

Incorporating El Niño Southern Oscillation (ENSO)-Induced Climate Variability for Long-Range Hydrologic Forecasting and Stream Water Quality Protection

by

Suresh Sharma

A dissertation submitted to the Graduate Faculty of
Auburn University
In partial fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 4, 2012

Keywords: Climate variability; Watershed model;
Data-driven model; Water quality model

Copyright 2012 by Suresh Sharma

Approved by

Puneet Srivastava, Co-chair, Associate Professor of Biosystems Engineering
Xing Fang, Co-chair, Associate Professor of Civil Engineering
Latif Kalin, Associate Professor of School of Forestry and Wildlife Sciences
Evaden Brantley, Assistant Professor of Agronomy and Soils

Abstract

El Niño Southern Oscillation (ENSO) has been found to have a strong predictable effect on streamflow in different parts of the world. Since ENSO has potential relationship to seasonal and inter-annual variability in streamflow, identifying the potential linkage between streamflow and ENSO and applying that linkage to data-driven model can improve the streamflow simulation and forecasting. That is, streamflow forecasting using sea surface temperature (SST) can be useful in ENSO-affected regions. In addition, ENSO might have teleconnection with stream water quality. One of the sources of stream water quality degradation is point source pollution which is regulated under the “National Point Discharge Elimination System (NPDES)” permitting process. Since conventional NPDES permits do not consider seasonal to inter-annual climate variability, they are either under-protective or over-protective of stream water quality. Application of ENSO information in NPDES permitting can be useful for stream water quality protection. Further, ENSO signals can be utilized to predict total organic carbon (TOC) loads, which form disinfection byproducts during chlorination of drinking water. Therefore, the specific objectives of this research were: (i) to demonstrate that the data-driven model, Adaptive Neuro Fuzzy Inference System (ANFIS) that incorporates SST and sea level pressure (SLP) can simulate streamflow as good as the Loading Simulation Program C++ (LSPC++), (ii) to quantify the long-range streamflow forecasting skill of the ANFIS model with the fusion of SST and predicted climate data, (iii) to demonstrate how ENSO information can be incorporated to improve NPDES permitting system in a complex river system using a watershed linked

hydrodynamic and water quality model, and (iv) to predict TOC loads quantitatively in different ENSO phases using climate data and ENSO indices, and forecast TOC load qualitatively using a fuzzy logic approach. It was found that: (i) the performance of the ANFIS model was comparable to the LSPC model, (ii) streamflow forecast using SST at one month lead time was satisfactory, (iii) ENSO information was useful for regulating point sources for stream water quality protection, and (iv) the TOC load was correlated with the ENSO phase; therefore, TOC load was predicted both quantitatively and qualitatively in different ENSO phases.

Acknowledgments

I would like to express my gratitude to my advisor Dr. Puneet Srivastava for providing me the opportunity to pursue PhD under his guidance. I am grateful for his continuous encouragement, support, motivation and proper direction to complete this study.

I would also like to express my sincere gratitude to Dr. Xing Fang for serving as a Co-chair and providing valuable guidance, suggestion, and feedback for this research. I am grateful to Dr. Latif Kalin for his valuable guidance, technical inputs, and suggestion. I am thankful to Dr. Evaden Brantley for her encouragement and valuable feedback. I am also grateful to Dr. Luke Marzen for serving as a university reader.

Special thank goes to Dr. Sushil Adhikari and Ms. Linda Newton for their help in different ways. I am also indebted to my present and past colleagues in my research group including Dr. Sumit Sen, Mr. Anand Gupta, Mr. Jasmeet Lamba, Mr. Pratap Mondal, Ms. Golbahar Mirhosseini, Ms. Vaishali Sharda, Ms. Gena Johnson, and Mr. Subhasis Mitra for their companionship. I am also indebted to Mr. Brian Watson and Ms. Jamie Childer from Tetra-tech for their technical suggestions. Special thanks go to Mr. Lynn Sisk and Mr. Charlie Reynolds from the Alabama Department of Environmental Management (ADEM).

Finally, I would like to sincerely thank my wife, Ms. Uma Paudel, for her technical support for data downscaling, processing and analysis. Besides, I am thankful to her continuous support and encouragement to complete this research. I would also like to thank my other family members for their help and support.

Table of Contents

Abstract	ii
Acknowledgments.....	iv
List of Tables	xi
List of Figures	xiii
List of Abbreviations	xvii
Chapter 1. Introduction	1
1.1 Background.....	1
1.2 El Niño Phase.....	2
1.3 La Niña phase	3
1.4 Neutral (Normal) Phase	3
1.5 Evolution of ENSO.....	3
1.6 ENSO in the Southeast USA	4
1.7 Impact of ENSO on Hydrologic Cycle and Water Quality	4
1.8 Objectives and Hypothesis.....	5
1.9 Specific Research Objectives.....	6
1.10 Organization of the Study	6
1.11 References.....	11
Chapter 2. Climate Variability-Based Streamflow Simulations Using Neuro-Fuzzy Computational Techniques	13
2.1 Abstract.....	13

2.2 Introduction.....	14
2.3 Theoretical Background.....	17
2.3.1 Wavelet Analysis	18
2.3.2 Continuous Wavelet Transforms (CWT).....	19
2.3.3 Cross Wavelet Transform (XWT)	20
2.3.4 Wavelet Coherence (WTC).....	21
2.3.5 Cross-correlation Analysis.....	21
2.3.6 Adaptive Neuro Fuzzy Inference System (ANFIS)	22
2.3.6.1 Concept	22
2.3.6.2 Estimation of Parameters	25
2.3.6.3 Subtractive and Fuzzy C -Mean Clustering for Structure Identification	26
2.4 Methodology	28
2.4.1 Study Area and Data	28
2.4.2 Input Data Selection.....	28
2.4.3 Model Development and Implementation	30
2.4.4 Comparison with Watershed Model	32
2.4.5 Loading Simulation Program C++ (LSPC) Model	33
2.4.6 LSPC Model Configuration	34
2.4.7 Hydrologic Model Calibration and Validation	35
2.5 Result and Discussion.....	35
2.5.1 Wavelet Analysis	35
2.5.2 Cross-wavelet Analysis.....	36
2.5.3 Wavelet Coherence Analysis	37

2.5.4 Performance of the ANFIS Model.....	37
2.5.5 Performance of LSPC Model.....	38
2.5.6 Model Comparison.....	38
2.6 Summary and Conclusion.....	41
2.7 References.....	56
Chapter 3. Long-Range Hydrologic Forecasting in El Niño Southern Oscillation-Affected Coastal Watersheds.....	61
3.1 Abstract.....	61
3.2 Introduction.....	62
3.3 Theoretical Consideration.....	65
3.3.1 ANFIS Model.....	65
3.3.2 ANFIS Model Development.....	65
3.3.3 Estimation of Parameters.....	66
3.4 Material and Methods.....	68
3.4.1 Study Area.....	68
3.4.2 CFSv2 Model.....	68
3.4.3 Weather Generator Approach.....	69
3.4.4 Input Data and Preprocessing.....	71
3.4.4.1 Input Data Selection.....	71
3.4.5 Model Training, Validation and Testing.....	72
3.4.6 Bias Correction.....	73
3.4.6.1 Quantile Mapping Method.....	74
3.5 Result and Discussion.....	75
3.6 Summary and Conclusion.....	78

3.7 References.....	86
Chapter 4. Incorporating Climate Variability for Point Source Discharge Permitting In a Complex River System	90
4.1 Abstract.....	90
4.2 Introduction.....	91
4.3 Theoretical Background.....	94
4.3.1 Ammonia Nitrogen: Basic Concept	94
4.3.2 EPA Approach for Controlling in stream Ammonia Nitrogen	95
4.3.3 National Criteria for Ammonia in Fresh Water	96
4.3.4 Ammonia Permitting Approach.....	97
4.3.5 Dissolved Oxygen Modeling Concept.....	98
4.3.6 Dissolved Oxygen Model	99
4.3.7 Climate Variability and Water Quality	100
4.4 Materials and Methods.....	100
4.4.1 Study Area	100
4.4.2 Overall Modeling Approach	101
4.4.3 Watershed Model (LSPC++)	102
4.4.4 Watershed Model Configuration and Input Data.....	103
4.4.5 Streamflow Simulation	105
4.4.6 Stream Temperature Simulation	105
4.4.7 Simulation of Pollutant and Source Assessment	106
4.4.8 Nutrients, Chlorophyll a, BOD5 and DO Simulation in the Watershed Model	107
4.4.9 Hydrodynamic and In-stream Water Quality Model	108
4.4.10 Hydrodynamic Model Configuration.....	109

4.4.11 In stream Water Quality Model Configuration and Calibration	110
4.4.11.1 Input Data for EPD-RIV1 Model.....	110
4.5 Result and Discussion.....	112
4.5.1 Model Performance.....	112
4.5.2 Effect of ENSO on Streamflow, Temperature and DO	113
4.5.3 ENSO and Ammonia Permit.....	115
4.5.4 ENSO for the Identification of Critical Conditions	118
4.5.5 ENSO Signal for Critical Conditions.....	119
4.6 Summary and Conclusion.....	120
4.7 Acknowledgement	122
4.8 Refrences	136
Chapter 5. Predicting Total Organic Carbon Load with El Niño Southern Oscillation Phase Using Hybrid and Fuzzy Logic Approaches	141
5.1 Abstract.....	141
5.2 Introduction.....	142
5.3 Study Area and Data	145
5.3.1 TOC Load Generation.....	146
5.3.2 ENSO and ENSO Indicators.....	147
5.3.3 Selection of ENSO Indices	148
5.4 Data-Driven Modeling Approach	149
5.4.1 Artificial Neural Network	150
5.4.2 Principal Component Regressions with Artificial Neural Network (PCR-ANN)	151
5.4.3 Fuzzy Logic Approach.....	152
5.5 Model Evaluation Criteria	154

5.6 Results and Discussion	155
5.6.1 ENSO Correlation with TOC Load.....	155
5.6.2 PCR- ANN Model Training and Testing.....	156
5.6.3 Effect of Climate Variability Caused by ENSO	158
5.6.4 Fuzzy Logic Modeling.....	159
5.6.5 Fuzzy Logic Model Calibration, Validation and Testing	160
5.6.6 Comparison of PCR-ANN and Fuzzy Logic Models	162
5.6.7 TOC Load Forecast.....	163
5.7 Summary and Concluding Remarks	164
5.8 Acknowledgement	165
5.9 References.....	178
Chapter 6. Conclusion and Recommendation.....	186
6.1 Summary.....	186
6.1.1 Conclusion of Objective I.....	186
6.1.2 Conclusion of Objective II.....	187
6.1.3 Conclusion of Objective III	187
6.1.4 Conclusion of Objective IV	188
6.1.5 Limitations of the Study	189
6.1.6 Suggestion for Future Work	189

List of Tables

Table 2.1. Data used for the study area with their sources and format.....	43
Table 2.2. Inputs for ANFIS model (monthly data).....	43
Table 2.3. The calibrated model parameters in LSPC model	43
Table 2.4 El Niño and La Niña years (Dec-April) from 1950 to 2003.....	44
Table 2.5. Performance of the ANFIS model in different modeling stages (monthly simulation from 1952-2002).....	44
Table 2.6. Model simulation in daily scale and performance comparison (1990-1996)	44
Table 2.7. Performance comparison of two models at monthly time scale (period 1990-1996)..	45
Table 2.8. Performance comparison of ANFIS and LSPC model during a period of LSPC model calibration and validation (monthly simulation).....	46
Table 2.9. Performance of LSPC and ANFIS model during 50 years simulation at monthly time step	46
Table 2.10. Model comparison for 25 ENSO events (monthly simulation)	46
Table 3.1. Input parameters showing best correlation with streamflow in Chickasaw Creek	80
Table 3.2. Streamflow forecast after bias correction using weather generator approach and CFSv2 data.....	80
Table 4.1. Data used for the study with their source and information.....	123
Table 4.2. Statistical parameters measuring the performance of the watershed model	123
Table 4.3. Correlations of observed streamflow, simulated stream temperature and DO with Niño 3.4 indexes since 1950.	123
Table 4.4. Ammonia limit in different ENSO phase	124
Table 5.1. Multi Linear Regression model performance with and without using ENSO information.....	166

Table 5.2. Hidden neuron selection for the model.....	166
Table 5.3. The architecture of the ANN model used in this study.....	167
Table 5.4. The PCR-ANN model performance in different stage: a) PCR-ANN model with Niño 3.4 index and TNI, and b) PCR-ANN model without Niño 3.4 index and TNI.....	167
Table 5.5. Rules used for fuzzy logic model, (a) JFM and AMJ, and (b) ASO	168
Table 5.6. Fuzzy logic model calibration, validation and testing	169
Table 5.7. Post validation of the qualitative TOC load forecast.....	169

List of Figures

- Figure 1.1. Figure describing the different ENSO phases; (a) El Niño phase, (b) La Niña phase and (c) Neutral phase.(Source: http://www.pmel.noaa.gov/tao/elNiño/Niño_normal.html) 10
- Figure 2.1. ANFIS architecture for two inputs (X and Y) with two rules and two membership functions (P1, P2 and Q1, Q2) for each rule..... 47
- Figure 2.2. Chickasaw Creek watershed in Mobile County of South Alabama, USA, showing the USGS gaging station and land use distribution within the watershed. The land use classification is as per NLCD land use categories. 48
- Figure 2.3. Cross correlation function (CCF) of the sea surface temperature with: a) surface runoff, and (b) base flow at different lagging months. Negative lag indicates the second series is lagged to first series. Here first series is sea surface temperature and second series is either surface runoff or base flow. 49
- Figure 2.4. (a) The wavelet power spectrum with in the cone of influence for monthly streamflow, (b) monthly SST anomalies in Niño 3.4 region. The red contour line indicates high wavelet power. The right figure indicates the Global Wavelet Spectrum (GWS). The dashed blue line indicates the 95% confidence limit. X-axis indicates the time period, and Y-axis indicates the period of occurrence. Red eye land corresponds to higher power (in the right side); for which, the period of occurrence is 2 to 7 years. It suggests that SST varies significantly with in a frequency of 2 to 7 years from a period of 1968 to 2005. 50
- Figure 2.5. (a) Cross wavelet spectrum between monthly streamflow and monthly SST, and (b) Cross wavelet spectrum between monthly streamflow and monthly SLP difference at Tahiti and Darwin. Black outline indicates that the relationship is significant to 95% within the region. The right vertical band associated with each figure indicates the global wavelet power. The clockwise “arrow pointing” indicates the same phase (positive) relationship, and anti-clockwise arrow indicates the opposite relationship (negative). The left panel suggests that the strong relationship with high power is observed during 1970 to 1982 with in the frequency of 2 to 5 years; however, this frequency is approximately 3 to 8 years from 1995 to 2000. 51
- Figure 2.6. Left figure indicates the wavelet coherence analysis (WTC) between streamflow and SST and the right figure indicates the WTC between streamflow and SLP difference (Tahiti and Darwin). Clockwise arrow indicates the in-phase relationship and anti-clock wise..... 51

Figure 2.7. Schematic diagram of the Stanford Watershed Model adapted for LSPC model. Number corresponding to each circle denotes the order of removing water to satisfy ET (Source: LSPC User’s Manual, Tetra tech, 2009).	52
Figure 2.8. Monthly streamflow simulated for 52 years using ANFIS model compared with observed monthly streamflow.....	53
Figure 2.9. LSPC model calibration and its performance comparison with ANFIS model from 1990 to 1996 (monthly streamflow).	53
Figure 2.10. LSPC model validation and comparison with ANFIS model during LSPC model validation phase.	54
Figure 2.11. Statistical analysis of monthly streamflow simulated using: (a) ANFIS model and (b) LSPC model.	54
Figure 2.12. Monthly streamflow simulation using LSPC and ANFIS models corresponding to historic La Niña and El Niño events (Table 3) since 1950. Dots and triangles represent ANFIS and LSPC simulation during La Niña and El Niño events.....	55
Figure 3.1. ANFIS structure.....	81
Figure 3.2. Chickasaw Creek watershed in Mobile County of South Alabama.	82
Figure 3.3. Comparison of observed and CFSv2 forecast mean monthly precipitation and temperature (forecast lead time is one month and period is 1982 to 2009).....	83
Figure 3.4. Comparison of observed precipitation and temperature with weather generated precipitation and temperature (average monthly).	84
Figure 3.5. Streamflow forecast at one month lead time using CFSv2 retrospective forecast data.	84
Figure 3.6. Streamflow forecast at one month lead time using weather generated climate data..	84
Figure 3.7. Probabilistic streamflow forecast at one month lead time using CFSv2 data.	85
Figure 3.8. Streamflow forecast at three months lead time using weather generated climate data.	85
Figure 4.1. Schematic showing major processes influencing DO in streams and rivers (EPA, 1997).	125
Figure 4. 2. Chickasaw Creek watershed in Mobile County of South Alabama with USGS gage, waste water treatment plant, water quality station CS1 (Latitude = 30.78224, Longitude = -88.07248) and cross section of the river.	126

Figure 4.3. Schematic diagram representing the river system modeling and analysis with ENSO for ammonia permit. EL and LN imply El Niño and La Niña, respectively.....	127
Figure 4.4. Observed and simulated a) daily calibrated streamflow, b) monthly calibrated streamflow, and c) monthly validated streamflow. EL and LN represents El Niño and La Niña respectively.	128
Figure 4.5. Observed DO response for various water quality parameters at monitoring station (CS1).	129
Figure 4.6. Observed and LSPC simulated (a) stream temperature, (b) total nitrogen, (c) total phosphorus, and (d) ammonia at the water quality station CS1 (Figure4. 2). EL and LN stands for El Niño and LN, respectively.	130
Figure 4.7. Observed and simulated a) BOD, (b) stream temperature, (c) DO at monitoring station CS1 using LSPC linked EPD-RIV1 water quality model.	131
Figure 4.8. Box plot showing effect of ENSO (El Niño, La Niña and neutral phases) on (a) streamflow (Dec-April), (b) stream temperature (Dec-April), and (c) DO concentration (Dec-April).	132
Figure 4.9. Box plot showing effect of ENSO on (a) DO concentration (May-July), (b) streamflow (Aug-Oct), and (c) DO concentration (Aug-Oct).	132
Figure 4.10. 7days consecutive low flows for the period Dec-April in the El Niño phase for (a) Chickasaw Creek, (b) Perdido, and (c) Fish River watershed.	133
Figure 4.11. (a) Aug-Oct 7days consecutive low flows in La Niña phase in Chickasaw Creek watershed, (b) chart of precipitation index and 7 days consecutive low flows.	133
Figure 4.12. Pattern of streamflow in two different ENSO phases (average taken over 50 years).	134
Figure 4.13. a) Autocorrelation of the streamflow b) autocorrelation of base flow c) cross correlation of ENSO with SST anomalies.	
Figure 4.14. Figure showing the association of average 7 days consecutive low flows with Niño 3.4 index (ENSO). Primary horizontal axis (lower) and primary vertical axis (left) is for 7 days low flow of each month and secondary vertical axis (Right) is for Niño 3.4 index. EL and LN represent El Niño and La Niña phase, respectively.	135
Figure 5.1. Map of the study area showing the land use distribution in different sub-basins and USGS gauging station in the Big Creek watershed.	170
Figure 5.2. Calibration and validation of a) monthly streamflow, and b) monthly TOC loads generated by the LSPC model.	171

Figure 5.3. Schematic representation of Fuzzy inference system used in this study..... 171

Figure 5.4. Monthly variation in average TOC loads (averaged over 55 years) in different ENSO phases..... 172

Figure 5.5. Box plot showing total monthly average TOC loads (ton) for (a) January-February-March (JFM) season, (b) April-May-June (AMJ) season, and (c) Aug-Sept-Oct (ASO) season in different phases of ENSO..... 173

Figure 5.6. Calibrated membership function for the fuzzy logic model for precipitation, ENSO phase, and TOC load: (a) MF's for ENSO (For all seasons), (b) MF's for JFM (precipitation, TOC), (c) MF's for AMJ season (precipitation, TOC) (d) MF's for ASO season (PCP, TOC temperature), (e) MF's for neutral season (precipitation, TOC, temperature). L: "low", M: "medium", H: "High", V: "Very", E: "Exceptional". 177

List of Abbreviations

ANFIS	Adaptive Neuro Fuzzy Inference System
ANN	Artificial Neural Network
CFSv2	Climate Forecast System Version 2
DEM	Digital Elevation Model
ENSO	El Niño Southern Oscillation
EPD-RIV1	Environmental Protection Division-River1
HCR	Hydrograph Controlled Release
HEC-RAS	Hydrologic Engineering Center-River Analysis System
LSPC	Loading Simulation Program C++
MLR	Multi-Linear Regression
NCDC	National Climatic Data Center
NLCD	National Land Cover Datasets
NOAA	National Oceanic and Atmospheric Administration
NPDES	National Pollutant Discharge Elimination System
PCR	Principal Component Regression
SST	Sea Surface Temperature
SLP	Sea Level Pressure

Chapter 1. Introduction

1.1 Background

In the past decade, water resources experts have shown increasing interest in quantifying the impact of climate variability on hydrology and water quality. This growing interest has advanced the science to the extent that short-range climate forecasts can be utilized for water management and stream water quality protection to reduce the climate-induced risk and vulnerabilities.

Recently, scientists have shown their interest in utilizing climate variability resulting from a number of ocean-atmosphere phenomena that operate at seasonal to decadal time scales. Some of these phenomena are: El Niño Southern Oscillation (ENSO), North Atlantic Oscillation (NAO), Pacific Decadal Oscillations (PDO), and Pacific and North American (PNA) Oscillation. Among these global climate patterns, water resources scientists have disproportionately focused on ENSO for two reasons: (i) ENSO has strong association with local and global climate patterns, and (ii) ENSO can be forecasted with higher degree of certainty. Generally, various climate models are used for ENSO prediction. The ENSO prediction based on the consensus of various models are more skillful than the single model prediction systems. Also, the model predictions for ENSO are relatively more reliable for extreme ENSO events. Readers can find the reliability of ENSO prediction and status of ENSO forecast skill in different articles (e.g. Kirtman et al. 2001; Tang et al. 2005).

ENSO is a coupled ocean-atmospheric phenomenon caused by complex interaction among different climatic variables such as clouds, storms, winds, oceanic temperatures and ocean currents along the equatorial Pacific. It is a natural process which initiates in the Pacific Ocean and has severe consequences leading to the extreme climatic conditions not only in the Southeast USA but also in different parts of the world. ENSO is divided into El Niño, La Niña and Neutral phases. These ENSO phases are determined by different ENSO indices using Sea surface temperatures (SST), changes in Sea level pressures (SLPs) and wind patterns in the equatorial Pacific Ocean.

1.2 El Niño Phase

The term El Niño is derived from a Spanish word which means “The Little Boy or Christ Child.” The name was suggested to represent the phenomenon that starts at the beginning of the year, or Christmas time. El Niño occurs when warm water from the western Pacific flows toward the Eastern Pacific due to the weakening of trade winds (Figure 1.1 a). This leads to the flattening of the sea level, developing warm surface water off of the coast of South America, and increasing the water temperature in the Eastern Pacific. This increase in water temperature in the Pacific Ocean tends to change the atmospheric weather at the local and global scale. The hot and humid air over the ocean accelerates the thunderstorms. As the warm ocean water shifts toward the Eastern Pacific, it also carries the clouds and rainstorms. Thus, rain which is supposed to occur over the Indonesian tropical forest will be transferred over the Peruvian deserts leading to drought in the Western Pacific and heavy rainfall in the South American Pacific.

1.3 La Niña phase

The term La Niña implies "The Little Girl" in Spanish. La Niña is interpreted as anti-El Niño, or simply "a cold event" or, "a cold episode." The cool conditions of La Niña are characterized by a shallow equatorial thermocline in the east and strong trade winds blowing to the west. This causes heat to concentrate in the Western tropical Pacific, which also strengthens both convection and westerly winds that move back to the East (Figure 1.1 b). A strong air circulation in the lower atmosphere is introduced by these conditions.

La Niña causes the cooling of SST periodically every 3 to 5 years in the central and east-central equatorial Pacific. In contrast to El Niño conditions, La Niña conditions are characterized by low SST and produce the climate variations that are opposite of El Niño conditions.

1.4 Neutral (Normal) Phase

In a Neutral phase, tropical winds blow from the East Pacific to West Pacific, which causes warm water ponding, or build up (Figure 1.1 c). Trade winds also pull cold waters to the Central Pacific from the Western Ecuadorian Coast.

1.5 Evolution of ENSO

The terminology ENSO was developed over different historical time periods. The term El Niño was given in the 19th century by a Peruvian fisherman after he noticed warm water during Christmas every few years. In 1969, scientists realized the periodic shift of pressure difference along the equatorial Pacific which was later named the Southern Oscillation (SO). This shift results in different climatic patterns at local and global scales. The SO is the mean sea level pressure difference across the Pacific basin between Tahiti and Darwin located in Ecuador and

Australia, respectively. Thus, El Niño condition is defined as the negative SO Index (SOI) when the surface pressure at Darwin is higher than that of Tahiti. Conversely, La Niña condition is defined as the positive SOI when the surface pressure at Darwin is lower than that of the pressure at Tahiti. In 1969, climatologists realized that El Niño is the oceanic and SO is the atmospheric component of the same phenomenon, and the name ENSO was given to represent the integrated effect of both components.

1.6 ENSO in the Southeast USA

Three states, Alabama, Georgia and Florida experience relatively stronger ENSO effects compared to other states in the Southeast USA (Keener et al. 2010). The El Niño condition in the Southeast USA is characterized by lower temperatures and higher rainfall in winters compared to neutral and La Niña years. La Niña winters are characterized by higher temperature and lesser precipitation (Kiladis and Diaz 1989; Hansen and Maul 1991; Schmidt et al. 2002). El Niño repeats approximately every 3 to 7 years (Rasmusson and Wallace 1983) which is followed by La Niña commencing roughly about a year or so after El Niño. The two successive events may occur within the time interval of 2 to 10 years (Kahya and Dracup 1993).

1.7 Impact of ENSO on Hydrologic Cycle and Water Quality

ENSO has strong teleconnection on surface air temperature and precipitation which has been well reported in several past studies (Chiew et al. 1998; Keener et al. 2007). Both precipitation and temperature are key driving input parameters for water balance over time and space, and therefore, influences the hydrological cycle. Temperature controls evapotranspiration from the land surface, which directly affects the water balance of a watershed. Further, it has a

large influence on in-stream water quality. Since increased temperature also accelerates chemical reactions, it affects biological activity, dissolved oxygen levels, photosynthesis, and build up of toxicity. Temperature of a water body affects the overall water quality.

Similarly, variations in precipitation over daily, monthly, and annual scales influence hydrologic cycle at the watershed scale. Due to this reason, ENSO has been found to have teleconnection with streamflow and flooding besides temperature and precipitation in different parts of the world (Handler 1990; Piechota and Dracup 1996; Chiew et al. 1998; Rajagopalan and Lall 1998; Barsugli et al. 1999; McCabe and Dettinger 1999; Kulkarni 2000; Pascual et al. 2000; Hansen et al. 2001; Roy 2006; Keener et al. 2007). ENSO information is reported in term of ENSO indices using different indicators such as SST and SLP. The National Center for Environmental Prediction (NCEP) of the National Oceanic and Atmospheric Administration (NOAA) can provide reliable SST anomalies and associated ENSO forecasts up to nine months in advance.

In this regard, this research is targeted at using ENSO forecasts issued by NOAA for advancing hydrological sciences and stream water quality protection.

1.8 Objectives and Hypothesis

The overall goal of this research was to conduct a comprehensive study of how climate variability represented by ENSO can be used to solve important water resources problems. In this research, I investigated the application of ENSO information for hydrologic simulation and forecasting, stream water quality protection, and drinking water treatment. The specific research objectives are listed as follows.

1.9 Specific Research Objectives

- (i) To demonstrate that the performance of Adaptive Neuro-fuzzy Inference System (ANFIS) incorporated with SST and SLP is comparable to the watershed model, Loading Simulation Program C++ (LSPC++) in an ENSO-affected watershed.
- (ii) To quantify the long-range streamflow forecasting skill of the ANFIS model by the fusion of SST with predicted climate data from two different approaches (climate model approach, ENSO-conditioned weather generator approach).
- (iii) To demonstrate how ENSO information can be incorporated to improve NPDES permitting system in a complex river system using a watershed linked hydrodynamic and water quality model.
- (iv) To predict TOC loads quantitatively in different ENSO phases using hybrid approach (Principal Component Regression-Artificial Neural Network), and forecast TOC loads qualitatively using a fuzzy logic approach.

1.10 Organization of the Study

Each of the above mentioned objectives are addressed in a separate chapter. Since the dissertation is written in journal paper format, the literature review associated with each objective are presented in the beginning of each chapter. More importantly, since each journal paper requires basic discussion of study area, data and model descriptions, readers may find some redundancy in contents.

In order to make this dissertation concise, coherent and focused, a journal paper which resulted from my PhD work and published online on March 19, 2012 (“Deriving Spatially-Distributed Precipitation Data Using the Artificial Neural Network and Multi-Linear Regression

Models”) in the *Journal of Hydrologic Engineering* ([doi:http://dx.doi.org/10.1061/\(ASCE\)HE.1943-5584.0000617](http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0000617)) has not been included in this dissertation.

In Chapter 2, I present the application of ENSO for streamflow simulation in coastal watersheds. Hydrological modeling of the flat coastal watershed is still challenging as the conventional hydrological modeling does not perform well in, flat terrain, coastal watersheds (Sheridan et al., 2010). In this context, finding a suitable approach of hydrological modeling in ENSO-affected coastal watersheds that can be utilized as a forecasting tool is important for water resources management to effectively deal with hydrologic drought. Since several past studies reported that coastal watershed hydrology in the Southeast USA is driven by SST in the equatorial Pacific, I hypothesize that SST can be utilized for streamflow simulation and forecasting.

First, I discuss the potential teleconnection between streamflow variation and climate variation using various mathematical tools, including wavelet, cross wavelet, wavelet coherence and cross-correlation techniques. Then, I present the application of this linkage into Adaptive Neuro-fuzzy Inference System (ANFIS) for streamflow simulation. In the next step, I evaluate ANFIS model performance against a watershed model, LSPC for streamflow simulations for different historic ENSO events and different seasons. Finally, I compare the performance of the two models for different climatic conditions, seasons and periods using a comprehensive model comparison approach. This research will be submitted for potential publication in the *Journal of Hydrological Sciences*.

In Chapter 3, I present the application of SST and predicted climate data by two different approaches: (i) ENSO-conditioned weather generated data and (ii) climate model (CFSv2) data for streamflow forecasting using ANFIS model at one- to three-month lead time. I post-validate streamflow forecasting results using 7 years of observed data. This research is a part of a continuous research which I want to further evaluate and compare with other approaches that are currently used by the National Weather Service (NWS) River Forecast Centers (RFCs). This paper entitled “Long-Range Hydrologic Forecasting in El Niño Southern Oscillation-Affected Coastal Watersheds: Challenges and Opportunities” will be submitted for potential publication in the *Journal of Hydrologic Engineering*.

In Chapter 4, I present how ENSO information can be used for point source permitting in a complex river system. Since study related with climate variability and its impact on water quality requires long term data sets, I simulate long term (50 years) in-stream water quality using a calibrated and validated watershed model linked with a hydrodynamic and a water quality model. In the next step, I present an extensive statistical analysis to find the link between climate variation due to ENSO and in-stream water quality variation. More importantly, I analyze different critical low streamflow/low DO and high streamflow /DO condition for point source restriction and more point sources assimilation, respectively, to improve the conventional NPDES permitting process. This paper is accepted with revision for publication in the *Transactions of the ASABE*.

In Chapter 5, I discuss the relationship between ENSO phase and watershed scale TOC load. In the first part of this research, I use an LSPC⁺⁺ model to generate long term simulated total organic carbon (TOC) loads data by segregating the effect of land use change from the

climate variability effect. Then, I develop four Principal Component Regression-Artificial Neural Network (PCR-ANN) and four fuzzy logic models, with different model architectures, for accurate estimation of TOC loads. Proper estimation of TOC loads will be helpful in reducing treatment costs associated with disinfection byproduct (DBP) removal. Further, I demonstrate how fuzzy logic approach can be applied for qualitative forecasting of TOC loads. Since there is no consensus among scientists and engineers about the selection of ENSO indices, this research applies linear approach of combining two indices to decide on the best ENSO indices and applies non-linear approach for their application in the models. This paper is currently in review for potential publication in the *Journal of Hydrologic Engineering*.

Chapter 6 briefly presents conclusions derived from this research and recommendations for future work.

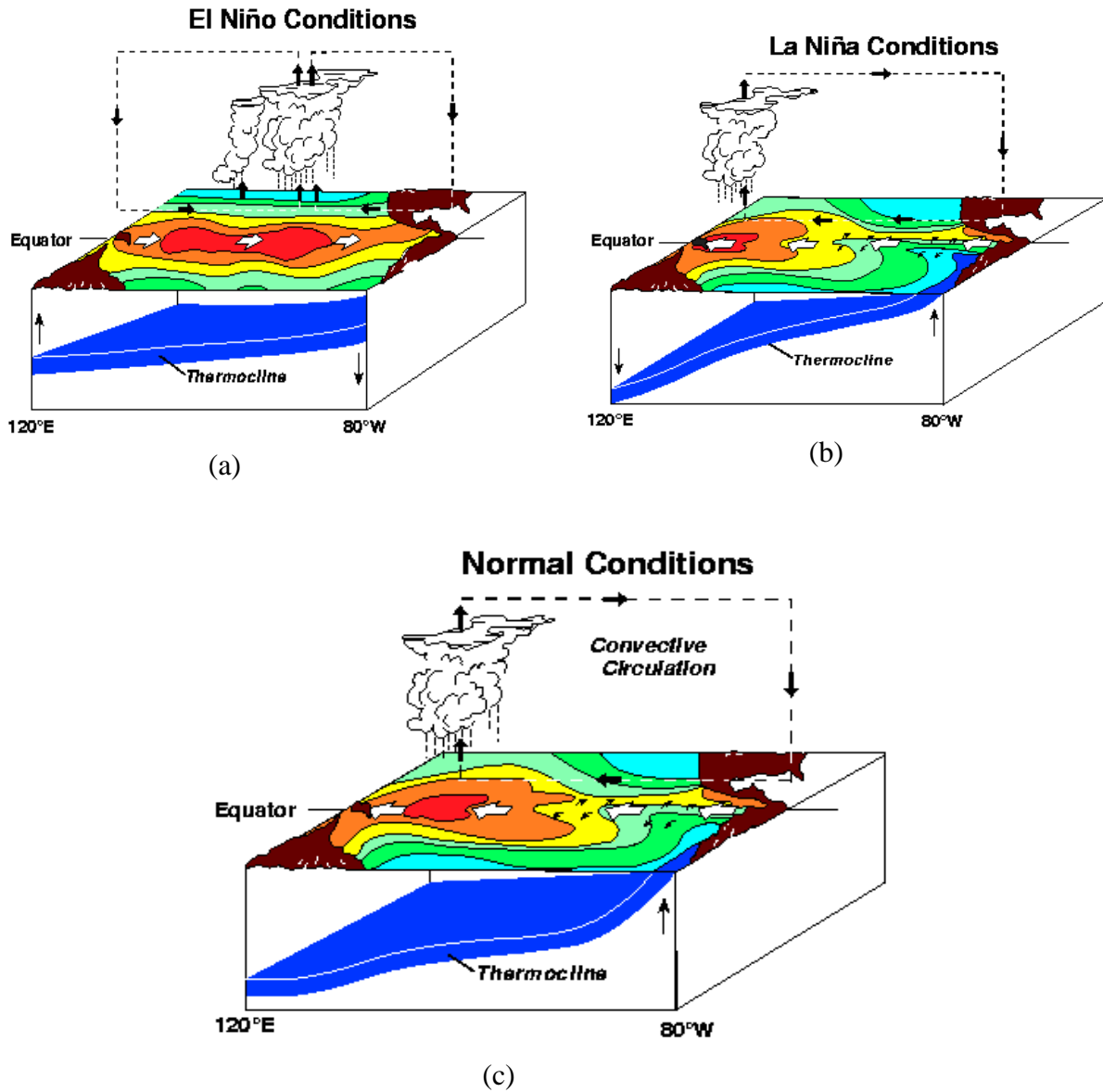


Figure 1.1. Figure describing the different ENSO phases; (a) El Niño phase, (b) La Niña phase and (c) Neutral phase. (Source: http://www.pmel.noaa.gov/tao/elNiño/Niño_normal.html)

1.11 References

- Barsugli, J. J., Whitaker, J. S., Loughe, A. F., Sardeshmukh, P. D., and Toth, Z. (1999). "The effect of the 1997/98 El Niño on individual large-scale weather events." *Bull. Amer. Meteor. Soc.*, 80, 1399–1411.
- Chiew, F., Piechota, T., Dracup, J., and McMahon, T. (1998). "El Niño/Southern Oscillation and Australian rainfall, streamflow and drought: Links and potential for forecasting." *Journal of Hydrology*, 204(1-4), 138-149.
- Handler, A. (1990). "USA corn yields, the El Niño and agricultural drought: 1867–1988." *International Journal of Climatology*, 10(8), 819-828.
- Hansen, D. V., and Maul, G. A. (1991). "Anticyclonic current rings in the eastern tropical Pacific Ocean." *Journal of Geophysical Research*, 96(C4), 6965-6979.
- Hansen, J., Jones, J., Irmak, A., and Royce, F. (2001). "El Niño-southern oscillation impacts on crop production in the southeast United States." *ASA Special Publication*, 63, 55-76.
- Kahya, E., and Dracup, J. A. (1993). "US streamflow patterns in relation to the El Niño/Southern Oscillation." *Water Resources Research*, 29(8), 2491-2503.
- Keener, V., Ingram, K., Jacobson, B., and Jones, J. (2007). "Effects of El Niño/Southern Oscillation on simulated phosphorus loading in South Florida." *Transactions of the ASAE*, 50(6), 2081-2089.
- Keener, V., Feyereisen, G., Lall, U., Jones, J., Bosch, D., and Lowrance, R. (2010). "El-Niño/Southern Oscillation (ENSO) influences on monthly NO₃ load and concentration, stream flow and precipitation in the Little River Watershed, Tifton, Georgia (GA)." *Journal of Hydrology*, 381(3-4), 352-363.
- Kiladis, G. N., and Diaz, H. F. (1989). "Global Climatic Anomalies Associated with Extremes in the Southern Oscillation." *Journal of Climate*, 2, 1069-1090.
- Kirtman, B., Shukla, J., Balmaseda, M., Graham, N., Penland, C., Xue, Y., and Zebiak, S. (2001). "Current status of ENSO forecast skill: a report to the CLIVAR working group on seasonal to interannual prediction."
- Kulkarni, J. (2000). "Wavelet analysis of the association between the southern oscillation and the Indian summer monsoon." *International Journal of Climatology*, 20(1), 89-104.
- McCabe, G. J., and Dettinger, M. D. (1999). "Decadal variations in the strength of ENSO teleconnections with precipitation in the western United States." *International Journal of Climatology*, 19(13), 1399-1410.

- Pascual, M., Rodó, X., Ellner, S. P., Colwell, R., and Bouma, M. J. (2000). "Cholera dynamics and El Niño-southern oscillation." *Science*, 289(5485), 1766.
- Piechota, T. C., and Dracup, J. A. (1996). "Drought and regional hydrologic variation in the United States: Associations with the El Niño-Southern Oscillation." *Water Resources Research*, 32(5), 1359-1373.
- Rajagopalan, B., and Lall, U. (1998). "Interannual variability in western US precipitation." *Journal of Hydrology*, 210(1-4), 51-67.
- Rajagopalan, B., and Lall, U. (1999). "A k-nearest-neighbor simulator for daily precipitation and other weather variables." *Water Resources Research*, 35(10), 3089-3101.
- Rasmusson, E. M., and Wallace, J. M. (1983). "Meteorological aspects of the El Niño/southern oscillation." *Science*, 222(4629), 1195.
- Roy, S. S. (2006). "The impacts of ENSO, PDO, and local SSTs on winter precipitation in India." *Physical Geography*, 27(5), 464-474.
- Schmidt, N., and Luther, M. E. (2002). "ENSO impacts on salinity in Tampa Bay, Florida." *Estuaries and Coasts*, 25(5), 976-984.
- Tang, Y., Kleeman, R., and Moore, A. M. (2005). "Reliability of ENSO dynamical predictions." *Journal of the atmospheric sciences*, 62(6), 1770-1791.

Chapter 2. Climate Variability-Based Streamflow Simulations Using Neuro-Fuzzy Computational Techniques

2.1 Abstract

The climate variability manifested by the coupled oceanic and atmospheric phenomenon, El Niño Southern Oscillation (ENSO), has strong predictable effects on temperature and precipitation. Therefore, ENSO has been found to have a strong teleconnection with streamflow and flooding in different parts of the world. ENSO phase and strength are commonly reported by various ENSO indices using sea surface temperatures (SSTs) and sea level pressures (SLPs) in different regions of the equatorial Pacific. Since ENSO has potential to change streamflow, it is important to examine the correlation of streamflow with SSTs and SLPs, and then translate these correlations into a mathematical model to better represent the effect of ENSO on streamflow. The hypothesis of this research is that a model will better simulate streamflow in different ENSO phases if SSTs and SLPs are explicitly utilized in rainfall-runoff modeling. This hypothesis was tested by developing an Adaptive Neuro-Fuzzy Inference System (ANFIS) model and evaluating its performance against a watershed model, Loading Simulation Program C⁺⁺ (LSPC). First, the correlation of streamflow with SSTs and SLPs was studied using continuous wavelet, cross wavelet, wavelet coherence analysis, and cross correlation techniques in the Chickasaw Creek watershed of South Alabama. Then, SSTs and SLPs were used in ANFIS model for streamflow simulations. The model was trained, validated and tested using 50 years of observed data. The statistical parameters to measure the performance of the model suggested that the ANFIS model simulated streamflow were comparable to the LSPC model simulated streamflow over the 50

year simulation period. In addition, the performance of the ANFIS model to simulate the hydrologic drought and high flows corresponding to ENSO phases were as good as LSPC model. This approach of hydrologic simulation demonstrates the potential of neuro-fuzzy computational technique for improved streamflow simulations because of the direct application of observed SSTs and SLPs (in the equatorial Pacific).

2.2 Introduction

A large scale coupled oceanic and atmospheric phenomenon resulting into an oscillation in the equatorial Pacific ocean, El Niño Southern Oscillation (ENSO) has significant influences on inter-annual variation in temperature, precipitation and streamflow of different parts of the world (Piechota and Dracup 1996; Chiew et al. 1998; Rajagopalan and Lall 1998; Barsugli et al. 1999; McCabe and Dettinger 1999; Kulkarni 2000; Pascual et al. 2000; Hansen et al. 2001; Roy 2006; Keener et al. 2007). Further, the impact of ENSO on water resources (Thomson et al. 2003; Keener et al. 2007) and its prospective to create the temporal variation in annual streamflow (Stahl and Demuth 1999; Mosley 2000) have been reported in different past studies.

ENSO is a non-stationary, dominant pattern of short-term climate variations and has recently been the subject of scientific interest. Different ENSO phases are identified using different ENSO indicators, such as, Sea surface temperature (SSTs) and associated sea level pressures (SLPs) in the equatorial Pacific Ocean. By identifying the potential teleconnection between streamflow and these indicators, it may be possible to explicitly utilize these indicators into rainfall-runoff modeling to better capture the impact of ENSO on streamflow. Streamflow simulation is one of the most important topics due to its wide applications in different hydrological issues, such as, flood and drought forecasting, design of hydraulic structures, and

water quality modeling. Further, streamflow simulation is a complicated process, which is greatly influenced by the selection of linear and non-linear parameters (Singh 1988). Accordingly, various approaches and methods have been proposed for streamflow simulations at different historical time period.

These approaches can be categorized into process-based models, conceptual models, or data-driven/black-box models (Jayawardena et al. 2006). Process-based models include distributed and semi-distributed watershed models. Conceptual models include those models in which water balance dynamics are represented by empirical functions (Jayawardena et al. 2006), whereas, data-driven models are the black-box type models and lack an explicit representation of the governing physical processes. Some of the data-driven models are Auto Regressive Integrated Moving Average (ARIMA) (Box et al. 1970), Artificial Neural Network (ANN) (ASCE 2000a, b), and soft computing techniques including the fuzzy logic approach (Zadeh 1965).

Different studies in the past have reported that none of these approaches are considered superior to another because none of the approaches can be expected to outperform all other approaches for all type of watersheds under all conditions (Jayawardena et al. 2006). Rather, the selection of the modeling approach depends on a number of factors, such as, the availability of data, objective of the study, cost, watershed type, and users' knowledge and expertise (Jayawardena et al. 2006; Srivastava et al. 2006). However, data-driven models have been widely accepted and applied for hydrological modeling (Srivastava et al. 2006) due to their simplicity and user friendly nature.

Over the years, researchers have found some limitations of the traditionally-adopted data-driven models as well. Although data-driven computational techniques, such as, ANN and fuzzy

logic offer advantage over conventional modeling (Nayak et al. 2004) and have been proven to be effective, recently, scientists are showing their interest in combining these two approaches. As a matter of fact, neuro-fuzzy system known as Adaptive Neuro-Fuzzy Inference System (ANFIS) (Jang 1993) has been developed combining ANN and fuzzy logic approaches. When combined together, the individual strengths of both ANN and fuzzy logic models is further enhanced resulting into powerful intelligent systems. This study explores the ANFIS model to simulate the streamflow using potential ENSO indicators of the equatorial Pacific. Readers can find a few past studies that have utilized SSTs in data-driven modeling for streamflow and drought predictions (Khalil et al. 2005; Farokhnia et al. 2011).

Since data-driven models are black boxes in nature without explicitly knowing any physics involved in it, the process becomes further black box when the inputs are selected without understanding the physical significance of the inputs or without sensitivity analysis (Hearty and Gibney 2008). Hence, this research is targeted to exploring mathematical tools, for example, continuous wavelet analysis, cross wavelet analysis, wavelet coherence and cross correlation techniques, to identify physically-significant model inputs so that the black box nature of the data-driven modeling can be eliminated to some extent.

Since a majority of past studies were limited mainly to identify the correlation of ENSO with streamflow and hydroclimatic variables (Xu et al. 2005; Hendon et al. 2007), the novel contribution of this research is that it attempts to find linkage between climate and streamflow variations, and then explicitly incorporates that linkage into a Neuro-fuzzy computational technique. I hypothesize that by integrating SSTs and SLPs with other climatic variables, models would adequately represent streamflow variations corresponding to different ENSO phases (El Niño, La Niña and Neutral). This hypothesis is supported by different past studies that

demonstrated impact of ENSO on water resources (Thomson et al. 2003; Keener et al. 2007). Besides, I will evaluate the response of the ANFIS model against the historic La Niña and El Niño events.

Because ANFIS model has been successfully applied in several hydrologic (Mukerji et al. 2009; Pramanik and Panda 2009; Yan et al. 2010) and water quality modeling (Yan et al. 2010) studies, it can be inferred that it has a good potential for streamflow simulations. In addition, I compare the performance of ANFIS model with a watershed model for different ENSO events. Therefore, the research objectives are to: (1) identify the potential connection between streamflow variation and climate variation, (2) incorporate this connection (SST, SLP) into an ANFIS model to develop a rainfall-runoff model, and (3) evaluate the ANFIS model performance against a watershed model, LSPC for streamflow simulations corresponding to different historic ENSO events.

2.3 Theoretical Background

ENSO, which is measured using various ENSO indicators (SST, SLP), is considered as one of the most reliable phenomenon for relating inter-annual climate variability in streamflow at the local as well as global scales (Ropelewski and Halpert 1986). In order to confirm the linkage between observed hydrological time series with SLP and SST anomalies (ENSO time series) and also to determine whether the connection between the two series is stationary or recurrent, I experimented with the wavelet analysis, continuous wavelet transform (CWT), cross-wavelet transform (XWT) and wavelet coherence (WTC) analyses (Grinsted et al. 2004) for ENSO time series and the observed hydrological time series (streamflow). Wavelet analysis is a process of

disintegrating a time series into time and frequency space to see the time-frequency distribution between two given time series and to analyze how their relationship changes over a time period. The basic theories about the wavelet analysis are briefly discussed in the following section.

2.3.1 Wavelet Analysis

Statistically, hydrological time series are non-stationary (Coulibaly and Baldwin 2005). The series may comprise of prevailing periodic signals that can vary both in amplitude and frequency during a historical time period. In wavelet analysis, relationship between two time series is examined to determine both the prevailing modes of variability and their mode of variation over a time period. It allows for the identification of dominant localized variations of powers, and also explains the period when the variance between two time series is highest for a given frequency. The real motive of applying the wavelet analysis technique is to compute and envisage the statistical changes in the SST anomalies and streamflow during historical time period. Wavelet analysis is better than the commonly used Fourier transform for the analysis of a signal in frequency space at a global scale because former is scale independent, efficient and accurate to accomplish an analysis of non-stationary hydrological time series data sets such as SST and SLP (Torrence and Compo 1998). Using the wavelet analysis, the frequency of occurrence and the amplitude of ENSO phases (El Niño, La Niña and Neutral) can be detected over a multi-decadal time period. Details about the wavelet analysis are available in literatures (Torrence and Compo 1998; Keener et al. 2010).

The application of CWT are preferable, especially for a time series that does not follow the normal distribution (Grinsted et al. 2004). Hydrological time series fall under this category. The XWT was constructed using two CWTs that depict their mutual power and phase

relationship in time and frequency space. In addition, I will define and describe WTC which determines the local correlation between two times series despite the low sharing power.

