenarios. Additionally, the application of paleo data may

provide promising results as data-driven models show higher efficiency for wide range of input data.

## CHAPTER 3

# INTERRELATIONSHIP BETWEEN REGIONAL STREAMFLOW AND OCEAN-ATMOSPHERIC CLIMATE VARIABILITY OF THE UNITED STATES: A SINGULAR VALUE DECOMPOSITION AND WAVELET APPROACH

### 3.1 Introduction

According to Intergovernmental Panel on Climate Change (IPCC, 2014), recent climate change has affected the human and natural systems all over the world. A few of the observed effects of climate change are increasing global temperature, diminishing snow and ice cover, and rising sea level. Alteration of regional hydrological cycle and subsequent effect on quality and quantity of water resources are some of the major ramifications of changing climate (Gleick, 1989). The magnitude, frequency, and intensity of precipitation is changing across the globe. The availability of water is also becoming uncertain in response to climate change over the years (Middelkoop et al., 2001). Variation in timing and magnitude of streamflow, change in groundwater flow and natural reservoirs can affect water resources planning and management. The effects of climatic variability on streamflow have important implications on water management system and the knowledge of such variability may support water management decisions (Miles et al., 2000). Since streamflow is a major component of water cycle streamflow variability can indicate overall variation of water cycle within a region (Tamaddun et al., 2017). The interrelationship between streamflow and the climate change is pivotal for reliable prediction of several hydrologic processes and mitigating climate change induced disasters.

Previous research efforts have shown that streamflow variability is linked with large-scale circulation patterns of the oceanic-atmospheric system (Redmond and Koch, 1991; Cayan and Webb, 1992; Dettinger and Diaz, 2000). Several teleconnection patterns are known to

influence the hydrologic variability on a local as well as global scale (Wallace and Gutzler, 1981; Kahya and Dracup, 1993; Barlow et al., 2001). These teleconnection patterns varying both spatially and temporally are found to influence the variability of the atmospheric circulation throughout the United States. The major oceanic-atmospheric oscillations that have influence on hydrology of U.S. include El Nino Southern Oscillation (ENSO), North Atlantic Oscillation, Arctic Oscillation, Pacific-North American pattern, Madden-Julian Oscillation (MJO), Pacific Decadal Oscillation (PDO), and Atlantic Multidecadal Oscillation (AMO). ENSO effect has shown strong predictive ability and its global influence on hydro-climatic anomalies is also evident (Webster et al., 1998; Ward et al., 2010). Stronger association of U.S. hydrology with ENSO has been understood and studied extensively (Ropelewski et al., 1987; Redmond and Koch, 1991). For PDO, the period of oscillation and the area of influence is bigger than ENSO (Trenberth and Fasullo, 2007). The AMO has a periodicity of 60-80 years and it primarily affects air temperature and rainfall pattern across the northern hemisphere (Enfield et al., 2001). The PDO in conjunction to AMO is found to be associated with majority of multidecadal drought patterns in the United States (McCabe et al., 2004). MJO is a moving intraseasonal oscillation primarily causes anomalous precipitation pattern in Indian and Pacific Ocean. Even though these major oceanic-atmospheric oscillations have widespread effects on the climate system, they may not explain the hydrologic variability in every watershed and offer limited predictive information (Grantz et al., 2005). Consideration of entire oceanic region in the analysis reduces spatial biases and provides wider scope of possible teleconnections (Tootle and Piechota, 2006).

Several researchers have identified the influence of different climate variables that drives the hydrological cycle of a region. Among the variables sea surface temperature (SST), atmospheric pressure, humidity, and wind speed directly influence the mechanism of hydrologic

cycle. Heating of the ocean, evaporation of water, formation of clouds, movement of clouds, and final precipitation are driven by the abovementioned hydro-climatological variables. Streamflow variability can be attributed to any change in temperature, pressure, wind speed or humidity within the hydrological cycle. To understand the physical mechanism governing the hydrologic variability a composite analysis using SST, pressure level, zonal wind speed, and humidity is preferred. Several documented literature have shown strong relationships between the U.S. streamflow and the Pacific/ Atlantic SST variability (Tootle and Piechota, 2006; Pascolini-Campbell et al., 2017). Another important oceanic-atmospheric variable, which have strong coupled relationship with streamflow variability is atmospheric pressure level. The atmospheric pressure is expressed in terms of equivalent height at specific pressure level termed as geopotential height in the current study. Many studies have shown the existence of strong correlation between the geopotential height at 500-mbar pressure level ( $Z_{500}$ ) and climate variability (Wallace and Gutzler, 1981; Dwttinger and Diaz, 2000; Corte-Real et al., 1995). Wallace and Gutzler (1981) pointed that the teleconnection pattern at 500 mbar pressure level is stronger than other levels. Two additional climate variables utilized by the research are zonal east-west wind speed ( $U_{500}$ ) and specific humidity ( $SH_{500}$ ) both measured at 500-mbar pressure level. Pathak et al. (2017) used  $U_{500}$  to find the teleconnection between snow water equivalents of the western United States and the Pacific and the Atlantic Ocean. The objective behind the selection of these variables is to evaluate the interconnection of SST,  $Z_{500}$ ,  $SH_{500}$ , and  $U_{500}$  in driving the streamflow variability. Furthermore, all these climate variables are interconnected and dependent on one another in driving the hydrological cycle. For example, variation of SST may cause wind speed or specific humidity to change, which may eventually affect precipitation

patterns of the affected watershed. In addition, inclusion of these variables may broaden the scope of the analysis.

Several techniques are available to find the interrelationship and dependency for multivariate analysis. One of the popular tools to examine the spatio-temporal association between two variables is the singular value decomposition (SVD). SVD has been extensively used to find the correlation between streamflow and various climate data. Bretherton et al. (1992) used various statistical analysis including SVD technique to find coupled pattern of various climate data and concluded SVD is simple and efficient. Wavelet transform is another technique that has been used to evaluate variance and isolate important information from nonlinear and non-stationary hydroclimatic data with complex periodicities (Tamaddun et al., 2017). Continuous wavelet transform (CWT) is a type of wavelet transform used to determine the variability in data while cross wavelet transform (CWT) and wavelet coherency (WTC) are employed for evaluation of correlation between two different time series data. Interested readers may also refer to Coulibaly and Baldwin (2005), and Labat (2010) for detail descriptions of wavelets in hydro-climatic data analyses.

The motivation behind this research was to evaluate the hydrologic responses in each of the geographic regions of the conterminous U.S. defined from 2014 National Climate Assessment (NCA) under changing climate conditions. Previous studies (Tootle and Piechota, 2006; Sagarika et al., 2015) have used SVD to find coupled relationship between streamflow and climatic variability of the conterminous United States. However, the studies have not focused the variability on regional level separately. In addition, consideration of the interlinkage between the climate variables has not received much attention in previous studies. The primary objectives of this research were to evaluate:

1. The association between conterminous U.S. streamflow and climate variables of the Pacific and Atlantic Oceans and to analyze the driving mechanism of streamflow variability due to these climate variables.
2. The correlation between regional streamflow and three important predefined indices, namely, the ENSO, PDO, and AMO.

In the current research, SVD was applied between April-August streamflow of each of the NCA regions and September-November and December-February SST data of both the oceans. The SVD-SST relation was then correlated with  $Z_{500}$ ,  $SH_{500}$ , and  $U_{500}$  data separately. Moreover, a wavelet approach was also proposed to study the correlation between U.S. streamflow and ENSO, PDO, and AMO separately. Utilization of principal predefined indices together may explain the majority of the streamflow variability of the region. The combination of both the SVD and wavelet approach for the comprehensive study of the U.S. streamflow with respect to climate variables aims to investigate the interdependency of streamflow with climate variability and change at regional level.

### 3.2 Study Area and Data

Table 4: NCA regions and the constituent states

<b>Great Plains</b>	<b>Midwest</b>	<b>Northeast</b>	<b>Southeast</b>	<b>Southwest</b>	<b>Northwest</b>
Texas	Ohio	Maine	Virginia	California	Idaho
Oklahoma	Indiana	Vermont	Kentucky	Nevada	Oregon
Kansas	Michigan	New Hampshire	Tennessee	Utah	Washington
Nebraska	Illinois	Massachusetts	Arkansas	Arizona	
South Dakota	Wisconsin	New York	Mississippi	New Mexico	
North Dakota	Missouri	Rhode Island	Alabama	Colorado	
Wyoming	Iowa	Connecticut	Louisiana		
Montana	Minnesota	New Jersey	Georgia		
		Pennsylvania	Florida		
		Maryland	South Carolina		

### 3.2.1 Study area

The conterminous U.S. is geographically distinguished into six major regions i.e. Great Plains, Midwest, Northeast, Northwest, Southeast, and Southwest region from NCA. The detail description and map of the regions can be viewed from NCA website

(<https://nca2014.globalchange.gov/>). Table 4 illustrates the NCA regions and constituent states.

Figure 9 illustrates the map of all the NCA regions and the analyzed streamflow stations.

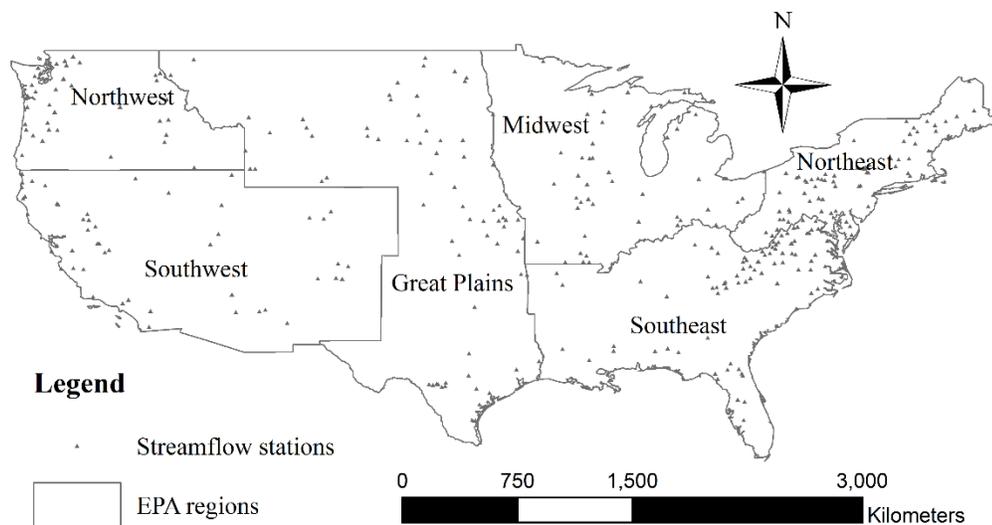


Figure 9: Map showing NCA regions and streamflow stations

### 3.2.2 Data

The dataset for the analysis comprises of streamflow dataset for 350 unimpaired streamflow stations and climate dataset represented by SST, Z<sub>500</sub>, SH<sub>500</sub>, and U<sub>500</sub> of the Pacific and Atlantic Oceans along with ENSO, PDO, and AMO climate indices.

### **3.2.2.1 Streamflow data**

Unimpaired streamflow data are used such that the variability of streamflow could only be attributed to climate change effects. United States Geological Survey (USGS) Hydro-Climate Data Network 2009 (HCDN-2009) provided streamflow dataset for 743 unimpaired streamflow stations throughout the country (Lins et al., 2015). Based on the spatial distribution of stations and availability of continuous data, streamflow data with different time range were selected for each region. Selection is intended to cover the maximum number of streamflow stations that can have potential to show wide variability within a region. Even though the starting years of the streamflow data were different, the ending year of the dataset was kept at 2015 for all the regions. The covered data range varied from 55 to 65 years. Out of the 743 stations, 350 streamflow stations were considered (Table 5). The mean monthly streamflow data from April to August were extracted from USGS website (<http://www.usgs.gov/>) and summed for the analysis.

### **3.2.2.2 Climate variability data**

The climate data were composed of mean monthly values of SST,  $Z_{500}$ ,  $SH_{500}$ , and  $U_{500}$  of the Pacific and the Atlantic Ocean. The SST data were extracted from National Oceanic and Atmospheric Administration (NOAA) Physical Sciences Division (<http://www.esrl.noaa.gov/>). The mean monthly SST data were extracted from  $2^{\circ} \times 2^{\circ}$  grid cells and the spatial extent of SST data in the Pacific Ocean was  $100^{\circ}E$  to  $80^{\circ}W$  longitude and  $30^{\circ}S$  to  $70^{\circ}N$  latitude. The extent for the Atlantic Ocean was  $80^{\circ}W$  to  $20^{\circ}W$  longitude and  $30^{\circ}S$  to  $70^{\circ}N$  latitude. The mean monthly SST data were divided into two periods: September to November of the previous year and December to February of the current year. For example, if streamflow was analyzed for April-August of 2005, monthly average SST data for September to November of 2004, and December

2004 to February 2005 were used for the analysis. This division created two lead-time cases i.e., 1-month lead-time and 4-month lead-time. Since each region had different range for streamflow data, all the climate data were utilized accordingly to develop above-mentioned time-lagged relationship.

NOAA Physical Science Center (<http://www.esrl.noaa.gov/>) provided the mean monthly  $Z_{500}$ ,  $U_{500}$ , and  $SH_{500}$  data. These data were obtained from  $2.5^\circ$  by  $2.5^\circ$  grid cell for both the oceans and the spatial extent and division of data were kept the same as that of SST data. The monthly indices of ENSO (December-February average) and AMO (annual average) were extracted from the NOAA (<https://www.esrl.noaa.gov/>) online databases for their respective indices. The online database of Joint Institute for the Study of the Atmosphere and Ocean (<http://www.jisao.washington.edu/>) provided the monthly PDO (December-February average) indices.

### **3.3 Methodology**

The analyses composed of two parts: (i) SVD analysis to find the spatio-temporal correlation between regional streamflow and SST/ $Z_{500}$ / $SH_{500}$ / $U_{500}$  of the Pacific and Atlantic Oceans; (ii) CWT analysis to determine the spatio-temporal association between regional streamflow and the ENSO, PDO, and AMO indices across multiple timescale bands.

#### **3.3.1 Singular value decomposition**

SVD is a data mining technique, which involves reduction of data dimensionality primarily used to explain mutual dependence and variance of huge series of data (Praus, 2005). Among other statistical tools, SVD is considered a simple, precise, and robust method for revealing correlation between two fields (Bretherton et al., 1992). From the viewpoint of linear algebra, SVD is related to the Eigen decomposition and it is useful in analyzing rectangular

matrices (Abdi, 2007). It is well known for its stability and convergence even for ill conditioned problems (Praus, 2005). The SVD analysis yields singular values, which contain very valuable information about the properties of the matrix. Isolation of the most important modes of data is calculated based on squared covariance fraction (SCF). Similar to SCF, normalized squared covariance (NSC) indicates the correlation between two fields averaged over all the grid points. The NSC value ranges from 0 to 1 with maximum value for perfect correlation between two variables.

