RICE UNIVERSITY

Development of a Distributed Water Quality Model Using Advanced Hydrologic Simulation
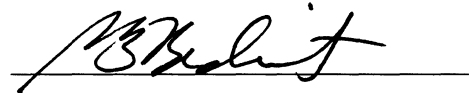
by

Aarin Teague

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
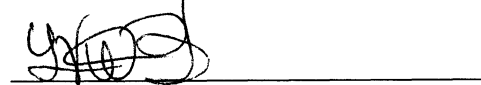REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy
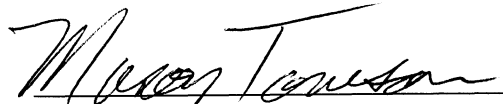
APPROVED, THESIS COMMITTEE:

Philip Bedient,
Herman and George R. Brown
Professor of Civil Engineering
Civil and Environmental Engineering

Dale Sawyer,
Professor Earth Science

Leonardo Duenas-Osorio,
Assistant Professor Civil and
Environmental Engineering

Mason Tomson,
Professor Earth Science

HOUSTON, TEXAS

August 2011

ABSTRACT

Cypress Creek is an urbanizing watershed in the Gulf Coast region of Texas that contributes the largest inflow of urban runoff containing suspended solids to Lake Houston, the primary source of drinking water for the City of Houston. Historical water quality data was statistically analyzed to characterize the watershed and its pollutant sources. It was determined that the current sampling program provides limited information on the complex behaviors of pollutant sources in both dry weather and rainfall events. In order to further investigate the dynamics of pollutant export from Cypress Creek to Lake Houston, fully distributed hydrologic and water quality models were developed and employed to simulate high frequency small storms.

A fully distributed hydrologic model, $Vflo^{TM}$, was used to model streamflow during small storm events in Cypress Creek. Accurately modeling small rainfall events, which have traditionally been difficult to model, is necessary for investigation and design of watershed management since small storms occur more frequently. An assessment of the model for multiple storms shows that using radar rainfall input produces results well matched to the observed streamflow for both volume and peak streamflow.

Building on the accuracy and utility of distributed hydrologic modeling, a water quality model was developed to simulate buildup, washoff, and advective transport of a conservative pollutant. Coupled with the physically based $Vflo^{TM}$ hydrologic model, the pollutant transport model was used to simulate the washoff and transport of total suspended solids for multiple small storm events in Cypress Creek Watershed. The output of this distributed buildup and washoff model was compared to storm water

quality sampling in order to assess the performance of the model and to further temporally and spatially characterize the storm events. This effort was the first step towards developing a fully distributed water quality model that can be widely applied to a wide variety of watersheds. It provides the framework for future incorporation of more sophisticated pollutant dynamics and spatially explicit evaluation of best management practices and land use dynamics. This provides an important tool and decision aid for watershed and resource management and thus efficient protection of the sources waters.

TABLE OF CONTENTS

List of Tables

List of Figures

Chapter 1 : Introduction

Lake Houston is an important source of drinking water for the City of Houston, with approximately 300,000 cubic meters of water withdrawn daily (Chellam, 2008), to provide drinking water for approximately 1 million customers (Smyer, 2008). This is projected to increase to 1,360,000 cubic meters per day by 2030 (Chellam et al , 2008). Unfortunately, the lake experiences seasonal algal blooms and stratification during warm weather. This eutrophication is associated with nutrient inflow and suspended solids from the seven watersheds draining into the lake. Increasing urbanization within the watersheds is expected to increase urban runoff with loads of nutrients, suspended solids, and bacteria. The combination of nutrient enrichment combined with bacterial impairment increases the cost of water treatment for the drinking water purification plant on Lake Houston.

In order to address the rising water treatment costs, source protection measures need to be implemented within the watersheds draining into the lake. Seven watersheds, encompassing 5,021 km^2, drain into the lake (See Figure 1-1). Cypress Creek, the most highly urbanized of these watersheds, is impaired for bacteria(TCEQ, 2008a) and listed on the 2008 303-d concerns list for nutrient enrichment (TCEQ, 2008b). Because of its contribution of urban and agricultural runoff to the lake, knowledge of the water quality in Cypress Creek is necessary for protection of the City of Houston's water supply.
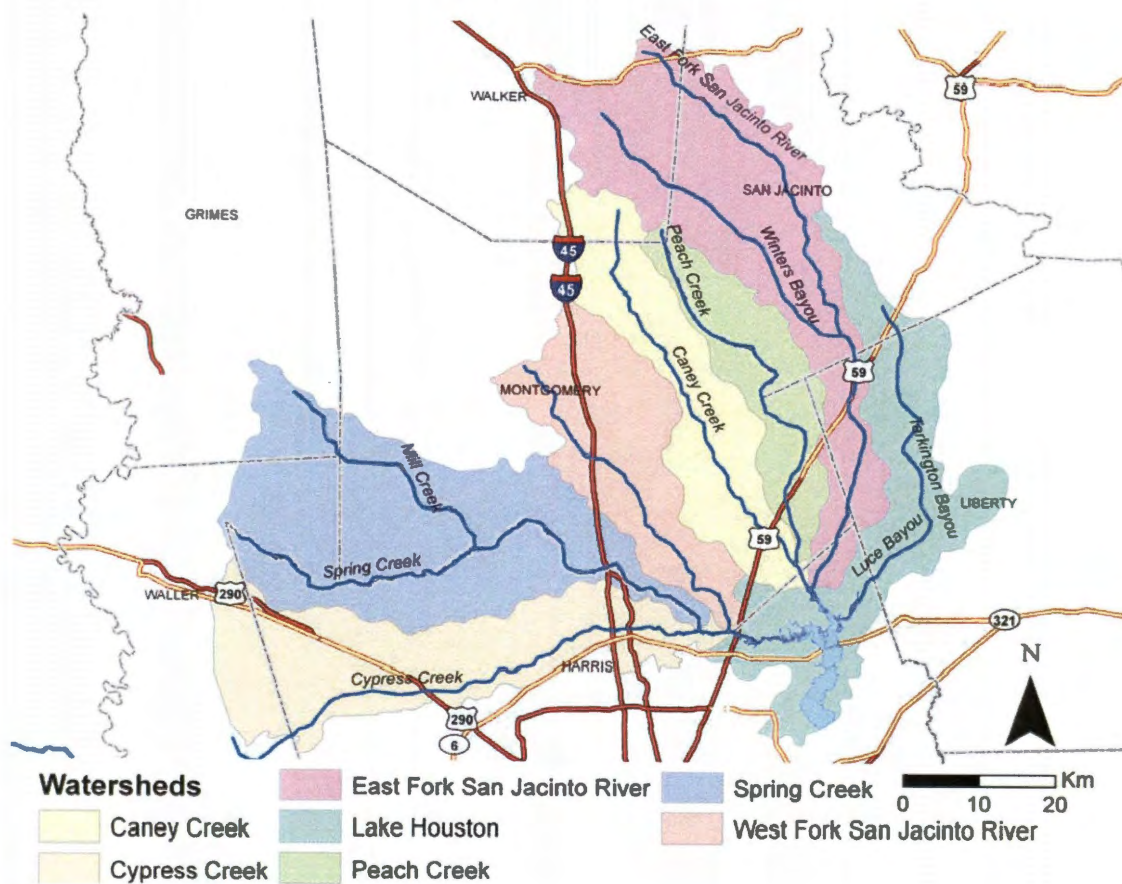
Figure 1-1 . Watersheds flowing into Lake Houston

Previous efforts to address the pollutant loading to Cypress Creek were based on statistical analysis of historical water quality data (Miertshcin & Associates, Inc., 2009). Limited water quality modeling has been performed to assess pollutant transport during storm events. By developing and applying a fully distributed pollutant washoff and transport model, the pollutant loading to Cypress Creek can be further investigated and characterized. Ultimately, water resource management and watershed protection can be assisted by distributed hydrologic and water quality modeling.

Current water quality models use lumped approaches which parameterize the watershed by aggregating similar spatial areas and applying values to the assumed homogeneous area (Bicknell et al., 2001). This limits the utility of the models due to the lack of spatially explicit pollutant source and transport information. The development of a fully distributed pollutant washoff and transport model takes advantage of the advancements in physically based hydrologic modeling, radar rainfall technology, and GIS spatial data processing. With the improved accuracy of hydrologic prediction (Vieux, 2004), the current limitation of some lumped models (Singh et al., 2005) in simulating small storms can be overcome. This project focused on the model development to simulate small storms due to the higher frequency of events and thus greater impact on water quality. Traditional lumped model approaches overestimate streamflow in smaller events (Chen et al., 1995; Singh et al., 2005). To overcome the limitations of lumped water quality models, a physically based, fully distributed model can be coupled with an independent pollutant washoff and transport model. This provides improved accuracy of hydrologic simulation as well as the utility of detailed spatial information on pollutant transport.

1.1. Objectives

The goal of the proposed project is to develop a water quality model using distributed hydrologic modeling for the simulation of pollutant buildup, washoff, and transport in Cypress Creek watershed. The developed model simulates the movement of pollutants through a watershed during rainfall events in order to provide a tool for addressing the large export of TSS from the influent watersheds to Lake Houston after storms (Matty et al., 1987; Sneck-Fahrer et al., 2005). This water quality model could then be applied to

other watersheds, notably the other watersheds flowing into Lake Houston, for a comprehensive storm water management in the greater Lake Houston watershed. Even further, the pollutant washoff and transport model could be applied for other watersheds with different slope and soil characteristics in order to develop a robust simulation program for the investigation of a variety of water quality problems.

This is accomplished by the following objectives:

Objective 1: Evaluate historical water quality data of pollutant loads to Lake Houston for both low flow and storm events, using statistical techniques and load duration curves to characterize pollutant and watershed behavior.

Objective 2: Develop a rainfall runoff model for Cypress Creek incorporating antecedent moisture condition and soils data, calibrated for small storm using NEXRAD radar rainfall.

Objective 3: Create a fully distributed pollutant washoff and transport model for Cypress Creek Watershed and link with $Vflo^{TM}$ hydrologic data output.

Objective 4: Collect water quality samples during storm events to assess the pollutant loads in Cypress Creek throughout the rising and falling limbs of a hydrograph.

Objective 5: Calibrate and validate pollutant washoff and transport model using the stormwater concentrations of TSS.

1.2. Significance

The City of Houston (COH), the fourth largest city in the U.S relies primarily on Lake Houston for drinking water for 3.5 million customers. The Lake has water quality

concerns from environmental and public health standpoints which have resulted in an increase in the cost of drinking water treatment. These concerns were investigated in the early 1980's as part of a comprehensive lake study assessing pollutant loading (Bedient et al, 1980), and found that significant loading was entering Lake Houston from Cypress Creek Watershed. Further studies in the same time period found that Cypress Creek was a major contributor of *E. coli*, nutrients, and suspended sediment loads to Lake Houston (Newell, 1981). During storm events, significant urban runoff flows from Cypress Creek into Lake Houston (Matty et al., 1987, Sneck-Fahrer, 2005). Due to the significant impact that storm-related pollutant export from Cypress Creek has on Lake Houston's water quality, it is important to investigate stormwater quality (Oden and Graham, 2008). These studies have found that periodic single grab sampling was found to be inadequate for useful estimation of the pollutant export from the watershed.

During the past three decades of increased urban development, the aforementioned water quality problems have persisted and intensified. This research is an effort to continue addressing the need for source water protection, by providing the best geo-spatial science, advanced models, and datasets to predict pollutant loading to Lake Houston from the influent watersheds. A fully distributed model of pollutant washoff and transport will provide an estimation of the pollutant concentration throughout a storm event. This can provide a future tool for spatially explicit analysis of pollutant sources and transport during rainfall events. In an effort to proactively protect water resources, the model can be used to evaluate the effects of different management strategies, land use changes, or climate scenarios. This water quality model was developed to use the output from any

fully distributed hydrologic model and runs independently so that it can be further tested and applied in other watersheds with different physical attributes in the future.

A fully distributed approach was undertaken for modeling pollutant transport and modeling, in order to provide greater spatial resolution than is provided by currently available lumped models. One of the limitations of lumped water quality models is the challenges in modeling the hydrology of small storm events. It is important to model the hydrology of small storms, because the high frequency of these events makes their impact on water quality greater than low frequency, high magnitude storms. Utilizing distributed hydrologic modeling takes advantage of the improved accuracy achieved by distributed rainfall-runoff simulation.

1.3. Description of the Study Area

Cypress Creek is a 797 km^2 (308 mi^2) watershed north of the city of Houston in north Harris County with the upstream, western portion in Waller County. It flows 80 km (50 river miles) to Lake Houston and is a complex watershed with a variety of land uses and covers. The western upstream part of the watershed is undeveloped primarily as cultivated agricultural fields. The eastern portion of the watershed has primarily residential development and is home to most of 216,000 residents (ESRI, 2000). Furthermore, the watershed has experienced rapid urbanization in the past decade, losing much of its forest cover to residential development. Based on the 2002 Land Cover analysis performed by the Houston-Galveston Area Council (H-GAC, 2002), low and high intensity development accounted for approximately 16% of the watershed. This

development increased to approximately 36% by 2008 (H-GAC, 2008). Additionally,

forested areas decreased from 22% in 2002 to 11% in 2008 whereas grasslands decreased

from 51% to 33% (See Table 1-1 and Figure 1-3).

Table 1-1. Land use and cover change from 2002 to 2008 in Cypress Creek Watershed

	Percentage Land Use							
Land Cover	Developed	Cultivated	Grassland	Forest	Wetland	Bare	Open Water	Other
11333								
2002	4.2	8.3	72.5	4.7	7.9	0.3	2.1	
2008	6.5	69.7	10.1	0.7	8.5	2.9	1.5	0.2
Change in % Land Cover	2.2	61.4	-62.4	-4.0	0.5	2.5	-0.6	0.2
11332								
2002	12.0	6.3	55.0	19.6	5.5	0.2	1.4	
2008	29.7	34.1	13.5	5.0	10.6	3.2	1.2	2.6
Change in % Land Cover	17.7	27.8	-41.5	-14.6	5.1	3.0	-0.2	2.6
11328								
2002	28.9	0.3	33.0	33.7	2.2	0.7	1.1	
2008	65.5	5.2	9.6	6.6	7.1	1.5	0.7	3.9
Change in % Land Cover	36.6	4.8	-23.4	-27.2	4.8	0.9	-0.4	3.9
Total Watershed								
2002	16.5	4.5	51.0	21.2	3.7	2.6	0.4	
2008	36.8	2.5	32.7	11.2	11.0	2.2	2.5	1.1
Change in % Land Cover	20.3	-2.1	-18.3	-10.0	7.3	-0.4	2.1	1.1
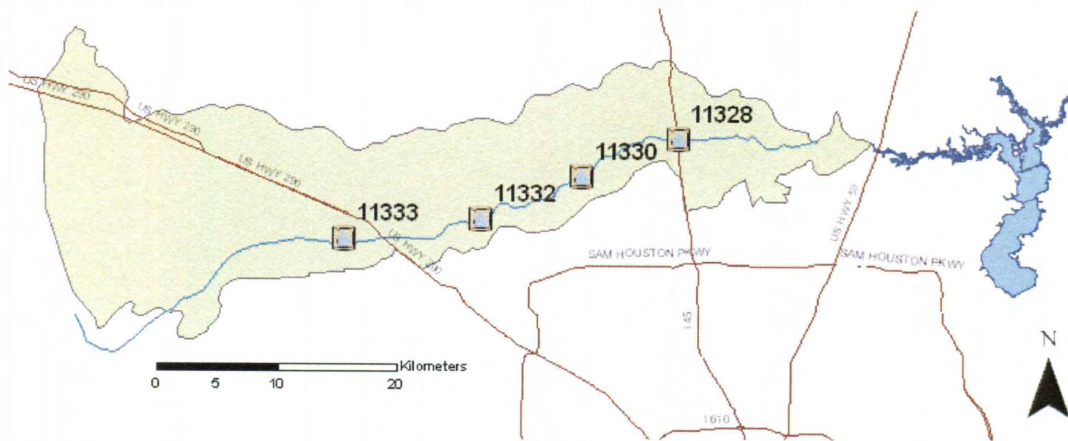
Figure 1-2. Locations of water quality stations in Cypress Creek

The changes in land use primarily occurred in the downstream and central portions of the watershed (See Figure 1-3). In the upstream portion of Cypress Creek (gauged by 11333), the primary change in land use was a minor loss in forest (Table 1-1 and Figure 1-3), although there was a reclassification of pastures from grassland to cultivated. In the center of the watershed (gauged by 11332), there were greater losses in forest cover as well as grassland (See Table 1-1). By 2008, these areas (See Figure 1-3), were developed. In the downstream area (gauged by 11328), forest was converted to developed and residential. The land cover modification from forested to developed has significantly increased the impervious cover throughout the middle to downstream areas of the watershed.
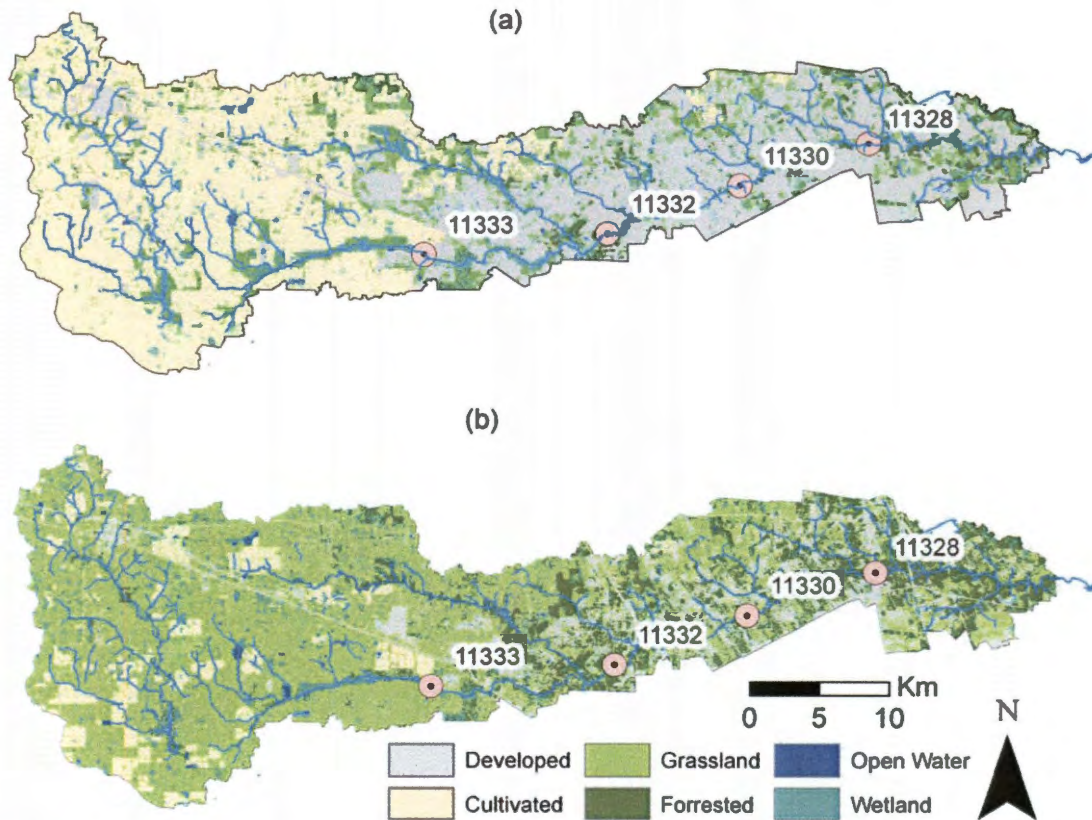
Figure 1-3. Land cover for Cypress Creek in (a) 2008 and (b) 2002

Cypress Creek watershed is relatively flat with sandy loam soils. The major soil group is

a Wockley series (See Figure 1-4), permeable alfisol (NRCS, 1976). Using the Natural

Resource Conservation Service (NRCS) soil survey, the soil characteristics were

processed in order to create spatial datasets for soil infiltration characteristics. The values

assigned to each general soil class are in Table 1-2.
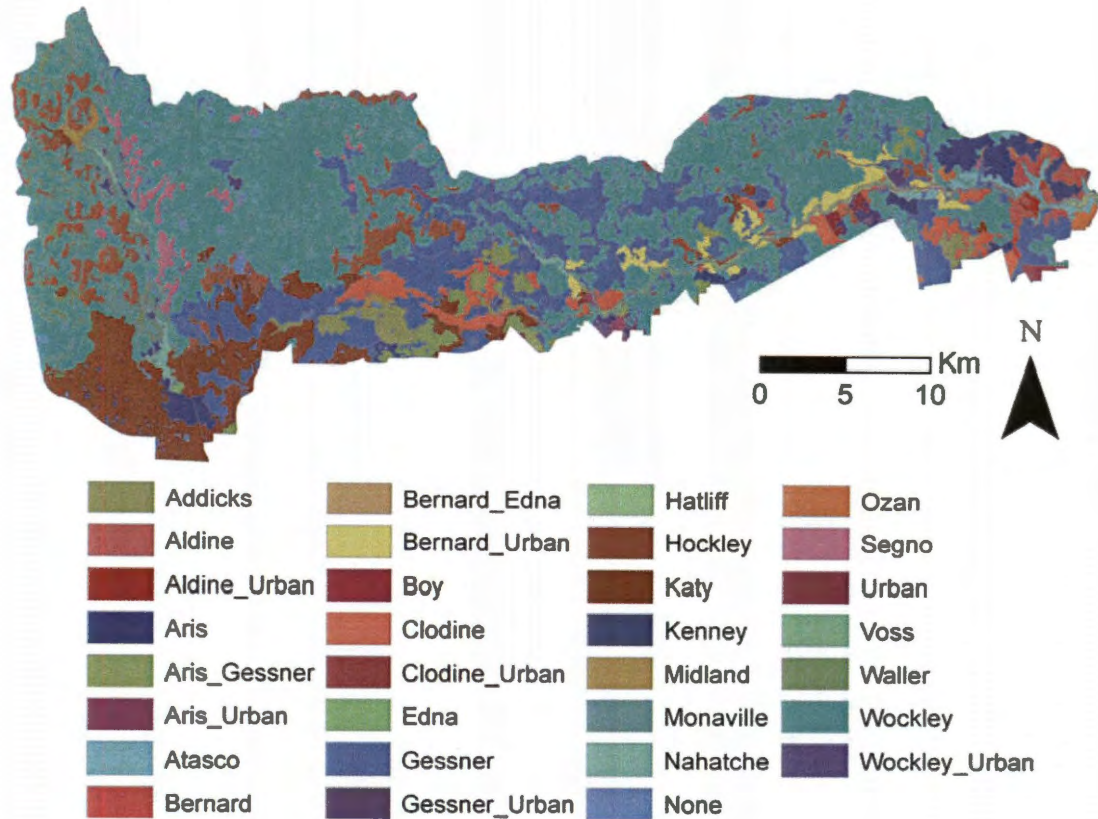
Figure 1-4. Cypress Creek soil taxonomy (NRCS, 2006)

Table 1-2. Infiltration parameters of soils in Cypress Creek (NRCS, 2006)

Soil Type	Effective Porosity (cm/cm)	Wetting Front Suction (cm)	Hydraulic Conductivity (cm/hr)
Loamy Sand	0.401	6.130	3.302
Sandy Loam	0.412	11.010	3.302
Sand	0.417	4.950	10.160
Silty Clay Loam	0.432	27.300	0.254
Loam	0.434	8.890	3.302
Clay Loam	0.390	20.880	0.254

In the areas close to the stream, or the near riparian area, as well as the center of the

watershed have a moderate hydraulic conductivity and higher effective porosity (Figure

1-5). The implication of urbanization in the region with these characteristics is that the

increase in impervious cover will increase the runoff and decrease erosion potential. On the other hand, soils in the upstream and overland portion of the watershed have a high hydraulic conductivity and mid-effective porosities (Figure 1-5). The 67% higher hydraulic conductivities in these overland areas results in lower runoff potential than the near riparian areas. The spatial variability of soil properties further highlights the need for distributed water quality modeling to select and design best management practices (BMPs) that are appropriate for the hydrology, soil properties, and land cover.
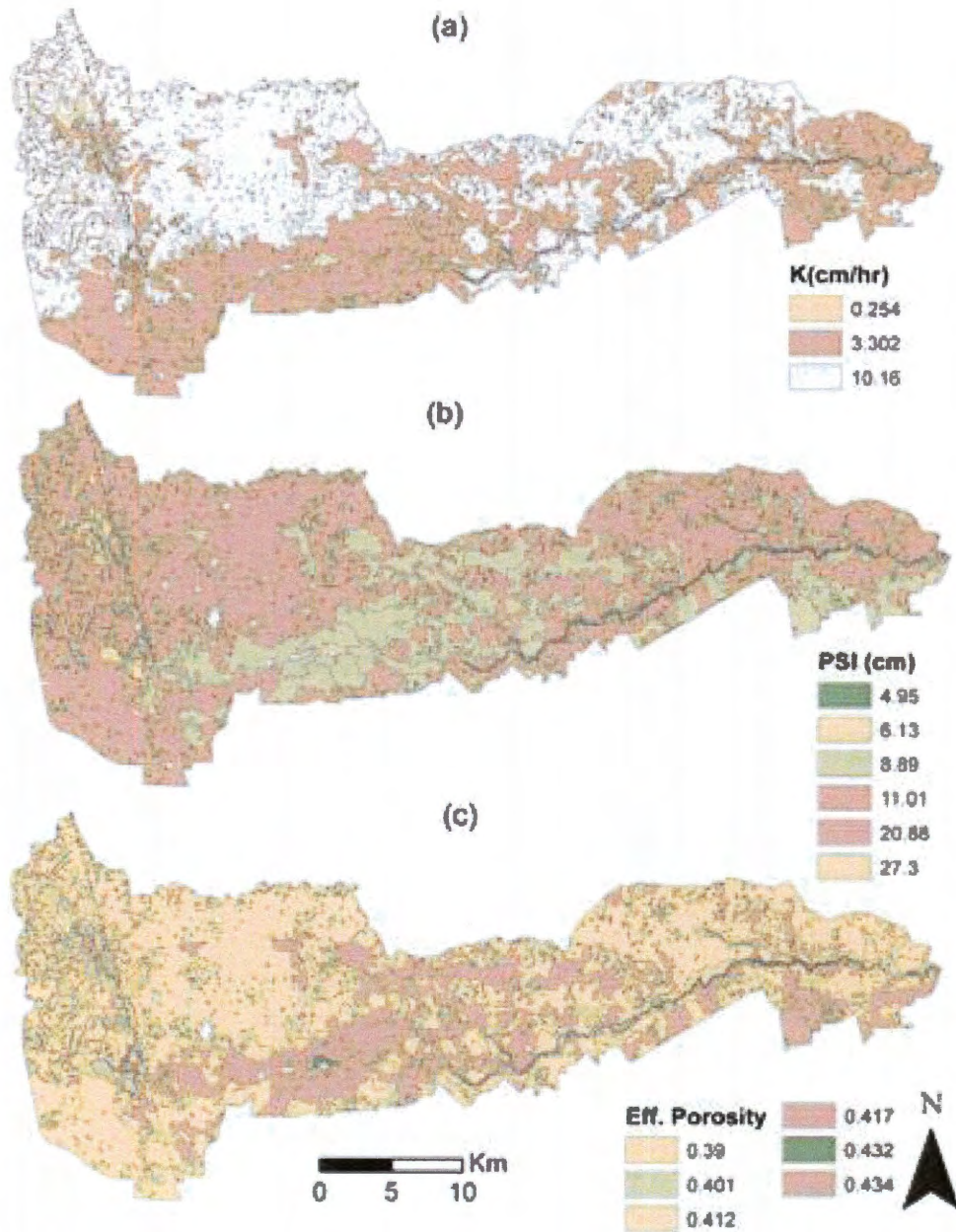
Figure 1-5. Cypress Creek soil (a) hydraulic conductivity, (b) wetting front suction, and (c) effective porosity

1.4. Water Quality in the Study Area

Water quality data have been collected intermittently within Cypress Creek since1980.

Monitoring stations were operated by the City of Houston, Water Quality Control and

Health and Human Services. This historical data has been analyzed with various statistical methods including trend analysis, multivariate analysis, and load duration curve analysis.

A comparison of the dry flow and storm flow loading of total suspended solids (TSS) , *E. coli*, total dissolved solids (TDS), total phosphorus, chloride, and nitrate found that storm loading of TSS and *E. coli* are much greater than low flow loading (See Appendix A). However for the other constituents, storm flow loading does not overwhelm the dry flow loading. This suggests that simulation of runoff related transport of TSS and *E. coli* is key to addressing these water quality impairments.

Trend analysis was performed using the Mann Kendall and Seasonal Kendall trend test in order to establish whether the concentration and loading rates of the previously mentioned constituents has increased during the time period of intense urban development (See Appendix B for methodological details and conclusions). Increasing trends in nitrate, TDS, *E. coli*, and chloride were found for the station in the down stream urbanized portion of the watershed. In contrast the only trend detected was an increasing trend in chloride at the station located in the center of the watershed. The increasing trend in chloride, which is attributed to wastewater treatment plants (Sawyer et al., 2006), would indicate an increasing influence of the permitted discharges. The trends identified indicate that the increased urban runoff and waste water discharge resulting from the urbanization in the downstream and center of the watershed could potentially be linked with the water quality degradation in Cypress Creek.

The analysis of the historical water quality data indicates an intricate mix of pollutant sources which are active in both low flow and storm flow stream conditions. The complexity of the pollutant loading to the stream during various flow conditions denotes the need for further analysis of pollutant loading via load duration curves and multivariate analysis as well as detailed modeling of storm flow loading.

1.5. Summary

Cypress Creek is a rapidly urbanizing watershed which is key to protecting the source of drinking water for the City of Houston. The watershed's soil and slope characteristics mean that this urbanization will result in increases in runoff and pollutant loading. This project's overarching goal was to investigate the export of pollutants from Cypress Creek Watershed, in order to support water resource protection and address degradation of water quality in Lake Houston. Analysis of water quality data collected during a period of rapid urbanization illustrates the need for advanced hydrologic and pollutant transport modeling. Further statistical analysis of the historical water quality data, hydrologic modeling, and simulation of pollutant washoff and transport were conducted to meet the objectives of this study.

1.6. Organization of this Document

This document is the compilation of various article published throughout the research process. The reader will find three separate manuscripts, that at the time of submission of the dissertation were at various stages of publication, including (Chapter 2) Targeted Application of Seasonal Load Duration Curves using Multivariate Analysis in Two

Watersheds Flowing into Lake Houston, published in the Journal of American Water Resources Association; (Chapter 3) Radar Rainfall Application in a Distributed Hydrologic Modeling for Cypress Creek Watershed, Texas submitted to the Journal of Hydrological Engineering; and (Chapter 4) Modeling of Pollutant Washoff and Transport Using Fully Distributed Hydrologic Modeling.

Chapter 2 : Targeted Application of Seasonal Load Duration Curves using Multivariate Analysis in Two Watersheds Flowing into Lake Houston

Aarin Teague[1], Philip B. Bedient[2], Birnur Guven[3]

Abstract: Water quality is a problem in Lake Houston, the primary source of drinking water for the City of Houston, Texas, due to pollutant loads coming from the influent watersheds, including Spring Creek and Cypress Creek. Statistical analysis of the historic water quality data was developed in order to understand the source characterization and seasonality of the watershed. Multivariate analysis including principal component, cluster, and discriminant analysis provided a custom seasonal assessment of the watersheds so that loading curves may be targeted for season specific pollutant source characterization. The load duration curves have been analyzed using data collected by the USGS with corresponding City of Houston water quality data at the sites to characterize the behavior of the pollutant sources and watersheds. Custom seasons were determined for Spring and Cypress Creek watersheds and pollutant source characterization compared between the seasons and watersheds.

2. Introduction

Water quality assessment is often based upon sampling for numerous water quality parameters at a limited set of conditions (Smith *et al.*, 1997). Most notably, water quality assessments through regulatory regimes include limited storm water sampling (Strobl and Robillard, 2008; Park *et al.*, 2006) which has the potential of sampling bias and does not provide a complete understanding of the stream conditions. Watershed protection efforts often include a hydrologic modeling component (Shirmohammadi *et al.*, 2006), which allows for the evaluation of varying scenarios, the optimization of resource allocation, and the selection of best management practices (Refsgaard *et al.*, 2005; Santhi *et al.*, 2006; Jayakrishnan *et al.*, 2005).

The purpose of this paper is to compare two watersheds that drain to Lake Houston, near the city of Houston. The two watersheds were compared through a combination of multivariate analysis techniques and load duration curves. This framework was used to assess seasonality and sources of nitrates, total phosphorus, and *E. coli*.