2.3.2 Continuous Wavelet Transforms (CWT)

The CWT transforms a time series into time and frequency space, and also analyzes localized recurrent oscillations in the time series. It is preferable to inspect time series that are linked together somehow. For this, we can characterize any wavelet function $\mu_0(n)$ for any time series, x_n ($n = 0 \dots N-1$) for time spacing δt , having zero mean and localized in time and frequency space. The choice and selection of the wavelet function depends upon the data series. The Morlet wavelet function is used for the analysis. This function depends on a non-dimensional time parameter, η and the non-dimensional frequency, w_o (default value 6).

$$\psi_0(\eta) = \pi^{-1/4} e^{iw_o n} e^{-\frac{\eta^2}{2}} \quad (2.1)$$

The continuous wavelets transform $W_n(s)$ of a discrete sequence x_n , which is a translated version of $\psi_0(\eta)$ is given by the following equation.

$$W_n(s) = \sum_{n'=0}^{n-1} X_{n'} \psi * \left[\frac{(n'-n)\delta t}{s} \right] \quad (2.2)$$

where s is the wavelet scale, n is the localized time index, n' is the translated time index, ψ is the normalized wavelet, and $(*)$ is the complex conjugate. Here, the null hypothesis is that the signal is created by a static process for a given wavelet spectrum (P_k), and the statistical significance of wavelet power can be derived for this null hypothesis. Since time series are generally autocorrelation in characteristics, these can be better simulated by a first order autoregressive (AR1) method (Grinsted et al. 2004). Following is the equation given for the Fourier power spectrum of AR1 (Allen and Smith 1996).

$$P_k = \frac{1 - \alpha^2}{|1 - \alpha e^{-2i\pi k}|^2} \quad (2.3)$$

where α is autocorrelation at lag-1 and k is a Fourier frequency index. For continuous wavelet analysis, the codes from <http://paos.colorado.edu/research/wavelets/> were used.

2.3.3 Cross Wavelet Transform (XWT)

Although high common power can be seen in the wavelet power spectra, cross-wavelet analysis will: (i) find a distinct region due to the direct analysis of two series with high shared power; and (ii) demonstrate the phase correlation (positive or negative). For two given time series, X and Y , with different wavelet transforms $W_n^X(s)$ and $W_n^Y(s)$, the cross wavelet transform is defined as

$$W_n^{xy}(s) = W_n^X(s)W_n^{Y*}(s) \quad (2.4)$$

where $*$ represents the complex conjugate. The cross-wavelet power can be defined as $|W_n^{XY}|$. XWT scans the regions in time-frequency space to assess a reliable relationship between the time series with high shared power. Theoretically, the cross wavelet power for two given time can be mathematically represented as follows (Torrence and Compo 1998).

$$D\left(\frac{|W_n^x(s) W_n^{y*}(s)|}{\sigma_x \sigma_y} < p\right) = \frac{Z_\nu(p)}{\nu} \sqrt{p_k^x p_k^y} \quad (2.5)$$

where p_k^x p_k^y and $Z_\nu(p)$ are the theoretical Fourier spectra and the level of confidence associated with the probability p for a pdf, which is represented by the Chi-square value. Similarly, σ_x and σ_y are the respective standard deviations and $\nu = 1$ for real wavelets, and $\nu = 2$ for complex wavelets.

2.3.4 Wavelet Coherence (WTC)

In addition to the cross-wavelet transform, wavelet-coherence transform are required to evaluate the local co-variance of the two time-series. The WTC indicates the regions where two time series are connected in time frequency space but may or may not have shared high power. Since XWT will lose the significance to visualize the shared power, WTC will find larger significant areas compared to XWT. The wavelet-coherence transform for two given time series (Grinsted et al. 2004) is given as:

$$R_n^2(s) = \frac{|S(s^{-1}W_n^{xy}(S))|^2}{S(S^{-1}|W_n^x(S)|^2) \cdot S(S^{-1}|W_n^y(S)|^2)} \quad (2.6)$$

where S is the smoothing operator for the given type of wavelet function. This equation closely represents a correlation coefficient indicating that the wavelet coherence is actually a localized correlation in time-frequency space. Monte Carlo methods were used for the evaluation of the statistical significance level for the wavelet coherence. The complete package for continuous wavelet analysis, XWT and WTC analysis were implemented using MATLAB code developed by Aslak Grinsted (<http://www.pol.ac.uk/home/research/waveletcoherence/>).

2.3.5 Cross-correlation Analysis

Cross correlation is a standard method of estimating the degree of correlation between the two series. The lagged cross-correlation between SST anomalies and the hydrological time series can be computed using following equation.

$$R = \frac{\sum_i^N [(x(i) - X_m) * (y(i-j) - y_m)]}{\sqrt{\sum_i^N (y(i-j) - y_m)^2} \sqrt{\sum_i^N (x(i) - x_m)^2}} \quad (2.7)$$

where N is the length of the time series, $x(i)$ and $y(i)$ are the two series for SST anomalies and observed streamflow at different time lags ($j=0, \dots, N-1$), and x_m and y_m are the corresponding means. The cross correlation of SST with streamflow and SST with precipitation was determined. This mathematical relationship was utilized to develop an ANFIS model. The basic concept of the ANFIS model is discussed in the following section.

2.3.6 Adaptive Neuro Fuzzy Inference System (ANFIS)

2.3.6.1 Concept

Many studies in the past have demonstrated that intelligent computational technique such as ANN and fuzzy logic approaches are competent for hydrological modeling though each approach has its own limitations. For example, the ANN lacks the explanatory power; similarly, the fuzzy logic approach is more intuitive and relies on user's expertise for the selection of the fuzzy set and membership function. Recently, combining these approaches, such as, Adaptive Neuro-Fuzzy Inference System (ANFIS) have been tried to eliminate shortcomings of individual approaches. Due to its capability of combining the verbal aspects of a fuzzy system with the quantitative aspect of a neural network, ANFIS has been found to be a more efficient modeling tool than the two independent models (i.e., ANN and fuzzy logic) to capture inherent non-linear processes (Jang 1993). Hence this model has been extensively applied in hydrological (Mukerji et al. 2009; Pramanik and Panda 2009) and water quality modeling (Yan et al. 2010).

ANFIS is a multi-layer feed-forward network that utilizes neural network technique and fuzzy logic to map an input vector to an output vector by using back propagation or combined algorithm. Input-output vector is expressed in a fuzzy inference system (FIS) by using number of fuzzy IF-THEN rules which designates the local response of the input and output mapping.

Several types of fuzzy inference system have been proposed, depending upon the fuzzy reasoning and fuzzy rules applied (Mamdani and Assilian 1975; Sugeno and Takagi 1983). Readers may refer an article by Jang (1993) for the detail on the model structure.

A generic example of first degree Sugeno FIS using two inputs and one output is demonstrated in the Figure 2.1. The architecture of ANFIS comprises of five layers with two rules, and two membership functions (MFs) for each input (Figure 2.1).

The Sugeno-fuzzy models comprising two fuzzy if-then rules can be written as follows.

$$\text{Rule 1: If } X \text{ is } P_1 \text{ and } Y \text{ is } Q_1, \text{ then } f_1 = a_1 X + b_1 Y + r_1$$

$$\text{Rule 2: If } X \text{ is } P_2 \text{ and } Y \text{ is } Q_2, \text{ then } f_2 = a_2 X + b_2 Y + r_2$$

where, P_1 and P_2 are the MFs for input X and Q_1 and Q_2 are the MFs for input Y . a_1, b_1, r_1 and a_2, b_2, r_2 are the parameters for the output function.

The brief operation of ANFIS model is explained in the following section.

Layer 1 In this layer, each node with membership functions is called an adaptive node.

$$O_i^1 = \mu_{P_i}(X), \text{ for } i = 1, 2 \quad (2.8)$$

$$O_i^1 = \mu_{Q_{i-2}}(Y), \text{ for } i = 3, 4 \quad (2.9)$$

Here, X and Y are crisp inputs and P_i and Q_{i-2} are fuzzy sets (subjective or qualitative label such as ‘very low’, ‘low’ or ‘high’) related with these nodes, categorized by the appropriate MFs.

The MFs could be any functions such as Gaussian, bell-shaped, generalized bell-shaped, triangular and trapezoidal-shaped. For simplicity, in this example, a generalized bell-shaped built-in membership function is used to compute output.

$$O_i^1 = \mu_{pi}(X) = \frac{1}{1 + \left| \frac{X - c}{a} \right|^{2b}} \quad (2.10)$$

where a is the starting point of the curve, b is generally a positive value and the parameter 'c' is associated with the center of the curve.

Layer 2 In this layer, the incoming signals are multiplied using operator (AND, OR) to obtain a single crisp output representing the result of the predecessor (firing strength) of that rule. Hence, the outputs O_i^2 of this layer can be expressed as the products of those corresponding layer.

$$O_i^2 = w_i = \mu_{Pi}(X)\mu_{Qi-2}(Y), \quad \text{for } (i) = 1, 2 \quad (2.11)$$

Layer 3 In this layer, every node is labeled as N , and the ratio of firing strength corresponding to each i^{th} rule's to the summation of firing strength that corresponds to each rules' is computed.

$$O_i^3 = \bar{w}_i = \frac{w_i}{w_1 + w_2} \quad (2.12)$$

where $i = 1, 2$

Layer 4 In this layer, the input contribution of each i^{th} rule for the output of the model is computed using the following equation.

$$O_i^4 = \bar{w}_i f_i = \bar{w}_i \sum_{i=1}^2 (a_i X + b_i Y + r_i) \quad (2.13)$$

where \bar{w} is the output from layer 3 and $[a_i, b_i, r_i]$ are the model parameter.

Layer 5 The final output is computed as the combination of entire receiving signals

$$O_i^5 = \sum_i^2 \bar{w}_i f_i = \frac{\sum_i^2 w_i f_i}{\sum_1^2 w_i} \quad (2.14)$$

The objective of this approach is to train adaptive networks in order to derive unidentified functions using training data, and determine the accurate value of the parameters. ANFIS model applies a hybrid-learning algorithm (combination of the gradient descent method and the least-square method) to update the model parameters.

2.3.6.2 Estimation of Parameters

Based on the simple architecture of ANFIS model (Figure 2.1) (designed for the given input parameters), the ANFIS structure can be further extended to several other inputs. For the present study, $SF = f(SST, SLP, T, PCP)$, where, SF, PCP and T stand for streamflow, precipitation and temperature, respectively. For simplicity, an example of two inputs with two rules is presented in this manuscript. Assuming P1:P2 and Q1:Q2 are the respective membership functions for inputs (SST and precipitation (PCP)) and a1:b1, a2:b2 are the parameters associated with the output function, resultant output can be computed as linear combinations of subsequent parameters.

The equation is simplified as follows.

$$SF = \bar{w}_1(SST)a1 + (\bar{w}_1 PCP)b1 + \bar{w}_1 r1 + (\bar{w}_2 SST)a2 + (\bar{w}_2 PCP)b2 + \bar{w}_2 r2 \quad (2.15)$$

The equation can be written in Matrix form as follows.

$$A = \begin{bmatrix} \overline{w_1} SST_1 & \overline{w_1} PCP_1 & \overline{w_1} & \overline{w_2} SST_1 & \overline{w_2} PCP_1 & \overline{w_2} \\ \overline{w_1} SST_2 & \overline{w_1} PCP_2 & \overline{w_1} & \overline{w_2} SST_2 & \overline{w_2} PCP_2 & \overline{w_2} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \overline{w_1} SST_n & \overline{w_1} PCP_n & \overline{w_1} & \overline{w_2} SST_n & \overline{w_2} PCP_n & \overline{w_2} \end{bmatrix}$$

$$Z = \begin{bmatrix} a_1 \\ b_1 \\ r_1 \\ a_2 \\ b_2 \\ r_2 \end{bmatrix}, \text{ and } B = \begin{bmatrix} SF_1 \\ SF_2 \\ SF_3 \\ \cdot \\ \cdot \\ \cdot \\ SF_n \end{bmatrix}$$

$$AZ = B \tag{2.16}$$

where SF stands for streamflow and Z is an unknown matrix, derived from the parameter sets, which can be solved using following equation.

$$Z = (A^T A)^{-1} A^T B \tag{2.17}$$

where A^{-1} is the inverse of matrix ‘‘A’’ and A^T is the transpose of the matrix ‘‘A’’. There are two frequently used training methods to solve this matrix: (1) back-propagation (BP) algorithm (Bishop 1995), and (2) hybrid learning algorithm (Jang 1993). For this analysis, I used the hybrid learning algorithm combining both the least-squares method and the BP algorithm. The hybrid approach trains the data rapidly, and also converges much faster. Readers can refer to the article by Jang et al. (1997) for the mathematical description of the hybrid learning algorithm.

2.3.6.3 Subtractive and Fuzzy C -Mean Clustering for Structure Identification

For large datasets and good quality data covering the wide variations in the feature space, fuzzy systems can have better generalization due to the use of more fuzzy rules. However, for a small range of datasets, which is more common in streamflow, large number of rules cannot be

derived from the training data as it may easily over fit the system and rule out the possibility of generalization. Hence, clustering approach is required for both the effective partition of the input space and the reduction of the number of rules. Several clustering methods have been suggested to organize the data and construct the rules. Some of the widely used approaches are grid partition (Jang and Sun 1995), subtractive fuzzy clustering (Chiu 1996) and fuzzy c-mean (FCM) (Dunn 1973). In this analysis, both FCM and subtractive clustering algorithm (SCA) were experimented. FCM is a technique in which each data point is categorized into a cluster based on membership function (Bezdek 1981). It divides the data from multidimensional space into a particular cluster number. Cluster center and the membership function were decided through the repeated iteration by minimizing the objective function which was defined as a distance between data point and cluster center. SCA is more suitable when users do not have a clear idea of cluster number required to be used for a given data sets. The subtractive clustering algorithm (Chiu 1996) was introduced to generate a FIS using minimum number of rules. This algorithm is based on the amount of the density points in the data space, and the approach is to find areas in the data space having higher densities of data points. The algorithm tries to find the point with the highest number of neighbors as the center of the cluster. For a given set of data points from X_1, \dots, \dots, X_n , the density measure at a data point X_j is given by the following equation.

$$d_j = \sum_{i=1}^n e^{-\frac{\|X_j - X_i\|^2}{(\frac{r_a}{2})^2}} \quad (2.18)$$

where r_a is a data cluster radius. The equation indicates that those data points which have high neighboring data points will be considered as a cluster center. The data used for this study and study area are discussed in the following section.

2.4 Methodology

2.4.1 Study Area and Data

This study was conducted in the Chickasaw Creek watershed (Figure 2.2) of Mobile County of Southern Alabama (USA) in the Mobile River Basin. The watershed is 714 km² in size and starts at Citronelle, AL in north, and eventually drains into Mobile Bay. The watershed is dominated by the coastal plain geology with maximum elevation range of 13.11 m to minimum 0 m above the mean sea level. The annual precipitation in the watershed (1651 mm) is relatively higher than in the other parts of Alabama. Air temperature and precipitation datasets were available from the closest weather station, the Mobile Regional Airport (Coop ID-015478), downloaded from National Climatic Data Center (NCDC). SST and SLP difference (Tahiti and Darwin) were available since 1950, and downloaded from NOAA's NCDC and National Center for Environment Prediction (NCEP) website. Streamflow data, recorded since 1952, were available at the USGS gage (02471001) which drains 357 km² of the watershed. The data used in the study with their sources and formats are reported in Table 2.1.

2.4.2 Input Data Selection

Selection of suitable Membership Functions (MFs) with proper number and rules are essential for a successful mapping of input-output relationship in FIS. Besides, the selection of input variables is a crucial part of ANFIS modeling. Optimum inputs should be selected such that it best captures the input and output relation. Accordingly, the cross correlation technique and wavelet analysis were experimented to decide the appropriate inputs. The cross correlation of surface runoff (SR) and base flows (BF) with SST is demonstrated in Figure 2.3. Similarly, the wavelet analysis suggested that both SST and SLP could be significant model inputs for

streamflow simulation (Figure 2.4), particularly for the ENSO affected region. The relationship between streamflow with SST and SLP are shown in Figure 2.5 and Figure 2.6 through the XWT and WTC. The detail findings from wavelet analysis will be presented in “result and discussion” section of this manuscript. The potential inputs selected for ANFIS model development are rainfall, SST, SLP, surface air temperature, land use and soil.

Since streamflow is an integral form of quick response of the watershed i.e. surface runoff and the slow response components i.e. base flow, it is always preferable to partition the streamflow into base flow and surface runoff components (Kalin and Hantush 2006). Accordingly, I used the base flow filter program (<https://engineering.purdue.edu/~what/>), developed at Purdue University for online hydrograph separation, and developed the separate SR and BF model. The model inputs for surface runoff model were “estimated surface runoff”, temperature, SST and SLP at different lead times. The estimated surface runoff was computed from precipitation using the SCS curve number equation (Bosznay 1989). The Curve Number (CN) was estimated based on the SSURGO soil (SSURGO, 2010) data base and 2001 land cover data set (NLCD, 2010). Likewise, the model inputs for base flow model were SST, SLP, air temperature and precipitation at different lead times.

In the next step, different inputs for summer (May-Nov) and winter (Dec-April) in Multi Linear Regression (MLR) model were experimented to identify the significant model inputs. Though wavelet analysis and cross correlation technique suggested the streamflow variation due to variation in SST and SLP over a certain period of time, it was important to further evaluate whether SSTs and SLPs were the significant model inputs besides temperature and precipitation or not. In fact, precipitation and temperature are also affected by SST and SLP. For this, MLR model was used, and the model depicted that the parameters were significant at 90% confidence

interval (P value < 0.1). In addition, the sensitivity analysis was carried out, and the result suggested that SST and SLP were some of the most sensitive model inputs. The optimal lead time and number of inputs were decided after the sensitivity analysis. Various combinations of inputs were experimented at different lead times to determine the best inputs. Any additional inputs beyond the optimum sets did not improve the model performance.

The analysis suggested that both SST and SLP data were important, especially when the ENSO shows the clear signature with streamflow (winter, spring and early fall). Since the same sets of inputs for two seasons (winter and summer) were obtained after sensitivity analysis, the single model was developed representing for winter and summer seasons because large training data sets would be beneficial to improve the model performance. However, each model consists of the separate SR and BF model.

2.4.3 Model Development and Implementation

The most sensitive sets of inputs giving the best result in MLR model were employed into ANFIS model. The generalized equation in ANFIS model for surface runoff model is given by equation 2.20.

$$SR(t) = f(SST(t), SLP(t), T(t), SR_E(t), SR_E(t-1)) \quad (2.20)$$

where, SR is surface runoff, t is monthly time step, SR_E is estimated surface runoff using SCS curve number method based on spatially averaged precipitation, land use/land cover and soil. The model inputs for both SR and BF model are presented in Table 2.2.

MATLAB fuzzy logic toolbox was selected for ANFIS simulation using 50 years of observed monthly data; 80% of the total data was used for model training, 10% for model validation and 10% for model testing. Both SCA and FCM clustering method were experimented

for the selection of the initial model parameters. For this particular study, SCA was found to be a best approach to cluster the data sets. The parameter to be fixed for SCA, were clustering radius, acceptance ration, rejection ration. I varied the clustering radius from 0.3 to 1 using 0.05 as step size to determine the optimal parameters which were ascertained by minimizing the root mean squared error. Other parameters such as the membership function, range of influence of the cluster center, acceptance ratio, and rejection ratio of the model were optimally ascertained through a repeated trial and error. “Generalized bell-shaped” built-in membership function was found to be appropriate for each fuzzy set in the fuzzy system after experimenting different membership function types such as Gaussian, trapezoidal, triangular, and sigmoidal. The parameters of the membership function were fixed through the application of hybrid learning algorithm. The number of epochs required for ANFIS training was optimally determined (190 for BF model and 80 for SR model). In order to avoid the overtraining of the data, the model simulation was stopped when the validation error started increasing. This was confirmed by evaluating its performance against checking data sets and using proper epoch number.

The model performance of each ANFIS model was measured through the three non-dimensional measures such as R-square, Nash-Sutcliffe efficiency, *NSE* (Nash and Sutcliffe 1970), the root mean square error (*RMSE*) and mass balance error (*MBE*). The detail descriptions about the statistical parameters to measure the model performance are available and given in many articles (Moriassi et al. 2007). The mathematical expressions for the relative goodness of fit statistics are given as follows.

$$R^2 = \left[\frac{\sum_{i=0}^n (Q_{obs,i} - \bar{Q}_{obs}) (Q_{sim,i} - \bar{Q}_{sim})}{\sum_{i=0}^n (Q_{obs,i} - \bar{Q}_{obs})^{-0.5} \sum_{i=0}^n (Q_{sim,i} - \bar{Q}_{sim})^{-0.5}} \right]^2 \quad (2.21)$$

R^2 varies from 0 to 1 which indicates the proportion of the total variances in the observed data. The higher value indicates the higher degree of collinearity.

$$NSE = 1 - \frac{\sum_{i=1}^n (Q_{sim,i} - Q_{obs,i})^2}{\sum_{i=1}^n (Q_{obs,i} - \bar{Q}_{obs})^2} \quad (2.22)$$

NSE is a measure of how well the observed and simulated data fits and its coefficient varies from $-\infty$ to 1. The perfect model has a NSE value of 1.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Q_{obs,i} - Q_{sim,i})^2}{n}} \quad (2.23)$$

$RMSE$ is a measure of how simulated data are close to the observed data.

$$MBE = \frac{(\bar{Q}_{sim} - \bar{Q}_{obs})}{\bar{Q}_{obs}} \quad (2.24)$$

MBE gives the percentage bias between simulated and observed data. Here, $Q_{obs, i}$ and $Q_{sim,i}$ are observed and simulated streamflow for each i^{th} observation and N is the number of observations. Similarly, \bar{Q}_{obs} and \bar{Q}_{sim} are the mean observed and simulated streamflow. Besides, the performance of ANFIS model for streamflow simulation was compared with watershed model.

2.4.4 Comparison with Watershed Model

The performance of the model was further evaluated by comparing the performance of ANFIS model with watershed model (LSPC++) (Shen et al. 2005). The modeling paradigms in two models are completely different. ANFIS is based on the learning approach, which is accomplished through the known set of association between inputs and output datasets used for model training, whereas, LSPC is based on physical processes related to the rainfall input and its transformation into runoff at the watershed outlet. In rainfall-runoff modeling, the objective is to

emulate the phenomenon such that rainfall is converted into runoff. The brief description of the LSPC model and its application in Chickasaw Creek watershed is explained in the following section.

2.4.5 Loading Simulation Program C⁺⁺ (LSPC) Model

The Loading Simulation Program in C⁺⁺ (LSPC), a C⁺⁺ version of the widely used Hydrologic Simulation Program-Fortran (HSPF) (Bicknell et al. 2001), is a watershed model for simulating streamflow and stream hydraulics. Though the LSPC's algorithms are identical to those of the HSPF, LSPC model is more efficient and flexible. The model has been considered as one of the most advanced hydrologic and watershed loading model. The hydrologic portion of the model is based on the Stanford Watershed Model (Crawford and Linsley 1966). The hydrologic processes in LSPC model is conceptualized with the schematic diagram as shown in Figure 2.7 (USEPA 2009). The diagram shows the water budget diagram with the water interchanging process from surface to subsurface including three flow path; surface, interflow and groundwater flow.

Precipitation is the input to the system (watershed) which is partitioned at the first decision node for the interception and water reached on the land surface using the parameter CEPC (Figure 2.7). In the next decision node, surface and lower zone storage will be divided using the parameter, INFILT. The surface storage contributes to the three components: i) upper zone storage which partly contribute to lower zone, partly to the groundwater/inactive groundwater storage, and rest to the ET, ii) interflow storage, which eventually contributes to the channel flow, iii) overland flow. The lower zone storage may partly contribute to groundwater storage or inactive groundwater storage (deep percolation) or partly lost through

evapotranspiration. The contribution from upper zone and lower zone storage to ground water is divided into active and inactive groundwater by using a coefficient (DEEPPFR). Active ground water storage can contribute to streamflow as base flow and losses as ET. ET can occur from the entire zone except inactive groundwater zone. The overland flow, interflow, and base flows eventually contribute to streamflow.

2.4.6 LSPC Model Configuration

Hydrological processes in LSPC model are governed by certain hydrological parameters. Watershed parameters such as length, slope, stream network and sub-watershed area were determined through the interactive watershed delineation using high resolution (10 m) Digital Elevation Model (DEM) (DEM 2010). Land use and soil-related parameters were extracted using Land cover data set of year 2001 (NLCD 2010) and high resolution soil data (SSURGO) (SSURGO 2010), respectively. LSPC model was configured to simulate a series of hydrologically connected sub-watersheds, with defined geometry, soil and land use characteristics. Each sub watershed area contributed runoff to their respective reach, where the cumulative flow was routed downstream to the watershed outlet. LSPC model utilizes different sets of hydrologic parameters for surface and subsurface hydrologic analyses in different sub watersheds based on the soil type and land use categories. Chickasaw Creek watershed is predominantly characterized by hydrological soil group A, B and D. The watershed land cover was dominated by the forest. The land use was classified as low, medium, and industrial urban (13%), deciduous, evergreen and mixed forest (47.4%), woody and herbaceous wetland (18.6%), range shrubland, grassland herbaceous and hay (19.4%), and the rest as water, south western

range and agricultural land (1.6 %). The information about the land use data, soil data and USGS gage is reported in Table 2.1.

2.4.7 Hydrologic Model Calibration and Validation

The streamflow calibrations were carried out using the observed streamflow recorded since 1/1/1990 to 12/31/1996 at USGS gage station (station ID02471001). Since 2001 land cover data sets were used, the calibrated model parameters were applied to an independent time period (1/1/1997 to 12/31/2002) for model validation. This period was selected as a validation period to evaluate the model performance for the latest land use condition. The simulation was started from 1/1/1985, permitting long spin up period for the model in order to minimize the effect of unknown initial moisture conditions, and to stabilize the hydrological conditions.

The watershed parameters, difficult to measure, were calibrated within a physically possible range, through the repeated trial and error procedure until the simulated flow closely approximated with the observed data. The model parameters adjusted during hydrologic model calibration are presented in Table 2.3. Once the model is calibrated and validated, the model was run for another 40 years to simulate the long term streamflow. LPSC was run in hourly time step, and simulated flows were aggregated into daily and monthly time scale for model comparisons and computing corresponding statistical error parameters (Tables 2.5, 2.6, and 2.7).

2.5 Result and Discussion

2.5.1 Wavelet Analysis

The monthly wavelet power spectra of SST and streamflow time series are shown in Figure 2.4. The wavelet analysis for streamflow and SST indicated that the common wavelet

power is observed for a common period but with distinct discontinuities. The distinct shared features in wavelet power between two time series (SST and streamflow series) were experienced in the 3 to 7 year band. These 3 to 7 years of frequency band also varied at different historic time period, and mostly dominated from period 1972 to 1992. The most prominent SST variation within the wavelet power spectrum is detected in 1998 during the strongest El Niño phenomenon, which is one of the strongest El Niño periods encountered since 1950. The major ENSO events since 1950 are reported in Table 2.4. Both series also had high power in the common longer period which was experienced from 1965 to 1982 within a period of 3 to 7 years. However, power for SST is not at 5% significance level. To ensure that the results are not merely coincidence, the cross wavelet transform (XWT) is used. The wavelet spectrum of SLP is not included in this manuscript to keep the manuscript concise.

2.5.2 Cross-wavelet Analysis

XWT, which exposes region with high common power, was constructed from the two CWTs. The cross-wavelet transform between SST and streamflow indicated that significant power was shared from 1970 to 1982 (Figure 2.5). These were phase locked positively. The cross-wavelet transform of streamflow and SST (Figure 2.5) indicated a significant area extending from 1970 to 1982 with a shared 2 to 5 years period of in-phase periodicity. The cross wavelet transform of SLP difference (Tahiti and Darwin) with streamflow indicated that SF and SLP demonstrate the common wavelet power although with a frequent cutoff. The two time series demonstrated the strong relationship (5% significance level) (Figure 2.5) corresponding to the La Niña encountered in 1975 and El Niño in 1998. Also, the significant in phase correlation was found for one year of frequency.

2.5.3 Wavelet Coherence Analysis

Generally, larger area in the wavelet coherence can be expected. Figure 2.6 shows the large area compared to XWT. However, the sharing power was not as high as in XWT. The wavelet-coherence transform between SST and streamflow show very similar significant areas for a period of around one year indicating high correlation between SST and streamflow. The large significant area was obtained with the SLP and streamflow.

2.5.4 Performance of the ANFIS Model

The overall performance of simulated streamflow model i.e. combined *SR* and *BF* model, with the actual observed data, is presented in Table 2.5. The model adequately simulated stream flows with satisfactory performance in each stage of training, validation and testing (Figure 2.8). From the visual inspection, it can be concluded that the model simulation matches well with the observed SF. Further, the ANFIS model was extended to a daily scale with some modification in its input data sets. For the *SR* model, while simulating at a daily scale, I used the same inputs that I used for the monthly *SR* model. However, base flows at the daily scale may depend on the precipitation characteristics of the preceding few weeks due to the time lag of the ground water contribution. Therefore, the combination of precipitation sets at different lead times was experimented with in ANFIS model until the model performance improved. It is interesting to mention that SLP, SST and SST (t-1) were the most sensitive inputs for the base flow model as well. The comparison of both model performances at daily scale is presented in the Table 2.6, which indicates that streamflow simulation by ANFIS model is as good as *LSPC* model.

2.5.5 Performance of LSPC Model

The performance of the LSPC model was promising for monthly simulation and satisfactory for daily simulation. The model performance was satisfactory demonstrating that hydrological parameters were able to capture the dynamics of the system. The following section describes the performance of LSPC model with ANFIS Model.

2.5.6 Model Comparison

Model results after calibration were assessed through the comparison of simulated and observed streamflow in terms of water budget, storm flow comparison with respect to volume and peak, low flow and high flow period (Table 2.7) and flow duration curve, etc. The statistical parameters, such as, NSE , R^2 and MBE are listed in the Table 2.8. The performance of ANFIS model was compared with the LSPC model for the same period of model simulation. Since LSPC model was calibrated for a period 1/1/1990 to 12/31/1996 (Figure 2.9), and validated for a period from 1/1/1997 to 12/31/2002 (Figure 2.10), two models were compared in those consecutive periods and found that the performance of the ANFIS model was comparable to the performance of the LSPC model (Table 2.8). This comparison was based on the long term data sets used for ANFIS model training starting from 1950.

In order to see the response of ANFIS model for shorter-term data sets and make realistic comparison with the LSPC model, additional ANFIS model was developed using 13 years data sets from 1990 to 2002 ensuring same period of datasets for both model development. ANFIS model performance in training and validation period was similar to corresponding LSPC model calibration ($NSE= 0.85$ for ANFIS and $NSE=0.83$ for LSPC) and validation ($NSE =0.73$ for

ANFIS and $NSE=0.71$). Now onwards, all other discussion and comparison with LSPC model in this manuscript are based on ANFIS model developed using 50 years datasets.

The LSPC model simulation (50 years) was stretched back to 1947 using 5 years of warm up period. The model performance corresponding to long term simulation against the ANFIS model for the same period was compared, and found that the ANFIS model is as good as the LSPC model (Table 2.9). In addition, I analyzed median observed and median simulated flow for each month and compared using the box plot diagram (Figure 2.11). It is noteworthy to mention that the response of the two models varied significantly from season to season. The median observed and simulated flows from ANFIS model closely matched from January to August, and demonstrated significant difference for other months (September to December). However, the response of LSPC model is almost opposite leading to the significant difference from February to June, and close resemblance with the observed streamflow for other months (July to December). This indicates to the fact that ANFIS model can simulate streamflow resembling the actual observed flows for most of the year and LSPC can simulate better for low flow period. Because SST and SLP were directly implemented into a mathematical model (ANFIS), it was interesting for me to explore how this model would simulate streamflow at corresponding major historic ENSO events since 1950. Moreover, I wanted to see how the conventional watershed model would simulate in a given condition, and also wanted to evaluate the best approach for streamflow simulation in the ENSO-affected region. For this, I classified the streamflow in different historic ENSO phases. The classification of ENSO phase was based on the Niño 3.4 index which is calculated using the 3 month consecutive average of ERSST.v3b SST anomalies in the Niño 3.4 region (5°N - 5°S , 120° - 170°W) (Trenberth and Stepaniak 2001).

The ANFIS model was compared with LSPC model separately for La Niña and El Niño events, and found that ANFIS performed better than the LSPC model at different ENSO events (Table 2.10) due to the training of ANFIS model with long term data sets.

Since SST, SLP and trade wind index are the parameters related with basic ENSO phenomenon in the equatorial Pacific, trade wind index was also a sensitive input for the model developed from 1979 to 2011 (not shown). This indicates to the fact that the hydrologic analysis in ENSO-affected coastal area can be better addressed using SST, SLP and trade wind index. Because I wanted to run the model for 50 years to evaluate the performance of the model in different ENSO events, I could not include trade wind index as model inputs as these data were available only after 1979.

ANFIS demonstrated its competent performance over the LSPC model in daily scale as well, which implies that the model can be extended to lower temporal scales (hourly). However, the influence of climate variability in streamflow was more distinct at monthly scales.

Having analyzed the two model performance in different scale for different events, it can be concluded that ANFIS can simulate watershed response as good as the LSPC model due to the direct incorporation of the dominant ENSO phenomenon. I also wanted to confirm that the actual improvement over the model simulation was due to the application of SST/SLP. To confirm this, I tested the winter season (JFM) using a simple MLR model as the ENSO signal was clear during this period. I found that model improved in a winter season due to the direct application of SST and SLP in the model. It is clearly inferred the fact that application of SST and SLP is beneficial only for a period when ENSO shows a better signature with streamflow. Readers can find the detailed analysis regarding the importance of SST in data-driven model in the ENSO-affected region in the fifth chapter of this dissertation.

The nearest precipitation gauge station to the Chickasaw Creek watershed is located at 20 miles away from the watershed boundary. In this context, I need to clearly infer the fact that both models (ANFIS and LSPC) suffered due to the lack of spatially distributed precipitation data. This was clear with the close inspection of precipitation and corresponding streamflow over a long period of record. Some of the streamflow events did not correspond to the observed precipitation. There could be many reasons for the unusual streamflow for the given precipitation record. Some of the major reasons are: (1) the precipitation recorded at the station was erroneous on that day, (2) the precipitation measured at the station did not match with the actual precipitation that occurred in the watershed, (3) there could be error while measuring the observed data (precipitation and streamflow). Despite of lack of large number of precipitation stations within the watershed boundary, the overall model performance were satisfactory.

2.6 Summary and Conclusion

In this paper, careful selection of inputs in a data-driven model (e.g., ANFIS) was highlighted, and in particular, wavelet analysis and cross-correlation techniques were used to identify and determine the teleconnection between SST and streamflow. The cross correlation analysis indicated that SST and precipitation has a lag correlation. The cross-wavelet transform and wavelet-coherence transform explored the physical teleconnection between SST and streamflow.

I implemented the SST and SLP difference between Tahiti and Darwin directly in the ANFIS model as model inputs and investigated its performance over 50 years of simulation. The comprehensive streamflow comparison was carried out with observed streamflow using different statistical criteria at different periods and seasons. Besides, observed and simulated streamflow

were compared in different La Niña and El Niño events throughout the 50 years simulation. The model simulated the monthly streamflow satisfactorily with reasonable accuracy. The ANFIS model was compared with the LSPC model at different temporal scale for the same input data sets. ANFIS model performance was comparable to the LSPC model at both monthly and daily scales. Moreover, the better performance of ANFIS model to capture the low and high flows during winter La Niña and El Niño events suggest the application of SST and SLP for long term continuous streamflow simulation. As a matter of fact, the research opens a new window for future studies to develop the neuro-fuzzy computational technique by using long term recorded climate data in equatorial Pacific for rainfall-runoff modeling to capture the dominant climate pattern due to ENSO phenomenon. As discussed in different literatures, the superiority of one modeling approach over the other cannot be generalized. However, this research point out the application of data-driven modeling approaches using SST and SLP data in the ENSO-affected coastal region. More importantly, it provides viable and a better alternative for streamflow simulation and justifies the relevancy of traditional data-driven modeling approach.

Since rainfall-runoff modeling that is applicable in much of the US, is not applicable for the coastal plain, this approach is still relevant for the coastal watershed affected by ENSO climate pattern.

Table 2.1. Data used for the study area with their sources and format

Data	Source	Additional information
Land use, NLCD 2001	www.AlabamaView.org	
Soil	Soildatamart.nrcs.usda.gov	SSURGO soil data base
DEM	www.seamless.usgs.gov , www.datagateway.nrcs.usda.gov	10-m resolution
Weather gage station	NOAA, National Climatic Data Center , http://ncdc.noaa.gov	Coop ID-015478
Streamflow	www.al.water.usgs.gov	(USGS gage 02471001)
Hydrologic (stream network) data	www.aces.edu/waterquality/gisdata	
GIS layers/tiger files	www.AlabamaView.org	

Table 2.2. Inputs for ANFIS model (monthly data).

Model	Model inputs					
SR Model	SST (t-1)	SLP _{diff} (t)	T(t-1)	T(t)	SR _{est} (t-1)	SR _{est} (t)
BF Model	SST(t)	SLP _{diff} (t)	T(t-1)	P(t-1)	P(t)	

Note: SR_{est}, T(t), P(t), and SST(t), SLP_{diff}(t) denote “estimated surface runoff”, temperature, precipitation, sea surface temperature, and sea level pressure difference between Tahiti and Darwin at the month “t”, respectively.

Table 2.3. The calibrated model parameters in LSPC model

Parameter	Description	Units	Typical	Calibrated Value	Remarks
LZSN	Lower zone nominal soil moisture storage	Inches	2.0-15.0	9	An increase will provide more chances for ET and decreases flow
INFILT	Index to infiltration capacity	Inches hr ⁻¹	0.001 to 0.50	0.01-0.08	An increase will cause shift of surface runoff to base flow
Kvary	Variable ground water recession	Inch ⁻¹	0.0-0.5	0.25	An increase will cause quick ground water depletion
AGWRC	Base ground water recession	none	0.850-0.999	0.992-0.998	An increase will cause the flattened base recession

Table 2.4 El Niño and La Niña years (Dec-April) from 1950 to 2003.

La Niña Year	El Niño Year
1950	1958
1951	1966
1955	1969
1956	1973
1968	1983
1971	1987
1974	1988
1975	1992
1976	1995
1985	1998
1989	2003
1996	
1999	
2000	

Table 2.5. Performance of the ANFIS model in different modeling stages (monthly simulation from 1952-2002)

Statistics	Streamflow Simulation		
	Training (1952-1989)	Validation (1990-1996)	Testing (1997-2002)
RMSE (m ³ /s)	2.63	2.37	3.7
NSE	0.8	0.85	0.74
MBE	0%	3.8%	8.8%

Table 2.6. Model simulation in daily scale and performance comparison (1990-1996)

Statistics	LSPC	ANFIS
RMSE(m ³ /s)	9.9	7.0
NSE	0.486	0.68
MBE	2%	0.6%

Table 2.7. Performance comparison of two models at monthly time scale (period 1990-1996)

Errors (Simulated-Observed)	ANFIS Error Statistics (%)	LSPC Error Statistics (%)	Recommended Criteria (adopted by Tetra tech)
Error in total volume	-0.18	1.82	10
Error in 50% lowest flows	19.79	39.56	10
Error in 10% highest flows	-18.64	-26.65	15
Seasonal volume error – Summer	20.63	12.95	30
Seasonal volume error – Fall	28.79	2.35	30
Seasonal volume error - Winter	-9.99	-5.58	30
Seasonal volume error - Spring	-13.84	6.19	30
Error in storm volumes:	0.65	-54.82	20
Error in summer storm volumes	80.93	-72.54	50

Table 2.8. Performance comparison of ANFIS and LSPC model during a period of LSPC model calibration and validation (monthly simulation)

Statistics	LSPC Model Calibration (1990 to 1996)	
	LSPC	ANFIS
NSE	0.83	0.85
MBE	1.8%	3.8%
R ²	0.83	0.86
Statistics	LSPC Model Validation (1997 to 2002)	
	LSPC	ANFIS
NSE	0.71	0.74
MBE	-4.0%	8.8%
R ²	0.71	0.76

Table 2.9. Performance of LSPC and ANFIS model during 50 years simulation at monthly time step

Statistics	LSPC	ANFIS
NSE	0.67	0.79
MBE	1%	0.6%
R ²	0.67	0.79

Table 2.10. Model comparison for 25 ENSO events (monthly simulation)

For La Niña and El Niño events (since 1952)		
Statistics	LPSC-simulated	ANFIS-Simulated
NSE	0.67	0.83
MBE	-4.02%	3.10
R ²	0.68	0.83
14 La Niña events (Table 2.4)		
NSE	0.67	0.86
MBE	5%	6%
R ²	0.68	0.87
11 El Niño events (Table 2.4)		
NSE	0.61	0.76
MBE	4.90%	-3.12%
R ²	0.66	0.77

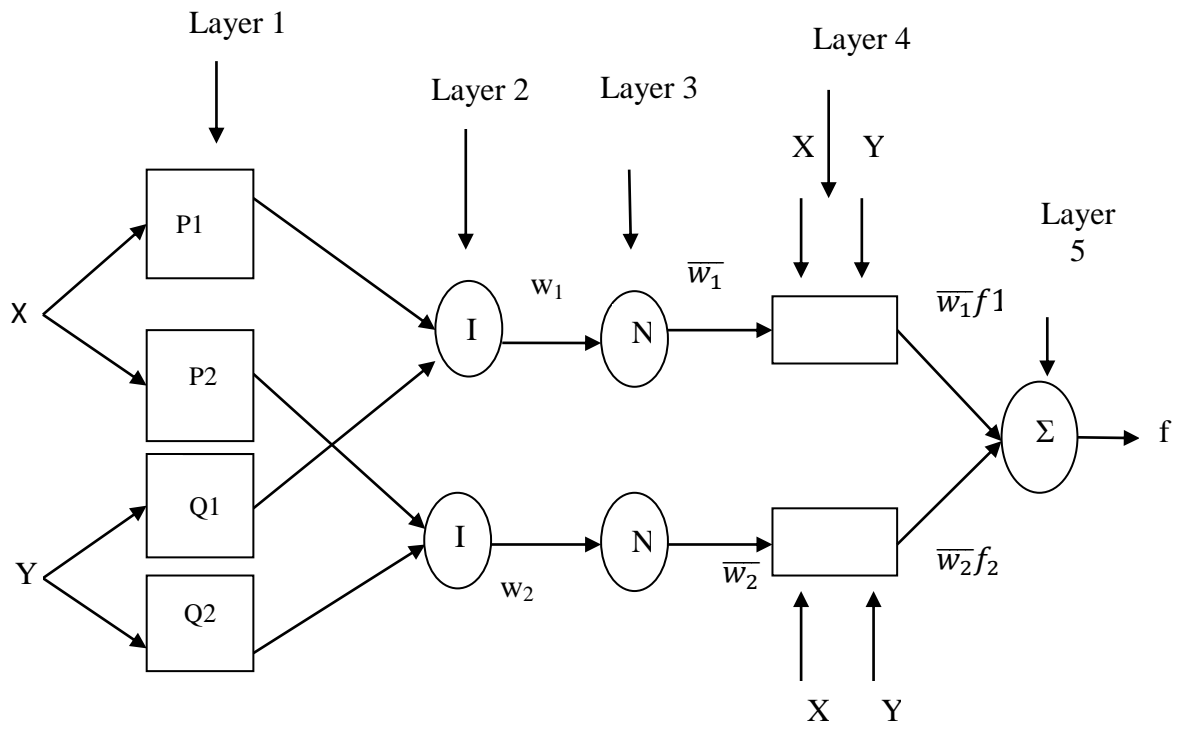


Figure 2.1. ANFIS architecture for two inputs (X and Y) with two rules and two membership functions (P1, P2 and Q1, Q2) for each rule.

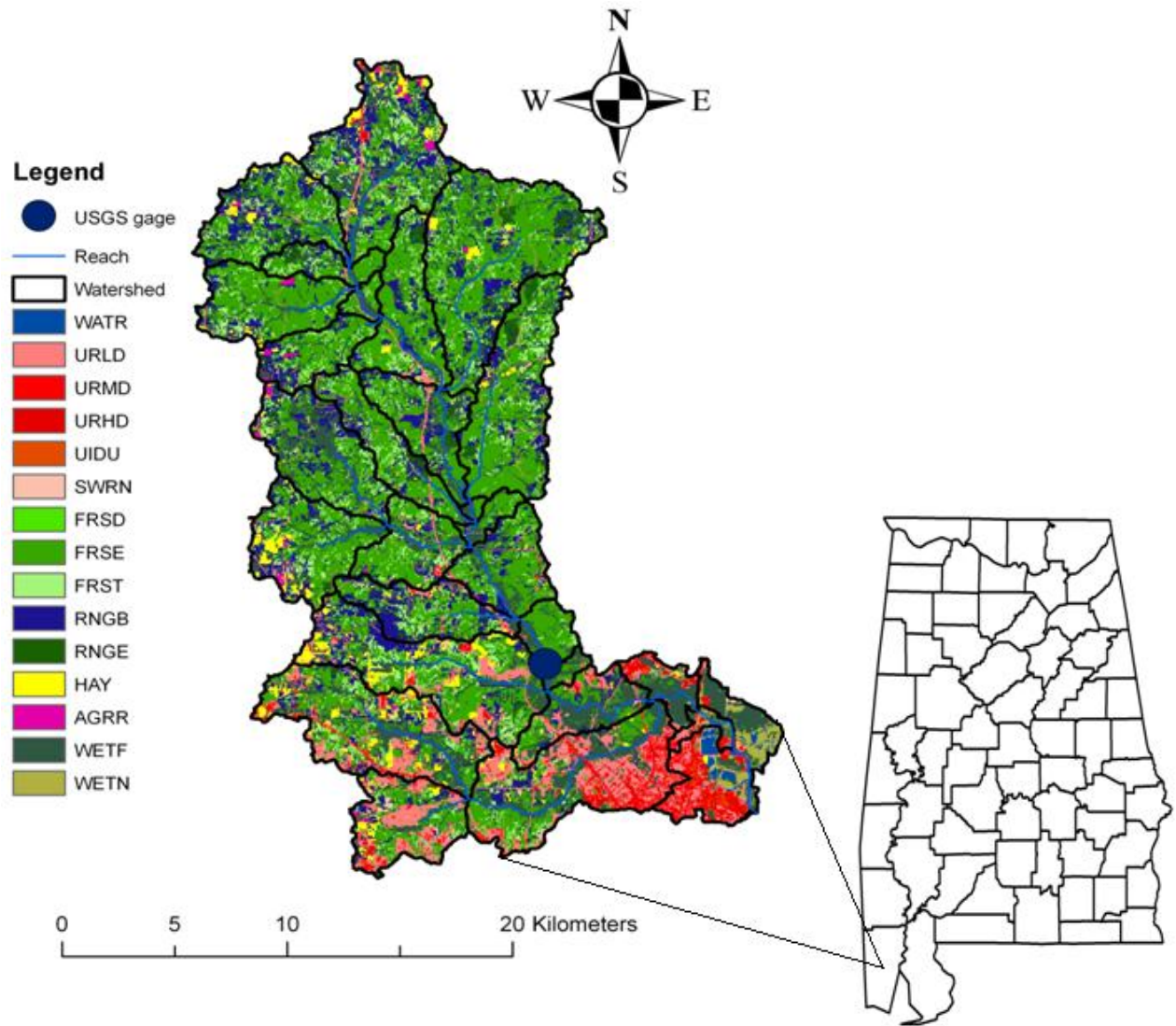


Figure 2.2. Chickasaw Creek watershed in Mobile County of South Alabama, USA, showing the USGS gaging station and land use distribution within the watershed. The land use classification is as per NLCD land use categories.

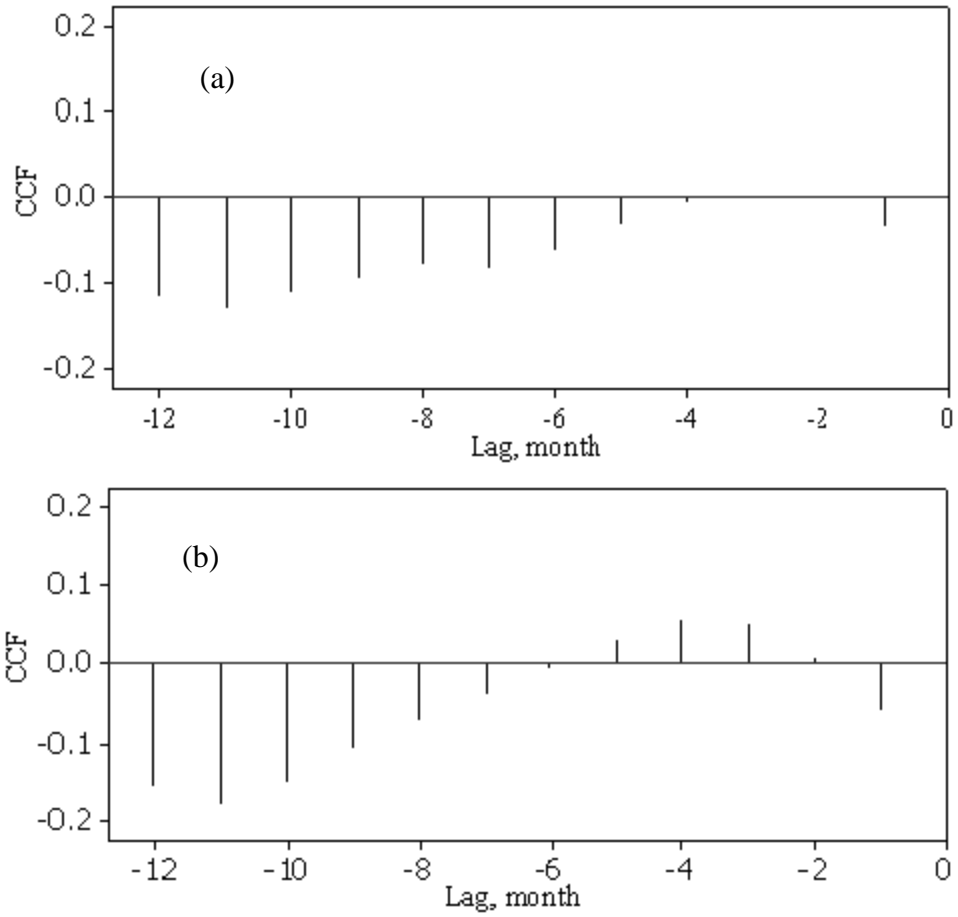


Figure 2.3. Cross correlation function (CCF) of the sea surface temperature with: a) surface runoff, and (b) base flow at different lagging months. Negative lag indicates the second series is lagged to first series. Here first series is sea surface temperature and second series is either surface runoff or base flow.