The use of SVD for the research is briefly described below. First, standardized SST anomalies matrix and standardized streamflow matrix was developed and cross-covariance matrix generated to apply SVD. SVD yields two singular vector matrices called left and right matrix and one diagonal matrix with singular value. In this study, left singular vector was SST matrix and right singular vector was streamflow matrix. Temporal expansion series (TES) of left field was developed by projecting the column elements of left vector into standardized SST anomalies matrix while TES of right field was developed by projecting the column elements of right vector into standardized streamflow matrix. Finally, left (right) heterogeneous correlation map was developed by correlating left (right) vector with right (left) TES field. The TES of SST was then correlated separately with standardized  $Z_{500}/SH_{500}/U_{500}$  data. Correlating TES of SST with other climate variables helps in retaining the initial relationship of SST with streamflow. This approach therefore allows to evaluate the interconnection of streamflow variability with hydro-climate variables in conjunction with SST influence.

### **3.3.2 Continuous wavelet transforms and its derivatives**

Application of CWT and its derivatives used in this study were based on the steps suggested by Torrence and Compo (1998), Torrence and Webster (1999), and Labat (2005). The

variance of a time-series data, known as the high power in the wavelet spectrum, was detected across continuous timescale bands using CWT from Foufoula-Georgia and Kumar (Foufoula-Georgiou et al, 1995). For each region, a principal component analysis (PCA) was conducted among all the streamflow stations and the first principal components (FPC) were selected to be the representatives of the regional streamflow patterns. Table 5 provides the number of stations and the variance from FPC for each region obtained from the PCA. According to Percival and Walden (2006), the Morlet function is the most fitting mother wavelet function in dealing with hydro-climatic data, hence, it was chosen in the study as the wavelet function.

To determine the association between two time series data across multiple timescale bands, two derivatives of CWT namely cross wavelet transform (XWT) and wavelet coherency (WTC) analyses have been used. XWT and WTC determine the covariance and correlation, respectively, between the two time series. Besides the association between the data, XWT and WTC also provide information regarding the relative phase relationship, i.e., in-phase, anti-phase, lagged response, and simultaneous response, across the various bands along the study period (Jevrejeva et al., 2003; Grinsted et al., 2004). Using the Monte Carlo simulation approach (Wallace et al., 1993), WTC is also capable of quantifying the correlation between two time series even at lower common power (lower association). Hence, the current study presented the results obtained from the WTC analyses only. The wavelet significance was at 5% against the red noise (Torrence and Compo, 1998).

Table 5: The number of stations, data range, and the total percentage of variance explained by FPC

SN	EPA regions	No. of stations	Data range	Total variance explained by FPC (%)
1	Great Plains	69	1960-2015	59.8
2	Midwest	45	1955-2015	49.7
3	Northeast	66	1955-2015	58.3
4	Northwest	38	1950-2015	80.1
5	Southeast	82	1960-2015	58.9
6	Southwest	50	1960-2015	71.8

### 3.4 Results and Discussion

The results and discussion section are also separated into two sections each for SVD and CWT analysis.

#### 3.4.1 SVD analysis

The results of the SVD analysis is presented in this section. Figure 10-15 shows the heterogeneous correlation map, which have 95% significance level, developed for each region along with significantly correlated streamflow stations. The heterogeneous correlation maps developed from the SVD analysis show significant regions for SST, Z<sub>500</sub>, SH<sub>500</sub>, and U<sub>500</sub> of both Pacific and Atlantic Ocean for two different lead-time cases. For clarity, each significant region was given a name and number. For example: SST-1, Z-2, SH-3, U-4 and so on. First mode SVD results were considered as it explained most of the variability. Table 6 presents the SCF and NSC value for both lead-time cases. The result for each region is explained in four subsections as: (i) Pacific 4-month lead-time, (ii) Pacific 1-month lead-time, (iii) Atlantic 4-month lead-time, and (iv) Atlantic 1-month lead-time.

### 3.4.1.1 Great Plains

For 4-month lead-time case, five SST significant reasons were found in the Pacific Ocean. SST-2 region found near the ENSO region had the largest spatial extent. The region bounded in between 5°N-5°S and 170°E-120°W has been reported as typical ENSO region (Alexander et al., 2002). Figure 10a shows that the positively correlated significant regions had smaller number of streamflow stations in the southernmost part of the region especially in the southern Texas area. The positive-negative relationship between the significant region and the streamflow stations has some ability to explain the predictability as well as variability of streamflow. The arrangement of each of figure in vertical direction was ordered such that the SST increase or decrease is accompanied by variation of pressure and specific humidity of the region represented as  $Z_{500}$  and  $SH_{500}$  respectively. Finally, the effect of  $U_{500}$  wind was considered in the variation of streamflow seen in the streamflow station. This physical relationship among the climate variables and the driving mechanisms of hydrological cycle can be understood in the sequence. The increase in SST in SST-2 region was accompanied by increase in the pressure zone near to northwestern US states and south Asian region shown in Z-4 and Z-2 region respectively. The increase in the SST and pressure increased the specific humidity around equatorial region and ENSO region shown by SH-2. This overall increase in the pressure and specific humidity strengthened the wind circulation over the study area represented by U-4 region. The positive correlation between the SVD regions and the streamflow stations shown by red color depicted that the streamflow volume increased in the southern streamflow stations of Texas. Similarly, the negative signal was observed in the SST-1, SST-3, and SST-4 regions. The decrease of SST was accompanied by dropping of pressure in Z-3 and Z-5 region. The cooling SST phase supported to decrease the specific humidity around eastern Asian regions and wind

stress was in diminishing trend near to northwestern US coast and the streamflow was decreasing in the northern streamflow stations of the Montana, North Dakota and Wyoming.

Table 6: SCF and NSC values obtained from SVD analysis for different lead-time cases

Climate variability	Pacific Ocean				Atlantic Ocean			
	1		4		1		4	
Lead-time (months)	SCF (%)	NSC (%)	SCF (%)	NSC (%)	SCF (%)	NSC (%)	SCF (%)	NSC (%)
EPA regions								
Great Plains	49.8	3.2	46.7	3.3	46	2.8	59.7	2.9
Midwest	55.4	2.6	53.7	2.4	62.1	2.4	64.3	2.5
Northeast	45.6	1.2	50.8	1.5	55.7	1.5	55.8	1.5
Northwest	89.1	6.2	89.1	5.7	71.1	2.7	72.4	2.6
Southeast	53.3	1.9	49.6	1.9	59.4	2.2	66.3	2.3
Southwest	73.4	3.7	72.8	3.1	46.1	2.6	66.5	2.5

Figure 10b suggests that the decrease in the lag time from 4-month to 1-month resulted in increase of significant regions. Furthermore the spatial extent of the positively correlated regions was also bigger as compared to the 4-month lead-time case. Both scenarios had shown consistent SST regions that have been teleconnected with the streamflow stations of the Great Plain. The warming phase of SST was accompanied by the increase in the pressure zones significantly throughout the equatorial regions represented as Z-1 that might have increased humidity over Alaska, west equatorial region, and central US regions. Wind stress was also going higher in central Pacific regions. The overall effect showed positive correlation in stations of Kansas, Nebraska and Texas. Similar correlation was obtained by Tootle and Piechota (2006) when SVD applied for Pacific SST streamflow of Great Plain stations. The previously identified cooling SST regions were consistent and spatial extent was also bigger. This can be accompanied by reduction of atmospheric pressure and specific humidity in Z-2 and SH-6 and SH-4 regions respectively.

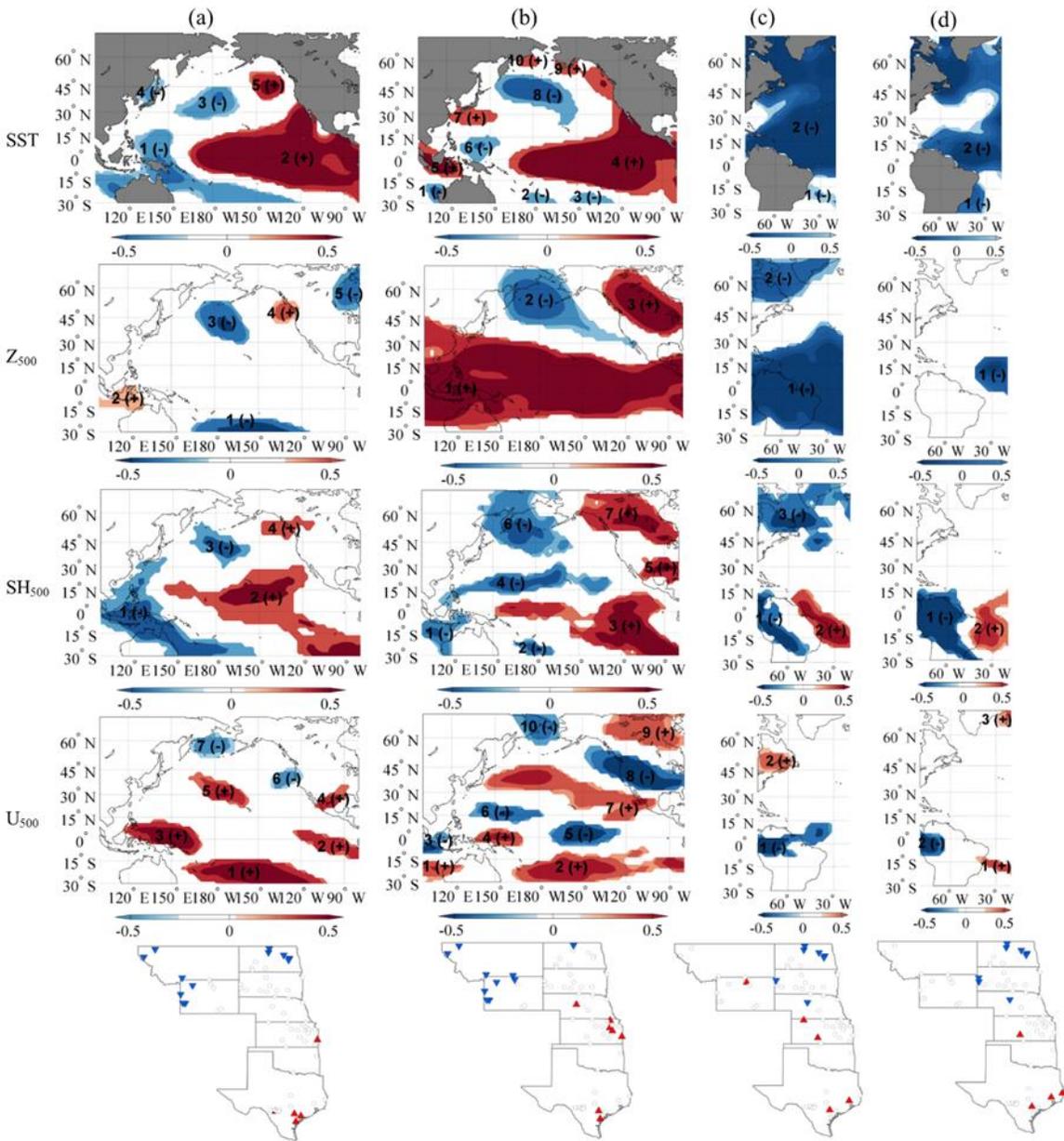


Figure 10: Heterogeneous correlation map developed for (a) Pacific September-November (b) Pacific December-February (c) Atlantic September-November (d) Atlantic December-February SST,  $Z_{500}$ ,  $SH_{500}$ , and  $U_{500}$  with Great Plains April-August streamflow. Significant regions with Positive (negative) correlations are represented by red (blue). Significant streamflow stations represented by red upward (blue downward) triangles.

Only one significant SST region was found throughout the Atlantic Ocean as shown in Figure 10c. The SST cooling phase had positive correlation in stations of North and South Dakota and vice-versa. Two significant  $Z_{500}$  regions were detected across South America and

mid-Atlantic regions while SH-3 and SH-1 regions were found significant and positively correlated with streamflow stations. The significant teleconnected U<sub>500</sub> region was a band like area across Northern part of South America while U-2 region above east coast of Canada showed negative correlation with the streamflow stations. The overall connection of these climate variables resulted in decrease of streamflow in Texas, Kansas and Wyoming while the increased streamflow was observed in Nebraska, South Dakota, and North Dakota.

Similar to 4-month lead-time the response to streamflow variation was alike for 1-month lead-time case. The spatial extent of the significant regions of each climate variable was found to be similar in this scenario shown in Figure 10d. The increasing (decreasing) of streamflow in northern stations of the Great Plains was correlated with increasing (decreasing) of SST, Z<sub>500</sub>, SH<sub>500</sub>, and U<sub>500</sub> regions represented by SST-2, Z-1, SH-1, and U-2 regions respectively. Smaller number of streamflow stations were found to be correlated with climate variables with similar signal which can be seen as SH-2 region was only significantly correlated with streamflow stations in Texas.

The warming (cooling) phase of SST in the ENSO region was found to be the most important phenomenon in affecting the streamflow variability in the Great Plains. The increasing trend of Pacific SST was linked to decreasing trend of streamflow in the northern parts of the Great Plains while increase of streamflow was observed for the southernmost regions. With increasing trend of Pacific SST, pressure, humidity and westerlies were also increased in most of the regions. However, Atlantic climate variables showed positive correlation with streamflow in the northern part and negative correlation in southern regions.

#### **3.4.1.2 Midwest**

SST-2 region was found to be positively correlated with streamflow stations in southern part of Midwest regions while fewer streamflow stations in Michigan showed negative correlation with the SST region shown in Figure 11a. The cooling SST phase was linked to

decrease of pressure throughout the Pacific Ocean as shown in Z-1 region. A similar drop of specific humidity region was observed near to the west coast of the United States. However, two specific humidity regions showed opposite correlation with most of the streamflow stations. A band of U-wind region was identified along the mid equatorial Pacific region which also showed positive correlation with most of the streamflow stations in the Midwest region.

Opposite to the 4-month lead-time case, the 1-month lead-time case identified SST regions which had negative correlation with the streamflow in the Midwest region as shown in Figure 11b. Majority of significant  $Z_{500}$  region showed positive correlation with the streamflow of the southern stations while Z-2 region which was in between Russia and Alaska showed positive signal for northern stations. Both positive and negative signals were present for the specific humidity regions which was also reflected on streamflow stations showing both negative and positive correlation. U-7 region showed opposite signal to SST region while U-6 region showed same signal for SST region. These U-wind regions could be the driving regions for the variability of streamflow in the Midwest region. The cooling SST phase was linked to increase of streamflow in the southern part while it was linked to decrease of streamflow in the northern part of the region.