The appropriate study areas for testing hydrologic models can be determined by comparing different watersheds. Identification of similar and dissimilar watersheds provides a basis for selection of watersheds that can be appropriately compared for a variety of scenarios through hydrologic models. A distinct part of this identification is understanding the influence of seasonality and source characterization, which is important for the appropriate application of water quality models by resource managers in development of watershed protection plans. In particular, the selection of Best

Management Practices (BMPs) requires knowledge of the seasonality and character of sources in order for BMPs to be structured to fit the seasons during which certain pollutant sources are primarily active.

Assessing the influence of seasonality and source characterization can be accomplished through linking multivariate analysis with load duration curves. Multivariate statistical techniques, including principal component analysis, cluster analysis, and discriminant analysis are used as unbiased methods in analyzing water quality data including data reduction and interpretation (Suk and Lee, 1999) while load duration curves are used to characterize violations of the water quality standard by the stream flow condition at which the violations occurred (Babbar-Sebens and Karthikeyan, 2009). These methods have been widely applied for the characterization and evaluation of temporal and spatial variations caused by natural and anthropogenic processes (Panda *et al.*, 2006; Alberto et al; 2001; Bengraine and Marhaba, 2003; Singh et al, 2004; Shrestha and Kazama, 2007; Najafpour et al, 2008), including the identification of seasonality and its effects on water quality parameters (Vega et al., 1998; Shrestha et al., 2008; Ouyang et al., 2006). Principal Component Analysis (PCA) is an unbiased pattern recognition technique used to decrease the dimensionality of the dataset without loss of variability (Mahloch *et al.*, 1974; Parinet *et al.*, 2004). Cluster Analysis (CA) uses the information gleaned from principal component analysis to classify samples of principal components into clusters of like members (Boyer et al. 1997). With this unbiased cluster analysis, the parameters or variables which have the greatest power to sort samples into clusters are determined through Discriminant Analysis (Singh *et al.*, 2005).

Through multivariate techniques, a large dataset of historical water quality data can be reduced using PCA to its most important factors (Ouyang, 2005; Haag and Westrich, 2002) and the temporal variation in water quality can be assessed (Razmkhah *et al.*, 2010) to determine clusters of months that have similar water quality characteristics (Kumar et al., 2009). This unbiased, custom determination of seasons provides a novel temporal framework to classify water quality samples that is unique to each watershed. Additionally, identification of the parameters which discriminate between seasons and the underlying correlations between parameters provide insight into the influence of seasonality on the water quality (Koklu et al., 2010). Based on these seasons, the water quality dataset can be segmented, and load duration curves can be assessed by season for time specific source characterization.

Load duration curves are plots of actual pollutant loading to a stream superimposed on the allowable loading to the stream. This technique is often used in the development of watershed protection plans as well as in the calculation of total maximum daily loads (TMDLs) (Ward *et al.*, 2009; USEPA, 2007). Load duration curves give insight into the patterns of loading throughout a variety of flow conditions, notably through the characterization of pollutant sources as point or non-point sources (Johnson *et al.*, 2009).

The research presents a novel approach to characterizing pollutant sources that are active during a specific water quality season. A framework developed using a combination of multivariate techniques and load duration curves was used to evaluate the water quality

characteristics of two watersheds, Spring Creek and Cypress Creek, discharging into

Lake Houston, a source of the City of Houston's drinking water. The objective of this

paper is to compare these two watersheds based on (1) determination of watershed

specific seasons, (2) season-specific source characterization through load duration curves,

and (3) discriminating parameters and associated correlations. Multivariate analysis was

performed on the available water quality and stream-flow data and load duration curves

were developed for *E. coli*, nitrates, and total phosphorus. The developed framework can

be used to select appropriate watersheds to be used for future hydrologic modeling efforts

and to improve water quality monitoring.

2.1. Background

2.1.1. Study Area

Lake Houston was the primary source of drinking water for the City of Houston, with

approximately 300 million liters of water withdrawn daily (Chellam *et al.,* 2008).

Unfortunately the lake was listed as impaired for bacteria and with concerns for nutrient

enrichment and chlorophyll on the Texas Commission on Environmental Quality (TCEQ)

2008 303-d list (TCEQ, 2008b; TCEQ, 2008c). The lake experiences seasonal algal

blooms attributed to high levels of nutrients draining into the lake from the watersheds.

Spring and Cypress Creek drain to Lake Houston (Figure 2-1), covering an area of 1,964

km^2 north of the city of Houston. These two watersheds are part of the San Jacinto River

Basin located west of the lake and were rapidly urbanizing. The comparison of land use

in the two watersheds is presented in Table 2-1. Of all the watersheds draining to Lake

Houston, Cypress Creek watershed had the most active urban development (Liscum and

East, 2000) and contributed the greatest nutrient loading to the lake (Sneck-Fahrer et al.,

2005). The nutrient loading from Cypress Creek was associated with eutrophication

within the Lake and thus was implicated in the challenges of treating water to meet

drinking water standards (Oden and Graham, 2008).
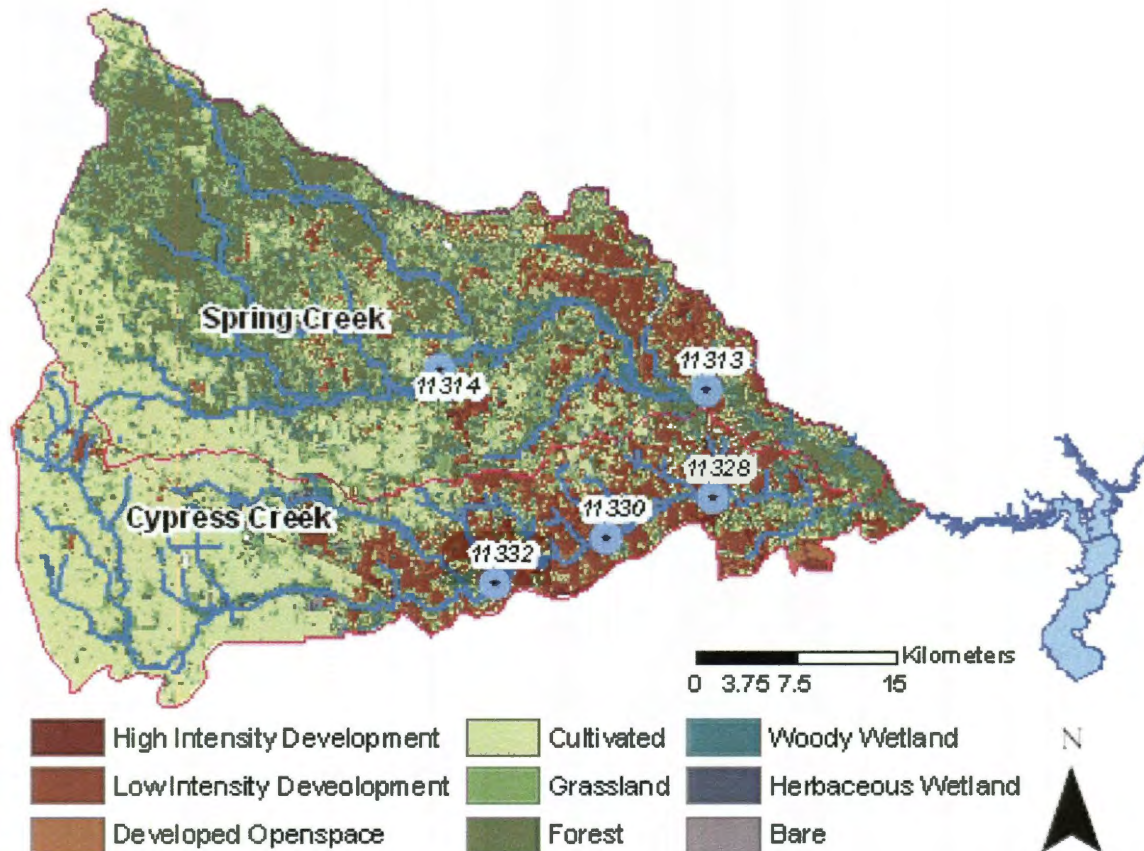
Figure 2-1.Watersheds flowing into Lake Houston with labeled water quality monitoring stations

Table 2-1. Land use breakdown of Spring Creek and Cypress Creek Watersheds

Land Use	Spring	Cypress	Total (km^2)
High Density Development	4.08%	7.93%	111
Low Intensity Development	3.17%	6.01%	85
Open Development	0.34%	6.01%	52
Cultivated	35.51%	53.19%	838
Grass & Shrub Lands	50.57%	20.13%	753
Forest	1.31%	1.21%	25
Woody Wetland	2.49%	2.16%	29
Herbaceous Wetland	1.00%	3.00%	36
Bare Land	1.53%	0.36%	21
Open Water	0%	0%	0
Total (km^2)	**1174**	**792**	**1966**

The USGS streamflow gages, water quality stations, stream impairments, concerns and

number of permitted outflows are outlined for each watershed in Table 2-2. Both

watersheds were classified for contact recreation, public water supply, and high aquatic

life use (USEPA, 2009), making their bacterial impairments and concerns for nutrient

enrichment of particular concern for protecting the drinking water source for the City of

Houston. Potential bacterial sources include failing septic systems, illicit stormwater

connections, parking lot storm water runoff, agricultural runoff, pet waste, and avian

wildlife populations (H-GAC, 2004a and 2004b). Nutrient pollution comes from these

potential sources in addition to treated wastewater outfalls, runoff with fertilizers applied

to lawns, golf courses, and croplands.

Table 2-2 .Comparison of watersheds by monitoring and impairment

| Watershed | USGS Streamflow Gage | WQ Gages | 303-d List | | # Permitted Wastewater Outfalls[c] | Permitted Wastewater Outflow |
			Impairments[a]	Concerns[b]		(cfs)
Cypress Creek	08069000, 08068740, 08068800	11328, 11332, 11330	Depressed DO, Bacteria	Nitrates, Total Phosphorus, Orthophosphorus, Impaired Habitat	101	107[c]
Spring Creek	08068500, 08068275	11313, 11314	Depressed DO, Bacteria	Nitrates, Total Phosphorus, Orthophosphorus, Impaired Habitat	14	35[d]

a. (TCEQ, 2008b), b. (TCEQ,2008c), c. (H-GAC, 2008) d.(USEPA, 2010)

Physical characteristics of the two watersheds, including watershed length, soils, and slopes, are compared in Table 2-3. Both Spring and Cypress watersheds have primarily sandy loam soils, leading to less erosion potential than other regional watersheds with clay soils (NRCS, 2006). Therefore urbanization and other changes in impervious cover increases the runoff rate within these watersheds. Both of these streams were in their natural state in the upper portion of the watershed which was primarily undeveloped (H-GAC, 2008). In contrast the stream channels in the lower, urbanized portions of the watersheds, have been widened with some concrete present for erosion control. The highly urbanized, lower, eastern part of Cypress Creek, upstream of the most downstream water quality monitoring station (station 11328 in Figure 2-1), watershed contained most of Cypress Creek's 101 permitted wastewater outfalls. A majority of these outfalls were small package plants serving Municipal Utility Districts (MUDs) (H-GAC, 2004a), representing a large number of point sources. Monthly sampling of the average outfall discharge (H-GAC, 2009) was similar to the low flow stream-flow (USGS, 2010) thus potentially linking low-flow pollutant loading with these and other point sources.

Table 2-3. Watershed characteristics

Watershed	Stream Length (km) [a]	Tributaries	Soils[b]	Average Overland Slope[c]
Cypress Creek	78	Little Cypress Creek, Snake Creek, Mound Creek, Faulkley Gully, Turkey Creek	Fine Sandy Loams	0.18%
Spring Creek	111	Willow Creek, Walnut Creek, Panther Branch, Mill Creek, Brushy Creek, Bear Creek	Loamy Fine Sand	0.31%

a. Calculated from TSARP LIDAR Data (TSARP, 2005), b. STATSGO (NRCS, 2006),
c. Calculated from Tx Elevation Dataset (USGS, 2007)

2.1.2. Analysis Techniques

2.1.2.1. Principal Component Analysis

Principal component analysis (PCA) is a group of pattern recognition techniques (Simeonov et al., 2003) that are used to reduce the dimensionality of a data set, while retaining the largest possible variability of the original dataset (Singh et al., 2004). PCA is based on the eigenvector decomposition of the covariance or correlation matrix (Bengraine and Marhaba, 2003; Morales, 1999). Principal components form the best linear approximation of the original variables (Dechesne et al., 2005) and are orthogonal or noncorrelated to each other (Li and Zhang, 2010) while producing maximum variance (Helena, 2000). For example in Table 2-4, the variables are transformed into principal components which reduce the dataset and reflect the influence of factors that incorporate multiple variables. According to the Kaiser criterion, only principal components with an eigenvalue greater than one should be retained (Liu et al., 2003). PCA is sensitive to outliers, missing data, and poor linear correlations between variables due to poorly distributed variables (Sarbu and Pop, 2005).

Table 2-4. Example of data reduction by principal components

Variables		Principal Component
Temperature	Temp	$PC1 = -0.41 * EC + 0.52 * Temp + 0.47 * Su + 0.5 * Exc$
Sulfate Concentration	Su	
Percent Exceedance of Flow	Exc	
E. coli Concentration	EC	$PC2 = -0.59 * EC + 0.78 * TSS$
Total Suspended Solids	TSS	

Principal component analysis has been used to empirically identify the main processes of

nutrient transport for development of simplified diagnostic models (Petersen *et al.*, 2001),

identify useful pollution indicators and delineate polluted areas (Wu and Wang, 2007),

determine pollutant source apportionment (Simeonov *et al.*, 2003), and to discriminate

the individual effects of season and anthropogenic activity on water quality (Vega *et al.*,

1998). PCA is most often used to interpret large datasets for characterization and data

reduction, as it provides information on the most meaningful parameters which describe

the whole data set and summarizes the statistical correlations among variables with

minimal loss of the original information (Helena *et al.*, 2000).

2.1.2.2. Cluster Analysis

Cluster Analysis is an unsupervised pattern recognition method that groups samples into

clusters based on similarity of the samples' characteristics (Lee *et al.*, 2004; Zhou *et al.*,

2007). This technique exposes intrinsic structure and underlying behaviors of a dataset

with no prior assumptions concerning the data (Vega *et al.*, 1998). The goal is for the

clusters to exhibit high intra-cluster homogeneity and high inter-cluster heterogeneity

(Shrestha and Kazama, 2007). Hierarchical methods, including Ward's method, form

clusters sequentially, starting with the most similar of objects then with each step forming

higher clusters, or clusters with greater membership, until a single clusters containing all the samples is obtained (Alberto *et al.*, 2001; Gupta *et al.*, 2009). Ward's method uses an analysis of variance (ANOVA) approach to evaluate the similarity of clusters with the goal of minimizing the sum of squares of any two clusters (Venugopal *et al.*, 2009). The results of cluster analysis assist in interpreting large datasets and identifying patterns (Vega *et al.*, 1998). The optimal number of clusters can be determined by finding the local maxima in the pseudo F statistic (DeGaetano, 1996). Cluster analysis has been used to optimize water quality monitoring strategies (Zhou *et al.*, 2007), characterize hydro-chemical regimes of groundwater (Suk and Lee, 1999), and determine sources of fecal pollution (Hagedorn *et al.*, 1999).

2.1.2.3. Discriminant Analysis

Discriminant Analysis (DA), also called supervised pattern recognition or canonical variate analysis (Shin and Fong, 1999), determines the variables that discriminate between clusters of observations. This technique is used to calculate discriminant functions for describing the differences between clusters, to predict cluster membership of observations, and ultimately data reduction. Given prior knowledge of observation cluster membership, DA determines the significance of different variables (Ellison et al., 2009) by analyzing dependence using canonical correlation. Forward stepwise DA, a specific DA technique, is a process where variables are included in a discriminant function one at a time, starting with the greatest significance, until no changes to the discriminant function are achieved (Singh *et al.*, 2004). At each step, an F-test from

analysis of covariance of the selected set of discriminating variables is performed with a

significance level of 0.15 for a variable to enter or leave the function.

DA has been used to identify sources of fecal pollution using antibiotic resistance

patterns (Harwood *et al.*, 2000), predict biologic conditions of a benthic environment

(Shin and Fong, 1999), develop eco-region classifications for water quality patterns

(Ravichandran *et al.*, 1996), and to confirm anthropogenic origin of nutrients in aquifers

(Lambraikis *et al.*, 2004). When evaluating data for each watershed, DA identifies the

variables that best differentiate between clusters of months or seasons.

2.1.2.4. Load Duration Curves

Load duration curves (LDC) are constructed by first evaluating flow duration curves.

Historical streamflow data are ranked in descending order and the percent exceedance is

calculated (rank/total number of points). The streamflow (ft^3/s) is then plotted versus the

percent exceedance. A flow duration curve provides information about the percentage of

time a particular streamflow value was exceeded over some historical period, thus

providing a hydrologic "signature" of a catchment (Cigizoglu and Bayazit, 2000). Load

duration curves are an extension of flow duration curves, where water quality violations

are characterized by the flow condition at the time of occurrence. For a load duration

curve, the allowable load is calculated by multiplying the streamflow data by the water

quality standard concentration for streams and represents the theoretical mass loading

rate of pollutant that the stream can recieve and remain in compliance with the water

quality standard. Actual loads are calculated by multiplying the measured concentration

of the water quality constituent by the streamflow that occurred at the time of measurement The allowable and actual load are then plotted against the percent exceedance of the corresponding streamflow (Bonta and Cleland, 2007).

Actual loads that fall below the allowable load curve are considered to be in compliance with water quality standards whereas points above the line indicate water quality violations (Figure 2-2). The LDC is then examined to determine the flow conditions where a majority of water quality standard violations occur. Flow conditions are divided into different categories of flow: 0 to 10% exceedance , High Flows; 10-40% exceedance , Moist Conditions; 40-60%, Mid-Range Flows; 60-90% exceedance, Dry Conditions; and 90-100% exceedance, Low Flows (Morrison and Bonta, 2008; USEPA, 2007).
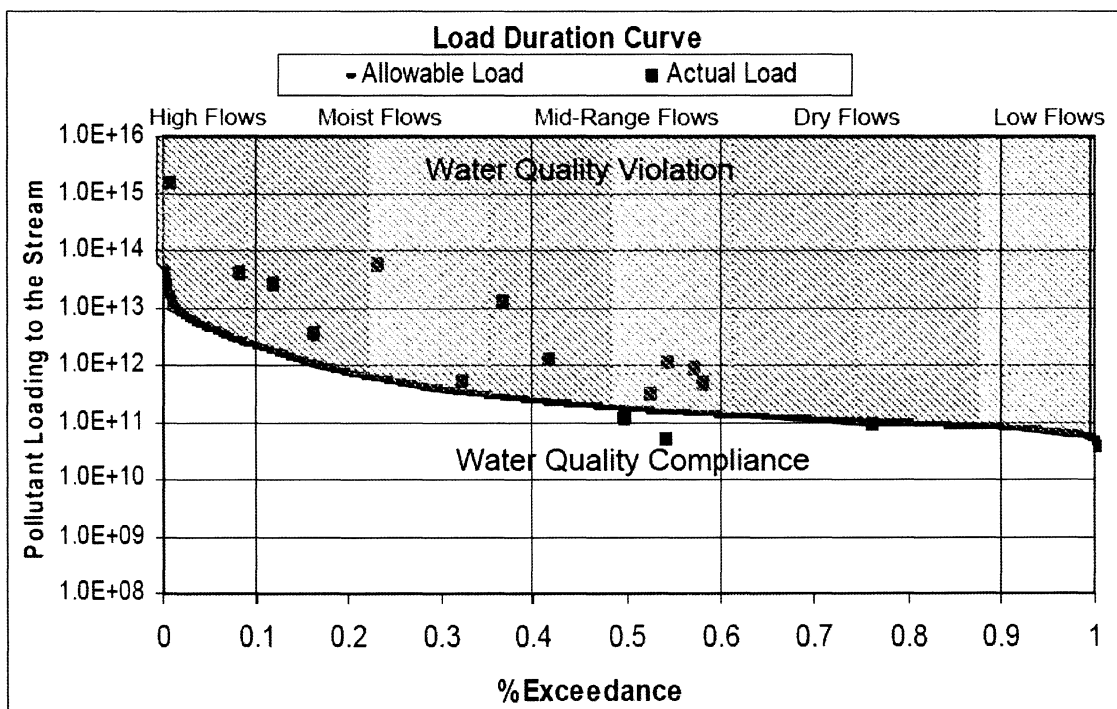
Figure 2-2. Example of a load duration curve

Water quality standard violations that occur near the high flow category are associated with rainfall events with the pollutant source characterized as non-point. In contrast, violations that occur near low flows are associated with dry weather with the pollutant sources characterized as point sources (USEPA, 2007). This technique has been used in addressing water quality concerns in Total Maximum Daily Load (TMDL) programs (Cleland, 2003).

2.2. Data and Methods

Streamflow data were acquired from the United States Geologic Survey (USGS, 2010) and water quality data collected from 2000 to 2008 by the City of Houston, was acquired through the Houston-Galveston Area Council (H-GAC, 2008) water quality monitoring website from the water quality monitoring stations in Figure 1. The water quality monitoring stations were located with USGS streamflow gauges on bridges across the streams in the downstream portion of the watershed. The water quality parameters used for analysis are listed in Table 2-5, along with summary statistics by stations. For the multivariate analysis, a dataset for each watershed was amassed with the water quality data sorted according to month and the median value for each water quality parameter calculated for each monitoring station. The mean of the monitoring station median values for each water quality parameter was determined and a z-scale transformation was applied to the monthly medians data for each parameter so that the variables have a zero mean and unit variance, a requirement for principle component analysis (Ouyang et al, 2006). The normality of the transformed variables was then tested using the Kolmorogov-Smirnov test with 95% or higher confidence. The variable Percent Flow

Exceedance is the percent exceedance of the streamflow associated with the streamflow at the time of water quality sample collection. This provides a measure of the streamflow condition that can be compared between watersheds with differences in flow.

Table 2-5. Variables assessed in multivariate analysis

| | Spring Creek Watershed | | | | | | Cypress Creek Watershed | | | | | | | | |
| | 11313 | | | 11314 | | | 11328 | | | 11330 | | | 11332 | | |
Parameter	Count	Median	Std Dev	Count	Median	Std Dev	Count	Median	Std Dev	Count	Median	Std Dev	Count	Median	Std Dev
Flow (CMS)	72	2.2	29.1	65	0.7	11.3	198	1.8	15.5	21	1.7	7.9	102	0.5	9.0
% Flow Exceedance	72	0.4	0.3	65	0.3	0.3	198	0.1	0.1	21	0.4	0.3	102	0.5	0.3
Temperature (°C)	72	21.9	6.2	65	20.6	5.9	194	22.3	6.3	21	25.8	4.5	101	21.7	6.4
Conductivity (µΩ)	-	-	-	-	-	-	126	586.5	274.1	21	677.0	258.2	92	433.5	244.4
Dissolved Oxygen (mg/L)	60	7.2	2.3	51	6.9	2.7	182	7.7	1.9	20	7.2	1.1	98	7.4	1.9
pH	47	7.6	0.7	56	7.8	0.6	151	7.7	0.5	19	7.8	0.5	97	7.6	0.4
Total Alkalinity (mg/L)	21	86.0	48.0	32	46.0	24.8	-	-	-	-	-	-	-	-	-
Total Suspended Solids (mg/L)	40	36.0	98.0	47	19.0	27.4	141	29.0	63.9	40	16.5	49.0	77	21.0	47.8
Total Dissolved Solids (mg/L)	-	-	-	-	-	-	106	427.0	132.4	15	475.0	131.5	78	333.5	123.4
Total Organic Carbon (mg/L)	26	9.7	4.1	37	10.0	4.3	-	-	-	-	-	-	-	-	-
Chloride (mg/L)	46	44.8	26.1	56	33.0	18.0	144	60.0	35.5	21	86.0	35.0	97	46.0	30.7
Sulfate (mg/L)	46	12.1	6.8	56	7.0	9.5	193	19.0	8.6	21	20.0	7.5	97	14.0	6.7
Ammonia (mg/L)	-	-	-	-	-	-	128	0.1	0.1	21	0.1	0.1	97	0.1	0.2
Nitrate (mg/L)	46	2.0	2.8	56	0.3	1.1	122	4.3	4.0	18	6.0	3.9	59	2.9	2.4
Total Phosphorus (mg/L)	44	0.7	0.5	50	0.2	0.3	64	1.5	0.8	18	1.8	1.0	78	0.9	0.6
Ortho-Phosphorus (mg/L)	45	0.3	0.4	54	0.0	0.3	-	-	-	-	-	-	-	-	-
E. coli (MPN/dL)	32	229.0	1041.1	40	277.5	3602.4	121	720.0	10121.2	19	820.0	1885.1	95	213.0	3649.0

* Variables used in analysis reflect the available data.

2.2.1. *Multivariate Analysis*

The statistical software SAS (SAS, 2003) was used to perform each of the multivariate analyses. First, principal component analysis was applied to the normalized dataset of each watershed in order to identify the underlying factors having the most influence on the variability of the dataset. The first step in principal component analysis was to calculate the correlation matrix (Bengraine and Marhaba, 2003) using the transformed dataset. Then eigenvector decomposition was performed on the correlation matrix (Morales et al., 1999). The corresponding eigenvectors were used to create the weighted linear combination of variables or principal components (Singh et al., 2004). The number of principal components in each dataset was selected using the Kaiser criteria (Thyne *et al.*, 2004).

The calculated principal components for each month were employed in Ward's clustering analysis, an agglomerative hierarchical clustering technique (Astel et al., 2006). Using squared Euclidean distances as a measure of similarity, the most similar elements are sequentially grouped in clusters (Alberto *et al.,* 2001). This step-by-step method uses analysis of variance (ANOVA) to minimize the sum of squares of the potential clusters at each step (Zhou, et al. 2007). The optimal number of clusters was evaluated using the local maxima of the pseudo-F statistic for selection of the appropriate level of clustering (DeGaetano, 1996).

Cluster membership was used with the original transformed data in forward stepwise

discriminant analysis in order to identify the variables most influential to determining

cluster membership. Discriminant analysis builds linear functions using the most

influential variables to predict cluster (Muxika et al, 2007) membership by sequentially

adding each variable to the function in a forward stepwise procedure (Shrestha et al.,

2008). At each step the influence of the variable to the predictive power of the

discriminant function is assessed and the variables reducing the predictive power

removed form the function. This is evaluated using an F-test at each step with a

threshold of 0.15 for addition or deletion of a variable to the discriminant function and

the process stopped until no variables can be added or deleted.

These variables identified by discriminant analysis to have the most influence on cluster

membership were then utilized to repeat principal components and cluster analyses, in

order to refine the cluster membership. The final clusters were then analyzed with the

original transformed data using Duncan's Multiple Range test. The cluster means for

each variable in the discriminant function were then compared at the 95% confidence

level. Then based on the mean comparison the clusters were characterized as having

high, medium, or low values for each parameter.

2.2.2. *Targeted Load Duration Curves*

The final cluster membership was then used to group the raw water quality data into

"seasons". Streamflow data were obtained from the USGS for the most downstream

gauge listed in Table 2-2, for the period 2000 through 2008 and the percent accidence

calculated by ranking the samples, largest to smallest, and then dividing by the total number of samples. The streamflow data were then multiplied by the water quality standard in order to calculate the allowable pollutant loading into the stream. The allowable load was based on the Texas Commission on Environmental Quality (TCEQ) adopted water quality geometric mean standard for streams of 126 MPN/dl (Most Probable Number per 100mL) for *E. coli*, and the screening criteria of 0.69 mg/l and 1.95 mg/l for total phosphorus and nitrate (2008a). These parameters were chosen for LDC analysis because these pollutants were identified as watershed impairments and concerns.

The raw water quality data reported by the Houston- Galveston Area Council for the downstream water quality stations (See Figure 2-1) were then used with the corresponding USGS reported streamflow measurement at the time of water quality sample collection to calculate the actual load. The actual loads were then segmented according to the clusters of months determined by cluster analysis for each watershed. For each watershed and cluster of months, load duration curves were developed for *E. coli*, nitrates, and total phosphorus loads. The curves were then examined for season specific source characterization. If the actual loading exceeded the allowable loading near dry to low flow conditions in the flow exceedance range greater than 60% , the pollutant sources were characterized as point sources. On the other hand if the violation occurred near moist to high flow conditions, or in an flow exceedance range less than 40%, the pollutant sources were characterized as non-point sources. In the cases where load duration curve had violations in multiple flow conditions the pollutant sources were characterized as both point and non-point. The violations were examined to identify the

flow condition in which a majority of the violations occur. When a preponderance of violations occur in the moist to high flow conditions, the pollutant sources are characterized as mostly non-point with some point sources. Likewise, when the preponderance of the violations occur in the dry to low flow conditions with some violations in the higher flow conditions, the pollutant sources are characterized as mostly point sources with some non-point sources present. When the violations are more evenly present near both the high and low flow conditions, the sources are characterized as either point and non-point or non-point and point, depending on which condition in which there were more violations.

The raw water quality dataset was sorted according to the seasons determined by the cluster analysis. Pairwise linear correlation coefficients were then calculated for each season's raw water quality data with complete records for the variables of streamflow, total suspended solids, total dissolved solids, nitrate, total phosphorus, chloride,and *E. coli* concentration. The parameters that were found to be significantly correlated, at $p \leq 0.05$, were compared between the seasons to evaluate the temporal variation of water quality parameters (Wu et al., 2009) .

2.3. Results

A summary of the results of multivariate analysis guided source characterization based upon the LDCs for both watersheds is presented in Figure 2-3. Correlations between water quality parameters in each season are presented in Table 2-6.

Source Characterization

(a)

Cluster 1	E. Coli	Non-Point
	Nitrates	Non-Point & Point
	Total Phosphorus	Point

Cluster 2	E. Coli	Mostly Point
	Nitrates	Point & Non-Point
	Total Phosphorus	Mostly Point

Outliers	E. Coli	Non-Point
	Nitrates	Point
	Total Phosphorus	Point

(b)

Cluster 1	E. Coli	Non-Point & Point
	Nitrates	Point
	Total Phosphorus	Non-Point & Point

Cluster 2	E. Coli	Point & Non-Point
	Nitrates	Non-Point & Point
	Total Phosphorus	Point & Non-Point

Outliers	E. Coli	Non-Point
	Nitrates	Point
	Total Phosphorus	Point

Figure 2-3.Spring Creek (a) and Cypress Creek (b) season specific characterization LDC results

Table 2-6 .Seasonal correlation between water quality variables using raw data from the downstream water quality stations

Variables		Spring Creek (11313)			Cypress Creek (11328)		
			Cluster			Cluster	
		Outliers	1	2	Outliers	1	2
	Months	Aug, Nov	Dec,Jan, Feb, Mar	Apr, May, Jun, Jul, Sep, Oct	Feb, Dec	Jan,Apr, Jun, Aug, Sep	Mar, May, Jul, Oct, Nov
# Samples		7	16	24	8	28	30
Flow	Total Suspended Solids	0.97*	0.55*	0.21	0.51*	0.63*	0.47*
	Total Dissolved Solids	No Data	No Data	No Data	-0.76*	-0.62*	-0.59*
	Nitrate	-0.72*	-0.48*	-0.47*	-0.66*	-0.42*	-0.41*
	Total Phosphorus	-0.51	-0.35*	-0.43*	-0.65*	-0.28	-0.53*
	Chloride	-0.85*	-0.38*	-0.58*	-0.67*	-0.42*	-0.63*
	E. coli	-0.25	0.17	0.003	0.58*	0.75*	0.79*
Total Suspended Solids	Total Dissolved Solids	No Data	No Data	No Data	-0.57*	-0.49*	-0.62*
	Nitrate	-0.73*	-0.58*	-0.33	-0.69*	-0.30	-0.36*
	Total Phosphorus	-0.63	-0.34	0.14	-0.65*	-0.28	-0.52*
	Chloride	-0.98*	-0.56*	-0.38*	-0.52	-0.41*	-0.59*
	E.coli	-0.26	0.61*	-0.10	0.78*	0.97*	0.61*
Total Dissolved Solids	Nitrate	No Data	No Data	No Data	0.98*	0.99*	0.95*
	Total Phosphorus	No Data	No Data	No Data	0.95*	0.74*	0.96*
	Chloride	No Data	No Data	No Data	0.94*	0.87*	0.86*
	E.coli	No Data	No Data	No Data	-0.36	-0.41*	-0.40*
Nitrate	Total Phosphorus	0.85*	0.83*	0.59*	0.94*	0.86*	0.56*
	Chloride	0.83*	0.94*	0.89*	0.98*	0.96*	0.69*
	E.coli	-0.87*	0.26	-0.37	0.38	-0.18	-0.15
Total Phosphorus	Chloride	0.83*	0.78*	0.74*	0.85*	0.89*	0.91*
	E.coli	-0.70*	0.35	-0.34	0.19	-0.11	-0.13
E.coli	Chloride	0.40	0.36	-0.41	0.03	-0.19	-0.12

* p<0.05

2.3.1. Spring Creek Watershed

Spring Creek cluster analysis divides the temporal data into three clusters of months

(Figure 2-3a). The first cluster includes the cool weather months of December, January,

February, and March; which Duncan's Multiple Range Test identifies as having high *E.*

coli concentration in comparison to other seasons. The second cluster includes the warm

weather months of April, May, Jun, July, September, and October. The outlier months

were August and November. The outlier months were identified as statistically different

than the rest of the dataset as defined by an extreme standard score with less than a 10%

probability density. Comparatively, Houston's climate generally has a short cool season

and a long warm to hot season with most of rain falling in late spring and early fall (NOAA, 2010). The rainy season generally includes months in the second cluster. Discriminant analysis shows that *E. coli*, temperature, sulfate, total suspended solids, and percent exceedance accounted for 85% of the variability between these clusters.