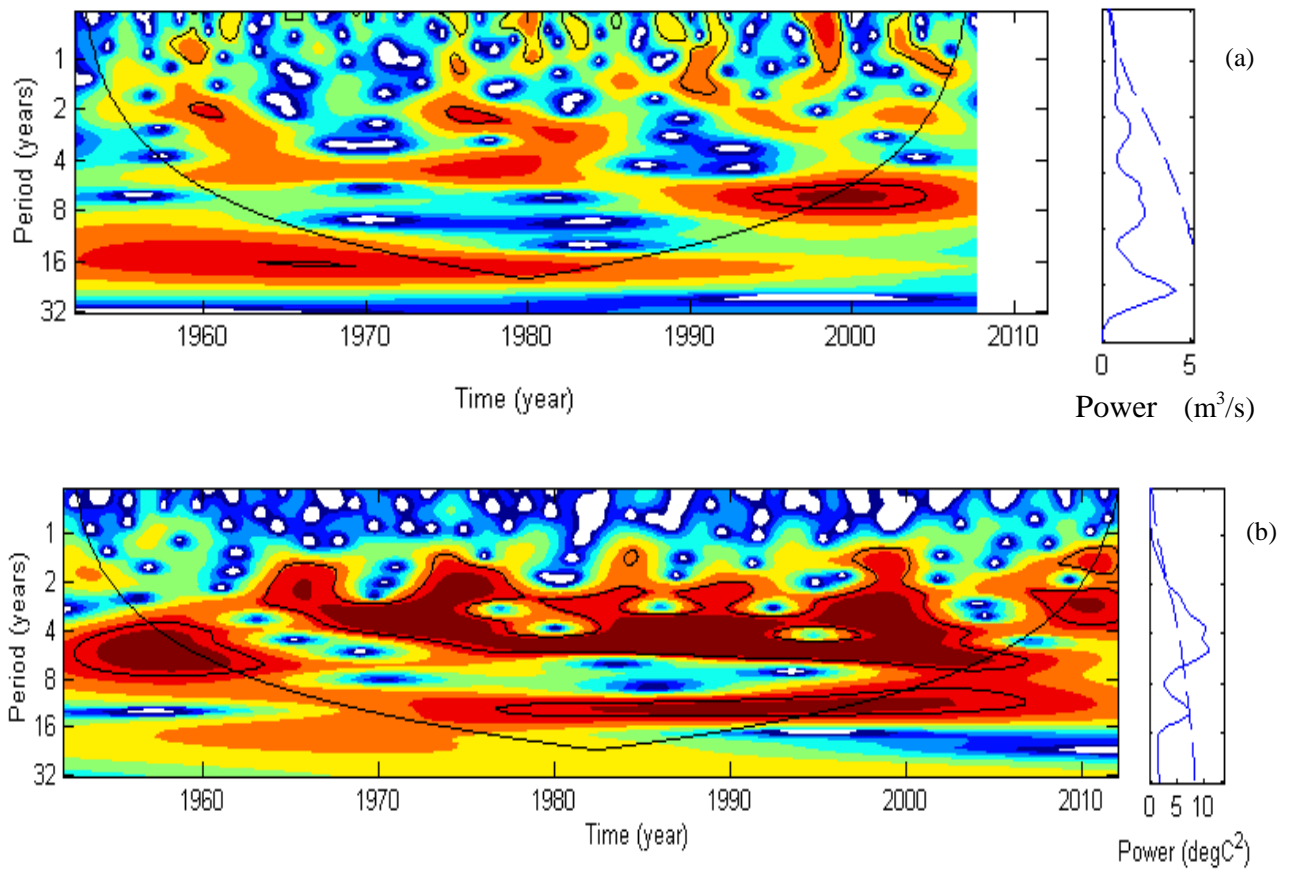


Figure 2.4. (a) The wavelet power spectrum with in the cone of influence for monthly streamflow, (b) monthly SST anomalies in Niño 3.4 region. The red contour line indicates high wavelet power. The right figure indicates the Global Wavelet Spectrum (GWS). The dashed blue line indicates the 95% confidence limit. X-axis indicates the time period, and Y-axis indicates the period of occurrence. Red eye land corresponds to higher power (in the right side); for which, the period of occurrence is 2 to 7 years. It suggests that SST varies significantly with in a frequency of 2 to 7 years from a period of 1968 to 2005.

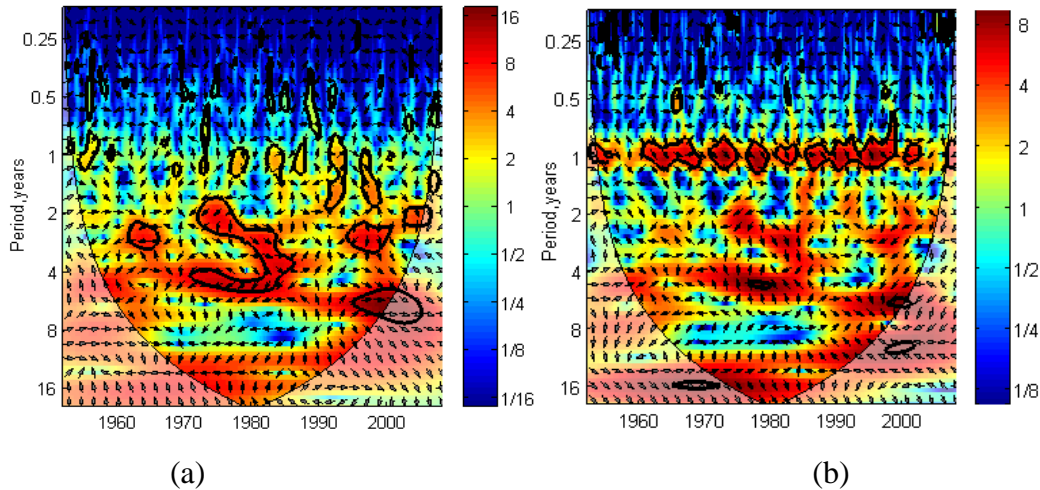


Figure 2.5. (a) Cross wavelet spectrum between monthly streamflow and monthly SST, and (b) Cross wavelet spectrum between monthly streamflow and monthly SLP difference at Tahiti and Darwin. Black outline indicates that the relationship is significant to 95% within the region. The right vertical band associated with each figure indicates the global wavelet power. The clockwise “arrow pointing” indicates the same phase (positive) relationship, and anti-clockwise arrow indicates the opposite relationship (negative). The left panel suggests that the strong relationship with high power is observed during 1970 to 1982 with in the frequency of 2 to 5 years; however, this frequency is approximately 3 to 8 years from 1995 to 2000.

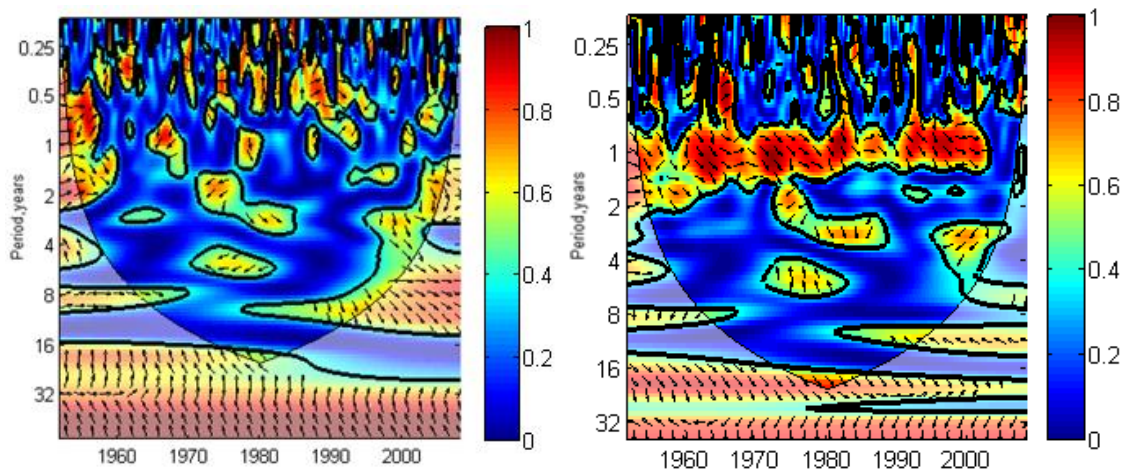


Figure 2.6. Left figure indicates the wavelet coherence analysis (WTC) between streamflow and SST and the right figure indicates the WTC between streamflow and SLP difference (Tahiti and Darwin). Clockwise arrow indicates the in-phase relationship and anti-clock wise arrow indicates the out of phase relationship. The black outline inside the cone of influence represents the areas significant to 90% confidence interval. Right vertical panel associated with each figure indicates the power. The left figure indicates that streamflow and SST are correlated with power (~0.6) with different frequencies of occurrence at different times.

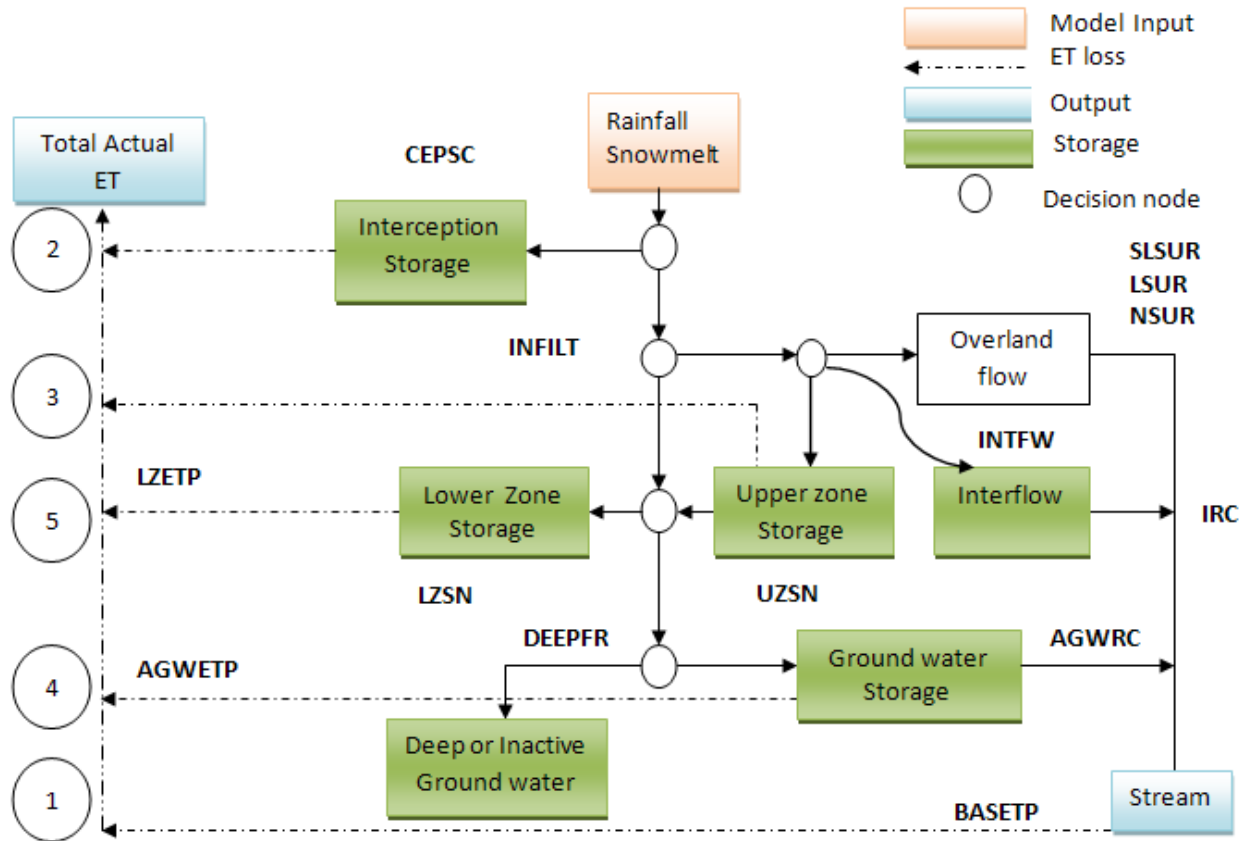


Figure 2.7. Schematic diagram of the Stanford Watershed Model adapted for LSPC model. Number corresponding to each circle denotes the order of removing water to satisfy ET (**Source: LSPC User's Manual, Tetra tech, 2009**).

INFILT, infiltration parameter; INTFW, interflow parameter; CEPSC, interception storage capacity; LZSN, lower zone storage capacity; UZSN, upper zone storage capacity; NSUR, Manning's surface roughness; LSUR, surface runoff length; SLSUR, surface slope; IRC, interflow recession constant; AGWETP; active groundwater ET parameter; AGWRC; active groundwater recession; BASETP; base flow ET parameter; DEEPFR, fraction to inactive groundwater; LZETP, lower zone ET parameter;

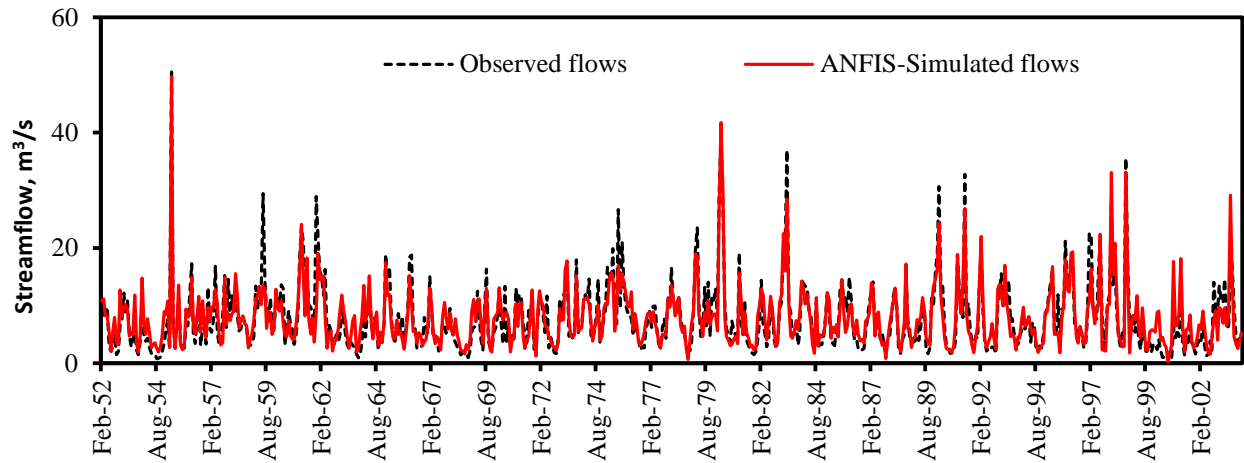


Figure 2.8. Monthly streamflow simulated for 52 years using ANFIS model compared with observed monthly streamflow.

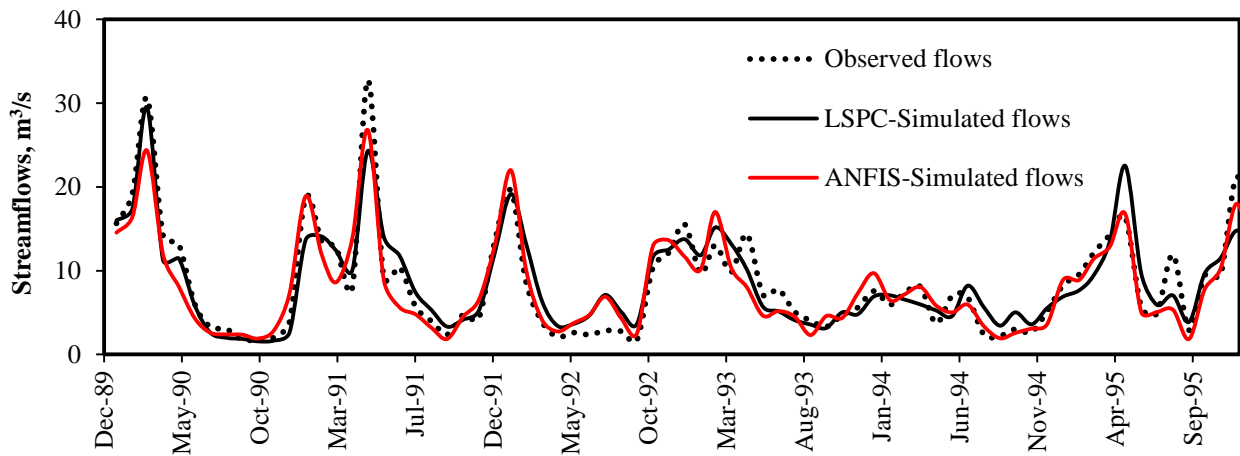


Figure 2.9. LSPC model calibration and its performance comparison with ANFIS model from 1990 to 1996 (monthly streamflow).

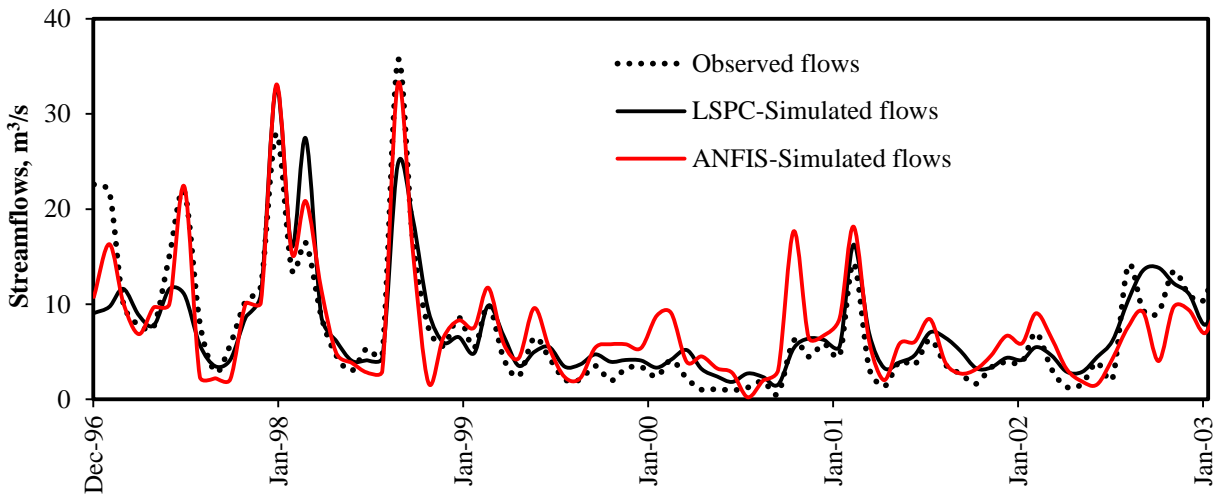


Figure 2.10. LSPC model validation and comparison with ANFIS model during LSPC model validation phase.

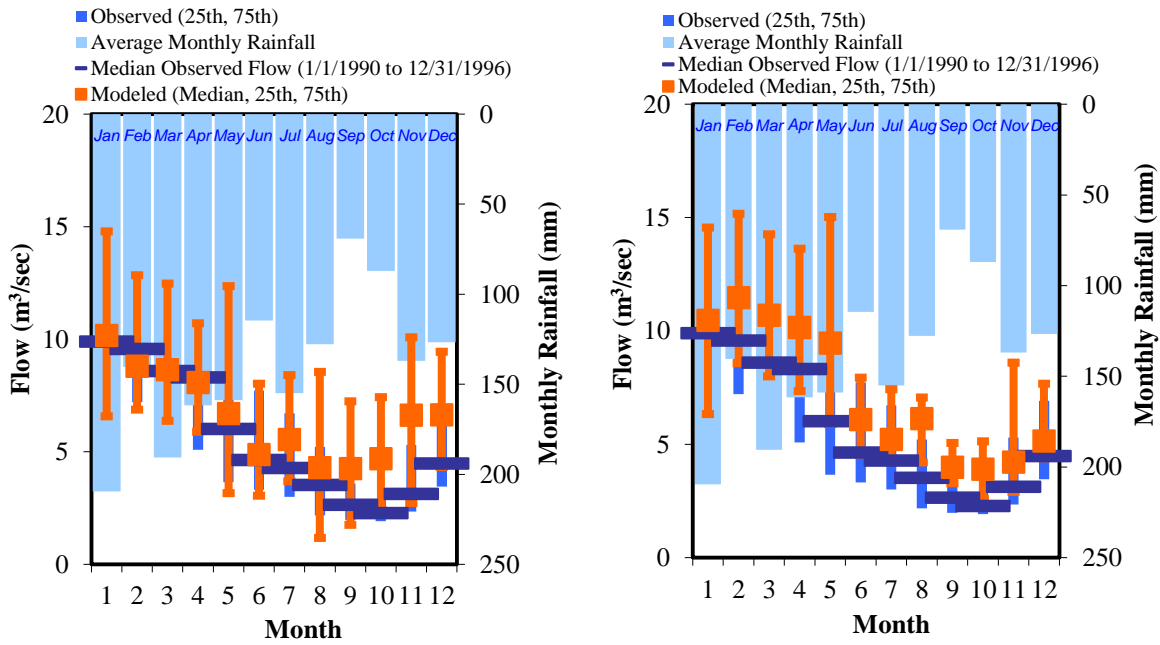


Figure 2.11. Statistical analysis of monthly streamflow simulated using: (a) ANFIS model and (b) LSPC model.

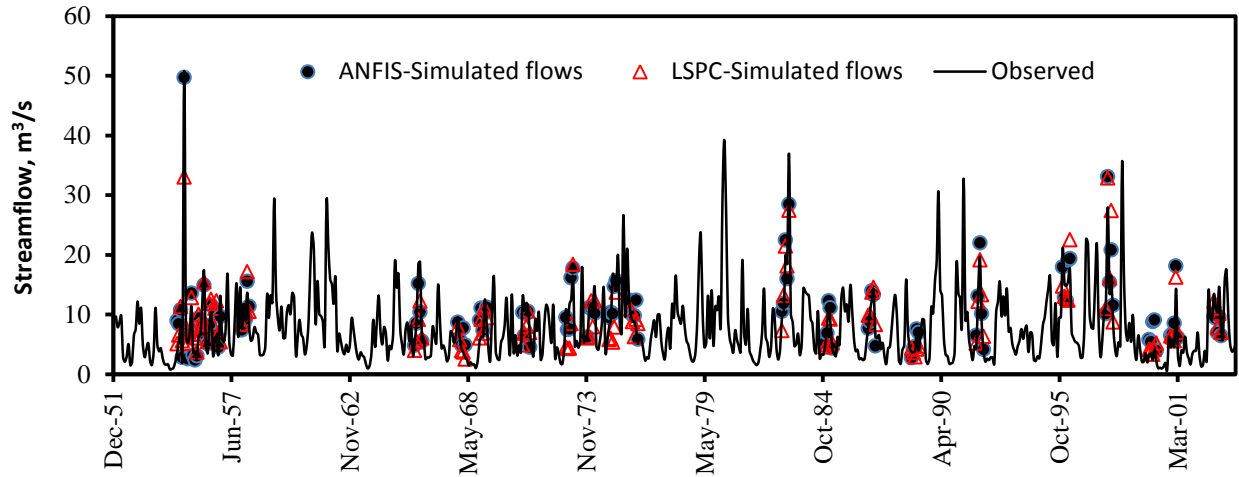


Figure 2.12. Monthly streamflow simulation using LSPC and ANFIS models corresponding to historic La Niña and El Niño events (Table 3) since 1950. Dots and triangles represent ANFIS and LSPC simulation during La Niña and El Niño events.

2.7 References

- Allen, M. R., and Smith, L. A. (1996). "Monte Carlo SSA: Detecting irregular oscillations in the presence of colored noise." *Journal of Climate*, 9(12), 3373-3404.
- ASCE (2000a). "Artificial neural networks in hydrology. I: Preliminary concepts." *Journal of Hydrologic Engineering*, 5(2), 115-123.
- ASCE (2000b). "Artificial neural networks in hydrology. II: Hydrologic Applications." *Journal of Hydrologic Engineering*, 5(2), 124-137.
- Bezdek J. C. (1981): "Pattern recognition with fuzzy objective function algorithms", Plenum Press, New York
- Barsugli, J. J., Whitaker, J. S., Loughe, A. F., Sardeshmukh, P. D., and Toth, Z. (1999). "The effect of the 1997/98 El Niño on individual large-scale weather events." *Bull. Amer. Meteor. Soc.*, 80, 1399–1411.
- Bicknell, B., Imhoff, J., Kittle Jr, J., Jobs, T., Donigian Jr, A., and Johanson, R. (2001). "Hydrological Simulation Program–Fortran: HSPF, Version 12 User’s Manual." *AQUA TERRA Consultants, Mountain View, California.*
- Bishop, C. (1995). "Neural networks for pattern recognition, 18 pages." Clarendon Press.
- Bosznay, M. (1989). "Generalization of SCS curve number method." *Journal of Irrigation and Drainage Engineering*, 115, 139.
- Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1970). *Time series analysis*, Holden-day.
- Chiew, F., Piechota, T., Dracup, J., and McMahon, T. (1998). "El Niño /Southern Oscillation and Australian rainfall, streamflow and drought: Links and potential for forecasting." *Journal of Hydrology*, 204(1-4), 138-149.
- Chiu, S. (1996). "Method and software for extracting fuzzy classification rules by subtractive clustering." *IEEE*, 461-465.
- Coulibaly, P., and Baldwin, C. K. (2005). "Nonstationary hydrological time series forecasting using nonlinear dynamic methods." *Journal of Hydrology*, 307(1-4), 164-174.
- Crawford, N. H., and Linsley, R. K. (1966). "Digital simulation in hydrology' stanford watershed model." Technical Report 39, Dept. of Civil Engineering, Stanford university, California
- DEM. (2010). "Digital Elevation Model." WWW data downloaded from www.seamless.usgs.gov.

- Dunn J. C. (1973). "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters", *Journal of Cybernetics* 3: 32-57
- Farokhnia, A., Morid, S., and Byun, H. R. (2011). "Application of global SST and SLP data for drought forecasting on Tehran plain using data mining and ANFIS techniques." *Theoretical and Applied Climatology*, 1-11.
- Grinsted, A., Moore, J., and Jevrejeva, S. (2004). "Nonlinear processes in geophysics application of the cross wavelet transform and wavelet coherence to geophysical time series." *Nonlinear processes in geophysics*, 11(5/6), 561-566.
- Hansen, J., Jones, J., Irmak, A., and Royce, F. (2001). "El Niño-southern oscillation impacts on crop production in the southeast United States." *ASA Special Publication*, 63, 55-76.
- Hearty, Á. P., and Gibney, M. J. (2008). "Analysis of meal patterns with the use of supervised data mining techniques—artificial neural networks and decision trees." *The American journal of clinical nutrition*, 88(6), 1632.
- Hendon, H. H., Thompson, D. W. J., and Wheeler, M. C. (2007). "Australian rainfall and surface temperature variations associated with the Southern Hemisphere annular mode." *Journal of Climate*, 20(11), 2452-2467.
- Jang, J. S. R. (1993). "ANFIS: Adaptive-network-based fuzzy inference system." *Systems, Man and Cybernetics, IEEE Transactions on*, 23(3), 665-685.
- Jang, J. S. R., and Sun, C. T. (1995). "Neuro-fuzzy modeling and control." *Proceedings of the IEEE*, 83(3), 378-406.
- Jang, J. S. R., Sun, C. T., & Mizutani, E. (1997). *Neuro-fuzzy and soft computing: "A computational approach to learning and machine intelligence."* Eaglewood cliffs,NJ: Prentice-Hall.
- Jayawardena, A., Muttill, N., and Lee, J. (2006). "Comparative analysis of data-driven and GIS-based conceptual rainfall-runoff model." *Journal of Hydrologic Engineering*, 11, 1.
- Kalin, L., and Hantush, M. M. (2006). "Hydrologic modeling of an eastern Pennsylvania watershed with NEXRAD and rain gauge data." *Journal of Hydrologic Engineering*, 11, 555.
- Keener, V., Ingram, K., Jacobson, B., and Jones, J. (2007). "Effects of El Niño/Southern Oscillation on simulated phosphorus loading in South Florida." *Transactions of the ASAE*, 50(6), 2081-2089.

- Khalil, A. F., McKee, M., Kemblowski, M., and Asefa, T. (2005). "Basin scale water management and forecasting using artificial neural networks." *JAWRA Journal of the American Water Resources Association*, 41(1), 195-208.
- Kulkarni, J. (2000). "Wavelet analysis of the association between the southern oscillation and the Indian summer monsoon." *International Journal of Climatology*, 20(1), 89-104.
- Mamdani, E. H., and Assilian, S. (1975). "An experiment in linguistic synthesis with a fuzzy logic controller." *International Journal of Man-Machine Studies*, 7(1), 1-13.
- McCabe, G. J., and Dettinger, M. D. (1999). "Decadal variations in the strength of ENSO teleconnections with precipitation in the western United States." *International Journal of Climatology*, 19(13), 1399-1410.
- Moriasi, D., Arnold, J., Van Liew, M., Bingner, R., Harmel, R., and Veith, T. (2007). "Model evaluation guidelines for systematic quantification of accuracy in watershed simulations."
- Mosley, M. P. (2000). "Regional differences in the effects of El Niño and La Niña on low flows and floods." *Hydrological Sciences Journal*, 45(2), 249-267.
- Mukerji, A., Chatterjee, C., and Raghuwanshi, N. S. (2009). "Flood forecasting using ANN, neuro-fuzzy, and neuro-GA models." *Journal of Hydrologic Engineering*, 14, 647.
- Nash, J., and Sutcliffe, J. (1970). "River flow forecasting through conceptual models part I--A discussion of principles." *Journal of Hydrology*, 10(3), 282-290.
- Nayak, P., Sudheer, K., Rangan, D., and Ramasastri, K. (2004). "A neuro-fuzzy computing technique for modeling hydrological time series." *Journal of Hydrology*, 291(1-2), 52-66.
- NLCD. (2010). "National Land Cover Dataset." WWW data downloaded from www.epa.gov/mrlc/nlcd-2001.html.
- Pascual, M., Rodó, X., Ellner, S. P., Colwell, R., and Bouma, M. J. (2000). "Cholera dynamics and El Niño -southern oscillation." *Science*, 289(5485), 1766.
- Piechota, T. C., and Dracup, J. A. (1996). "Drought and regional hydrologic variation in the United States: Associations with the El Niño-Southern Oscillation." *Water Resources Research*, 32(5), 1359-1373.
- Pramanik, N., and Panda, R. K. (2009). "Application of neural network and adaptive neuro-fuzzy inference systems for river flow prediction." *Hydrological Sciences Journal*, 54(2), 247-260.
- Rajagopalan, B., and Lall, U. (1998). "Interannual variability in western US precipitation." *Journal of Hydrology*, 210(1-4), 51-67.

- Ropelewski, C., and Halpert, M. (1986). "North American precipitation and temperature patterns associated with the El Niño/Southern Oscillation (ENSO)." *Mon. Weather Rev.:(United States)*, 114(12).
- Roy, S. S. (2006). "The impacts of ENSO, PDO, and local SSTs on winter precipitation in India." *Physical Geography*, 27(5), 464-474.
- Shen, J., Parker, A., and Riverson, J. (2005). "A new approach for a Windows-based watershed modeling system based on a database-supporting architecture." *Environmental Modelling & Software*, 20(9), 1127-1138.
- Singh, V. P. (1988). "Hydrologic Systems. Volume I: Rainfall-Runoff Modeling." *Prentice Hall, Englewood Cliffs New Jersey*. 1988. 480.
- Srivastava, P., McNair, J. N., and Johnson, T. E. (2006). "Comparison of process based and artificial neural network approaches for streamflow modeling in an agricultural watershed." *JAWRA Journal of the American Water Resources Association*, 42(3), 545-563.
- SSURGO. (2010). "Soil Survey Geographic Database." WWW data downloaded from soildatamart.nrcs.usda.gov
- Stahl, K., and Demuth, S. (1999). "Linking streamflow drought to the occurrence of atmospheric circulation patterns." *Hydrological Sciences Journal*, 44(3), 467-482.
- Sugeno, M., and Takagi, T. (1983). "Multi-dimensional fuzzy reasoning." *Fuzzy Sets and Systems*, 9(1-3), 313-325.
- Thomson, A. M., Brown, R. A., Rosenberg, N. J., Izaurralde, R. C., Legler, D. M., and Srinivasan, R. (2003). "Simulated impacts of El Niño /Southern Oscillation on United States water resources." *JAWRA Journal of the American Water Resources Association*, 39(1), 137-148.
- Torrence, C., and Compo, G. P. (1998). "A practical guide to wavelet analysis." *Bulletin of the American Meteorological Society*, 79(1), 61-78.
- Trenberth, K. E., and Stepaniak, D. P. (2001). "Indices of El Niño evolution." *Journal of Climate*, 14(8), 1697-1701.
- USEPA (2009). "Loading Simulation Program in C++ (LSPC) Version 3.1 User's Manual." *United States Environmental Protection Agency, Region 4, Sam Nunn Atlanta Federal Center, 61 Forsyth Street, SW, Atlanta, GA 30303-8960.*
- Xu, Z., Takeuchi, K., Ishidaira, H., and Li, J. (2005). "Long-term trend analysis for precipitation in Asian Pacific river basins." *Hydrological processes*, 19(18), 3517-3532.

Yan, H., Zou, Z., and Wang, H. (2010). "Adaptive neuro fuzzy inference system for classification of water quality status." *Journal of Environmental Sciences*, 22(12), 1891-1896.

Zadeh, L. A. (1965). "Fuzzy sets*." *Information and control*, 8(3), 338-353.

Chapter 3. Long-Range Hydrologic Forecasting in El Niño Southern Oscillation-Affected Coastal Watersheds

3.1 Abstract

Streamflow forecasting is essential for the proper management of water resources, especially when severe droughts cause water resources scarcity. Streamflow forecasting using physically based or conceptual hydrologic models is a commonly used approach. However, conventional hydrologic models do not perform well in flat terrain coastal watersheds, and hence forecasting streamflow using such models may not be an appropriate choice. Also, conceptual models rely on the predicted climate data which are at times unrealistic and depart significantly from actual observed data resulting in an unreliable forecast. Since Sea surface temperature (SST) at Niño 3.4 region has a potential teleconnection with streamflow in the El Niño Southern Oscillation (ENSO)-affected regions, streamflow forecasting skill of a model can be enhanced using SST in data-driven models. In fact, conceptual models cannot incorporate SST data as an input. Therefore, in this study, an Adaptive Neuro Fuzzy Inference System (ANFIS) was used to infuse SST data (from equatorial Pacific) with predicted precipitation and temperature for streamflow forecasting with one to three months lead time. For forecasted climate data, I utilized two different methods: (i) conditioned weather sequences and (ii) climate data from Climate Forecast System Version 2 (CFSv2) model. The ANFIS model was developed using the long term climate data (since 1952), and the streamflow forecasting was initiated in 1982. The forecasted streamflow, after systematic error correction, were post-validated with observed

streamflow from 1982 to 1988. The streamflow forecasting at one month lead time was found to be better than that of three months lead time. This research concludes that the climate model approach will be a better choice for moderate size watershed for streamflow forecast at one month lead time. Conversely, the weather generator approach will be more suitable for streamflow forecasting at three months lead time. This is especially true for low flow conditions. Also, streamflow forecasting in the ENSO-affected coastal region can be enhanced using SST data predicted at equatorial Pacific.

3.2 Introduction

Uncertainty in water availability caused by inter-annual climate variability has led water managers to look for advanced techniques that utilize climate forecast for proper management of water resources. Streamflow forecasts are crucial for water resource managers for optimal allocation of water resources for multiple purposes (e.g. irrigation, water supply, hydropower generation and downstream requirements), especially when severe droughts intensify water resource scarcity. Suitable methods of forecasting streamflow, particularly low flows at appropriate temporal scale are needed in order to properly harness the limited water resources at watershed scale. For example, for long-term planning of water resources, longer lead time forecasts, preferably at monthly and seasonal scales are needed. Several approaches have been introduced for long-range hydrologic forecasting.

Traditionally, regression methods (Hsieh et al. 2003) or hydrologic time series (Krstanovic and Singh 1991) models have been used for streamflow forecasting. For the application of time series models, continuous data without any missing record is required. The time series approach is not plausible in case the few days' data are missing. Recently, application

of data-driven approaches for streamflow forecast, considering inter-annual climate variability resulting from a number of ocean-atmosphere phenomena that operate at seasonal to decadal time scales, have been introduced. For example, streamflow was forecasted in Iberian River, located in Spain, using SST anomalies at the seasonal scale (Gimiz-Fortis et al. 2010), and Columbia River discharge was forecasted at annual scales using various oceanic-atmospheric indices such as Pacific Decadal Oscillation (PDO) (Hamlet and Lettenmaier 1999), North Atlantic Oscillation (NAO) (Araghinejad et al. 2006), and Atlantic Multi-decadal Oscillation (AMO) (Rogers and Coleman 2003; Kalra and Ahmad 2009). Gutierrez and Dracup (2001) studied the relationship between ENSO and Columbia River discharge using systematic cross-correlations and concluded that several climate indices can be fused for predicting streamflow in the Columbia River. Similarly, Wang and Eltahir (1999) studied ENSO effect on river flows in the Nile and found that ENSO information is the most prominent predictor for long-range forecasting. Majority of these streamflow forecasting studies were carried out at seasonal and annual scales and ruled out the possibility of forecasting streamflow at monthly scale using climate indices. In this regard, fusion of SST with the precipitation and temperature seems to be beneficial for monthly streamflow forecasting. This hypothesis is justified by an earlier experiment of Eldaw et al. (2003) using SST of the current year with the precipitation of the previous year for quantitative long-range forecasting in the Blue Nile River. Using previous year precipitation is not a plausible approach, especially when climate variability has a potential to cause significant variations in precipitation from year to year. One of the approaches might be a climate model-based approach, which is operational at the National Weather Service (NWS) (Wood and Lettenmaier 2006). For example, Wood et al. (2002) utilized the predicted climate data from the Global Spectral Model (GSM) developed at the National Climatic and

Environmental Prediction (NCEP) for long-range hydrologic forecasting. Yet, predicted climate data from climate model are at times very unrealistic and significantly departs from the actual observed data (Wang et al. 2010; Yuan et al. 2011a) degrading the overall quality of the resulting forecast. Therefore, fusion of SST with predicted climate data could be beneficial for improved streamflow forecasting. However, the conceptual model that has been conventionally applied for streamflow forecasting; for example, widely used Variable Infiltration Capacity (VIC) model (Liang et al. 1996) cannot incorporate SST as an input. More importantly, the conventional hydrologic models have met several challenges, and are not fully capable to address flat-terrain coastal hydrology (Sheridan 2002). In fact, SST is a dominant component for driving coastal watershed hydrology (Keener et al. 2010). Therefore, this research explores the Adaptive Neuro-Fuzzy Inference System (ANFIS) model (Jang 1993) for the fusion of SST and predicted climate data derived from two different approaches: i) the Climate Forecast System Version 2 (CFSv2) (Saha et al. 2006) model, which is the latest version of climate model operational at NCEP; and (ii) the conditioned weather generated data (Clark et al. 2004; Baigorria and Jones 2010). The specific research objective of this study is to evaluate the climate model and the ENSO-conditioned weather generator approach for average monthly streamflow forecasting at one to three months lead time using the ANFIS model. More specifically, this research evaluates the low flow forecasting skill of the ANFIS model using these two approaches.

ANFIS is a combination of ANN and fuzzy logic and is better at capturing inherent non-linear processes than the two independent models (i.e., ANN and fuzzy logic). The theory behind the ANFIS model structure is briefly discussed in the following section.

3.3 Theoretical Consideration

3.3.1 ANFIS Model

Many studies in the past have demonstrated that intelligent computational techniques such as ANN and Fuzzy Logic approaches are efficient for hydrologic analysis (Mukerji et al. 2009; Pramanik and Panda 2009) and water quality modeling (Yan et al. 2010). In order to eliminate the shortcomings of each models, the concept of combining two models evolved. As a matter of fact, a network-based fuzzy logic approach (ANFIS model) which incorporates ANN and fuzzy logic approach has become common in recent years.

ANFIS is a feed-forward network consisting of multiple layers that utilizes neural network technique and fuzzy logic to map an input vector to an output vector by using back propagation or combined algorithm. Input–output vector is expressed in a fuzzy inference system (FIS) by using several fuzzy IF–THEN rules that represent the local behavior of the input and output mapping. ANFIS model has been extensively applied in water resources and water quality modeling (Mukerji et al. 2009; Pramanik and Panda 2009; Yan et al. 2010). The details of the model structure have been given in Jang (1993).

3.3.2 ANFIS Model Development

Figure 3.1 shows an example of first order Sugeno FIS for two inputs and one output. ANFIS architecture consists of five layers with two rules and two membership functions (MFs) for each input (Figure 3.1).

For two fuzzy if-then rules, the Sugeno-fuzzy models can be expressed as follows:

Rule 1: If X is P1 and Y is Q1, then $f1 = a_1 x + b_1 y + r1$

Rule 2: If x is P2 and Y is Q2, then $f2 = a_2 x + b_2 y + r2$

Where, P1:P2 and Q1:Q2 are the MFs for inputs X and Y, respectively, and a_1, b_1, r_1 and a_2, b_2, r_2 are the parameters for output functions.

After a series of operation in each layer, the final output is computed as the synopsis of all received signals, which can be represented as:

$$O_i^5 = \sum_i^2 \bar{w}_i f_i = \frac{\sum_i^2 w_i f_i}{\sum_1^2 w_i} \quad (3.1)$$

The brief operation of the ANFIS model is explained in Chapter 2. Hybrid approach was applied to derive unknown parameters by using training datasets to train adaptive networks.

3.3.3 Estimation of Parameters

Based on a simple architecture of ANFIS (Figure 3.1), the ANFIS model can be further extended to several other inputs. The relationship between streamflow and input variables can be expressed as:

$$SF = f(SST, T(t), PCP(t), PCP(t-1), PCP(t-2), PCP(t-3), WS(t-1), ET(t-1)) \quad (3.2)$$

Where SF, T, PCP, WS, ET and t denote streamflow, temperature, precipitation, wind speed, Evapotranspiration and time step (monthly), respectively.

For simplicity, ANFIS structure using two input datasets are presented in this manuscript. Assuming P1:P2, Q1:Q2 are the respective membership functions for inputs (SST, PCP) and $a_1:b_1, a_2:b_2$ are the parameters of the output function, resultant output can be expressed as a linear combination of subsequent parameters as follows.

$$SF = \sum_i^8 \bar{w}_i f_i = \bar{w}_1 f_1 + \bar{w}_2 f_2 \quad (3.3)$$

For two rules, the equation is simplified as follows.

$$SF = \bar{w}_1 (SST)a1 + (\bar{w}_1 PCP)b1 + +\bar{w}_1 r1 + (\bar{w}_2 SST)a2 + (\bar{w}_2 PCP)b2 + \bar{w}_2 r2 \quad (3.4)$$

The equation can be written in the matrix form as:

$$AZ = B \quad (3.5)$$

where matrix “A” is given as follows:

$$A = \begin{bmatrix} \bar{w}_1 SST_1 & \cdot & \cdot & \bar{w}_1 PCP(t)_1 & \bar{w}_1 & \bar{w}_2 SST_1 & \cdot & \cdot & \bar{w}_2 PCP(t)_1 & w_2 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \bar{w}_1 SST_2 & \cdot & \cdot & \bar{w}_1 PCP(t)_2 & w_1 & \bar{w}_2 SST_2 & \cdot & \cdot & \bar{w}_2 PCP(t)_2 & w_2 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \bar{w}_1 SST_n & \cdot & \cdot & \bar{w}_1 PCP(t)_n & w_1 & \bar{w}_2 SST_n & \cdot & \cdot & \bar{w}_2 PCP(t)_n & w_2 \end{bmatrix}$$

$$Z = \begin{bmatrix} a_1 \\ b_1 \\ r_1 \\ a_2 \\ b_2 \\ r_2 \end{bmatrix}, \text{ and } B = \begin{bmatrix} SF_1 \\ SF_2 \\ SF_3 \\ \cdot \\ \cdot \\ \cdot \\ SF_n \end{bmatrix},$$

where Z is an unknown matrix, to be derived from the parameter sets, which can be solved using following equation.

$$Z^* = A^T (A^T A)^{-1} B \quad (3.6)$$

where A^{-1} is the inverse of matrix “A” and A^T is the transpose of the matrix “A”. The most commonly used training methods for solving this matrix are (1) back-propagation (BP) algorithm (Bishop 1995) and (2) hybrid learning algorithm (Jang 1993). For this study, we used the hybrid approach which combines both the least-squares method and the BP algorithm. The hybrid approach is the most efficient method which trains the data rapidly and also converges much faster due to reduction of dimension of the search space in BP algorithm. The detail mathematical descriptions of the hybrid learning algorithm are given by Jang et al. (1997).

3.4 Material and Methods

3.4.1 Study Area

This study was carried out in the Chickasaw Creek watershed of Mobile County in South Alabama (Figure 3.2). The watershed is characterized by mixed land use categories such as forested, agricultural, industrial and high density residential. The watershed is 250 sq mile in size. The Creek starts in the vicinity of the City of Citronelle and discharges to the Mobile River.

The NCDC climate station nearest to the watershed (Mobile Regional Airport - 015478) has 60 years of observed precipitation and temperature data. The USGS gage 02471001, located near Kushala, has recorded streamflow data for the last 60 years since 1952. The average discharge is 270 cfs and 7Q10 low flow is 27 cfs, taken from ADEM report prepared in 1997 for Chickasaw Creek watershed. Generally, maximum flow occurs in April and low flows occur in October.

3.4.2 CFSv2 Model

The Climate Forecast System (CFS) is a combined oceanic, land, and atmospheric modeling approach for dynamic seasonal prediction system developed at the Environmental Modeling Center at NCEP. The model represents interactions among the oceans, land and atmosphere and came in effect at NCEP in 2004 (Saha et al. 2006). The CFSv2 model is considered a significantly improved model over the previous operational coupled model at NCEP. The model incorporates a number of physical process such as cloud-aerosol radiation, oceanic and sea ice processes, land surface and atmosphere, land data assimilation system, etc. (Saha et al. 2010).

The CFSv2 became operational at NCEP in March 2011. The CFSv2 is a coupled model of different sub-models such as the NCEP Global Forecast System, the Geophysical Fluid Dynamics Laboratory Modular Ocean model, two layer sea ice model, and the four layer NOAA land surface model (Yuan et al. 2011). The CPC seasonal outlook for climatic and hydrologic forecast are based on the CFSv2 model calibration and bias correction after the retrospective forecast of CFSv2 over a period from 1982 to 2010. The prediction skill of the CFSv2 for precipitation and surface air temperature was evaluated by Yuan et al. (2011). The study suggested that the CFSv2 had significantly better skill for surface air temperature and precipitation prediction compared to CFSV1 and is comparable to the European Center for Medium Range Weather Forecast (ECMWF).

The CFSv2 model run starts four times at the interval of six hours of a day, and each run extends up to nine months. Similarly, the monthly mean precipitation and temperature forecast is initiated at an interval of five days (i.e., six times a month). Each time it is associated with four runs; that is, every run at six hour interval resulting in 24 ensembles of monthly precipitation and temperature.

3.4.3 Weather Generator Approach

Weather generators are statistical methods for generating synthetic daily weather data (Wilks and Wilby 1999; Schoof 2008; Baigorria and Jones 2010) representing the future possible climatic conditions and can be used as model inputs for hydrologic forecasts. Weather generator typically uses a random number from the respective probability distribution function and rescales as per the statistical characteristics of the data from the corresponding station (Richardson and Wright 1984). In order to generate synthetic climate data, there are several weather generator

approaches (Richardson 1981; Young 1994; Rajagopalan and Lall 1999). In fact, most of these weather generator approaches have an under-prediction issue, while generating the precipitation, and may not fully address the inter annual variability of precipitation (Schoof 2008). Therefore, I utilized the Geospatial Temporal weather generator (GiST) (Baigorria and Jones 2010) to generate the conditioned weather sequences for different ENSO phases. In other words, I translated ENSO forecasts issued by Climatic Prediction Center (CPC) into daily realizations of rainfall based on historical ENSO trend. The weather generated data were conditioned according to ENSO pattern that I encountered since 1950; for example, I generated precipitation in 1983 (El Niño period) using all the precipitation that I encountered during the historical El Niño period. Likewise, I generated the precipitation in 1985 (La Niña period) using all historical La Niña period. These ENSO-conditioned weather generation approach is consistent with the approach described by Clark et al. (2004) and generated relatively better precipitation for those respective periods (evaluated through the correlation of observed and generated data). However, neutral period is relatively unpredictable because of high variability in precipitation in neutral conditions (not shown). It is noteworthy to mention that the precipitation in the neutral period is not consistent with the precipitation experienced in previous neutral periods. This is because anticipated neutral conditions depend on the initial condition. That is, a neutral condition preceded by an El Niño condition is different than that of a neutral condition preceded by a La Niña condition. Therefore, climate prediction during 1990 to 1995, which was mostly dominated by neutral period, except an El Niño event starting from mid-1991 and ending in mid-1992, was not at the expected level. Since every neutral condition is different in characteristic, I conclude that neutral conditions should not be conditioned; rather, it is appropriate to use all the historical data sets irrespective of the ENSO phases. It is important to mention that the climate model

predictions for precipitation and temperature for the neutral period are also not as good as that of the La Niña and El Niño periods.

3.4.4 Input Data and Preprocessing

The meteorological data processing (Metadapt) tool (Tetra Tech, Inc., 2007) was used to preprocess the input data to derive other climate data, such as, cloud cover, solar radiation, and potential evapotranspiration (PET). PET and solar radiation were calculated using Hamon method (Haith and Shoenaker 1987). Hamon method utilizes maximum and minimum temperature to compute potential ET. Likewise, it utilizes latitude, longitude and cloud cover to compute solar radiation.

3.4.4.1 Input Data Selection

In order to find out the most sensitive inputs, I experimented with different inputs such as wind speed, cloud cover, solar radiation and ET (at one month lead time) apart from precipitation and temperature in a Multi Linear Regression (MLR) model. Solar radiation is estimated using cloud cover and is crucial for stream temperature and snow simulation. Since snowfall in coastal regions of Alabama is rare, both cloud cover and solar radiation were not significant at 5% significance level. ET at one month lead time was a sensitive input because higher ET in the immediate preceding month will cause less surface runoff for the forthcoming month. Wind speed at one month lead time is probably not needed if the precipitation at one month lead time is included. However, the model suffered because of inadequate representation of spatial variability of the precipitation; that is, a single precipitation station that was utilized for simulation was located 20 miles away from the watershed outlet and also the missing data were replaced using

the data from nearby stations (COOP-ID 01583, 01084). This could be the reason that MLR model depicted wind speed (one month lead time) as the significant model input (p-value < 0.05).