The significant SST regions in the Atlantic region was found near to east coast of Canada and north east coast of Brazil as shown in Figure 11c. These regions were positively correlated with streamflow of Midwest except the stations in Michigan. The increasing phase of SST was also linked to increase of pressure throughout the identified  $Z_{500}$  regions. Furthermore, majority of the  $SH_{500}$  regions also showed positive correlation with streamflow. Smaller pockets of  $U_{500}$  regions were found to be positively correlated to streamflow in the southern stations while northern stations showed negative correlation.

Both 4-month and 1-month lead-time case had shown SST significant regions that were correlated to streamflow stations in a similar way as seen from Figure 11d. Spatial extent and position of the significant regions for other climate variables also looked identical for these two lead-time cases. The cooling SST phase was accompanied by the decrease of pressure in Z-1 region which in turn linked to decrease of specific humidity in SH-3 and SH-4 regions. However, U<sub>500</sub> region (U-2) showed opposite signal with SST, Z<sub>500</sub> and SH<sub>500</sub> regions. This resulted in decrease in streamflow in the southern most stations while there was increase of streamflow in the northern stations of the Midwest.

Similar to Great Plains, Midwest regions showed both positive and negative correlation with streamflow within the same region. The Pacific equatorial band was the primary significant region which showed strong association with Midwest streamflow. The cooling SST phase was linked to decrease of streamflow in most of the stations except the northern most stations. These northernmost stations showed decreased streamflow with increase of SST in the Atlantic Ocean. The increased precipitation over great lakes area when SST was below average could be attributed to polar jet stream. The storm track shifted towards northern region and over the great lakes region and caused increased precipitation.

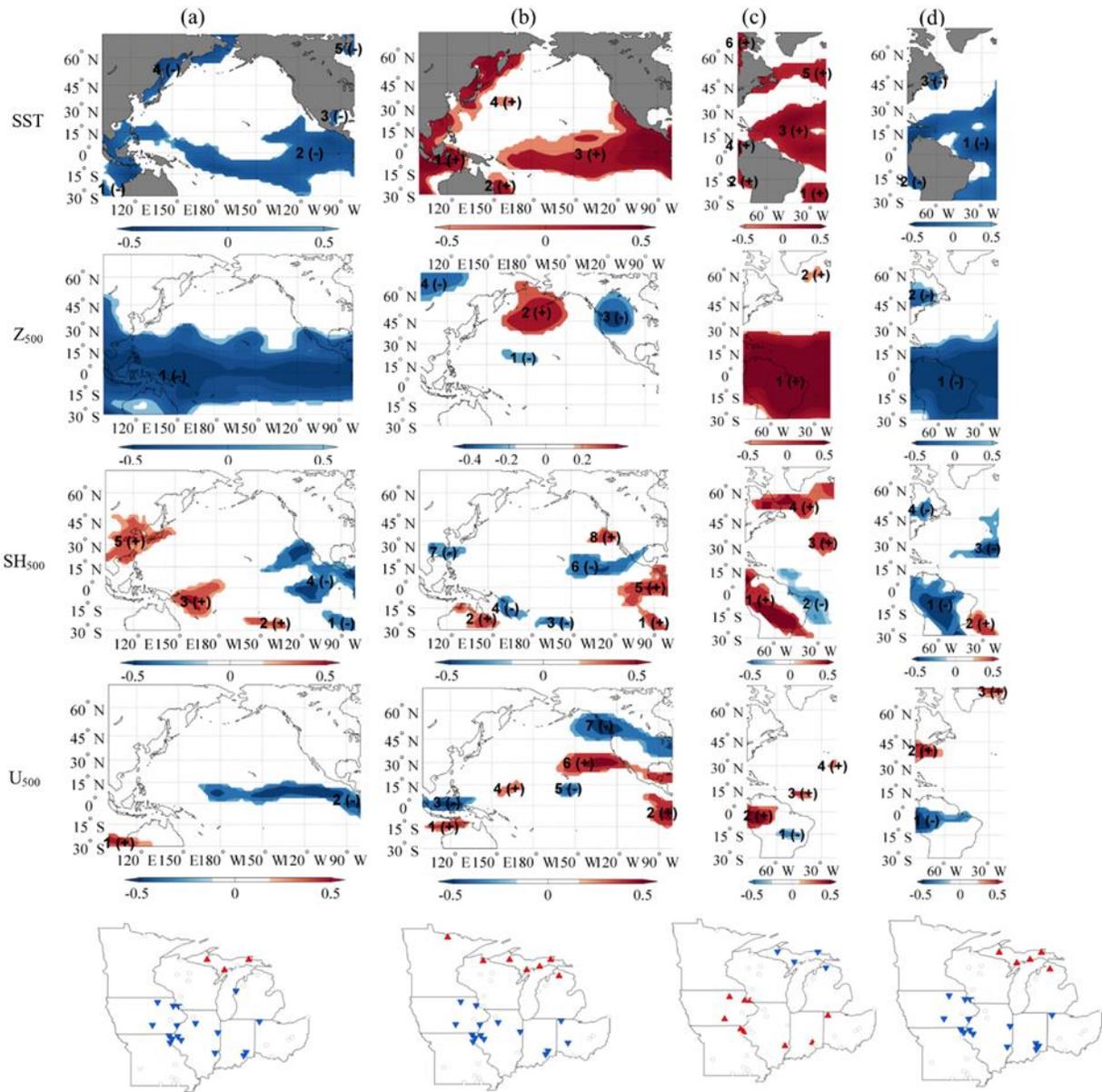


Figure 11: Heterogeneous correlation map developed for (a) Pacific September-November (b) Pacific December-February (c) Atlantic September-November (d) Atlantic December-February SST,  $Z_{500}$ ,  $SH_{500}$ , and  $U_{500}$  with Midwest April-August streamflow. Significant regions with Positive (negative) correlations are represented by red (blue). Significant streamflow stations represented by red upward (blue downward) triangles.

### 3.4.1.3 Northeast

The SST significant regions were sparsely located around south eastern Asian region as seen from Figure 12a. The decrease in SST resulted in reduced streamflow throughout the

Northeast region. The decreased SST region was related with decrease in atmospheric pressure in Z-1 along with increase in specific humidity represented as SH-2 and SH-3 with both increasing and decreasing of wind stress in different region. Only SST and Z<sub>500</sub> regions had strong positive correlation with streamflow in Northeast region. Mixed signals were observed for both SH<sub>500</sub> and U<sub>500</sub> regions.

The generated significant regions and the respective signal of correlation with the streamflow stations remained the same as compared with 4-month lead-time case. The identified SST significant regions were found in east coast of Korean Peninsula and Indonesia as shown in Figure 12b. The increased SST in those regions was associated with increase in pressure zones above north Mid-North Atlantic Ocean shown as Z-2 region. Increase of both of SST and Z<sub>500</sub> was associated with decrease of SH<sub>500</sub> as in SH-3 while the wind stress over western US was found to increase too. The overall connection of these climate variables resulted in increase of streamflow all over the Northeast regions.

Based on the SCF and NSC value, Atlantic Ocean had pronounced effect on variability of streamflow in Northeast region as compared to the Pacific Ocean. However, fewer number of streamflow stations showed correlation with the climate variables. SST-1 region was found just above Northeast region while significant Z<sub>500</sub> region was found off the cost of eastern Canada as shown in Figure 12c. The decrease of SST was accompanied by decrease of pressure and specific humidity over the region as represented by Z-2 and SH-1 respectively. The signal for wind stress was found to be dispersed with both positive and negative correlation. The combined effect of the variables showed decrease of streamflow in the study area.

Figure 12d shows the similar correlation pattern as observed in September-November case. One significant SST region was identified which was located off the coast of South

America and decrease of SST was associated with decrease of pressure but the observed significant region had smaller spatial extent. Furthermore, the location of the reduced pressure area was found in south eastern coast of Brazil. The decrease of SST was related with decrease of specific humidity and wind stress near to the Northeast regions. The overall connection resulted in decrease of streamflow in the region.

Northeast region was little affected by the Pacific SST or  $Z_{500}$ . Specific humidity and U-wind showed stronger correlation compared to SST and  $Z_{500}$  of the Atlantic with the Northeast streamflow. The result was in accordance with Rajagopalan et al. (2000) where streamflow of the Northeast region had shown limited correlation with climate variability of the Pacific Ocean.

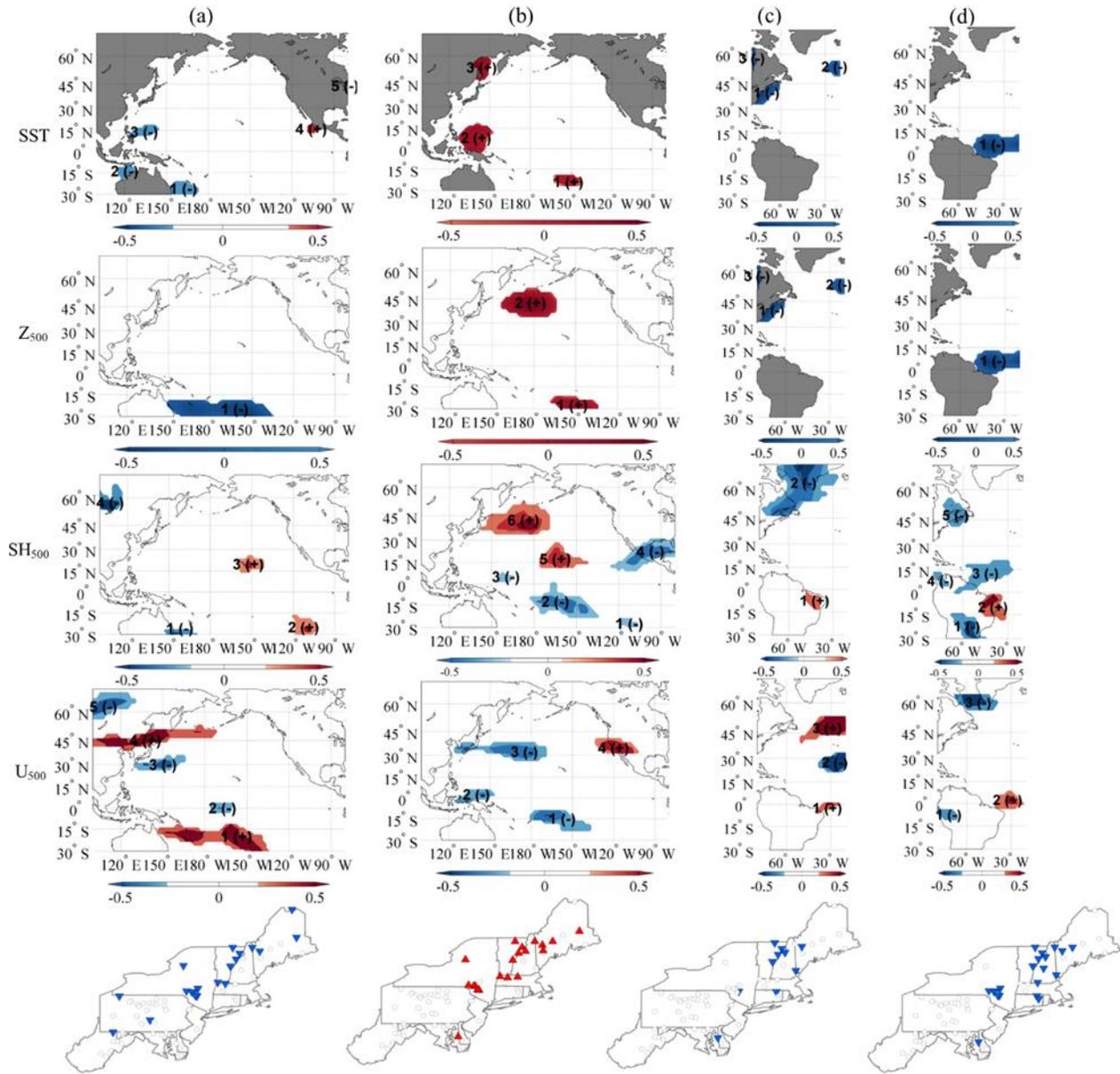


Figure 12: Heterogeneous correlation map developed for (a) Pacific September-November (b) Pacific December-February (c) Atlantic September-November (d) Atlantic December-February SST,  $Z_{500}$ ,  $SH_{500}$ , and  $U_{500}$  with Northeast April-August streamflow. Significant regions with Positive (negative) correlations are represented by red (blue). Significant streamflow stations represented by red upward (blue downward) triangles.

### 3.4.1.4 Northwest

It can be seen SST-2 region had the largest spatial extent occupying most of the equatorial mid Pacific region as shown in Figure 13a. The SST region was showing negative

correlation with streamflow stations in the Northwest region. The cooling SST phase was linked to decreasing atmospheric pressure while small pockets of increasing pressure zones were also present i.e., Z-3 and Z-5 regions. However, SH<sub>500</sub> significant regions were showing positive correlation with the streamflow. The wind stress over northwest region were also showing increasing trend while rest of the significant wind zones were showing decreasing trend. The overall connection of climate variables resulted in increase of streamflow in all the stations.

The SCF was same while NSC value was greater in 1-month lead-time case. Similar to 4-month lead-time case, SST-2 region had larger spatial extent and was negatively correlated with the streamflow in northwest region as seen from Figure 13b. The cooling SST phase was accompanied by the increased atmospheric pressure seen in Z-2 region while specific humidity in mid Pacific was also increasing (SH-3). It can be seen from U-6 region that the U-wind over northwest region had same signal as that of streamflow stations. The increase in specific humidity and wind might have favored the increased streamflow in the region.

It can be seen from Figure 13c that there were fewer number of SST significant regions in the Atlantic Ocean which had smaller spatial extent. SST-1 region was negatively correlated to streamflow stations in the region. The increased SST region was accompanied by increase of atmospheric pressure seen in Z-1 region. Sparse number of specific humidity and U-wind region were observed in the Atlantic region. The influence of Atlantic specific humidity and U-wind stress was minimum over streamflow variability within Northwest region.