Using the seasons determined by multivariate analysis, season specific load duration curves were developed for *E. coli*, nitrate, and total phosphorus loading. For *E. coli* loading into Spring Creek, the warm weather months (Cluster 2) months show violations of the water quality standard mostly in the mid-range to dry conditions, leading to a characterization of mostly point sources with some non-point sources (Figure 2-4a). Sources of nitrates were characterized as both point and non-point sources and total phosphorus sources were characterized as mostly point sources (Figure 2-3a). This characterization is supported by a weak negative correlation between flow and nutrients and a lack of correlation between flow and *E. coli* (Table 2-6).
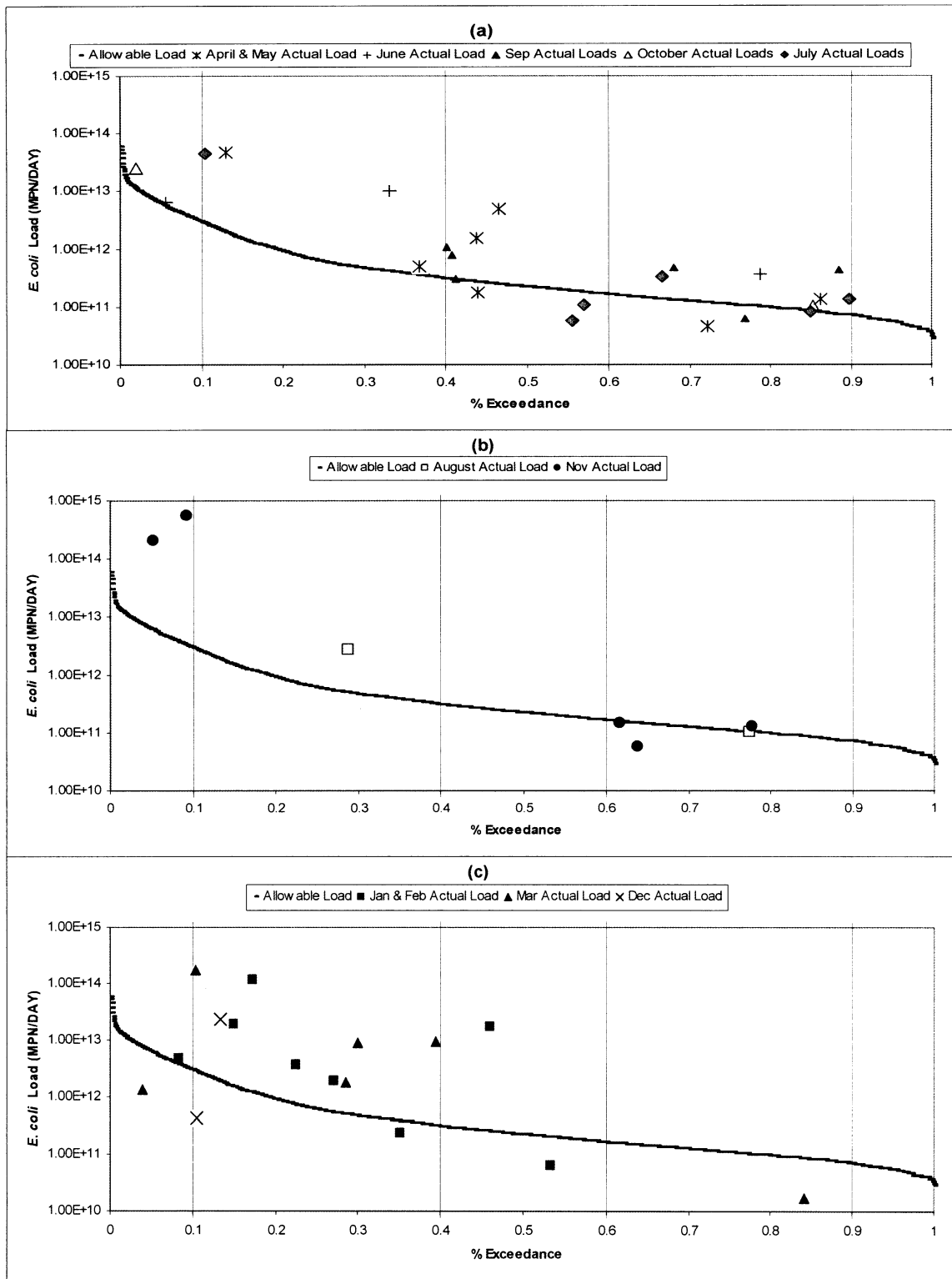
Figure 2-4. Load duration curves for characterization of *Escherichia coli* sources in Spring Creek in (a) Cluster 2, (b) Outliers, and (c) Cluster 1.

The load duration curve for the outlier points show a majority of the violations in the high flow and moist conditions (Figure 2-4b) leading to a characterization of non-point sources. The sources of nutrients are characterized as point sources (Figure 2-3a). This characterization is supported by a lack of correlation between flow and *E. coli* and a negative correlation between flow and nutrients (Table 2-6).

The cool season months have violations of the water quality standard primarily in the high flow to moist conditions (Figure 2-4c) leading to a characterization of *E. coli* sources as non-point. The nutrient sources in the cool season months were characterized as point sources with some impact from non-point sources (Figure 2-3a). The characterization is supported by a negative correlation between nutrients and flow along with a strong positive correlation between nutrients and chloride (Table 2-6) an indication of a common source for both nutrients and chloride, which traditionally has been associated with waste water contamination (Sawyer et al., 2006), a key point source within the watershed .

2.3.2. *Cypress Creek Watershed*

According to multivariate analysis, Cypress Creek water quality data is best classified into three clusters or seasons (Figure 2-3b). The first cluster includes the months of January, April, June, August and September. The results of Duncan's Multiple Range Test characterize this season as having "low streamflow" in comparison to the other clusters of months. The second season includes the months of March, May, July, October, and November and has low phosphorus concentrations in comparison to the

other seasons. The outlier months are February and December, which are characterized as high flow and high *E. coli* concentrations in comparison to the other seasons. This non-intuitive segmentation, which does not follow climatic variation, is thought to be the result of the flow conditions at which samples were taken in these periods, which is slightly skewed to mid-range flows. Discriminant analysis determined that the parameters of dissolved oxygen, flow, sulfate, conductivity, *E. coli* concentration, total suspended solids, and total phosphorus concentration were the most discriminating parameters to sort data into seasons and thus accounted for approximately 83% of the variability between the seasons.

Analysis of the LDC for total phosphorus and nitrate in the outlier of February and December months leads to a point source characterization, whereas *E. coli* sources for the same period were characterized as non-point. This was supported by analyzing the correlations between water quality parameters (Table 2-6). Flow and nutrients were strongly negatively correlated while flow and *E. coli* were positively correlated indicating that low flows occurred with high nutrient loading and high flows occurred with high *E. coli* loading. At the same time, chloride and total dissolved solids were negatively correlated to flow and positively correlated to nutrients. Thus the positive correlation between chloride and nutrients supports a common source between chloride and nutrients. As such, the lack of correlation between *E. coli* and chloride along with a positive correlation among total suspended solids, flow, and *E. coli* indicates different sources for *E. coli* and chloride for the outlier months.

The load duration curve for the first cluster on months reveals violations of the nitrate criterion primarily in the mid-range flows and dry conditions leading to a source characterization of mostly point sources with some non-point sources. A negative correlation between flow and nitrate indicates a point source. Furthermore, a positive correlation between nitrate and chloride indicates a common source between nitrates and chloride. In contrast, the load duration curves for *E. coli* and total phosphorus show violations of the water quality standard throughout all flow conditions. Thus the sources are characterized as both non-point and point.

The load duration curves for total phosphorus and nitrate loading in the second cluster of months show that the water quality criteria is violated primarily in the dry and mid-range flow with some in the moist conditions. Thus the sources of phosphorus and nitrate are characterized as both point and non-point. Likewise, the *E. coli* violations occurred throughout all flow conditions, leading to a mixed point and non-point source characterization. Furthermore the correlation between water quality parameters does not provide strong support of a distinct characterization of sources.

2.4. Discussion

The process of developing season specific load duration curves based upon multivariate analysis provided a framework for characterizing sources in Spring Creek, where as the results for Cypress Creek highlight the need for improved sampling programs that takes into account flow condition when water quality samples are taken in order to address both

low flow and high flow conditions. Both point and non-point sources were identified as being active in specific time periods. In both watersheds the outlier months exhibited point sources for nutrients and non-point sources for *E. coli*. In addition, *E. coli* concentration was a discriminating variable between the seasons for both watersheds. However, total phosphorus concentration was a discriminating variable for differentiating between seasons in Cypress Creek but not Spring Creek. This is reflected in the point source characterization of phosphorus loading to Spring Creek in all of the seasons. In contrast, Cypress Creek has non-point sources of phosphorus present in both seasons, but not in outlier months.

When examining the correlations between water quality variables in each season, both watersheds show positive correlation between nutrients and chloride (Table 2-6). In addition, for a majority of the clusters, there is positive correlation between flow and total suspended solids as well as negative correlation between flow and nutrients. Therefore for the majority of the time, both watersheds have nutrient sources associated with waste water treatment plants.

The results suggest that in Spring Creek, where non-point sources are responsible for the violations of the *E. coli* standard during cold weather months, watershed protection plans implementing best management practices should take into account the seasonal variability of vegetation based BMPs. In addition, warm season characterization of point sources of nutrients and *E. coli* suggests that watershed protection plans should consider

bacterial re-growth downstream from wastewater treatment plants, thus allocating greater resources to improve effluent from point source discharges.

However, the results of this analysis are only effective with sampling programs that reflect the range of flow conditions of the stream and take into account seasonality. The results of the analysis can identify groups of months in which the sampling disproportionately represents a particular flow condition. For instance, Figure 2-4c, shows that the sampling in cluster 2 of Spring Creek disproportionately represents the higher flows. The improved sampling program should include both routine low flow sampling and storm sampling that encompassed the wide range of flows of storm hydrographs. Thus the presented framework allows for the identification of components of the sampling program that can be improved.

2.5. Summary and Conclusions

Two watersheds near Houston, Texas were compared through a framework of multivariate analysis and load duration curves. Water quality data were analyzed using principal component, cluster, and discriminant analysis. The data were segmented into clusters or "seasons" of months with similar water quality conditions that are characteristic of the watershed. Duncan's multiple range test was used to compare the water quality data between these clusters through identification of parameters which distinguish each cluster. These custom seasons were then used to target load duration curves to characterize pollutant sources. Correlations between parameters in each season were then examined to further characterize the pollutant sources. Spring and Cypress

Creek watersheds were compared through this framework in order to understand the influence of seasonality and source characterization for each of the watersheds. Water quality data describing Spring Creek were segmented into three identifiable seasons that reflect the climate of the region and the pollutant sources characterized for *E. coli* and nutrients. In contrast, the seasons determined for Cypress Creek were less intuitive and do not reflect the climate of the region.

Water quality datasets, while often containing large amounts of data, are comprised of observations in a limited set of stream conditions. In order to select appropriate study areas for the development of water quality models, watersheds should be evaluated to understand the similarities and differences. The presented framework of analysis provides a method for identification of seasonality and characterization of pollutant sources. In addition, the understanding gained through this framework of comparison can be used to evaluate and improve the sampling efforts currently in place. The application of this comparison has provided the basis for future hydrologic and water quality modeling efforts in an attempt to better understand stream processes as they impact the poor water quality of Lake Houston.

Chapter 3 : Radar Rainfall Application in a Distributed Hydrologic Modeling for Cypress Creek Watershed, Texas

Aarin Teague, Jason Christian, and Philip Bedient
Submitted to the Journal of Hydrologic Engineering

Abstract

Recent advances in hydrologic models have depended on the use of radar rainfall input in physically based, fully distributed models. Previous research conducted for case studies near Houston, Texas have focused on the use of radar rainfall for large storm events such as tropical storms and hurricanes. A fully distributed model, $Vflo^{TM}$, was used to model streamflow during small storm events in the Cypress Creek Watershed, near Houston, Texas. Two events were simulated both with rain gage corrected radar data and exclusively with rain gages, while a third event was modeled exclusively with rain gage data. The modeled streamflow was then compared, using peak streamflow, time to peak, and volume streamflow, to the USGS observed streamflow to evaluate the model performance between radar and rain gage input. A comparison of the models for the events shows that the radar input results better match the observed streamflow for the streamflow volume and peak streamflow.

3. Introduction

A significant goal of current research to improve the use of distributed hydrologic modeling is to improve the ability to accurately predict and simulate streamflows (Carpenter et al., 2004). Several studies have identified the high degree of sensitivity that such hydrologic models have to the rainfall inputs (Sunet al., 2000 , Carpenter et al, 2001,

Koren et al., 1999). Radar rainfall provides high spatial and temporal resolution input data (Borga, 2002) that, when corrected for bias, has been shown to improve the accuracy of hydrologic model performance (Vieux and Bedient, 2004).

Hydrologic models are typical tools in the development of watershed protection plans, providing simulation of rainfall-runoff processes. Fully distributed hydrologic models provide the ability to simulate the spatial variability of hydrologic processes over the landscape of a watershed (Yilmaz et al., 2008). As such, fully distributed hydrologic models are important tools for assessing (1) the effects of land-use change, (2) the influence of geospatial inputs, and (3) the movement of pollutants and sediment (Smith et al., 2004). When combined with radar rainfall, fully distributed hydrologic models improve the accuracy of hydrologic prediction (Vieux et al., 2009). Recent advances in computing resources and availability of radar rainfall data have allowed for improvements in rainfall-runoff modeling (Delrieu et al., 2009).

The objective of this study was to present a case study comparing the performance of a fully distributed physically based hydrologic model using either radar rainfall or rain gage data input for Cypress Creek Watershed, near Houston Texas. The hydrologic model was calibrated and evaluated specifically for small storms, which have been poorly studied past hydrologic research. This watershed is poorly monitored by rain gages (Figure 3-1); thus radar rainfall provides greater density of rainfall data input for the fully distributed model.
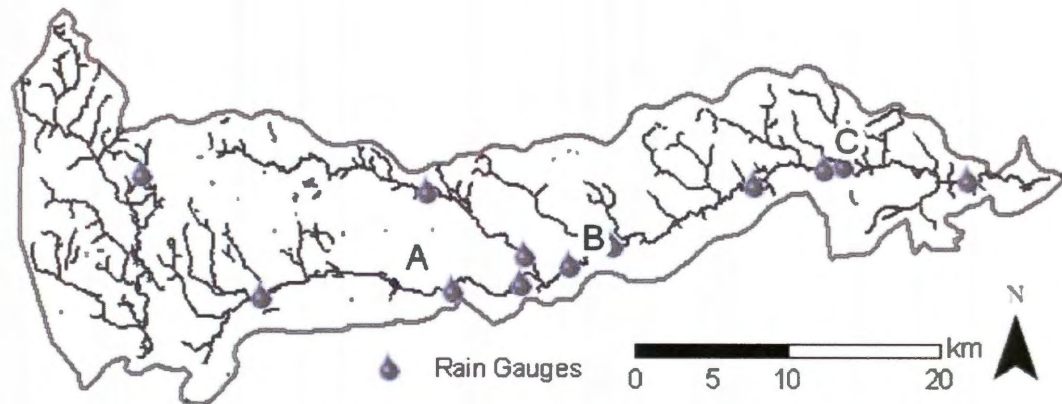
Figure 3-1.Rain gage network in Cypress Creek (rain gage density for area draining to A is 182km2 per rain gage, rain gage density for area draining to B is 79 km^2 per rain gage, and rain gage density for area draining to C is 67 km^2 per rain gage)

3.1. Background

3.1.1.Radar Rainfall

Radar derived rainfall data have been used in numerous applications for hydrologic modeling, including the use of a real-time flood alert system (Fang et al, 2009 ; Sharif et al, 2010), watershed assessment models for Total Maximum Daily Loads (TMDLs) (Wu et al, 2010), and pollutant transport models (Shaw et al, 2010). It has been found that radar rainfall provides a more accurate computation of the rising limb and peak streamflow of hydrographs than rain gages (James et al, 1993). Furthermore, streamflow volumes are better matched with the use of bias corrected radar data than with rain gages alone (Einfalt et al., 2004).

An important advantage of radar rainfall is that it provides information on the spatial distribution of rainfall, allowing for its use in fully distributed hydrologic models

(Carpenter et al, 2001). However radar rainfall alone is unable to accurately capture rainfall amounts (Kim et al, 2008). In order to correct for error in raw radar rainfall data, the radar must be calibrated to rain gages (Seo, 1998). The estimate of rainfall by radar when compared to rain gages can be biased by rain gage errors, radar errors, or the inherent difference between point estimates by rain gages and aerial estimations by radar scans (Ahnert, 1983). Evaluation of radar estimates with gage measurements have shown that large rainfall events tend to be underestimated where as small rainfall amounts are overestimated (Seo and Breidenbach, 2002).

Next Generation Weather Radar (NEXRAD) radar is provided by the National Weather Service (NWS) through an array of Weather Surveillance Radar-1988 Doppler (WSR-88D). NEXRAD uses a 10 cm wavelength to record reflectivity, radial velocity, and spectrum width of the reflected wave (Bedient et al., 2000). The measured reflectivity is used to calculate rainfall rate using the empirically based Z-R relationship (Krajewski and Smith, 2002). Vieux et al. (1998) showed that precipitation in the Houston area is adequately represented by the tropical Z-R relationship (Rosenfeld, 1993),

$$Z = 200R^{1.2} \quad , \tag{3-1}$$

where Z is the reflectivity (mm^6m^{-3}) and R is the rainfall rate ($mmhr^{-1}$). This approach has been used to reconstruct multiple large events in the Houston region (Bedient et al, 2007). Bias in the radar estimations is corrected by comparing the 24 hour rain accumulation estimated by the radar to that of the rain gage measurements, and then

adjusting the coefficient in the Z-R relationship (Equation 3-1) (Vieux and Bedient, 1998).

3.1.2. Fully Distributed Hydrologic Model

$Vflo^{TM}$ is a fully distributed hydrologic model developed by Vieux et al. as a refinement of $r.water.fea$ (Vieux and Gauer, 1994). It has been used extensively in the Houston and Texas Gulf Coast region to model and predict flooding from extreme rainfall events. The $Vflo^{TM}$ model has previously been used to model a tributary to Cypress Creek, Little Cypress Creek, in order to evaluate flood storage using a 100 year design storm (Fang et al, 2010). Fang et al. used $Vflo^{TM}$ in their flood alert system (FAS) to model real-time response with radar rainfall to forecast flooding in Brays Bayou for the Texas Medical Center of Houston (2009). In addition, $Vflo^{TM}$ models have been developed for White Oak Bayou (Safiolea et al, 2005), Horsepen Bayou, and Clear Creek, all watersheds in and around Houston, Texas. The modeling efforts in Brays Bayou were supported by an extensive rain gage network that monitored a mostly homogenous urban watershed. In contrast Cypress Creek watershed has a complex mix of land use and is relatively poorly monitored by a sparse rain gage network (See Fig 3-1).

$Vflo^{TM}$ uses finite element solutions of the kinematic wave equation for runoff routing. The solution for both overland and channel flow were derived from the Saint Venant equations for unsteady free surface flows. It is derived from the continuity and momentum equations (Borah, 2003). The one-dimensional continuity equation is

$$\frac{\partial Q}{\partial x} + \frac{\partial A}{\partial t} - q = 0, \qquad (3\text{-}2)$$

where Q is the flow rate, A is the cross-sectional area, q is the lateral inflow, x is length, and t is time. The momentum equation is simplified to

$$S_0 = S_f,$$

(3-3)

where S_0 is the slope (length/length) and S_f is the friction slope (length/length). The continuity and momentum equations are used to solve for discharge through

$$q = \alpha Q^\beta,$$

(3-4)

where β for overland flow is assumed to be 5/3. The conveyance factor α is

$$\alpha = \frac{k_m}{n}\sqrt{S_0},$$

(3-5)

where n is the Manning's coefficient, and k_m is the dimensionless kinematic flow number. Overland flow is calculated from the surface flow modeled by Manning's equation as

$$q = \frac{1}{n}S_f^{1/2}Bh^{5/3},$$

(3-6)

where is v is the flow velocity (length/time), S_f is the overland slope (length/length), B is the width of flow (length), h is the depth of flow (length), and n is the Manning's coefficient, which is based on surface characteristics (Vieux, 2004).

Runoff moves from overland cells into channel cells. Open channel flow can simplified to the form

$$q = \frac{\partial Q}{\partial t} + \alpha\beta Q^{\beta-1}\left(\frac{\partial Q}{\partial t}\right)$$

(3-7)

which takes into account the change in the ratio of flow depth to flow width. This formulation can then be solved by finite element analysis, which is an efficient way to transform partial differential equations into ordinary differential equations (Vieux, 2004). By translating the 2-D grid into 1-D finite elements, or partial discretization, the system becomes computationally more efficient. The result is a system of equations for each

element incorporating the boundary conditions of the grid cell, which can then be solved in matrix form by numerical methods.

The *Vflo*^{TM} model solves the Green & Ampt infiltration and saturation excess equations for runoff generation (Vieux, 2004). Geospatial data representing elevation, soils, and land use (Figure 3-2a, b, and c) are incorporated as parameters for the solution of these relationships. Precipitation input can be radar rainfall data, interpolated from rain gage data, or simulated design storms. The model is used to simulate runoff and other hydrologic quantities at any location within the study area, thereby supporting the generation of hydrographs for the selected locations in the watershed.
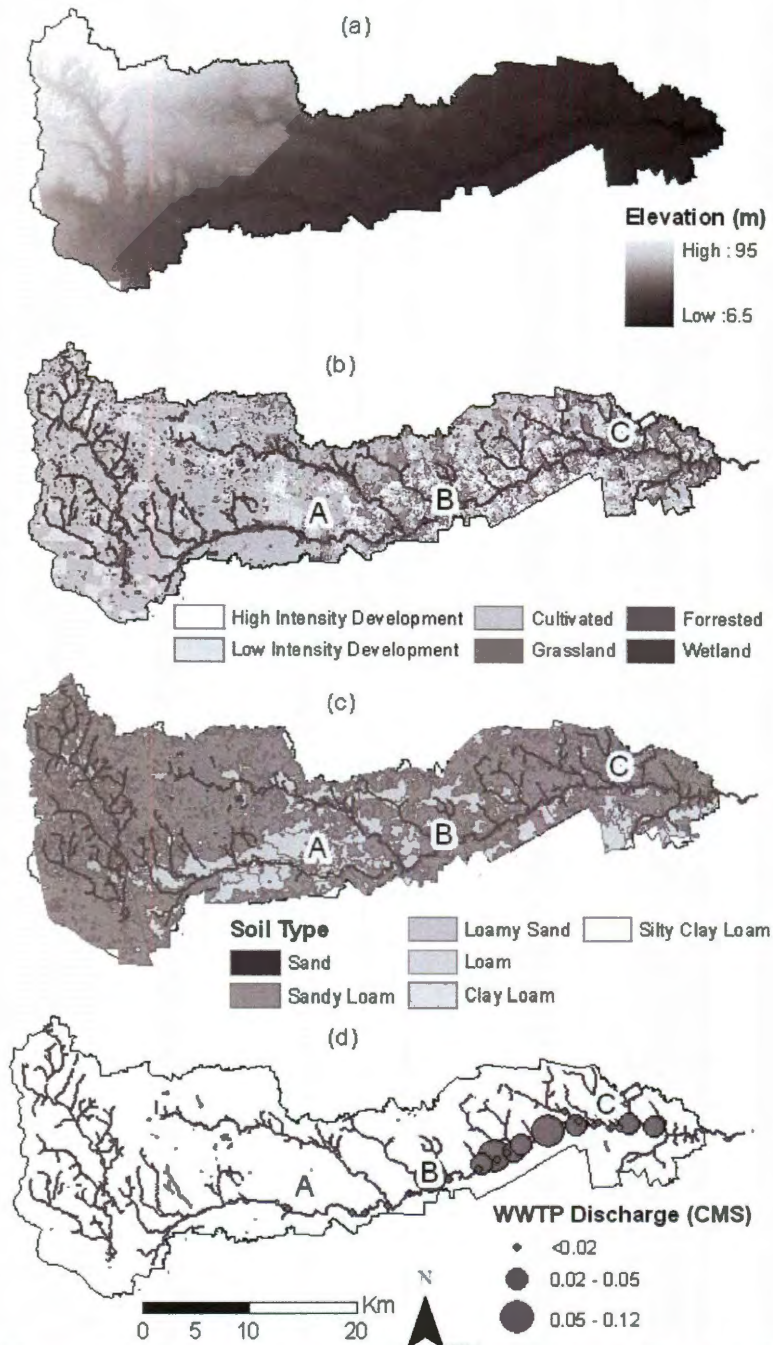
Figure 3-2. Data used to build *Vflo*^{TM} model including (a) elevation, (b) land use, (C) soils, and (d) waste water treatment plant discharge

3.2.   Study Area

Cypress Creek is a 797 km^2 (308 mi^2) watershed located north of the city of Houston and contained in north Harris and Waller counties. It flows for 50 river miles to Lake Houston, the primary source of drinking water for the City of Houston (Chellam, 2008). This watershed is the primary contributor of urban runoff and pollutant loading to Lake Houston (Sneck-Fahrer, 2005). The western upstream part of the watershed is undeveloped primarily as cultivated agricultural fields. The eastern portion of the watershed has primarily residential development and is home to most of 216,000 residents (ESRI, 2000). Cypress Creek watershed is relatively flat with sandy loam soils (Figure 3-2c) which have greater infiltration and less erosion potential. As a result, increases in impervious cover increase runoff.

The watershed has multiple USGS stream gages to monitor streamflow. Three of the gages were used in this study for evaluation and comparison of the modeled streamflow. The first gage USGS 0808740 (Station A in Figure 3-1) receives runoff from 363.4 km^2 (140.3 mi^2) of primarily grassland and agricultural areas. The second gage, USGS 08068800 (Station B), receives runoff from 550.2 km^2 (212.8 mi^2) of grassland, agricultural, forested, and low intensity residential areas. The third gage, USGS 08069000 (Station C), receives runoff from 737.7 km^2 (285.2 mi^2) of grassland, agricultural, forested, and low to high intensity residential areas.

Little Cypress Creek, a sub-area of the Cypress Creek Watershed, was previously modeled using $Vflo^{TM}$ by Fang et al. (2010) in order to assess the necessary flood storage

capacity required by urbanization. This work focused on large storms including a 100 year design storm. The lessons learned and associated datasets from this smaller scale study were instrumental in the modeling efforts of this current study.

3.3. Method

A *Vflo*^{TM} model was developed for Cypress using geospatial datasets as shown in Table 3-1. Each of the datasets was processed into grids consisting of 22 acre cells (or 300 meter on a side) in order to spatially represent the watershed. The 797 km^2 (308 mi^2) watershed is represented by a total of 25,070 cells.

Table 3-1 Cypress Creek *Vflo*^{TM} model data sources.

Data Type	Source	Data Processed
Elevation Data	Lidar -TSARP	Slope
		Flow Direction
		Flow Accumulation
Soils Data	Statsgo	Infiltration
		Hydraulic Conductivity
		Wetting Front
		Soil Depth
		Initial Saturation
		Impervious
Land Use Data	TSARP	Roughness
HEC RAS Cross Sections	TSARP	Channel Geometry
TWDB Lake Evaporation	TWDB	Evapotranspiration
Baseflow	H-GAC Permitted Outfalls, WWTP	

(TSARP -Tropical Storm Allison Recovery Project ;HEC RAS- Hydrologic Engineering Centers River Analysis System ; TWDB- Texas Water Development Board ; H-GAC - Houston-Galveston Area Council ; WWTP- Waste Water Treatment Plant)

A digital elevation model (DEM) created from Lidar data (Figure 3-2a) gathered by the Tropical Storm Allison Recovery Project (TSARP) in 2006 was processed in ArcView

using spatial analyst tools to create a slope grid using the process reported by Fang et al.

(2010) to create a flow direction grid.

$Vflo^{TM}$ uses the Green & Ampt equation (Vieux and Bedient, 2004) to solve for the

infiltration of water through the soil surface. This requires data including hydraulic

conductivity, wetting front, effective porosity, soil depth, initial saturation, abstraction,

and impervious cover. Soils data for Cypress Creek were taken from the NRCS soil

survey (2006), (Figure 3-2c). Using the percentage of sand, silt, and clay for each soil

classification, the effective porosity, wetting front, and hydraulic conductivity were

extrapolated (See Table 3-2). Soil depth was assumed to be the depth of the first layer of

soils as reported in the soil survey.

Table 3-2. Green & Ampt parameters based on soil type

Soil Class	Effective Porosity (cm^3/cm^3)	Wetting Front Suction (cm)	Hydraulic Conductivity (cm/hr)
Sand	0.417	4.95	11.78
Loamy Sand	0.401	6.13	2.99
Sandy Loam	0.412	11.01	10.90
Silty Clay Loam	0.432	27.30	0.10
Loam	0.434	8.89	0.34
Clay Loam	0.390	20.88	0.10

Land use data collected through the TSARP project (2006) (Figure 3-2b) was used to

determine the Manning's overland roughness coefficient, *n*. Roughness ranged from

0.012 to 0.15. In addition, each land use category was assumed to have a percent

impervious value.

The channel was specified by the use of cross section cells, where the most recent detailed cross section surveys from a HEC-RAS (Hydrologic Engineering Center, 2002) model from TSARP (2006) were used to delineate the channel. To simulate the additional streamflow produced by significant wastewater treatment plants, baseflow was added to the cells corresponding to the location of known outfalls. The flow rate of baseflow to act as a surrogate from effluent was estimated from average monthly monitored discharge rates reported by the Houston-Galveston Area Council (H-GAC, 2009). Figure 3-2d shows the locations of the outfalls.

The model was then calibrated at three locations within the watershed as shown in Figure 3-1 (Stations A, B, and C) for two rainfall events July 7, 2009 and September 22, 2009 with radar rainfall data by adjusting the roughness factor.

NEXRAD data collected by the National Weather Service at Dickinson, Texas, was used for the July 7, 2009 and September 22, 2009 events. Reflectivity data were processed by Vieux and Associates, in Norman, Oklahoma, using the tropical Z-R relationship to estimate the rainfall rate (Vieux and Bedient, 1998). In order to ensure the quality of the rainfall rate estimations the estimated R was calibrated to the rain gages in and around the watershed by adjusting of the coefficient in the Z-R relationship. Table 3-3 outlines the agreement of radar and rain gauge data before and after calibration adjustment. Rainfall data were delivered in ascii grids with a resolution of 1 km at 5 minute intervals. The total rainfall depths from the radar data for July 7, 2009 and September 22, 2009 are shown in Figures 3-3 and 3-5 respectively.

Table 3-3. Radar calibration statistics

Start	End	MFB Mean Field Bias	AD Average Difference (%)	CAD Calibrated Average Difference (%)	RD Relative Dispersion (%)
7/7/2009 4:05	7/8/2009 6:00	1.82	39.20	17.40	21.00
9/22/2009 17:05	9/22/2009 5:00	0.61	102.90	11.10	14.10
9/22/2009 5:05	9/22/2009 15:00	1.24	18.90	7.50	9.10
9/22/2009 15:05	9/23/2009 16:00	1.54	34.00	19.30	21.70
9/23/2009 16:05	9/25/2009 6:00	2.04	43.00	12.70	15.80

For each rainfall event the model was run for 48 hours past the end of the rain event. In addition, for three rainfall events, July 7, 2009, September 22, 2009, and August 16, 2010, the model was run using rain gage rainfall data. Rain gage data were acquired from the Harris county Office of Emergency Management (HCOEM) for the rain gages shown in Figure 3-1. The rain gage data were spatially distributed using an exponential weighting function by the $Vflo^{TM}$ model. The total rainfall depths estimated from rain gage data are shown for the three events are in Figures3- 4, 3-6, and 3-7.