Besides, for input data selection, I evaluated the correlation of streamflow with each input and those which depicted the best correlation were considered for possible model inputs (Table 3.1). Since the watershed is 357 km² in size, the precipitation from the immediate previous month is equally vital for simulating streamflow, particularly, due to delayed ground water contribution as base flows. For this, I experimented with precipitation at different lead times as model inputs in MLR model. The precipitation that occurred three months before had no impact on the model (zero correlation), and therefore, monthly precipitation occurring within the immediate past three months (PCP (t-1), PCP (t-2), PCP (t-3)) were included in the model. Here, PCP and t stand for precipitation and monthly time step, respectively. The correlation of SST with streamflow for all lengths of record was low (Table 3.1). This is due to the different ENSO characteristics in different seasons. For example, SST is strongly correlated with streamflow in winter season ($r=0.48$) when only ENSO events are considered. This correlation is 0.14 for spring season but -0.17 for fall season. ENSO essentially shows opposite characteristics in Aug-Oct (ASO) as compared to Jan-March (JFM) and April-June (AMJ). This indicates that SST is correlated with streamflow most of the season throughout the year but the strength and characteristics of the correlation varies from season to season.

3.4.5 Model Training, Validation and Testing

The ANFIS model simulation was implemented using the MATLAB platform. Initial parameters were estimated using fuzzy subtractive clustering method. The model parameters

(clustering radius, membership function) were optimized through repeated trial and error procedure sought by maximizing Nash-Sutcliffe efficiency for the given datasets. ANFIS model was trained with 60 epochs. Each model training was optimized with its minimum validation error (maximum validation performance) in order to avoid the overtraining of the model. The streamflow simulation was started from 1/1/1952 and the model was initiated to forecast streamflow for 7 years from 1/1/1982. The Nash-Sutcliffe efficiency (NSE) for training (1952 to 1976) and validation of the model (1976 to 1981) was 0.61 and 0.8, respectively. The streamflow forecast was post validated for a period from 1/1/1982 to 12/31/1988. Since the majority of the meteorological data (wind speed, cloud cover etc.) were not available after 1996, the model simulation was simply terminated on 12/31/1995. The streamflow forecast from 1989 to 1995 was utilized for bias correction.

3.4.6 Bias Correction

Bias correction was implemented to remove the systematic error for the enhancement of the forecasting skill (Hashino et al. 2006). From the comparison of observed streamflow to the forecasted streamflow for the 14-year period (1982 to 1995), some error was found systematically associated with the forecast (under prediction of the flow) for both data (CFSv2 and weather generator) sources (not shown). The regional, temporal and spatial discrepancy between predicted and observed precipitation and temperature was reflected in streamflow. In order to resolve this issue, the mean predicted streamflow was rescaled with the observed streamflow between 1989 to 1995, and the forecasting error associated with the climate model and the weather generator data during this period was adjusted for another time period (1982 to 1988). There are many bias correction methods discussed in the literature for correcting

systematic discrepancy. I used the quantile mapping method which has been extensively used for bias correction in streamflow forecasting (Wood et al. 2002).

3.4.6.1 Quantile Mapping Method

In this method, bias is generally removed using the cumulative distribution function (CDF) of the observed and analogous historical simulation of streamflow. The predicted and observed streamflow data for the same span of the length of the record were used to develop a “Quantile map”. At first, the CDF of the predicted streamflow to be corrected (1982 to 1988) was computed. Similarly, the CDF of observed and predicted streamflow over a same length of the record (1988 to 1996) was determined. Then, the bias error was transformed through the cumulative distribution function using unitary method. Alternatively, one can calculate the difference for each individual quantile in the forecasted streamflow data (1988 to 1996), and can adjust that difference to the same quantile in the data between 1982 to 1988. There are mainly two reasons to select the period from 1982 to 1988 as a period of forecast and utilize data from 1988 to 1996 for bias correction: (i) the CFSv2 forecast starts from 1982; (ii) the period from 1982 to 1988 covers El Niño, La Niña and neutral periods, and I wanted to evaluate the performance of the model in different ENSO phases. Alternatively, bias correction from 1982 to 1988 can be applied to the period from 1988 to 1996. The bias correction eliminated the systematic error associated with the forecast and improved the model performance in terms of Mass Balance Error (MBE). There are two commonly used approaches for bias correction: (i) first, execute a bias correction for climate data and then apply bias corrected data in model for streamflow forecast called preprocessing; (ii) alternatively, apply climate data in ANFIS model without bias correction and perform a bias correction for the streamflow after realizing the

systematic error which is called post processing. Finally, in order to be consistent with the Ensemble Streamflow Prediction (ESP) operational at NWS, I implemented the second approach though our analysis suggested that results were essentially same irrespective of the approaches I used.

The limitation of the quantile mapping approach is that the error will be adjusted for future as per the climatic condition (wet periods or dry periods) that was encountered during the historical simulation. For example, an “extremely wet hydrologic condition” in the past will lead to the extreme wet condition for the future because the transformation process will tend to replicate the extremity of the historical simulation to the future output due to the mapping of extremely observed value. Therefore, identification of analogous period in historical simulation with the future expected period (wet and dry) is crucial for bias correction.

3.5 Result and Discussion

The CFSv2 reanalysis and reforecast data spanning from 1982 to 2009 were plotted against the observed data at the local station, the Mobile Regional Airport (Coop ID-015478) as shown in Figure 3.3. The predicted precipitation shows better correlation with observed precipitation in the winter season ($r=0.44$) compared to April-June ($r=0.22$), July-Sept ($r=0.38$) and Oct-Dec ($r=0.35$). Temperature shows a good agreement with the observed temperature ($r=0.96$), which is consistent with previous studies (Yuan et al. 2011b). Similarly, the precipitation and temperature generated using the weather generator approach and its temporal correlation with observed data is shown in Figure 3.4. The Ensemble of retrospective SST forecast from the CFSv2 model at one month and three month lead time was compared with observed SST data derived from Extended Reconstructed Sea surface temperature (ERSST.v3b)

analysis. The CFSv2 forecast skill for SST at one month lead time, after bias correction, was relatively better ($r=0.77$) than that of three month lead time ($r=0.49$).

Figure 3.5 shows the streamflow forecast at one month lead time using the CFSv2 model data starting from 1982 and ending in 1988. The statistical parameter to measure the performance of forecast (Table 3.2) indicates that the model was satisfactorily forecasting at one month lead time and Figure 3.5 suggests that the forecast is satisfactory for both high and low flows. The statistical parameters measuring the performance of the model were RMSE (Root Mean Square Error), R-Square, PE (percentage error of the average observed and forecasted streamflow) and CO (ratio of standard deviation of the forecasted and observed streamflow). Figure 3.6 shows the average monthly streamflow forecast at one month lead time using the ENSO conditioned weather generated data. The streamflow forecast at one month lead time using the CFSv2 data was relatively better than the weather generator approach (Table 3.2). In fact, overall CFSv2 forecast for precipitation at one month lead time was better than the weather generated precipitation data. This assessment was based on the correlation of observed precipitation with the CFSv2 precipitation ($r = 0.35$) and the weather generated precipitation ($r = 0.22$). Overall, CFSv2 data are better than weather generated data except for few events. For example, the average monthly precipitation predicted by the CFSv2 model for March 1983 deviated slightly from the observed precipitation. Likewise, the weather generator under predicted precipitation for March 1983; therefore slightly degraded the overall forecast in terms of Nash-Sutcliffe efficiency (NSE) in that particular month. If this one month precipitation is ignored, the performance of weather generator approach is comparable to CFSv2 forecast (NSE=0.54). This indicates to the fact that the climate model precipitation downscaled at local station can furnish promising result for monthly forecast at one month lead time compared to the

weather generated data; at the same time, weather generator can be considered as an alternative approach.

It is not surprising to realize some unusual precipitation events deviated from the observed data when data from climate model were utilized at the local station. This uncertainty associated with the deterministic forecast justifies the relevancy of the probabilistic streamflow forecast (Figure 3.7). Figure 3.7 demonstrates the probability of streamflow occurring below that range using 24 ensembles of CFSv2 monthly average precipitation and temperature; that is, 98 percentile flows suggest that there is a 98 percent chance of the streamflow remaining below the corresponding streamflow magnitude. I tried to forecast Surface Runoff (SR) and Base Flow (BF) separately. However, I do not recommend for separate model to forecast SR and BF because users need to run two separate models each time using 24 ensemble families of CFSv2 forecasts for probabilistic streamflow forecast which is tedious and cumbersome. Essentially, the overall forecasting skill of the model did not improve simply because SR and BF were forecasted separately.

The streamflow forecast at three months lead time using weather generated data is shown in Figure 3.8. One month's lead time forecast was better than three months lead forecast irrespective of the approach I used (Table 3.2). The correlation of observed precipitation data with weather generated precipitation data ($r= 0.23$) was better than that of CFSv2 data ($r= 0.11$) for three month lead time. Since the forecasting skill of the CFSv2 model for precipitation beyond one month lead time decreased (not shown), weather generated data is recommended as opposed to CFSv2 data for streamflow forecasting at higher lead time (more than one month). The operational streamflow forecast implemented for Chickasaw Creek watershed suffered from the quality of precipitation forecasted by both approaches for three months lead time forecast;

that is, three month lead forecast is not as good as one month lead forecast. I need to clearly infer the fact that the model demonstrates its better skill for low flow prediction than high flow for three month lead time. There are two important points to mention regarding this inference: (i) for three month lead time, weather generator data were used which were relatively better for low flow, but had an issue of under prediction for the peak flow (Schoof 2008); and ii) as I discussed earlier, I am actually interested in monthly low flow forecast rather than monthly high flow forecast because the low flow condition is the most crucial condition when serious hydrologic droughts cause water resources scarcity. The forecasting skill of the model for 50 percentile low flow was evaluated with few more statistics such as Hit rate and False Alarm Ratio. Readers can find the details about these statistics in literature (Martina et al. 2006). The Hit rate (0.87) and False Alarm Ratio (0.30) suggests that the model is satisfactorily forecasting low flows. Since weather generator data are synthetically generated data for that year, it is essentially possible to estimate low flows at 9 months lead time assuming that the discrepancy in SST prediction for 3 months lead time and 9 months (SST can be predicted up to 9 months) lead time does not significantly affect forecast.

3.6 Summary and Conclusion

In this research, I presented a systematic assessment of the two approaches operational at NWS: (i) the weather generator approach, and ii) the climate model approach for streamflow forecasting for one to three months advance in time. I applied SST in the ANFIS model to incorporate the interannual climate variability for streamflow forecasting. I experimented with precipitation inputs at different lead times to simulate streamflow that correspond to base flow. In addition, I carried out the screening of the input variables to reduce the input data sets. The

best combinations of input data sets were fed into the ANFIS model, and the hybrid algorithm was applied to estimate the model parameters. The model was trained since 1952 to 1978 and validated from 1979 to 1981. The forecast was initiated since 1982 using two different sources of precipitation and temperature (the CFSv2 and the weather generator approach). The streamflow forecast after 1989 was applied for bias correction using quantile mapping methods. The performance of the model for one month lead time forecast was satisfactory, and the predictions were in good agreement with the observed streamflow, which was tested using different statistical criteria. The research concludes that CFSv2 model could be a better choice than ENSO conditioned weather sequences for a moderate size watershed for a one month lead time forecast. However, the CFSv2 data for three month lead time forecast is not a suitable approach because the prediction skill of climate model degrades with increased lead time. Therefore, this research categorically specifies that the ENSO conditioned weather generator is a better approach for long range hydrologic forecast, especially for low flow at three months lead time. Since the forecasting skill of the climate model is better at capturing both high flows and low flows, the CFSv2 model is recommended for one month lead time. Alternatively, the weather generator approach can be utilized for one month lead time forecast as a viable substitute to the climate model.

Table 3.1. Input parameters showing best correlation with streamflow in Chickasaw Creek

Inputs	Correlation	Remarks
PCP(t)	0.52	Precipitation at time (t)
WS(t-1)	0.36	wind speed at time (t-1)
PCP(t-1)	0.25	Precipitation at time (t-1)
PCP(t-2)	0.11	Precipitation at time (t-2)
PCP(t-3)	-0.06	Precipitation at time (t-3)
SST (t)	-0.10(0.48*)	Sea surface temperature at time (t)
T (t)	-0.28	Temperature at time (t)
ET (t-1)	-0.36	Evapotranspiration at time (t-1)

* Indicates the correlation in winter season. Summer and fall season the ENSO pattern is different. 't' indicates the monthly time step.

Table 3.2. Streamflow forecast after bias correction using weather generator approach and CFSv2 data

Weather generator				
Lead	correlation (γ)	RMSE	Percentage Error (PE)	CO
One month	0.75	4.4	10.5%	1.24
Three month	0.52	5.2	-9.8%	0.99
CFSv2 model				
One month	0.77	3.8	7 %	1.09

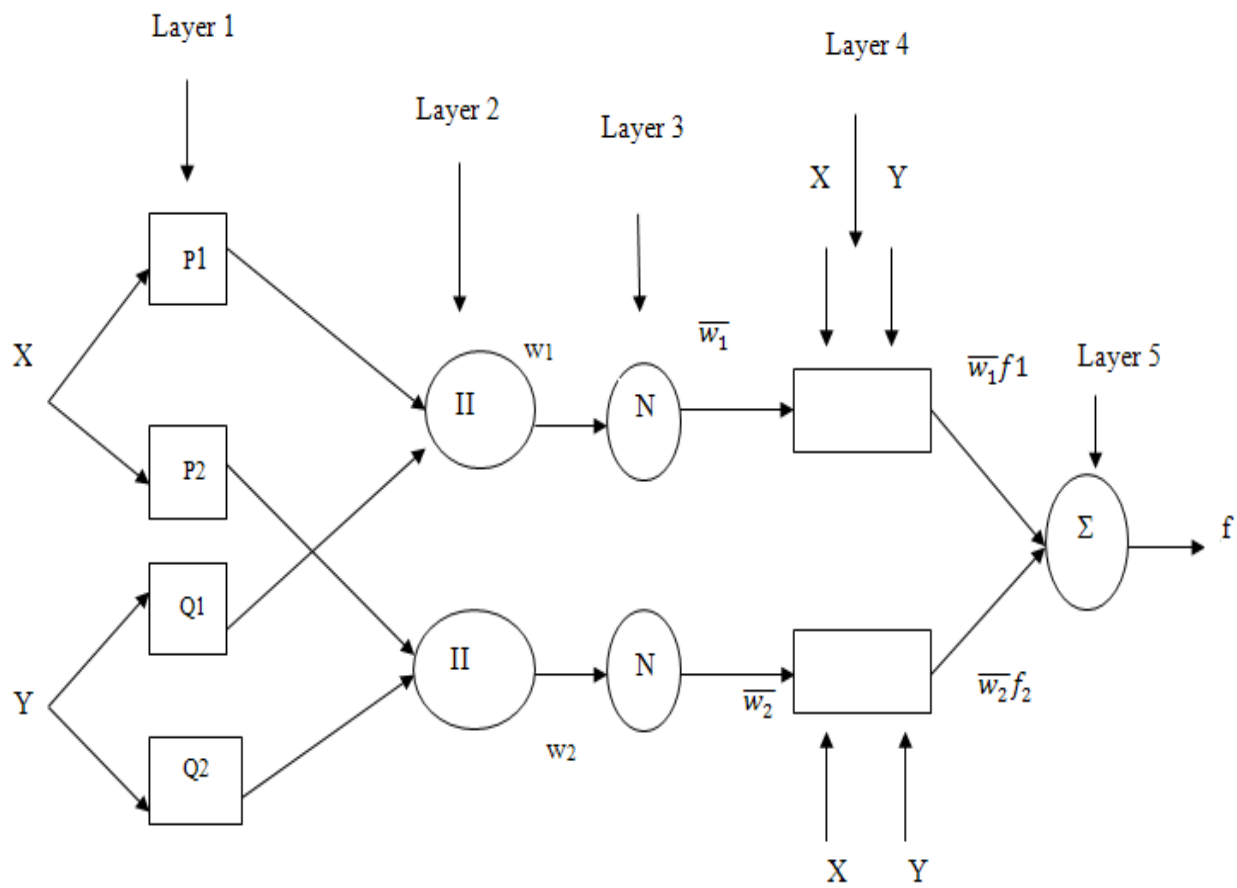


Figure 3.1. ANFIS structure

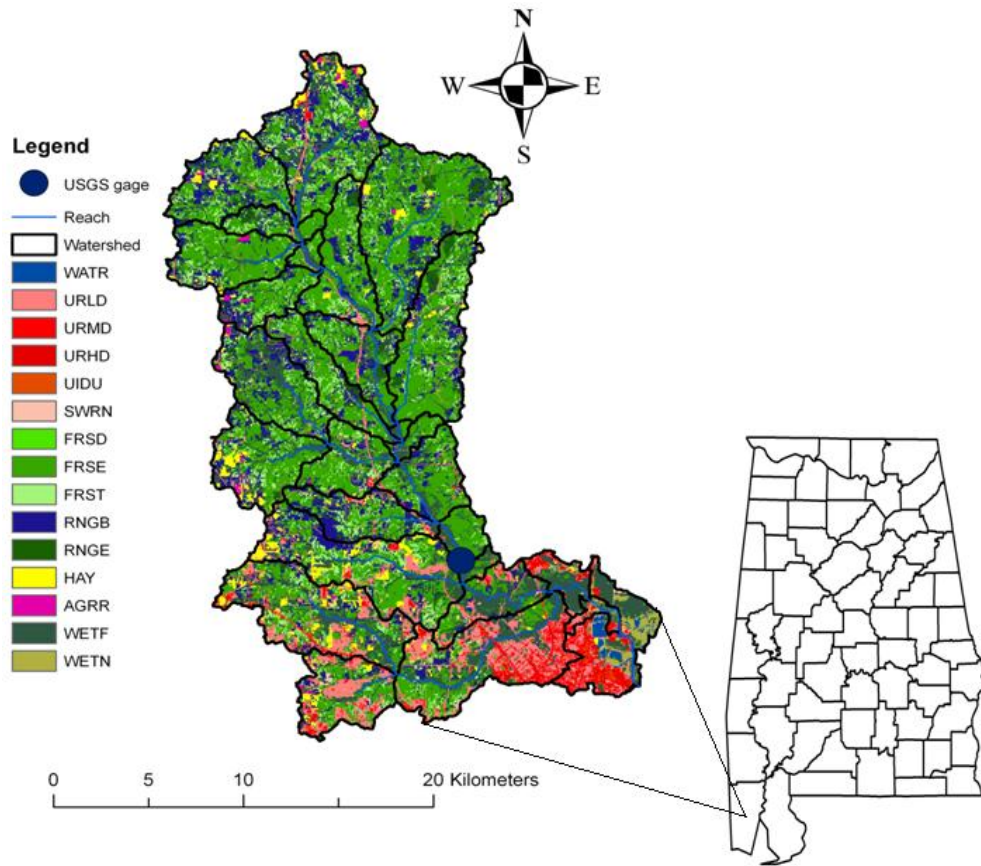


Figure 3.2. Chickasaw Creek watershed in Mobile County of South Alabama.

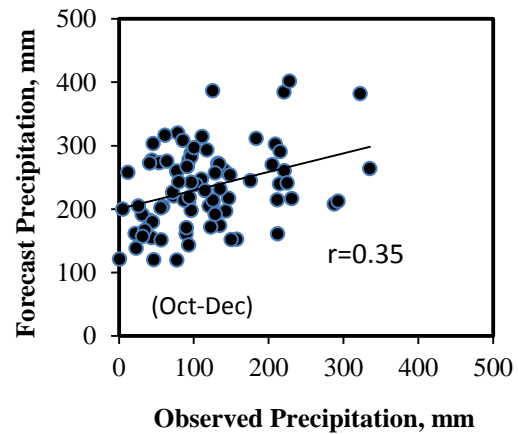
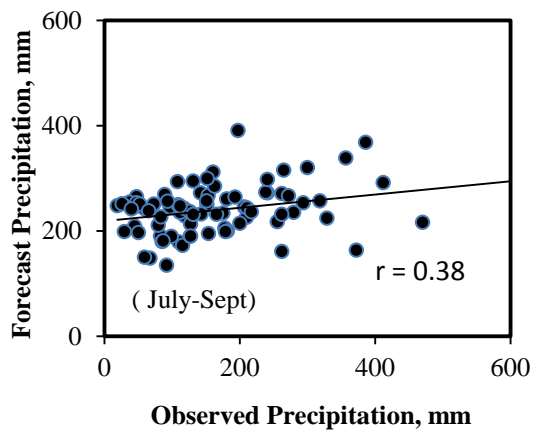
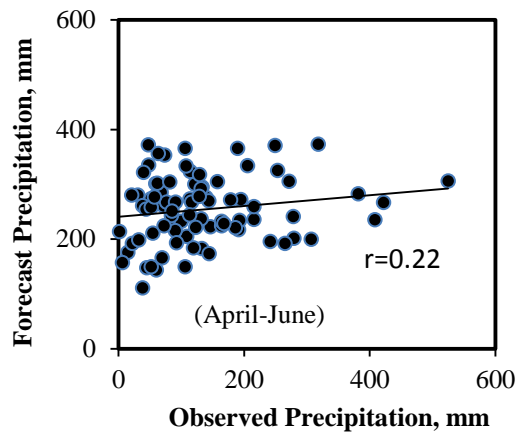
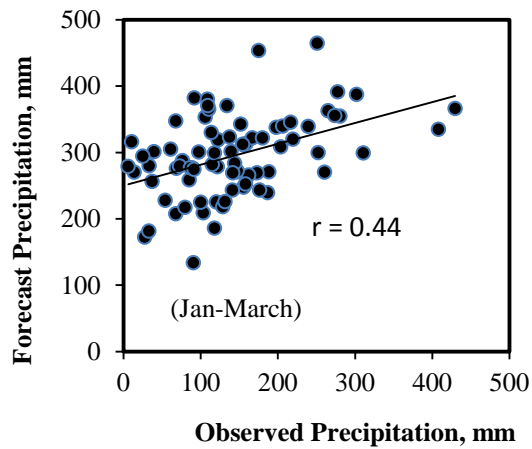
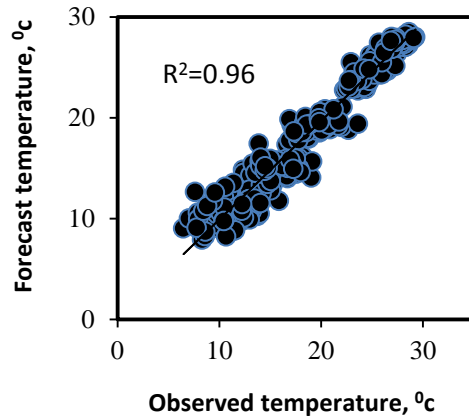
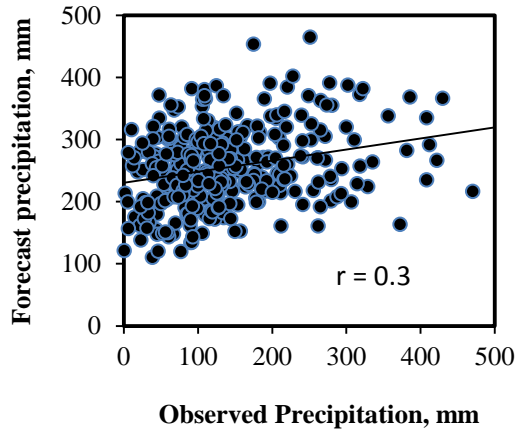


Figure 3.3. Comparison of observed and CFSv2 forecast mean monthly precipitation and temperature (forecast lead time is one month and period is 1982 to 2009).

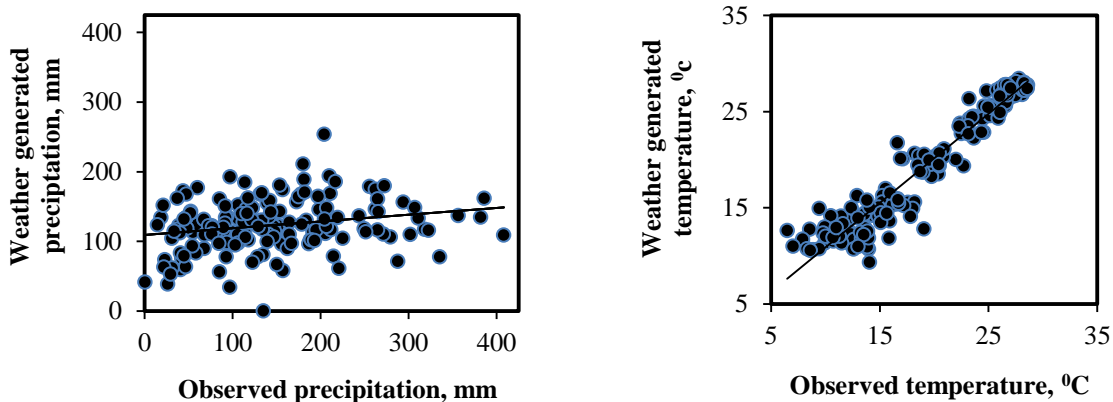


Figure 3.4. Comparison of observed precipitation and temperature with weather generated precipitation and temperature (average monthly).

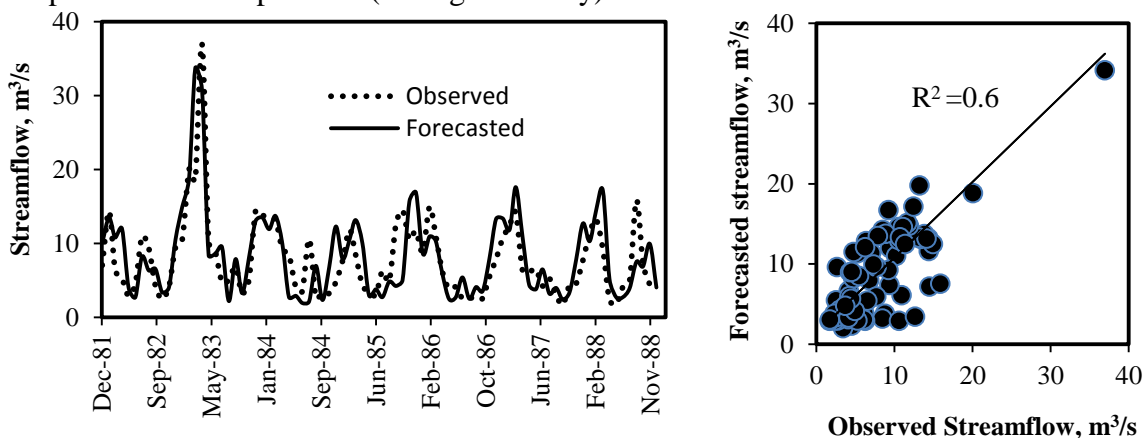


Figure 3.5. Streamflow forecast at one month lead time using CFSv2 retrospective forecast data.

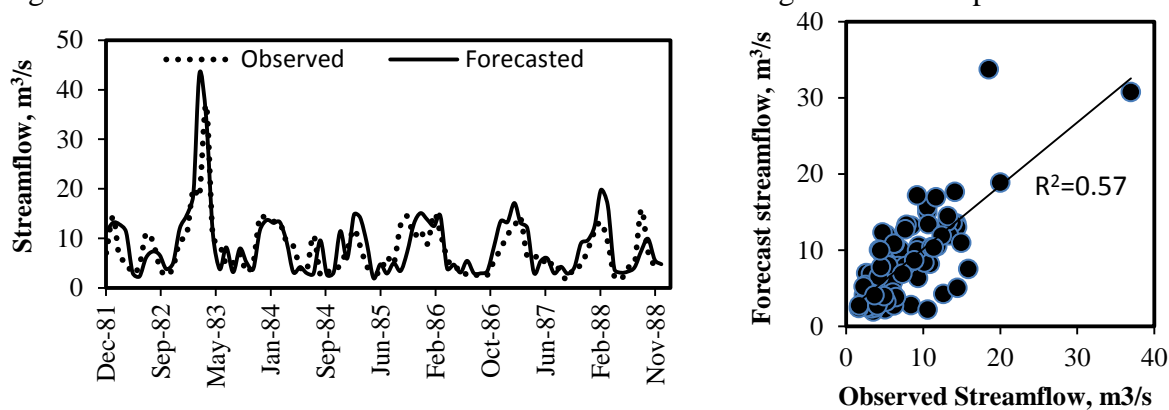


Figure 3.6. Streamflow forecast at one month lead time using weather generated climate data.

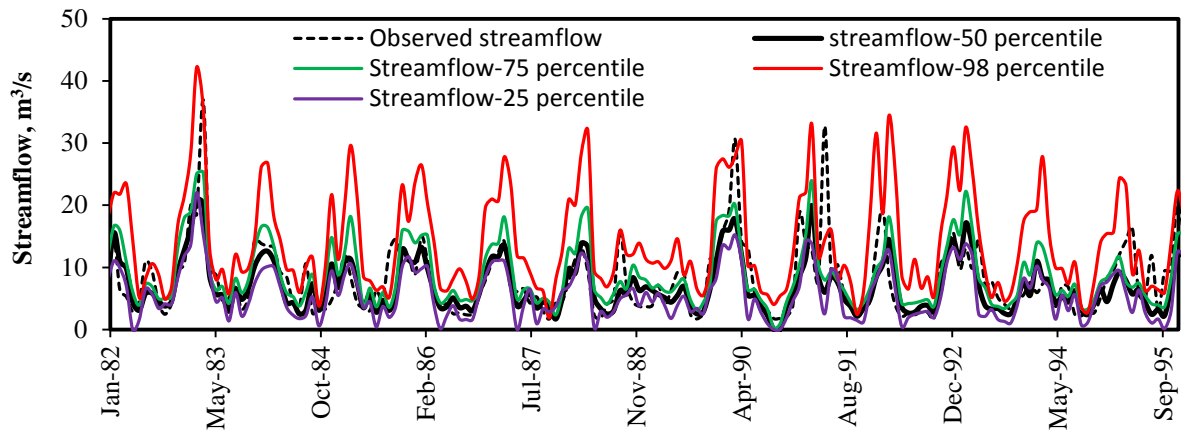


Figure 3.7. Probabilistic streamflow forecast at one month lead time using CFSv2 data.

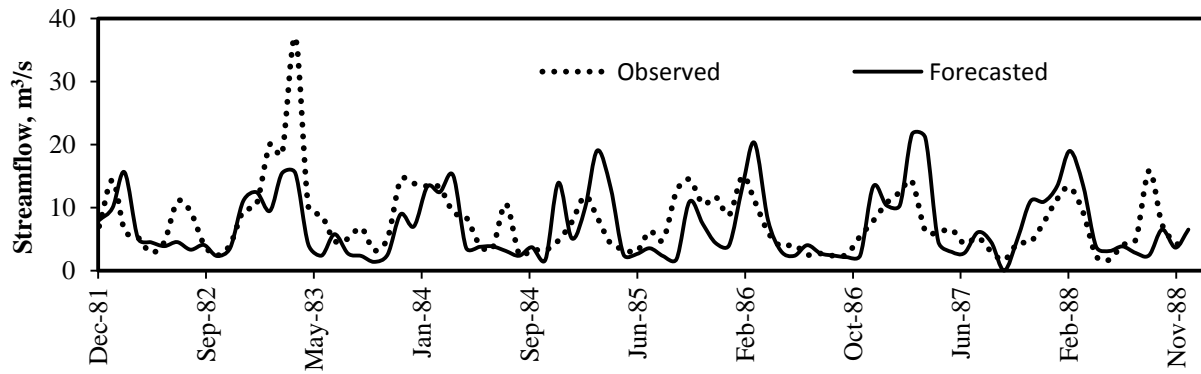


Figure 3.8. Streamflow forecast at three months lead time using weather generated climate data.

3.7 References

- Araghinejad, S., Burn, D. H., and Karamouz, M. (2006). "Long-lead probabilistic forecasting of streamflow using ocean-atmospheric and hydrological predictors." *Water Resources Research*, 42(3), W03431.
- Baigorria, G. A., and Jones, J. W. (2010). "GiST: A stochastic model for generating spatially and temporally correlated daily rainfall data." *Journal of Climate*, 23(22), 5990-6008.
- Bishop, C. "Neural Networks for Pattern Recognition, 1995, 18 pages." Clarendon Press.
- Clark, M. P., Gangopadhyay, S., Brandon, D., Werner, K., Hay, L., Rajagopalan, B., and Yates, D. (2004). "A resampling procedure for generating conditioned daily weather sequences." *Water Resour. Res.*, 40(2), W04304.
- Eldaw, A. K., Salas, J. D., and Garcia, L. A. (2003). "Long-range forecasting of the Nile River flows using climatic forcing." *Journal of Applied Meteorology*, 42(7), 890-904.
- Gim�iz-Fortis, S., Esteban-Parra, M., Trigo, R., and Castro-Dķez, Y. (2010). "Potential predictability of an Iberian river flow based on its relationship with previous winter global SST." *Journal of Hydrology*, 385(1-4), 143-149.
- Gutierrez, F., and Dracup, J. (2001). "An analysis of the feasibility of long-range streamflow forecasting for Colombia using El Niño-Southern Oscillation indicators." *Journal of Hydrology*, 246(1-4), 181-196.
- Haith, D. A., and Shoenaker, L. L. (1987). "Generalized watershed loading functions for streamflow nutrients." *JAWRA Journal of the American Water Resources Association*, 23(3), 471-478.
- Hamlet, A. F., and Lettenmaier, D. P. (1999). "Columbia River streamflow forecasting based on ENSO and PDO climate signals." *Journal of water resources planning and management*, 125(6), 333-341.
- Hashino, T., Bradley, A., and Schwartz, S. (2006). "Evaluation of bias-correction methods for ensemble streamflow volume forecasts." *Hydrology and Earth System Sciences Discussions*, 3(2), 561-594.
- Hsieh, W. W., Li, J., Shabbar, A., and Smith, S. (2003). "Seasonal prediction with error estimation of Columbia River streamflow in British Columbia." *Journal of water resources planning and management*, 129, 146.
- Jang, J. S. R. (1993). "ANFIS: Adaptive-network-based fuzzy inference system." *Systems, Man and Cybernetics, IEEE Transactions on*, 23(3), 665-685.

- Jang, J. S. R., Sun, C. T., & Mizutani, E. (1997). *Neuro-fuzzy and soft computing: "A computational approach to learning and machine intelligence."* Eaglewood cliffs,NJ: Prentice-Hall.
- Kalra, A., and Ahmad, S. (2009). "Using oceanic-atmospheric oscillations for long lead time streamflow forecasting." *Water Resources Research*, 45(3), W03413.
- Keener, V., Feyereisen, G., Lall, U., Jones, J., Bosch, D., and Lowrance, R. (2010). "El-Niño/Southern Oscillation (ENSO) influences on monthly NO₃ load and concentration, streamflow and precipitation in the Little River Watershed, Tifton, Georgia (GA)." *Journal of Hydrology*, 381(3-4), 352-363.
- Krstanovic, P., and Singh, V. (1991). "A univariate model for long-term streamflow forecasting." *Stochastic hydrology and hydraulics*, 5(3), 189-205.
- Liang, X., Wood, E. F., and Lettenmaier, D. P. (1996). "Surface soil moisture parameterization of the VIC-2L model: Evaluation and modification." *Global and Planetary Change*, 13(1), 195-206.
- Martina, M., Todini, E., and Libralon, A. (2006). "A Bayesian decision approach to rainfall thresholds based flood warning." *Hydrology and Earth System Sciences*, 10(3), 413-426.
- Moriasi, D., Arnold, J., Van Liew, M., Bingner, R., Harmel, R., and Veith, T. (2007). "Model evaluation guidelines for systematic quantification of accuracy in watershed simulations."
- Mukerji, A., Chatterjee, C., and Raghuvanshi, N. S. (2009). "Flood forecasting using ANN, neuro-fuzzy, and neuro-GA Models." *Journal of Hydrologic Engineering*, 14, 647.
- Pramanik, N., and Panda, R. K. (2009). "Application of neural network and adaptive neuro-fuzzy inference systems for river flow prediction." *Hydrological Sciences Journal*, 54(2), 247-260.
- Rajagopalan, B., and Lall, U. (1999). "A k-nearest-neighbor simulator for daily precipitation and other weather variables." *Water Resources Research*, 35(10), 3089-3101.
- Richardson, C. W. (1981). "Stochastic simulation of daily precipitation, temperature, and solar radiation." *Water Resources Res.*, 17(1), 182-190.
- Richardson, C. W., and Wright, D. A. (1984). "WGEN: A model for generating daily weather variables." *U.S. Department of Agriculture, Agricultural Research Service*, Washington, D.C., ARS-8, 88p.
- Rogers, J. C., and Coleman, J. S. M. (2003). "Interactions between the Atlantic Multidecadal Oscillation, El Niño/La Niña, and the PNA in winter Mississippi valley streamflow." *Geophys. Res. Lett.*, 30(10), 1518.

- Saha, S., Nadiga, S., Thiaw, C., Wang, J., Wang, W., Zhang, Q., Van den Dool, H., Pan, H. L., Moorthi, S., and Behringer, D. (2006). "The NCEP climate forecast system." *Journal of Climate*, 19(15), 3483-3517.
- Saha, S., Moorthi, S., Pan, H. L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., and Behringer, D. (2010). "The NCEP climate forecast system reanalysis." *Bulletin of the American Meteorological Society*, 91(8), 1015-1057.
- Schoof, J. (2008). "Application of the multivariate spectral weather generator to the contiguous United States." *Agricultural and Forest Meteorology*, 148(3), 517-521.
- Sheridan, J. (2002). "Peak flow estimates for coastal plain watersheds." *Transactions of the ASAE*, 45(5), 1319-1326.
- Tetra Tech. Inc. (2007). "User's manual for meteorological data analysis and preparation tool (Metadept)". Prepared for USEPA by tetra tech., 10306 Eaton Place, Suite 340, Fairfax, VA 22030.
- Wang, G., and Eltahir, E. A. B. (1999). "Use of ENSO information in medium-and long-range forecasting of the Nile floods." *Journal of Climate*, 12(6), 1726-1737.
- Wang, W., Chen, M., and Kumar, A. (2010). "An assessment of the CFS real-time seasonal forecasts." *Weather and Forecasting*, 25(3), 950-969.
- Wilks, D. S., and Wilby, R. L. (1999). "The weather generation game: a review of stochastic weather models." *Progress in Physical Geography*, 23(3), 329-357.
- Wood, A. W., Maurer, E. P., Kumar, A., and Lettenmaier, D. P. (2002). "Long-range experimental hydrologic forecasting for the eastern United States." *J. Geophys. Res.*, 107(D20), 4429.
- Wood, A. W., and Lettenmaier, D. P. (2006). "A test bed for new seasonal hydrologic forecasting approaches in the western United States." *Bulletin of the American Meteorological Society*, 87(12), 1699-1712.
- Yan, H., Zou, Z., and Wang, H. (2010). "Adaptive neuro fuzzy inference system for classification of water quality status." *Journal of Environmental Sciences*, 22(12), 1891-1896.
- Young, K. C. (1994). "A multivariate chain model for simulating climatic parameters from daily data." *Journal of Applied Meteorology*, 33, 661-671.
- Yuan, X., Wood, E. F., Luo, L., and Pan, M. (2011a). "A first look at Climate Forecast System version 2 (CFSv2) for hydrological seasonal prediction." *Geophys. Res. Lett.*, 38, L13402.

Yuan, X., Wood, E. F., Luo, L., and Pan, M. (2011b). "A first look at Climate Forecast System version 2 (CFSv2) for hydrological seasonal prediction." *Geophysical research letters*, 38(13), L13402.

Chapter 4. Incorporating Climate Variability for Point Source Discharge Permitting In a Complex River System

4.1 Abstract

The conventional point source discharge permitting approach, referred to as the National Pollutant Discharge Elimination System (NPDES), is based on either a regulatory low flow (hydrologic, biological, or seasonal) criterion or on a Hydrograph Controlled Release (HCR) approach. Regulatory low flows are often estimated using empirical equations because of the lack of historical flow data. Overestimated low flows may threaten water quality protection, while underestimated low flows can result in uneconomical wastewater treatment. Since uncertainty in low flow estimations is caused by climate variability, uncertainty in the permitting process can be reduced through explicit incorporation of climate information. Therefore, the objective of this study was to demonstrate how the NPDES permitting process can be improved through the incorporation of climate information. A dissolved oxygen (DO) model was developed for the Chickasaw Creek Watershed located in Southeast Alabama using the Loading Simulation Program C++ (LSPC) coupled with a hydrodynamic and water quality model (EPD-RIV1). Models were calibrated and validated for flow, stream temperature, DO and other water quality variables. DO and stream temperature variations were examined for the historic, climate variability-causing events of La Niña and El Niño, using a number of statistical criteria and non-parametric tests to develop toxicity and DO-based criterion for ammonia. The analysis identified December-April for El Niño and August-October for La Niña as periods of high assimilation.

May-July for La Niña and August-October for El Niño were identified as periods of restrictive point source discharge. Analysis suggested that El Niño Southern Oscillation (ENSO) forecasts provide sufficient warning for the impending drought for reducing point source discharges because low flows in summer months is a function of winter and spring sea surface temperature, precipitation, and streamflow due to its autocorrelation and cross correlation characteristics.

4.2 Introduction

State and federal agencies maintain surface water quality by regulating point and nonpoint source discharges through restrictions on the release of pollutants. Often, because of challenges associated with regulating nonpoint source discharges, point source discharges are regulated. However, regulating point source discharges also involves numerous technical issues because complete treatment of wastes is often expensive and impractical. Therefore, estimation of the dilution needed to regulate point source discharges is important.

Point source discharge is regulated using the National Pollutant Discharge Elimination System (NPDES) permits written by state environmental regulatory agencies for lake and stream water quality protection. To maintain stream water quality, some states have approved anti-degradation rules for water quality regulation, which suggest utilizing a variable loading scheme (e.g., the Hydrograph Controlled Release (HCR) approach) in order to take the benefit of increased assimilative capacity of streams during high flow periods. The HCR approach is practiced in the Southeast USA and is consistent with the Total Maximum Daily Load (TMDL) policy of the U.S. Environmental Protection Agency (USEPA) (Conrads et al. 2003). However, when extreme drought conditions persist for a long time, point source dischargers relying on the HCR approach may have to hold their discharge for a long time due to extended low flow

conditions in streams. This issue with the HCR approach can be better handled by incorporating reliable climate forecasts, such as, El Niño Southern Oscillation (ENSO) forecasts, which are issued by National Oceanic and Atmospheric Administration (NOAA) with a six to nine months lead time.

Another commonly adopted approach for NPDES permitting is based on regulatory low flows that are analyzed based on historical flow data. Low flow estimation is the quantitative connection between stream standards that maintain designated use of a water body and the permit limits that maintain effluent quality. Both of these components are equally important because of two reasons: (1) an overestimated low flow increases the risk that stream may not receive adequate protection for designated use and aquatic life and (2) an underestimated low flow will unnecessarily increase the cost of wastewater treatment. Accurate low flow estimation requires long term data sets that are rarely available. A few past attempts (Saunders III and Lewis Jr 2003; Saunders III et al. 2004) to assess the minimum years of record required for proper estimation of regulatory low flows examined the connection between climate variability and hydrological or biological low flow estimates and suggested that at least 10 to 20 years data are required for proper estimation of regulatory low flows. Estimation of low flows based on less than ten years of data gives biased results and threatens water quality protection.

Uncertainty associated with the permitting process due to limited data can be reduced through a better understanding and interpretation of the linkage between low flows and climate variability (Saunders III and Lewis Jr 2003). Since inter-annual climate variability resulting from the coupled ocean-atmosphere phenomena (ENSO) has a significant effect on streamflow in the Southeast USA (Ropelewski and Halpert 1986), climate variability information should be

explicitly utilized for the interpretation of water quality in rivers (Scarsbrook et al. 2003) and to improve the conventional approach of NPDES permitting.

The conventional approach of using 7Q10 for permitting may not capture extremely low flow conditions (i.e. hydrologic drought) due to two reasons: (1) estimation of the low flow condition at a specific site is sensitive to the extent of the data record (Saunders III and Lewis Jr 2003) and (2) there is always a possibility of encountering flows lower than 7Q10 because this flow is the non-exceedence probability for the ten year recurrence interval. This is the primary reason why point source discharges could not be properly regulated during the extreme droughts of the years 2000 and 2007 in the Southeast USA. Many fish and other forms of aquatic life were under stress and died due to extremely low DO levels, high stream temperatures, and low streamflow (Johnson et al. 2001).

High stream temperatures and high pH levels can cause ammonia nitrogen released from waste water treatment plants (WWTP) to become toxic to fish and other aquatic life. In order to protect fish and aquatic life, USEPA has devised an approach (equations) to tailor ammonia nitrogen discharge based on streamflow temperature, pH and streamflow. However, climate variability has the prospective of creating temporal variations in annual low flow (Stahl and Demuth 1999; Mosley 2000), stream temperature, and pH. Using NOAA ENSO forecasts, specifically the teleconnection of ENSO with DO level and stream temperature, it may be possible to tailor ammonia nitrogen discharge from WWTP's, especially when the conventional approach has the potential to miss an extreme hydrologic drought. Extreme hydrologic drought conditions may be overlooked and water quality protection may be threatened due to the inadequate representation of climate variability while estimating low flows. This is especially

true when low flow estimation is based on a regression equation developed for specific region (Kroll 1992; Ries and Friesz 2000).

Few past studies have explored the influence of climate variability on low flow estimations and how this affects stream water quality protection (Saunders III et al. 2004). Considering the significant effect of climate variability on water quality, the objective of this research was to demonstrate how the NPDES permitting process can be improved through the incorporation of climate information. The research explored extremely low flows and their auto correlation and cross correlation characteristics with ENSO. The research further analyzed the historic ENSO events together with streamflow, stream temperature, and DO to evaluate the toxicity and DO-based ammonia permit in different ENSO phases so that the period of high flows and low flows can be assessed for assimilation of pollutant discharges. The research also analyzed extreme high and low flow conditions for inter-seasonal transfer of pollutant loads, which is helpful for WWTP's operating under the HCR approach. Specifically, this research demonstrated the usefulness of integrating the use of climate information with conventional methods for ammonia nitrogen permitting.

4.3 Theoretical Background

4.3.1 Ammonia Nitrogen: Basic Concept

Ammonia is regarded as one of the most important contaminants in the aquatic environment. The primary reason is that it is highly toxic for aquatic life in surface water systems (Russo et al. 1985). Many effluents have to be treated extensively so that ammonia concentration limits in surface waters are not exceeded. The main sources of ammonia in surface waters are municipal or industrial wastes, in addition to agricultural runoff, nitrogen fixation and

animal excretion of nitrogenous wastes. Ammonia, generally expressed as total ammonia, consists of two components: ammonium (NH_4^+), which is more available and is not toxic, and non-dissociated or un-ionized ammonia (NH_3), which is toxic (Thurston et al. 1984a). The ratio of these species in a given aqueous solution depends on pH and temperature (Emerson et al. 1975; Erickson 1985; Thurston 1990; Wood and Evans 1993). If pH and temperature are known, ammonium and ammonia fraction can be calculated in freshwater based on salinity (Whitfield 1974; Hampson 1977) using the Henderson-Hasselbach equation.

$$\text{NH}_4^+ = \frac{\text{Total Ammonia}}{(1 + \text{antilog}(\text{pH} - \text{pK}))} = \text{Total Ammonia} - \text{NH}_3 \quad (\text{Wood 1993}) \quad (4.1)$$

$$\text{pk} = 0.09018 + \left(\frac{2729.92}{273.2 + T} \right) \quad (\text{Emerson et al. 1975}) \quad (4.2)$$

where T is temperature in $^{\circ}\text{C}$.

Higher pH and higher temperatures result in a higher proportion of total ammonia being present in its toxic form (NH_3). Therefore, depending on pH and temperature, fish can experience toxicity from exposure to external ammonia. The toxicity impact is more severe for starved fish, especially when fish are in stressful and exhaustive exercise conditions (Randall and Tsui et al. 2002).

4.3.2 EPA Approach for Controlling in stream Ammonia Nitrogen

To develop standards for the control of the ammonia toxicity in fish and aquatic life, acute and chronic ammonia toxicity criteria were devised by USEPA (1998, 1999). The acute criteria recommendation given by EPA is called the Criterion Maximum Concentration (CMC) and chronic criterion recommendation is called the Criterion Continuous Concentration (CCC). CMC is derived based on sets of LC50's or EC50's for various aquatic species. An LC50

represents the lethal concentration of a chemical that causes 50% mortality, whereas, an EC50 represents the 50% effect concentration when organisms are killed or effectively dead (USEPA, 2009). On the contrary, CCC is derived based on EC20 (defined same as EC50). Readers may refer to the latest “Update for Aquatic Life Ambient Water Quality Criteria for Ammonia-Fresh Water” (USEPA, 2009) for more details. The following section describes the EPA-prescribed, national criteria for ammonia in fresh water.

4.3.3 National Criteria for Ammonia in Fresh Water

The ammonia toxicity guidelines are based on acute criterion (CMC) of 2.9 or 5.0 mg N/L at pH 8 and temperature 25⁰C for the presence and absence of freshwater mussels, respectively. Likewise, the chronic criteria (CCC) are based on 0.26 or 1.8 mg N/L for the presence and absence of fresh water mussels, respectively, at pH 8 and temperature 25⁰C. For a given temperature and pH, the following conditions should be sufficient to protect freshwater aquatic life until and unless a remarkably delicate species is to be protected at a site (USEPA, 2009):

- (1) While freshwater mussels exist, total ammonia nitrogen concentration (the one-hour average in mg N/L) does not exceed the standard, on an average, more than once every three years.

$$CMC = 0.811 * \left(\frac{0.0489}{1+10^{(7.204-pH)}} + \frac{6.95}{1+10^{(pH-7.204)}} \right) * \min(12.09, 3.539 * 10^{0.036*(25-T)}) \quad (4.3)$$

- (2) While freshwater mussels exist, irrespective of the presence of fish early life stages, the chronic criteria, that is the total ammonia nitrogen (30-day average concentration in mg N/L), does not, on average, exceed the standard more than one time in every three years.

$$CCC = 0.744 * \left(\frac{0.0676}{1 + 10^{(7.688-pH)}} + \frac{2.912}{1 + 10^{(pH-7.688)}} \right) * (0.3443 * 10^{0.028*(25-Max(T,7))}) \quad (4.4)$$

(3) Further, the highest 4-day average during the 30-day period should be well below 2.5 times the CCC.