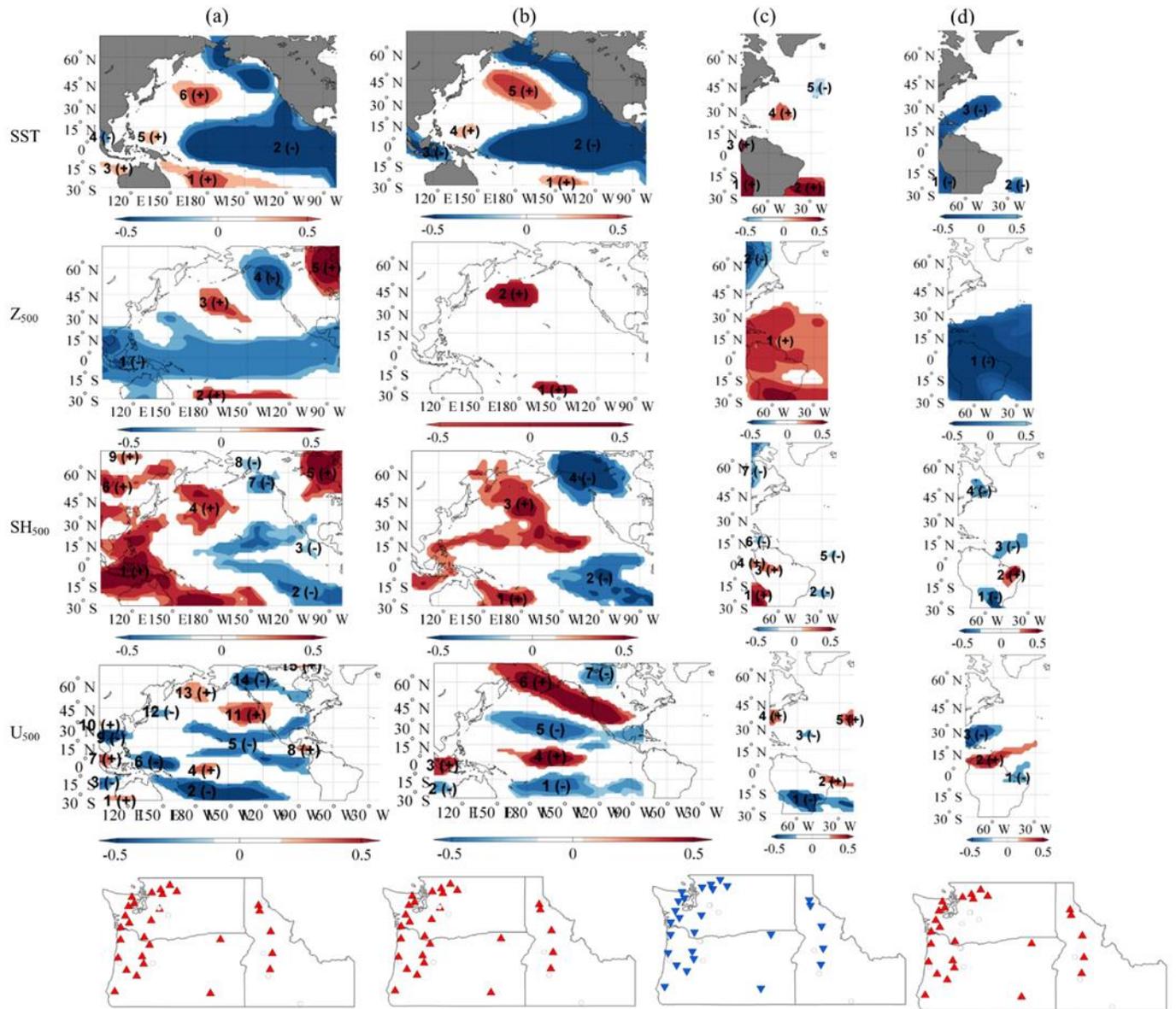


Figure 13: Heterogeneous correlation map developed for (a) Pacific September-November (b) Pacific December-February (c) Atlantic September-November (d) Atlantic December-February SST,  $Z_{500}$ ,  $SH_{500}$ , and  $U_{500}$  with Northwest April-August streamflow. Significant regions with Positive (negative) correlations are represented by red (blue). Significant streamflow stations represented by red upward (blue downward) triangles.

Similar to 4-month lead-time case, 1-month lead-time SST was negatively correlated to streamflow in the Northwest region. The cooling SST phase was linked with decreasing pressure zone ( $Z-1$ ). Similar to previous 4-month lead-time case, smaller number of significant regions of

specific humidity and U-wind were observed in 1-month lead-time case as shown in Figure 13d. The overall connection was increase of streamflow with decreasing SST in the Atlantic Ocean.

Several studies had shown direct influence of climate variability of western Pacific in the hydrologic responses of the Northwest (Rajagopalan et al., 2000; Aziz et al., 2010; Hunter et al., 2006). It is commonly observed in Northwest that cooling phase of ENSO causes increased snowfall. The Asian/Pacific jet stream carries moisture during winter and wind flow toward north western region is associated with increased snowfall leading to increased snowfall (Lamb et al., 2010). Redmond and Koch (1991) also observed increased precipitation in the Northwest region when SST in ENSO-like region was below average. The SCF and NSC value were also quite higher for the Northwest and the Southwest region as compared to other regions, which indicated that SST variability and consequent streamflow response was clear in those regions. The influence of the Atlantic Ocean was found to limited in streamflow variability of the Northwest region.

#### **3.4.1.5 Southeast**

Figure 14a indicates that different four pockets of significant regions were identified in eastern and mid Pacific Ocean. The increased SST was associated with majority of increased pressure zones. Significant regions of the specific humidity also depicted similar positive correlation with the streamflow. However, maximum number of negatively correlated wind stress regions were observed throughout the Pacific Ocean. The negatively correlated wind stress might have caused decrease in streamflow in the Southeast region. The spatial extent of  $U_{500}$  zones had also wider spatial extent compared to the regions of other climate parameter.

Figure 14b shows the SST significant regions had been shifted to different locations; one of which region (SST-2) showed similarity to ENSO. The ENSO like region had same signal

with the streamflow stations implying positive correlation between Southeast streamflow and Pacific equatorial SST region. The cooling SST was related with the increase of pressure zone (Z-1) and majority of specific humidity regions (SH-1, SH-2, and SH-5) also showed increase of humidity in the Pacific region.

Three SST significant regions were found near east coast of Canada, north coast of Brazil, and mid-Atlantic region. Sparse  $Z_{500}$  region was spotted in the Atlantic Ocean as seen from Figure 14c. The  $SH_{500}$  and  $U_{500}$  region were also in smaller spatial extent and the signal of all the climate parameters was same. However, the signal for streamflow station was opposite meaning cooling of SST was related with increase of streamflow in the region. The weak signal of  $Z_{500}$ ,  $SH_{500}$ , and  $U_{500}$  should be fully understood.

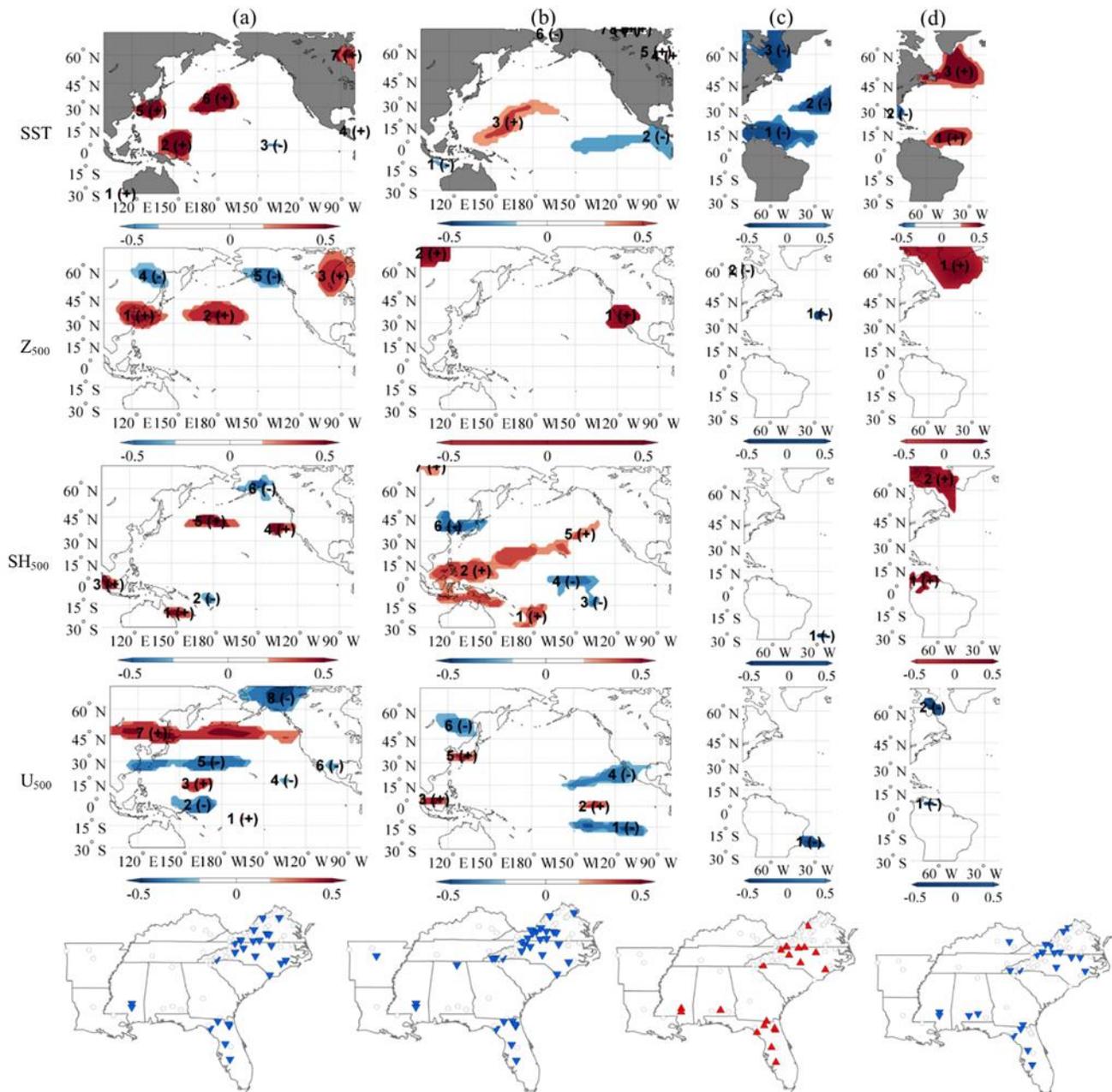


Figure 14: Heterogeneous correlation map developed for (a) Pacific September-November (b) Pacific December-February (c) Atlantic September-November (d) Atlantic December-February SST,  $Z_{500}$ ,  $SH_{500}$ , and  $U_{500}$  with Southeast April-August streamflow. Significant regions with Positive (negative) correlations are represented by red (blue). Significant streamflow stations represented by red upward (blue downward) triangles.

SST-3 region was found near to the east coast off Canada which had been associated with strong correlation of  $Z_{500}$  with same sign with SST as shown in Figure 14d. Significant region of

SH<sub>500</sub> were also showed similar signal while the signal of U<sub>500</sub> was opposite. When SST was to increase, then Z<sub>500</sub> and SH<sub>500</sub> was increasing while wind stress was diminishing making overall effect for streamflow to decrease. It should be noted that in all the four scenarios, the signal of wind stress was positively correlated with the streamflow.

Similar to Northeast region, the Southeast streamflow was more correlated with the U-wind and specific humidity compared to SST of the Pacific Ocean. Limited number of significant regions of SST and Z<sub>500</sub> were identified for the Southeast. Only northern Pacific region had shown teleconnection with streamflow variability, similar observation from Wang and Ting (2000) was observed. Atlantic SST and Z<sub>500</sub> were negatively correlated with the streamflow. Different pressure gradients which drive the circulation were the primary factors that affect streamflow variability in the region.

#### **3.4.1.6 Southwest**

It can be noted that the SCF value was quite higher than the other SVD regions. From Figure 15a it can be seen that the SST significant region (SST-2) was similar to ENSO region. SST-1 region was similar to the Hondo region as identified by previous researchers (Sagarika et al., 2015; Zhang et al., 1997). This particular region which had been found to influence the hydrology of western U.S. regions provided important streamflow predicting ability. The heterogeneous correlation map showed the increase of the SST was related with decrease of pressure in western US seen in Z-2 region. However, magnitude of specific humidity and U-wind were also increasing. These connections influenced to increase the streamflow in Southwest regions. SST, U<sub>500</sub> and SH<sub>500</sub> regions were dominantly correlated with streamflow in the region in positive signal while Z<sub>500</sub> regions were negatively correlated.

The SCF and NSC values were slightly greater for 1-month lead-time compared to 4-month lead-time period. SST regions for this case was also found identical in comparison to previous longer lead-time case even though the signal was reversed as shown in Figure 15b. The decrease of SST region was linked with increase of pressure above western coast of continental U.S. as seen in Z-2 region while Z-1 and Z-3 regions showed opposite signal i.e., pressure was decreasing. Decrease in SST and  $Z_{500}$  tend to decrease humidity over SH-2 and SH-4 regions while SH-3 region showed increasing trend. Similarly, equal number of positively and negatively correlated  $U_{500}$  regions were seen. The overall linkage of the climate variables tend to decrease streamflow in the southwest region. Compared to 4-month lead-time case, 1-month lead-time case had shown slightly good streamflow variability while the correlation shown by the significant regions did not change much.

SST-2 region was the dominant teleconnected region which showed negative correlation with the streamflow variability and was linked to Z-1 region which also showed same signal as shown in Figure 15c. Dominant number of positively correlated SH<sub>500</sub> and  $U_{500}$  regions were found in the Atlantic region by which influence streamflow was found to be increased when SST was reduced. It can be noted that cooling of Atlantic SST was favoring more streamflow in the region unlike cooling Pacific SST which was shown to favor less streamflow in southwest region.

As the lead-time decreased more number of significantly correlated regions were identified. Furthermore, the spatial extent of these regions were larger than that of longer lead-time case. From Figure 15d, it can be seen that three SST regions were positively correlated with streamflow in California while negatively correlated with streamflow in New Mexico. The cooling SST phase was accompanied by increasing pressure in Z-1 and Z-2 region. The

interconnection led to decrease of specific humidity over Atlantic Ocean and increase of U-wind in U-2 while decrease in U-1 region. The overall connection resulted increased of streamflow in California and decrease of streamflow in New Mexico.

Streamflow variability in the Southwest region was highly correlated to SST of ENSO region. The warmer phase of SST was accompanied by increase of humidity and U-wind in west Pacific region with increased streamflow throughout the region. The direct influence of Pacific anomalies in the streamflow of the Southwest was possibly due to close proximity of the region from the ocean-atmospheric activities. Previous researches had shown consistent result in which El Nino phase had brought increased precipitation over the southwest region (Grantz et al., 2005; Pathak et al., 2017). Increase in SST led to increased surface atmospheric pressure and strong wind leading to increased precipitation. Atlantic variables showed minimal effects in explaining the streamflow variability of Southwest.