For each model run, the results were then plotted as hydrographs to compare the modeled and observed streamflow. Observed streamflow was taken at three USGS stream gages in the watershed: 08068720 (A), 08068800 (B), and 08069000 (C). The total volume of streamflow, time of peak, and peak streamflow was compared to the USGS streamflow observed at the three stream gages. In addition, the runoff depth and rainfall depth were used to calculate the runoff ratio (k) as the ratio of runoff depth to rainfall depth for each event.

3.4. Results

The model results were compared for radar rainfall and rain Gage rainfall input. The hydrographs for the July 7, 2009 rainfall event using radar input are in Figure 3-3, where as the observed versus modeled hydrographs using rain gage data are in Figure 3-4. The observed streamflow and modeled streamflow for the September 22, 2009 event using radar rainfall input is in Figure 3-5 and using rain gage data is in Figure 3-6. The observed and modeled streamflow for the August 16, 2010 event are in Figure 3-7. Table 3-4 shows the differences in total streamflow volume and time difference in peak.
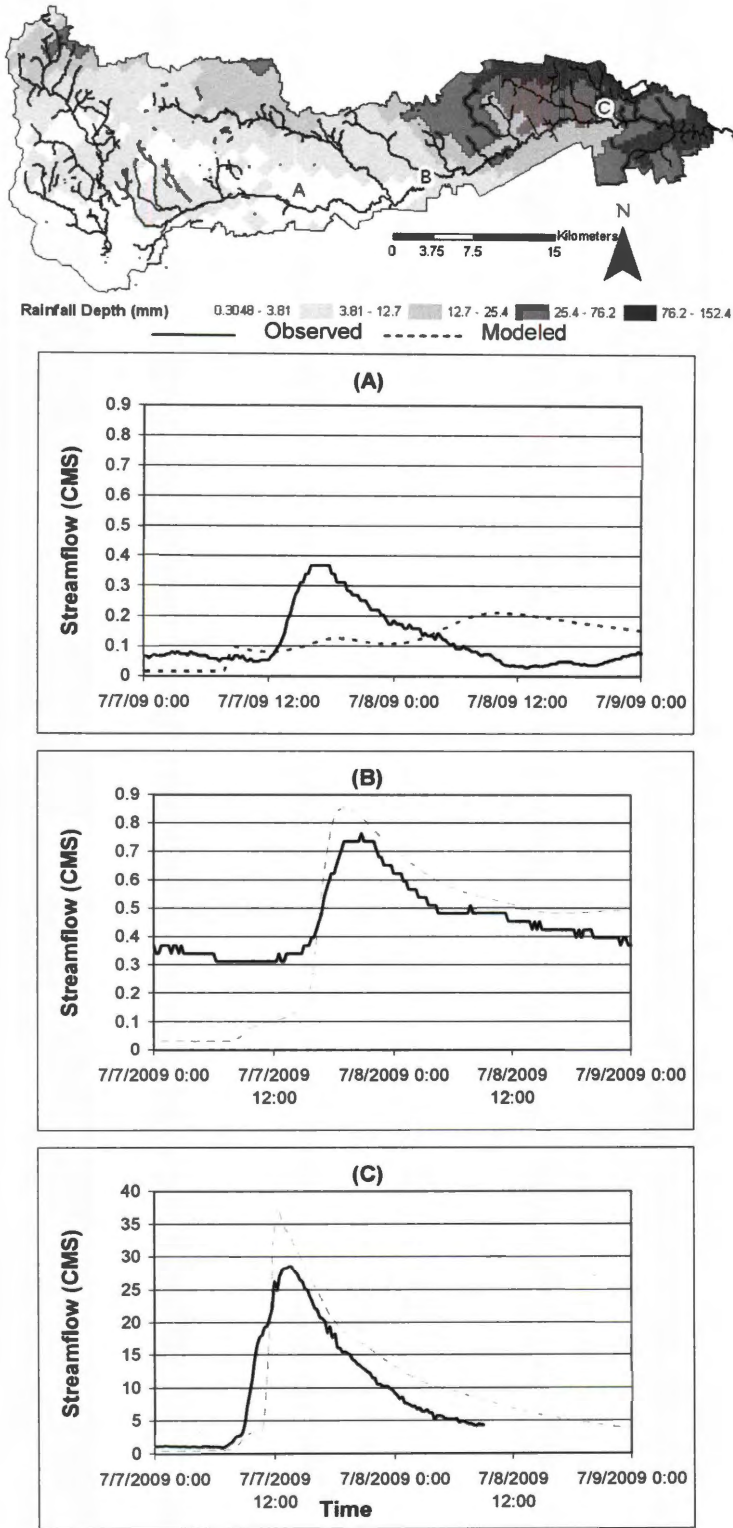
Figure 3-3. Modeled streamflow using radar rainfall for July 7, 2009 for stations (A), (B), and (C)
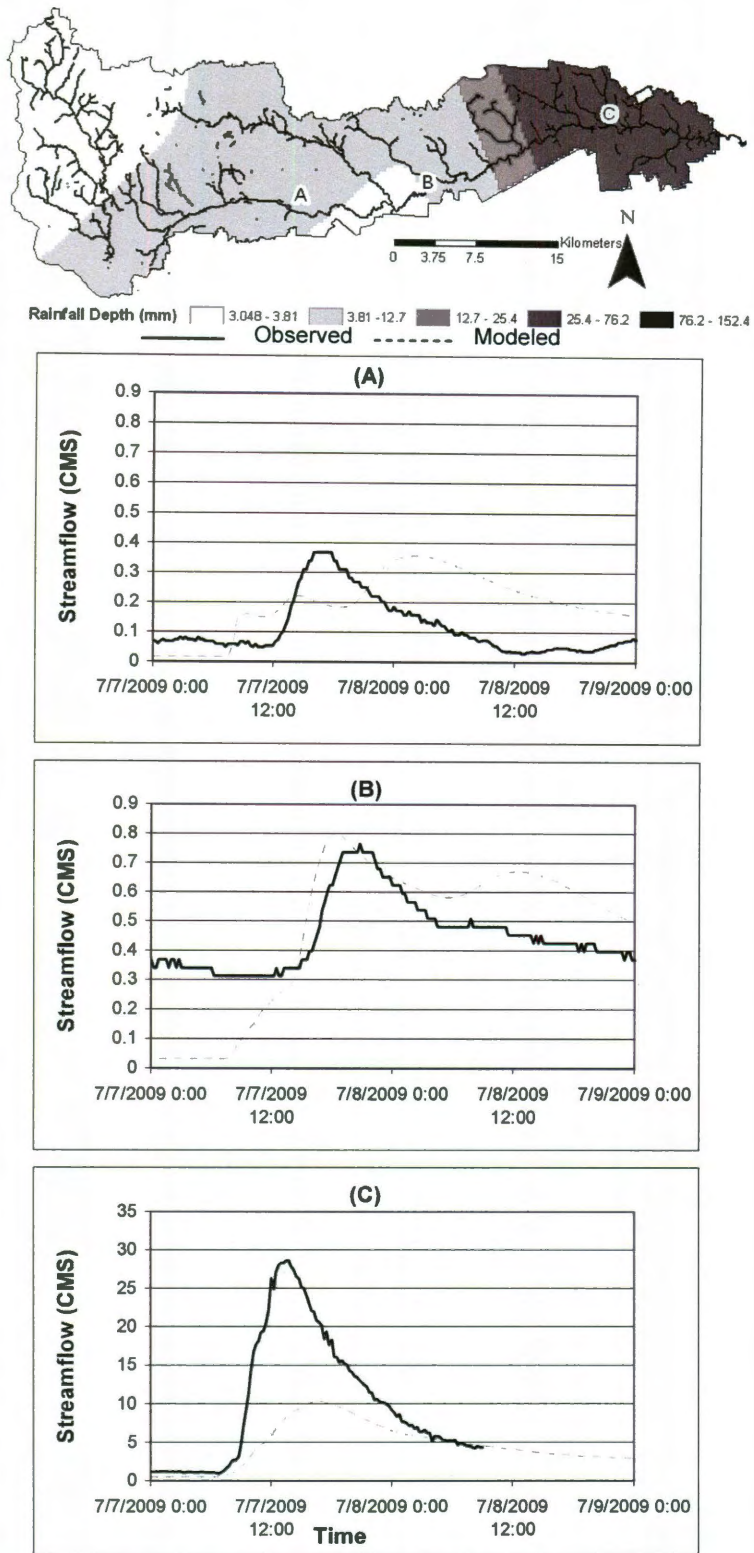
Figure 3-4. Modeled streamflow using rain gage rainfall data for July 7, 2009 for stations (A), (B), and (C)
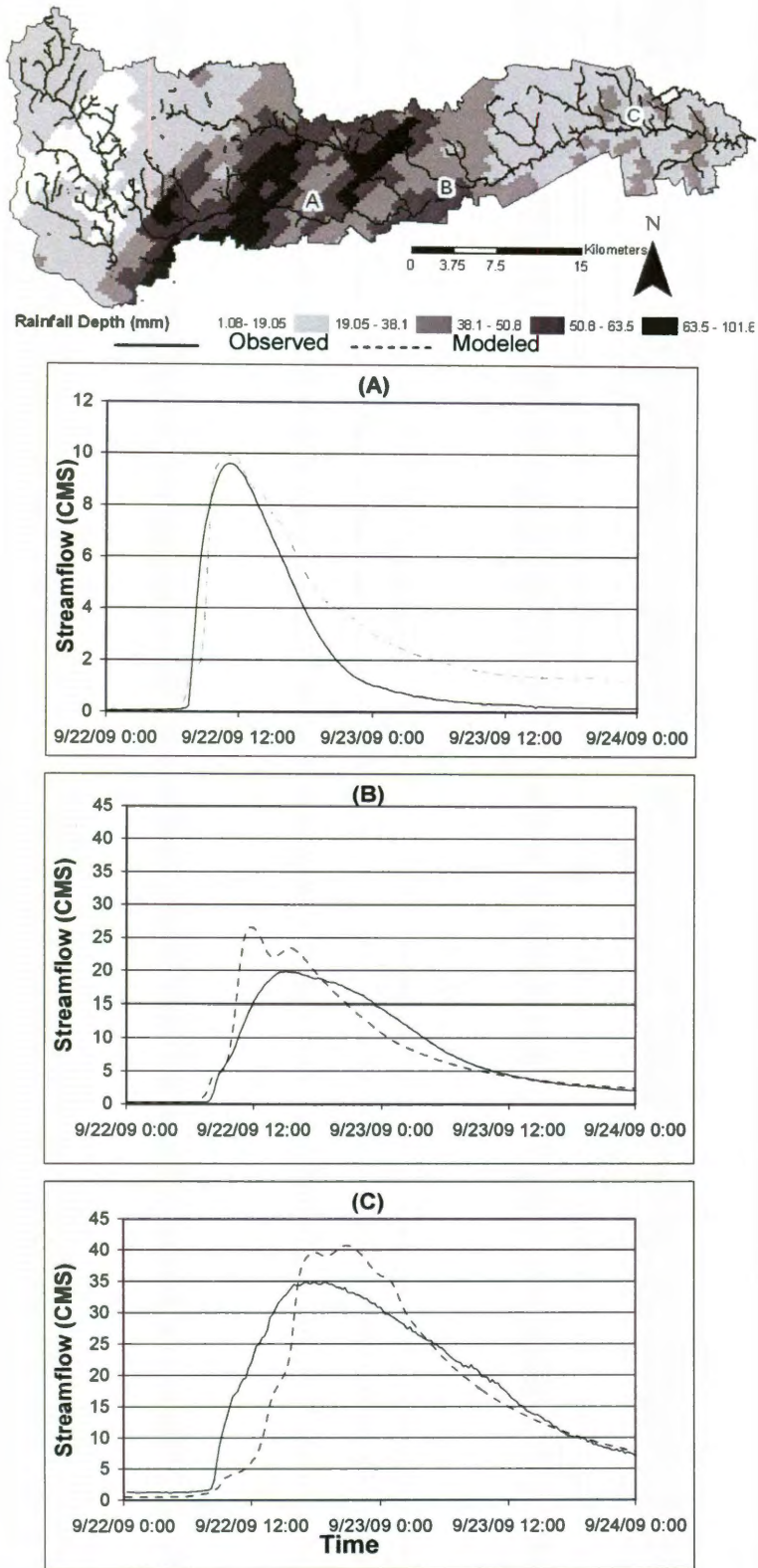
Figure 3-5. Modeled streamflow using radar rainfall data for September 22, 2009 for stations (A), (B), and (C)
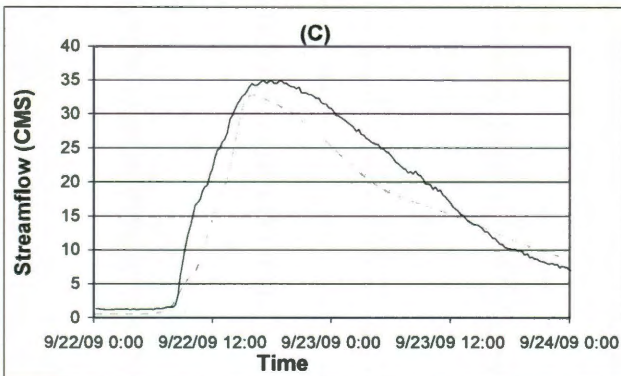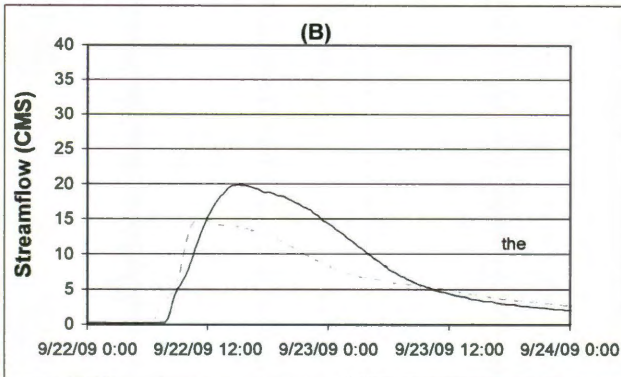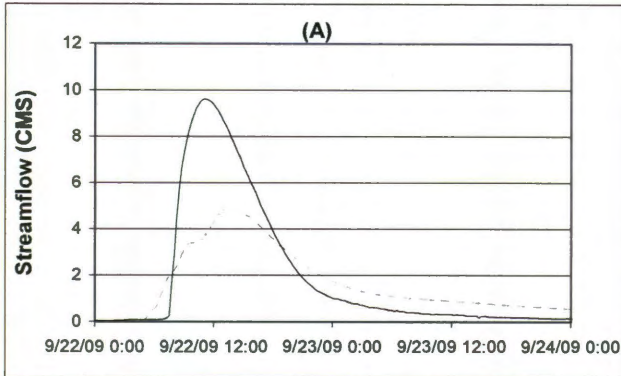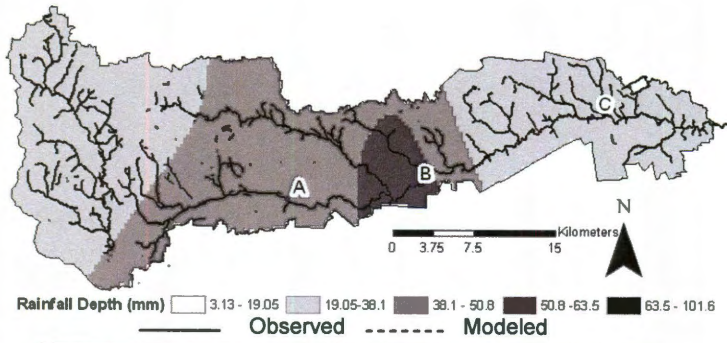
Figure 3-6. Modeled streamflow using rain gage rainfall data for September 22, 2009 for stations (A), (B), and (C)
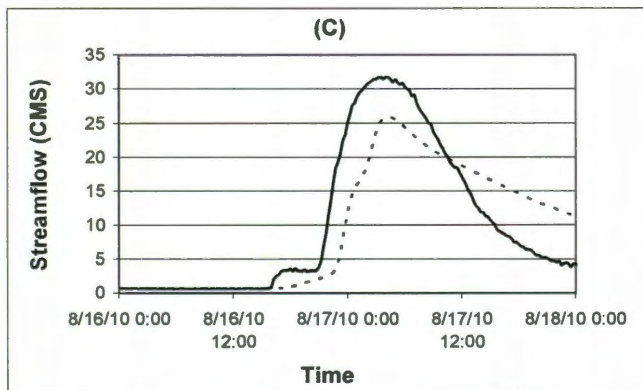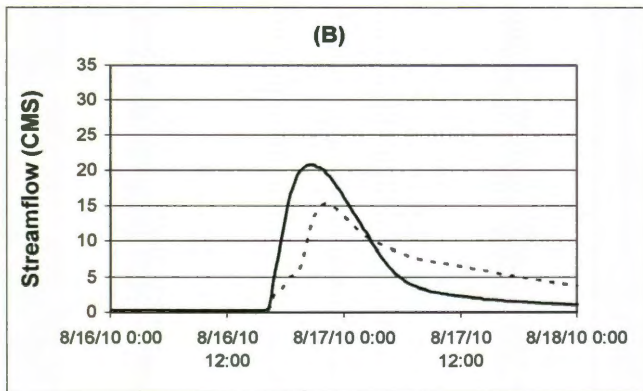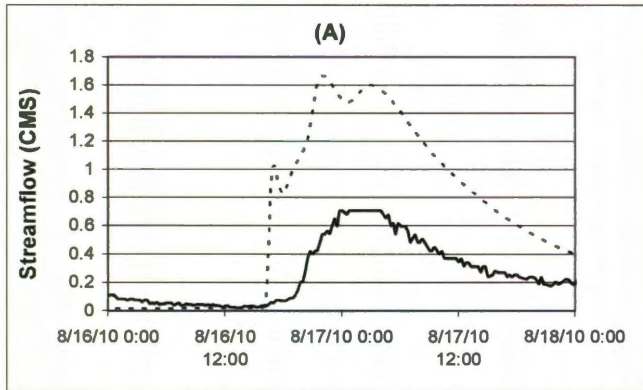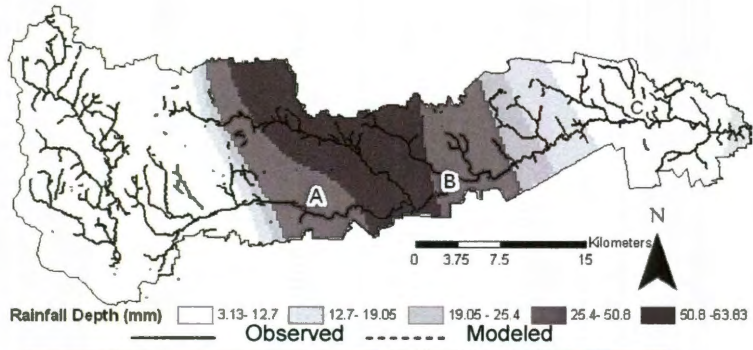
Figure 3-7. Modeled streamflow using rain gage rainfall data for August 16, 2010 for stations (A), (B), and (C

Table 3-4. Comparison of modeled and observed streamflow

Rainfall Event	% Volume Difference			Peak difference (hr)			Runoff Coeffient (mm/mm)		
	Station			Station			Station		
	A	B	C	A	B	C	A	B	C
Radar 7/7/2009	35.96	14.50	4.90	16.33	0.07	3.75	0.05	0.04	0.35
Rain Gage 7/7/09	54.00	12.00	44.00	5.33	7.25	3.33	0.05	0.10	0.40
Radar 9/22/09	43.30	3.00	8.00	0.08	3.25	3.92	0.07	0.10	0.17
Rain Gage 9/22/09	119.19	78.39	70.65	8.92	1.75	5.08	0.03	0.07	0.13
Rain Gage 8/16/2010	35.00	14.56	13.54	1.00	1.75	1.67	0.08	0.17	0.28
Average	57.49	24.49	28.22	6.33	2.81	3.55	0.06	0.10	0.27

The July 7, 2009 event was preceded by 15 days of dry weather. The total rainfall depth with radar input was 0.5 inches (12.7 mm), whereas the rainfall depth with rain gages was 0.28 inches (7.11 mm). Most of the rain fell in the downstream portion of the watershed (See Figures 3-3 and 3-4). With rain gage data, there was 44% difference in volume between the modeled and observed streamflow at the downstream point C. With the use of radar rainfall, the difference in volume was 4.9%. The difference in timing of the peak streamflow was 3.3 and 3.8 hours, respectively. At the midpoint of the watershed (point B), the volume differences were more similar with 14.5% for radar and 12% for rain gage data. The peaks matched well for the radar with a time difference of 0.0695 hours for radar input. There was a peak difference of 7.25 hours for rain gage input. The modeled runoff coefficient for the entire watershed, as estimated from station C, was calculated at 0.35 when radar was used to model the rainfall-runoff processes and was 0.40 when rain gages were used.

In the September 22, 2009 rainfall event, most the rain fell in the upstream portion of the watershed (see Figures 3-5 and 3-6). When modeled exclusively with rain gages, there is a 70% difference in modeled versus observed volume of streamflow and the time difference in peaks is 5 hours at the downstream gage (C). When modeled with radar

rainfall, the hydrographs for this event show an 8% difference volume at the most downstream point (C). The time differences in the peak streamflow at station C was 5.1 hours when modeled with rain gages and 3.9 hours when modeled with radar. The peak streamflow at this station is overestimated when modeled with Radar and underestimated with rain gage input. This over and underestimation of peak streamflow is also the result at stations A and B. At station B, there was a 3% volume difference when modeled with radar and 78% when modeled with rain gages. When modeled with radar, the shape of the hydrograph shows a double peak which is not present in the observed flow. The modeled streamflow for this station differs from the observed at this station by 3% for radar input and by 78% for rain gage data. However, the rain gage model peaked closer to the observed peak with a time difference of 1.8 hours whereas the radar model peaked with a 3.3 hour time difference. For the most upstream gage (Station A), the model differed from the observed volume of streamflow by 43% for radar and 119% for rain gage. The time of peak was well matched by the model using radar data input with a time difference of 0.1 hours. The time difference between the observed and rain gage modeled streamflow was 8.92 hours. The runoff coefficient varied from 0.17 to 0.13 when radar and rain gage data were used to model the rainfall-runoff processes.

The August 16, 2010 rainfall event was modeled exclusively with rain gage data (See Figure 3-7). The volume differences at station C showed a 13.5% difference in volume and a 1 hour time difference in peak streamflow. At the mid-point in the watershed (B), there was a 14.6 volume difference and 1.8 hour time difference in peak. At the downstream most point, the volume difference was 13.5% and the time difference in

streamflow peak was 1.7 hours. Using the rain gage data to model the rainfall-runoff

processes resulted in a calculated runoff coefficient of 0.28.

3.5. Discussion

A comparison of rain gage and radar rainfall for the July 7th event shows that the model

ultimately performs better at the downstream most station (C), which is where most of

the rainfall fell. It should be noted that the radar rainfall data estimated nearly twice the

amount of rainfall than the rain gage data. Furthermore the streamflow rate at the mid

and upper watershed was comparatively smaller than downstream with 0.4 and 0.9 cms

versus 28 cms. It is often very difficult to model streamflows in the lower range (Gan, et

al., 1997), especially where infiltration is a big factor, so this result follows previous

findings.

For the September 22 event, most of the rain fell in the upstream portion of the watershed

(Station A). However a comparison of the radar and rain gage rain total maps (See

Figures 3-5 and 3-6), shows that the greatest rain depths recorded by the NEXRAD radar

were in locations not monitored by rain gages. The model results best matched the

magnitude and timing of the peak best at the upstream station, A, when radar data were

used. However the volume differences were the greatest for the upstream section for

both radar and rain gage data. This illustrates the radar data's dependence on rain gage

data quality, as the radar data were calibrated to the rain gage observed data.

A comparison of the model results between the two events tested show that the model performs best in terms of timing and the shape of the hydrograph at the station nearest to the mass of the rainfall. Furthermore, the model more closely matches the observed peak, timing of peak, and volume of streamflow in events with greater rainfall depths. For instance, at the upstream portion of the watershed (station A), the model performed better for the September 22nd, event where the rainfall primarily fell upstream of station A, than for the July 7th event where the rainfall primarily fell near station C. Furthermore, the total depth of rainfall that fell in this portion of the watershed was greater. This follows other studies (Gan et al., 1997), which have determined that distributed hydrologic models simulate large rainfall events better than rainfall events with small depths.

Overall, the radar input improved the estimations of streamflow in terms of volume for both of the events. However it should be noted that the quality of radar data is intrinsically tied to the quality of the rain gage network because the radar is corrected based on the rain gage measurements. The upstream portion of Cypress Creek, monitored by Station C only has two rain gages. Thus this portion of the watershed has a rain gage density of 182 km^2 per rain gage. Rainfall is both over and underestimated in this data poor section. At station B, the rainfall is collectively monitored by 7 rain gages, with a density of 79 km^2 per rain gage. The downstream station, C, has a density of 67 km^2 per rain gage. Given the improved estimations of total streamflow volume at the down stream station in comparison to the upstream station, it can be concluded that the downstream portion of the watershed benefits from the higher density of rain gages.

When the calculated runoff coefficients were compared between the rain gage and radar modeled rainfall events, it can be seen that for both events, the runoff coefficients roughly agreed. Although there was a wide range in calculated values between the events, the coefficients fell within the published values for flat sandy soils for suburban residential and pasture areas (Haan et al., 1994). By comparing the magnitude of the estimated runoff coefficients, it can be observed that the rainfall runoff processes in the downstream section, which has the highest runoff coefficient, dominate the watershed response. Because of the sandy soils of the watershed, the residential and urban development, in the downstream portion of the watershed, dramatically increases the runoff in comparison to the undeveloped, upstream portion of the watershed.

3.6. Conclusion

A fully distributed model was created for the Cypress Creek Watershed in order to simulate rainfall runoff processes using both radar rainfall and rain gage data. The hydrographs of the modeled streamflow resulting from using the model with both radar and rain gages were then compared in terms of volume streamflow, peak time, and magnitude of peak. It was found that the model performed better with radar rainfall than with rain gages in terms of streamflow volume and peak flow. Between the compared events, the volume differences between modeled and observed streamflow varied by an order of magnitude between radar and rain gage datasets. The July 7th, 2009 event varied between 4.9% and 44% whereas the September 22nd, 2009 varied between 8% and 71% for radar and rain gage modeled storms, respectively. The modeled streamflow best

matched the observed streamflow for the station where the greatest depths of rainfall fell. However, the portions of the watershed with a low density of rain gages, to which the radar could be calibrated, performed poorly in comparison to portions of the watershed with a greater density of rain gages. In other words, the quality of radar rainfall data was dependent on the rain gage network in that the density of the rain gage network directly impacts the quality of the radar rainfall data. While radar represents a significant improvement in hydrologic analysis, it must be accompanied with a robust rain gage network to ensure quality of the data, especially for smaller storm events.

The use of radar rainfall in distributed hydrologic improves modeling of rainfall runoff processes in comparison to exclusive use of rain gages, especially in poorly gauged watersheds, such as Cypress Creek where the rain gage density ranged from 67 to 182 km^2 per rain gage. Advances in computing resources that allow for more efficient use of distributed models allow for greater use of radar rainfall in the modeling, management, and forecasting of water resources. Thus radar rainfall is a valuable data source for hydrologic study. The findings of this study to evaluate the use of radar rainfall to model small rainfall events will be used in future model development and analysis to include water quality.

Chapter 4 : Modeling of Pollutant Washoff and Transport Using Fully Distributed Hydrologic Modeling

Aarin Teague, Jason Christian, and Philip Bedient
Submitted July 2011 to Hydrological Processes

Abstract

Advances in hydrologic modeling have been shown to improve the accuracy of rainfall-runoff simulation and prediction. Building on the capabilities of distributed hydrologic modeling, a water quality model was developed to simulate buildup, washoff, and advective transport of a conservative pollutant. Classical washoff and transport relationships were utilized similarly to current lumped models. The spatially explicit model output provides greater spatial information on the dynamics of pollutant movement during storm events as well as a greater density of temporal information than current resource limited sampling data.

Coupled with the physically based *Vflo*^{TM} hydrologic model, the pollutant transport model was used to simulate the washoff and transport of total suspended solids for multiple storm events in Cypress Creek Watershed. The model was calibrated and applied to small storm events. Since small storms occur more frequently, accurately modeling small rainfall events, which have traditionally been difficult to model, is necessary for the investigation and design of watershed management practices. The output of the distributed buildup and washoff model was compared with storm water quality sampling in order to assess the performance of the model. For a majority of the storms modeled, the model performed with an acceptable degree of error using root mean

squared error and normalized mean squared error metrics. The model output was then analyzed to temporally and spatially characterize the storm events. This effort was the first step in developing a fully distributed water quality model. As such it provides the framework for the incorporation of more sophisticated pollutant dynamics and a spatially explicit evaluation of best management practices and land use change.

4. Introduction

Stormwater quality in streams continues to present challenges for protecting water resources. Pollutant transport models are valuable tools for the investigation and management of watersheds (Mannina and Viviani, 2010) as well as design and evaluation of measures to protect the sources of water supplies. Physically based water quality models combine mathematical models of buildup and washoff (Avellaneda et al., 2009). These complex processes, which vary temporally and spatially, create technical challenges for the development of stormwater quality models (Dotto et al., 2010).

Current water quality models such as Hydrologic Simulation Programmed in Fortran (HSPF) and Storm Water Management Model (SWMM), both lumped models, are capable of simulating single rainfall events. Lumped models parameterize the watershed by aggregating similar spatial areas and applying values to the assumed homogeneous area (Bicknell et al., 2001). Although HSPF provides well matched simulation of high flows, intermediate storm flows are generally underestimated and low flows are generally overestimated (Chen et al., 1995; Singh et al., 2005). Physically-based, distributed models have been introduced as a way to improve predictive capability by including spatially variable physical parameters (Min and Wise, 2010). Advances in GIS provide

physical data that make it possible to model hydrologic and water quality processes with greater accuracy, detail, and utility. While lumped models lack the spatial refinement to simulate the effects of specific best management features or land use changes (Nikolaidis et al., 1998), distributed models provide the necessary detail to simulate important watershed features that effect water quality.

Previously developed water quality models such as SWMM and HSPF have well established methods which use a lumped approach for simulating the washoff and transport processes. SWMM was designed to model urban watersheds (Huber et al., 1975; Rossman, 2004), and thus is best used for homogeneous urban areas. In contrast, HSPF is capable of modeling watersheds with mixed developed and undeveloped land uses. While HSPF can simulate single events with a user-defined timestep, it was designed for continuous simulation (Borah et al., 2006). Both HSPF and SWMM lack the spatial refinement to model the effect of specific best management practices (BMPs) such as street sweeping, vegetative filter strips, or detention basins (Shoemaker, et al., 2005) in a spatially explicit manner. To provide more comprehensive modeling, the Better Assessment Science Integrating point and Nonpoint Sources (BASINS) tool incorporates HSPF with other hydrologic and water quality models. By coupling other models with HSPF, such as Kineros, which simulates sediment transport (Guber et al., 2009), some BMPs can be modeled.

Recent advances in hydrologic modeling provide the basis for development of models which incorporate the necessary spatial refinement to accomplish detailed pollutant

transport simulation. However, few water quality models take advantage of fully distributed hydrologic models or the use of radar rainfall. The improvements in simulation and prediction of streamflow, (Vieux et al., 2004) as well as the utility provided by spatially explicit pollutant data, warrant further effort to refine water quality modeling. By applying well established modeling approaches for washoff and transport to advanced hydrologic simulation, a foundation for a fully distributed water quality model has been developed in order to take advantage of the improved accuracy of hydrologic modeling. This is the first step towards building a water quality model that can be used to model BMPs, land use changes, and pollutant loading in a spatially explicit manner.

The objective of this study was to develop a pollutant washoff and transport model that is coupled with a physically based, fully distributed hydrologic model. This was accomplished through the application of a hydrologic model, $Vflo^{TM}$ for the study watershed, Cypress Creek Watershed. The output of the hydrologic model was then used as input for an independent washoff and transport model. The model was used to simulate total suspended solids (TSS) and was compared to storm water sampling data, in order to assess model performance for small rainfall events. The model output was then used to further analyze the dynamics of the storm events by evaluating the existence of a first flush behavior and the spatial distribution of pollutant washoff loading.