4.3.4 Ammonia Permitting Approach

Two criteria, toxicity-based criteria and DO-based criteria are important for permitting ammonia nitrogen into fresh water systems. Toxicity-based ammonia limits are determined using the Ammonia Toxicity Protocol and the General Guidance for Writing Water Quality Based Toxicity Permits. These protocol and guidelines are based on CMC or CCC criteria, the selection basis for which is the Stream Dilution Ration (SDR). The SDR is defined as

$$SDR = \frac{Q_w}{7Q_{10} + Q_w} \quad (4.5)$$

where 7Q10 represents seasonal, 7 day, consecutive low flows with a 10 year recurrence interval calculated separately for summer and winter, and Q_w is the facility design flow. If the SDR is less than 1%, the water body is considered stream-dominated and the CMC will be applied to determine the ammonia toxicity limitations. Otherwise, the water body is considered effluent-dominated and CCC will be applied. Ammonia toxicity limitations for summer and winter are determined based on allowable summer and winter stream ammonia nitrogen (Equation 4.3 or 4.4) using the following equation:

$$\begin{aligned} & \text{NH}_3 - \text{N toxicity limit} \\ &= \frac{[(\text{Allowable Instream NH}_3 - \text{N}) * (7Q_{10} + Q_w)] - [(\text{Headwater NH}_3 - \text{N}) * 7Q_{10}]}{Q_w} \quad (4.6) \end{aligned}$$

The next step is to determine the DO-based ammonia limit using the DO model. It is important to make sure that the minimum DO level is maintained in the stream after possible

nitrification for a given release of ammonia nitrogen as the point source. The permit is established based on the lesser of the toxicity-based ammonia and DO-based ammonia limits. Therefore, the development of a DO model which takes waste load allocation into consideration is vital for establishing a point source permit for ammonia nitrogen. The following section briefly describes the theory behind the DO modeling.

4.3.5 Dissolved Oxygen Modeling Concept

Dissolved oxygen concentration is one of the most commonly used indicators of lake, stream, and river health conditions. Aquatic organisms are under stress when DO drops below 4 mg/L. Under continuous hypoxic or anoxic conditions, most of the aquatic organisms perish and, hence, most states want to maintain a daily DO concentration average of 5 mg/l, with no less than 4 mg/l at all times.

DO concentrations in water bodies are manifested depending upon the sources and sinks for DO. The sources of DO in water bodies are: a) atmospheric re-aeration, b) oxygen production due to photosynthesis, and c) DO in incoming tributaries/branches. Similarly, the sinks of DO are: a) oxidation of carbonaceous waste material, b) oxidation of nitrogenous waste material, c) the sediment oxygen demand of water bodies, d) oxygen consumption during respiration by aquatic plants, and e) the chemical oxygen demand (COD). In addition, an incoming tributary can be a source or sink of DO. Figure 4.1 shows a schematic representing major processes governing DO concentration in streams. Mathematical representation of DO concentration in a stream or river can be found in literatures (e.g. Cox 2003).

4.3.6 Dissolved Oxygen Model

DO concentration in a stream decreases due to the addition of waste. A stream also gains DO due to re-aeration as it moves in a downstream direction. However, as more and more waste is added, the decreasing rate of DO concentration is greater than the increasing rate of the DO. When the DO deficit reaches its maximum, it is called the critical DO deficit. Once the DO concentration attains its lowest point, DO starts to increase until it reaches atmospheric equilibrium. This pattern results in the appearance of a curve called DO sag. The sag curve equation which is included in the profile of the DO deficit is given by the Total Streeter-Phelps Model (Lung 2001):

$$D_t = \frac{K_d L_o}{K_a - K_r} (e^{-k_r t} - e^{-k_a t}) + \frac{K_n N_o}{K_a - K_n} (e^{-k_r t} - e^{-k_a t}) + \frac{SOD}{K_a H} (1 - e^{-k_a t}) - \frac{P}{K_a} (1 - e^{-k_a t}) + \frac{R}{K_a} (1 - e^{-k_a t}) + \frac{ZOD}{K_a H} (1 - e^{-k_a t}) + D_o e^{-k_a t} \quad (4.7)$$

where k_d is the effective deoxygenation rate, k_a is the reaeration coefficient, L_o is the BOD, K_r is the overall loss rate, K_n is the overall oxidation rate of nitrogenous BOD, N_o is the initial value of organic nitrogen, P is photosynthesis, R is respiration, ZOD is zebra mussels oxygen demand, H is depth and D_o is the initial oxygen level.

Compared to winter, summer DO is very low due to: (1) the solubility of oxygen decreases significantly with an increase in temperature, and (2) re-aeration decreases due to low streamflow. Besides seasonal variation in DO, DO and ammonia levels in a stream may be affected by climate variability caused by periodic, non-stationary phenomenon such as ENSO. The following section briefly discusses climate variability and its impact on water quality.

4.3.7 Climate Variability and Water Quality

ENSO has been developed as one of the most reliable phenomena for relating inter-annual climate variability in terms of temperature and precipitation on both a local and global scale (Ropelewski and Halpert 1986). ENSO is a coupled, ocean-atmospheric phenomenon that occurs in the equatorial Pacific Ocean and the atmosphere above it and results in varied climatic effects in different parts of the world (Roy 2006). The terms “El Niño” and “La Niña” describe the respective warming and cooling of sea surface temperatures off the shores of the West Coast of South America (QAcetuno 1992). Low frequency climate forcing, such as ENSO, has been found to have strong predictable effects on temperature, precipitation and streamflow (Handler 1990; Handler 1994; Piechota and Dracup 1996; Chiew et al. 1998; Rajagopalan and Lall 1998; Barsugli et al. 1999; McCabe and Dettinger 1999; Kulkarni 2000; Pascual et al. 2000; Hansen et al. 2001; Roy 2006) and water quality (Keener et al. 2007; Marcé et al. 2010) in different parts of the world. Considering the potential link between ENSO and stream water quality and also considering that NOAA can provide reliable ENSO forecasts, this study hypothesizes that ENSO forecasts can be successfully used in NPDES permitting for better protection of aquatic life in streams and rivers.

4.4 Materials and Methods

4.4.1 Study Area

This study was conducted in the Chickasaw Creek Watershed (Figure 4.2) located in Mobile County of South Alabama, near the Mississippi state border in the Mobile River Basin. The watershed, which is 714 km² in size and 45.9 km in length, starts at Citronelle, AL, in the northern part of the state and drains into Mobile Bay. The watershed is dominated by Coastal

Plain geology with an elevation range from a maximum of 43 ft to a minimum of 0 ft at the watershed outlet. The annual precipitation in the watershed is 65 inches, which is relatively higher than North and Central Alabama. The section of the Creek between Eight Mile Creek and the Mobile River receives the highest combined point source loading of total nitrogen, total phosphorus, and BOD (Figure 4.2). There were many point source dischargers, such as, large pulp and paper mills and chemical manufacturing plants, before 1990. Currently, except for the municipal WWTP, discharges from most of them are diverted to the Mobile River. Therefore, currently, there is only one WWTP (Stanley Brooks WWTP- AL0055204) with active NPDES permit discharging into the Chickasaw Creek. The most significant impairment to water quality (hypoxia) in the creek is due to low DO concentration. Historical monitoring efforts suggest a severe threat of low DO concentration downstream of the confluence with Eight Mile Creek. As per ADEM technical report, A segment of the creek from the mouth of stream to US highway 43 is classified as having Agricultural and Industrial (A&I) use. Other parts of the creek are classified as having Fish and Wildlife and Public Water Supply use.

4.4.2 Overall Modeling Approach

In order to best represent the unsteady and dynamic characteristics of the Chickasaw Creek, a one-dimensional hydrodynamic model (EPD-RIV1) (EPA, 2002) was linked with the watershed model Loading Simulation Program C⁺⁺(LSPC) (Shen et al. 2005). I linked the models to develop the nutrient and DO models for the Chickasaw Creek watershed using the best available data and standard modeling practices. The LSPC model was explored to distinguish the sources of nutrients and magnitude from the watersheds in different phases of ENSO (La Niña, El Niño, and neutral). The linked models were run for eleven years to capture five La Niña and

five El Niño years in order to calibrate and validate the model. Seasonal precipitation and temperatures in south Alabama vary with each ENSO phase, especially in winter, spring and summer months (data not shown).

For quantifying the long term impact of climate variability on DO and stream temperature, the model simulations were extended to 55 years, starting from 1950, under a steady state assumption using the calibrated parameters. For ammonia nitrogen permitting, toxicity-based and DO-based criteria should be satisfied. To test against the toxicity-based criteria, stream temperature, pH and streamflow were used and, to test against the DO-based criteria, the only water quality variable DO was used (daily average 5 mg/l and 4 mg/l at all times). Since modeling pH is a relatively complicated process and I was not confident about the modeling capacity of the adopted watershed model for pH, I considered the conventional modeling approach of using average pH (Smith R.L. 2002). Of the above-mentioned stream characteristics, only daily streamflow data was available from a USGS gauging station (02471001) since 1952. Stream temperature and DO were simulated using the calibrated and validated models. For tailoring ammonia nitrogen permit for different ENSO phases, instead of all year average, these parameters were averaged separately for La Niña and El Niño phases for use in Equation 4.6 for different types of fishes and aquatic life. A schematic diagram representing the river system modeling and analysis with ENSO is shown in Figure 4.3.

4.4.3 Watershed Model (LSPC⁺⁺)

Watershed modeling is important for quantifying point source loading from a WWTP and nonpoint source loading for the watershed for DO modeling. The hydrological characteristics that vary in spatial and temporal scale within a watershed should be represented properly. For

this, the LSPC model, which is a version of the widely used Hydrologic Simulation Program-Fortran (HSPF) (Bicknell et al. 2001) model but written in C++, was used. LSPC's algorithms are identical and no different than those of the HSPF model but are more efficient and flexible. The model has been applied for several TMDL developments and is generally considered to be one of the most advanced hydrologic and watershed loading models. The hydrologic portion of the model resembles the Stanford Watershed Model (Crawford and Linsley 1966). The model has a capacity to simulate watershed hydrology, pollutant transport from point and nonpoint sources, stream hydraulics, in-stream water quality, DO, nutrients, and algae. The model has been customized for simulation of other pollutants, such as fecal coliform bacteria, temperature, sediment, and other general water quality, and has been tested at numerous times by TMDL studies of inland and coastal basins (Henry et al., 2002a; Shen et al., 2002a, b).

4.4.4 Watershed Model Configuration and Input Data

Hourly precipitation and other climatological data, such as cloud cover, dew point temperature, solar radiation, wind speed, air temperature, and evapotranspiration, are some of the most sensitive inputs for the model. Climate data are needed on an hourly time scale for the appropriate representation of hydrologic response. The NCDC climate station (Coop ID-015478) of the Mobile regional airport provided these data, and missing climate data were obtained from a nearby climate stations (Coop ID-01583, Coop ID-1084). The meteorological data processing (Metadapt) tool (Tetra Tech, Inc., 2007), was utilized to prepare the climate data in a specific format for the LSPC model. The input data, with their sources and formats, used in this study are summarized in Table 4.1.

For the Chickasaw Creek watershed, the LSPC model was configured to simulate a series of hydrologically-connected sub-watersheds, each of which was characterized with defined geometry, soil and land use characteristics. Each sub-watershed area contributed runoff and nutrient load to the corresponding reach, where the cumulative flow and pollutant loads were routed downstream, eventually contributing as input to the EPD-RIV1 in-stream model. A 10 m resolution, digital elevation model (DEM, 2010) was used in ArcGIS for the watershed delineation and extraction of the stream network. The 2001 land cover data set (NLCD, 2010) and high resolution soil data (SSURGO) (SSURGO, 2010) were utilized to acquire land use and soil-related parameters, respectively. The land use was categorized as low, medium, and industrial urban (13%), deciduous, evergreen and mixed forest (47.4%), woody and herbaceous wetland (18.6%), range shrubland, grassland herbaceous and hay (19.4%), and the remaining 1.6 % as water, south western range and agricultural land. The watershed soil is characterized predominantly by hydrological soil groups A, B and D.

The LSPC model utilizes different sets of hydrologic parameters for surface and subsurface hydrologic analyses in different sub watersheds, depending on the soil type and land use categories. Long term, streamflow data, recorded since 1952, were available at the USGS gage (02471001), which drains 357 km² of the watershed. In-stream water quality data were collected by the Alabama Department of Environmental Management (ADEM) for stream temperature, nutrient, chlorophyll a, BOD, ammonia nitrogen and DO simulation. ADEM-monitored water quality data at station CS₁ (Latitude = 30.78224, Longitude = -88.07248) (Figure 4.2) were available since early 1980. The information about the location of the treatment plant, land use data, soil data, USGS gage and the water quality sampling stations are reported in Table 4.1.

4.4.5 Streamflow Simulation

Streamflow simulations were carried out for 16 years using USGS-gage observed data from 1990 to 2005 (Figure 4.4). Streamflow simulations were started from 1/1/1985 which corresponded to a 5-year warm up period. The reason for using a long spin-up period is to minimize the effect of unknown initial conditions, such as antecedent moisture, initial ground water table height, etc., and stabilize the hydrologic component of the model. The model was calibrated for the period of 1990 to 1996 and validated for the period of 1997 to 2005. Streamflow was calibrated at daily as well as monthly time scales. Since model calibration is an iterative procedure in which simulated and observed data are compared to produce the best agreement between these two datasets throughout the calibration period, for calibration at specified locations in the watershed, the model parameters were adjusted within the physically possible ranges until the resulting predictions best fit the observed data.

4.4.6 Stream Temperature Simulation

Since water temperature is an important parameter for simulating biochemical transformation and DO, I calibrated water temperature after the hydrologic calibration. The period for temperature calibration and validation was from 1997 to 2003 and 1990 to 1996, respectively. The selection of this period was based on the availability of the data and its correspondence with the hydrologic calibration and validation. Stream temperature was affected by three flow paths (land surface, interflow and groundwater) of water from the upland areas, as well as, stream heat budget interactions with the atmosphere. The land surface layer temperature was estimated using a regression equation as a function of air temperature. The interflow and groundwater temperatures were estimated using the mean difference from air temperature and a

smoothing factor. The model-simulated stream temperatures were compared with the observed stream temperatures.

4.4.7 Simulation of Pollutant and Source Assessment

Sources of pollutants contributed by a watershed can be broadly categorized into point and nonpoint. There is one major point source, Stanley Brooks WWTP, in this particular study (Figure 4.2). The point source was taken into account in the LSPC model by using time series inputs for flow and concentrations. ADEM carried out the comprehensive assessment of the point source discharge at different times from the Stanley Brooks WWTP. Since the point source discharge data were not available for the entire modeling time period, the average monthly values from the available data were given as an input to the LSPC model.

Nonpoint sources of pollutants are diffused and are from diverse locations. Generally, they involve a buildup of pollutants on land surfaces that wash-off during rain events from agricultural, pasture and urban land. Daily atmospheric data, such as ammonia and nitrate, were taken from National Atmospheric Deposition Program (NADP). These data were available for a short period (year 2010) from the Alabama-Mississippi border monitoring location (MS12) in the vicinity of the watershed. I compared these data with the data recorded at the Black Belt Research & Extension Center (AL10) located in Dallas County, Central Alabama, which has furnished long term data sets (18 years) for the model calibration period. The difference in the observed data between these two stations were nominal, and hence I developed a regression equation based on the data monitored at station AL10 and available data at MS12 to transfer the atmospheric deposition load from station AL10 to the study area. Model calibration parameters

adopted in the Mobile Bay LSPC model and the Flint River watershed model (Tetra tech., 2010) were used as starting points for simulation of pollutants and source assessment.

4.4.8 Nutrients, Chlorophyll a, BOD5 and DO Simulation in the Watershed Model

Since numerous physical and chemical processes affect the interplay between the nutrients, phytoplankton, and carbonaceous material and affect the DO level in a stream, I simulated a number of water quality variables. Figure 4.1 shows the kinetics of nutrient cycling and interactions with dissolved oxygen. The LSPC model was configured to simulate all of the mechanisms pertaining to the nutrient cycle. BOD5 was simulated in LSPC using the BOD5 decay rate, BOD5 release from sediment, and benthic source of BOD5 due to scouring of the sediment. The simulation of BOD5 was not as good as DO (not shown) and did not adequately simulate low and high BOD, which indicates to the fact that a watershed model, such as LSPC, may at times not accurately represent the physical, chemical, and biological processes for BOD simulation. The possibility of analytical uncertainty associated with the observed BOD5 data cannot be ignored. However, since BOD5 did not have an association with the observed DO (Figure 4.5) in the study watersheds and did not show a consistent trend with the streamflow and temperature (not shown), I anticipate DO simulations will not be affected by BOD5 simulation in this particular watershed. Since nitrification of ammonium can be a sink for DO through bacterial transformation, ammonium was simulated for DO calibration. Moreover, the model was calibrated for all constituents: ammonia, total nitrogen, total phosphorus, BOD5, Chlorophyll-a, and DO from both qualitative and quantitative perspective.

4.4.9 Hydrodynamic and In-stream Water Quality Model

EPD-RIV1 is a “cross-sectionally averaged”, one-dimensional, hydrodynamic as well as water quality model (EPA 2002; USEPA 2004) for rivers and estuaries. The Georgia Environmental Protection Division (EPD) developed the model EPD-RIV1 from CE-Qual-Riv1 model (Bedford et al. 1982) for the Chattahoochee River Modeling Project. This model, originally developed in 1982 (Dortch 1990; Martin et al. 2002) is a continuation of CE-Qual-Riv1 model (USACE 1995) by the U.S. Army Corps of Engineering Waterways Experiment Station and is used for hydraulic and unsteady flow simulation (Herb and Stefan 2010). Although the model was originally intended to analyze waste load allocation, including provisions of Total Maximum Daily Loads (TMDLs) under dynamic and unsteady river flow conditions, it can be used on riverine systems with steady flows as well. In addition, the model can simulate complex rivers and streams, especially when they encounter dams, reservoirs, significant lateral flows and tidal influences, provided that the stratification is limited to one direction.

The model consists of two components: a hydrodynamic component (EPD-RIV1H), which estimates streamflow, channel depths, flow velocities, water surface elevations and other hydraulic characteristics which are then used to solve the St. Venant equations, the governing equation which utilizes a four point, implicit finite difference numerical scheme (Martin et al. 2002); and a water quality component (EPD-RIV1Q), which can simulate sixteen state water quality variables (temperature, carbonaceous biochemical oxygen demand (CBOD1), CBOD2, nitrogenous biochemical oxygen demand, DO, organic nitrogen, ammonia nitrogen, nitrate + nitrite nitrogen, organic phosphorus, phosphates, algae, dissolved iron, dissolved manganese, coliform bacteria, arbitrary constituent 1, arbitrary constituent 2) (Martin et al. 2002). For modeling purposes, the model assumes that the water body is one-dimensional (longitudinal)

with uniform velocity over the cross-section, and well-mixed laterally and vertically. The model simulates conventional pollutants using comprehensive water quality algorithms and is fully capable of simulating the impacts of macrophytes on DO and nutrient transformation. However, some of the processes, for example, sediment transport processes (scour and deposition), and the associated effect on water quality are not simulated by the model even though it simulates sediment oxygen demand and nutrient release.

4.4.10 Hydrodynamic Model Configuration

Chickasaw Creek and its tributaries downstream from the USGS gage (Figure 4.2) are represented using ten different cross sections based on the geometric properties of the stream network, computational requirements, and distribution of the point and nonpoint sources. The hydraulic model consists of one major branch, Eight Mile Creek. The data for the hydrodynamic model development consists of cross-sectional information for river segments at different locations, downstream boundary conditions, point sources, and observed flows at USGS gauging stations for model calibration. The bathymetry of the stream was obtained from the US Army Corps of Engineers at the downstream end, near to the confluence with Mobile Bay. In addition, the cross sectional information was collected from the USGS and Alabama Department of Transportation's (ALDOT) Bridge Department. The river cross-section was further verified using the LIDAR data obtained from the City of Mobile, AL and the FEMA HEC-2 flood-forecast studies. The Manning's roughness of 0.035 was used for the channel and was further validated by the report prepared by FEMA flood forecast studies for the Chickasaw Creek. The upstream boundary conditions were achieved from the USGS gage data. Hourly streamflow were derived from the USGS gauging station after LSPC model calibration and validation, and were

used as an upstream boundary condition. Point source discharge from the Stanley Brook WWTP was used as a boundary condition.

In addition, the hydrodynamic model requires downstream boundary conditions and initial conditions. Since the observed downstream water level data were not available, I utilized the unsteady state simulation module in HEC-RAS to derive downstream boundary conditions. The steady state flow simulation module in HEC-RAS was utilized to derive the initial flow depth at different river cross-sections (initial condition) for both the hydrodynamic and water quality models. The approach of deriving downstream boundary conditions using HEC-RAS has been successfully applied in various other projects (Herb and Stefan 2008; ADEM 2006). Many HEC-RAS simulations were run to determine the downstream rating table for the hydrodynamic as well as water quality models using different flow rates. Additional cross-sections in the stream were generated using the HEC-RAS interpolation function. HEC-RAS uses the same one-dimensional unsteady flow St.Venant equations to estimate water surface elevations for a given discharge (Te Chow 1959) as those used in the hydrodynamic modeling of the EPD-RIV1 model.

4.4.11 In stream Water Quality Model Configuration and Calibration

4.4.11.1 Input Data for EPD-RIV1 Model

The hydrodynamic linkage file prepared by the EPD-RIV1 model was transferred to the water quality component while performing water quality simulations. The EPD RIV1 model requires several site-specific, climatic parameters for water quality simulations in a specific format. The same meteorological station (Coop ID-015478) located at the Mobile regional airport was utilized. The time series data needed for the model calibration and validation

includes precipitation, cloud cover, wind speed, solar radiation, wet bulb temperature, and dry bulb temperature. Climate input data were prepared in a model-required format using the Metadapt tool.

I employed a direct mapping scheme to link the hydrodynamic and water quality models, resulting in eighteen segments (Nineteen cross-sections) with the same geometry applied in the hydrodynamic model. The major information needed for water quality modeling was the initial conditions, boundary conditions, external loadings and internal sources and sinks. In addition, the water quality model uses loadings from tributaries, point sources, etc. The locations of the boundary conditions for water quality modeling were the same as those in the hydrodynamic model. The constant initial conditions at each cross section were provided for the nine state variables (temperature, CBODu1, Org-N, NH₃-N, NO₃-N, Org-P, Ortho-P, DO, CBODu2). The model run was extended for an additional period of six months to achieve steady state for both the hydrodynamic and the water quality models. The spin up period chosen for the hydrodynamic model was very short because it was found from a past series of sensitivity analyses that the impact of initial conditions quickly dampens, suggesting that the model output is independent from the initial arbitrary value chosen; this ensures model stability (Zou et al. 2006). The boundary conditions for the major point source were specified based on the grab samples furnished by ADEM and further verified by the Discharge Monitoring Report (DMR) issued by ADEM. The flow as point source discharge was also specified as a boundary condition for the hydrodynamic modeling although the flows from these point source dischargers did not significantly affect downstream hydrodynamics.

Once I configured the model, I explored the governing mechanisms primarily responsible for DO variations in the Chickasaw Creek watershed. Since long term, observed datasets were

available, I simply plotted each entire, observed variable such as total nitrogen, total Kjeldahl nitrogen, total phosphorus, chlorophyll-a, BOD5, ammonia nitrogen, stream temperature and streamflow against DO (Figure 4.5). The scatter plot shows that streamflow and stream temperature are the most important variables affecting DO. The model outputs were compared to the observed data for the various parameters. The main parameters subjected to calibration were stream temperature, BOD, ammonia nitrogen, and DO. The model was calibrated in a stepwise manner to adjust the model parameters within a reasonable range to adequately reproduce the observed data.

4.5 Result and Discussion

4.5.1 Model Performance

Model performance was assessed using a number of non-dimensional measures, such as the Nash-Sutcliffe coefficient of Efficiency (*NSE*) (Nash and Sutcliffe 1970), Mass Balance Error (*MBE*), and the coefficient of determination (R^2), because there is not a single best statistical measure to check the performance of a model's outputs against observed data. Details on the statistical parameters measuring performance of a model can be found in several pieces of literature (Moriassi et al. 2007). Interested readers can also refer to the articles by Donigian, Jr et al. (1984) and Lumb et al. (1994) for various methods of assessing model adequacy.

Daily, monthly, seasonal, and total modeled flows (not shown) were compared with observed data to measure model performance. The calibrated model parameters were applied to an independent time period (1997 to 2005) for model validation to make ensure that the calibrated parameters can be applied in a wide range of conditions. Model validation was satisfactory, demonstrating that hydrological parameters were able to capture the system

dynamics (Figure 4.4). The statistical parameters, such as, NSE , R^2 and MBE are listed in the Table 4.2.

Similarly, the simulated water temperatures closely resembled the observed temperatures and captured the seasonal variations (Figure 4.6). R^2 for the stream temperature was 0.73, indicating a reasonable model performance.

In addition to the statistical measures, the LSPC model performance for water quality calibration was judged through visual inspection using the best professional judgment of model fitting to the observed data. Figure 4.6 shows the comparison between the observed and the simulated water quality variables at a monitoring station. This figure indicates that the model was representing the seasonal variation in water quality of the system well. Similarly, the statistical parameters measuring the performance of the model were found to be satisfactory (Table 4.2). The performance statistics indicate that the model satisfactorily predicted water quality constituents for the calibration period. Figure 4.7 shows the BOD, stream temperature and DO simulation using the LSPC linked EPD-RIV1 model. In general, it was found that the model captured the temporal distribution of water quality constituents satisfactorily.

4.5.2 Effect of ENSO on Streamflow, Temperature and DO

Using the calibrated and validated models, I quantified the impact of ENSO on DO and stream temperature. The classification of ENSO phase was based on the Niño 3.4 index which is calculated based on the 3 month running average of ERSST.v3b sea surface temperature (SST) anomalies in the Niño 3.4 region (5°N-5°S, 120°-170°W) (Trenberth and Stepaniak 2001). The watershed experienced several El Niño (1991-1992, 1994-1995, and 1997-1998, 2004-2005) and La Niña (1995-1996, 1999-2000, and 2000-2001) years during the model calibration and

validation period. The association of nutrient load with ENSO phase varied from season to season. El Niño years contributed significant winter and spring streamflow in years 1992, 1995 and 1998 (Figure 4.4). These years corresponded to a higher nutrient load (total nitrogen, total phosphorus) and a relatively higher rate of winter and spring DO (Figure 4.6). Conversely, La Niña years, especially in 1999 and 2000, contributed less streamflow in winter seasons. Therefore, these years corresponded to less nutrient load, higher stream temperature, and less DO in winter and spring seasons. The impact of persistent drought caused by La Niña from late 1998 to August 2000 resulted in low streamflow ($0.51 \text{ m}^3/\text{s}$), high stream temperature ($27 \text{ }^\circ\text{C}$) and low DO (2.5 mg/l) at the USGS gage. This substantially low flow was experienced by a few nearby watersheds of South Alabama (Perdido and Fish River watersheds) as well (not shown). When a stream encounters very low and calm flows, reaeration decreases and temperature increases, resulting in low DO levels. The point source contribution during this period was relatively higher, suggesting that it has more influence during La Niña (dry) years compared to El Niño (wet) years.

The impact of nonpoint sources on DO was also evaluated under La Niña and El Niño conditions separately. The nonpoint source has more effect on DO variation in the El Niño period than in the La Niña period (not shown).

I also analyzed the correlation between streamflow, stream temperature, and DO with ENSO using 55 years of ENSO information and model-simulated data. Because a stream's DO level is a function of streamflow and temperature, ENSO shows strong correlation with all these variables. The analysis was focused in two seasons, i.e., December to April and August-October, because ENSO greatly affects precipitation in these seasons in the Southeast (Keener et al. 2007). Figures 4.8 and 4.9 show the variations in spatially averaged streamflow, stream

temperature and DO in different ENSO phases in different seasons. The Kendall tau (Kendall 1938) and Pearson correlation (Rodgers and Nicewander 1988) for streamflow, stream temperature and DO with ENSO in two seasons (Dec to April and Aug to October) is reported in Table 4.3, suggesting that there is a significant correlation with ENSO phase. Similarly, differences in El Niño and La Niña were evaluated at a significance level of p-value <0.05. Figure 4.8 depicts that significant differences exist in flow, stream temperature and DO in different ENSO phases during December-April. (This period is considered as winter in the conventional approach of NPDES permitting). Our results strongly suggest that DO, streamflow and temperature are directly linked to non-stationary climate modes, ENSO.

The ENSO signature in summer (May-July) becomes distinct only during strong ENSO events that perpetuate for a number of years. The DO variation (p-value < 0.05) during May to July in different ENSO phases for the strong ENSO events (i.e. La Niña or El Niño is experienced throughout a year) is given in Figure 4.9. Similarly, the streamflow and DO variations in the August-October season (this season is considered to be the summer season for conventional permitting purposes) in different ENSO phases is demonstrated in Figure 4.9. The box plot shows a relatively higher degree of variability in August-October, indicating a slight overlap of the inter quartile range in each box plot of streamflow and DO, respectively.

4.5.3 ENSO and Ammonia Permit

This above analysis clearly indicated that in two seasons, winter (December to April) and summer (August to October), variations in the streamflow, DO and stream temperature can be attributed to ENSO. Therefore, I wanted to evaluate the differences in ammonia permit in two different climatic conditions (El Niño and La Niña) in these two seasons. As I discussed in the

literature section, I relied on a USEPA-prescribed equation (equation 4.6) to determine the allowable ammonia nitrogen for different seasons, which is based on 7Q10 low flows. Since the seasonal 7Q10 have been adopted by different state agencies such as the South Carolina Department of Environment and Natural Resources, the Alabama Department of Environmental Management (ADEM), etc., I explored the climatic conditions that can accommodate higher assimilation, and also the climatic conditions that demand stricter regulation using seasonal 7Q10. For this, I estimated 7Q10 in both seasons of December-April and August-October. This approach of seasonal analysis is practicable and consistent with the USEPA's TMDL policy (Conrads et al. 2003). The estimated 7Q10 (1.90 m³/s) using log-Pearson III distribution and 7 days consecutive annual low streamflow corresponding to different El Niño phases (marked with a triangle) are shown in the Figure 4.10. The El Niño streamflow always exhibited strong association with higher streamflow, and even the lowest streamflow encountered in El Niño periods over 55 years of historical records (2.53 m³/s) were substantially higher than the adopted 7Q10. This was further confirmed in two additional watersheds: a) the Fish River watershed and b) the Perdido watershed (Figure 4.10). 7Q10 calculated separately using El Niño years for Chickasaw Creek watershed is 2.58 m³/s. This allows 28% more permissible discharge in the Chickasaw Creek watershed in winter season (Table 4.4) after satisfying the minimum DO requirement of 5 mg/l. This result was derived using average pH in different ENSO phases. However, the correlation of pH with ENSO phases during Jan-Feb-March, using Kendall tau (0.14) and Pearson correlation (0.10) test for the observed data since 1980, depicted that climate variability may cause slight variations in pH. The monthly variation of pH has been documented in several past studies (Araoye 2009). Incorporating climate-induced variability in pH into point source permitting can be a part of future research. Table 4.4 indicates that, for a particular type of

fish, the La Niña condition closely resembles the adopted 7Q10, suggesting that La Niña represents the lowest flow condition, which is further explained by Figures 4.9 and 4.11. Besides interseasonal variation, the significant variation in stream characteristics within the season (intra-seasonal variation) in different ENSO phases was observed. I also clearly observed the possibility of releasing more pollutants in the La Niña phase of the ASO season (associated with higher streamflow, higher DO but with a great deal of variability) because the continuation of La Niña in the successive season (winter La Niña, i.e., critical condition) will require stricter regulations. Figure 4.12 illustrates the possibility of storing the pollutant in a previous season and releasing it in the following season, which is especially true when we encounter an El Niño period in the Aug-Sept-Oct season and NOAA predicts the continuation of El Niño phase for the consecutive season. This provides an opportunity to reduce the pollutant load in El Niño, August-October (critical condition), and transfer the pollutant to the successive El Niño season (high assimilative period). This approach of inter-seasonal pollutant transfer to the next season, utilizing the prior knowledge of ENSO forecasts and without compromising the minimum water quality threshold, is particularly suitable for the flow-based treatment plant or the treatment plant operating under a hydrograph controlled release approach. This approach is still useful for small, community-based, waste water treatment systems and sometimes eliminates the need for the further treatment. The system involves using devices for the measurement of the water quality threshold, velocity of streamflow, etc. to discharge the pollutant based on the instantaneous flows. Therefore, when the impending drought is extended for a number of years, the system will need to continuously hold the pollutant and this can be managed properly using ENSO information. The potential link of ENSO forecasts to variations in streamflow, stream

temperature, and DO provides an opportunity to release the pollutant based on the assimilative capacity of the stream.

4.5.4 ENSO for the Identification of Critical Conditions

Identification of the critical condition for NPDES permitting and TMDL development is an important but very challenging issue. For this, I divided the summer into two seasons: May-July and August-October to identify the critical conditions. I recommend different 7Q10 for two seasons as more assimilation can be achieved in May-July ($7Q_{10} = 1.1 \text{ m}^3/\text{s}$), and more strict criteria should be adopted from August-October ($7Q_{10} = 0.88 \text{ m}^3/\text{s}$). This is consistent with the past TMDL studies in Alabama suggesting that August-October is a more critical period. ENSO signature demonstrates two other critical conditions in different seasons. The La Niña period in May-July is characterized by substantially less DO than the El Niño period in this season (Figure 4.9). This period would be a critical situation for both toxicity and DO limits if the La Niña perpetuates for a number of years. Similarly, El Niño in August-October is characterized with less DO. This hypoxic condition in the stream will be further detrimental to the aquatic life (critical condition) as the stream experiences increased temperature and decreased streamflow simultaneously.

August-October is a period when ENSO demonstrates a relatively better response than May-July. This is consistent with previous research (e.g. Keener et al. 2007). Figure 4.9 illustrates that La Niña in the August-October season tends to have higher streamflow but with a high degree of variability and produces the lowest streamflow during this season as well. This is particularly attributed to summer thunderstorms. Therefore, releasing more pollutant in this period is risky. The lowest streamflow for La Niña in August-October was encountered in 1954.

I further explored this occurrence and discovered the following. The watershed concurrently experienced extremely low precipitation in the winter and spring months of 1954. The lowest streamflow in August 1954 were due to the influence of the low precipitation encountered in the spring and winter months. For this, I introduced a new index (the ratio of precipitation from January-July to annual precipitation) and developed a chart (Figure 4.11), which plots the average of 7 days of consecutive low flows for a La Niña year with this index. This suggests that the low streamflow in the August-October season is weakly correlated ($R^2 = 0.39$) with the precipitation characteristics of the preceding season. The lower streamflow can be expected if the precipitation index is less than 0.5 (Figure 4.11). This indicates that climatic conditions of the immediate preceding season should be thoroughly interpreted before releasing more pollutant in the current season. The tendency of extremely low flow to be auto correlated with the previous season's precipitation/streamflow was studied comprehensively using autocorrelation and cross correlation functions and is described in the following section.

4.5.5 ENSO Signal for Critical Conditions

When streamflow become extremely low (less than anticipated 7Q10) due to extreme meteorological drought conditions, the streamflow become primarily a function of the previous season/month's underground storage. This is further confirmed using the autocorrelation graph of base flow, which closely resembles the low flows condition in the stream, demonstrating an autocorrelation even after three months (one season) (Figure 4.13). This autocorrelation is true for streamflow as well (Figure 4.13). Further evaluation of the cross correlation function between low streamflow and the preceding ENSO characteristics show that streamflow in the August-October season manifest cross correlation with the SST anomalies in Niño 3.4 region (Niño 3.4

index) in the winter and spring seasons (Figure 4.13) with 0.17 for two season lag and 0.13 for one season lag. Hence, it can be concluded that the ENSO characteristics provide sufficient clues to understanding the streamflow characteristics in August-October. This is further explained by Figure 4.14 which demonstrates that substantially low streamflow (lesser than adopted 7Q10) were encountered in August 2000 due to the continuation of La Niña since 1998. The reason for extreme low flows in 2007 can also be described by the ENSO characteristic that was experienced over that period. There are primarily two reasons: 1) the watershed experienced El Niño in 2006 in this season, characterized by lowest streamflow and this is consistent with previous research which indicates that seasonal streamflow are a function of ENSO characteristics of the previous year (Gim�iz-Fortis et al. 2010) and also the function of previous streamflow in that season, and 2) winter drought experienced in 2007. This autocorrelation characteristic of streamflow led to extremely low flows in 2007 (Aug-Oct) in Chickasaw and other creeks (Figure 4.14). Hence, ENSO information gives sufficient warning for the impending drought condition. When the drought perpetuates for a long time, the streamflow will be considerably less than the adopted 7Q10, and may require further reduction in the point source discharges. Extremely low flows will be a function of winter and spring streamflow, SST and precipitation.

4.6 Summary and Conclusion

In this paper, I investigated the connection between climate variability and point source permitting for water quality protection of the Chickasaw Creek watershed of South Alabama. The study was carried out using a non-stationary climate mode, ENSO, and its impact on simulated DO, stream temperature and streamflow. The DO and stream temperature were

simulated using the LSPC and hydrodynamic and water quality model, EPD-RIV1. I further analyzed the long term, observed streamflow with different ENSO phases. Various non-parametric tests and statistical analyses were performed to detect correlation between ENSO and the simulated stream temperature, DO and observed streamflow over a period of 55 years. Analysis suggested that stream temperature, DO and streamflow are correlated with ENSO.

The conventional method of point source permitting neither fully utilizes the assimilative capacity of the stream nor captures extreme drought. Hence, the specific objective of this research was to demonstrate how the short term climate information can be used for improved point source discharge permitting. Ammonia nitrogen permitting from a waste water treatment plant was investigated in this case study. Because the unionized form of ammonia nitrogen becomes toxic at a certain temperature and pH, the toxicity limit and DO limit in the stream was used to regulate its discharge. Criterion Maximum Concentration (CMC) and Criterion Continuous Concentration (CCC) suggested by the EPA were used to set the severe and acute criteria involving fresh water muscles. Three inter seasonal dry periods, such as La Niña winter (December-April), La Niña summer (May-July) and El Niño fall (August-October), were identified as periods of critical conditions, and wet periods, such as El Niño winter (December-April) and La Niña fall (August-October) were identified as periods of high streamflow and assimilation. However, the La Niña fall exhibited a great deal of variability. It was found that the El Niño winter can assimilate 28% more ammonia nitrogen than what is allowed using conventional seasonal 7Q10 in this season. This period can be utilized to assimilate the preceding season's (El Niño in August-October) waste, because El Niño in August-October generally represents the critical condition, provided that the pollutant could be stored temporarily.

Most often, drought perpetuates for a long time before the severe hydrological drought is encountered. The summer low flow was found to be autocorrelated with spring and winter streamflow, and cross correlated with spring and winter SST anomalies. Therefore, ENSO provides sufficient warning for an impending drought (critical condition). In addition, most of the total maximum daily loads (TMDLs) are developed to satisfy applicable water quality standards for critical conditions. Identification of critical conditions in the water body caused by climate variability is a major step in capturing the worst case scenario, leading to the protection of aquatic life, and maintenance of the designated use. ENSO can be a useful tool for TMDL allocations and NPDES permitting in the future.

The potential link between interannual climate variability caused by ENSO can be utilized in NPDES permitting, especially when impending droughts due to ENSO can be projected a few months in advance. This can avoid the uncertainty associated with low flow estimation due to limited data. The research demonstrates the potential application of climate information in NPDES permitting for the improvement of water quality and protection of aquatic life.

4.7 Acknowledgement

The authors would like to truly acknowledge Mr. Lynn Sisk and Mr. Reynolds Charlie of the ADEM water quality branch division for providing necessary data and information. The authors would also like to express their sincere thanks to Ms. Jamie Childers and Mr. Brian Watson of Tetra Tech, Inc., for their technical suggestions and proper guidance in model-development.

Table 4.1. Data used for the study with their source and information

Data	Source	Additional information
Land use, NLCD 2001	www.AlabamaView.org	
Soil	Soildatamart.nrcs.usda.gov	SSURGO soil data base
DEM	www.seamless.usgs.gov , www.datagateway.nrcs.usda.gov	10-m resolution
Weather gage station	NOAA, National Climatic Data Center , http://ncdc.noaa.gov	Coop ID-015478
Streamflow	www.al.water.usgs.gov	(USGS gage 02471001)
Hydrologic (stream network) data	www.aces.edu/waterquality/gisdata	
GIS layers/tiger files	www.AlabamaView.org	
Water quality data	Alabama Department of Environmental Management (ADEM)	Station (30.7822, -88.0725)
Point source data	Alabama Department of Environmental Management (ADEM)	WWTP (30.78,-88.099)

Table 4.2. Statistical parameters measuring the performance of the watershed model

	Calibration			Validation		
	NSE	R ²	MBE	NSE	R ²	MBE
Streamflow (daily)	0.31	0.49	7.3%	0.49	0.32	2.1%
Streamflow (monthly)	0.64	0.69	7.4%	0.84	0.84	2.1%
TN (monthly)	0.81	0.85	17.9%	0.65	0.67	5.5%
TP (monthly)	0.8	0.81	-10.6%	0.56	0.67	-16.0%

Note: NSE, MBE and R² imply Nash-Sutcliffe Efficiency, Mass Balance Error and Coefficient of Determination.

Table 4.3. Correlations of observed streamflow, simulated stream temperature and DO with Niño 3.4 indexes since 1950.

Statistical test	December-April			August-October			May-Nov
	Streamflow	Temperature	DO	Streamflow	Temperature	DO	DO
Kendall tau (τ)	0.21 ($\alpha = 0.06$)	-0.12 ($\alpha = 0.3$)	0.33 ($\alpha = 0.000$)	-0.16 ($\alpha = 0.09$)	-0.18 ($\alpha = 0.05$)	-0.12 ($\alpha = 0.2$)	0.26 ($\alpha = 0.05$)
Pearson Correlation	0.36 ($\alpha = 0.03$)	-0.22 ($\alpha = 0.18$)	0.53 ($\alpha = 0.000$)	-0.3 ($\alpha = 0.02$)	-0.25 ($\alpha = 0.07$)	-0.17 ($\alpha = 0.2$)	0.47 ($\alpha = 0.00$)

Note: α represents the p-value

Table 4.4. Ammonia limit in different ENSO phase

ENSO phase	Winter season			% difference in Toxicity limit
	Temperature, 0 C	% Difference in Allowable NH3-N	7Q10 streamflow	
La Niña	14.2	3.00%	1.93	2.1%
El Niño	12.9	4.70%	2.58	28.0%
All phase	13.7		1.91	

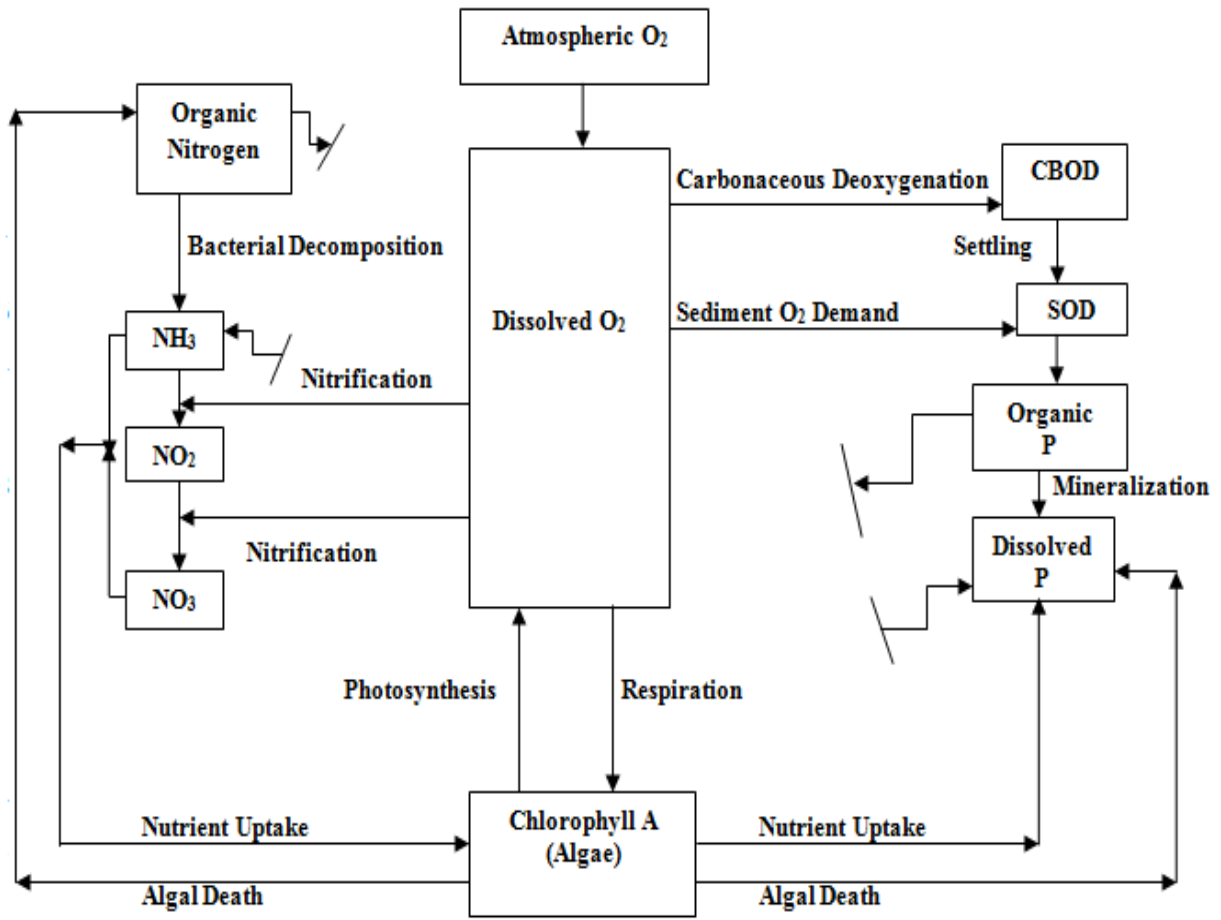


Figure 4.1. Schematic showing major processes influencing DO in streams and rivers (EPA, 1997).

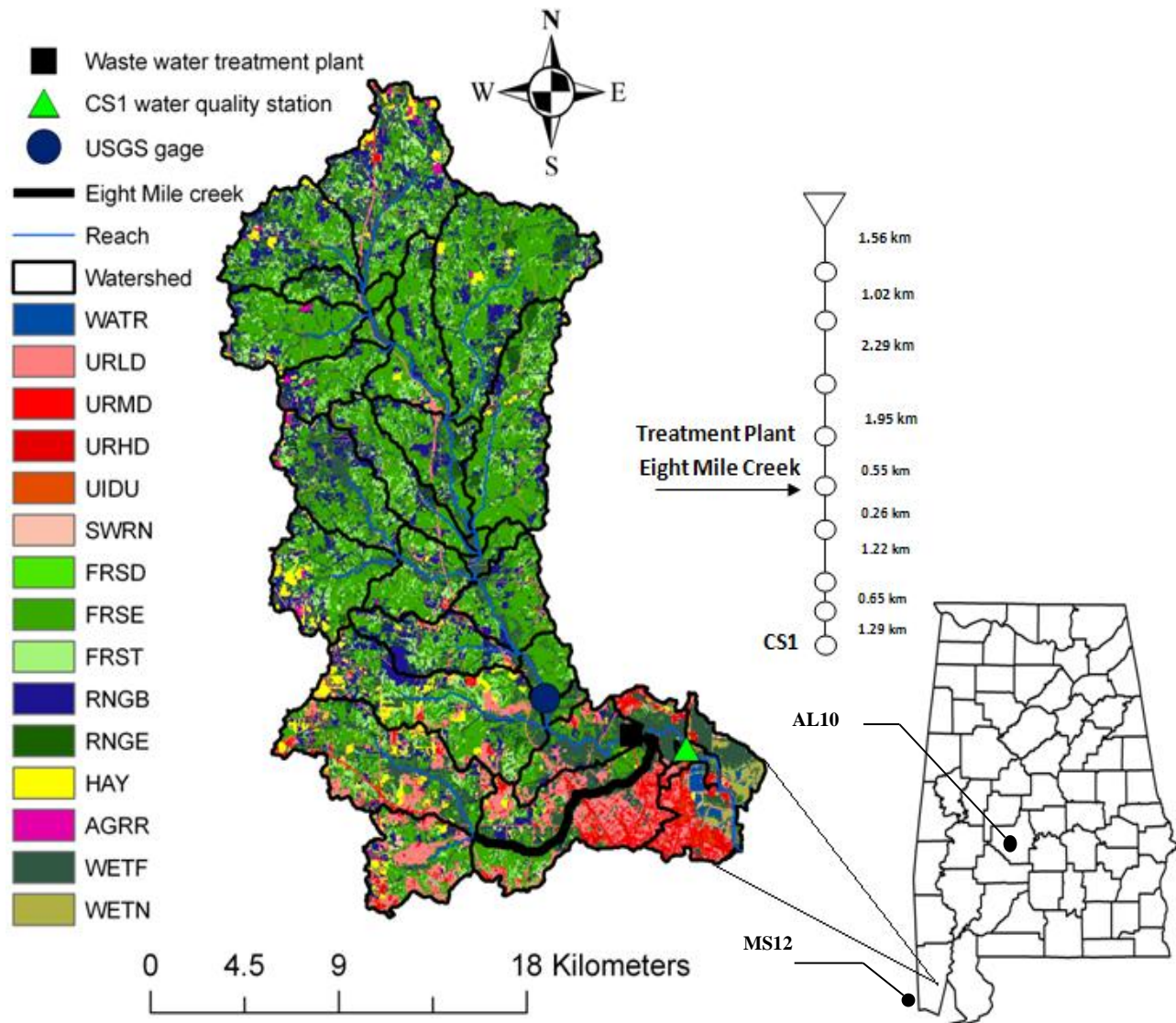


Figure 4. 2. Chickasaw Creek watershed in Mobile County of South Alabama with USGS gage, waste water treatment plant, water quality station CS1 (Latitude = 30.78224, Longitude = -88.07248) and cross section of the river.