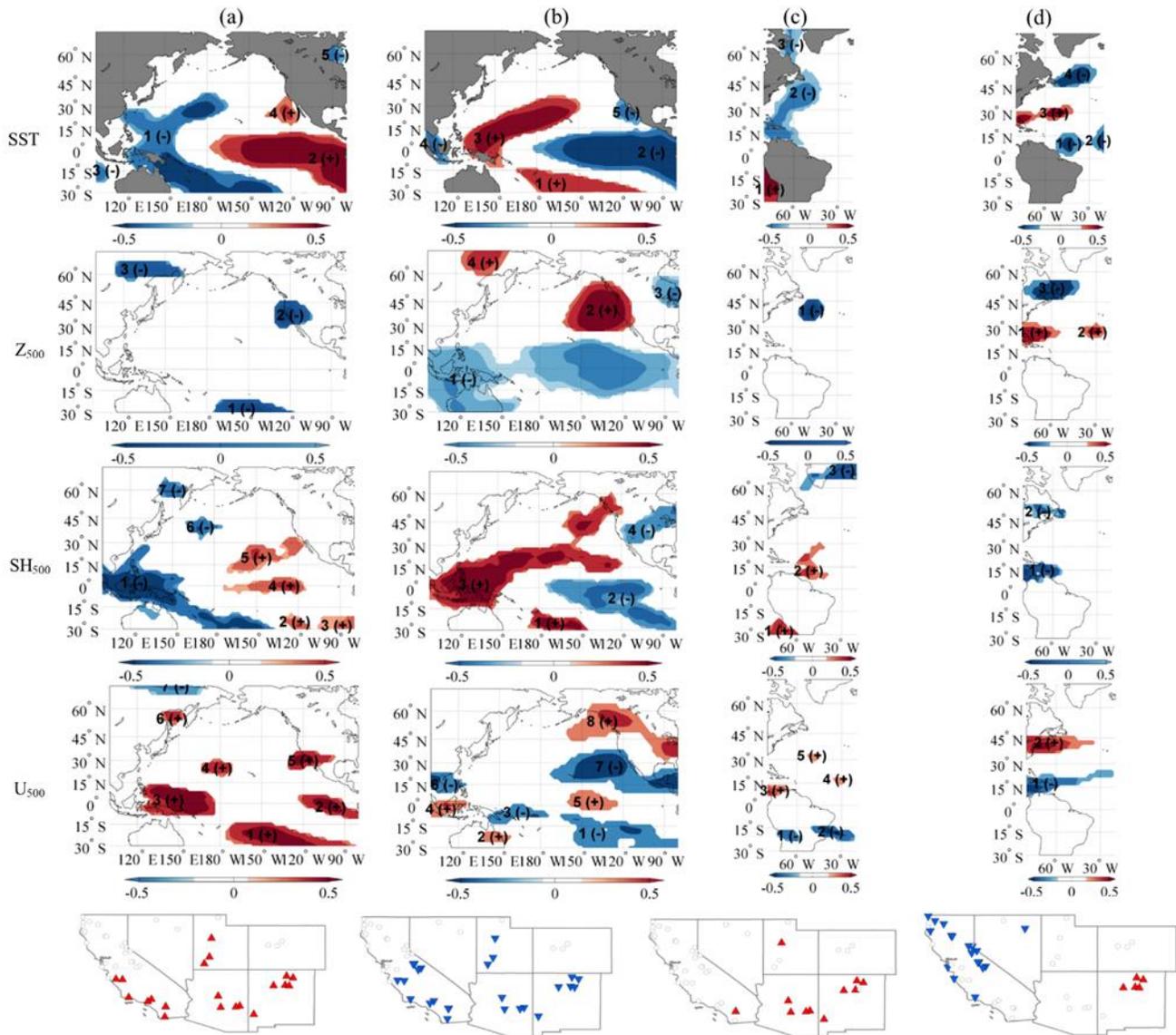


Figure 15: Heterogeneous correlation map developed for (a) Pacific September-November (b) Pacific December-February (c) Atlantic September-November (d) Atlantic December-February SST,  $Z_{500}$ ,  $SH_{500}$ , and  $U_{500}$  with Southwest April-August streamflow. Significant regions with Positive (negative) correlations are represented by red (blue). Significant streamflow stations represented by red upward (blue downward) triangles.

### 3.4.2 CWT analysis

The variability of data explained by the CWTs of ENSO, PDO, and AMO are presented in Figure 16. Figure 17 contains the CWTs of each of the regional streamflow patterns, where the FPC obtained from the PCA represents the regional streamflow. The results revealed the inherent

nature of variance and periodicity across multiple timescale bands over the study period in each of the indices and among the regional streamflow patterns. The association between the regional streamflow and the ENSO, PDO, and AMO indices obtained through the WTC of the involved CWTs are provided in Figures 18, 19, and 20, respectively. The wavelet coherency, scaled from 0 to 1, a measure of correlation, was calculated based on Torrence and Webster (1999).

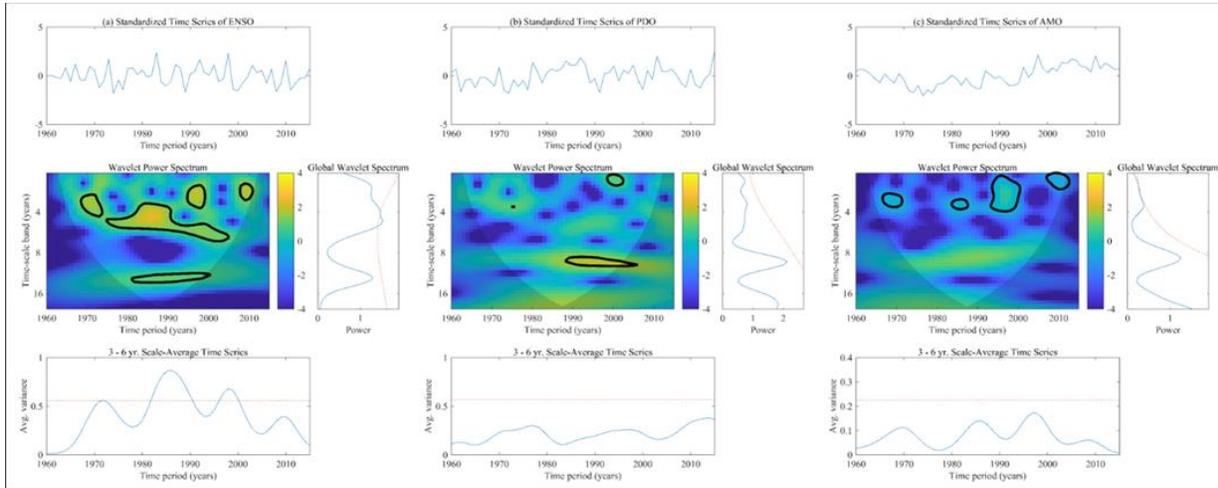


Figure 16: Continuous wavelet transforms (CWT) along with the global wavelet spectrum and the 3-to-6-years scale average time series of (a) ENSO, (b) PDO, and (c) AMO. Warmer (yellow) color represents higher variance in data. Zones with significant variance against 5% red noise are delineated by black contour lines.

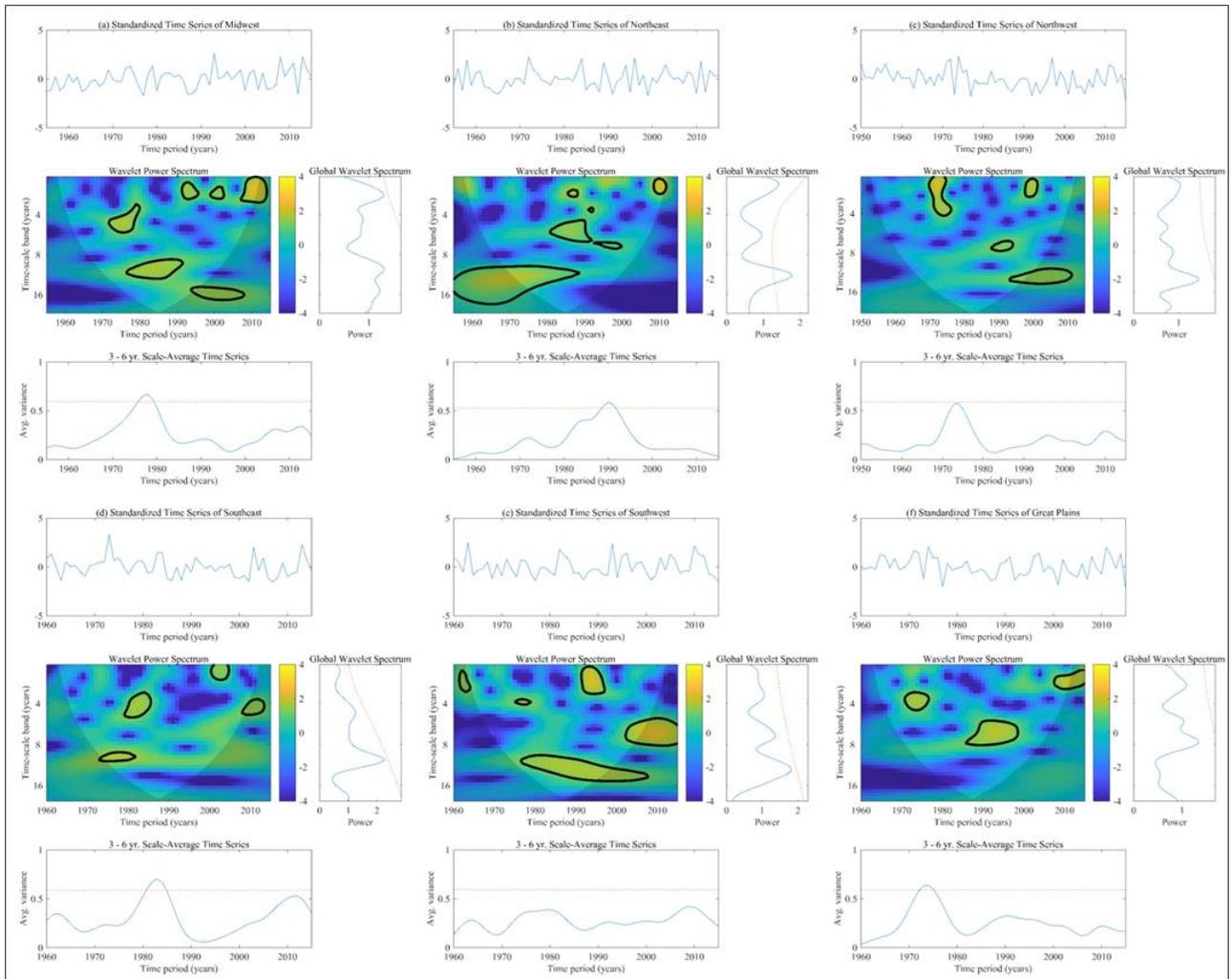


Figure 17: Continuous wavelet transforms (CWT) along with the global wavelet spectrum and the 3-to-6-years scale average time series of (a) Midwest, (b) Northeast, (c) Northwest, (d) Southeast, (e) Southwest, and (f) Great Plains streamflow. Warmer (yellow) color represents higher variance in data. Zones with significant variance against 5% red noise are delineated by black contour lines

### 3.4.2.1 Correlation between ENSO and Regional Streamflow

Figure 18a shows the WTC between ENSO and the Midwest streamflow, which illustrates that these two time series were highly correlated beyond the 16-year band from 1955 to 1990 and below the 4-year band from 1990 to 2015. The majority of arrows indicated an anti-phase relationship between ENSO and the Midwest streamflow by pointing to the left. A tendency of arrows pointing upwards was also noticeable, especially at higher bands. Arrows

pointing upwards indicate a lag of  $90^\circ$  (one quarter of a cycle) between ENSO and the streamflow. The correlation between ENSO and the Northeast streamflow (Figure 18b) was found to be higher from 1995 to 2015 in the 8-to-12-year band and from 1980 to 1990 around the 4-year band. The arrows were observed to be pointing in the opposite directions in the two significant zones of higher correlation, which indicated that the relative phase relationship between ENSO and the Northeast streamflow were not uniform across the study period.

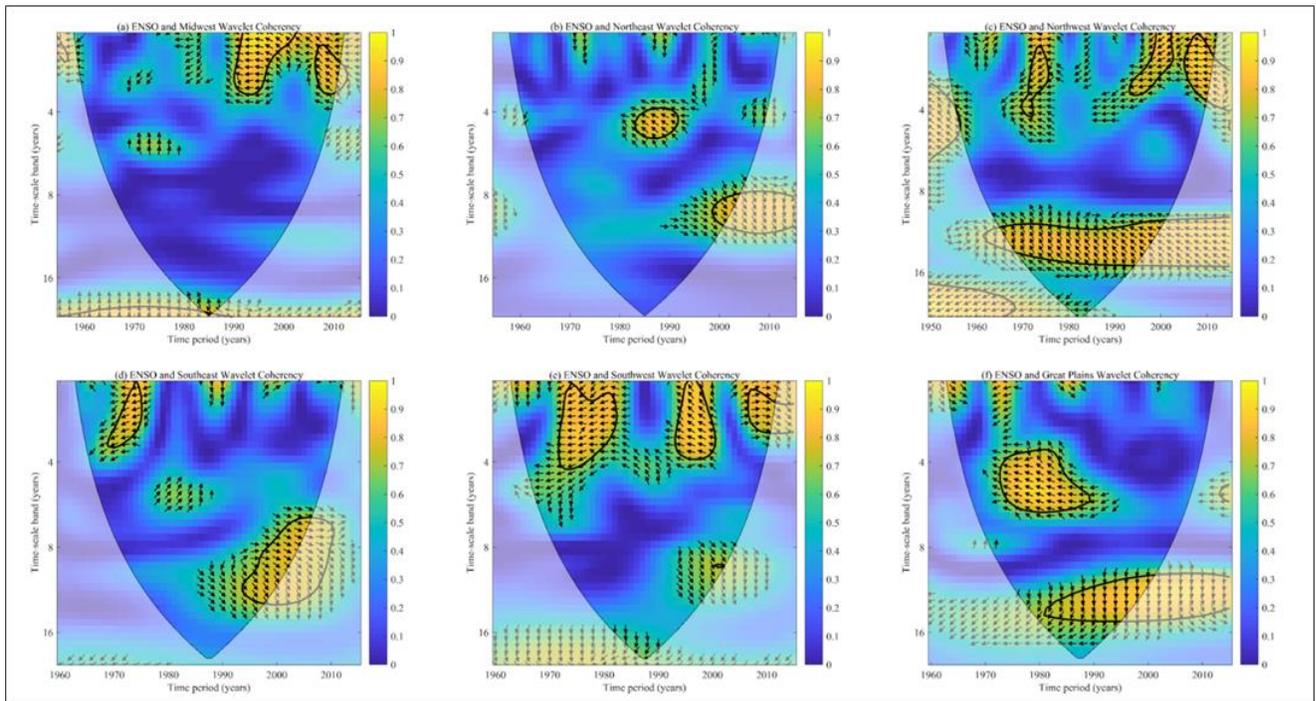


Figure 18: Wavelet coherence (WTC) spectrums between ENSO and the regional streamflow of (a) Midwest, (b) Northeast, (c) Northwest, (d) Southeast, (e) Southwest, and (f) Great Plains. Warmer (yellow) color represents higher correlation between the two time series. The arrows indicate the relative phase relationship within the significant zones of higher correlation.