4.1. Background

4.1.1. Fully Distributed Hydrologic Modeling

Rainfall-runoff processes were simulated using the physically based hydrologic model, $Vflo^{TM}$. While there are a variety of models that provide distributed simulation and prediction of watershed hydrology, such as MIKE SHE (Vasquez, 2002), TUFLOW (Syme, 2001), TOPNET (Bandaragoda, 2004), and Gridded Surface Subsurface Hydrologic Analysis (GSSHA) (Byrd, 2005), $Vflo^{TM}$ was selected due to extensive application within the Texas Gulf Coast region (Fang et al., 2010; Fang, et al., 2008; Safiolea, 2006, Duncan, 2011; Vieux and Bedient, 2004) and proven performance in low-slope urbanized watersheds (Vieux and Vieux, 2006). Vieux has published detailed documentation of the model and its applications (2004). The kinematic wave analogy is used to route runoff from overland flow to channel flow, according to a flow direction grid derived from a digital elevation model (DEM). Infiltration is calculated using the Green & Ampt Equation (Kim et al., 2008) and overland flow is modeled using Manning's equation (Vieux et al., 2009). An advantage of using this model is that geospatial data representing elevation, soils, and land use are incorporated as parameters to solve these relationships. It uses the Galerkin's formulation of finite elements for the solution of the kinematic wave analogy (Vieux, 2004) with a finite difference solution to time discretization subject to the Courant condition (Vieux et al, 2009).

4.1.2.Pollutant Buildup

The type and rate of pollutant buildup is dependent on land use, human activities, and season (Overton and Meadows, 1976). The accumulation of a pollutant on a surface can be calculated by different relationships such as linear, power, exponential, and Michaelis-Menton function (Barbe et al, 1996). Among the different modeling options used by models, such as SWMM and HSPF, is the linear buildup function. The rate of accumulation of a pollutant can be modeled using the linear function as

$$\frac{dP}{dt} = C \qquad (4\text{-}1)$$

where P is the pollutant mass, t is time, and C is the constant rate of accumulation. Observed pollutant loading shows that there is an upper limit on the amount of buildup, often influenced by degradation, wind, or human activity (Alley and Smith, 1981). In addition, runoff does not completely remove pollutant from the land surface, leaving residual mass. While the linear function may not always be adequate to describe buildup (Chen and Adams, 2007), it is considered to be an appropriate model in the absence of extensive water quality sampling data.

4.1.3.Pollutant Washoff

Washoff is the process of removal of soluble and particulate pollutants by rainfall and runoff (Vaze and Chiew, 2003). Falling raindrops create turbulence and overland flow loosens particles from the surface, transporting the particles through the watershed with the water flow. Storm water quality models traditionally conceptualized the washoff

process as driven by the energy of raindrop impact or the shear stress of the runoff (Brodie and Rosswall, 2007).

The washoff rate is the first order differential equation (Soonthornnonda et al., 2008, Butcher, 2001, Stieber et al., 1999)

$$\frac{dP}{dt} = -kP \tag{4-2}$$

which describes the rate at which pollutant mass, P, is removed from the land surface relative to the coefficient, k. Both SWMM and HSPF assume that this coefficient would vary in direct proportion to the rate of runoff (Butcher, 2003; Barbe et al., 1996) over the subcatchment, r (depth/time), as

$$k = \left(\frac{2.30}{WSQOP} \right) r \tag{4-3}$$

where $WSQOP$ is the runoff depth that results in washoff of 90% of solids from the land surface (Deliman et al., 1999). Typical values for k range from 10.16 mm (0.4 in.) to 17.78 mm (0.7in) for impervious surfaces (Butcher, 2003; Sartor and Boyd, 1972), and are generally greater than 25.4 mm (1in) for rougher pervious surfaces (Yagow et al., 2001).

4.1.4. Pollutant Transport

Pollutants can be transported through advection, dispersion, or diffusion. However, due to the time scale of a single storm, fully entrained particles and solutes are transported overwhelmingly by shallow overland flow (Singh, 2002a). Therefore, transport can be modeled based exclusively on advection (Bicknell, 2001). As a result, pollutant transport

by overland flow can be modeled using the dynamic equations of free-surface flow, also known as Saint Venant equations (Akan, 1987). This may not accurately represent the natural environment, where pollutants are subject to other transport, biochemical, and biological processes. Despite these limitations, it is assumed that pollutant transport by diffusion and dispersion as well as biochemical reactions are negligible (Singh, 2002a; Singh, 2002b), and transport is sufficiently modeled by advection.

Advective transport in shallow overland flow can be adequately approximated by the kinematic wave analogy (Singh, 2002a; Singh, 2002b, Akan, 1987), similar to the approach for modeling hydrology. Kinematic wave is a simplification of the Saint Venant equations and is basically a mass balance of the pollutant movement in runoff, run-on, rainfall deposition, pollutant flux from the land surface, and accumulation of pollutant in the overland flow (Akan et al., 2000). Mathematically this takes the one dimensional form

$$\frac{\partial(Ch)}{\partial t} + \frac{\partial(CQ)}{\partial x} + \frac{\partial P}{\partial t} = C_R I \tag{4-4}$$

where C is the concentration of the pollutant in runoff (mass/volume), C_R is the concentration in rainfall (mass/volume), Q is the overland flow rate (volume/time), h is the depth of runoff (length), P is the mass of pollutant on the surface of the land (mass/area), I is the intensity rainfall (depth/time), t is time, and x is the length in the x flow direction (Akan, 1987). The first term is the change in mass flux of the pollutant in the runoff over time. The second term is the net flux of pollutant in the runon and runoff, over the distance dx. The third term is the change in mass of pollutant per area of land

surface over time. The term on the right hand side is the mass of pollutant falling on the surface during rainfall.

4.2.   Methodology

The development of the washoff and transport model is dependent on the fully distributed hydrologic model, $Vflo^{TM}$ (Vieux Inc., 2011). This rainfall-runoff model output, in the form of spatially explicit discharge data for each time step of the simulation, served as the input of the independent washoff and transport model (See Figure 4-1) and was then analyzed using ArcMap 9.3 (ESRI, 2008). Both the hydrologic and washoff and transport model were developed for the study watershed, Cypress Creek, and used to simulate the rainfall-runoff processes coupled with washoff and convective transport of TSS for lower flow rainfall events. The simulated TSS concentration was then compared to observations of TSS collected during the modeled storm events to assess the performance of the washoff and transport model.
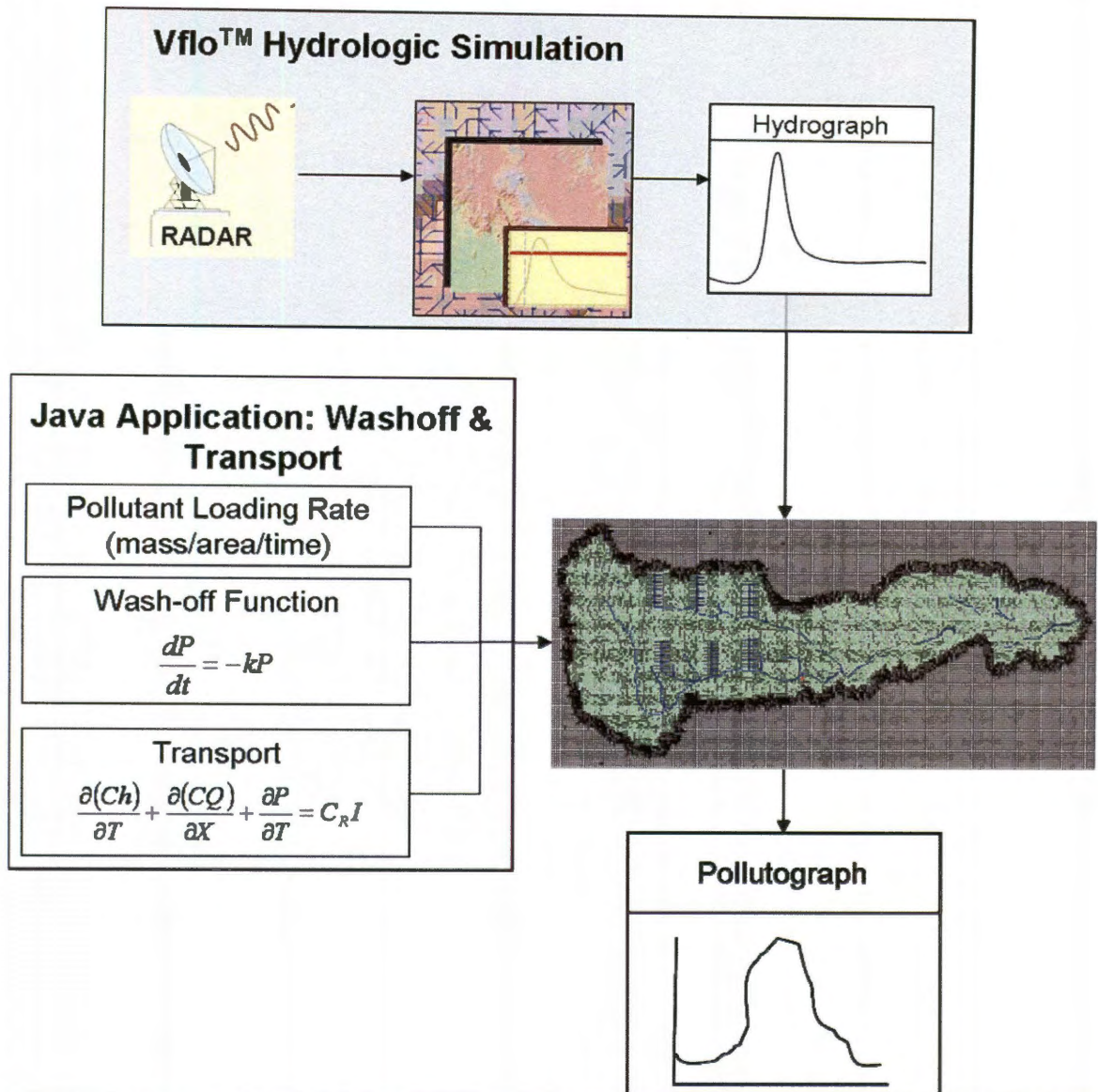
Figure 4-1. Development of water quality model coupled with distributed hydrologic modeling

4.2.1. Hydrologic Model Development

A $Vflo^{TM}$ model of Cypress Creek watershed (Figure 4-2) was previously developed using Lidar elevation, NRCS soils data, and Tropical Storm Allison Recovery Project (TSARP) land use and cross section datasets (Teague et al., 2011). Four storm events were simulated using the $Vflo^{TM}$ model. The characteristics of the events are described in

Table 4-1, and the spatial rainfall distributions are shown in Figure 4-3. The storms

varied between 7.64 mm and 39.4 mm of average rainfall, and are thus considered small

or minor storm events in the Texas Gulf Coast region. Two of the events, the July 7,

2009 and September 22, 2009 storms were modeled using radar rainfall data that was

calibrated to the rain gauge network. The other two events were modeled using an

exponential interpolation of data from the rain gauge network. The resultant discharge

and depth of flow data from the simulation of the storm events served as the hydrologic

basis of the pollutant washoff and transport model described below.

Table 4-1. Storm event characteristics

Storm Event	Duration (hr)	Average Rainfall Depth (mm)	Maximum Intensity (mm/hr)	Average Intensity (mm/hr)	Observed Streamflow Peak (CMS)
7-Jul-09	7.00	12.70	79.25	1.81	28.60
22-Jul-09	3.25	12.20	35.56	3.75	6.77
22-Sep-09	9.75	39.37	40.64	4.04	34.83
7-Jul-10	10.00	7.62	40.64	0.76	31.15
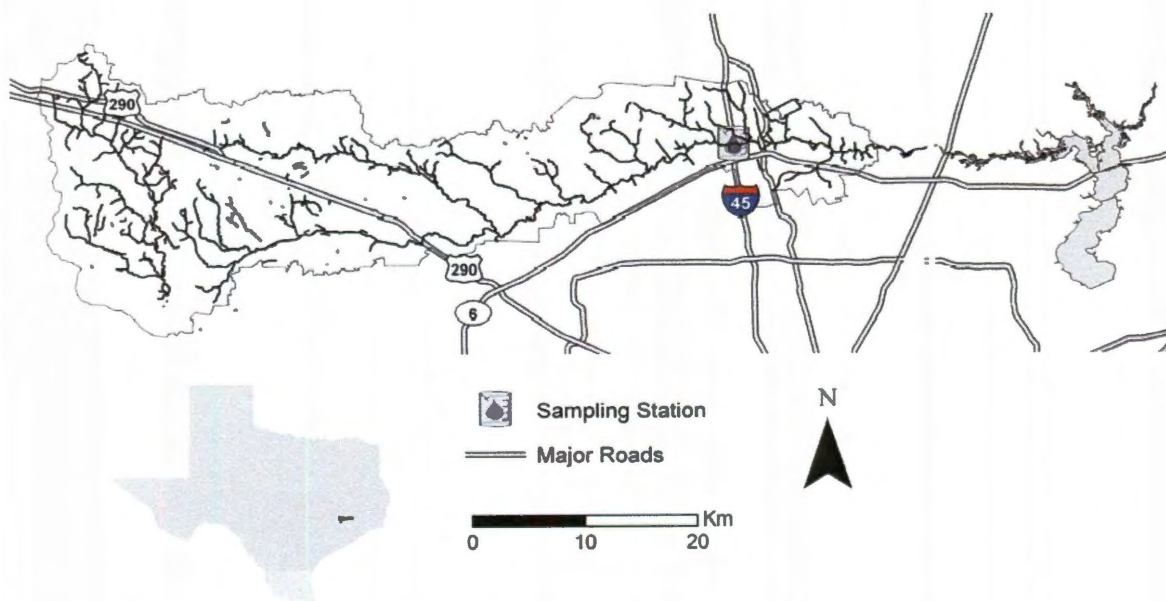
Figure 4-2. Cypress Creek Watershed on the Texas Gulf Coast
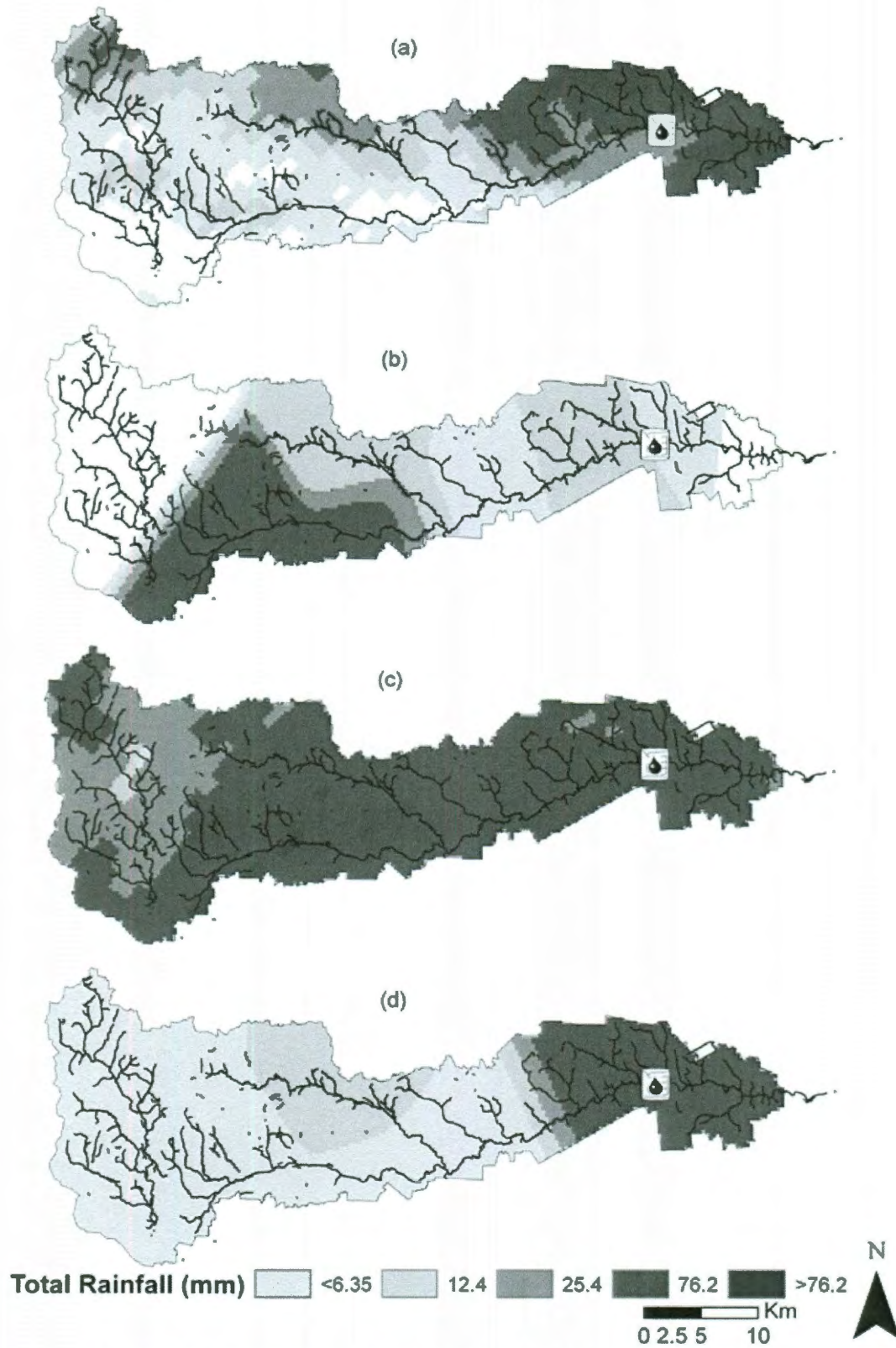
Figure 4-3 . Spatial distribution of total rainfall for (a) July 7, 2009, (b) July 22, 2009, (c) September 22, 2009, and (d) July 7, 2010

4.2.2. Washoff - Transport Model Formulation

The washoff and transport model was developed in the Eclipse development environment

using Java (Eclipse, 2001). The application uses the kinematic wave analogy to

calculate the mass of TSS in the runoff from each cell in a grid representing the

watershed. This mass balance approach accounts for the washoff, deposition, and the

pollutant runon from other cells, so that the pollutant discharge from each cell can be

determined (See Figure4- 4).
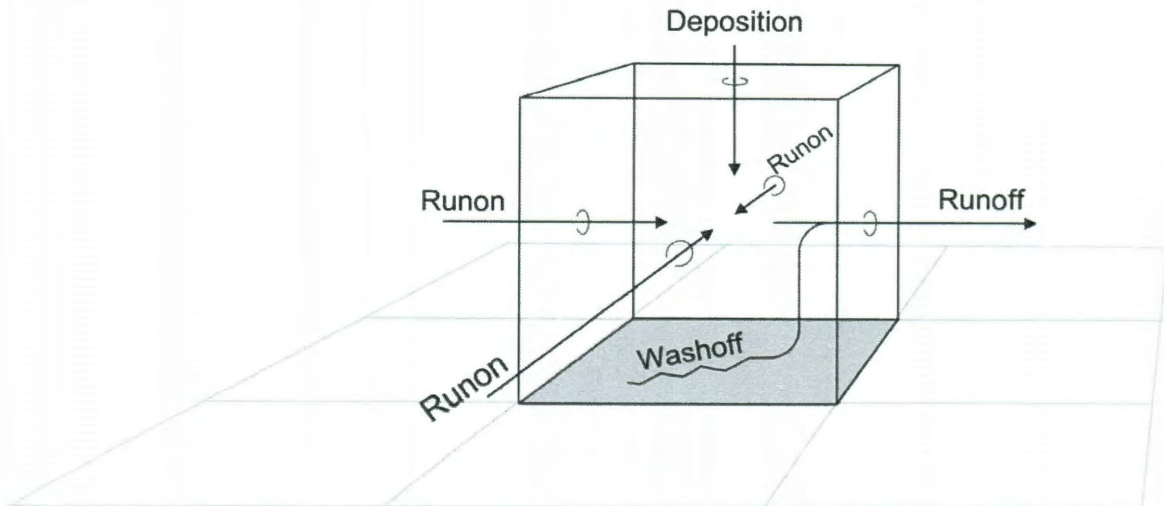
Figure 4-4 . Mass balance of pollutant solved over a grid

Pollutant washoff and transport were modeled based upon the previously discussed

conceptual framework using the same grid as the $Vflo^{TM}$ simulation. The algorithm to

accomplish this is conceptualized within the following steps: (1) Estimate the pollutant

loading to the land surface; (2) Determine the mass of pollutant entering the runoff in

terms of pollutant washoff, point source loading, and pollutant runon; (3) Use the

continuity equation to calculate the concentration of pollutant in the flow leaving each

cell. (4) Repeat for the next time step. The result of the algorithm is a distributed simulation of the mass of pollutant in runoff throughout the watershed over time.

4.2.2.1. Buildup

The first step of the algorithm, to calculate the pollutant loading to the surface, uses a land use dataset to determine the loading rate. When possible, locally available data regarding the population and distribution of pollutant sources should be used to create spatially explicit estimates of the potential loading rate (Teague et al., 2009). The HSPF default settings were used to assign loading rates based on land use according to Table 4-2 (Bicknell et al., 2001). The time since the last storm with a depth greater than 6.35 mm (0.25 in), the depth assumed to result in runoff (Butcher, 2003), was uniformly applied to the loading rates throughout the watershed to estimated the pollutant buildup on the land surface.

Table 4-2. Buildup rates for total suspended solids based on land use

Land Use	TSS (kg/Ha/yr)
Residential	560
Industrial	560
Cropland and Pasture	2242
Herbaceous Rangeland	56
Mixed Rangeland	56
Forest Land	56
Water	28
Forested Weltand	28
Non-Forested Wetland	28
Bare	22

*(Bicknell et al., 2001)

4.2.2.2.   *Washoff and Transport*

The mass of pollutant washed off the land surface can be calculated for each time step using

$$P_t = P_0 \left(1 - e^{\left(-\frac{2.30}{WSQOP} r \right)} \right)$$ (4-5)

where P_0 is the initial pollutant mass per unit area on the land surface, and P_t is the mass per unit area of the pollutant washed off the land surface. Runon, or the mass entering the cell through the runoff from other cells, is calculated using,

$$Runon(i, j, t) = \sum_{i,j} QC\Delta t .$$ (4-6)

The required initial condition was assumed to be 20 mg/L for cells representing a channel and the near riparian area based on initial water quality sampling data. Overland cells were given an initial concentration of 1 mg/L (Deng et al. 2005).

Point source flows include the known effluent from permitted discharges (H-GAC, 2009), that were primarily municipal utility district (MUD) waste water treatment plant effluent discharges (H-GAC, 2009). For the purposes of this study, it was assumed that the treatment plants would discharge at the reported monthly average flow rate and at a concentration of 13 mg/L (TCEQ, 2010). It was known that the point source behavior was highly variable and was thought to contribute to the pollutant loading to the stream during storm events, however there was no available data to estimate the influence of the timing and magnitude of such events. Therefore, the point sources were represented by the average monthly discharge, $Q(i,j,t)$ and the assigned concentration of the effluent, $C(i,j,t)$ by the relationship

$$Mass(i,j,t) = \sum Q(i,j,t)C(i,j,t)\Delta t \qquad (4\text{-}7)$$

for each grid cell in the model, were *Mass(i,j,t)* is the mass loading within each timestep

for each grid cell.

Advection of a fully entrained constituent was simulated according to the method used in

the ADVECT subroutine of HSPF using the continuity equation (Bicknell et al., 2001).

The model solves for the concentration, *C*, of the pollutant at each time step through

$$C(i,j,t) = \frac{\left(\sum mass_{in} + \left(C(i,j,t-1)*\left(h(i,j,t)(\Delta x)^2 - Svol(t-1)\right)\right)\right)}{\left(h(i.j.t)(\Delta x)^2\right)+ Evol(t))} \qquad (4\text{-}8)$$

where *h* is the depth of runoff, *Svol* is the outflow component of the discharge based on

the start of the timestep, and *Evol* is the outflow component of the discharge based on the

end of the timestep. The $\Sigma mass_{in}$ is the sum of the mass flowing into the grid cell,

including washoff from the land surface, runon, point source flows, and pollutant in the

precipitation. This study assumed that there was no TSS in the rainfall.

The application described was applied for the study watershed, Cypress Creek, described

later in the methods section, using distributed geospatial datasets including flow

direction, channels, land use, and point source effluent discharge rates. The flow

direction data were derived from Lidar elevation dataset (TSARP, 2006). Land use data

(HGAC, 2008) were used to parameterize the washoff coefficient and total suspended

solids loading rate, as shown in Table 4-2, using the default values suggested by the

USEPA's HSPF model. Depth and flow rate of the water, the output of the hydrologic

model, were used to simulate the movement of TSS by calculating the mass balance of TSS in each grid cell.

4.2.3. Calibration

The washoff and transport model was calibrated by adjusting the loading rate of total suspended solids onto the land surface. The initial pollutant mass in or near riparian grid cells, or the cells that correspond to the channels, was multiplied by a loading factor. The loading factor was adjusted until the pollutographs matched the sampling data for the two calibration storms. In other words, the initial mass of pollutant in areas with higher rates of runoff discharge was multiplied by this loading factor. The final calibrated loading factor was spatially varied, depending on the average velocity of the runoff in the storm, with a loading factor of 25 applied to areas with velocity greater than 0.1 m/s. This was considered an acceptable approach because of the higher potential for pollutant contribution from riparian and near riparian areas, which capture loading from previous runoff events (Newham et al., 2005). The calibration storms were the July 7, 2009 and September 22, 2009 rainfall events. The calibrated model was then used to simulate TSS washoff and transport for two additional events, July 22, 2009 and July 7, 2010.

4.2.4. Analysis of Results

The outputs from the java application were matrices of estimated pollutant concentration for the watershed grid at each time-step of the simulation. The results were post-processed in ArcMap for visualization and analysis purposes. The estimated concentration at the down-stream water quality sampling station (Figure 4-2) was

selected and plotted as a time series, in order to compare the results to the sampled concentration.

The performance of the model was assessed using the root mean squared error (RMSE) and normalized mean square error (NMSE) in order to quantify the difference between the modeled and the sampled observations. The RMSE was calculated by

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(y_{p_i} - y_i\right)^2}{n}} \qquad (4\text{-}9)$$

where n was the number of samples, y_i was the actual observed value, and y_{pi} was the modeled value (Stow et al., 2003). The RMSE is the average model error. The normalized mean square error is in turn calculated as

$$NMSE = \frac{\sum_{i=1}^{n}\left(y_i - y_{p_i}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \qquad (4\text{-}10)$$

where $\bar{y}$ is the mean of the observed data (Castelli et al., 2003). This performance indicator represents the sum of squared errors normalized by the estimated variance of the data (Chau, 2003). Values of less than or equal to 1 for the NMSE metric are generally considered to indicate a good model and the goal is for the RMSE to be as low as possible (Poli and Cirillo, 1993).

The flow averaged concentration of TSS for each event, or event mean concentration, *EMC*, (Sansalone, 1996), was calculated using both the observation and modeled data

$$EMC = \frac{M}{V} = \frac{\int\limits_0^n C(t)Q(t)dt}{\int\limits_0^n Q(t)dt}$$
(4-11)

where M is the total mass of constituent over the entire event duration , V is the total

volume of flow over the entire event duration, $Q(t)$ is the streamflow, $C(t)$ is the

constituent concentration, and n is total number of time steps in the event duration

(Huber, 1993). The EMC calculated from the observed and modeled data were compared

to assess agreement between modeled and observed data. This is especially critical in

resource limited studies, where only a limited number of water quality samples are

available.

Stormwater monitoring data is often reported as EMCs (Charbeneau and Barret, 1998),

however it doesn't provide an indication of the temporal aspect of the pollutant loading,

such as the first flush phenomenon (Sansalone, 1996). The first flush concept describes

the disproportionate mass of pollutant that is loaded in the stream during the initial stages

of a storm event (Hathaway and Hunt, 2011). The observation of first flush is

inconsistent in urban watersheds, anddependent on storm size, rainfall intensity,

watershed characteristics, hydrologic conditions, and transport factors (Deletic, 1998).

There are a variety of definitions and methodologies to describe first flush, with mass

based approaches most commonly used for evaluation. The mass based first flush is

indicated by greater delivery of the constituent mass during the rising limb of the runoff

hydrograph (Sansalone and Cristina, 2004). The percent total modeled mass of

suspended sediment and volume of streamflow is calculated by

$$V(t) = \frac{\int_0^k Q(t)dt}{\int_o^n Q(t)dt} * 100 \tag{4-12}$$

$$M(t) = \frac{\int_1^k Q(t)C(t)dt}{\int_1^n Q(t)C(t)dt} * 100 \tag{4-13}$$

Where $V(t)$ is the percent of total volume runoff, $M(t)$ is the percent of total mass in the streamflow, $Q(t)$ is the streamflow, and $C(t)$ is the constituent concentration through the modeled station, k is the sample time, and n is the total time of event. Using the modeled TSS concentration and streamflow, the $M(t)$ and $V(t)$ were calculated and then plotted for each event, in order to examine the temporal behavior of the mass loading in relation to the volume of streamflow, focusing on the period prior to peak streamflow (Berretta and Sansalone, 2011).

In addition, the sum of mass washed off and the point source loading for each storm was determined by subtracting the mass remaining on the land surface at the end of the event from the estimated mass of TSS present on the surface at the start of rainfall and then adding the estimated total mass loading for the duration of the event from the point sources using the spatial analyst feature in ArcMap. This analysis was examined in order to compare the sources of pollutant loading for each of the modeled storms.

4.2.5. Water Quality Sampling

In order to validate and calibrate the water quality results of the washoff and transport model, storm water quality samples were collected at a downstream sampling point. Water samples were collected as grab samples for rainfall events in the 1.5 through 3 inch range at the most downstream gauge of Cypress Creek watershed before it flows into Lake Houston (See Figure 4-2). Samples were preserved on site and then analyzed by the City of Houston Water Quality Laboratory. Total suspended solids were measured using EPA method 160.2. Sample collection timing and interval was guided using radar rainfall and advance hydrologic modeling, in order to attempt capturing the data reflecting a range of samples that reflect the storm water pollutant washoff and transport throughout the watershed.