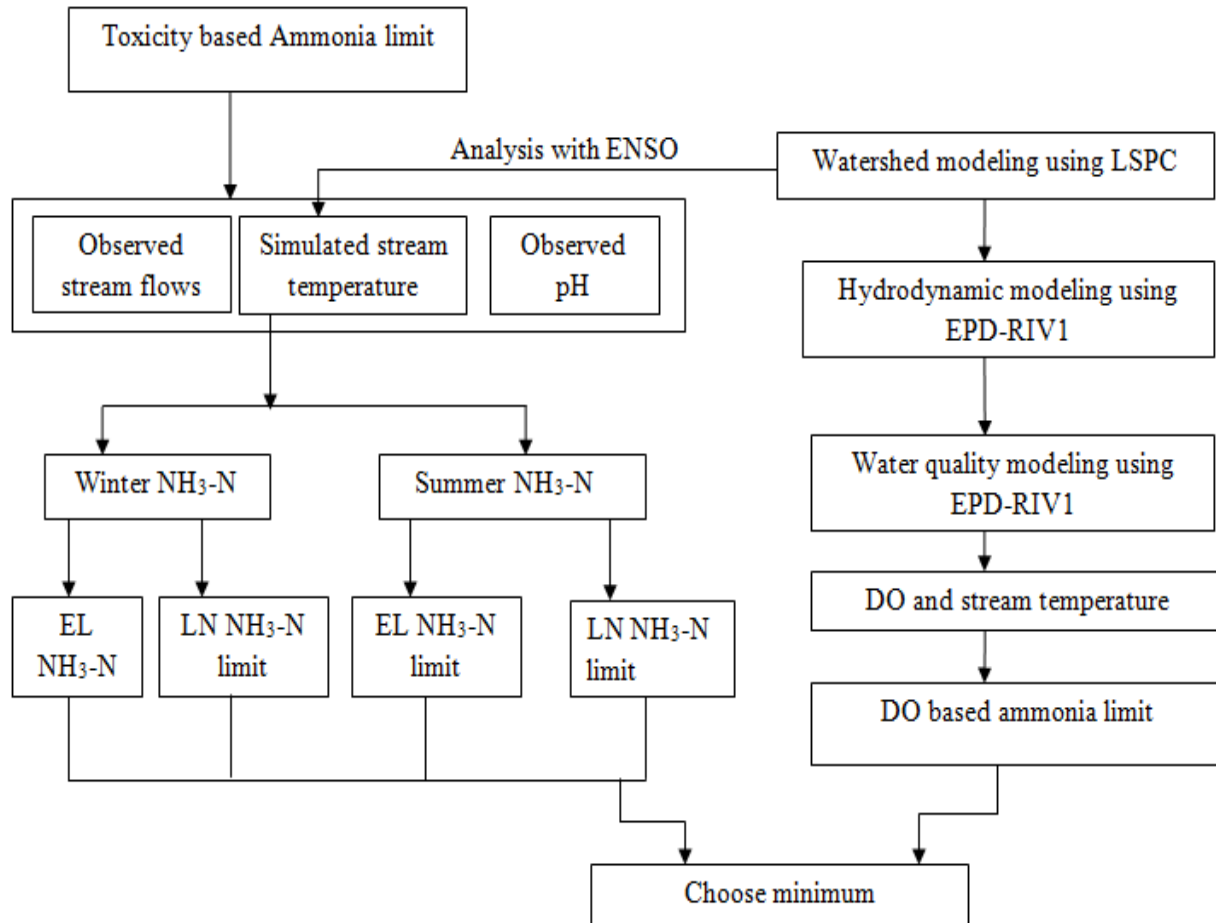


Figure 4.3. Schematic diagram representing the river system modeling and analysis with ENSO for ammonia permit. EL and LN imply El Niño and La Niña, respectively.

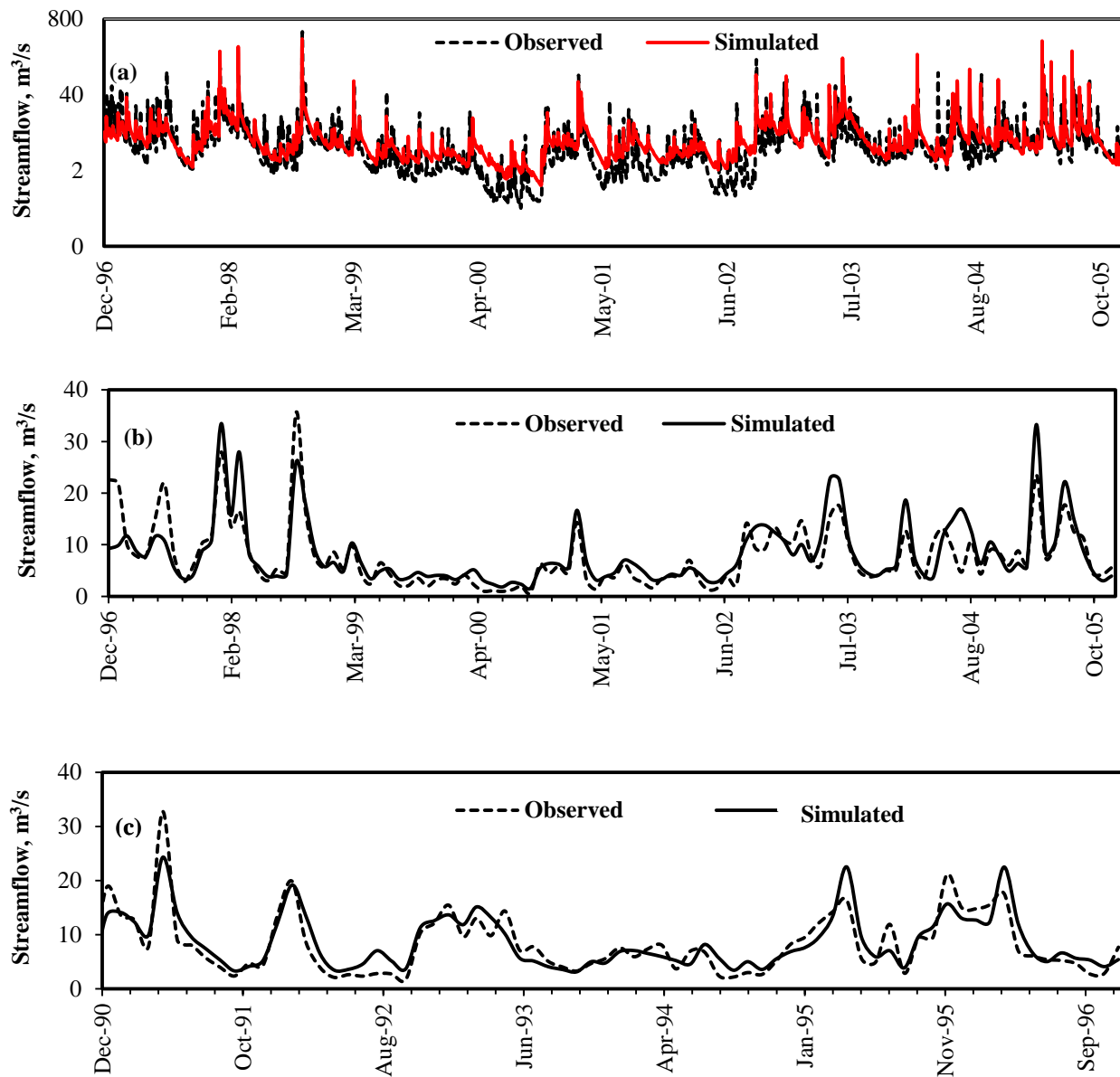


Figure 4.4. Observed and simulated a) daily calibrated streamflow, b) monthly calibrated streamflow, and c) monthly validated streamflow. EL and LN represents El Niño and La Niña respectively.

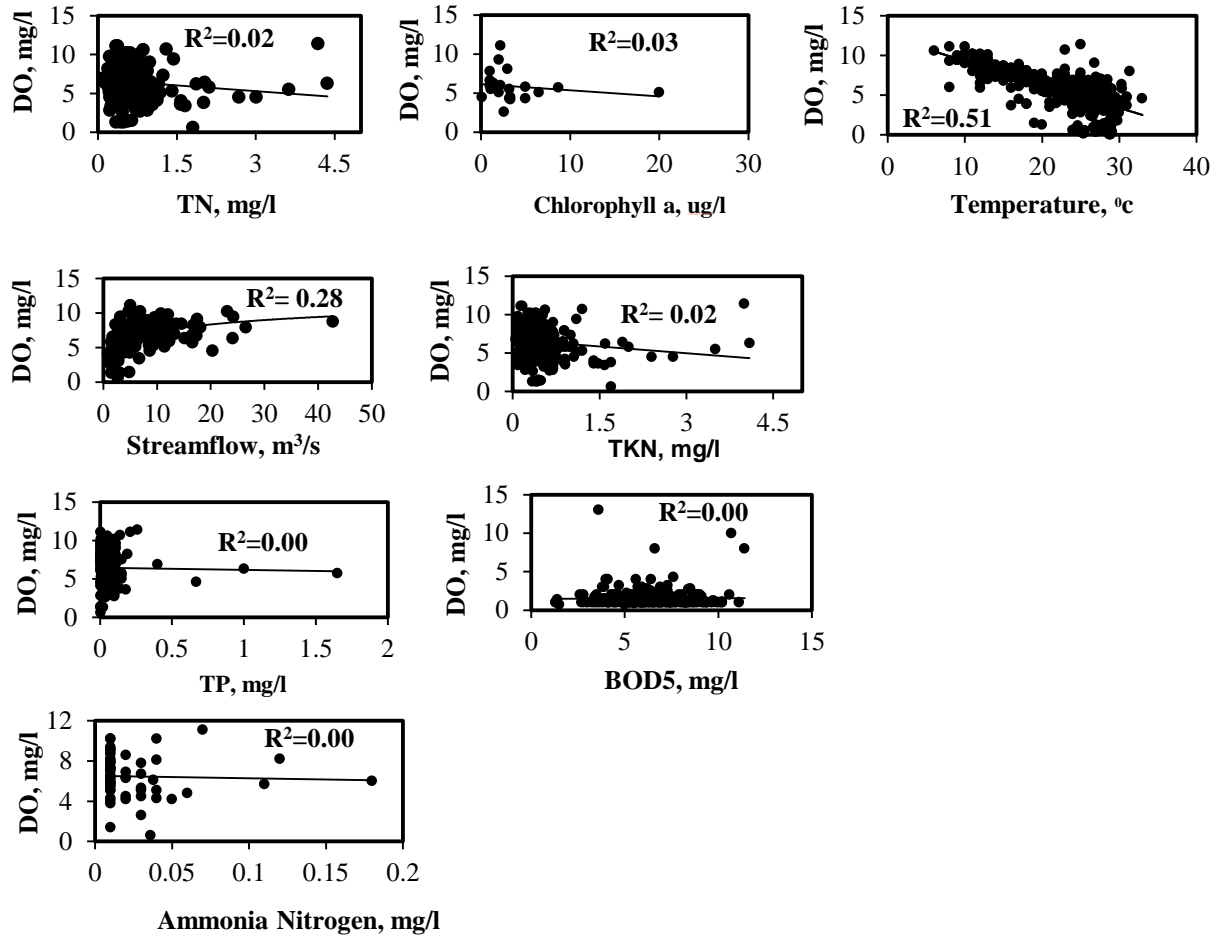


Figure 4.5. Observed DO response for various water quality parameters at monitoring station (CS1).

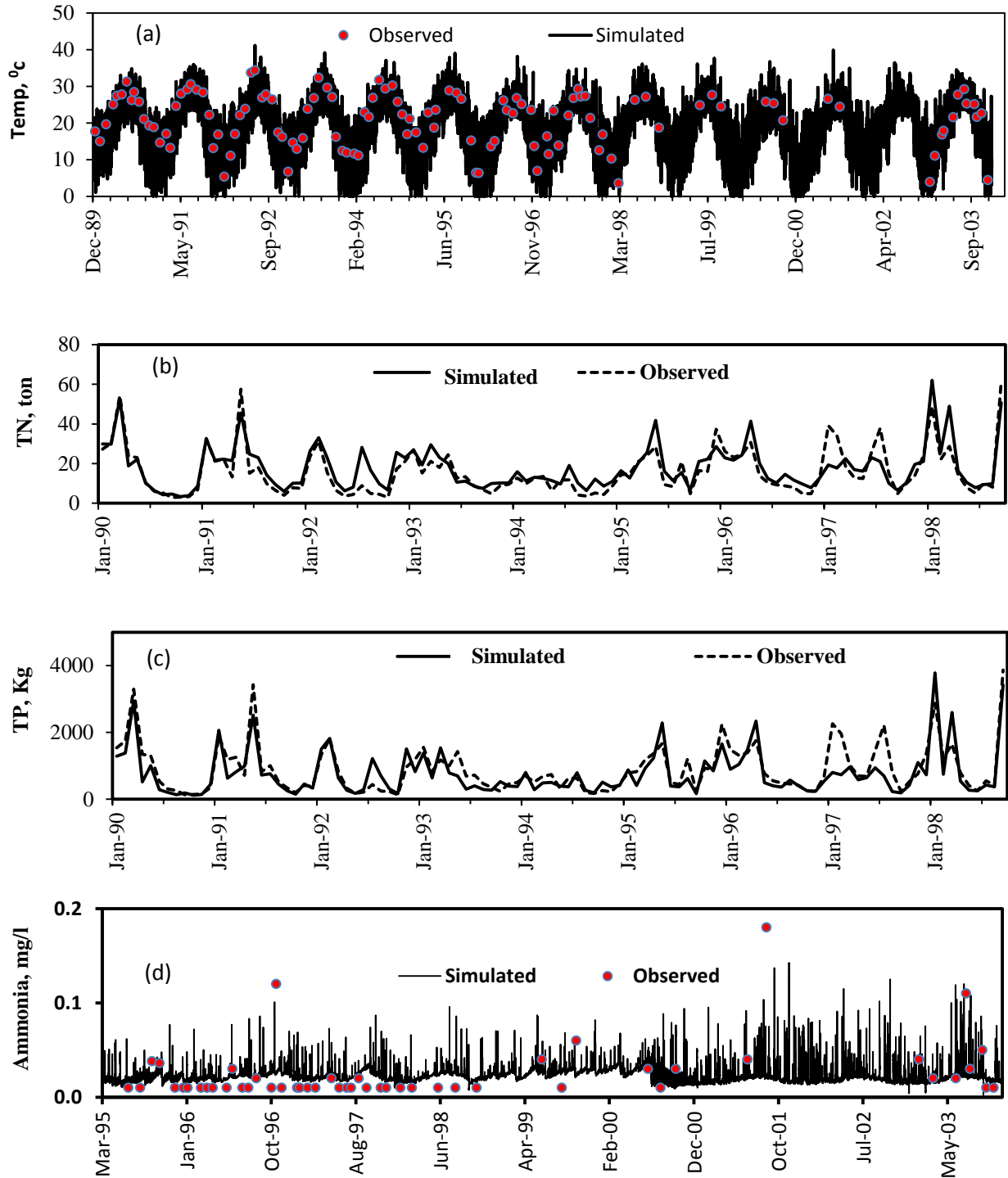


Figure 4.6. Observed and LSPC simulated (a) stream temperature, (b) total nitrogen, (c) total phosphorus, and (d) ammonia at the water quality station CS1 (Figure4. 2). EL and LN stands for El Niño and LN, respectively.

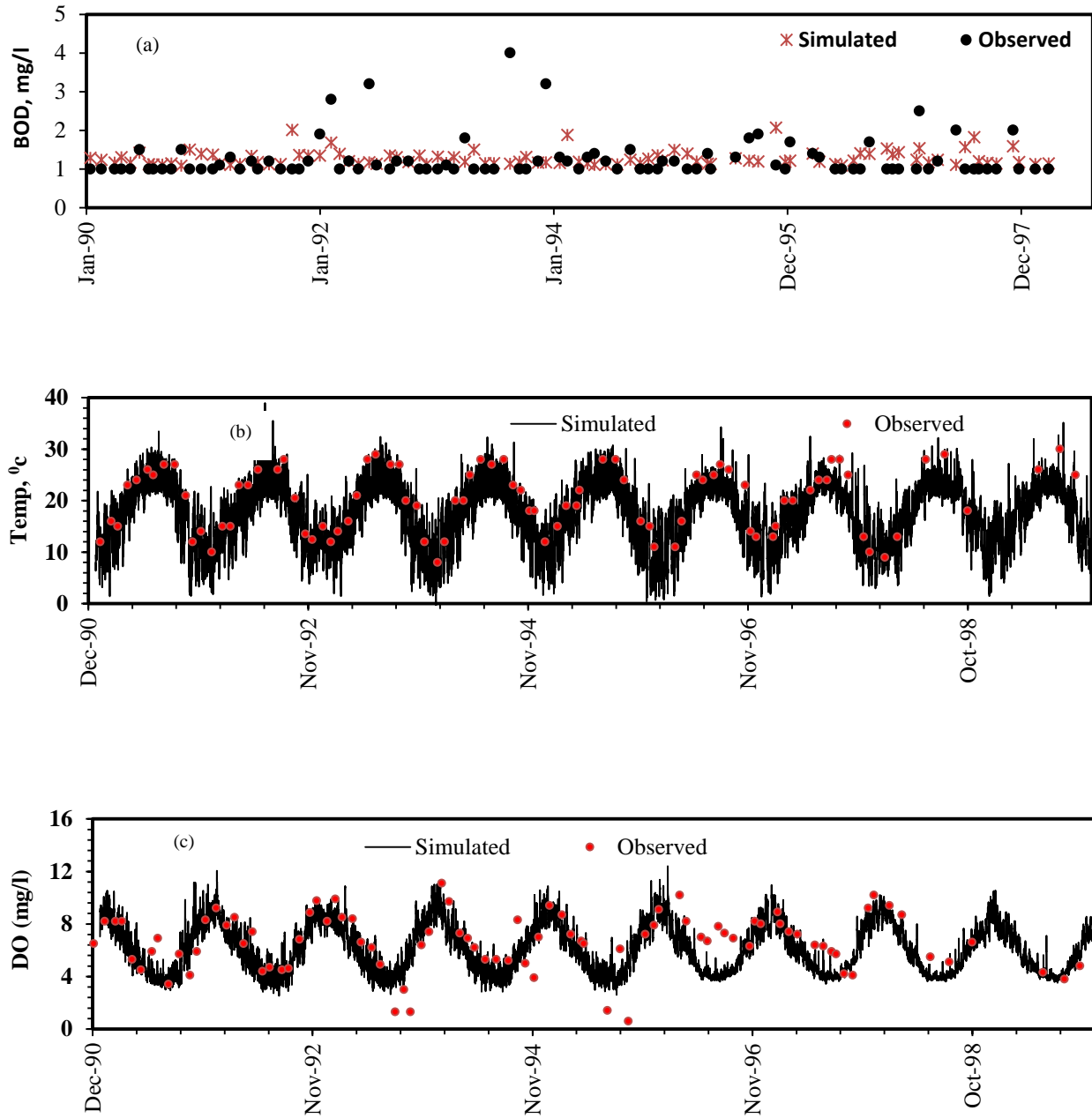


Figure 4.7. Observed and simulated a) BOD, (b) stream temperature, (c) DO at monitoring station CS1 using LSPC linked EPD-RIV1 water quality model.

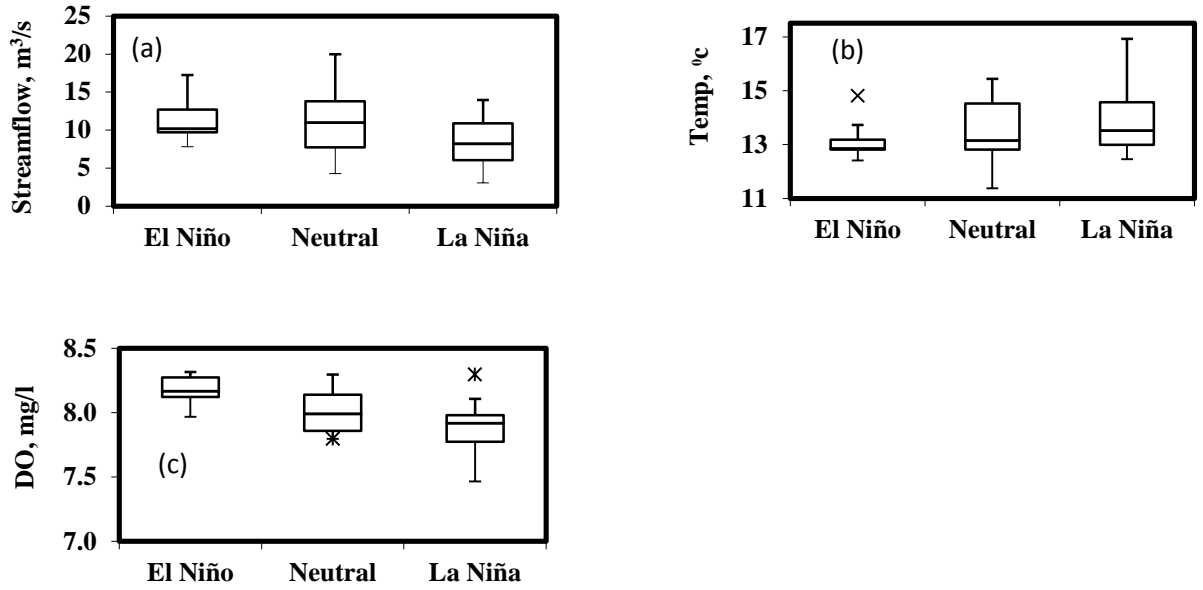


Figure 4.8. Box plot showing effect of ENSO (El Niño, La Niña and neutral phases) on (a) streamflow (Dec-April), (b) stream temperature (Dec-April), and (c) DO concentration (Dec-April).

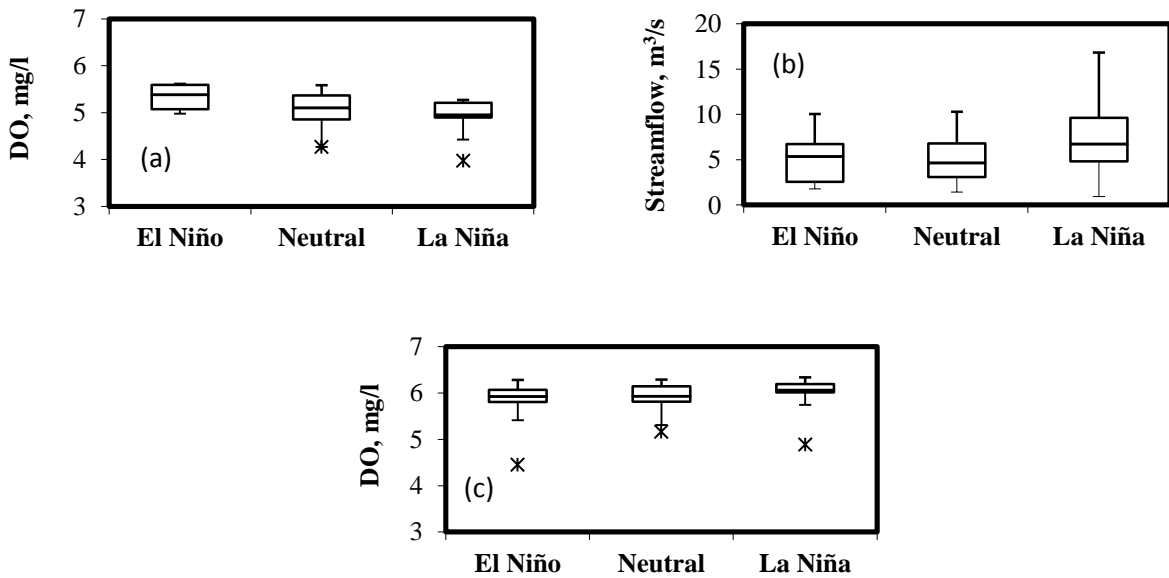


Figure 4.9. Box plot showing effect of ENSO on (a) DO concentration (May-July), (b) streamflow (Aug-Oct), and (c) DO concentration (Aug-Oct).

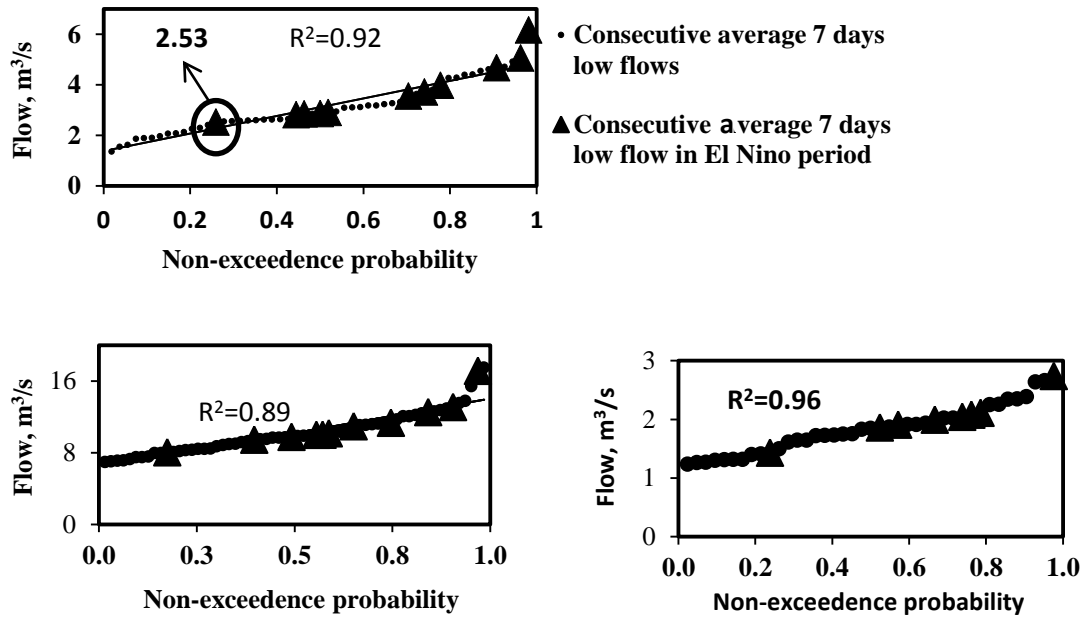


Figure 4.10. 7days consecutive low flows for the period Dec-April in the El Niño phase for (a) Chickasaw Creek, (b) Perdido, and (c) Fish River watershed.

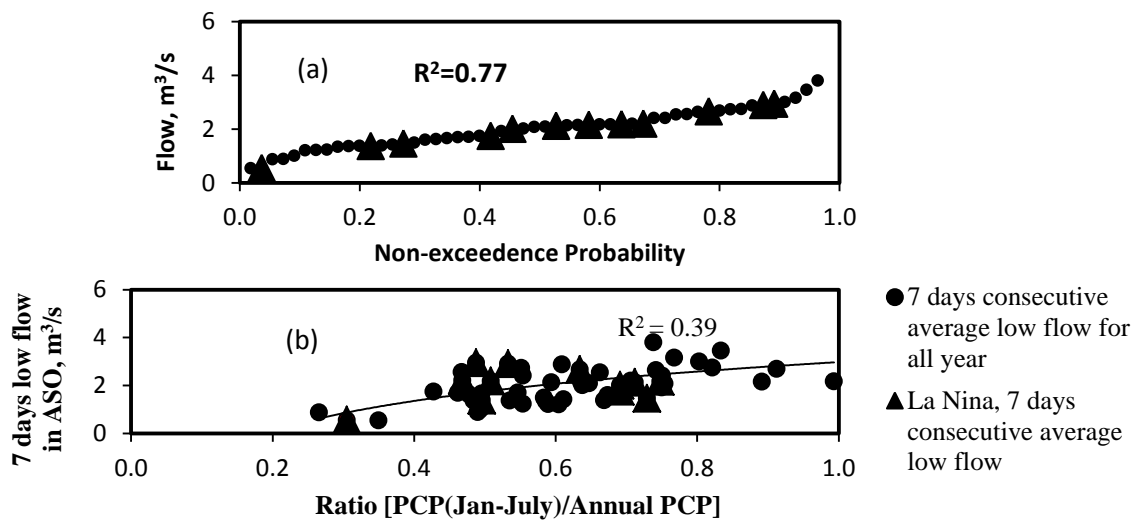


Figure 4.11. (a) Aug-Oct 7days consecutive low flows in La Niña phase in Chickasaw Creek watershed, (b) chart of precipitation index and 7 days consecutive low flows.

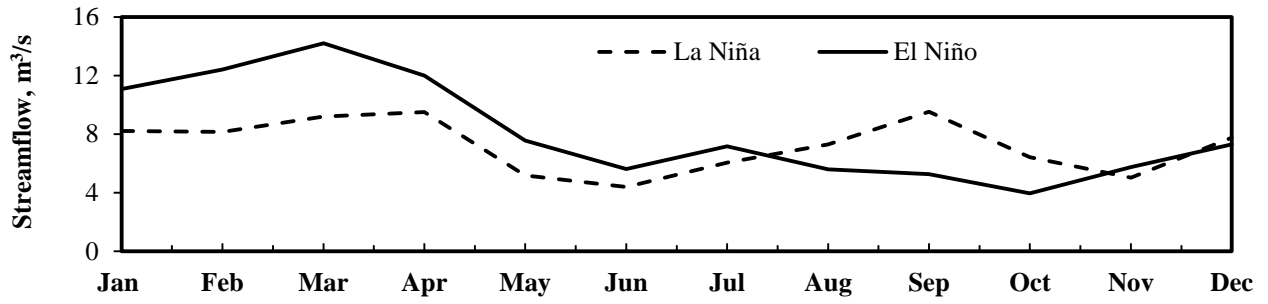


Figure 4.12. Pattern of streamflow in two different ENSO phases (average taken over 50 years).

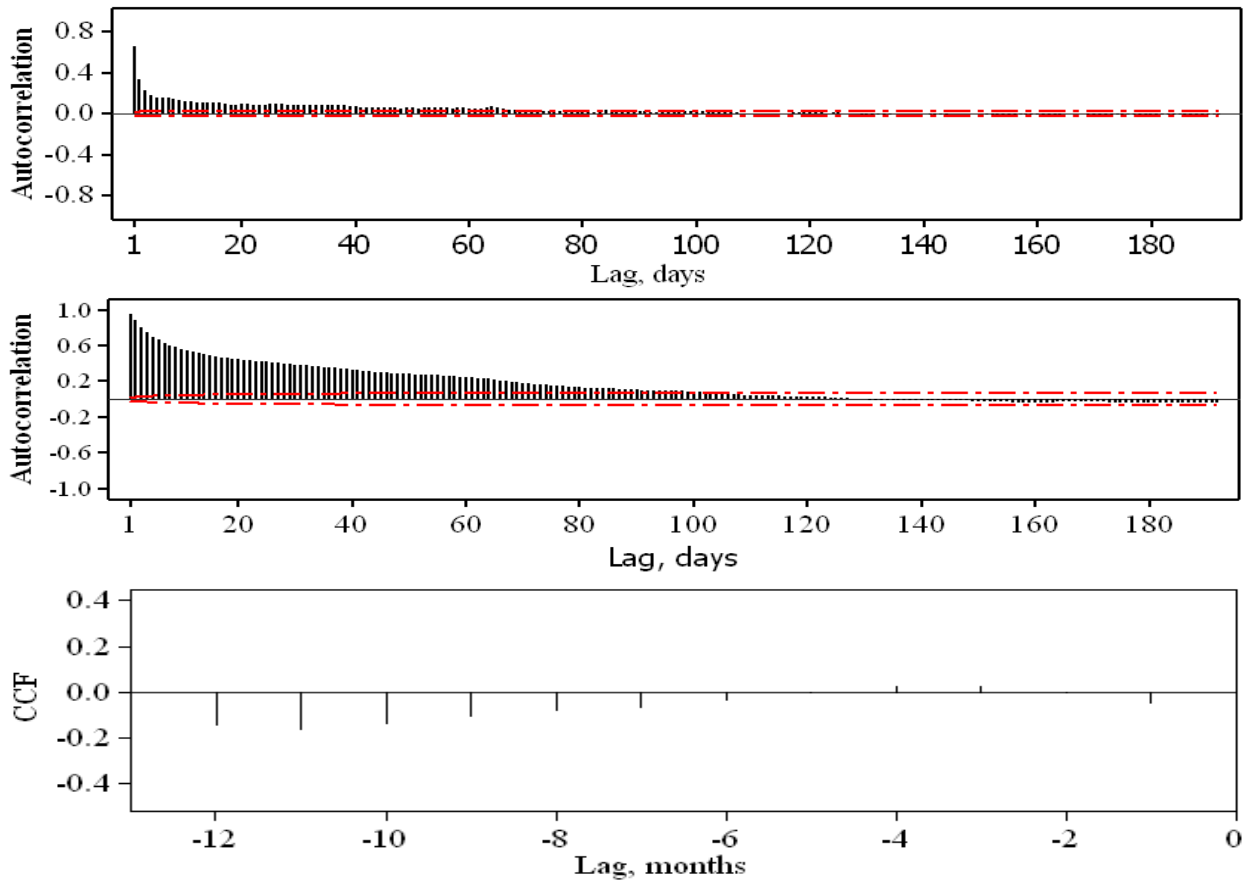


Figure 4.13. a) Autocorrelation of the streamflow b) autocorrelation of base flow c) cross correlation of ENSO with SST anomalies.

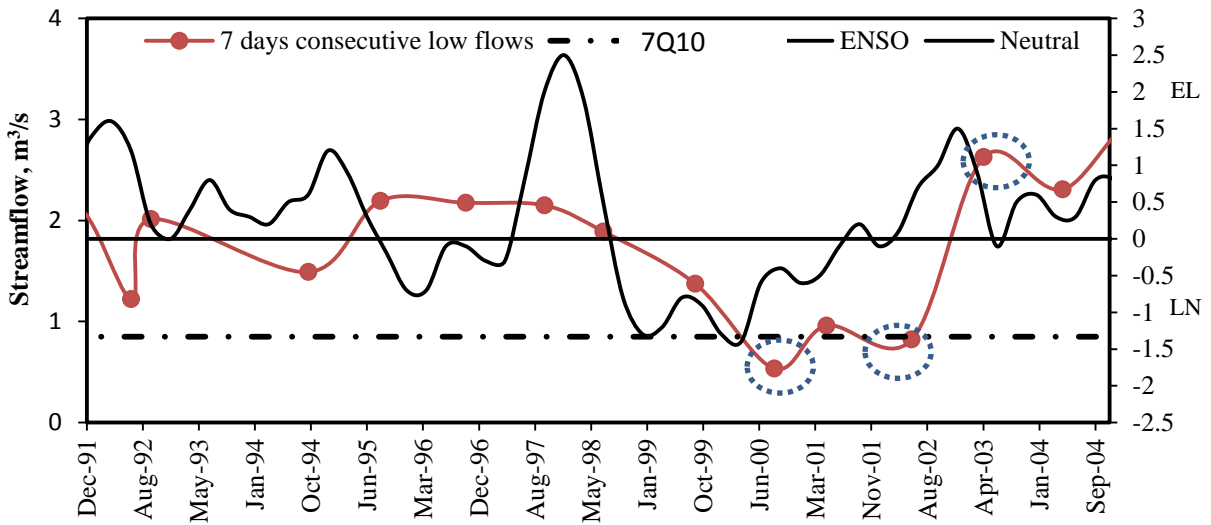


Figure 4.14. Figure showing the association of average 7 days consecutive low flows with Niño 3.4 index (ENSO). Primary horizontal axis (lower) and primary vertical axis (left) is for 7 days low flow of each month and secondary vertical axis (Right) is for Niño 3.4 index. EL and LN represent El Niño and La Niña phase, respectively.

4.8 References

- Aceituno, P. (1992), El Niño, the southern oscillation, and ENSO: confusing names for a complex ocean-atmosphere interaction. *Bulletin of the American Meteorology Society*, 73, 483–485.
- ADEM (2006). "Final nutrient total maximum daily loads (TMDLs) for the Cahaba River Watershed." *Alabama Department of Environmental Management, Water Quality Branch, Water Division*, Montgomery, AL.
- Araoye, P. (2009). "The seasonal variation of pH and dissolved oxygen (DO₂) concentration in Asa lake Ilorin, Nigeria." *International Journal of Physical Sciences*, 4(5), 271-274.
- Barsugli, J. J., Whitaker, J. S., Loughe, A. F., Sardeshmukh, P. D., and Toth, Z. (1999). "The effect of the 1997/98 El Niño on individual large-scale weather events." *Bull. Amer. Meteor. Soc.*, 80, 1399–1411.
- Chiew, F., Piechota, T., Dracup, J., and McMahon, T. (1998). "El Niño/Southern Oscillation and Australian rainfall, streamflow and drought: Links and potential for forecasting." *Journal of Hydrology*, 204(1-4), 138-149.
- Conrads, P. A., Martello, W. P., and Sullins, N. R. (2003). "Living with a large reduction in permitted loading by using a hydrograph-controlled release scheme." *Environmental monitoring and assessment*, 81(1), 97-106.
- Cox, B. (2003). "A review of currently available in-stream water-quality models and their applicability for simulating dissolved oxygen in lowland rivers." *The Science of the Total Environment*, 314, 335-377.
- Crawford, N. H., and Linsley, R. K. (1966). "Digital simulation in hydrology's stanford watershed model ." Technical Report 39, *Dept. of Civil Engineering, Stanford University, California*.
- DEM. (2010). "Digital Elevation Model." WWW data downloaded from www.seamless.usgs.gov.
- Donigian Jr, A., Imhoff, J., Bicknell, B., and Kittle Jr, J. (1984). "Application guide for the hydrological simulation program-FORTRAN EPA 600."
- Dortch, M. (1990). "CE-QUAL-RIV1: A dynamic, one-dimensional (longitudinal) water quality model for streams. user's manual." DTIC Document.
- Emerson, K., Russo, R. C., Lund, R. E., and Thurston, R. V. (1975). "Aqueous ammonia equilibrium calculations: effect of pH and temperature." *Journal of the Fisheries Board of Canada*, 32(12), 2379-2383.

- Erickson, R. J. (1985). "An evaluation of mathematical models for the effects of pH and temperature on ammonia toxicity to aquatic organisms." *Water research*, 19(8), 1047-1058.
- Gómez-Fortis, S., Esteban-Parra, M., Trigo, R., and Castro-Díez, Y. (2010). "Potential predictability of an Iberian river flow based on its relationship with previous winter global SST." *Journal of Hydrology*, 385(1-4), 143-149.
- Hampson, B. (1977). "Relationship between total ammonia and free ammonia in terrestrial and ocean waters." *Journal du Conseil*, 37(2), 117.
- Handler, A. (1990). "USA corn yields, the El Niño and agricultural drought: 1867–1988." *International Journal of Climatology*, 10(8), 819-828.
- Hansen, J., Jones, J., Irmak, A., and Royce, F. (2001). "El Niño-southern oscillation impacts on crop production in the southeast United States." *ASA Special Publication*, 63, 55-76.
- Henry, T., Beck, M., Campbell, P., Montali, D., Ludwig, J. Shen, Parker, A., (2002b). Metals and pH TMDL development for the Tygart Valley. Watershed 2002, *WEF Specialty Conference Proceedings on CD-ROM*, Fort Lauderdale.
- Herb W. and Stefan H.(2008). "A flow and temperature model for the Vermillion River, Part I: Model development and base flow conditions." *Minnesota Pollution Control Agency* St. Paul, Minnesota. Project Report No. 517
- Herb, W., and Stefan, H. (2010). "Projecting the impact of climate change on coldwater stream temperatures in Minnesota using equilibrium temperature models."
- Keener, V., Ingram, K., Jacobson, B., and Jones, J. (2007). "Effects of El Niño/Southern Oscillation on simulated phosphorus loading in South Florida." *Transactions of the ASAE*, 50(6), 2081-2089.
- Kendall, M. G. (1938). "A new measure of rank correlation." *Biometrika*, 30(1/2), 81-93.
- Kroll, C. N. (1992). "Regional geohydrologic-geomorphic relationships for the estimation of low-flow statistics." *Water Resources Research*, 28(9), 2451-2458.
- Kulkarni, J. (2000). "Wavelet analysis of the association between the southern oscillation and the Indian summer monsoon." *International Journal of Climatology*, 20(1), 89-104.
- K.W. Bedford, R.M. Sykes and C. Libicki, (1982). "A dynamic water quality model for stormwater assessment." *US Army Corps of Engineers*, waterway experiment station, MS, USA.

- Lumb, A. M., McCammon, R. B., Kittle, J. L., and Division, G. S. W. R. (1994). *Users manual for an expert system (HSPEXP) for calibration of the Hydrological Simulation Program--Fortran*, US Geological Survey.
- Lung, W. S. (2001). *Water quality modeling for wasteload allocations and TMDLs*, John Wiley & Sons Inc.
- Marcé, R., RODRÍGUEZ- ARIAS, M. À., García, J. C., and Armengol, J. (2010). "El Niño Southern Oscillation and climate trends impact reservoir water quality." *Global Change Biology*, 16(10), 2857-2865.
- Martin, J. L., Wool, T., and Olson, R. (2002). "A dynamic one-dimensional model for hydrodynamics and water quality." *EPDRiv1 Version*, 1.
- McCabe, G. J., and Dettinger, M. D. (1999). "Decadal variations in the strength of ENSO teleconnections with precipitation in the western United States." *International Journal of Climatology*, 19(13), 1399-1410.
- Moriassi, D., Arnold, J., Van Liew, M., Bingner, R., Harmel, R., and Veith, T. (2007). "Model evaluation guidelines for systematic quantification of accuracy in watershed simulations."
- Mosley, M. P. (2000). "Regional differences in the effects of El Niño and La Niña on low flows and floods." *Hydrological sciences journal*, 45(2), 249-267.
- NLCD. (2010). "National Land Cover Dataset." WWW data downloaded from www.epa.gov/mrlc/nlcd-2001.html
- Nash, J., and Sutcliffe, J. (1970). "River flow forecasting through conceptual models part I--A discussion of principles." *Journal of Hydrology*, 10(3), 282-290.
- Pascual, M., Rodó, X., Ellner, S. P., Colwell, R., and Bouma, M. J. (2000). "Cholera dynamics and El Niño-southern oscillation." *Science*, 289(5485), 1766.
- Piechota, T. C., and Dracup, J. A. (1996). "Drought and regional hydrologic variation in the United States: Associations with the El Niño-Southern Oscillation." *Water Resources Research*, 32(5), 1359-1373.
- Rajagopalan, B., and Lall, U. (1998). "Interannual variability in western US precipitation." *Journal of Hydrology*, 210(1-4), 51-67.
- Ries, K. G., and Friesz, P. J. (2000). *Methods for estimating low-flow statistics for Massachusetts streams*, US Dept. of the Interior, US Geological Survey.
- Rodgers, J. L., and Nicewander, W. A. (1988). "Thirteen ways to look at the correlation coefficient." *American Statistician*, 59-66.

- Ropelewski, C., and Halpert, M. (1986). "North American precipitation and temperature patterns associated with the El Niño/Southern Oscillation (ENSO)." *Mon. Weather Rev.:(United States)*, 114(12).
- Roy, S. S. (2006). "The impacts of ENSO, PDO, and local SSTs on winter precipitation in India." *Physical Geography*, 27(5), 464-474.
- Russo, R.C. Russo, Ammonia, nitrite and nitrate, G.M. Rand, S.R. Petrocelli (1985) Editors, *Fundamentals of Aquatic Toxicology*, Hemisphere Publishing Corporation, Washington, DC (1985), pp. 455–471.
- Saunders III, J. F., and Lewis Jr, W. M. (2003). "Implications of climatic variability for regulatory low flows in the South Platte River Basin, Colorado." *JAWRA Journal of the American Water Resources Association*, 39(1), 33-45.
- Saunders III, J. F., Murphy, M., Clark, M., and Lewis Jr, W. M. (2004). "The influence of climate variation on the estimation of low flows used to protect water quality: A nationwide assessment." *JAWRA Journal of the American Water Resources Association*, 40(5), 1339-1349.
- Scarsbrook, M. R., McBride, C. G., McBride, G. B., and Bryers, G. G. (2003). "Effects of climate variability on rivers: Consequences for long term water quality analysis." *JAWRA Journal of the American Water Resources Association*, 39(6), 1435-1447.
- Shen, J., Wang, Harry, Sisson, G.M.,(2002a). "Application of an integrated watershed and tidal prism model to the Poquoson coastal embayment." *Special Report in Applied Marine Science and Ocean Engineering*, No. 380.
- Shen, J., Sullines, N., Park, A., (2002b). "Mobile Bay TMDL development, linking inland and estuarine systems." *Coastal Water Resources, American Water Resources Association, Spring Specialty Conference*, May 13–15, 2002, New Orleans, LA, pp.313–318.
- SSURGO. (2010). "Soil Survey Geographic Database." WWW data downloaded from soildatamart.nrcs.usda.gov
- Stahl, K., and Demuth, S. (1999). "Linking streamflow drought to the occurrence of atmospheric circulation patterns." *Hydrological sciences journal*, 44(3), 467-482.
- Stahl J.K. and Smith R. L. (2002). "Total maximum daily load analysis for Limekiln Brook, Danbury, Connecticut." *State of Connecticut, Department of Environmental Protection*.
- Te Chow, V. (1959). *Open-channel hydraulics*, McGraw-Hill College.

- Tetra Tech. Inc. (2007). "User's manual for meteorological data analysis and preparation tool (Metadapt)". Prepared for USEPA by tetra tech., 10306 Eaton Place, Suite 340, Fairfax, VA 22030.
- Tetra Tech. Inc. (2010). "Flint River watershed modeling report." Prepared by Tetra Tech. Inc. for U.S. Environmental Protection Agency.
- Thurston, R. V. (1990). "Ammonia toxicity to fishes." Environmental Research Laboratory, office of research and development, US Environmental Protection Agency, 183.
- Trenberth, K. E., and Stepaniak, D. P. (2001). "Indices of El Niño evolution." *Journal of Climate*, 14(8), 1697-1701.
- USACE (1995). "CE-QUAL-RI V1: A dynamic, one-dimensional (longitudinal) water quality model for streams, in User's Manual, Instruction Report EL-95-2, US Army Engineer Waterways Experiment Station, Vicksburg, MS.
- USEPA (1997). Technical guidance manual for developing total maximum daily loads, Book 2, Part 1, Section 2.3.3. EPA 823-B-97-002. Washington, D.C.: U.S. EPA, Assessment and Watershed Protection Division.
- USEPA (2004). "Total maximum daily load evaluation for the Coosa River in the Coosa River Basin for dissolved oxygen. "Submitted by the Georgia Department of Natural Resources Environmental Protection Division."
- USEPA (2009). Draft 2009 update aquatic life ambient water quality criteria for ammonia, freshwater.
- Whitfield, M. (1974). "The hydrolysis of ammonium ions in sea water: A theoretical study." *J. Mar. Biol. Assoc. UK*, 54, 565-580.
- Wood, C., and Evans, D. (1993). "Ammonia and urea metabolism and excretion [in fish]." *CRC Marine science series*. 379-426.

Chapter 5. Predicting Total Organic Carbon Load with El Niño Southern Oscillation Phase Using Hybrid and Fuzzy Logic Approaches

5.1 Abstract

During drinking water treatment, chlorine reacts with Total organic carbon (TOC) to form Disinfection byproducts (DBP), some of which can be carcinogenic. Additional treatment required to remove TOC increases the treatment cost significantly. There are two main sources of TOC in a water supply reservoir: (1) the watershed draining to the reservoir, and 2) the internal loading within the reservoir. Out of the two sources, watershed TOC load can be significant especially when the watershed has large wetland areas. The TOC load in the Southeast can be affected by the climate variability phenomenon called El Niño Southern Oscillation (ENSO). Reliable TOC load prediction in different ENSO phases can help reduce the additional treatment cost required for DBP removal. The objectives of this study were to quantify the effect of ENSO on watershed TOC loads and develop data-driven modeling approaches for TOC load prediction. Four Principal Component Regression-Artificial Neural Network (PCR-ANN) and four fuzzy logic models, with different model architectures, were developed for predicting watershed TOC loads using temperature, precipitation, Niño 3.4 index and Trans Niño index (TNI). This study concludes that PCR-ANN models are suitable for estimating real-time TOC loads and fuzzy logic models are suitable for qualitative forecast of TOC loads at one month lead time. In addition, the study highlights the importance of incorporating ENSO information into the data-driven models in the ENSO-affected region.

5.2 Introduction

Total organic carbon (TOC) is a water quality variable of specific interest in water supply reservoirs because it can form Disinfection byproducts (DBP) (Elias et al. 2011). There are primarily two sources of TOC in a water supply reservoir: a) watershed and b) internal loading of the reservoir (Elias et al. 2011). Out of these two sources, a watershed can contribute high TOC loads, especially when the watershed has large areas in wetlands (Morrison et al. 2006; Gergel et al. 1999). Organic carbon can be elevated in soils, sediments and streams due to various anthropogenic influences and natural processes. Watershed organic carbon load can be contributed by diverse inputs, such as clear cuttings, agricultural patterns, animal waste applications, and different land use practices (Moore 1989).

During drinking water treatment, TOC reacts with chlorine and forms DBP (Reckhow et al. 1990; Pomes et al. 1999; Aiken et al. 1995) due to chemical disinfection (Singer and Chang 1989; Singer 1994). Therefore, TOC in source water is considered as one of the indicators of DBP formation. A study by USEPA (2005) has suggested that a TOC concentration greater than certain threshold can increase the formation of DBP significantly, some of which are reported as carcinogenic (Moore 1989; USEPA 2005). USEPA (2005) has illustrated a potential linkage between bladder cancer and exposure to chlorinated drinking water. Around 70,530 cases of bladder cancer have been estimated by the American Cancer Society (ACS 2010) as occurring each year in United States. In order to address this alarming situation, there are two options: (1) minimize DBP formation by reducing TOC in the source through watershed management (Walker Jr 1983; Canale et al. 1997), or treat drinking water in the treatment plants. Because acquiring land and monitoring the watershed for TOC source protection may incur huge costs, a large number of treatment plants (2260) use additional treatment to reduce TOC (USEPA 2005).