In comparison to any other regions, ENSO was found to be highly correlated with the Northwest streamflow (Figure 18c). Zones of significant correlation were found from 1950 to 1970 beyond the 16-year band, from 1960 to 2015 in the 8-to-16-year band, and at multiple occasions around and below the 4-year band across the study period. The arrows were pointing towards the left showing an anti-phase relationship between ENSO and the Northwest

streamflow. The WTC of ENSO and the Southeast streamflow (Figure 18d) showed higher correlation from 1990 to 2010 in the 6-to-16-year band and at a few occasions of short durations below the 4-year band. The arrows in significant zones were observed to be pointing in different directions even in the same zone of higher correlation, which suggests that the phase relationships of ENSO and the Southeast streamflow were not uniform across the study period.

ENSO was strongly correlated to the Southwest streamflow (Figure 18e) in the lower timescale bands, i.e., the bands below 4-year, from 1970 to 1980, from 1990 to 2000, and from 2005 to 2015. The arrows representing relative phase relationship were found to be pointing at different directions across the study period indicating a non-uniform phase relationship between ENSO and the Southwest streamflow. ENSO and the Great Plains streamflow (Figure 18f) showed strong correlation in the 8-to-16-year band from 1980 to 2015 and in the 4-to-8-year band from 1970 to 1990. Higher correlation in the lower timescale bands were observed in few cases. The arrows representing the relative phase relationship were not found to be uniform across the different timescale bands, though their directions were uniform within a significant zone.

#### **3.4.2.2 Correlation between PDO and Regional Streamflow**

The WTC between PDO and the Midwest streamflow (Figure 19a) showed higher correlation only around or below the 4-year band. The significant zones with higher correlation were found from 1955 to 1985, from 2000 to 2005, and from 2010 to 2015. The relative phase relationship before 1990 showed an anti-phase relationship between PDO and the Midwest streamflow. From 2000 and onwards, the Midwest streamflow lagged PDO by  $90^\circ$  (one quarter of a cycle). PDO showed higher correlation with the Northeast streamflow (Figure 19b) at intermittent intervals around and below the 4-year timescale, especially from 1990 to 2015.

Higher correlation was also observed from 2000 to 2015 in the 10-to-12-year band. Bands beyond the 16-year timescale also showed higher correlation but were not statistically significant. The arrows did not show any uniform pattern across the study period. However, the arrows beyond the 16-year band were mostly pointing to the right showing an in-phase association of PDO with the Northeast streamflow

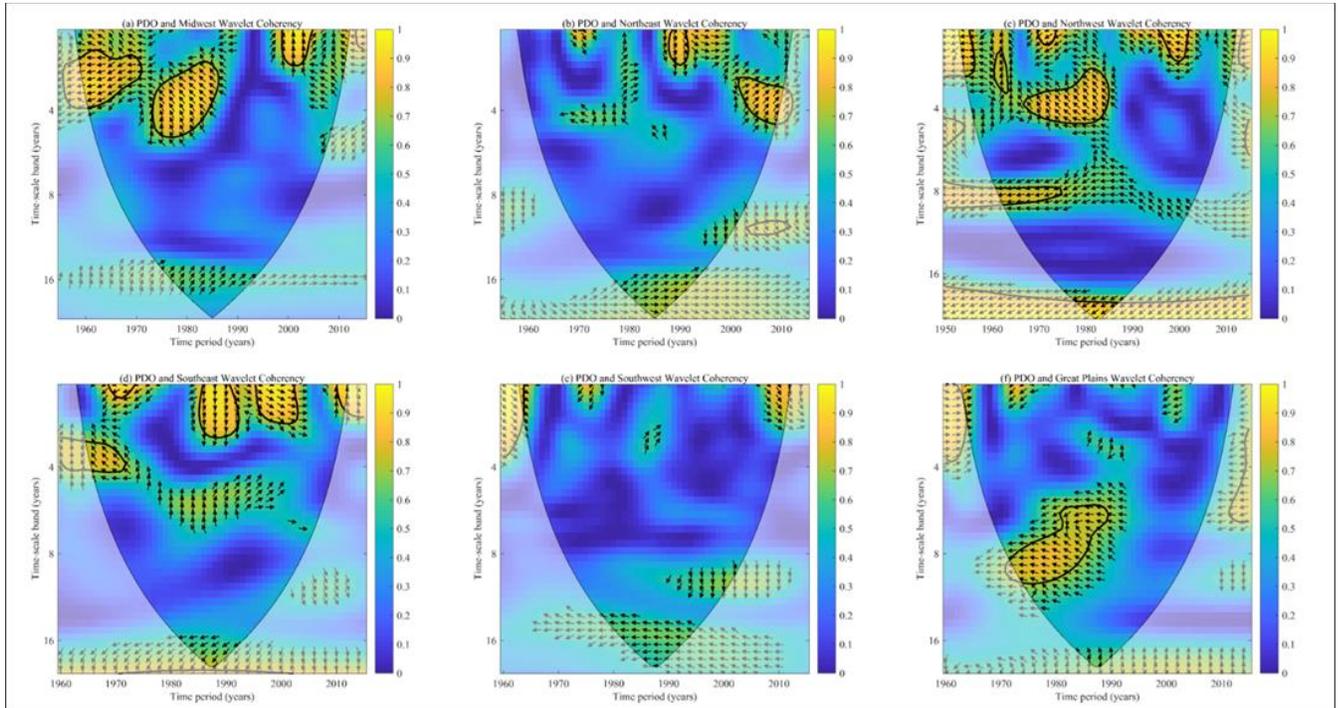


Figure 19: Wavelet coherency (WTC) spectrums between PDO and the regional streamflow of (a) Midwest, (b) Northeast, (c) Northwest, (d) Southeast, (e) Southwest, and (f) Great Plains. Warmer (yellow) color represents higher correlation between the two time series. The arrows indicate the relative phase relationship within the significant zones of higher correlation.

PDO was strongly correlated to Northwest streamflow (Figure 19c) in comparison to other regions. Higher correlations were observed across each timescale bands at separate duration of time. A continuous zone of significant correlation expanding throughout the study period (from 1950 to 2015) was observed beyond the 16-year band. Higher correlation was also observed around the 8-year timescale from 1950 to 1980 and at multiple occasions around and

below the 4-year band. Higher correlations at lower timescales were more dominant before 1980s. The relative phase relationship suggested an anti-phase relationship in most of the significant zones. PDO was higher correlated to the Southeast streamflow (Figure 19d) at timescales below the 4-year band at multiple intervals, especially after 1985, and at timescale beyond the 16-year band from 1970 to 2000. The arrows did not show any uniform pattern across the study period as they were found to be pointing in different directions even within the same timescale band.

PDO did not show much correlation with the Southwest streamflow (Figure 19e) across the study period. Higher correlations were only observed from 1960 to 1965 and from 2005 to 2015 below the 4-year timescale. A lagged in-phase relationship between PDO and the Southwest streamflow in the significant zones of higher correlation was observed. PDO showed higher correlation with the Great Plains streamflow (Figure 19f) below the 4-year timescale from 1960 to 1965, in the 6-to-10-year band from 1970 to 1990, and in the 3-to-7-year band from 2010 to 2015. The arrows were pointing to different directions across the timescale bands with significant zones with higher correlation, which indicated that the relative phase relationship between PDO and the Great Plains streamflow were not uniform for the analyzed duration. For all regions, bands around and beyond the 16-year timescale indicated higher correlation with PDO. However, the bands were not found to be statistically significant. Even though certain regions showed significant correlation beyond the 16-year band, e.g., the Northwest region, the wavelet spectrums implied presence of higher correlation even at higher bands (possibly in the 16-to-32-year band).

### 3.4.2.3 Correlation between AMO and Regional Streamflow

AMO showed higher correlation with the Midwest streamflow from 1980 to 2000 and from 2010 to 2015 around and below the 4-year band (Figure 20a). No presence of higher correlation was observed in the higher timescales. An anti-phase relationship was found between AMO and the Midwest streamflow in the zones with significant correlation. The WTC between AMO and the Northeast streamflow showed presence of higher correlation from 2000 to 2015 in the 6-to-10-year band and below and around the 4-year band at intermittent intervals after 1980s to 2015 (Figure 20b). The relative phase relationships were found to be pointing at opposite directions in the two major zones of significant correlation, however, the arrows within a significant zone uniform in direction.

Significant correlation between AMO and the Northwest streamflow were observed from 1950 to 1970 in the 8-to-12-year band and from 2010 to 2015 around and below the 4-year band (Figure 20c). There was also presence of significant correlation at intermittent intervals of shorter durations from 1970 to 2010 in the 4-to-8-year band. Presence of high correlation was also observed beyond the 16-year band. An anti-phase relationship between AMO and the Northwest streamflow was observed mostly. Among the regions, AMO showed the highest correlation with the Southeast streamflow (Figure 20d). Significant correlation was identified throughout the period, i.e., from 1960 to 2015 beyond the 16-year band. Strong correlation was also observed from 1960 to 1970 and from 2000 to 2015 in the 6-to-8-year band and at multiple intervals of shorter duration below the 4-year band. Non-uniform phase relationship was observed in the significant zones across the study period, however, the arrows in each significant zone were found to be pointing in the same direction.

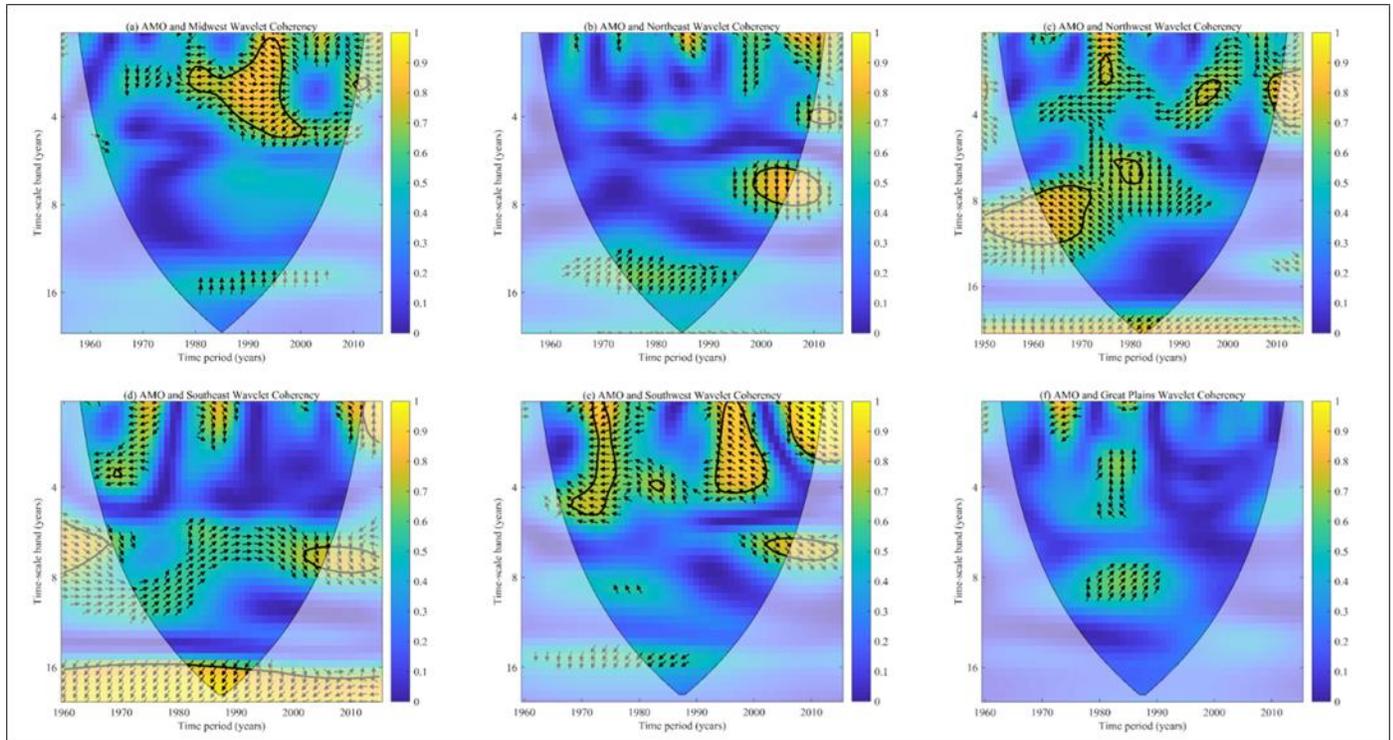


Figure 20: Wavelet coherence (WTC) spectrums between AMO and the regional streamflow of (a) Midwest, (b) Northeast, (c) Northwest, (d) Southeast, (e) Southwest, and (f) Great Plains. Warmer (yellow) color represents higher correlation between the two time series. The arrows indicate the relative phase relationship within the significant zones of higher correlation.

The WTC between AMO and the Southwest streamflow showed higher correlation only at the lower timescale bands, i.e., around and below the 4-year band and in the 6-to-8-year band. The intervals of with higher correlation were from 1970 to 1980 and from 1990 to 2015. No presence of higher correlation was observed in the higher timescale band. The arrows indicating the relative phase relationships were found to be pointing in different direction in the various significant zones indicating a non-uniform phase relationship between AMO and the Southwest streamflow across the study period. AMO did not show much correlation with the Great Plains streamflow across the study period. Among the regions, Great Plains was the only region that did not show any zone of significance across the study period. There were a few instances of slightly

high correlation before 1990s below and around the 4-year and 8-year band, but none of them were found to be significant.