4.2.6. Study Area

The model was tested using the study watershed, Cypress Creek is a 797 km^2 (308 mi^2) watershed north of the city of Houston on the Texas Gulf Coast (Figure 4-2). It flows 80 river km (50 miles) to Lake Houston, which serves as the primary source of drinking water for the City of Houston (Chellam et al., 2008). This watershed has been the primary contributor of urban runoff and pollutant loading to Lake Houston (Sneck-Fahrer et al., 2005), creating challenges in treating the water for potable use. Although rapidly urbanizing, Cypress Creek is a complex watershed with varied land use. The western upstream sections of the watershed are primarily agricultural, forest, and pasture land use. In contrast, the eastern, downstream portion of the watershed is residential and urban

development. Cypress Creek watershed is relatively flat with sandy loam soils, which have a greater infiltration potential and less erosion potential

4.3.   Results

Radar rainfall and rain gauge data was used to model the rainfall-runoff processes in Cypress Creek for four storm events. A comparison of the modeled versus observed streamflow is shown in Figure 4-5. The hydrologic model was evaluated at multiple points within the watershed (Teague, 2011). Overall the model simulated the hydrology of the watershed at an acceptable level. The magnitudes of the streamflow peaks were well matched, but the model generally was late in timing. The difference between modeled and observed streamflow total volume ranged between 7% and 18% (Table 4-3). The July 7, 2009 storm, which exhibited the greatest difference in volume, was a higher intensity storm primarily center in the downstream portion of the watershed. The greatest percent difference in the peak streamflow was exhibited by the July 22, 2009 storm which was the shortest duration storm event that occurred primarily in the center of the watershed.
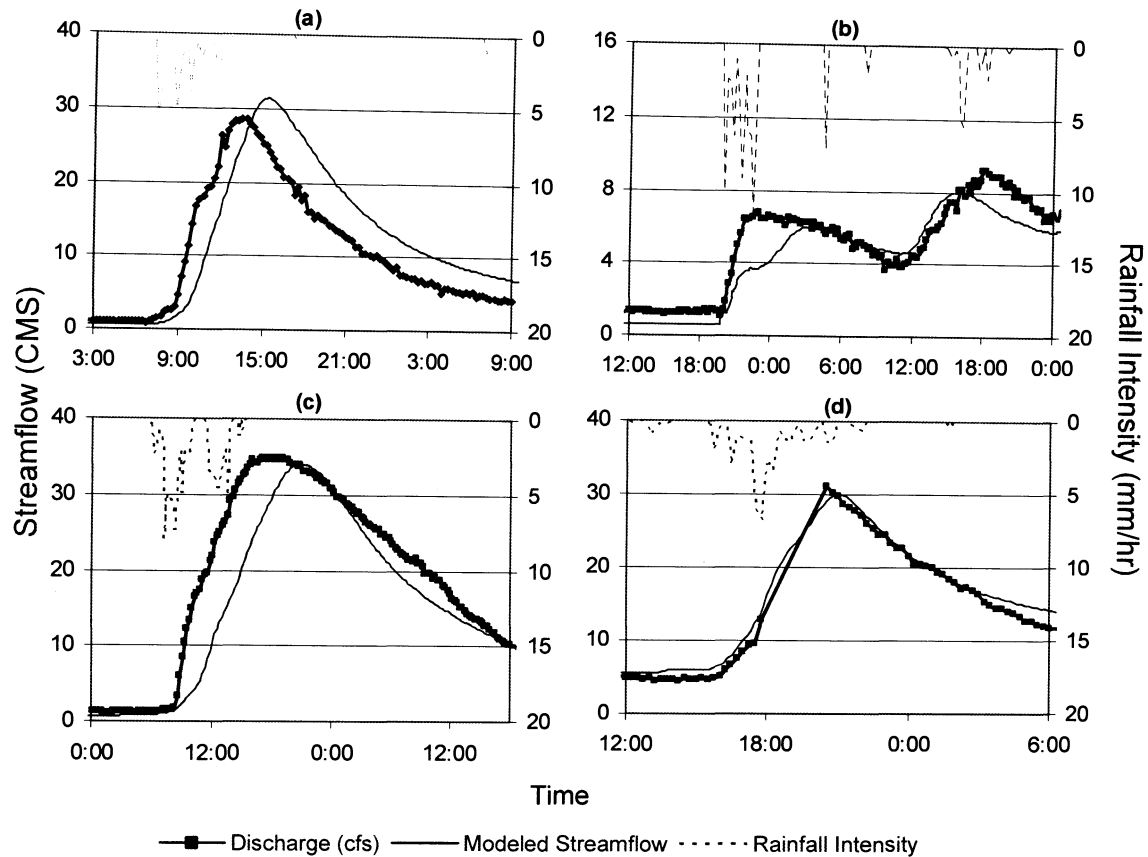
Figure 4-5. Modeled and observed streamflow for (a) July7, 2009, (b) July 22, 2009, (c) September 22, 2009, and (d) July 7, 2010

Table 4-3 . Results of hydrologic modeling

Storm Event	Observed Peak (CMS)	Modeled Peak (CMS)	Volume Difference (%)	Difference in Peak Streamflow (%)
7-Jul-09	28.60	31.40	18.34	9.8
22-Jul-09	6.77	5.40	-13.87	-20.2
22-Sep-09	34.83	34.07	-15.39	-2.2
7-Jul-10	31.15	29.97	7.64	-3.8

The simulated hydrology then served as the input to the washoff and transport model.

The simulated concentration of TSS is plotted along with the storm water quality

sampling data as pollutographs in Figure 4-6, in order to assess the model performance.
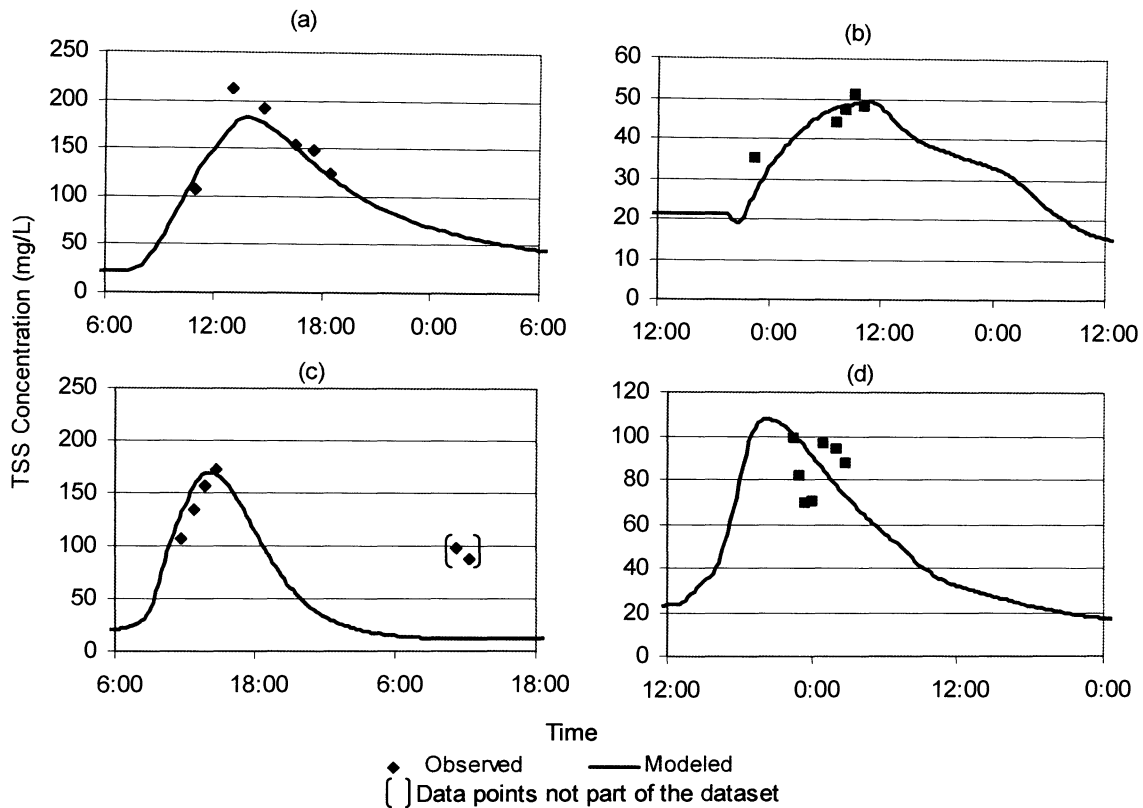
Figure 4-6. Modeled and observed TSS concentrations for (a) July 7, 2009, (b) July 22, 2009, (c) Septermber 22, 2009, and (d) July 7, 2010

The first event, July 7, 2009 (a), matched the approximate shape of the pollutographs from the observed data, but underestimated the maximum concentration 15%. The July 22, 2009 event (b) again matched the approximate shape of the observed pollutograph, and approximately matched the magnitude of the samples. The simulated TSS concentration of the September 22, 2009 event (c) matched the observed data on the rising limb of the hydrograph. However, samples taken late on the falling limb of the hydrograph were grossly underestimated. All of the TSS measurements for the fourth event (d), on July 7, 2010 were taken on the falling of the hydrograph. The modeled data

was roughly in range of the observed data, but did not capture the variability of the sample results.

The metrics of RMSE and NMSE were used to evaluate the agreement of the modeled and observed data and are reported in Table 4-4 along with the EMCs calculated from both the modeled and observed data. The best match of the modeled rainfall event was for the July 22, 2009 storm. The greatest amount of error was exhibited by the September 22, 2009 event, due primarily to the two late sampling points on the falling limb of the hydrograph. Due to the time lag between the peak of the streamflow and these sampling points, it is suspected that these sample points are not valid for inclusion in the dataset. When the values are excluded from the dataset, the model shows good agreement, according to the NMSE and a 7.6% difference in the EMC. Overall, multiple storms have NMSE of less than one, making the washoff and transport model an acceptable simulation of TSS for these character of events. One event, July 7, 2010, had a NMSE of greater than one, indicating poor model performance within the small time scale of sample collection.

Table 4-4. Assessment of the washoff and transport model for the modeled rainfall events

Storm	RMSE (mg/L)	NMSE(mg/L)	Modeled EMC (mg/L)	Observed EMC (mg/L)
7-Jul-09	19.34	0.61	146.49	158.59
22-Jul-09	4.78	0.26	41.74	39.18
22-Sep-09	18.27	0.30	136.11 (54.03)*	127.11, (140.24)*
7-Jul-10	16.98	2.56	88.87	83.91

* With non-applicable data points removed, ( ) calculated with non-applicable data points

The entire dataset of modeled versus observed TSS concentrations were plotted in Figure 4-7. Overall the modeled data matches fairly well, as evidence by a linear regression with a slope of 0.98 and r^2 of 0.87. From this, it can be concluded that the TSS concentration is slightly underestimated by the distributed washoff and transport model.
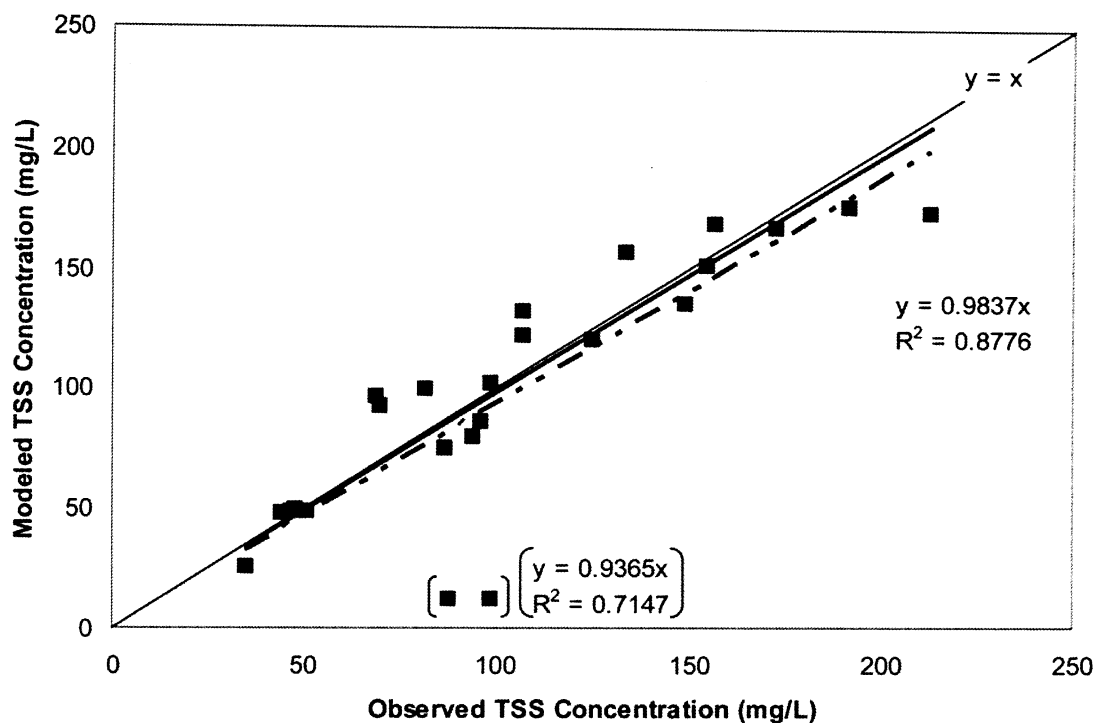
( ) With non-applicable points included
Figure 4-7. Overall modeled versus Observed TSS

The difference in modeled EMC versus observed EMC ranged from 3.6% to 61.4%. The best match was for the July 22, 2009 storm, and the largest difference was with the September 22, 2009 storm. This latter discrepancy is mainly due to the underestimation on the falling limb of the hydrograph. It is important to note that due to the limited number of samples, the observational data is an incomplete picture of the TSS loading

dynamics. Therefore large differences in the EMC reflect uncertainty in both the model as well as the limited data.

The temporal dynamics of the simulated mass loading of TSS was analyzed by the calculation of the percent mass loading and the percent volume streamflow and then plotted versus the normalized time of the storm shown in Figure 4-8. The model of the July 7, 2009 event (a) shows that prior to the peak in streamflow, the percentage of mass loading exceeded that of the percent of volume streamflow. However the differences were not substantial, and thus the first flush was only weakly exhibited in the washoff model. The July 22, 2009 (b) model does not exhibit first flush phenomenon at all, as evidenced by the cumulative percent volume exceeding the cumulative percent mass loading throughout the duration of the storm. This event did not have a lengthy prior buildup period and occurred soon after a previous event, which would have washed off much of the TSS built up on the surface. Furthermore, this was a low intensity event with a uniform rainfall distribution over a longer duration than the other three modeled events. In fact, the mass loading and volume are roughly proportional, so the event can be characterized as "flow- limited" where the critical factors limiting the washoff process is the flow rate and volume of water (Ma et al., 2010). The event on July 7, 2010 (d) also did not exhibit high mass loading during the rising limb of the hydrograph. In contrast, third event on September 22, 2009 (c), showed distinctive first flush phenomenon, with the percent mass loading disproportionate to the volume streamflow, especially prior to peak streamflow. In this case, the critical limiting factor in the washoff process was the mass of pollutant on the surface available for transport, and the event is characterized as "mass limited" (Sheng et al, 2008).
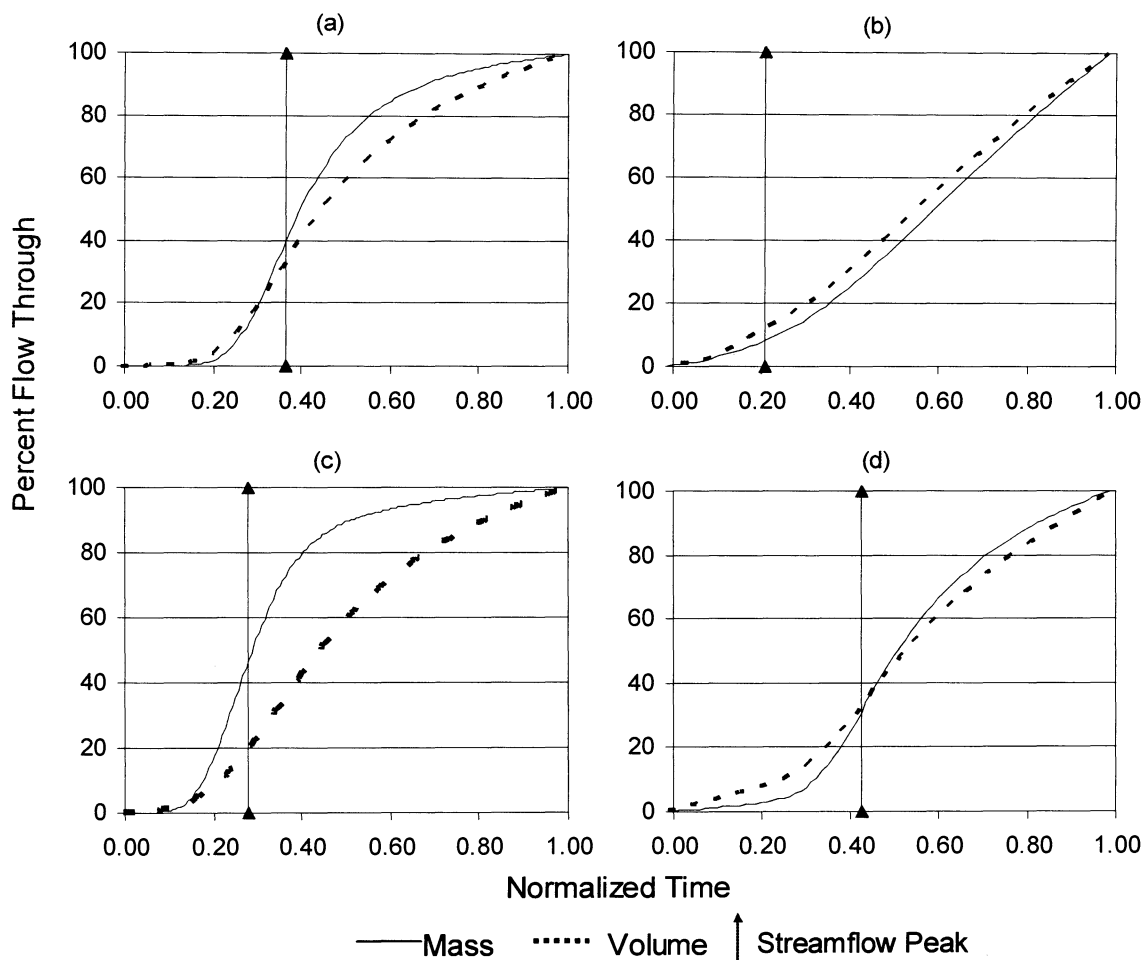
Figure 4-8. Percent mass and volume flow through curves (a) July 7, 2009, (b) July 22, 2009, (c) Septermber 22, 2009, and (d) July 7, 2010

The fully distributed data output of the washoff transport model was used to visualize the distributed loading of TSS resulting from both washoff and point sources for the duration of the storm event, seen in Figure 4-9. For the July 7, 2009 storm (a), most of the loading occurred in the urban, downstream portion of the watershed. A similar spatial loading pattern was exhibited for the July 7, 2010 storm (d), although the there was a more extensive area of washoff in the upstream, agricultural portion of the watershed. The July

22, 2009 (b) event had a more even distribution of loading in the downstream and middle of the watersheds, which are primarily urban and suburban developed areas. In contrast, the September 22, 2009 event (c), which exhibited strong first flush characteristics, had most of the loading in middle to upper-middle portions of the watershed. At the time of the event, there was extensive urban development and residential construction occurring in this area of the watershed, potentially contributing to the TSS loading during the rainfall event. By identifying areas that contribute greater pollutant mass to the stream during storm events, the source of pollutants can be spatially estimated. This would then provide the basis for spatially located BMPs to address the pollutant loading and transport.
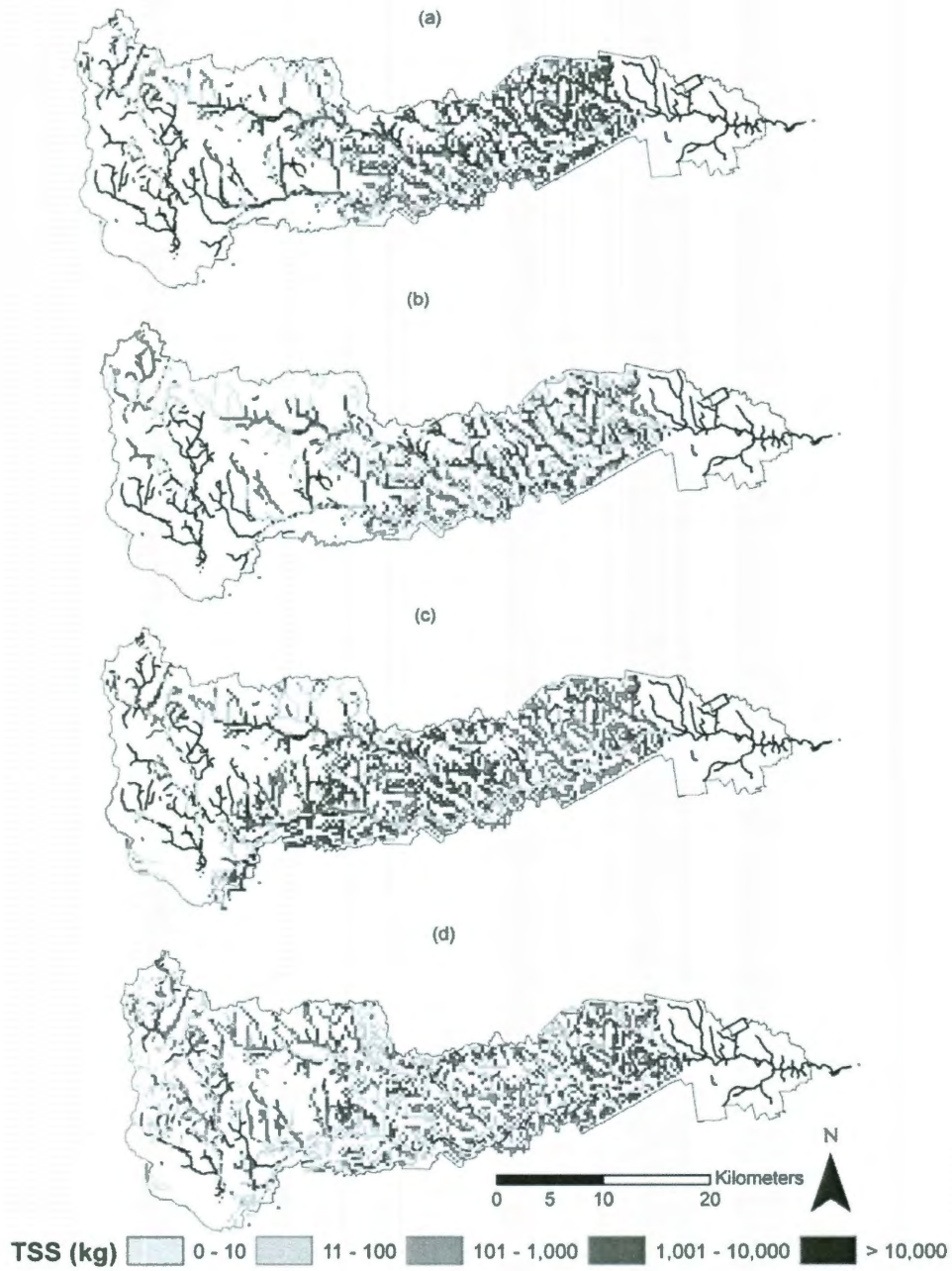
Figure 4-9. Spatial distribution of total washoff and waste loading to Cypress Creek during (a) July 7, 2009, (b) July 22, 2009, (c) September 22, 2009, and (d) July 7, 2010

4.4.   Discussion

The washoff and transport model exhibited good performance for a majority of these small rainfall events when compared to a storm water quality sampling of TSS. In general, the rising limb of the hydrograph is matched better than the falling limb of the hydrograph; however the limited number of observations restricts the assessment of model performance.

In the case of the September 22, 2009 storm, the model performed well on the rising limb of the pollutographs. However it did not match well for the two data points past the peak, which were sampled at 24 hours past the hydrologic peak, when the influence of point sources is greater. These two points account for the higher RMSE as well as the discrepancy in the modeled versus observed EMC. However, when these points are not included in the dataset, the model performs adequately, similar to the other storms. This lower flow, receding streamflow regime has been difficult to simulate and predict (Singh et al., 2005; Krause et al., 2005). Furthermore, point sources are a source of uncertainty. During storms, the hydraulic shock of the event results in a decrease in treatment plant performance as well as solids washout (Lessard and Beck, 1990). Therefore, the behavior of the many point sources within the watershed is unknown and could be a source of error in the simulation of pollutant transport.

The model also performed well for the July 7, 2009 and July 22, 2009 events. The highest RMSE was for the July 7, 2009 event, but the observed TSS for this event was of higher magnitude than the other events, and the samples were taken during a short

interval of the total storm simulation. The event had a longer build-up period, and greater maximum intensity of rainfall than any of the other storms. The modeled and observed EMC differed by 5.9% and 3.85% for the July 7, 2009 and July 22, 2009 storms, respectively. Even though the model performed poorly for the July 7, 2010 event, the observed and modeled EMCs differed only by 13.4%. Comparing all the modeled and observed data, Figures 4-6 and 4-7, shows that there is good general agreement. The modeled data provided a greater duration and density of data points than was feasible to collect given the limited resources for storm water sampling. Ideally, further storm water sampling at a greater frequency during storm events would be conducted to more extensively evaluate model performance.

The modeled TSS concentration throughout each of the events was then analyzed to further study the temporal dynamics of the mass loading of TSS in Cypress Creek. It was shown that the events varied from mass limited to flow limited events. The September 22, 2009 event strongly exhibited a first flush behavior where as the July 22, 2009 event exhibited no first flush phenomenon. The comparison of cumulative percent mass and volume streamflow is often analyzed, for BMP design purposes, for the control of the storm water pollution. The goal is to determine the volume of water that must be captured by the BMP in order to capture a fraction of the pollutant loading (Kayhanian and Stenstrom, 2005). These curves will vary from event to event based on variations of the hyetograph, antecedent buildup duration, and point source behavior (Betrand-Krajewski et al, 1998). Because Cypress Creek is a large and complex watershed without evident first flush, BMP selection, placement, and design should be considered on the

sub-watershed scale (Sansalone and Cristina, 2004) targeting the spatially variable sources and associated variable travel times. The fully distributed nature of the washoff and transport model allows for this level of detailed analysis. Further development of this fully distributed water quality model will provide a valuable tool for resource management and planning. Because this model allows the user to develop and analyze the data at any point in the watershed, this analysis can be implemented at multiple locations within the watershed for spatially targeted BMP design and placement. For instance, BMPs such as vegetated filter strips and retention basins can be sized and the expected performance quantified. This allows for the optimal location for the greatest impact by BMPs to be determined.

The distributed output of the washoff and transport was analyzed to identify the regions of high loading for each of the simulated events, thus providing an estimate of the primary source areas. Two of the events, which did not exhibit first flush phenomenon, had the highest loading from the downstream, urban areas of the watershed. In contrast, the event which distributed first flush phenomenon had the greatest loading in the center of watershed. This fully distributed output provides a foundation for future model development such as BMP functions, sediment transport routines, or non-conservative pollutant mass balance modules.

When analyzing the fully distributed output from the washoff and transport model, in terms of washoff mass and TSS concentration, it must be remembered, that due to resource limitation, the model was only calibrated to a single down stream location, thus

limiting the assessment and validation of the model. Future studies should incorporate storm water quality sampling at multiple locations within the watershed. By sampling more extensively both spatially and temporally, the washoff and transport model can be improved and evaluated with greater confidence.

Furthermore, future development of the model should include other transport processes, so that the model can be extended to other non-conservative constituents. However the current framework is an appropriate foundation for developing a fully distributed water quality model. The benefit of extending and improving this model, include the ability to spatially locate best management practices (BMPs) and simulate their impact. In addition the influence of land use change, wetland loss, and low impact developments can be assessed in a spatially explicit manner. Moreover a fully distributed pollutant transport model can be linked with spatially explicit estimates of pollutant loading (Teague et al, 2010) in order to analyze pollutant source populations and areas in a risk based framework. Future effort should include application of the model to simulate pollutant transport within different watersheds with different topography, soils, and land use characteristics to verify the robustness of the model.

4.5.   Conclusions

A distributed pollutant buildup, washoff, and transport model was developed in order to build upon the advancements in fully distributed hydrologic modeling. Coupled with the physically based rainfall-runoff model, $Vflo^{TM}$, this water quality model ran independently as a Java application. The model was applied to simulate the transport of

TSS in Cypress Creek watershed during rainfall events. Multiple events were simulated and the model was found to perform satisfactorily. The model's spatial and temporal output was then used to calculate the mass flow to analyze the first flush behavior of the watershed and spatial loading for each of the storms.

The framework of modeling the pollutant dynamics as represented in a distributed watershed grid, is a first step towards advanced modeling of pollutant transport. Despite the current limitations of this model, it provides a foundation for implementation of more complex pollutant dynamics and watershed features. This is the first step towards applying advanced hydrologic modeling and GIS technologies for the development of a fully distributed water quality model. Future efforts should include: further application within watersheds with different physical characteristics to ensure the robustness of the model, further development of the model to simulate the effect of BMPs, and the inclusion of the transport of specific non-conservative pollutants and other transport processes. This will allow for spatially explicit investigation of best management practices and land use evolution, both of which are key questions for resource management and planning.

Chapter 5 : Conclusions

In order to address the problem of water quality degradation in Lake Houston that results from pollutant loading from Cypress Creek watershed, a variety of statistical and modeling analyses were performed. The overall objective of this analysis was to develop a framework to investigate and address the pollutant loading to Cypress Creek and subsequent transport to Lake Houston.

Multivariate analysis, including principal component, cluster, and discriminant analysis, was used to determine the underlying seasonal pattern of the water quality data and then target load duration curves. This framework, combined with analysis of the correlations between water quality variables, was used to characterize pollutant sources. The analysis suggests that Cypress Creek has a complex mix of pollutant sources. Furthermore, the sources are not specifically tied to a climatic season. The presented method was shown to provide interpretation of large, complex, water quality datasets for improved decision support of BMP selection and resource management.

With the knowledge of the complexity of pollutant sources in Cypress Creek, stormwater quality was further investigated with hydrologic and water quality modeling. The analysis focused on small storm events, which are more frequent, and account for a majority of pollutant loading from Cypress Creek to Lake Houston.

A fully distributed hydrologic model was developed using the physically based model, $Vflo^{TM}$. NEXRAD radar rainfall or data from the existing rain gauge network was used to simulate high frequency small storms with less than two inches average rainfall. Based on a comparison of storm simulations using both radar rainfall and rain gauge data input, it was concluded that the spatial and temporal resolution of radar data improved the accuracy of hydrologic performance.

The results from the $Vflo^{TM}$ hydrologic simulation were then used as input into the independent pollutant washoff and transport model. The output of the water quality model was distributed estimations of total suspended solids concentration (TSS), as well as mass of TSS washed off the land surface at each time step of the simulation. The simulated concentration at the downstream water quality station was then compared to observed concentrations of TSS. For the four storms simulated, the model matched the observed TSS concentrations fairly well. The model output was then used to characterize the storm characteristics. It was determined that only one of the four storms exhibited a first flush behavior based on model output. Furthermore, the distributed washoff data was analyzed to determine the exported mass of TSS throughout the watershed. In each of the storms, the near riparian areas exhibited the greatest amount of washoff.

The presented pollutant washoff and transport model provides the foundation for future development of a water quality application that simulates pollutant dynamics within a watershed. Advanced hydrologic modeling and GIS datasets allow for a spatially explicit

analysis environment which can be used for sophisticated analyses of land use changes, best management practices, and other watershed characteristics.

Overall, the presented analyses have been shown to provide valuable tools for evaluation of water quality data and simulation of pollutant movement during rainfall events. These applications are the first step towards future development of a more comprehensive approach to managing water quality degradation and threats to the sustainability of our water resources.

Chapter 6 : References

Ahnert, P. R., M. D. Hudlow, E. R. Johnson, and D. R. Greene, (1983). Proposed "on-site" precipitation processing system for WSR- 88D. *Preprints, 21st Conf. on Radar Meteorology, Edmonton, AB, Canada, American Meteorological Society.*, 378–385.

Akan, A., and R. Houghtalen. 2003. *Urban Hydrology, Hydraulics, and Stormwater Quality*. John Wiley and Sons, Inc. Hoboken, NJ:345-347.

Akan, O. 1987. Pollutant washoff by overland flow. *Journal of Environmental Engineering*. 113(4):811-823.

Akan, O., F. Schafran, P. Pommerenk, and L. Harrell. 2000. Modeling storm-water runoff quanitity and quality from marine drydocks. *Journal of Environmental Engineering*. 126(1):5-11.

Albek, M., B. Ögütveren, and E. Albek. 2004. Hydrological modeling of Seydi Suyu watershed (Turkey) with HSPF. *Journal of Hydrology*. 285(1-4): 260-271.

Alley, W. and P. Smith. 1981. Estimation of accumulation for urban runoff quality modeling. *Water Resources Research*. 17(6):1657-1664.

Alberto, W., M. Diaz, M. Ame, F. Silvia, C. Andrea, and M. Bistoni, 2001. Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality a case study: Suquia River Basin (Cordoba-Argentina). *Water Research*. 35(12):2881-2894.

Alley, W. and P. Smith. 1981. Estimation of accumulation for urban runoff quality Modeling. *Water Resources Research*. 17(6):1657-1664.

Ambrose, R., T. Wool, and J. Martin. 1993. The water quality analysis simulation program, WASP5 Part A: Model Documentation. *US Environmental Protection Agency.* Acessed at ftp://ftp.hydroanalysisinc.com/MDEPWASP/MDEPWASP%20Manual%20A.pdf. Accessed on February 2, 2010.

Astel, A., M. Biziuk, A. Przyjazny, and J. Namiesnik, 2006. Chemometrics in monitoring spatial and temporal variations in drinking water quality. *Water Research*. 40(8):1706-1716.

Atasoy, M., R. Palmquist, and D. Phaneuf. 2006. Estimating the effects of urban residential development on water quality using microdata. *Journal of Environmental Management*. 79(4) : 399-408.

Avellaneda, P., T. Ballestero, R. Roseen, and J. Houle. 2009. On parameter estimation of urban storm-water runoff model. *Journal of Environmental Engineering.* 135(8):595-608.

Babbar-Sebens, M. and R. Karthikeyan, 2009. Consideration of sample size for estimating contaminant load reductions using load duration curves. *Journal of Hydrology.* 372:118-123.

Bandaragoda, C., D. Tarboton, and R. Woods. 2004. Application of TOPNET in the distributed model intercomparison project. *Journal of Hydrology.* 298(1-4):178-201.