In this context, estimating TOC load from the watershed is vital for the water quality managers to be able to maintain the desired concentration in the water supply reservoir. Because several studies in the past (Correll et al. 2001; Chang and Carlson 2005) have reported seasonal variations in TOC loads, these loads should be predicted separately for each season.

In addition to seasonal variability, TOC load in the Southeast USA could be affected by interannual climate variability (Elias 2010), resulting from the coupled oceanic and atmospheric phenomenon called El Niño Southern Oscillation (ENSO). This hypothesis is supported by several studies in the past which show a strong influence of ENSO on precipitation, temperature and streamflow (Chiew et al. 1998; Piechota and Dracup 1996; Rajagopalan and Lall 1998; Handler 1990; Kulkarni 2000; Hansen and Maul 1991; Pascual et al. 2000; Keener et al. 2007; Roy 2006; Barsugli et al. 1999; McCabe and Dettinger 1999). The proper prediction of TOC load in different ENSO phases should be helpful in minimizing the additional treatment cost required to treat DBP. Therefore, TOC load correlation with ENSO phase should be evaluated. If a good correlation between TOC load and ENSO phase exists, ENSO information can be used to forecast TOC load. In this case, I can predict TOC loads with a single input without running a watershed model. However, predicting TOC load based upon ENSO phase would be possible only for specific seasons. One of the best approaches to predict TOC load for all seasons is to utilize the seasonal climatic forecast (temperature and precipitation) and ENSO information provided by the National Climatic Environmental Prediction (NCEP) of the National Oceanic and Atmospheric Administration (NOAA).

NCEP has recently started using the second generation, climate forecast system version 2 (CFSv2) with increased spatial and temporal resolution to improve seasonal climate (temperature and precipitation) forecasts (Yuan et al. 2011). These inputs can be used to predict TOC loads

using fully calibrated, watershed models. However, the calibration and validation of watershed models demand a high level of expertise, time, and detailed watershed information (Srivastava et al. 2006). Further, model development might be difficult due to the non-linear and fuzzy characteristics of the underlying processes, especially when large uncertainties are associated with the watershed model parameters (Tayfur et al. 2003). In order to address this issue, data-driven models that utilize temperature, precipitation and ENSO indices may be a better choice. The reason for emphasizing the data-driven models with these particular inputs is due to the capability of the CFSv2 model regarding surface air temperature and precipitation predictions (Yuan et al. 2011). A few attempts in the past have explored data-driven modeling approaches using different climate indices for seasonal (G miz-Fortis et al. 2010; Karamouz and Zahraie 2004; Piechota and Dracup 1996; Awadallah and Rousselle 2000; Grantz et al. 2005; Tootle et al. 2007) and annual scales by combining them with different oceanic-atmospheric indices (Kalra and Ahmad 2009), such as, Pacific Decadal Oscillation (PDO) (Hamlet and Lettenmaier 1999), North Atlantic Oscillation (NAO) (Araghinejad et al. 2006), and Atlantic Multidecadal Oscillation (AMO) (Rogers and Coleman 2003). However, these studies were limited to streamflow forecasting on seasonal and annual scales, suggesting that climate indices alone are not sufficient for simulation on a monthly scale. For monthly TOC load predictions, ENSO information and climatic variables were utilized as inputs for fuzzy logic and hybrid models. Hybrid models combine the Principal Component Regression with an Artificial Neural Network (PCR-ANN). Hence the terminology referred to as ANN in this manuscript should be considered as part of a hybrid (PCR-ANN) model.

This research explores the potential of using hybrid and fuzzy logic approaches to predict TOC load for both quantitative and qualitative scales. The specific research objectives are to: (i)

to incorporate ENSO information for TOC loads prediction using hybrid approach ii) to make a qualitative forecast of TOC loads in different ENSO phases using fuzzy logic approach.

5.3 Study Area and Data

The study was focused in the Big Creek watershed in Mobile County of South Alabama (Figure 5.1), which is located near the Mississippi State border in the Mobile River Basin. The watershed, which drains to the Converse Reservoir, is located in the Escatawpa hydrologic cataloging unit (8-digit hydrologic unit code: 03170008). This watershed is a major source of TOC load to the Converse Reservoir, a primary source of drinking water for the city of Mobile (Elias 2010). The watershed is located by the coastal plain ecoregion with elevation ranging from 55 ft to 102 ft. The average annual precipitation is 65 inches, which is relatively higher than that of northern and central Alabama.

The long term, climate data (52 years) were available at the climate station (Coop ID-015478) from the National Climatic Data Center (NCDC). Streamflow data recorded since 1990 were available at the USGS gage (Station Id 02479945) which drains 31 sq. miles of the watershed. Stream water quality data were collected by the USGS, Auburn University (AU), and the Mobile Area Water and Sewer System (MAWSS) for different projects at different times from 1990 to 2005. The watershed is mainly comprised of deciduous, evergreen and mixed forest (45%), woody and herbaceous wetland (14%), range shrubland (23%), and grassland herbaceous and hay (11%). The remaining watershed land (17%) includes agricultural and urbanized areas. The land use practice in the watershed has been stable since 1990 (Srivastava et al. 2010). The watershed soil is characterized predominantly by Troupbenndel (46 %). Other soil groups are Troupeidel (22%), Bama (10 %), Heidel (9%), Notcher (8%), and Troup (5%). High resolution

soil data from the Soil Survey Geographic (SSURGO) (SSURGO 2010) soil database was used to acquire soil-related parameters. Digital Elevation Model (DEM) (DEM 2010) of 10 m resolution was used for watershed delineation and extraction of the stream network for watershed modeling. The following section briefly describes TOC load generation under static land use conditions.

5.3.1 TOC Load Generation

Since I wanted to demonstrate the effect of climate variability on TOC loads using long term data sets, I generated TOC load data for another 40 years using a calibrated and validated watershed model.

Since the observed data indicates the resultant effect of land use and climate variability (Keener et al. 2007), long term data sets that don't include the land use effect are needed. It is recommended to use the simulated data instead of observed data to demonstrate the effect of climate variability on TOC load simulation, and this is the best way to look for the direct link between TOC loads with ENSO without confounding it with land use/land cover effect. Therefore, I selected 15 years of stable periods in terms of land use/land cover data to calibrate and validate the loading simulation program C++ (LSPC) model.

The Loading Simulation Program C++ (LSPC) model (Henry et al. 2002a; Shen et al. 2002 a, b), which has been widely used for TMDL developments, was used for TOC load simulation. LSPC is one of the most advanced hydrologic and watershed loading models. The SSURGO soils data and 2001 National Land Cover Data (NLCD) (NLCD 2010) were used as model inputs to simulate a series of hydrologically connected sub watersheds, characterized by defined geometry, soil and land use characteristics. Hourly climate data are needed for an

appropriate representation of a hydrologic response. The NCDC climate data station (COOP ID - **015478**) located in Mobile Regional Airport was utilized for these data. Any missing data were obtained from the nearest stations. A USGS gage (02479945) was used for the evaluation of both streamflow and TOC load simulation. The LSPC model was successfully calibrated ($NSE=0.71$) from November 1997 to December 2005 and validated ($NSE=0.72$) from January 1990 to October 1997, respectively, for monthly streamflow (Figure 5.2). Similarly, the TOC loads were calibrated ($NSE=0.62$) and validated ($NSE=0.33$) for the same period (Figure 5.2). I could have obtained very good model performance ($NSE > 0.8$), both for calibration and validation, as manifested by the systematic error in Figure 5.2. However, I considered that getting a higher magnitude of NSE does not necessarily result in a good model. Since I compared the model-simulated result with the Loadest-generated data, the observed data seems to demonstrate a higher value for the peak-simulated load. The Loadest relies on regression equations (Runkel et al. 2004) which sometimes have a tendency to over-or under-predict, depending upon the type of equation chosen. Therefore, instead of adjusting the simulated results, I simply overlaid these results onto the Loadest-generated data and evaluated the NSE. The model parameters were very close to the parameters adopted in the Mobile Bay Project (Childers 2009). The model parameters also matched those of a previous study in this watershed (Elias 2010). The next step, after generating the TOC load, was to study the impact of climate variability due to ENSO on TOC load simulation. The following paragraph discusses ENSO and various ENSO indicators.

5.3.2 ENSO and ENSO Indicators

ENSO is a coupled, atmosphere-ocean phenomenon occurring at interannual time scales, due to the complex interplay of different climatic variables such as clouds, storms, winds,

oceanic temperatures and oceanic currents along the Equatorial Pacific Ocean (Trenberth and Stepaniak 2001).

Several ENSO indices are discussed in the literature for the classification of the phase and strength of ENSO events. Some of the ENSO indices are the Southern Oscillation Index (SOI) (Trenberth and Shea 1987; Trenberth 1984; Chao and Philander 1993), Japanese Meteorological Agency (JMA) (Hanley et al. 2003), and Multivariate ENSO Index (MEI) (Wolter and Timlin 1993). Examples of the indices used to define ENSO phases are Sea surface temperature (SST) indices in different regions of the Equatorial Pacific, such as Niño-1+2, Niño-3, Niño-4, and Niño 3.4 etc. Finding the most suitable index for one's specific needs is the most important step, as there is no consensus among scientists and climatologists as to which of the ENSO indicators is the best at capturing the ENSO phases (Hanley et al. 2003). Hanley et al. (2003) carried out a study using all the aforementioned indices and did not find a single index that is superior at capturing the ENSO phases. Some indices are more responsive to La Niña and less responsive to El Niño and others have the exact opposite characteristics. The study suggests combining the indices that show better results in El Niño with the index that demonstrates better performance in La Niña, using either a linear or non-linear approach.

5.3.3 Selection of ENSO Indices

I used a linear approach to evaluate the performance of ENSO indices, but used a non-linear approach to combine the indices for their application as inputs in the data-driven models. A Multi Linear Regression (MLR) model was applied to find the best fit for the TOC load and different ENSO indices using the long term TOC data sets. Different possible options suggested by Hanley et al. (2003) were evaluated for finding the potential indicators in the Equatorial

Pacific (not shown). These combinations are consistent with the studies carried out by Trenberth and Stepaniak (2001) and Hanley et al. (2003).

The combined ENSO indices Niño 3.4 and Trans-Niño index (TNI) resulted in the largest R^2 value (0.42) with TOC load, and therefore were considered to be the best choices of ENSO indices for the proposed study. This finding is consistent with a previous study (Trenberth and Stepaniak 2001), which emphasizes using at least two indices for optimal characterization of El Niño and La Niña events. The Niño 3.4 index is developed based on SST anomalies in the Niño 3.4 region (5°N -5°S and 170°W - 120°W) of the Pacific, which is regarded as a region of strong correlation with sea level pressure and temperature anomalies. TNI is the gradient in SST, which is the difference between normalized anomalies of SST in the Niño 1+2 and Niño 4 regions. Readers are suggested to refer to the article by Hanley et al. (2003) for more information. The Niño 3.4 index (Zubair et al. 2003) was available from 1950 to 2010 from the NOAA Climatic Prediction Center. TNI was taken from the University Corporation of Atmospheric Research (UCAR) under the Climate and Global Dynamic Division (CGD) of the National Center for Atmospheric research (NCAR). These indices were applied as inputs in data-driven model, which is briefly discussed in the following paragraph.

5.4 Data-Driven Modeling Approach

The data-driven models are based on the relationship between input and output records without looking into the details of the physical phenomenon. Many studies emphasize the usefulness of the empirical or the parametric types of modeling rather than the process-based or physically-based models (Pramanik and Panda 2009) due to limited data sources for physically-based modeling. Data-driven models such as ANN and fuzzy logic are explored in this study.

The ANN and fuzzy logic models with their successful applications against physically-based model have been tested several times in the past (Jain and Indurthy 2003; Abedi-Koupai et al. 2009; Shrestha and Simonovic 2010). The application of ANN models varies and includes water quality modeling (Kalin et al. 2010), ground water modeling (Trichakis et al. 2009), streamflow modeling (Srivastava et al. 2006), rainfall prediction (Wong et al. 2003), hydrologic processes (Bowden et al. 2004; Sudheer et al. 2002), and rainfall-runoff processes (Govindaraju and Rao 2000). Likewise, the fuzzy logic approach has been applied to sediment modeling (Mitra et al. 1998; Tayfur et al. 2003; Kisi 2006; Mianaei and Keshavarzi 2010), precipitation (Maskey and Price 2004), event-based rainfall-runoff modeling (Tayfur and Singh 2006), water supply (Mahabir et al. 2003), and reservoir operation (Deka and Chandramouli 2009). In the fuzzy logic approach, imprecise models can be developed based on human knowledge, skills and understanding, which can better handle the uncertainties associated within the underlying system through the user's prior knowledge. Both data-driven models (ANN and fuzzy logic) are widely used for forecasting due to their proven forecasting ability. Various forecasting approaches using ANN models (Hsu et al. 1995; Shamseldin 1997; Thirumalaiah and Deo 2000) and fuzzy logic models (Hundecha et al. 2001; Özelkan and Duckstein 2000; Chang et al. 2005) are available. I apply these data-driven approaches to develop a model to utilize ENSO forecasts and other climatic variables to predict future TOC loads.

5.4.1 Artificial Neural Network

The ANN is formulated with the inspiration of a biological neural network (ASCE 2000a). It imitates the human brain function of obtaining information through the learning process. ANN can solve complex, non-linear problems involving hydrological, water quality, groundwater and several other water resource-related issues (ASCE 2000a; Raghuwanshi et al.

2006; Srivastava et al. 2006) due to its capacity to learn, memorize and generalize from the given datasets. Its unique capacity to self-organize makes it possible to reproduce output for given inputs. The details of the ANN model and their mathematical aspects have been discussed in different articles (Govindaraju 2000; ASCE 2000a).

5.4.2 Principal Component Regressions with Artificial Neural Network (PCR-ANN)

MLR is one of the simplest approaches to express the independent variables with the response variables linearly. Although the MLR approach has been extensively used in many research and engineering applications, its application can be seriously questioned when the independent variables are correlated with each other (Cureton and D'Agostino 1983; Weisberg 1985; Fritts 1991; Jennrich 1995). The multi-collinearity or correlation among the independent variables makes it difficult to correctly identify the most significant parameters in the model. One of the approaches to remove such multi-collinearity is to apply multivariate analysis, such as Principal Component Regression (PCR) (Hidalgo et al. 2000; Xuan et al. 2010). PCR has been widely applied to streamflow forecasting (e.g., Eldaw et al. 2003). The new variables, after using principal component analysis, are appropriate to use as predictors because the complications due to multi-collinearity are removed. In addition, the hybrid method using the PCR with ANN is considered an even better approach to use as opposed to using a single method (either PCR or ANN) because it captures unique features in the data sets. Several studies in the past (Clemen 1989; Zhang 2003) have suggested using combined methods from different models for improved prediction over the predication of individual models. Readers are suggested to refer to the article by Al-Alawi et al.(2008) for more details.

The PCR-ANN method is a combined or hybrid approach which blends a linear PCR model with a nonlinear (ANN) model. The relationship between linear and nonlinear components can be expressed as follows:

$$Y_t = L_t + N_t \quad (5.1)$$

where L is the linear component and N the non-linear component. L is computed using the PCR technique of using different numbers of principal components (PCs). In the next step, the residuals are computed from the linear (PCR) model which contains the non-linear component.

$$R = Y_t - \bar{L}_t \quad (5.2)$$

where Y_t is the observed value, and \bar{L}_t is the simulated value from the PCR model. The ANN is applied for residual extracting PC1, PC2, and PC (n) by principal component analysis.

$$R = f[\text{PC1}, \text{PC2}, \text{PC}(n)] + \varepsilon_t \quad (5.3)$$

Finally, the ANN model is employed for the resulting residuals (R). The simulated result using the ANN model is expressed by \bar{U}_t

$$Y_t = \bar{L}_t + \bar{U}_t \quad (5.4)$$

Hence, both PCR and ANN models are utilized to estimate the linear and non-linear components, respectively. The fitted residual data from the neural network model can be combined with the PCR model. Readers are suggested to refer to the article by Zhang (2003) for detailed procedures of the hybrid approach.

5.4.3 Fuzzy Logic Approach

Our motive in exploring the fuzzy logic approach was to develop a qualitative assessment and make a subjective forecast. This kind of approach will be more relevant than a deterministic approach from the implementation perspective. Besides, decisions regarding the real world

systems for monitoring and controlling TOC load are not based on a particular magnitude of TOC load.

Implementing agencies devise rules for watershed and water supply reservoir protection based on categorization of loads under subjective (qualitative) range (e.g. “low load,” “medium load,” and “high load”). Qualitative forecasting is a better choice if the inputs for the forecasting model are not deterministic (i.e., probabilistic) in nature. Another advantage of using the fuzzy logic approach is its capacity to deal with the input data which are “fuzzy” in nature. Although climate prediction is improving due to advances in climate sciences, it has certain degree of uncertainty. The fuzzy logic approach can handle these kinds of uncertainties (Shrestha and Simonovic 2010).

Fuzzy logic is derived from the fuzzy set theory which classifies the objects with flexible boundaries considering the membership function as a matter of degree. Fuzzy logic is a mathematical procedure based on the If-then rule system for mapping the human way of thinking in a computational way (Tayfur et al. 2003; Kisi 2006; Panigrahi and Mujumdar 2000). There has been tremendous advancement in the “fuzzy” concept and operational algorithm since its origination by Zadeh (1965). Utilizing the human expert’s knowledge and experience, it can map the set of inputs into output utilizing a fuzzy inference method (Chau et al. 2005). Input variables are generally categorized as “low,” “medium” and ‘high and the fuzzy rules are developed based on the user’s knowledge.

The basic concept behind fuzzy logic is the fractional possessions of any object to different subsets rather than entirely belonging to a single set. The degree of belongingness to a set is expressed quantitatively by a membership function for which value is assigned between 0 and 1. Fuzzy membership function can take many forms, but the triangular function with equal

base width is the simplest and most preferable (Russell and Campbell 1996). Values are entirely expressed on a qualitative scale in linguistic terms with their associations represented in terms of If-Then rules. The fuzzy system can be conceptualized with four basic components: i) fuzzification, ii) fuzzy base rule, iii) fuzzy output engine, and iv) defuzzification (Tayfur and Singh 2006) as shown in Figure 5.3. Interested readers can refer to the articles by Jantzen et al. (1999), Tayfur et al. (2003) and Kisi (2006) for detailed information about the fuzzy logic algorithm.

5.5 Model Evaluation Criteria

The model performance is always evaluated using several non-dimensional measures because there is not a single, best statistical criterion to measure the performance of a model's simulated outputs with actual data. These model performance measures are the Nash-Sutcliffe Coefficient of Efficiency (*NSE*) (Nash and Sutcliffe 1970), coefficient of determination (R^2), Mean Square Error (*MSE*), and Mass Balance Error (*MBE*). In addition, the Akaike Information Criterion (*AIC*) is generally used to find optimal ANN architectures (Kalin et al. 2010; Ren and Zhao 2002; Qi and Zhang 2001). Detailed information regarding these statistical measures are available in different publications (Qi and Zhang 2001; Srivastava et.al 2006; Kalin and Hantush 2006).

5.6 Results and Discussion

5.6.1 ENSO Correlation with TOC Load

In order to predict the TOC load on a monthly scale in each season, I developed seasonal models by dividing the data sets into three seasons, with respective La Niña/El Niño conditions. The model for the neutral conditions, referred to as neutral model hereafter, was developed irrespective of the season. There are many reasons for developing a separate neutral model and seasonal models that represents La Niña and El Niño conditions. The primary reason is that the ENSO phase and its correlation characteristics (positive or negative) with TOC load vary seasonally throughout the year (Figure 5.4). For example, in summer months, the ENSO signal is not very distinct, and the effect of ENSO on TOC loads in the August-September-October (ASO) season is opposite to the effect in Jan-Feb-March (JFM) (Figure 5.4). Another reason is that Niño 3.4 index and Trans Niño index, to which I refer as ENSO indicators in this manuscript, are not significant parameters in ASO seasons and the neutral phase. This is true for all the neutral conditions irrespective of the seasons. The final reason for developing seasonal models was to make models fully capable to utilize the NCEP seasonal forecasts (temperature, precipitation and SST anomalies), which NCEP forecasts with one season lead time.

In order to detect the significance of the input parameters in different seasons, I developed the MLR model before employing inputs into the PCR- ANN model. The MLR model depicted that SST anomalies and precipitation are significant for JFM whereas temperature was not significant. This was evaluated at a significance level of $p \leq 0.05$ (Table 5.1). This indicates that temperature-dependent biological activity does not have a substantial influence on TOC simulation in JFM, and therefore I can ignore temperature in this season. Precipitation and temperature were found to have a significance level of $p \leq 0.05$ for both the April-May-June

(AMJ) and ASO seasons whereas SST anomalies were not significant in the ASO season. This indicates to the fact that the ENSO phenomenon has little influence on TOC simulation in ASO and AMJ, and TOC loads in these seasons are primarily driven by precipitation and biological activity pertaining to temperature. In addition, Table 5.1 shows that the SST anomalies were significant at $p\text{-value} > 0.05$ (95% CI) in AMJ.

These four models (JFM, AMJ, ASO and neutral) covered most of the seasons for the entire year with all ENSO phases except July, November and December. Since the ENSO characteristics in December closely resemble the ENSO characteristics in January, the JFM model can be utilized when making the TOC load estimation for December. For July and November, the Neutral model can be utilized for all ENSO phases because the TOC load in La Niña and El Niño phases are not significantly different than the loads in Neutral phase.

5.6.2 PCR- ANN Model Training and Testing

Although the statistical parameters that measure the performance of the MLR models were satisfactory, I developed PCR-ANN models because of two reasons: (1) the explanatory variables are not independent as the SST anomalies are also correlated with precipitation and temperature, and I wanted to remove co-linearity among the parameters in the input data vectors, and (2) the PCR-ANN model is superior to the MLR model in addressing the non-linear nature of the system, which was suggested by many studies in the past (Abudu et al. 2011). Once I determined the residuals using the actual data and principal component regression in Minitab 16, I fitted it into ANN.

I divided the data sets into three components: 60% data for training, 20% for validation and 20% for testing. The La Niña and El Niño months from 1950 to 2005 were arranged with

their respective precipitation and temperature data sets. As discussed earlier, I chose four inputs (temperature, precipitation, Niño 3.4, and TNI). I did not experiment with other input datasets because, by reducing irrelevant inputs and computational complexity, more efficient models can be developed (Bowden et al. 2005).

The next step was to use those optimum input datasets and determine the optimum number of hidden neurons. The optimum number of hidden neurons can be estimated using the rule of thumb. The number of nodes recommended is in a range of $2n+1$ to $2\sqrt{n+m}$, where 'n' is the number of input nodes and 'm' is the number of output nodes (Qi and Zhang 2001; Fletcher 1993). The numbers of hidden neurons were determined, as discussed in the following paragraph, by writing a separate code in MATLAB.

I experimented with different hidden neuron trials by varying the hidden neurons up to 10, with 50 simulations for each hidden neuron, which resulted in 500 simulations. The number 10 was chosen using the empirical method proposed by Fletcher (1993). The hidden neurons corresponding to the best performance in this trial were selected for the model input (Table 5.2). In addition, I tested the model performance with randomly ordered data sets prepared using a random number generator to ensure that the model was performing well for all kinds of data sets over the period of training, validation and testing. However, by using randomly ordered data, the data used for testing and validation did not remain within the range of the data used in the training stage. Finally, the model was developed for the regular data sets, which attests to the fact that proper datasets selection for training is one of the most crucial parts of the ANN model development.

Input data were normalized before being applied to the model. By data normalization, the model treats the entire range of data equally and improves training (Srivastava et al. 2006). Data

training in ANN is a process in which the given outputs are compared to the expected outputs which minimizes the root mean squared error (Srivastava et al. 2006). Since overtraining of the model reduces its predictive capability (Amiri and Nakane 2009), training beyond the maximum cross-validation performance is not recommended (Srivastava et al. 2006). I utilized the commonly used, feed forward, back propagation algorithm. Three models with their model architectures, hidden neurons, input layers, and datasets are listed in Table 5.3. Similarly, the model outputs at each stage of training, validation and testing are reported in Table 5.4.

5.6.3 Effect of Climate Variability Caused by ENSO

In order to evaluate the effect of climate variability, the model was reconstructed by utilizing temperature and precipitation as model inputs and eliminating ENSO indices from the input space. As the input parameters are reduced in the input vectors, the same model architecture was not appropriate. For this, I redesigned the model architecture and determined the proper number of hidden neurons using the same methodology discussed earlier. The model performance in JFM, including the ENSO indices in the input vector, is significantly better than the model performance when ENSO indices were not used from the input space (Table 5.4). This is consistent with the preliminary assessment I did using the MLR model (Table 5.1). In the next step, I tested the ANN model performance in the AMJ season by applying the same approach that I discussed for JFM. The model performance in this season, without considering the ENSO indices in the input space, is slightly better than the model with ENSO indices in its input space, especially in the testing stage (Table 5.4). This suggests that the effect of ENSO is relatively weak in this season compared to the JFM season. The ANN models were developed for the ASO season as well as neutral periods. The model performance in different stages of training,

validation and testing are tabulated in Table 5.4. The statistical parameters measuring the performance of the model indicated that the model was adequately simulating TOC loads after explicit incorporation of ENSO information into the data-driven models. Since the parameter significance test using the MLR model suggested that ENSO did not depict a significant influence on TOC loads in the ASO season and Neutral period, it was not useful to conduct a model comparison without ENSO information in these seasons. It is noteworthy to conclude that climate variability caused by ENSO has a strong influence on TOC loads in JFM, and some influence in AMJ. This implies that temperature and precipitation alone cannot adequately simulate pollutant transport processes in the coastal areas of the Southeast USA. Incorporating ENSO indices explicitly into data-driven models can capture the dominating influence of ENSO in model simulations, which is particularly true for a period (JFM and AMJ) when the ENSO signal has a better signature on TOC loads. This may not be true for the ASO season as the consistent effect of ENSO on TOC loads was not observed during this period.

5.6.4 Fuzzy Logic Modeling

For the fuzzy logic models, I used rainfall, temperature and ENSO information as inputs. Inputs were fuzzified into fuzzy subsets, and fuzzy rules were developed based on 55 years of historical data and expert knowledge. Model complexity increases for each increase in new inputs due to the increase in the number of rules to be addressed (Mahabir et al. 2003). Hence, only the parameters that were significant at $p \leq 0.05$ were considered as model inputs.

Fuzzy logic toolbox in MATLAB platform R2009A (version 7.8) was chosen for the fuzzy logic expert system. The input variables were combined using the AND function and the

fuzzy operator “minimum.” For defuzzification, the centroid (center of gravity) method was applied.

5.6.5 Fuzzy Logic Model Calibration, Validation and Testing

The precipitation, ENSO phase, and TOC loads were fuzzified into (fuzzy) subsets based on the available data sets. The fuzzy subsets and the membership functions (MF’s) are shown in Table 5.5 and Figure 5.6, respectively. Also, the range of precipitation, temperature and TOC load for different subsets are shown in Figure 5.6. The input and output vectors were fuzzified with many smaller subsets because better accuracy can be expected with more subsets. In the JFM-fuzzy logic model, nine linguistic terms were selected to address the input and output variables (Table 5.5).

Since the ENSO characteristics and pattern are not identical in the JFM and AMJ seasons, fuzzy partitions could be different for different seasons. Therefore, the TOC load and PCP was again fuzzified intuitively with the triangular membership function for AMJ (Table 5.5). Fuzzy rules were applied for input and output with possible fuzzy inference procedures in a descriptive way using If-Then format. I devised different fuzzy rules for the same range of precipitation resulting in different ranges of TOC loads for the La Niña and El Niño phases. This is due to the difference in precipitation characteristics in different phases of ENSO. Since JFM encompasses three months, the TOC load in March is somehow related to the precipitation characteristics of January and February due to a lagging effect. The observed data also exhibits a higher TOC load in the El Niño phase than in the La Niña phase for the same range of precipitation. This is also true for AMJ season. The same data sets used for ANN model training, validation and testing were employed for fuzzy logic model construction using Mamdani fuzzy

rules (Mamdani 1974). The calibrated MF's for the JFM- fuzzy logic- as well as the AMJ- fuzzy logic model are shown in Figure 5.6.

The fuzzy logic model was applied to both the ASO season and neutral period. The clear pattern of TOC load with ENSO phases was not observed during the ASO season (Figure 5.5). Instead, it illustrated a great deal of variability. The TOC load exhibited greater association in the La Niña phase, but this finding was not consistent. For the same range of precipitation during this season, the TOC load was different, indicating a strong association with temperature. Therefore, taking into account the parameter significance test performed using the MLR model (p -value < 0.001 for temperature and > 0.1 for ENSO indices), the ASO model was developed based on temperature, precipitation and a stronger ENSO signal. Similarly, for the neutral model, only temperature and precipitation ($p < 0.05$) were utilized because ENSO indices were not significant ($p > 0.10$).

The fuzzy MF's for precipitation, TOC, ENSO, and temperature for this season are shown in Figure 5.6. Twenty seven rules were devised for nine different ranges of precipitation with three different ranges of temperature (not shown). The calibrated fuzzy logic model was applied using an independent time period to validate the model for the validation data sets. The model was applied to predict the TOC load in the testing phase to test whether the membership function calibrated produced reasonable results in another time period. Model testing ensured that the model could be applied to a wide range of conditions. Calibration, validation and testing of the fuzzy logic models (JFM, AMJ, ASO and neutral), and their respective model performance criteria at each stage are tabulated in Table 5.6.

5.6.6 Comparison of PCR-ANN and Fuzzy Logic Models

The model performance of the PCR- ANN model and fuzzy expert systems for all phases of training, validation and testing are shown in the Table 5.4 and Table 5.6, respectively. The overall performance of the PCR-ANN and fuzzy logic models are satisfactory. However, the fuzzy logic model indicated sudden failure for particular data sets when sufficient rules were not incorporated. Hence, the performance of fuzzy logic model is not consistent in all stages (Table 5.6). The PCR-ANN model had better model performance in all stages of model training, validation and testing.

The two models have different ways of mapping the input-output relationship. The PCR-ANN model uses the numerical weight, which the model determines based on the nature, pattern, and characteristics of the data, whereas, the fuzzy logic model establishes the relationship between input and output through categorization of the data sets into “low,” “medium,” and “high.” The PCR-ANN model was sensitive to the magnitude of the ENSO indices (large negative value, i.e. -2.1 to large positive value, i.e., 2.5), and produced the output with respect to its magnitude. However, the generic ENSO information (El Niño and La Niña information irrespective of the indices’ values) were fed into the fuzzy logic model with a defined set of different rules for El Niño and La Niña. The input data sets for the fuzzy logic were decided after assessing the parameter significance using the MLR model ($p < 0.05$). A suitable model was constructed through the trial and error procedure.

The fuzzy logic approach can be better for predicting TOC loads than PCR-ANN when input variables are limited. Fuzzy logic algorithm may not be a better choice for a large number of input variables due to the difficulty in establishing the rule base. It becomes much more complicated than PCR-ANN for a large number of input variables. This was observed in the

model development for the ASO season in which, stronger ENSO events were incorporated besides temperature and precipitation.

Both PCR-ANN and fuzzy logic models performed well over the period of model training, validation, and testing. The PCR-ANN model adequately simulated with consistent and relatively better result than fuzzy logic model in different stages of modeling. The statistical parameters measuring the performance of the models in different seasons for PCR-ANN and fuzzy logic models are presented in table 5.4(a) and Table 5.6, respectively.

5.6.7 TOC Load Forecast

Since fuzzy logic can handle the uncertainties associated with the forecast, I used this approach instead of PCR-ANN model for forecasting TOC load at one month lead time. I used the forecasted precipitation and temperature data from the Reanalysis and Reforecast of CFSv2 model, and ENSO information from Climatic Prediction Center (CPC) of National Weather Service (NWS). CFSv2 model is already operational at NCEP and climate data are available at 0.9 degree spatial resolution. The readers can find details about CFSv2 data in an article by Saha et al. (2006).

The daily forecast of precipitation and temperature data were downscaled at local grid. In the next step, the mean precipitation and temperature from the set of 24 ensemble family of precipitation and temperature were applied in fuzzy logic model to forecast the TOC load and compare with the actual observed data during 3 years (1992 to 1994) of post validation period. The selection of the period is based on the available observed data sets. The TOC load was forecasted both qualitatively and quantitatively. However, the quantitative forecast of TOC load was not satisfactory. In fact, the deterministic forecast of TOC load is still challenging and not

recommended in a context that the hydrologic forecasting skill has met several challenges and not fully developed yet. The qualitative forecast of TOC load has been tabulated in Table 5.7, which indicates that 80% forecast are in consistent with the range of the actual observed data.

5.7 Summary and Concluding Remarks

In this study, two data-driven models, PCR-ANN and fuzzy logic, were developed. The PCR-ANN model was developed to predict the TOC load using real time data at different ENSO phases and the fuzzy logic model was developed to forecast the TOC loads at one month lead time. The objectives of the study were to quantify the effect of climate variability on TOC loads and develop a simple TOC prediction model for different ENSO phases of each season. Simulated long-term TOC data sets were used to quantify the effect of ENSO on TOC load. This was accomplished with a stable land use pattern to differentiate the effect of ENSO from the land use effect.

Fuzzy logic model has a tendency to show the sudden failure for specific data sets if the sufficient rules are not developed. In order to generalize the fuzzy logic model for a wide range of data sets, several combinations of natural processes are to be included in the model training stage. In addition, it can be fairly concluded that MF's other than the triangular function might be appropriate for the study in which too many inputs are involved. The benefit of the fuzzy logic model is that the model results can be applied to the similar ENSO-affected region on a qualitative scale, which avoids the necessity of site specific parameters.

For the real time prediction, PCR-ANN was found to be more reliable than fuzzy logic as it incorporates different architectures to train the input vectors. The PCR-ANN model fitted the data very well with negligible mass balance error. However, the fuzzy logic model is better

choice, especially, for forecasting purpose as it can translate CFSv2 climate forecast into a quantitative and qualitative TOC load forecast. In brief, there are three clear findings of the study: (1) the data-driven model using multiple ENSO indices (Niño 3.4 indices and TNI), temperature and precipitation can adequately predict the TOC load, and is also one of the new approaches in water quality modeling in the ENSO affected region; (2) the model performance will be enhanced if I include multiple ENSO indices in the input vectors, which is particularly true for the season, when the TOC load is correlated with ENSO phase; and (3) PCR-ANN is a better model for real time prediction and fuzzy logic approach is better for forecasting TOC load at qualitative scale.

Both approaches seem to be strongly relevant for TOC simulation with these particular inputs because the CFS forecasts maintain a high level of forecasting skills for ENSO indices, temperature and precipitation on monthly and seasonal scales. Predicting the seasonal fluctuation in TOC load in different ENSO phases will be beneficial for the minimization of additional cost required to treat the drinking water for TOC before chlorination.

5.8 Acknowledgement

The authors wish to acknowledge the funding provided by the NOAA-RISA and SARP program for this work. The authors would also like to acknowledge Dr. Sabahattin Isik for his help in developing the Artificial Neural Network model.

Table 5.1. Multi Linear Regression model performance with and without using ENSO information

With ENSO				Without ENSO			
JFM model							
Predictors	P-value	R-square	Adj. R-square	Predictors	P-value	R-square	Adj. R-square
Temp.	0.8	0.7	0.69	Temp.	0.174	0.57	0.56
PCP	0			PCP	0		
Niño 3.4	0						
Trans Niño	0						
AMJ model							
Temp.	0	0.65	0.63	Temp.	0	0.63	0.62
PCP	0			PCP	0		
Niño 3.4	0.05						
TNI	0.006						
ASO model							
Temp.	0.001	0.54	0.52				
PCP	0						
Niño 3.4	0.954						
TNI	0.202						
Neutral							
Temp.	0	0.58	0.57				
PCP	0						
Niño 3.4	0.234						
TNI	0.213						

Note: Temp. and PCP denote temperature and precipitation respectively. In Neutral phase and ASO season, there is no comparison as ENSO indices were not significant.

Table 5.2. Hidden neuron selection for the model

Season	Input Layer	With ENSO			Input Layer	Without ENSO		
		NHN [†]	NSE	AIC		NHN [†]	NSE	AIC
JFM	4	9	0.67	11.85	2	10	0.47	11.99
AMJ	4	6	0.78	10.85	2	7	0.82	10.13
ASO	4	10	0.66	11.34	NA	NA	NA	NA
Neutral	4	8	0.68	10.94	NA	NA	NA	NA

Note: NHN[†] implies number of hidden neurons

Table 5.3. The architecture of the ANN model used in this study

ANN Model	Input layers		Hidden neurons		Training data sets	Validation data sets	Testing data sets	Output neuron
	With ENSO	Without ENSO	With ENSO	Without ENSO				
JFM	4	2	9	10	51	17	17	1
AMJ	4	2	6	7	46	14	14	1
ASO	4	2	10	NA	52	17	18	1
Neutral	4	2	8	NA	198	65	64	1

Table 5.4. The PCR-ANN model performance in different stage: a) PCR-ANN model with Niño 3.4 index and TNI, and b) PCR-ANN model without Niño 3.4 index and TNI

(a)

Season	Training			Validation			Testing		
	<i>NSE</i>	<i>MBE</i>	R^2	<i>NSE</i>	<i>MBE</i>	R^2	<i>NSE</i>	<i>MBE</i>	R^2
JFM	0.93	-1.1%	0.94	0.74	5.8%	0.77	0.67	1.1%	0.68
AMJ	0.86	-8.2%	0.89	0.76	11.9%	0.83	0.64	-13.1%	0.67
ASO	0.34	-12.0%	0.38	0.61	5.4%	0.62	0.71	-5.4%	0.82
Neutral	0.67	-8.0%	0.69	0.69	0.0%	0.69	0.64	-3.0%	0.66

(b)

Model	Training				Validation				Testing			
	<i>NSE</i>	<i>MBE</i>	R^2	<i>AIC</i>	<i>NSE</i>	<i>MBE</i>	R^2	<i>AIC</i>	<i>NSE</i>	<i>MBE</i>	R^2	<i>AIC</i>
JFM	0.69	-2.0%	0.69	11.13	0.42	10.0%	0.44	12.6	0.66	14.0%	0.74	12.9
AMJ	0.84	2.0%	0.84	11.55	0.75	8.0%	0.77	11.13	0.57	4.0%	0.77	12.49
ASO		NA				NA				NA		
Neutral		NA				NA				NA		

Table 5.5. Rules used for fuzzy logic model, (a) JFM and AMJ, and (b) ASO

(a)				
ENSO Phase	JFM		AMJ	
	PCP	TOC load	PCP	TOC load
La Niña	VVL	VL	VVL	VL
	VL	VL	VL	VL
	L	L	L	VL
	M	M	M	VL
	MH	MH	MH	VL
	H	MH	H	MH
	VH	H	VH	MH
	VVH	VVH	VVH	MH
El Niño	VVL	L	VL	VL
	L	M	L	L
	M	M	M	M
	MH	H	MH	M
	H	VVH	H	MH
	VH	VVH	VH	VH
	VVH	EH	VVH	VVH
(b)				
ENSO	PCP	TOC load		
Strong La Niña	VVL	M		
Strong La Niña	VL	MH		
Strong La Niña	L	H		
Strong La Niña	M	H		
Strong EL Niño	VL	VVL		
Strong EL Niño	L	VL		
Strong EL Niño	M	L		
<i>For any phase of ENSO</i>				
Temperature	PCP	TOC load		
VH	M	M		
VH	H	H		
<i>For Any temperature (except strong ENSO events)</i>				
Temperature	PCP	TOC load		
NA	VL	L		
	L	M		
	M	MH		
	VH	VH		
	VVH	VVH		
	EH	EH		

Table 5.6. Fuzzy logic model calibration, validation and testing

Model	Calibration			Validation			Testing		
	R^2	NSE	MBE	R^2	NSE	MBE	R^2	NSE	MBE
JFM	0.58	0.57	-2.0%	0.47	0.46	7.1%	0.78	0.75	-9.7%
AMJ	0.78	0.74	-4.0%	0.46	0.32	25.1%	0.85	0.85	11.9%
ASO	0.38	-0.05	0.7%	0.61	0.45	31.5%	0.82	0.77	20.1%
Neutral	0.64	0.63	-0.0%	0.67	0.65	7.4%	0.69	0.66	8.9%

Table 5.7. Post validation of the qualitative TOC load forecast

Date	Observed	Forecasted	Date	Observed	Forecasted
1/1/1999	MH	M*	7/1/2000	VVL	VVL
2/1/1999	M	M	8/1/2000	VVL	VVL
3/1/1999	M	M	9/1/2000	VVL	VVL
4/1/1999	VL	VL	10/1/2000	VVL	VVL
5/1/1999	VL	VL	11/1/2000	VL	VL
6/1/1999	L	L	12/1/2000	VL	L*
7/1/1999	VL	VL	1/1/2001	L	L
8/1/1999	VVL	VVL	2/1/2001	L	L
9/1/1999	VVL	VVL	3/1/2001	VH	VH
10/1/1999	VL	VVL*	4/1/2001	L	L
11/1/1999	VVL	VVL	5/1/2001	VVL	VVL
12/1/1999	VL	VL	6/1/2001	VL	VL
1/1/2000	VL	VL	7/1/2001	VVL	VL*
2/1/2000	L	L	8/1/2001	L	L
3/1/2000	L	L	9/1/2001	VL	VL
4/1/2000	VVL	VVL	10/1/2001	VVL	VL*
5/1/2000	VVL	VVL	11/1/2001	VVL	VL*
6/1/2000	VVL	VL*			

* Represents the qualitative forecast mismatched with qualitative range of observed data

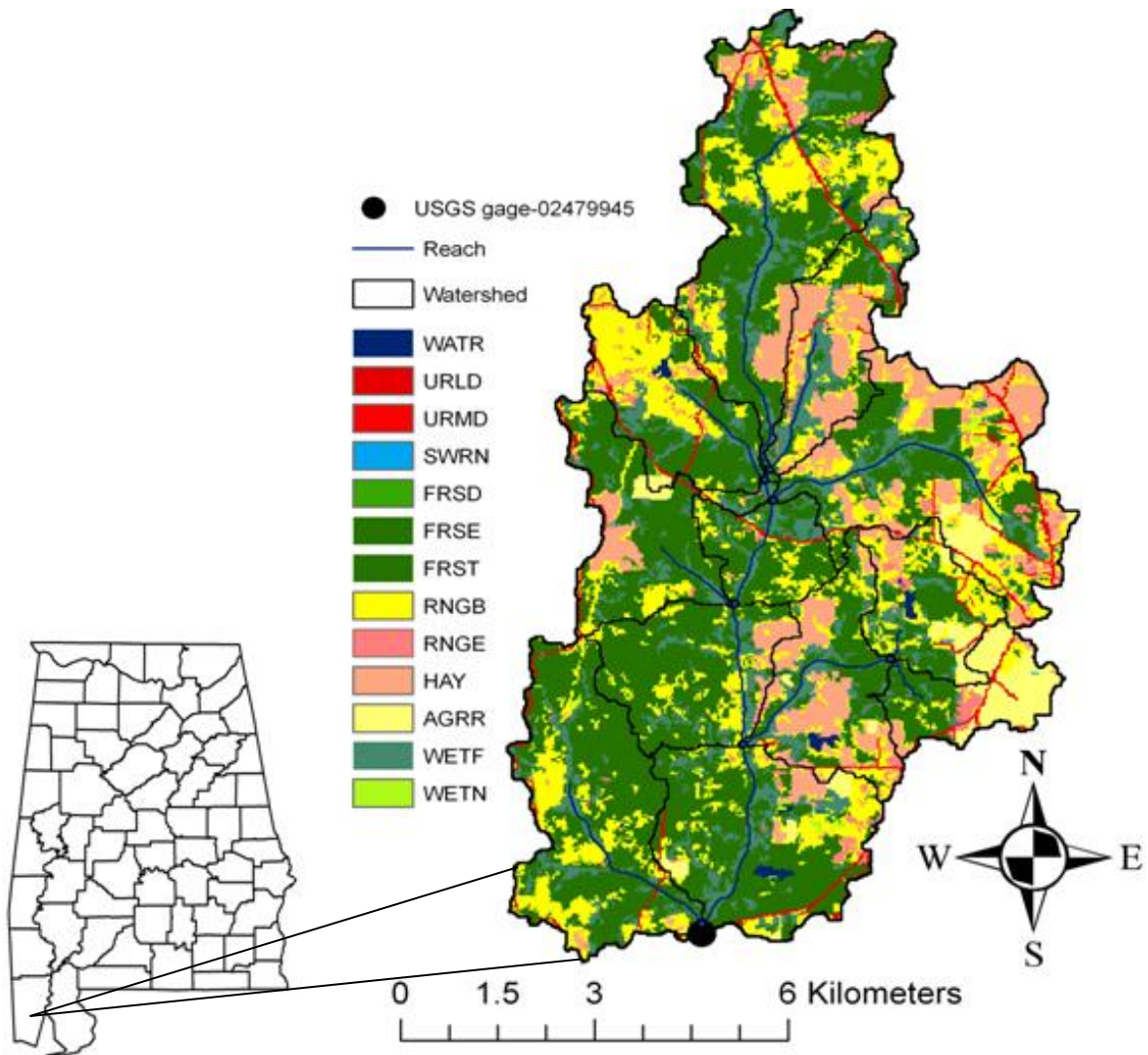


Figure 5.1. Map of the study area showing the land use distribution in different sub-basins and USGS gauging station in the Big Creek watershed.

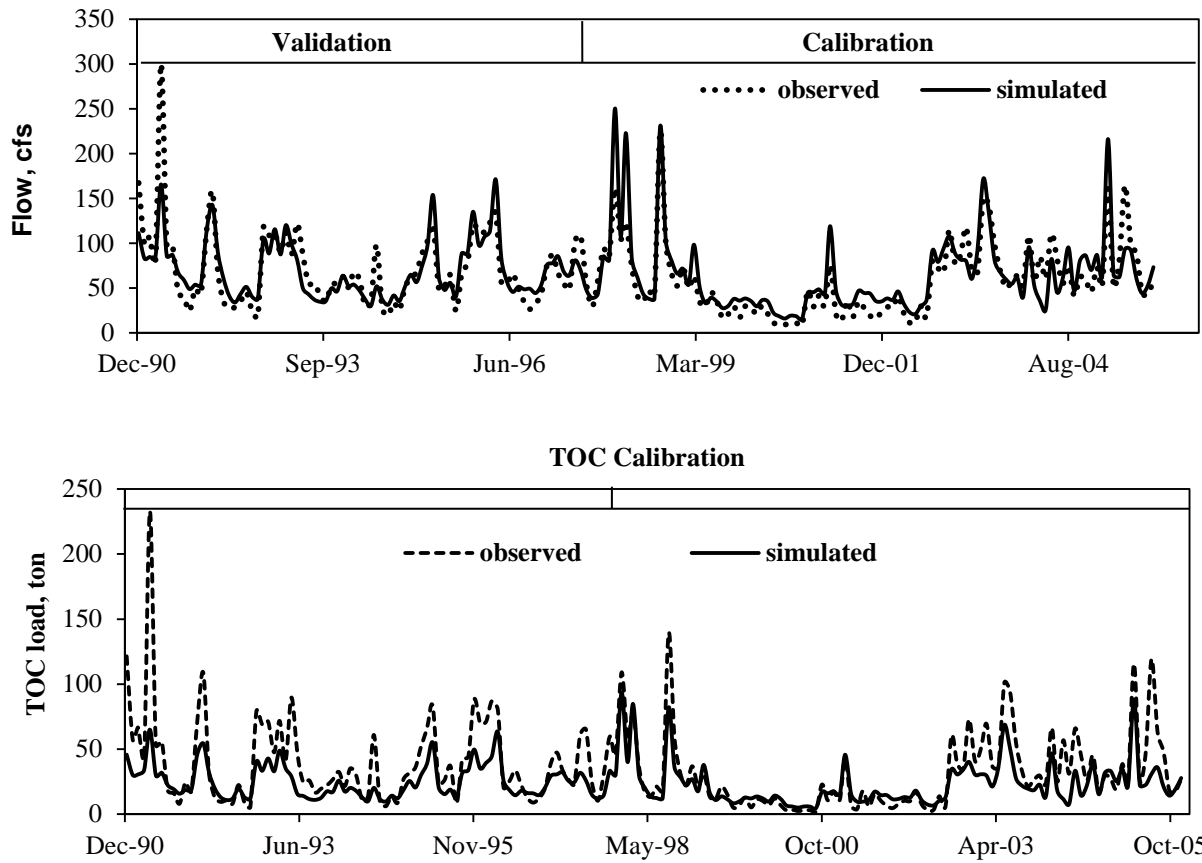


Figure 5.2. Calibration and validation of a) monthly streamflow, and b) monthly TOC loads generated by the LSPC model.

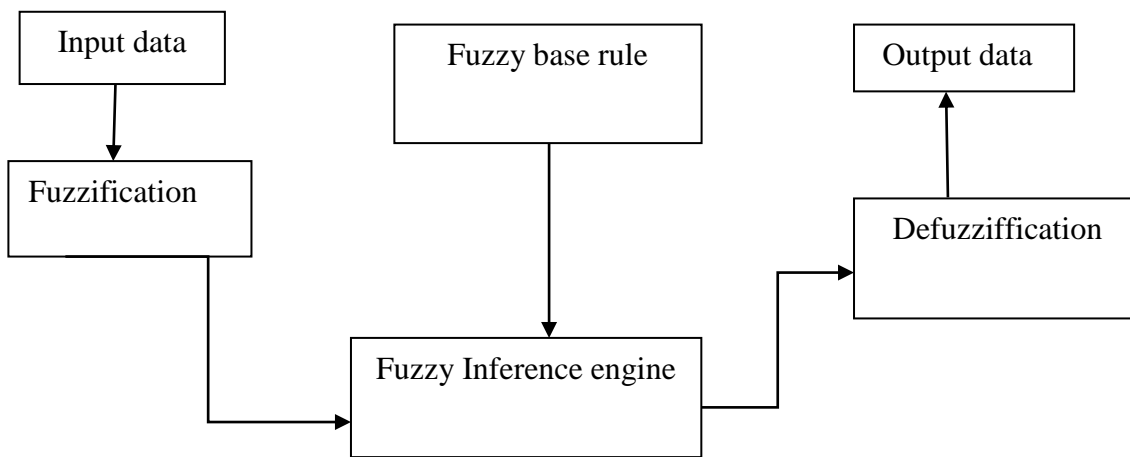


Figure 5.3. Schematic representation of Fuzzy inference system used in this study.