### 3.5 Conclusions

The interconnection of climate variabilities and the regional streamflow of the continental US were analyzed in this study. SVD analyses were implemented on April-August streamflow of six NCA regions with September-November and December-February climate variabilities represented by SST,  $Z_{500}$ ,  $SH_{500}$ , and  $U_{500}$  of the Pacific and Atlantic Oceans. The WTC analyses of ENSO/PDO/AMO and the regional streamflow patterns revealed the most significant timescale bands that affected their variation over the study period. The major findings of the analyses are given below:

1. The warming (cooling) phase of SST in the ENSO like region was found to be the most important phenomenon in affecting the streamflow variability in the Great Plains, Midwest, and Southwest region
2. The streamflow variability was found to be different in Northwest region. The cooling SST in the ENSO like region showed increasing trend of streamflow in the Northwest region, unlike other regions.
3. SST,  $SH_{500}$  and U-wind were primarily responsible for streamflow variation in the Northeast and Southeast region.
4. The correlation between ENSO and the regional streamflow revealed significant variations among the regions. The Northwest region was highly correlated in comparison to any other regions.
5. PDO revealed unique associations with the regional streamflow. Among the regions, the Northwest showed the highest correlation with PDO across multiple timescale bands.

6. Among the regions, Southeast streamflow was found to be highly correlated with AMO.

The SVD analyses identified several new teleconnected regions of SST,  $Z_{500}$ ,  $SH_{500}$ , and  $U_{500}$  in the Pacific and Atlantic Oceans, which may provide some explanations of the hydrological variations observed in the United States. The information from identified significant regions of climate variables can be used for long lead-time streamflow forecasting which in turn can be helpful for water managers in effectively allocating water to competing users in regional scale. Additionally, the influence of  $SH_{500}$  and  $U_{500}$  on streamflow was found to be prominent throughout the regions. Though the data used in this study have hardly been utilized in previous studies, the results suggested that the correlations between  $SH_{500}/Z_{500}$  and streamflow were equally as higher as that of SST and  $Z_{500}$ . The wavelet approach evaluated the correlation of regional streamflow with three major predefined indices together, such approach has received little attention in the previous research efforts. Unlike the association with ENSO, which had relatively stronger correlation with majority of the regional streamflow time series after 1990, PDO did not show any such pattern. However, AMO did not show much correlation beyond higher timescale bands. Future works should look into evaluating the association between streamflow and PDO at higher timescale bands, since PDO showed a decadal to multi-decadal oscillatory cycle in the 20th century (Trenberth and Fasullo, 2007). Use of longer period of data may provide wider scope of association of predefined indices with streamflow in higher timescale bands.

## CHAPTER 4

### CONTRIBUTIONS AND RECOMMENDATIONS

#### 4.1 Summary

The knowledge of climate variability and change on the regional streamflow can provide valuable information for addressing various water resource management issues. The driving mechanism of hydrological cycle is complex and uncertain. Full comprehension involving physical reasoning can be cumbersome and expensive, a simple approach using statistical methods can be helpful in understanding the idea of hydrologic variability. The current work investigated the relationship between the regional streamflow of the United States and hydro-climatological variables along with predefined indices. The aforementioned research questions are addressed in this study.

Research question #1: How does the proposed modeling framework improve the lead-time of the streamflow forecast? The model used SVD techniques to find spatial-temporal correlations between climate variables and streamflow of the basin and the best correlations are screened through non-parametric test and finally taken as input to predict streamflow by SVM method. The framework showed superior forecasting ability as several model performance parameter gave higher values.

Research question #2: How is the streamflow within a basin associated with ocean-atmospheric climate variability? To better represent the climate variability, wide range of hydro-climatological variables were studied and analyzed by SVD. The strong correlation of U-wind and specific humidity along with geopotential height and SST with streamflow showed important predictive ability of these variables. Furthermore, wavelet approach clearly showed the effect of ENSO, AMO and PDO on streamflow of majority of the regions.

## 4.2 Contributions

The research proposed a novel modeling framework to forecast streamflow forecasting and to evaluate the response of streamflow due to climate variability and change. Regional analysis of streamflow with respect to climatic indices are also performed. The major contributions can be shown below:

1. The study has used an extended set of climate data i.e.,  $U_{500}$  and  $SH_{500}$  in addition to SST and  $Z_{500}$  data over the entire Pacific and Atlantic Ocean. The inclusion of these data led to identify associated significant regions for the six NCA regions and showed equally competent potential for explaining streamflow variability of the region. These climate data have received little attention in previous research efforts.
2. The modeling framework of forecasting research is new: the utilization of SVD temporal expansion series of climate variables and employing non-parametric approach for screening the variables used as input for SVM model to forecast streamflow is a novel approach in the field of hydrology.
3. The proposed model forecasted streamflow volumes up to 13 months in advance with excellent forecasting ability at each lead-time scenario.
4. The regional SVD analysis presented some ideas about interconnection between climate variables. Previous SVD studies had not considered the combined effect of climate variables together.

## 4.3 Limitations

The study tried to investigate the relationship between climate change and streamflow pattern in the United States. The study considered six regions to investigate the relationship

while the results showed some streamflow variability within the region. The streamflow variability could be better analyzed if more number of basins were considered. Further, this research used limited number of climate variables. Inclusion of several other variables together could improve the results. The result may be affected by several uncertainties involved in the process due to complex nature of hydrological cycle driven by climate variability and change.

#### **4.4 Recommendations for future work**

Future study can consider several climate variables together to increase the scope of the research. Future work, may explore extended lead-time scenarios. Additionally, the application of paleo data may provide promising results as data-driven models show higher efficiency for wide range of input data. Physical model should also be given high priorities. Moreover, unlike the association with ENSO, which had relatively stronger correlation with majority of the regional streamflow time series after 1990, PDO did not show any such pattern. Future works should look into evaluating the association between streamflow and PDO at higher timescale bands.

## REFERENCES

- Abdi, H., 2007. Singular value decomposition (SVD) and generalized singular value decomposition. *Encyclopedia of Measurement and Statistics*. Thousand Oaks (CA): Sage, 907-912.
- Ahmad, S., A. Kalra, and H. Stephen, 2010. Estimating soil moisture using remote sensing data: A machine learning approach. *Advances in Water Resources* 33(1): 69-80.
- Astuti, W., R. Akmeliawati, W. Sediono, and M.J.E.Salami, 2014. Hybrid technique using singular value decomposition (SVD) and support vector machine (SVM) approach for earthquake prediction. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(5), 1719-1728.
- Babovic, V., M. Keijzer, and M. Bundzel, 2000. From global to local modelling: a case study in error correction of deterministic models. In *Proceedings of Hydroinformatics 2000*, Vol. 4, No. 5.
- Barlow, M., S. Nigam, and E.H. Berbery, 2001. ENSO, Pacific decadal variability, and US summertime precipitation, drought, and stream flow. *Journal of climate*, 14(9), 2105-2128.
- Bartolino, J.R., and J.C. Cole, 2002. Ground-water resources of the middle Rio Grande basin, New Mexico (No. 1222). US Geological survey.
- Bhandari, S., A. Kalra, K. Tamaddun, and S. Ahmad, 2018. Relationship between ocean-atmospheric climate variables and regional streamflow of the conterminous United States. *Hydrology*, 5(2), 30 (DOI: 10.3390/hydrology5020030)

- Bhandari, S., A. Jobe, B. Thakur, A. Kalra, S. Ahmad, 2018. Flood Damage Reduction in Urban Areas with Use of Low Impact Development Designs. *World Environmental and Water Resources Congress 2018 (pp. 52-61)*.
- Bhandari, R., R. Thakali, G. A. D. Kandissounon, A. Kalra and S. Ahmad, 2017. Effects of Soil Data Resolution on the Simulated Stream Flow and Water Quality: Application of Watershed-Based SWAT Model. *World Environmental and Water Resources Congress 2018*.
- Bhandari, R., R. Parajuli, G. R. Lamichhane, A. Alyami, A. Kalra, S. Ahmad, and R. Gupta, 2018. Utilizing Civil Geo-HECRAS Capabilities for Floodplain Mapping of Colorado River in Texas during Hurricane Harvey. *World Environmental and Water Resources Congress 2018*.
- Booker, J.F., A.M. Michelsen, and F.A. Ward, 2005. Economic impact of alternative policy responses to prolonged and severe drought in the Rio Grande Basin. *Water Resources Research* 41(2).
- Bretherton, C.S., C. Smith, and J.M. Wallace, 1992. An intercomparison of methods for finding coupled patterns in climate data. *Journal of Climate* 5(6): 541-560.
- Cayan, D. R., 1992. El Niño/southern oscillation and streamflow in the western United States. *El Niño, Historical and Paleoclimatic Aspects of the Southern Oscillation*, 29-68.
- Çimen, M., and O. Kisi, 2009. Comparison of two different data-driven techniques in modeling lake level fluctuations in Turkey. *Journal of hydrology*, 378(3), 253-262.
- Christensen, N.S., A.W. Wood, N. Voisin, D.P. Lettenmaier, and R.N. Palmer, 2004. The effects of climate change on the hydrology and water resources of the Colorado River basin. *Climatic change* 62(1-3): 337-363.

- Corte-Real, J., X. Zhang, and X. Wang, 1995. Large-scale circulation regimes and surface climatic anomalies over the Mediterranean. *International Journal of Climatology*, 15(10), 1135-1150.
- Coulibaly, P., and C.K. Baldwin, 2005. Nonstationary hydrological time series forecasting using nonlinear dynamic methods. *Journal of Hydrology*, 307(1-4), 164-174.
- Dettinger, M. D., and H.F. Diaz, 2000. Global characteristics of stream flow seasonality and variability. *Journal of Hydrometeorology*, 1(4), 289-310.
- Dettinger, M.D., D.R. Cayan, H.F. Diaz, and D.M. Meko, 1998. North-south precipitation patterns in western North America on interannual-to-decadal timescales. *Journal of Climate* 11(12): 3095-3111.
- Dibike, Y. B., 2000. Machine learning paradigms for rainfall-runoff modelling. *Hydroinformatics* 2000.
- Dibike, Y. B., S. Velickov, D. Solomatine, and M. B. Abbott, 2001. Model induction with support vector machines: introduction and applications. *Journal of Computing in Civil Engineering*, 15(3), 208-216.
- Du, K., Y. Zhao, and J. Lei, 2017. The incorrect usage of singular spectral analysis and discrete wavelet transform in hybrid models to predict hydrological time series. *Journal of Hydrology*, 552, 44-51.
- Foufoula-Georgiou, E., P. Kumar, T. Mukerji, and G. Mavko, 1995. Wavelets in geophysics. *Pure and Applied Geophysics*, 145(2), 374-375.
- Gleick, P. H., 1989. Climate change, hydrology, and water resources. *Reviews of Geophysics*, 27(3), 329-344.

- Grantz, K., B. Rajagopalan, M. Clark, and E. Zagona, 2005. A technique for incorporating large-scale climate information in basin-scale ensemble streamflow forecasts. *Water Resources Research* 41(10).
- Grinsted, A., J.C. Moore, and S. Jevrejeva, 2004. Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear processes in geophysics*, 11(5/6), 561-566.
- Gupta, H.V., S. Sorooshian, and P.O. Yapo, 1999. Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. *Journal of Hydrologic Engineering* 4(2): 135-143.
- Intergovernmental Panel on Climate Change. (2015). *Climate change 2014: mitigation of climate change* (Vol. 3). Cambridge University Press.
- Jevrejeva, S., J.C. Moore, and A. Grinsted, 2003. Influence of the Arctic Oscillation and El Niño-Southern Oscillation (ENSO) on ice conditions in the Baltic Sea: The wavelet approach. *Journal of Geophysical Research: Atmospheres*, 108(D21).
- Jobe, A., S. Bhandari, A. Kalra, and S. Ahmad, 2017. Ice-Cover and Jamming Effects on Inline Structures and Upstream Water Levels. *World Environmental and Water Resources Congress 2017* (pp. 270-279).
- Jobe, A., A. Kalra, and E. Ibendahl, 2018. Conservation Reserve Program effects on floodplain land cover management. *Journal of environmental management*, 214, 305-314.
- Kahya, E., and J. A. Dracup, 1993. US streamflow patterns in relation to the El Niño/Southern Oscillation. *Water Resources Research*, 29(8), 2491-2503.

- Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, L., and Y. Zhu, 1996. The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American meteorological Society*, 77(3), 437-471.
- Kalra, A., T. C. Piechota, R. Davies, and G. A. Tootle, 2008. Changes in US streamflow and western US snowpack. *Journal of Hydrologic Engineering*, 13(3), 156-163.
- Kalra, A., and S. Ahmad, 2011. Evaluating changes and estimating seasonal precipitation for the Colorado River Basin using a stochastic nonparametric disaggregation technique. *Water Resources Research*, 47(5).
- Kalra, A., and S. Ahmad, 2012. Estimating annual precipitation for the Colorado River Basin using oceanic-atmospheric oscillations. *Water Resources Research*, 48(6).
- Kalra, A., S. Sagarika, P. Pathak, and S. Ahmad, 2017. Hydro-climatological changes in the Colorado River Basin over a century. *Hydrological Sciences Journal*, 62(14), 2280-2296.
- Kandissounon, G. A., A. Kalra, S. Ahmad, 2018. Integrating System Dynamics and Remote Sensing to Estimate Future Water Usage and Average Surface Runoff in Lagos, Nigeria. *Civil Engineering Journal*, 4(2), 378-393.
- Khedun, C. P., A.K. Mishra, J.D. Bolten, H.K. Beaudoin, R.A. Kaiser, J.R. Giardino, and V.P. Singh, 2012. Understanding changes in water availability in the Rio Grande/Río Bravo del Norte basin under the influence of large-scale circulation indices using the Noah land surface model. *Journal of Geophysical Research: Atmospheres*, 117(D5).
- Kundzewicz, Z.W., L.J. Mata, N.W. Arnell, P. Döll, B. Jimenez, K. Miller, and I. Shiklomanov, 2008. The implications of projected climate change for freshwater resources and their management. *Hydrological Sciences Journal* 53:1, 2-10

- Labat, D., 2005. Recent advances in wavelet analyses: Part 1. A review of concepts. *Journal of Hydrology*, 314(1-4), 275-288.
- Labat, D., 2010. Cross wavelet analyses of annual continental freshwater discharge and selected climate indices. *Journal of Hydrology*, 385(1-4), 269-278.
- Lins, H.F., 2012. USGS hydro-climatic data network 2009 (HCDN-2009) (No. 2012-3047). US Geological Survey.
- Liong, S.Y., and C. Sivapragasam, 2002. Flood stage forecasting with support vector machines. *Journal of the American Water Resources Association (JAWRA)* 38(1): 173-186.
- Maheshwari, P., R. Khaddar, P. Kachroo, and A. Paz, 2016. Dynamic Modeling of Performance Indices for Planning of Sustainable Transportation Systems. *Networks and Spatial Economics*, 16(1), 371-393
- Marques, C. A. F., J. A. Ferreira, A. Rocha, J. M. Castanheira, P. Melo-Goncalves, N. Vaz, and J. M. Dias, 2006. Singular spectrum analysis and forecasting of hydrological time series. *Physics and Chemistry of the Earth, Parts A/B/C*, 31(18), 1172-1179.
- McCabe, G.J., M.A. Palecki, and J.L. Betancourt, 2004. Pacific and Atlantic Ocean influences on multidecadal drought frequency in the United States. *Proceedings of the National Academy of Sciences* 101(12): 4136-4141.
- Michelsen, A.M., and K. Wood, 2003. Water demand in the Paso del Norte region. Paper presented at Weather and Water on the Border: A Forum on Drought, Paso del Norte Water Task Force. El Paso, Tex.
- Middelkoop, H., K. Daamen, D. Gellens, W. Grabs, J.C. Kwadijk, H. Lang, and K. Wilke, 2001. Impact of climate change on hydrological regimes and water resources management in the Rhine basin. *Climatic change* 49(1-2): 105-128.