Barbe, D., J. Cruise, and X. Mo. 1996. Modeling the buildup and washoff of pollutants on urban watersheds. *Water Resources Bulletin,* American Water Resources Association. 32(3):511-519.

Bedient, P., A. Holder, J. Benavides, B. Vieux. 2003.Radar-Based flood warning system applied to Tropical Storm Allison. *Journal of Hydrologic Engineering.* 8(6):308-318.

Bedient, P., A. Holder, J. Thompson, and Z. Fang. (2007). Modeling of storm-water response under large tailwater conditions: Case study of the Texas Medical Center. *Journal of Hydrologic Engineering.* 12(3): 256-266.

Bedient P. B. Hobit, D. Gladwell, and B. Vieux. (2000). NEXRAD radar for flood prediction in Houston. *Journal of Hydrologic Engineering.* 5(3): 269-277.

Bedient, P., W. Huber, and B. Vieux. 2008. *Hydrology and Floodplain Analysis.* Fourth Edition. Prentice Hall. Upper Saddle River, NJ. : 688-692, 469-470, 322,.

Bedient, P., R. Olsen, E. Baca, C. Newell, J. Lambert, J. Anderson, P. Rowe, C. Ward, E. Quevedo, D. Krentz, and R. Hamilton. 1980. Environmental study of the Lake Houston watershed: Phase 1. Final report, June 1980. Prepared by Rice University for City of Houston, Houston, Texas.

Belhateche, D., M. Turco, Y. Wang, J. Bleth. Solar-Powered circulators improve Lake Houston water quality. Accessed at http://www.tawwa.org/TW07Proceedings/070412a/WaterQuality/SolarPoweredCirculators.pdf. Accessed on November 4, 2009

Bengraine, K. and T. Marhaba, 2003. Using principal component analysis. *Journal of Hazardous Materials.* B100:179-195.

Berretta C, and J. Sansalone. 2011. Hydrologic transport and partitioning of phosphorus fractions. *Journal of Hydrology.* 403:25-36.

Betrand-Krajewski, J., G. Chebbo, and A. Saget. 1998. Distribution of pollutant mass vs volume in stormwater discharges and the first flush phenomenon. *Water Research.* 32(8):2341-2356.

Bicknell, B., J. Imhoff, J. Kittle, T. Jobes, and A. Donigan. 2001. Hydrological simulation program—FORTRAN (HSPF): *User's Manual. Version 12.* Athens, Ga.: U.S. Environmental Protection Agency. Accessed at http://www.epa.gov/waterscience/basins/b3docs/HSPF.pdf. Acessed on November 11, 2009.

Bonta, J. and B. Cleland, 2007. Incorporating natural variability, uncertainty, and risk into water quality evaluations using duration curves. *Journal of the American Water Resources Association (JAWRA).* 39(6):1481-1496.

Borah, D., and M. Bera. 2003.Watershed-Scale hydrologic and non-point source pollution models: review of Mathematical Bases. *Transactions of ASAE.* 46(6):1553-1566.

Borah, D., J. G. Arnold, M. Bera, E. Krug, and X. Liang. 2007. Storm event and continuous hydrologic modeling for comprehensive and efficient watershed simulations. *Journal of Hydrologic Engineering.* 12(6): 605-616.

Borah, D., G. Yagow, A. Saleh, P. Barnes, W. Rosenthal, E. Krug, and L. Hauck. 2006. Sediment and nutrient modeling for TMDL development and implementation. *Transactions of the American Society of Agricultural and Biological Engineers.* 49(4):967-986.

Borga, M. (2002). Accuracy of radar rainfall estimates for streamflow simulation. *Journal of Hydrology.* 267(2002):26-39.

Bosch, N. 2008. The influence of impoundments on river nutrient transport: An evaluation using the soil and water assessment tool. *Journal of Hydrology.* 355(1-4):131-147.

Boyer, J., J. Fourqurean And R. Jones, 1997. Spatial characterization of water quality in Florida Bay and Whitewater Bay by multivariate analyses: Zones of similar influence. *Estuaries and Coasts.* 20(4):743-758.

Brodie, I. and C. Rosswall. 2007. Theoretical relationships between rainfall intensity and kinetic energy variants associated with stormwater particle washoff. *Journal of Hydrology.* 340:40-47

Brown, L., and T. Barnwell. 1987. Enhanced stream water quality models QUAL2E and QUAL2E-UNCAS: Documentation and user manual. Environmental Protection Agency. Environmental Research Laboratory. Office of Research and Development. Report EPA/600/3-87/007, U.S. EPA, Athens, GA, USA.

Butcher, J. 2003. Buildup, washoff, and event mean concentration. *Journal of the American Water Resources Association.* 39(6):1521-1528.

Byrd, A. GSSHA Software Primer- An introduction to getting up and running with GSSHA. U.S. Army Corps of Engineers, Coastal & Hydraulics Laboratory. Accessed at http://chl.erdc.usace.army.mil/chl.aspx?p=s&a=PUBLICATIONS;223&g=109. Accessed on July, 2011.

Carpenter, T. and K. Georgakakos.(2004). Impacts of parametric and radar rainfall uncertainty on the ensemble streamflow simulations of a distributed hydrologic model. *Journal of Hydrology.* 298(2004): 202-221.

Carpenter, T., K. Georgakakos, and J. Sperfslagea. (2001). One the parametric and NEXRAD-radar sensitivities of a distributed hydrologic model suitable for operational use. *Journal of Hydrology.* 253(2001):169-193.

Castelli, V., A. Thomasian, and C. Li. 2003. CSVD: Clustering and singular value decomposition for approximate similarity search in high-dimensional spaces. *IEEE Transaction on Knowledge and Data Engineering.* 15(3):671-685.

Chapra, S. 1997. *Surface Water Quality Modeling.* McGraw-Hill. New York: 486-499.

Charbeneau, R. and M. Barret. 1998. Evaluation of methods for estimating stormwater pollutant loads. *Water Environment Research.* 70(7):1295-1302.

Chau, K. 2003. Manipulation of numerical coastal flow and water quality models. *Environmental Modeling & Software.* 18(2003):99-108.

Chellam, S., R. Sharma, G. Shetty, and Y. Wei. 2008. Nanofiltration of pretreated Lake Houston water: Disinfection by-product speciation, relationships, and control. *Separation and Purification Technology.* 64(2):160-169.

Chen, J. and B. Adams. 2006. A framework for urban storm water modeling and control analysis with analytical models. *Water Resources Research.* 42. W06419.

Chen, J. and B. Adams. 2007. A derived probability distribution approach to stormwater quality modeling. Advances in Water Resources. 30(1):80-100.

Chen, Y., S McCutcheon, R. Carsel, A. Donigian, and J. Craig. 1995. Validation of HSPF for the water balance simulation of the Upper Grande Ronde watershed. Oregon, USA. *Proceedings of Man's Influence on Freshwater Ecosystems and Water Use.* July 1995.No. 230.

Cigizoglu, H., and M. Bayazit, 2000. A generalized seasonal model for flow duration curve. *Hydrological Processes*. 14(6):1053-1067.

Cleland, B.R., 2003. TMDL development from the "Bottom up". Part III: Duration curves and wet-weather assessments. In: 2003 National TMDL Science and Policy Conference. Water Environment Federation, Chicago, IL. https://engineering.purdue.edu/~ldc/JG/duration/PDF/TMDL_Development_from the_Bottom_UP_PartIII.pdf. Acessed May 2010.

Dechesne, M., S. Barraud, J. Bardin, 2005. Experimental assessment of stormwater infiltration basin evolution. *Journal of Environmental Engineering*. 131(7):1090-1098.

Deletic, A. 1998. The first flush load of urban surface runoff. *Water Research*. 32(8), 2462–2470.

Deliman, P., W. Pack, and E. Nelson. 1999. Integration of the hydrologic simulation program-Fortran (HSPF) watershed water quality model in the Watershed Modeling System (WMS). U.S. Army Corp of Engineers, Engineer Research and Development Center. *Technical Report W -99-2*.

DeGaetano, A, 1996. Delineation of mesoscale climate zones in the northeastern United States using a novel approach to cluster analysis. *American Meteorological Society*. 9 (August): 1765-1782.

Delrieu, G., I. Braud, A. Berne, M Borga, B. Boudevillain, F. Fabry, J Freer, E, Gaume, E. Nakakita, A, Seed, P. Tabary, and R Uijlenhoet. (2009). Weather radar and hydrology. *Advances in Water Resources*. 32(7): 969-974.

Deng, Z., J. Lima, and V. Singh. 2005. Transport-Based model for overland flow and solute transport: Parameter estimation and process simulation. *Journal of Hydrology*. 315(2005):220-235.

Desai, A. 2009. Real-Time characterization and modeling of *Escheria coli* Contamination in Urban Streams. Thesis Proposal, University of Houston. December 2009.

DiToro, D., J. Fitzpatrick, and R. Thomann. 1983. Documentation for Water Quality Analysis Simulation Program (WASP) and Model Verification Program (MVP). US Environmental Protection Agency. EPA-600/3-81-044.

Dotto, C., M. Kleidorfer, A. Deletic, T. Fletcher, D. McCarthy and W. Rauch. 2010. Stormwater quality models: Performance and sensitivity analysis. *Water Science and Technology*. 62.4: 837-843.

Duncan, B. 2011. The impact of Palustrine wetland loss on flood peaks: An application of distributed hydrologic modeling in Harris County, Texas. Masters of Science Thesis, Rice University

Eclipse. 2001. Eclipse IDE for Java Developers. Helios Service Release 2. Accessed at http://eclipse.org/

Einfalt, T., and K. Arnbjerg-Nielsen, G. Golz, N. Jensen, M. Quirmbach, G. Vaes, and B. Vieux. (2004). Towards a roadmap for use of radar rainfall data in urban drainage. *Journal of Hydrology.* 299(2004):186-202.

Ellison, C., Q. Skinner, and L. Hicks, 2009. Assessment of best management practice effects on rangeland stream water quality using multivariate statistical techniques. *Rangeland Ecology and Management.* 62:371-386.

Endreny, T. 2002. Manipulating HSPF to simulate pollutant transport in suburban system. *Proceedings ASAE TMDL Meeting.* Fort Worth, Texas. March, 2002.

ESRI. 2000. Census 2000 TIGER. Accessed at http://www.esri.com/data/download/census2000-tigerline/index.html. Accessed on Dec 2010.

ESRI. 2008. ArcMap 9.3. ESRI Inc. Redlands, California.

Fang, Z., A. Zimmer, P. Bedient, H. Robinson, C. Christian, and B. Vieux. 2010. Using a distributed hydrologic model to evaluate the location of urban development and flood control storage. *Journal of Water Resources Planning and Management-ASCE.* 136(5):597-601.

Fang, Z., P. Bedeint, J. Benavides, and A. Zimmer. 2008. Enhanced rader-based flood alert system and floodplain map library. *Journal of Hydrologic Engineering.* 13(10):926-938.

Farabi, H. (2005). Modelling the hydrological connection of forest roads as a source of sediment to stream. Proceedings of the 2005 International Conference on simulation and modeling. V. Kachitvichyanukul, U. Purintrapiban, and P. Utayopas, eds.

Gan, T., E. Dlamini, and G. Biftu. (1997). Effects of model complexity and structure, data quality, and objective functions on hydrologic modeling. *Journal of Hydrology.* 192(1997): 81-103.

Gourley, J. and B. Vieux. 2005. A method for evaluating the accuracy of quantitative precipitation estimates from hydrologic modeling perspective. *Journal of Hydrometeorology.* 6(2) : 115-133

Gourley, J. and B. Vieux. 2006. A method for identifying sources of model uncertainty in rainfall-runoff simulations. *Journal of Hydrology*. 327(1-2):68-80.

Guber, A., A. Yakirevich, A. Sadeghi, Y. Pachepsky, D. Shelton. 2009. Uncertainty evaluation of coliform bacteria removal from vegetated filter strip under overland flow condition. *Journal of Environmental Quality*. 38(4):1638-1644.

Gupta, I., S. Dhage, and R. Kumar, 2009. Study of variations in water quality of Mumbai coast through multivariate analysis techniques. *Indian Journal of Marine Sciences*. 38(2):170-177.

Haag, I. and B. Westrich, 2002. Processes governing river water quality identified by principal component analysis. *Hydrological Processes*. 16:3113-3130.

Haan, C. 2002. *Statistical Methods in Hydrology*. 2nd Edition. Iowa State Press. Ames, Iowa: 340-346.

Hagedorn, C. S. Robinson. J. Filtz, S. Grubbs. T. Angier, and R. Reneau, 1999. Determining sources of fecal pollution in a rural Virginia watershed with antibiotic resistance patterns in fecal streptococci. *Applied and Environmental Microbiology*. 65(12): 5522-5531.

Harwood, V., J. Whitlock, and V. Withington, 2000. Classification of antibiotic resistance patterns of indicator by discriminant analysis: Use in predicting the source of fecal contamination in subtropical waters. *Applied and Environmental Microbiology*. 66(9):3698-3704.

Hathaway, J. and W. Hunt. 2011. Evaluation of first flush for indicator bacteria and total suspended solids in urban stormwater runoff. *Water Air Soil Pollution*. 217:135-137.

HCOEM.(2010). Rainfall Map. Harris County Office of Emergency Management. Accessed at http://www.hcoem.org/HCRainfall.aspx. Accessed on Novemeber 2010.

Heathcote, I. 2009. *Integrated Watershed Management: Principles and Practice*. Second Edition. John Wiley & Son, Inc. Hoboken, NJ. 243.

Helena, B., R. Pardo, M. Vega, E. Barrado, J. Fernandez, and L. Fernandex, 2000. Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. *Water Research*. 34(3):807-816.

H-GAC, 2004a Cypress Creek. find your watershed. Houston-Galveston Area Council. Acessed at http://ntis04.hgac.cog.tx.us/website/bsr06/Cypresscreek.pdf . Accessed on June 2009.

H-GAC, 2004b Spring Creek. find your watershed. Houston-Galveston Area Council. Acessed at http://ntis04.hgac.cog.tx.us/website/bsr06/Springcreek.pdf . Accessed on May 2010.

H-GAC, 2008. Houston-Galveston Coordinated Monitoring Program. Accessed at http://ntis04.hgac.cog.tx.us/website/mtpasp/asp/viewer.htm. Accessed on May 2010.

H-GAC. 2008. 2008 Land Cover Data Set. Regional Data & GIS Data Services, Houston-Galveston Area Council. Accessed June 2011, Acessed at http://www.h-gac.com/community/socioeconomic/land_use/default.aspx.

H-GAC. 2009. H-GAC Clean Rivers Program. Houston-Galveston Area Council. Accessed at http://webgis2.h-gac.com/CRPflex/. Accessed on September 1, 2009.

H-GAC, 2009. "Permitted Domestic WWTP". Clean Rivers Program, Houston-Galveston Area Council .

Hirsch, R., Slack, J.R., 1984. A nonparametric trend test for seasonal data with serial dependence. *Water Resources Research* 20 (6), 727–732.

Hirsch, R.M., Slack, J.R., Smith, R.A., 1982. Techniques of trend analysis for monthly water quality data. *Water Resources Research* . 18 (1), 107–121.

Huber, W., J. Heaney, M. Medina, W. Peltz, and H. Sheikh. 1975. Stormwater management model: User's manual, Version II. U.S. Environmental Protection Agency. EPA-670/2-75-017.

Huber, W. 1993. Contaminant transport in surface water. *Handbook of Hydrology*. D. Maidment (Ed). McGraw-Hill, Inc, New York, NY. p14.7.

Huber, W., L. Rossman, and R. Dickinson. 2005. EPA Storm Water Management Model, SWMM5. Chapter 14 in *Watershed Models*. Edited by V. Singh and D. Frevert. CRC Press, Boca Raton, FL. 2005.

Hydrologic Engineering Center (HEC). 2002. HEC-RAS, River Analysis System, U.S. Army Corps of Eng., Davis, Calif.

James, R., J. Martin, T. Wool., P. Wang. 1997. A sediment resuspension and water quality model of Lake Okeechobee. *Journal of the American Water Resources Association.* 33(3):661-678.

James, W. and A. Kuch. 1998. Sensitivity-calibration decision-support tools for continuous SWMM modeling: A fuzzy-logic approach". Chapter 9, *Modeling the Management of Stormwater Impacts*. Edited by W. James. Monograph 6. Proceedings of Conference on the Stormwater and Water Quality Management Modeling Conference, Toronto, 1997, Computational Hydraulics International, Guleph, Ontario.

James, W., C. Robinson, and J. Bell. 1993. Radar-assisted real time flood forecasting. *Journal of Water Resources Planning and Management*. 119(1):32-44.

Jayakrishnan, R., R. Srinivasan, C. Santhi, and J. Arnold, 2005. Advances in the application of the SWAT model for water resources management. *Hydrological Processes*. 19:749-762.

Johanson, R. D. Imhoff, and H. Davis. 1980. User's manual for hydrological simulation program- FORTRAN (HSPF). US Environmental Protection Agency, Athens GA. EPA-600/9-80-015.

Johnson, S., T. Whiteaker, and D. Maidment, 2009. A tool for automated load duration curve creation. *Journal of the American Water Resources Association*. 45(3):654-663.

Jonnalagadda, S. and G. Mhere. 2001. Water quality of the Odzi River in the eastern highlands of Zimbabwe. *Water Research*. 35(10): 2371-2376.

Kahya, E. and S. Kalayci. 2004. Trend analysis of streamflow in Turkey. *Journal of Hydrology*. 289:128-144.

Kayhanian, M. and M. Stenstrom. 2005. Mass loading of first flush pollutants with treatment strategy simulations. *Transportation Research Record: Journal of the Transportation Research Board*. 1904:133-143.

Kellershohn, D., and I. Tsanis. 1999. 3D eutrophication modeling of Hamilton Harbour: Analysis of remedial options. *Journal of Great Lakes Research*. 25(1):3-25.

Kim, B. S., Kim, B. K., & Kim, H. S. 2008. Flood simulation using the gauge-adjusted radar rainfall and physics-based distributed hydrologic model. *Hydrological Processes*, 22, 4400-4414.

King, K., J. Arnold, and R. Bingener. 1999. Comparison of Green-Ampt and Curve Number Methods on Goodwin Creek watershed using SWAT. *Transactions of ASAE*. 42(4):919-925.

Koklu, R., B. Sengorur, and B. Topal, 2010. Water quality assessment using multivariate statistical methods—A case study: Melen River System (Turkey). *Water Resources Managament*. 24:959-978.

Koren, V. B. Finnerty, J. Schaake, M. Smith, D. Seo, and Q. Duan. 1999. Scale dependencies of hydrologic models to spatial variability of precipitation. *Journal of Hydrology*. 217(3-4):285-302.

Krajewski, W., and J. Smith. 2002. Radar hydrology: Rainfall estimation. *Advances in Water Resources*. 25(2002)1837-1394.

Krause, P., D. Boyle, and F. Base. 2005. Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences*. 5:89-97.

Kumar, K., B. George, R. Kumar, P. Sajish, and S. Viyol.,2009. Assessment of spatial and temporal fluctuations in water quality of a tropical permanent estuarine system- Tapi, West Coast India. *Applied Ecology and Environmental Research*. 7(3):267-276.

Lambraikis, N., A. Antonakos, and G. Panagopoulos, 2004. The use of multicomponent statistical analysis in hydrogeological environmental research. *Water Research*. 38:1832-1872.

Lautz, L and D. Siegal. (2006). Modeling surface and groundwater mixing in the hyporheic zone using MODFLOW and MT3D. *Advances in Water Resources*. 29(2006)1618-1633.

Lee, L., C. Lu, and S. Kung, 2004. Spatial diversity of chlorine residual in a drinking water distribution system. *Journal of Environmental Engineering*. 130(11):1263-1268.

Lessard, P. and M. Beck. 1990. Operational water quality management: Control of storm sewage at a wastewater treatment plant. *Research Journal of the Water Pollution Control Federation*. 62(6):810-819.

Li, S., and Q. Zhang, 2010. Spatial characterization of dissolved trace elements and heavy metals in the upper Han River (China) using multivariate statistical techniques. *Journal of Hazardous Materials*. 176:579-588.

Liscum, F. and J. East. 2000. Estimated effects on water quality of Lake Houston from interbasin transfer of water from the Trinity River, Texas. U.S. Geological Survey. *Water-Resources Investigations Report* 00-4082, 50 p.

Liu, C., K. Lin, and Y. Kuo, 2003. Application of factor analysis in the assessment of groundwater quality in a Blackfoot Disease Area in Taiwan. *The Science of the Total Environment*. 313:77-89.

Lohani, V., D. Kibler, and J. Chanat. 2002. Constructing a problem solving environment tool for hydrologic assessment of land use change. *Journal of the American Water Resources Association.* 38(2):439-452.

Looper, J. P., Vieux, Baxter E., & Moreno, M. A. (2009). Assessing the impacts of precipitation bias on distributed hydrologic model calibration and prediction accuracy. *Journal of Hydrology.*

Lung, W. 2001. Water quality modeling for wasteload allocations and TMDLs. John Wiley 7 Sons, New York: 188-199.

Ma, J., G. Ying, and J. Sansalone. 2010. Transport and distribution of particulate matter phosphorus fractions in rainfall-runoff from roadway source areas. *Journal of Environmental Engineering.* 136(11):1197-1205.

Mahamoud, Y., R. Pamar, K. Wolfe, and J. Carleton. 2008. HSPF Toolkit: A tool for stormwater management at the watershed scale. Proceedings of the Water Environment Federation. Sustainability 2008: 421-421.

Mahloch, J. 1974. Multivariate techniques for water quality analysis. *Journal of the Environmental Engineering.* 100(5): 1119-1132.

Maidment. 2006. "ArcHydro Data Model". Center for Research in Water Resources, University of Texas at Austin. Accessed at http://www.crwr.utexas.edu/giswr/hydro/ArcHOSS/Downloads/index.cfm.

Mannina, G. and G. Viviani. 2010. Water quality modeling for Ephemeral Rivers: Model development and parameter assessment. *Journal of Hydrology.* 393(3-4):186-196.

Matty, J., J. Anderson, and R. Dunbar. 1987. Suspended Sediment Transport, Sedimentation, and Resuspension in Lake Houston, Texas: Implications for Water Quality. *Environmental Geology.* 10(3):175-186.

Millar, R. 1999. Analytical determination of pollutant wash-off parameters. *Journal of Environmental Engineering.* 125(10):989-992.

Miertschin & Associates, Inc. 2009. Technical support document for indicator bacteria Total Maximum Daily Loads, Lake Houston watershed. Prepared for Texas Commission on Environmental Quality. September 2009. Accessed on June 2011. Accessed at http://www.tceq.texas.gov/assets/public/implementation/water/tmdl/82lakehousto n/82-lakehouston-tsd.pdf

Min, J. and W. Wise. 2010. Depth-averaged, spatially distributed flow dynamic and solute transport modeling of a large-scaled, subtropical constructed wetland. Hydrological Processes. 24(19):2724-2737.

Moore, L., C. Chew, R. smith, and S. Sahoo. 1992. Modeling of best management practices on North Reelfoot Creek, Tennessee. *Water Environment Research.* 64(3):241-247.

Morales,M. P. Marti, A. Llopis, L. Campos, and S. Sagrado, 1999. An environmental study by factor analysis of surface seawaters in the gulf of Valencia (Western Mediterranean)". *Analytica Chimica Acta.* 394:109-117.

Morrison, M., and J. Bonta, 2008. Development of duration-curve based methods for quantifying variability and change in watershed hydrology and water quality. US Environmental Protection Agency Office of Research and Development. EPA/600/R-08/065. http://www.epa.gov/nrmrl/pubs/600r08065/600r08065.pdf. Acessed May 2010.

Muxika, I. and A. Borja, and J. Bald, 2007. Using historical data, expert judgment and multivariate analysis in assessing reference conditions and benthic ecological status according to the European Water Framework Directive. *Marine Pollution Bulletin.* 55: 16-29.

Nakamura, E. 1984. Factors affecting stormwater quality decay coefficient. Proceedings of the Third International Conference on Urban Storm Drainage. Goteburg, Sweden. June, 1984. 3: 979-988.

Najafpour,S., A. Alkarkhi, M. Kadir, and G. Najafpour, 2008. Evaluation of spatial and temporal variation in river water quality. *International Journal of Environmental Resources.*2(4):349-358.

Natural Resources Conservation Service (NRCS), United States Department of Agriculture, 2006. U.S. General Soil Map (STATSGO2) for Texas. http://soildatamart.nrcs.usda.gov Accessed April 26, 2010..

Newell, C. 1981. An assessment of point and nonpoint pollution loads into Lake Houston. Master's Thesis. Rice University, Houston, Texas.

Newham, L.T.H., J.C. Rutherford, and B.F.W. Croke. 2005. A conceptual model of particulate trapping in riparian buffers, *CSIRO Land and Water Technical Report* 21/05. CSIRO Land and Water, Canberra.

Nikolaidis, N. H. heng, R. Semagin, J. Clausen. 1998. Non-Linear response of a mixed land use watershed to nitrogen loading. *Agriculture, Ecosystems and Environment.* 67(1998):251-265.

NOAA, 2010. Houston: Extremes, normals and annual summaries. National Weather Service Weather Forecast Office. Accessed at: http://www.srh.noaa.gov/hgx/?n=climate_iah_normals_summary. Accessed May 2010.

Novotny, V. and G. Chesters. 1981. *Handbook of nonpoint pollution sources and management*. Van Nostrand Reinhold Company. New York, New York.

NRCS. 2006. NRCS Web Soil Survey. US. Department of Agriculture. Accessed at http://websoilsurvey.nrcs.usda.gov/app/.

Obropta, C. and J. Kardos. 2007. Review of urban stormwater quality models: Deterministic, stochastic, and hybrid approaches. *Journal of the American Water Resources Association*. 43(6):1508-1523.

Oden, T.D., and Graham, J.L., 2008. Watershed influences and in-lake processes—A regional-scale approach to monitoring a water-supply reservoir, Lake Houston Near Houston, Texas: U.S. Geological Survey Fact Sheet 2008-3003.

Overton, D. and M. Meadows. 1976. *Stormwater Modeling*. Academic Press. New York: 304-325.

Ouyang, Y, 2005. Evaluation of river water quality monitoring stations by principal component analysis. *Water Research*. 39:2621-2635.

Ouyang, Y., P. Nkedi-Kizza, Q. Wu, D. Shinde, and C. Huang, 2006. Assessment of seasonal variations in surface water quality. *Water Research*. 40:3800-3810.

Panda, U., S. Sundaray, P. Rath, B. Nayak, D. Bhatta, 2006. Application of factor and cluster analysis for characterization of river and estuarine water systems- A case study: Mahanadi River (India). *Journal of Hydrology*. 331:434-445.

Parinet, B.,A. Lhote, and B. Legube. 2004. Principal component analysis: An appropriate tool for water quality evaluation and management- Application to a tropical lake system. *Ecological Modeling*. 178:295-311.

Park, S., J. Choi, S. Wang, and S. Park, 2006. Design of a water quality monitoring network in a large river system using the genetic algorithm. *Ecological Modeling*. 199: 289-297.

Patry, G. and A. Kennedy. 1989. Pollutant washoff under noise-corrupted runoff conditions. *Journal of Water Resources Planning and Management*. 115(5):646-657.

Paul, S., R. Sirnivasan, J. Sanabria, P. Haan, S. Mukhtar, and K. Neimann. 2006. Groupwise modeling study of bacterially impaired watersheds in Texas:Clustering analysis. *Journal of the American Water Resources Association*. 42(4): 1017-1031.

Petersen, T., C. Sowells, H. Rifai, R. Stein. 2008. Comparisons of methods to calculate TMDLs for E. coli: Load duration curves, mass balance, and HSPF. Proceedings of the World Environmental Water Resources Congress 2008. 416.

Petersen, T., H. Suarex. Rifai, P. Jensen, Y. Su, and R. Stein. 2006. Status and trends of fecal indicator bacteria in twp urban watersheds. *Water Environment Research.* 78(12): 2340-2355.

Petersen W., L. Bertino, U. Callies, and E. Zorita, 2001. Process identification by principal component analysis of river water-Quality data. Ecological Modeling. 138(1-3:193-213.

Poli, A. and M. Cirillo. 1993. On the use of the normalized mean square error in evaluating dispersion model performance. *Atmospheric Environment.* 27A(15):2427-2434.

Qi, C. and S. Grunwald. 2005. GIS-Based hydrologic modeling in the Sandusky watershed using SWAT. *Transaction of ASAE.* 48(1): 169-180.

Ravichandran, S. R. Ramanibai, and N. Pundarikanthan, 1996. Ecoregions for describing water quality patterns in Tamiraparani Basin, South India. *Journal of Hydrology.* 178:257-276.

Razmkhah, H., A. Abrishamchi, and A. Torkian, 2010. Evaluation of spatial and temporal variation in water quality by pattern recognition techniques: A case study on Jajrood River (Tehran, Iran). *Journal of Environmental Management.* 91(4):852-860.

Refsgaard, J., H. Henriksen, W. Harrar, H. Scholten, and A. Kassahun, 2005. Quality assurance in model based water management- Review of existing practice and outline of new approaches. *Environmental Modeling and Software.* 20:1201-1215.

Riebscleager, K. 2008. Development and application of the Spatially Explicit Load Enrichment Calculation Tool (SELECT) for determining Total Maximum Daily Loads. Master's Thesis. Texas A&M University, College Station, Texas.

Robinson, H., Z. Fang, and P. Bedient. 2009. Distributed hydrologic model for flood prevention in the Yuna River watershed, Dominican Republic. *Proceedings of World Environmental and Water Resources Congress 2009: Great Rivers.* Kansas City. American Society of Civil Engineers.

Rosenfeld, D., D. Wolff, and D. Atlas. (1993). General probability-matched relationships between radar reflectivity and rain rate. *Journal of Applied Meteorology.* 32:50-72.

Rosenthal, W., R. Srinivasan, and J. Arnold. 1995. Alternative river management using a linked GIS-Hydrology model. Transactions of the ASAE. 38(3):783-790.

Rossman, L. 2004. Stormwater management model: User's manual Version 5.0. U.S. Environmental Protection Agency, National Risk Management Research Laboratory, Office of Research and Development.

Rothstein, E. 2001. Regional conversion to surface water supplies in Harris County, Texas. *2001 New Horizons to Drinking Water Annual Conference.* Washington, DC, June 2001 : 17-21.

Ryu, J. 2009.Application of HSPF to the Distributed Model Intercomparison Project: Case study. *Journal of Hydrologic Engineering.* 14(8): 847-857.

Safiolea, E., P. Bedient, and B. Vieux. 2005. Assessment of the relative hydrologic effects of land use change and subsidence using distributed modeling. Managing Watersheds for Human and Natural Impacts: Engineering, Ecological, and Economic Challenges .Proceedings of Watershed 2005. 178, 87 (2005).

Safiolea, E. 2006. Advanced techniques for distributed hydrologic modeling of an urban environment. PhD Dissertation. Rice University, Houston Texas. Accessed at http://proquest.umi.com/pqdlink?vinst=PROD&attempt=1&fmt=6&startpage=-1&ver=1&vname=PQD&RQT=309&did=1155557321&exp=11-11-2014&scaling=FULL&vtype=PQD&rqt=309&cfc=1&TS=1258054796&clientId=480. Accessed on November 12, 2009.

Sansalone, J. and C. Cristina. 2004. First flush concepts for suspended and dissolved solids in small impervious watersheds. *Journal of Environmental Engineering.*130(11):1301-1314. doi:10.1061/(ASCE)0733-9372(2004)130:11(1301)

Sansalone, J., J. Koran, J. Smithson, and S. Buchberger. 1996. Physical characteristics of urban roadway solids transported during rain evernts. *Journal of Environmental Quality.* 124(5): 427-440.

Santhi,C., R. Srinivasan, J. Arnold, and J. Williams, 2006. A modeling approach to evaluate the impacts of water quality management plans implemented in a watershed in Texas. *Environmental Modeling & Software.* 21(8):1141-1157.

Sarbu, C. and H. Pop, 2005. Principal component analysis versus fuzzy principal component analysis A case study: The quality of Danube water (1985-1996). *Talanta.* 65:1215-1220.

Sartor, J. and G. Boyd. 1972. Water pollution aspects of street surface contamination. U.S. Einvironmental Protection Agency. EPA R2-72-081.

SAS. 2003. SAS User's Guide: Statistics. Ver. 8. Cary, N.C.: SAS Institute, Inc.

Sawyer, C., P. McCarty, and G. Parkin. 2003. *Chemistry for Environmental Engineering and Science*. Fifth Edition. McGraw- Hill Publishing. New York: 77-80.