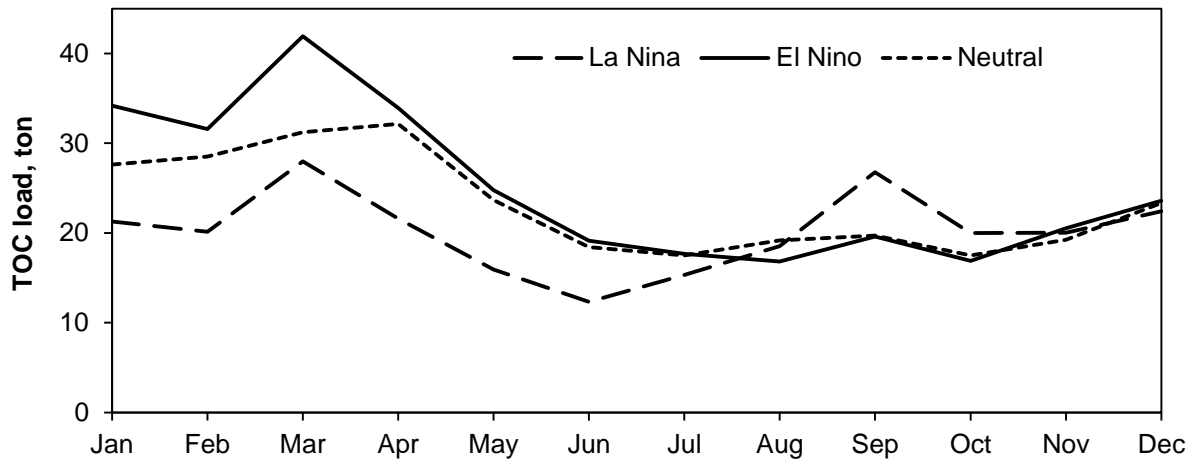
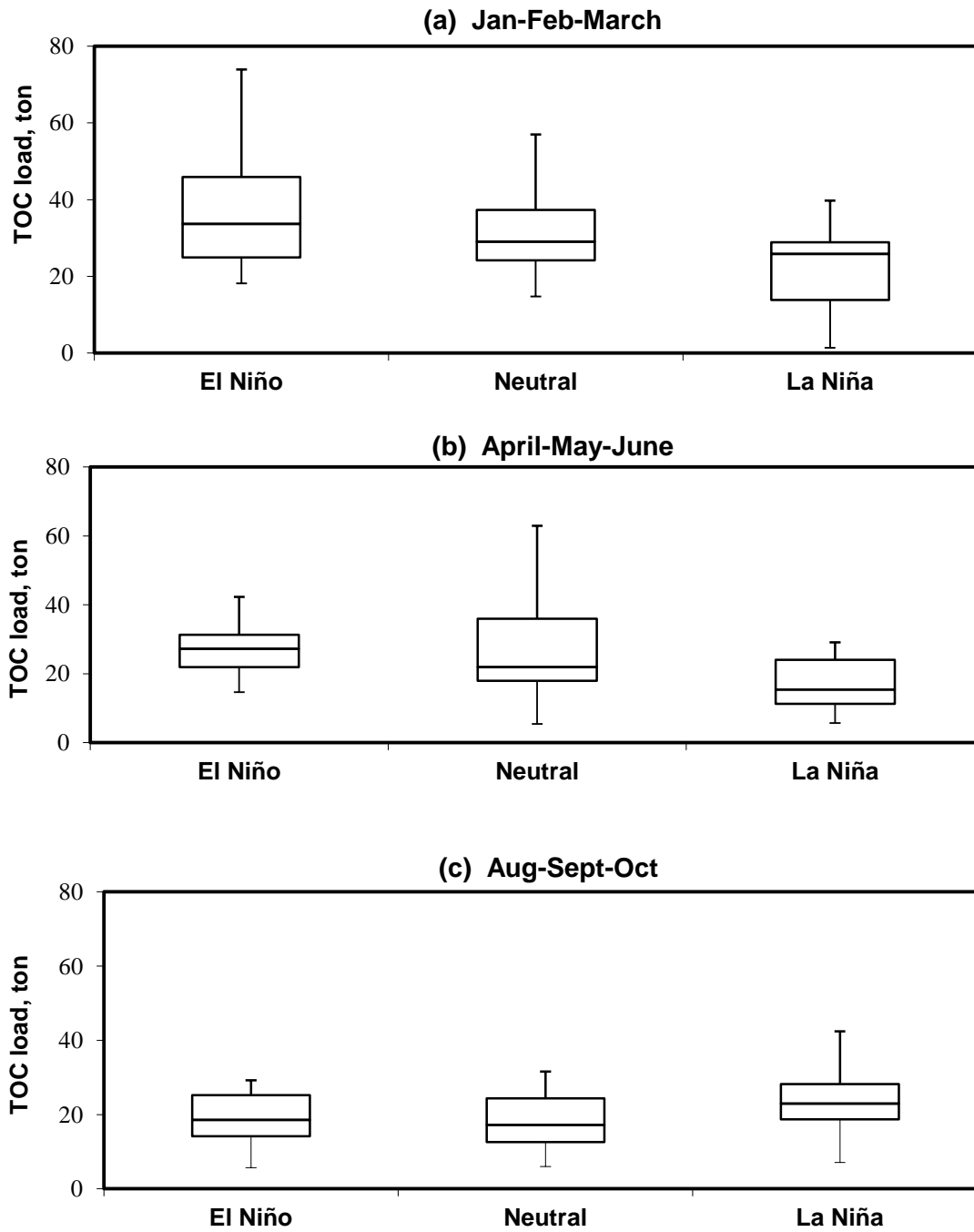


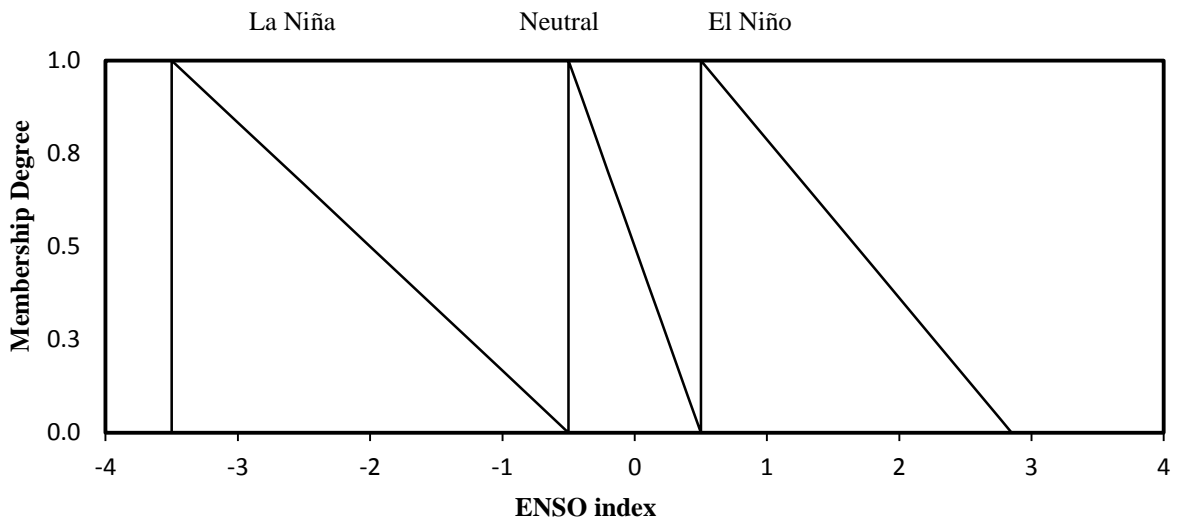
Figure 5.4. Monthly variation in average TOC loads (averaged over 55 years) in different ENSO phases.



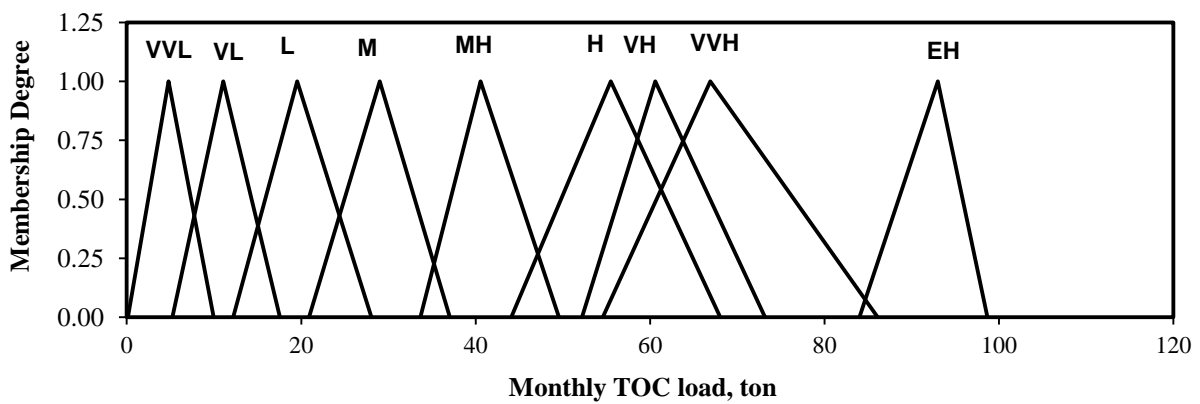
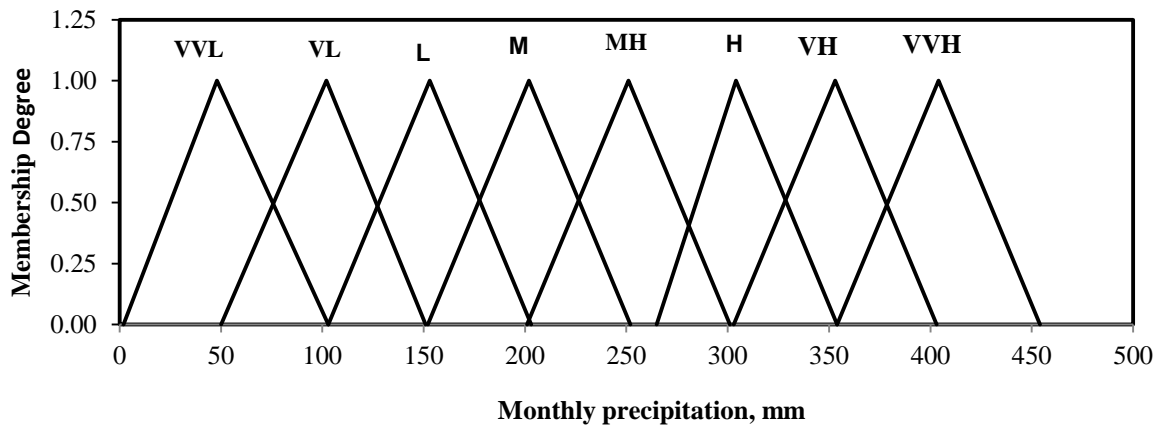
(c)

Figure 5.5. Box plot showing total monthly average TOC loads (ton) for (a) January-February-March (JFM) season, (b) April-May-June (AMJ) season, and (c) Aug-Sept-Oct (ASO) season in different phases of ENSO.

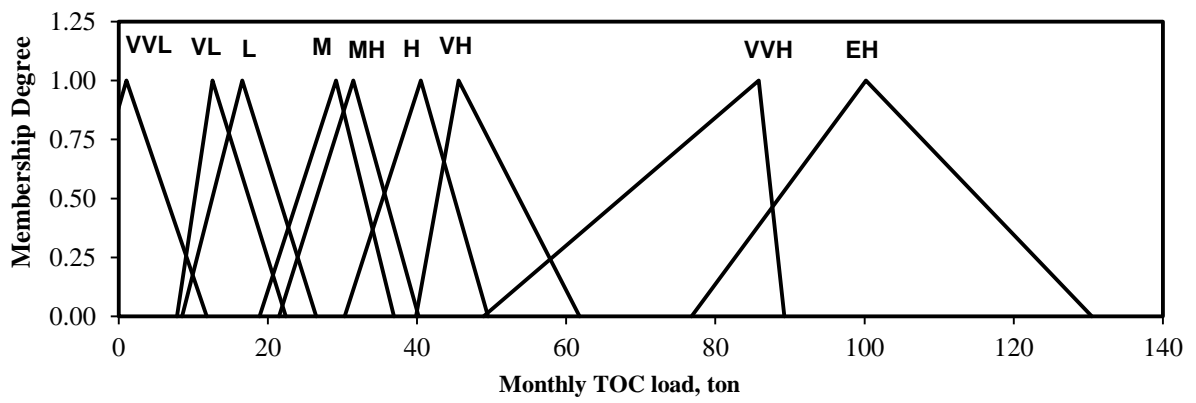
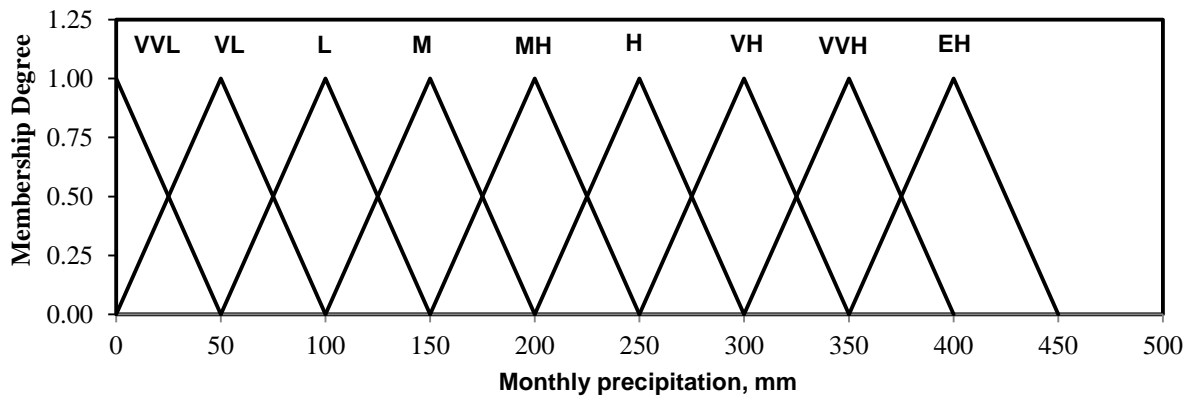
Note: Load is “total monthly load (ton)” averaged over a season.



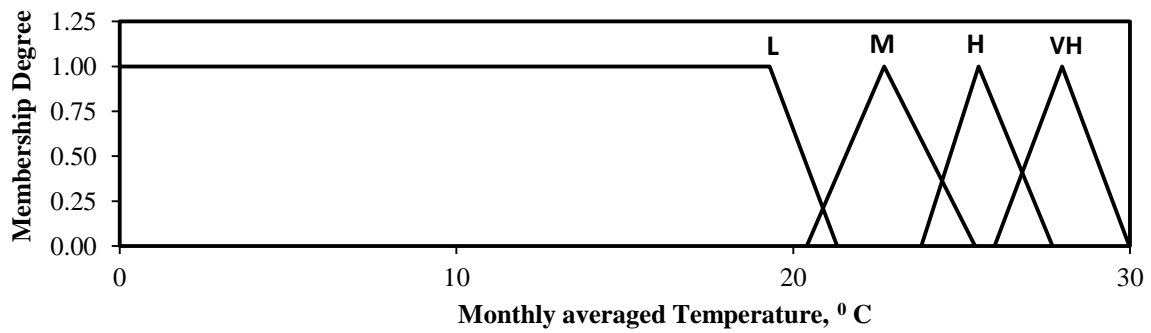
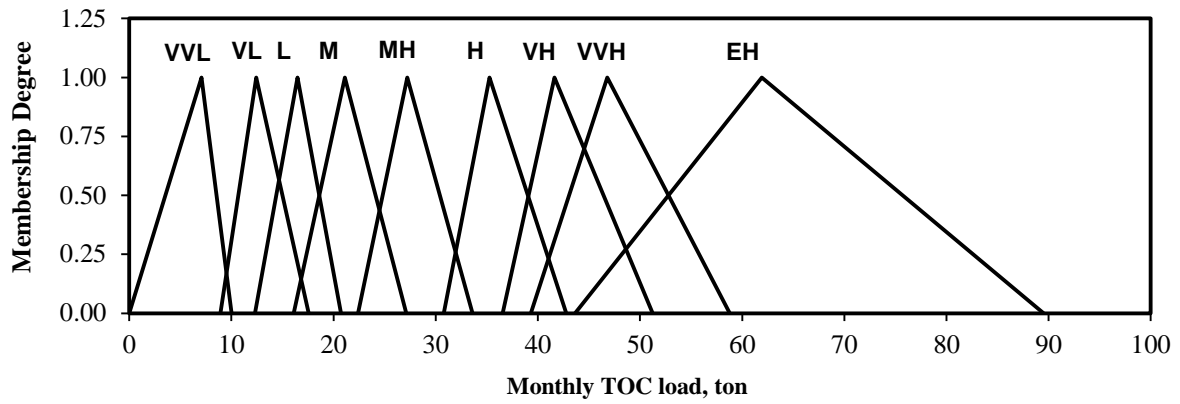
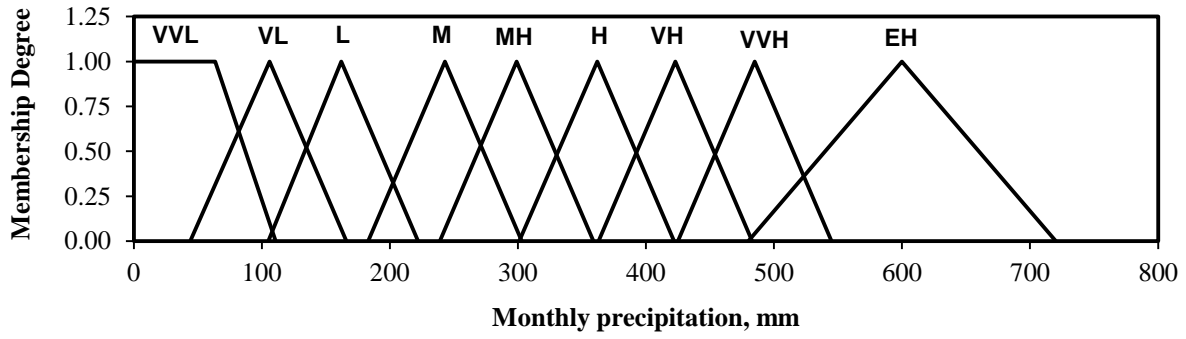
(a)



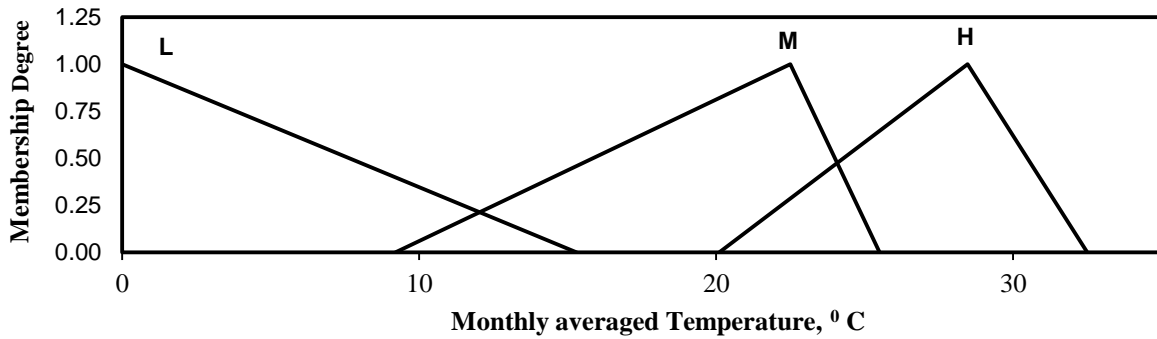
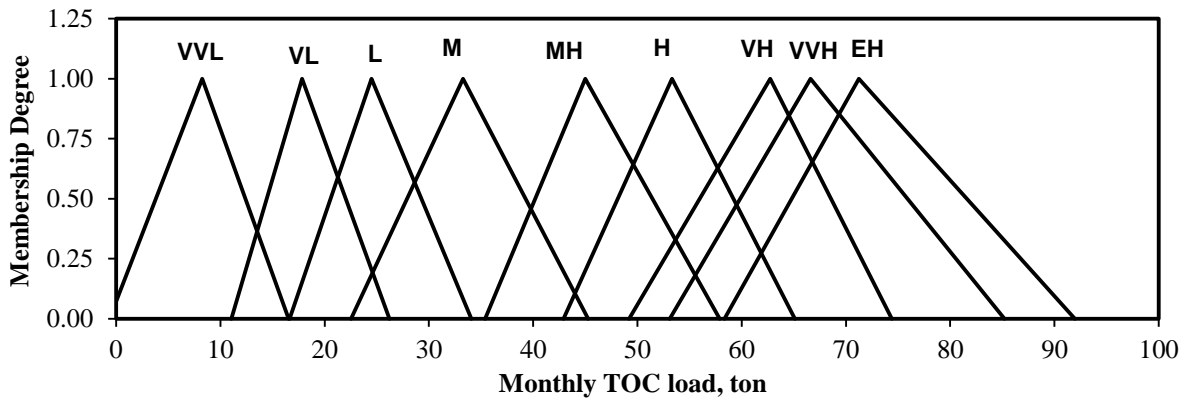
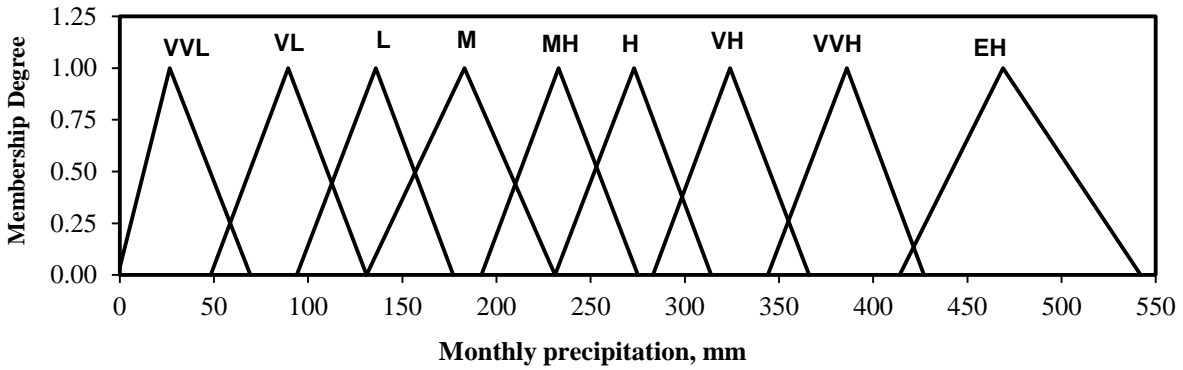
(b) (Above two panels)



(c) (Above two panels)



(d) (Above three panels)



(e) (Above three panels)

Figure 5.6. Calibrated membership function for the fuzzy logic model for precipitation, ENSO phase, and TOC load: (a) MF's for ENSO (For all seasons), (b) MF's for JFM (precipitation, TOC), (c) MF's for AMJ season (precipitation, TOC) (d) MF's for ASO season (PCP, TOC temperature), (e) MF's for neutral season (precipitation, TOC, temperature). L: "low", M: "medium", H: "High", V: "Very", E: "Exceptional".

5.9 References

- Abedi-koupai, J., Amiri M. J., Eslamian, S. S. (2009). "Comparison of artificial neural network and physically based models for estimating of reference evapotranspiration in greenhouse." *Australian Journal of Basic and Applied Sciences*, 3(3): 2528-2535.
- Abudu, S., King, J. P., and Bawazir, A. S. (2011). "Forecasting monthly streamflow of spring-summer runoff season in Rio Grande headwaters Basin using stochastic hybrid modeling approach." *Journal of Hydrologic Engineering*, 16, 384.
- Aiken, G. R., Cotsaris, E., (1995). "Soil and hydrology: their effect on NOM". *Journal of the American Water Works Association*, 87, 36–45. Analysis. Prentice-Hall, Englewood Cliffs.
- Al-Alawi, S. M., Abdul-Wahab, S. A., and Bakheit, C. S. (2008). "Combining principal component regression and artificial neural networks for more accurate predictions of ground-level ozone." *Environmental Modelling & Software*, 23(4), 396-403.
- Amiri, B., and Nakane, K. (2009). "Comparative prediction of stream water total nitrogen from land cover using artificial neural network and multiple linear regression approaches." *Polish Journal of Environmental Studies*, 18(2), 151-160.
- Araghinejad, S., Burn, D. H., and Karamouz, M. (2006). "Long-lead probabilistic forecasting of streamflow using ocean-atmospheric and hydrological predictors." *Water Resources Research*, 42(3), W03431.
- ASCE (2000a). "Artificial neural networks in hydrology. I: Preliminary concepts." *Journal of Hydrologic Engineering*, 5(2), 115-123.
- Awadallah, A., and Rousselle, J. (2000). "Improving forecasts of Nile flood using SST inputs in TFN model." *Journal of Hydrologic Engineering*, 5, 371.
- Barsugli, J. J., Whitaker, J. S., Lough, A. F., Sardeshmukh, P. D., and Toth, Z. (1999). "The effect of the 1997/98 El Niño on individual large-scale weather events." *Bull. Amer. Meteor. Soc.*, 80, 1399–1411.
- Bowden, R., Windridge, D., Kadir, T., Zisserman, A., and Brady, M. (2004). "A linguistic feature vector for the visual interpretation of sign language." *Computer Vision-ECCV 2004*, 390-401.
- Canale, R. P., Chapra, S. C., and Amy, G. L. (1997). "Trihalomethane precursor model for lake Youngs, Washington." *Journal of water resources planning and management*, 123, 259.
- Chang, H., and Carlson, T. N. (2005). "Water quality during winter storm events in Spring Creek, Pennsylvania USA." *Hydrobiologia*, 544(1), 321-332.

- Chang, L. C., Chang, F. J., and Tsai, Y. H. (2005). "Fuzzy exemplar-based inference system for flood forecasting." *Water Resources Research*, 41(2), W02005.
- Chao, Y., and Philander, S. (1993). "On the structure of the Southern Oscillation." *Journal of Climate*, 6(3), 450-469.
- Chau, K., Wu, C., and Li, Y. (2005). "Comparison of several flood forecasting models in Yangtze River." *Journal of Hydrologic Engineering*, 10, 485.
- Chiew, F., Piechota, T., Dracup, J., and McMahon, T. (1998). "El Niño/Southern Oscillation and Australian rainfall, streamflow and drought: Links and potential for forecasting." *Journal of Hydrology*, 204(1-4), 138-149.
- Childers, J.M. 2009. LSPC input file for Mobile Bay modeling project – TetraTech, Inc., 2110 Powers Ferry Road, Atlanta, GA, 30339.
- Clemen, R. T. (1989). "Combining forecasts: A review and annotated bibliography." *International Journal of Forecasting*, 5(4), 559-583.
- Correll, D. L., Jordan, T. E., and Weller, D. E. (2001). "Effects of precipitation, air temperature, and land use on organic carbon discharges from Rhode River watersheds." *Water, Air, & Soil Pollution*, 128(1), 139-159.
- Cureton, E. E., & D'Agostino, R. B. (1983). *Factor analysis: An applied approach*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Deka, P. C., and Chandramouli, V. (2009). "Fuzzy neural network modeling of reservoir operation." *Journal of water resources planning and management*, 135, 5.
- DEM.(2010). "Digital Elevation Model." WWW data downloaded from www.seamless.usgs.gov.
- Eldaw, A. K., Salas, J. D., and Garcia, L. A. (2003). "Long-range forecasting of the Nile River flows using climatic forcing." *Journal of Applied Meteorology*, 42(7), 890-904.
- Elias, E. H. (2010). "Valuing ecosystem services from forested landscapes: How urbanization influences drinking water treatment cost." Auburn University.
- Elias, E., Dougherty, M., Srivastava, P., and Laband, D. (2011). "The impact of forest to urban land conversion on streamflow, total nitrogen, total phosphorus, and total organic carbon inputs to the converse reservoir, Southern Alabama, USA." *Urban Ecosystems*, 1-29.
- Fletcher Ernie, D. (1993). "Forecasting with neural networks:: An application using bankruptcy data." *Information & Management*, 24(3), 159-167.
- Fritts, H.C. (1991). "reconstruting large-scale climatic patterns from tree-ring data": A diagnostic analysis, Univ. of Ariz. Press, Tucson.

- Gim�iz-Fortis, S., Esteban-Parra, M., Trigo, R., and Castro-D ez, Y. (2010). "Potential predictability of an Iberian river flow based on its relationship with previous winter global SST." *Journal of Hydrology*, 385(1-4), 143-149.
- Gergel, S. E., Turner, M. G., and Kratz, T. K. (1999). "Dissolved organic carbon as an indicator of the scale of watershed influence on lakes and rivers." *Ecological Applications*, 9(4), 1377-1390.
- Govindaraju, R. S., and Rao, A. R. (2000). *Artificial neural networks in hydrology*, Springer Netherlands.
- Grantz, K., Rajagopalan, B., Clark, M., and Zagona, E. (2005). "A technique for incorporating large-scale climate information in basin-scale ensemble streamflow forecasts." *Water Resources Research*, 41(10), W10410.
- Hamlet, A. F., and Lettenmaier, D. P. (1999). "Columbia River streamflow forecasting based on ENSO and PDO climate signals." *Journal of water resources planning and management*, 125(6), 333-341.
- Handler, A. (1990). "USA corn yields, the El Niño and agricultural drought: 1867–1988." *International Journal of Climatology*, 10(8), 819-828.
- Hanley, D. E., Bourassa, M. A., O'Brien, J. J., Smith, S. R., and Spade, E. R. (2003). "A quantitative evaluation of ENSO indices." *Journal of Climate*, 16(8), 1249-1258.
- Hansen, D. V., and Maul, G. A. (1991). "Anticyclonic current rings in the eastern tropical Pacific Ocean." *Journal of Geophysical Research*, 96(C4), 6965-6979.
- Hansen, J., Jones, J., Irmak, A., and Royce, F. (2001). "El Niño-southern oscillation impacts on crop production in the southeast United States." *ASA Special Publication*, 63, 55-76.
- Henry, T., Shen, J., Lahlou, M., Shoemaker, S., Parker, A., Ouyang, J. Yang, Ludwig, H.J. (2002a). "Mining data analysis system (MDAS)." Watershed 2002, *WEF Specialty Conference Proceedings on CDROM, Fort Lauderdale*.
- Hidalgo, H. G., Piechota, T. C., and Dracup, J. A. (2000). "Alternative principal components regression procedures for dendrohydrologic reconstructions." *Water Resources Research*, 36(11), 3241-3249.
- Hundecha, Y., Bardossy, A., and WERNER, H. W. (2001). "Development of a fuzzy logic-based rainfall-runoff model." *Hydrological Sciences Journal*, 46(3), 363-376.
- Jain, A., and Indurthy, S. K. V. P. (2003). "Comparative analysis of event-based rainfall-runoff modeling techniques—deterministic, statistical, and artificial neural networks." *Journal of Hydrologic Engineering*, 8, 93.

- Jantzen, J., Verbruggen, H., and Østergaard, J. J. (1999). "Fuzzy control in the process industry: Common practice and challenging perspectives." *Practical Applications of Fuzzy Technologies*, 3.
- Jennrich, R.I. (1995). *An Introduction to Computational Statistics; Regression analysis*. Inglewood Cliffs, NJ: Prentice Hall.
- Kalin, L., and Hantush, M. M. (2006). "Hydrologic modeling of an eastern Pennsylvania watershed with NEXRAD and rain gauge data." *Journal of Hydrologic Engineering*, 11, 555.
- Kalin, L., Isik, S., Schoonover, J. E., and Lockaby, B. G. (2010). "Predicting water quality in unmonitored watersheds using artificial neural networks." *Journal of environmental quality*, 39(4), 1429-1440.
- Kalra, A., and Ahmad, S. (2009). "Using oceanic-atmospheric oscillations for long lead time streamflow forecasting." *Water Resources Research*, 45(3), W03413.
- Karamouz, M., and Zahraie, B. (2004). "Seasonal streamflow forecasting using snow budget and El Niño-southern oscillation climate signals: Application to the Salt River Basin in Arizona." *Journal of Hydrologic Engineering*, 9, 523.
- Keener, V., Ingram, K., Jacobson, B., and Jones, J. (2007). "Effects of El Niño/Southern Oscillation on simulated phosphorus loading in South Florida." *Transactions of the ASAE*, 50(6), 2081-2089.
- Kisi, O. (2006). "Daily pan evaporation modelling using a neuro-fuzzy computing technique." *Journal of Hydrology*, 329(3-4), 636-646.
- Kulkarni, J. (2000). "Wavelet analysis of the association between the southern oscillation and the Indian summer monsoon." *International Journal of Climatology*, 20(1), 89-104.
- Mahabir, C., Hicks, F., and Fayek, A. R. (2003). "Application of fuzzy logic to forecast seasonal runoff." *Hydrological processes*, 17(18), 3749-3762.
- Mamdani, E.H.(1974) , "Application of fuzzy algorithms for control of simple dynamic plant," *Electrical Engineers, Proceedings of the Institution of* , vol.121(12),1585-1588.
- Maskey, S., and Price, R. K.(2004) "Assessment of uncertainty in flood forecasting using probabilistic and fuzzy approaches." *Sixth International Conference on Hydroinformatics*, World Scientific Publishing Company, 21-24 June, Singapore.
- McCabe, G. J., and Dettinger, M. D. (1999). "Decadal variations in the strength of ENSO teleconnections with precipitation in the western United States." *International Journal of Climatology*, 19(13), 1399-1410.

- Mianaei, S. J., and Keshavarzi, A. R. (2010). "Prediction of riverine suspended sediment discharge using fuzzy logic algorithms, and some implications for estuarine settings." *Geo-Marine Letters*, 30(1), 35-45.
- Mitra, B., Scott, H., Dixon, J., and McKimney, J. (1998). "Applications of fuzzy logic to the prediction of soil erosion in a large watershed." *Geoderma*, 86(3-4), 183-209.
- Moore, T. (1989). "Dynamics of dissolved organic carbon in forested and disturbed catchments, Westland, New Zealand: 1. Maimai." *Water Resources Research*, 25(6), 1321-1330.
- Morrison, J., Colombo, M. J., District, M., and Survey, G. (2006). Surface-water quality and nutrient loads in the Nepaug Reservoir watershed, Northwestern Connecticut, 1999-2001, US Geological Survey.
- Nash, J., and Sutcliffe, J. (1970). "River flow forecasting through conceptual models part I--A discussion of principles." *Journal of Hydrology*, 10(3), 282-290.
- Özelkan, E. C., and Duckstein, L. (2000). "Multi-objective fuzzy regression: a general framework." *Computers & Operations Research*, 27(7-8), 635-652.
- Panigrahi, D., and Mujumdar, P. (2000). "Reservoir operation modelling with fuzzy logic." *Water Resources Management*, 14(2), 89-109.
- Pascual, M., Rodó, X., Ellner, S. P., Colwell, R., and Bouma, M. J. (2000). "Cholera dynamics and El Niño-southern oscillation." *Science*, 289(5485), 1766.
- Piechota, T. C., and Dracup, J. A. (1996). "Drought and regional hydrologic variation in the United States: Associations with the El Niño-Southern Oscillation." *Water Resources Research*, 32(5), 1359-1373.
- Pomes, M. L., Green, W. R., Thurman, E. M., Orem, W. H., and Lerch, H. E. (1999). "DBP formation potential of aquatic humic substances." *Journal-American Water Works Association*, 91(3), 103-115.
- Pramanik, N., and Panda, R. K. (2009). "Application of neural network and adaptive neuro-fuzzy inference systems for river flow prediction." *Hydrological Sciences Journal*, 54(2), 247-260.
- Qi, M., and Zhang, G. P. (2001). "An investigation of model selection criteria for neural network time series forecasting." *European Journal of Operational Research*, 132(3), 666-680.
- Raghuwanshi, N., Singh, R., and Reddy, L. (2006). "Runoff and sediment yield modeling using artificial neural networks: Upper Siwane River, India." *Journal of Hydrologic Engineering*, 11, 71.
- Rajagopalan, B., and Lall, U. (1998). "Interannual variability in western US precipitation." *Journal of Hydrology*, 210(1-4), 51-67.

- Reckhow, D. A., Singer, P. C., and Malcolm, R. L. (1990). "Chlorination of humic materials: byproduct formation and chemical interpretations." *Environmental science & technology*, 24(11), 1655-1664.
- Ren, L., and Zhao, Z. (2002). "An optimal neural network and concrete strength modeling." *Advances in Engineering Software*, 33(3), 117-130.
- Rogers, J. C., and Coleman, J. S. M. (2003). "Interactions between the Atlantic Multidecadal Oscillation, El Niño/La Niña, and the PNA in winter Mississippi valley streamflow." *Geophys. Res. Lett.*, 30(10), 1518.
- Roy, S. S. (2006). "The impacts of ENSO, PDO, and local SSTs on winter precipitation in India." *Physical Geography*, 27(5), 464-474.
- Runkel, R.L., Crawford, C.G., Cohn, T.A (2004). "Load estimator (LOADEST): a FORTRAN program for estimating constituent loads in streams and rivers." *USGS techniques and methods book 4*, Chapter A5, 69p.
- Russell, S. O., and Campbell, P. F. (1996). "Reservoir operating rules with fuzzy programming." *Journal of water resources planning and management*, 122(3), 165-170.
- Saha, S., Nadiga, S., Thiaw, C., Wang, J., Wang, W., Zhang, Q., Van den Dool, H., Pan, H. L., Moorthi, S., and Behringer, D. (2006). "The NCEP climate forecast system." *Journal of Climate*, 19(15), 3483-3517.
- Seaber PR, Kapinos F. Paul, Knapp GL (1987) Hydrologic Unit Maps. U.S. Geological Survey Water-Supply Paper 2294. <http://water.usgs.gov/GIS/huc.html>. Accessed 4 May 2010.
- Shamseldin, A. Y. (1997). "Application of a neural network technique to rainfall-runoff modelling." *Journal of Hydrology*, 199(3-4), 272-294.
- Shen, J., Wang, Harry, Sisson, G.M., 2002a. "Application of an integrated watershed and tidal prism model to the Poquoson coastal embayment." *Special Report in Applied Marine Science and Ocean Engineering No. 380*.
- Shrestha, R. R. S. R. R., and Simonovic, S. P. S. S. P. (2010). "Fuzzy set theory based methodology for the analysis of measurement uncertainties in river discharge and stage." *Canadian Journal of Civil Engineering*, 37(3), 429-440.
- Shrestha, R. R., and Simonovic, S. P. (2010). "Fuzzy Nonlinear Regression Approach to Stage-Discharge Analyses: Case Study." *Journal of Hydrologic Engineering*, 15, 49.
- Singer P.C., and Chang S.D. (1989). "Research Report; Impact of ozone on the removal of particles, TOC and THM precursors." AWWA Research Foundation, Denver, CO.

- Singer, P. C. (1994). "Control of disinfection by-products in drinking water." *Journal of Environmental Engineering*, 120(4), 727-744.
- Srivastava, P., McNair, J. N., and Johnson, T. E. (2006). "Comparison of process based and artificial neural network approaches for streamflow modeling in an agricultural watershed." *JAWRA Journal of the American Water Resources Association*, 42(3), 545-563.
- Srivastava, P., Gupta, A. K., and Kalin, L. (2010). "An ecologically-sustainable surface water withdrawal framework for cropland irrigation: A case study in Alabama." *Environmental management*, 46(2), 302-313.
- Sudheer, K., Gosain, A., and Ramasastri, K. (2002). "A data-driven algorithm for constructing artificial neural network rainfall runoff models." *Hydrological processes*, 16(6), 1325-1330.
- Tayfur, G., and Singh, V. P. (2006). "ANN and fuzzy logic models for simulating event-based rainfall-runoff." *Journal of Hydraulic Engineering*, 132, 1321.
- Tayfur, G., Ozdemir, S., and Singh, V. P. (2003). "Fuzzy logic algorithm for runoff-induced sediment transport from bare soil surfaces." *Advances in water resources*, 26(12), 1249-1256.
- Thirumalaiah, K., and Deo, M. C. (2000). "Hydrological forecasting using neural networks." *Journal of Hydrologic Engineering*, 5, 180.
- Tootle, G. A., Singh, A. K., Piechota, T. C., and Farnham, I. (2007). "Long lead-time forecasting of US streamflow using partial least squares regression." *Journal of Hydrologic Engineering*, 12, 442.
- Trenberth, K. E. (1984). "Signal versus noise in the Southern Oscillation." *Monthly Weather Review*, 112(2), 326-332.
- Trenberth, K. E., and Stepaniak, D. P. (2001). "Indices of El Niño evolution." *Journal of Climate*, 14(8), 1697-1701.
- Trenberth, K., and SHEA, D. (1987). "On the evolution of the Southern Oscillation." *Monthly Weather Review*, 115(12), 3,078-096.
- Trichakis, I. C., Nikolos, I. K., and Karatzas, G. P. (2009). "Optimal selection of artificial neural network parameters for the prediction of a karstic aquifer's response." *Hydrological processes*, 23(20), 2956-2969.
- USEPA (2005). "Economic analysis for the stage 2 disinfectants and disinfection byproducts rule." *Prepared by The Cadmus Group, Inc. Contract 68-C-99-206. EPA 815-R-05-010. December 2005*
- Walker Jr, W. W. (1983). "Significance of eutrophication in water supply reservoirs." *Journal of the American Water Works Association*, 75(1), 38-42.

- Wang, W., Chen, M., and Kumar, A. (2010). "An Assessment of the CFS Real-Time Seasonal Forecasts." *Weather and Forecasting*, 25(3), 950-969.
- Weisberg R. (1985). *Applied Linear Regression*, John Wiley and Sons: New York.
- Weisheimer, A., Doblas-Reyes, F. J., Palmer, T. N., Alessandri, A., Arribas, A., Déqué, M., Keenlyside, N., MacVean, M., Navarra, A., and Rogel, P. (2009). "ENSEMBLES: a new multi-model ensemble for seasonal-to-annual predictions: Skill and progress beyond DEMETER in forecasting tropical Pacific SSTs." *Geophysical research letters*, 36(21), L21711.
- Wolter, K., and Timlin, M. S. "Monitoring ENSO in COADS with a seasonally adjusted principal component index." 52-57.
- Wong, K., Wong, P., Gedeon, T., and Fung, C. (2003). "Rainfall prediction model using soft computing technique." *Soft Computing-A Fusion of Foundations, Methodologies and Applications*, 7(6), 434-438.
- Xuan Z., Qishan W., Miao, Y., Jing W. (2010). "Combining principal component regression and artificial neural network to predict chlorophyll-a concentration of Yuqiao Reservoir's outflow". *Trans. Tianjin Univ.*, 16: 467-472.
- Yuan, X., Wood, E. F., Luo, L., and Pan, M. (2011). "A first look at Climate Forecast System version 2 (CFSv2) for hydrological seasonal prediction." *Geophysical research letters*, 38(13), L13402.
- Zadeh, L. A. (1965). "Fuzzy sets*." *Information and control*, 8(3), 338-353.
- Zhang, G. P. (2003). "Time series forecasting using a hybrid ARIMA and neural network model." *Neurocomputing*, 50, 159-175.
- Zubair, L., Rao, S. A., and Yamagata, T. (2003). "Modulation of Sri Lankan Maha rainfall by the Indian Ocean Dipole." *Geophys. Res. Lett*, 30(2), 1063.

Chapter 6. Conclusion and Recommendation

6.1 Summary

The overall goal of this research was to use ENSO information, both quantitatively and qualitatively, for water resources management. Particularly, this research explored the potential of ENSO for streamflow simulation, streamflow forecasting, point source permitting and TOC load prediction.

The research was conducted in two different watersheds (e.g. Chickasaw Creek and Big Creek) in Mobile County of South Alabama, which is one of the most ENSO-affected regions of Alabama. Four specific research objectives were explored using various watershed and data-driven models. The conclusions pertaining to each objective are summarized as follows.

6.1.1 Conclusion of Objective I

There were three small sub-objectives under the first objective of this research. The first sub-objective was to find the potential linkage between climate and stream flows and the next sub-objective was to apply that linkage to ANFIS model and make a historical streamflow simulation against different ENSO events. The third sub-objective was to compare simulation skill of ANFIS and LSPC model for streamflow simulation in ENSO-affected region using direct application of SST and SLP in the ANFIS model. I identified the potential teleconnection between streamflow variation and SST/SLP variation using wavelet analysis. The distinct mutual characteristics in wavelet power of the two time series (SST and streamflow series) were experienced in 3 to 7 year band. The cross-wavelet transform between SST and streamflow indicated that significant power was shared from 1970 to 1982. The cross correlation analysis suggested that SST and precipitation had a lag correlation. Then, I applied SST and SLP as

model inputs in SR and BF simulation. In the next step, I compared ANFIS model with LSPC model and found that ANFIS model simulated better in daily as well as monthly scale, partly due to the application of SST/SLP in model and partly due to the simulating skill of the ANFIS model.

6.1.2 Conclusion of Objective II

The objective of this research was to evaluate the predicted climate data by climate model (CFSv2) and ENSO-conditioned weather sequences in ANFIS model for streamflow forecasting at one to three month advance in time. I forecasted streamflow using 24 ensemble members of temperature and precipitation and the weather generated data separately. The forecast was followed by post-processing with systematic error correction using quantile mapping method. The forecast was post-validated with observed streamflow using 7 years of observed data. The streamflow forecasted at one month was relatively better than three months lead time. The streamflow forecasting using climate model was better for one month lead time. Conversely, the weather generator approach is a viable approach for three month lead time forecast, especially for low flow prediction.

6.1.3 Conclusion of Objective III

The specific objective of this research was to demonstrate how the ENSO information can be utilized for point source discharge permitting for stream water quality protection. For this, I investigated the connection between climate variability and point source permitting using a case study of waste water treatment plant discharging ammonia nitrogen in the Chickasaw Creek.

The DO and stream temperature were simulated using the LSPC and hydrodynamic/water quality model, EPD-RIV1. Various non-parametric tests and statistical analyses suggested that ENSO is correlated with the simulated stream temperature, DO and observed streamflow. The severe and acute criteria containing fresh water muscles were set using CMC and CCC criteria of EPA. For the strict regulation of point source discharge, three inter seasonal dry periods such as La Niña winter (December-April), La Niña summer (May-July) and El Niño fall (August-October) were identified as periods of critical condition. Whereas, El Niño winter (December-April) was identified as a period of high streamflow assimilation, which can assimilate 28% more ammonia nitrogen than what is allowed using conventional seasonal 7Q10 in this season. Generally, drought perpetuates for a long time before experiencing any severe hydrological drought. In this regard, ENSO information provides sufficient warning for an impending drought (critical condition) due to the auto correlation and cross correlation characteristics of streamflow with spring and winter SST anomalies.

6.1.4 Conclusion of Objective IV

The objectives of the study were to analyze the effect of climate variability on TOC loads and develop a simple TOC prediction model at different ENSO phases of each season, which is beneficial for improved estimation of TOC loads in different ENSO phase, and minimize the additional cost required for DBP removal. For this, simulated long-term TOC data sets were used using a stable land use pattern in order to segregate the effect of climate variability from the land use effect. I developed four fuzzy logic and four hybrid models and compared their performance. I found that PCR-ANN was suitable for real time TOC load prediction as it incorporates different architectures to train the input vectors. However, it was found that fuzzy

logic model could be a better choice for qualitative forecast of TOC load at one month lead time. There were three distinct findings worth mentioning. The first finding is that the data-driven model using multiple ENSO indices (Niño 3.4 indices and TNI), temperature and precipitation can be utilized to predict the TOC load with reasonable accuracy. Another finding was that the model performance in data-driven modeling approach would be enhanced if the multiple ENSO indices were included in the input vectors, which was particularly true for the season when the TOC load was correlated with ENSO phase. The third finding was that fuzzy logic approach is a better choice as it has a capacity to forecast TOC loads qualitatively.

6.1.5 Limitations of the Study

Though the impact of ENSO is well documented in different parts of the world, its impact in climate and water resources varies from place to place. The ENSO signature and magnitude of its correlation with hydro climatic variables is location specific, and the benefit due to the potential application of ENSO entirely depends upon the spatial location of the watershed. This will be reflected on the benefit that users derive from the application of ENSO to deal with water resources issues.

6.1.6 Suggestion for Future Work

This study suggests detecting the ENSO correlation with hydro climatic variable in a selected watershed before further investigation. More importantly, though climate science has advanced in recent years, certain degree of uncertainty is associated with ENSO forecasts. In majority of our study, uncertainty analysis was not performed, and therefore, researchers can

carry out uncertainty analysis to develop confidence in the application of climate forecast in water resources.

Each major objective that was discussed in the respective chapters can be further extended as a part of future research. For example, in chapter 2, we forecasted the streamflow using SST and predicted climate input in ANFIS; however, data-driven model like ANFIS does not allow us to see how the particular model inputs are playing the role in forecast. In order to understand the importance of SST for streamflow forecasting, regression model, though are not suitable approaches for forecast, can be a better choice, particularly to explore and quantify the effect of SST. Multi-linear regression model can reveal the importance and significance of parameter (SST) in forecast. Therefore, quantifying the role of SST for streamflow forecasting in ENSO affected watersheds of the Continental United States, preferably in MOPEX (Model Parameter Estimation Experiment) basins, using stepwise linear regression approach will be a potential research to pursue in future. Similarly, in chapter 4, I demonstrated the importance of applying climate information for NPDES permitting and its potential benefit over the conventional approach. Further research can be pursued to develop a decision support system (tools) by integrating ENSO information with conventional methods. In the fifth chapter, I developed the tools for quantitative and qualitative prediction of TOC loads. It can be further extended into the bigger system; for example, watershed coupled reservoir system so that the users can realize the maximum benefit of applying ENSO information to reduce the cost associated with the DBP removal.