- Miles, E. L., A.K. Snover, A.F. Hamlet, B. Callahan, and D. Fluharty, 2000. Pacific Northwest regional assessment: the impacts of climate variability and climate change on the water resources of the Columbia River Basin. *JAWRA Journal of the American Water Resources Association*, 36(2), 399-420.
- Moriasi, D.N., J.G. Arnold, M.W. Van Liew, R.L. Bingner, R.D. Harmel, and T.L. Veith, 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE* 50(3): 885-900.
- Nash, J.E., and J.V. Sutcliffe, 1970. River flow forecasting through conceptual models part I—A discussion of principles. *Journal of hydrology* 10(3): 282-290.
- Nijssen, B., G.M. O'Donnell, A.F. Hamlet, and D.P. Lettenmaier, 2001. Hydrologic sensitivity of global rivers to climate change. *Climatic change* 50(1-2): 143-175.
- Nourani, V., M. Komasi, and A. Mano, 2009. A multivariate ANN-wavelet approach for rainfall-runoff modeling. *Water resources management*, 23(14), 2877-2894.
- Nyaupane, N., R. Thakali, A. Kalra, L. Mastino, M. Velotta, S. Ahmad, 2017. Response of Climate Change on Urban Watersheds: A Case Study for Las Vegas, NV. *World Environmental and Water Resources Congress 2017*.
- Nyaupane, N., S. Bhandari, M. Rahaman, K. Wagner, A. Kalra, S. Ahmad, R. Gupta, 2018. Flood Frequency Analysis Using Generalized Extreme Value Distribution and Floodplain Mapping for Hurricane Harvey in Buffalo Bayou. *World Environmental and Water Resources Congress 2018 (pp. 364-375)*.
- Nyaupane, N., S. R. Mote, M. Bhandari, A. Kalra, S. Ahmad, 2018. Rainfall-Runoff Simulation Using Climate Change Based Precipitation Prediction in HEC-HMS Model for Irwin

Creek, Charlotte, North Carolina. *World Environmental and Water Resources Congress 2018*

Pahl-Wostl, C., 2007. Transitions towards adaptive management of water facing climate and global change. *Water resources management* 21(1): 49-62.

Pascolini-Campbell, M., R. Seager, A. Pinson, and B.I. Cook, 2017. Covariability of climate and streamflow in the Upper Rio Grande from interannual to interdecadal timescales. *Journal of Hydrology: Regional Studies*, 13, 58-71.

Pathak, P., A. Kalra, S. Ahmad, and M. Bernardez, 2016. Wavelet-aided analysis to estimate seasonal variability and dominant periodicities in temperature, precipitation, and streamflow in the Midwestern United States. *Water resources management*, 30(13), 4649-4665.

Pathak, P. (2016). *Seasonal variation in temperature, precipitation and streamflow across the Midwestern United States*. Southern Illinois University at Carbondale.

Pathak, P., M. Bhandari, A. Kalra, and S. Ahmad, 2016. Modeling Floodplain Inundation for Monument Creek, Colorado. *World Environmental and Water Resources Congress* (pp. 131-140).

Pathak, P., A. Kalra, S. Ahmad, 2017. Temperature and precipitation changes in the Midwestern United States: implications for water management. *International Journal of Water Resources Development*, 33(6), 1003-1019.

Pathak, P., A. Kalra, K. W. Lamb, W. P. Miller, S. Ahmad, R. Amerineni, R., and D. P. Ponugoti, 2018. Climatic variability of the Pacific and Atlantic Oceans and western US snowpack. *International Journal of Climatology*, 38(3), 1257-1269.

- Pathak, P., A. Kalra, K.W. Lamb, W.P. Miller, S. Ahmad, R. Amerineni, and D.P. Ponugoti, 2017. Climatic variability of the Pacific and Atlantic Oceans and western US snowpack. *International Journal of Climatology*.
- Percival, D. B., and A.T. Walden, 2000. Wavelet methods for time series analysis, vol. 4 of Cambridge Series in Statistical and Probabilistic Mathematics.
- Piechota, T. C., and J.A. Dracup, 1996. Drought and regional hydrologic variation in the United States: Associations with the El Niño-Southern Oscillation. *Water Resources Research*, 32(5), 1359-1373.
- Piechota, T.C., F.H. Chiew, J.A. Dracup, and T.A. McMahon, 2001. Development of exceedance probability streamflow forecast. *Journal of Hydrologic Engineering* 6(1): 20-28.
- Potts, J.M., C.K. Folland, I.T. Jolliffe, and D. Sexton, 1996. Revised “LEPS” scores for assessing climate model simulations and long-range forecasts. *Journal of Climate* 9(1): 34-53.
- Praus, P., 2005. Water quality assessment using SVD-based principal component analysis of hydrological data. *Water SA*, 31(4), 417-422.
- Rajagopalan, B., E. Cook, U. Lall, and B.K. Ray, 2000. Spatiotemporal variability of ENSO and SST teleconnections to summer drought over the United States during the twentieth century. *Journal of Climate* 13(24): 4244-4255.
- Redmond, K.T., and R.W. Koch, 1991. Surface climate and streamflow variability in the western United States and their relationship to large-scale circulation indices. *Water Resources Research* 27(9): 2381-2399.
- Ropelewski, C. F., and M.S. Halpert, 1987. Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation. *Monthly weather review*, 115(8), 1606-1626.

- Sagarika, S., A. Kalra, and S. Ahmad, 2015. Interconnections between oceanic–atmospheric indices and variability in the US streamflow. *Journal of Hydrology* 525: 724-736.
- Scholkopf, B., K. K. Sung, C. J. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik, 1997. Comparing support vector machines with Gaussian kernels to radial basis function classifiers. *IEEE transactions on Signal Processing*, 45(11), 2758-2765.
- Slack, J.R., and J.M. Landwehr, 1992. Hydro-climatic data network (HCDN); a U.S. Geological Survey streamflow data set for the United States for the study of climate variations, 1874-1988. U.S. Geological Survey, Open-File Report 92-129.
- Soukup, T.L., O.A. Aziz, G.A. Tootle, T.C. Piechota, and S.S. Wulff, 2009. Long lead-time streamflow forecasting of the North Platte River incorporating oceanic–atmospheric climate variability. *Journal of Hydrology* 368(1): 131-142.
- Stewart, I.T., D.R. Cayan, and M.D. Dettinger, 2004. Changes in snowmelt runoff timing in western North America under a business as usual climate change scenario. *Climatic Change* 62(1-3): 217-232.
- Tamaddun, K. A., A. Kalra, and S. Ahmad, 2016. Patterns and Periodicities of the Continental US Streamflow Change. In *World Environmental and Water Resources Congress 2016*(pp. 658-667).
- Tamaddun, K.A., A. Kalra, M. Bernardez, and S. Ahmad, 2017. Multi-Scale Correlation between the Western US Snow Water Equivalent and ENSO/PDO Using Wavelet Analyses. *Water Resources Management* 31(9): 2745-2759.
- Tamaddun, K. A., A. Kalra, and S. Ahmad, 2017. Wavelet analyses of western US streamflow with ENSO and PDO. *Journal of Water and Climate Change*, 8(1), 26-39.

- Tang, C., D. Chen, B.T. Crosby, T.C. Piechota, and J.M. Wheaton, 2014. Is the PDO or AMO the climate driver of soil moisture in the Salmon River Basin, Idaho?. *Global and Planetary Change*, 120, 16-23.
- Team, C. W., R.K. Pachauri, and L.A. Meyer, 2014. IPCC, 2014: climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the intergovernmental panel on Climate Change. *IPCC, Geneva, Switzerland*, 151.
- Thakali, R., A. Kalra, and S. Ahmad, 2016. Understanding the Effects of Climate Change on Urban Stormwater Infrastructures in the Las Vegas Valley. *Hydrology*, 3(4), 34.
- Thakali, R., R. Bhandari, G. A. A. D. Kandissounon, A. Kalra and S. Ahmad, 2017. Flood Risk Assessment Using the Updated FEMA Floodplain Standard in the Ellicott City, Maryland, United States. *World Environmental and Water Resources Congress 2017*.
- Thakali, R., A. Kalra, S. Ahmad, and K. Qaiser, 2018. Management of an Urban Stormwater System Using Projected Future Scenarios of Climate Models: A Watershed-Based Modeling Approach. *Open Water Journal*, 5(2), 1.
- Thakur, B., R. Parajuli, A. Kalra, S. Ahmad, and R. Gupta, 2017. Coupling HEC-RAS and HEC-HMS in Precipitation Runoff Modelling and Evaluating Flood Plain Inundation Map. *World Environmental and Water Resources Congress 2017*.
- Thakur, B., P. Pathak, A. Kalra, S. Ahmad, and M. Bernardez, 2017. Using Wavelet to Analyze Periodicities in Hydrologic Variables. *World Environmental and Water Resources Congress 2017*.
- Thakur, B., R. Parajuli, G. R. Lamichhane, A. Alyami, A. Kalra, S. Ahmad, and R. Gupta, 2018. *World Environmental and Water Resources Congress 2018*.

- Tootle, G.A., and T.C. Piechota, 2006. Relationships between Pacific and Atlantic Ocean sea surface temperatures and US streamflow variability. *Water Resources Research* 42(7).
- Torrence, C., and G.P. Compo, 1998. A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, 79(1), 61-78.
- Torrence, C., & Webster, P. J. (1999). Interdecadal changes in the ENSO–monsoon system. *Journal of Climate*, 12(8), 2679-2690.
- Torrence, C., and P.J. Webster, 1999. Interdecadal changes in the ENSO–monsoon system. *Journal of Climate*, 12(8), 2679-2690.
- Trenberth, K. E., and J. Fasullo, 2007. Water and energy budgets of hurricanes and implications for climate change. *Journal of Geophysical Research: Atmospheres*, 112(D23).
- U.S. Department of Interior, 2003. Water 2025: Preventing crises and conflict. Bur. of Reclam., Washington, D. C.
- Vapnik, V., 1995. The nature of statistical learning theory. *Springer*, New York.
- Vapnik, V., 1998. Statistical learning theory. *John Wiley*, New York.
- Wallace, J.M., and D.S. Gutzler, 1981. Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Monthly Weather Review* 109(4): 784-812.
- Wallace, J.M., C. Smith, and C.S. Bretherton, 1992. Singular value decomposition of wintertime sea surface temperature and 500-mb height anomalies. *Journal of climate* 5(6): 561-576.
- Ward, M.N., and C.K. Folland, 1991. Prediction of seasonal rainfall in the north nordeste of Brazil using eigenvectors of sea-surface temperature. *International Journal of Climatology* 11(7): 711-743.

- Ward, P. J., W. Beets, L.M. Bouwer, J.C. Aerts, and H. Renssen, 2010. Sensitivity of river discharge to ENSO. *Geophysical Research Letters*, 37(12).
- Webster, P. J., V.O. Magana, T.N. Palmer, J. Shukla, R.A. Tomas, M.U. Yanai, and T. Yasunari, 1998. Monsoons: Processes, predictability, and the prospects for prediction. *Journal of Geophysical Research: Oceans*, 103(C7), 14451-14510.
- Woodhouse, C.A., D.M. Meko, D. Griffin, and C.L. Castro, 2013. Tree rings and multiseason drought variability in the lower Rio Grande Basin, USA. *Water Resources Research* 49(2): 844-850.
- Zealand, C.M., D.H. Burn, and S.P. Simonovic, 1999. Short-term streamflow forecasting using artificial neural networks. *Journal of hydrology* 214(1): 32-48.

## VITA

Graduate School

Southern Illinois University Carbondale

Swastik Bhandari

Swastikbhandari7@gmail.com

Tribhuvan University, Kathmandu, Nepal

Bachelor of Engineering, Civil Engineering, August 2013

Thesis Paper Title: Analyzing the Relationship between Large Scale Climate Variability and  
Streamflow of the Continental United States

Major Professor: Dr. Ajay Kalra

Publications:

Bhandari, S., A. Kalra, K. Tamaddun, and S. Ahmad, 2018. Relationship between ocean-atmospheric climate variables and regional streamflow of the conterminous United States. *Hydrology*, 5(2), 30 (DOI: 10.3390/hydrology5020030)

Bhandari, S., A. Jobe, B. Thakur, A. Kalra, S. Ahmad, 2018. Flood Damage Reduction in Urban Areas with Use of Low Impact Development Designs. In *World Environmental and Water Resources Congress 2018 (pp. 52-61)*.

Nyaupane, N., S. Bhandari, M. Rahaman, K. Wagner, A. Kalra, S. Ahmad, R. Gupta, 2018. Flood Frequency Analysis Using Generalized Extreme Value Distribution and Floodplain Mapping for Hurricane Harvey in Buffalo Bayou. In *World Environmental and Water Resources Congress 2018 (pp. 364-375)*.

Jobe, A., S. Bhandari, A. Kalra, and S. Ahmad, 2017. Ice-Cover and Jamming Effects on Inline Structures and Upstream Water Levels. In *World Environmental and Water Resources Congress 2017* (pp. 270-279).

Bhandari S., and A. Kalra, 2016. Effect of Climate Change on Streamflow Variability in the Rio Grande River Basin. *22<sup>nd</sup> Annual Mid-American Environmental Engineering Conference (MAEEC)*, Southern Illinois University Edwardsville. (Poster: MAEEC-2016)