Scanlon, B., I. Jolly, M. Sophocleous, and L. Zhang. 2007. Global impacts of conversions from natural to agricultural ecosystems on water resources: Quantity vs quality. *Water Resources Research*. 43: WO3437.

Seo. D. (1998). Real-time estimation of rainfall fields using radar rainfall and rain gage data. *Journal of Hydrology*. 208(1998):37-52.

Seo, D., and J. Breidenbach. 2002. "Real-time correction of spatially nonuniform bias in radar rainfall data using rain gage measurements. *Journal of Hydrometeorology*. 3:93-111.

Shanahan, P., M. Henze, L. Koncsos, W. Rauch, P. Reichert, L. Somlyody, and P. Vanrolleghem. 1998. River water quality modeling: II. Problems of the art. *Water Science Technology*. 38(11):245-252.

Sharif, H., L. Sparks, A. Hassan, J. Zeitler, and H. Xie. (2010). Application of a distributed hydrologic model to the November 17, 2004, flood of Bull Creek Watershed, Austin, Texas. *Journal of Hydrologic Engineering*. 15(8):651-657.

Shaw, S. J. Stedinger, and M. Walter. (2010). Evaluating urban pollutant buildup/wash-off models using a Madison, Wiscosin Catchment. *Journal of Environmental Engineering*. 136(2):194-203.

Sheng, Y., G. Ying, and J. Sansalone. 2008. Differentiation of transport for particulate and dissolved water chemistry load indices in rainfall-runoff from urban source area watersheds. *Journal of Hydrology*. 361:144-158.

Shin, R., and K. Fong, 1999. Multiple discriminant analysis of marine sediment data. *Marine Pollution Bulletin*. 39(1-12):285-294.

Shirmohammadi, A., I. Chaubey, R. Marmel, D. Bosch, R. Munoz-Carpena, C. Dharmasri, A. Sexton, M. Arabi, M. Wolfe, J. Frankenberger, C. Graff, and T.Sohrabi, 2006. Uncertainty in TMDL models. *Transactions of ASABE*. 49(4):1033—1049.

Shoemaker, L., T. Dai, and J. Koenig. 2005. TMDL model evaluation and research needs. Tetra Tech, Inc for the U.S. Environmental Protection Agency. National Risk Management Research Laboratory, Cincinnati, Ohio. EPA/600/R-05/149.

Shrestha,S. and F. Kazama, 2007. Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji River Basin. Japan. *Environmental Modeling & Software*. 22:464-475.

Shrestha, S. F. Kazama, and T. Nakamura, 2008. Use of principal component analysis, factor analysis, and discriminant analysis to evaluate spatial and temporal variations in water quality of the Mekong River. Journal of Hydroinformatics. 10(1):43-56

Silva, L., and D. Williams. 2001. Buffer zone versus whole catchment approaches to studying land use impact on river water quality. *Water Research.* 35(14):3462-3472.

Simeonov, V., J. Stratis, C. Samara, G. Zachariadis, D. Voutsa, A. Anthemidis, M. Sofoniou, and T. Kouimtzis, 2003. Assessment of the surface water quality in Northern Greece. *Water Research.* 37:4119-4124.

Singh, V. 2002a. Kinematic wave solutions for pollutant transport by runoff over an impervious plane, with instantaneous or finite-period mixing. *Hydrological Processes.* 16(9):1831-1863.

Singh, V. 2002b. Kinematic wave solutions for pollutant transport over an infiltrating plane with finite-period mixing and mixing zone. *Hydrological Processes.* 16:2441-2477.

Singh, J., H. Knapp, J. Arnold, and M. Demissie. 2005. Hydrological modeling of the Iroquois River watershed using HSPF and SWAT. *Journal of the American Water Resources Association.* 41(2):343-360.

Singh, K., A. Malik, D. Mohan, S. Sinha, 2004. Multivariate statistical techniques of the evaluation of spatial and temporal variations in water quality of Gomti River (India)- A case study. *Water Research.* 38:3980-3992.

Singh, K., A. Malik, and S. Sinha, 2005. Water quality assessment and apportionment of pollution sources of Gomti River (India) using multivariate statistical techniques- A case study. *Analytica Chimica Acta.* 538:355-374.

Shaw, S. J. Stedinger, and M. Walter. (2010). Evaluating urban pollutant buildup/wash-off models using a Madison, Wiscosin Catchment. *Journal of Environmental Engineering.* 136(2):194-203.

Smith, R., G. Schwarz, R. Alexander, 1997. Regional interpretation of water-quality monitoring data. *Water Resources Research.* 33(12):2781-2798.

Smyer, S. 2008. History of the City of Houston's drinking water operations. City of Houston Public Works , Houston, TX. Acessed at http://documents.publicworks.houstontx.gov/documents/divisions/utilities/history of_drinking_water_operations.pdf. Accessed on Nov. 4, 2009.

Sneck-Fahrer, D., M. Milburn, J. East, and J. Oden. 2005. Water-Quality assessment of Lake Houston Near Houston, Texas, 2000-2004. US. Geologic Survey. Scientific Investigations Report 2005-5241.

Soonthornnonda, P., E. Christensen, L.Yang, and L. Jin. 2008. A washoff model for stormwater pollutants. *Science of The Total Environment* . 402(2-3): 248-256.

Stieber, P, M. Bues, and B. Granjean. 1999. Simple approach to characterizing a pollution event return period in urban hydrology, by means of hydrological parameters. *Hydrological Sciences Journal*. 44(2):183-198.

Stow, C., D. Roessler, M. Borsulk, and J. Bowen and K. Reckhow. 2003. Comparison of estuarine water quality models for Total Maximum Daily Load development in Neuse river estuary. *Journal of Water Resources Planning and Management*. 129(4):307-314.

Strobl, R. and R. Robillard, 2008. Network design for water auality monitoring of surface freshwaters: A Review. *Journal of Environmental Management*. 87:639-648.

Sun, X. and R. Mein, T. Keenan, and J. Elliot.(2000). Flood estimation using radar and raingage data. *Journal of Hydrology*. 239(1-4):4-18.

Suk, H. and K. Lee,1999. Characterization of a ground water hydrochemical system through multivariate analysis: Clustering into ground water zones. *Ground Water*. 37(3):358-366.

Syme, W.J. 2001. TUFLOW - Two & one-dimensional unsteady FLOW software for rivers, estuaries and coastal waters. The Institution of Engineers, Australia.

TCEQ, 2008a, 2008 Guidance for assessing and reporting surface water quality in Texas. Texas Commission on Environmental Quality, Surface Water Quality Monitoring Program. Austin, Tx. http://www.tceq.state.tx.us/assets/public/compliance/monops/water/08twqi/2008_guidance.pdf. Accessed on May 2010.

TCEQ. 2008b. Revisions to §307 - Texas surface water quality standards. Texas Commission on Environmental Quality. April 9, 2008. http://www.epa.gov/waterscience/standards/wqslibrary/tx/tx-wqs.pdf. Access on November 4, 2009.

TCEQ, 2008b, 2008 Texas water quality inventory and 303(d) list. Texas Commission on Environmental Quality. March 19, 2008. http://www.tceq.state.tx.us/assets/public/compliance/monops/water/08twqi/2008_303d.pdf. Access on May 2010.

TCEQ, 2008c, 2008 Texas water quality inventory: Water bodies with concerns for use attainment and screening levels. Texas Commission on Environmental Quality March 19, 2008. http://www.tceq.state.tx.us/assets/public/compliance/monops/water/08twqi/2008_concerns.pdf. Accessed on May 2010.

TCEQ. 2010. Procedures to implement the Texas surface water quality standards. Texas Commission on Environmental Quality, Water Quality Division. RG-194 . June 2010. Accessed June 25, 2011. Acessed at http://www.tceq.state.tx.us/assets/public/legal/rules/rule_lib/adoptions/RG-194.pdf.

Teague, A. 2007. Spatially explicit load enrichment calculation tool and cluster analysis for identification of *E. coli* sources in Plum Creek watershed, Texas. Masters Thesis. Texas A&M University, College Station, Texas.

Teague, A., R. Karthikeyan, M. Babbar-Sebens, R. Srinivasan, and R. Persyn., 2009. Spatially explicit load enrichment calculation tool to identify potential E.coli sources in watersheds. *Transactions of ASABE*. 54(4): 11009-1120.

Teague, A. and P. Bedient. 2011. Use of radar rainfall in an application of distributed hydrologic modeling for Cypress Creek watershed, Texas. *Journal of Hydrologic Engineering*. Submitted Dec 2010, Accepted.

Thomann, R., and J. Mueller. 1987. *Principles of Surface Water Quality Modeling and Control*. Harper & Row, Publishers. New York: 261-340.

Thyne, G., C. Guler, and E. Poeter, 2004. Sequential analysis of hydrochemical data for watershed characterization. *Ground Water*. 42(5):711-723.

Tong, S. and W. Chen. 2002. Modeling the relationship between land use and surface water quality. *Journal of Environmental Management*. 66:377-393.

TSARP, 2005. "Tropical Storm Allison Recovery Project" Accessed at http://www.tsarp.org/. Accessed April 2010.

TSARP. 2006. "Tropical Storm Allison Recovery Project". Accessed at http://www.tsarp.org. Accessed April 2010.

TWDB. 2004. Surface water, lake volumetric surveys. Texas Water Development Board Accessed January 15, 2005 at http://www.twdb.state.tx.us/data/surfacewater/surfacewater_toc.asp

U.S. Environmental Protection Agency (USEPA), 2007. An approach for using load duration curves in the development of TMDLs. Office of Wetlands, Oceans and Watersheds, EPA 841-B-07-006, Washington, DC. http://water.epa.gov/lawsregs/lawsguidance/cwa/tmdl/upload/2007_08_23_tmdl_ duration_curve_guide_aug2007.pdf. Acessed August 2010.

U.S. Environmental Protection Agency (USEPA), 2005. *National Hydrography Dataset.* http://www.horizon-systems.com/nhdplus/. Accessed on May 2010.

U.S. Environmental Protection Agency (USEPA), 2009. Revisions to §307 – Texas surface water quality standards. Texas Commission on Environmental Quality, Surface Water Quality Monitoring Program. Austin, Tx. http://www.epa.gov/waterscience/standards/wqslibrary/tx/tx-wqs.pdf. Accessed on April 2010.

U.S. Environmental Protection Agency (USEPA), 2010. EnviroFacts. http://www.epa.gov/enviro/index.html. Accessed on September 2010.

U.S. Geologic Survey (USGS), 2007. National elevation dataset. Digital elevation model technologies and applications. http://ned.usgs.gov/. Accessed April 2010.

U.S. Geologic Survey (USGS), 2010. USGS water data for Texas. http://waterdata.usgs.gov/tx/nwis/. Accessed on April 2010.

USGS. 2002. Water-Quality trends in suburban Houston, Texas, 1954-1997, as indicated by sediment cores from Lake Houston. US Geological Survey. USGS Fact Sheet 040-02.

Van Liew, M., T. Veith, D. Bosch, and J. Arnold. 2007. Suitability of SWAT for the Conservation Effects Assessment Project: Comparison on USDA agricultural research service watersheds. *Journal of Hydrologic Engineering.* 12(2):173-189.

Vasquez, R., L. Feyan, J. Feyen, and J. Refsgaard. 2002. Effect of grid size on effective parameters and model performance of the MIKE-SHE Code. *Hydrological Processes.* 16(2):355-372.

Vaze, J. and F. Chiew 2003. Study of pollutant washoff from small impervious experimental plots. *Water Resources Research.* 39(6): 1160-1169.

Vega, M. R. Pardo, E. Barrado, and L. Deban, 1998. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water Research.* 32(12):3581-3592.

Venugopal, T., L. Giridharan, and M. Jayaprakash, 2009.Application of chemometric analysis for identifying pollution sources: A case Study on the River Adyar, India. *Marine and Freshwater Research.* 60:1254-1264.

Vieux, B. 2004. *Distributed Hydrologic Modeling Using GIS.* Second Edition. Kluwer Academic Publishers, Boston: 182-219.

Vieux, B., and N. Gauer. 1994. Finite element modeling of storm water runoff using GRASS GIS . *Microcomputersin Civil Engineering.* 9(4):263-270.

Vieux, B. and P. Bedient. 1998. Estimation of rainfall for flood prediction from WSR-88D reflectivity: A case study, 17-18 October 1994. *Weather and Forecasting.* 13(2): 407-415.

Vieux, B. and P.Bedient. 2004. Assessing urban hydrologic prediction accuracy through event reconstruction. *Journal of Hydrology.* 229(3-4): 217-236.

Vieux, B., J. Park, and B. Kang. 2009. Distributed hydrologic prediction: Sensitivity to accuracy of initial soil moisture conditions and rainfall input. *Journal of Hydrologic Engineering.* 14(7):671-389.

Vieux, B. and J. Vieux. 2006. Evaluation of a physics-based Distributed hydrologic model for coastal island and unland hydrologic modeling. Coastal hydrology and processes: proceedings of the AIH 25th Anniversary Meeting & International Conference "Challenges in Coastal Hydrology and Water Quality. Edited by V. Singh. Water Resources Publications. 453-464.

Vieux, B., Z. Cui, and A. Gaur. 2004. Evaluation of a physics-based distributed hydrologic model of flood forecasting. *Journal of Hydrology.* 298(1-4):155-177. Vieux Inc. 2011. *Vflo*^{TM} 5.0. Norman, Oklahoma.

Vieux Inc. 2011. *Vflo*^{TM} 5.0. Norman, Oklahoma.

Ward, G. R. Srinivasan, H. Rifai, J. Mott, L. Hauck, G. Di Giovanni, K. Wagner, and A. Allan. 2009. *Bacteria Total Maximum Daily Load Task Force Final Report.* Texas Water Resources Institute. TR-341 2009.

Wu, J., R. Zou, and S. Yu. 2006. Uncertainty analysis for coupled watershed and water quality modeling systems. *Journal of Water Resources Planning and Management.* 132(5):351-361.

Wu, M. and Y. Wang, 2007. Using chemometrics to evaluate anthropogenic effects in Daya Bay, China. *Estuarine, Coastal and Shelf Science.* 72(4):732-742.

Wu, M., Y. Wang, C. Sun, H. Wang, J. Dong, and S. Han, 2009. Identification of anthropogenic and seasonality in Daya Bay, South China Sea. *Journal of Environmental Management.* 90(10):3082-3090.

Wu, T. D. Gilbert, H. Fuelberg, H. Cooper, D. Bottcher, and C. Reed. (2010) Doppler-derived rainfall data monitoring to support surface water modeling of TMDL. *Journal of Coastal Research.* 52 (Special Issue): 273-280.

Yagow, G., T., Dilaha, S. Mostaghimi, K. Brannan, C. Heatwole, and M. Wolfe. 2001. TMDL modeling of fecal coliform bacteria with HSPF. ASAE Meeting Paper No. 01-2066 St. Joseph, Mich.: ASAE

Yilmaz, K., H. Gupta, and T. Wagener. (2008). A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. *Water Resources Research.* 44(W09417).

Zimmer, A. 2007. The use of a distributed hydrologic model to evaluate the location of development and flood control reservoirs. Masters of Science Thesis, Rice University, accessed at http://gradworks.umi.com/14/55/1455302.html.

Zhou, F., Y. Liu, and H. Guo, 2007. Applications of multivariate statistical methods to water quality assessment of the water courses in Northwestern New Territories, Hong Kong. *Environmental Monitoring and Assessment.* 132:1-13.

Appendix A .   Water Quality Statistics

Table A- 1.Water Quality Statistics for Cypress Creek at IH-45 (Gauge 11328)

Overall	Streamflow	TDS (mg/L)	TSS (mg/L)	CHLORIDE (mg/L)	N03-N (mg/L)	PHOS-T (mg/L)	E COLI (MPN/dL)
Count	229	183	137	108	66	84	93
Conc. Median	54	29	438	65	4.195	1.458	580
Conc. StdDEV	560.92	74.12	133.01	35.63	4.58	0.84	8,268.61
Load Median (mg/s)		44.77	777.22	136.50	10.76	3.25	15,943.62
Load StdDEV (mg/s)		3,555.87	3,316.40	242.79	16.46	11.61	7,622,145.22

Low Flows (21-28 cfs)	Flow (cfs)	TSS	TDS	Chloride	N03-N	PHOS-T	E COLI
Count	33	28	27	7	3	3	3
Conc. Median	25	26.5	535	109	8.02	2.317	4100
Conc. StdDEV	2.19	9.67	46.80	8.37	1.59	0.20	2,036.30
Load Median (mg/s)		22.08	444.82	95.87	7.38	2.27	33,665.10
Load StdDEV (mg/s)		9.51	52.24	16.16	2.42	0.28	17,612.46

Dry Flows (29-47 cfs)	Flow (cfs)	TSS	TDS	Chloride	N03-N	PHOS-T	E COLI
Count	67	55	40	35	13	24	24
Conc. Median	36	23	481.5	94	6.89	2.061	215
Conc. StdDEV	5.38	16.33	77.78	18.04	2.07	0.65	597.68
Load Median (mg/s)		30.63	604.58	116.95	8.75	2.92	2,570.40
Load StdDEV (mg/s)		23.62	94.84	25.43	2.75	0.72	9,816.17

Mid Range's Flow (48-82 cfs)	Flow (cfs)	TSS	TDS	Chloride	N03-N	PHOS-T	E COLI
Count	49	39	22	28	22	25	27
Conc. Median	61	20	409	59	5.28	1.45	430
Conc. StdDEV	10.75	20.49	79.51	27.34	6.25	0.66	1,647.80
Load Median (mg/s)		41.13	896.50	135.86	11.03	2.80	8,289.54
Load StdDEV (mg/s)		51.78	156.92	53.13	11.88	1.32	31,730.00

Wet flows (83-662 cfs)	Flow (cfs)	TSS	TDS	Chloride	N03-N	PHOS-T	E COLI
Count	59	47	35	28	20	22	26
Conc. Median	183	48	291	34.72	1.37	0.69	1345
Conc. StdDEV	161.58	96.55	103.70	26.38	2.44	0.58	5,566.48
Load Median (mg/s)		354.72	1,857.47	208.54	12.93	5.67	147,007.60
Load StdDEV (mg/s)		1,123.43	892.12	240.51	8.45	7.27	523,905.60

High Flows (663-4350 cfs)	Flow (cfs)	TSS	TDS	Chloride	N03-N	PHOS-T	E COLI
Count	21	14	13	10	8	10	13
Conc. Median	1380	111	163	8.5	0.36	0.43	1900
Conc. StdDEV	1001.74	124.76	38.47	10.25	0.85	0.24	18,743.62
Load Median (mg/s)		4,127.28	8,204.93	411.26	15.58	19.14	698,649.00
Load StdDEV (mg/s)		9,355.39	6,351.44	509.22	38.98	23.18	18,649,393.51

Table A- 2.Water Quality Statistics for Cypress Creek at Steubner Airline Road (Gauge 11330)

Overall							
Month	Streamflow (cfs)	TDS (mg/L)	TSS (mg/L)	CHLORIDE (mg/L)	N03-N (mg/L)	PHOS-T (mg/L)	E COLI (MPN/dL)
Count	42	27	29	29	15	15	32
Conc. Median	71.00	391.00	17.00	83.00	4.32	1.65	1055.00
Conc. StdDEV	859.22	156.99	52.51	34.38	3.37	0.84	18211.21
Load Median (mg/s)		1,600	30.64	138.86	13.83	4.73	36,253
Load StdDEV (mg/s)		5,588	2656.53	130.39	8.66	8.00	7,865,247
Low Flows (<23 cfs)							
Month	Streamflow (cfs)	TDS	TSS	CHLORIDE	N03-N	PHOS-T	E COLI
Count	1	1	1	1	0	0	1
Median (mg/L)	23.00	499.00	17.00	102.00 -		-	100.00
StdDEV (mg/L)	-	-	-	-	-	-	-
Median (mg/L)		409.89	13.96	83.79 -		-	821.10
StdDEV (mg/L)		-	-	-	-	-	-
Dry Flows(23-50cfs)							
Month	Streamflow (cfs)	TDS	TSS	CHLORIDE	N03-N	PHOS-T	E COLI
Count	14	5	13	13	3	3	7
Median (mg/L)	33.50	477.00	12.00	93.00	7.11	2.65	320.00
StdDEV (mg/L)	4.80	64.05	5.77	9.14	1.39	0.38	201.20
Median (mg/L)		619	17.14	122.64	8.54	3.35	3,770
StdDEV (mg/L)		96	6.03	14.71	3.72	0.08	2,191
Mid Range's Flow (50-87 cfs)							
Month	Streamflow (cfs)	TDS	TSS	CHLORIDE	N03-N	PHOS-T	E COLI
Count	8	7	5	5	4	4	7
Median (mg/L)	64.38	440.43	24.00	70.02	6.94	1.86	954.29
StdDEV (mg/L)	11.87	74.50	18.15	22.65	2.15	0.29	717.49
Median (mg/L)		1,024	32.00	189.00	15.33	4.07	15,808
StdDEV (mg/L)		303	45.79	58.28	7.84	1.16	13,448
Wet flows (86-751 cfs)							
Month	Streamflow (cfs)	TDS	TSS	CHLORIDE	N03-N	PHOS-T	E COLI
Count	10	7	7	7	5	5	9
Median (mg/L)	213.00	283.00	47.00	40.00	1.92	0.73	1300.00
StdDEV (mg/L)	164.52	117.62	27.19	16.69	1.47	0.49	4162.90
Median (mg/L)		2,025	262.29	183.93	13.64	5.98	129,484
StdDEV (mg/L)		887	575.02	104.59	3.62	4.65	363,719
High Flows (752-14600 cfs)							
Month	Streamflow (cfs)	TDS	TSS	CHLORIDE	N03-N	PHOS-T	E COLI
Count	9	7	3	3	3	3	8
Median (mg/L)	1410.00	142.00	106.00	9.00	0.39	0.38	3850.00
StdDEV (mg/L)	1108.39	17.19	101.53	3.61	0.31	0.19	33406.52
Median (mg/L)		7,906	5337.86	352.50	14.35	15.10	2,731,050
StdDEV (mg/L)		8,026	5368.42	201.33	16.28	10.40	14,022,594

Table A- 3.Water Quality Statistics for Cypress Creek at Grant Road (Gauge 11332)

Overall						
	Flow (cfs)	TSS (mg/L)	TDS (mg/L)	Cl (mg/L)	N03-N (mg/L)	PHOS-T (mg/L)	E COLI (MPN/dL)
Count	77	77	58	77	48	59	76
Conc Median	18.00	21.00	299.50	44.00	1.63	0.81	275.00
Conc Std Dev	336.99	47.79	118.93	30.40	2.26	0.65	15205.85
Load Median		10.00	263.36	30.36	2.62	0.85	1,865
Load Std Dev		1170.67	2082.42	103.47	4.08	4.33	1,066,866

Low Flows (<6.8 cfs)						
	Flow (cfs)	TSS (mg/L)	TDS (mg/L)	Cl (mg/L)	N03-N (mg/L)	PHOS-T (mg/L)	E COLI (MPN/dL)
Count	14	14	7	14	6	6	14
Conc Median	6.20	12.50	453.00	83.00	5.29	2.00	110.00
Conc Std Dev	0.61	6.08	74.42	18.40	1.01	0.59	189.29
Load Median		2.62	101.46	17.96	1.08	0.40	254.78
Load Std Dev		1.37	17.36	4.21	0.27	0.15	451.40

Dry Flows (6.8 -17 cfs)						
	Flow (cfs)	TSS (mg/L)	TDS (mg/L)	Cl (mg/L)	N03-N (mg/L)	PHOS-T (mg/L)	E COLI (MPN/dL)
Count	20	20	16	20	12	15	20
Conc Median	10.50	16.00	393.50	63.00	4.01	1.30	165.00
Conc Std Dev	2.82	9.76	69.81	15.68	1.64	0.45	112.49
Load Median		6.46	150.66	22.41	1.82	0.56	547.07
Load Std Dev		5.36	51.36	7.53	0.86	0.12	557.71

Mid Range Flows (17-35 cfs)						
	Flow (cfs)	TSS (mg/L)	TDS (mg/L)	Cl (mg/L)	N03-N (mg/L)	PHOS-T (mg/L)	E COLI (MPN/dL)
Count	14	14	12	14	7	12	13
Conc Median	21.00	12.00	286.50	39.60	2.89	0.91	430.00
Conc Std Dev	6.60	13.28	98.54	21.83	2.38	0.52	2083.22
Load Median		9.50	244.28	30.04	2.33	0.77	2,610.69
Load Std Dev		11.46	52.40	12.05	1.53	0.26	22,318.91

Wet Flows (35-500 cfs)						
	Flow (cfs)	TSS (mg/L)	TDS (mg/L)	Cl (mg/L)	N03-N (mg/L)	PHOS-T (mg/L)	E COLI (MPN/dL)
Count	22	22	17	22	17	19	22
Conc Median	133.00	63.00	239.00	17.05	0.48	0.50	1065.00
Conc Std Dev	147.61	64.82	45.12	9.59	0.64	0.28	27570.66
Load Median		366.16	1,510.81	84.21	3.04	2.67	71,356.57
Load Std Dev		933.55	1,264.70	131.06	5.77	3.58	1,554,647.52

High Flows (500-4350 cfs)						
	Flow (cfs)	TSS (mg/L)	TDS (mg/L)	Cl (mg/L)	N03-N (mg/L)	PHOS-T (mg/L)	E COLI (MPN/dL)
Count	7	7	6	7	6	7	7
Conc Median	1110.00	85.00	146.00	7.00	0.20	0.27	3873.00
Conc Std Dev	346.76	23.87	25.99	0.76	0.05	0.05	2685.85
Load Median		3,006.40	6,920.84	237.86	6.49	10.70	1,274,454.09
Load Std Dev		1,619.45	2,256.66	77.59	2.52	4.28	1,611,094.58

Table A- 4.Water Quality Statistics for Cypress Creek at House-Hahl Road (Gauge 11333)

Overall	Flow (cfs)	TSS (mg/L)	TDS (mg/L)	Cl (mg/L)	N03-N (mg/L)	PHOS-T (mg/L)	E COLI (MPN/dL)
Count	136	111	108	40	21	21	55
Conc Median	5.80	15.00	326.00	48.50	1.58	0.69	200.00
Conc StdDev	130.37	29.35	177.47	88.05	2.65	0.90	7054.14
Load Median		1.83	63.04	8.46	0.42	0.15	511
Load StdDev		195.54	638.49	39.67	1.37	1.76	1,094,289
Low Flows (<0.78 cfs)	Flow (cfs)	TSS (mg/L)	TDS (mg/L)	Cl (mg/L)	N03-N (mg/L)	PHOS-T (mg/L)	E COLI (MPN/dL)
Count	19	16	15	2	1	1	2
Conc Median	0.19	15.00	509.00	92.50	9.31	3.45	215.00
Conc StdDev	0.28	8.63	120.00	17.68	-	-	35.36
Load Median		0.17	2.94	2.12	0.16	0.06	47
Load StdDev		0.27	4.76	1.02	-	-	7
Dry Flows (0.78-3.8 cfs)	Flow (cfs)	TSS (mg/L)	TDS (mg/L)	Cl (mg/L)	N03-N (mg/L)	PHOS-T (mg/L)	E COLI (MPN/dL)
Count	39	34	33	16	7	7	19
Conc Median	1.70	11.50	432.00	85.50	3.65	1.18	160.00
Conc StdDev	0.77	37.12	193.09	122.46	2.23	0.77	309.74
Load Median		0.90	25.14	5.25	0.22	0.11	91
Load StdDev		1.33	14.16	5.90	0.22	0.05	212
Mid Range Flow (3.8-11 cfs)	Flow (cfs)	TSS (mg/L)	TDS (mg/L)	Cl (mg/L)	N03-N (mg/L)	PHOS-T (mg/L)	E COLI (MPN/dL)
Count	25	20	21	8	5	5	11
Conc Median	6.00	8.00	339.00	37.50	1.58	0.45	110.00
Conc StdDev	2.27	9.65	115.50	9.19	0.36	0.27	114.60
Load Median		1.54	85.88	7.91	0.32	0.09	350
Load StdDev		3.39	40.17	1.60	0.06	0.05	341
Wet Flows (11-199 cfs)	Flow (cfs)	TSS (mg/L)	TDS (mg/L)	Cl (mg/L)	N03-N (mg/L)	PHOS-T (mg/L)	E COLI (MPN/dL)
Count	39	34	29	12	6	6	15
Conc Median	28.00	26.00	232.00	21.15	0.31	0.27	540.00
Conc StdDev	34.44	32.37	113.62	20.35	0.21	0.38	6599.34
Load Median		24.81	236.80	30.90	1.02	1.21	10,889
Load StdDev		120.09	312.95	45.90	0.79	0.81	239,931
High Flows (199-7640 cfs)	Flow (cfs)	TSS (mg/L)	TDS (mg/L)	Cl (mg/L)	N03-N (mg/L)	PHOS-T (mg/L)	E COLI (MPN/dL)
Count	14	7	10	2	2	2	8
Conc Median	361.00	44.00	126.00	6.50	0.20	0.28	930.00
Conc StdDev	172.18	13.81	36.09	2.12	0.00	0.08	15997.25
Load Median		627.61	2,023.55	136.68	4.49	5.84	133,982
Load StdDev		266.90	820.36	9.92	1.77	0.69	2,824,054

Appendix B . Analysis of Water Quality Trends and Pollutant Loading for Cypress Creek Watershed

B.1 Introduction

Lake Houston is the primary source of drinking water for the City of Houston. Over 146 billion gallons are treated annually for 2 million customers (City of Houston, 2008). The lake is impaired for bacteria (TCEQ, 2008) and has concerns for nutrients and Chlorophyll-A (TCEQ, 2008). The watersheds flowing into the lake are also impaired for bacteria and have concerns for nutrients and depressed dissolved oxygen. With rising water treatment costs, the degraded influent has become a key concern for the City of Houston and thus protecting the watersheds is a chief priority. In order to efficiently manage the water quality of Lake Houston, an understanding of the watersheds and the relationship between pollutant loads and the influent flows is required.

Statistical analysis was performed to characterize the water quality of Cypress Creek. This was accomplished by (A) comparing the median concentration and loading of storm and low flow conditions, (B) analyzing the temporal trends in both concentration and loading, (C) establishing the low flow stream profile to assess the influence of point sources in the downstream section of the watershed, and (D) comparing the storm flow related pollutant concentration to ascertain the pollutant relationships during rainfall events.

B.2 Background

Long-term water quality data is commonly used to detect trends in pollutants over time and as well as to identify, describe, and explain major factors that affect trends in water quality (Yu, 1993). The Mann-Kendall test, also known as Kendall's tau statistic is a

nonparametric test for monotonic trends in a time series (Kaha, 2004). The test is derived from a rank correlation test for two groups of observations (Haan, 2002). It is tolerant of outliers but requires that data are serially independent (Hamed, 1988). The Mann Kendall test is robust towards missing values, seasonal effect and non-normality (Larsen et al, 1999). Previous research has applied Mann-Kendall test to identify longer term effects of urbanization (Boeder, 2008) and to link stream flow with long term meteorological changes (Burn, 2002).

The null hypothesis assumed for the test states that the data ($x_1, \ldots x_n$) are a sample of n independent and identically distributed random variables, in other words the data does not have trend or serial correlation. The alternative hypothesis, H_a of a two sided test, states that the distribution of x_k and x_j are not identical for all k, j less than n with k not equal to j. Each value in the series $X(t+1)$ is compared to $X(t)$ and assigned a score $z(k)$ calculated as

$$z(k) = \begin{cases} 1 \ldots if X(t) > X(t+1) \\ 0 \ldots if X(t) = X(t+1) \\ -1 \ldots if X(t) < X(t+1) \end{cases}$$

The $z(k)$, score is then used to calculate the Mann-Kendall statistics by

$$S = \sum_{k=1}^{N(N-1)/2} z(k)$$

where N is the total number of observations. The test statistic for $N \geq 10$, where the samples include at least 10 years of data is

$$u_c = \frac{S+m}{\sqrt{V(S)}}$$

where $m=1$ if $S<0$ and $m=-1$ if $S>0$, and the variance, $V(S)$ is calculated by formula

$$V(S) = \frac{[N(N-1)(2N+5)]}{18}$$

The hypothesis of no trend is rejected if $|u_c|>z_{1-\alpha/2}$ where $\alpha=0.05$

The Seasonal Kendall test is applied to address situations where data displays seasonality (Hirsch, 1982). By dividing the data into seasonal groups, performing the Mann-Kendall test on each seasonal group, and then summing the results, the effect of seasonality is removed (Hirsch, 1984). In effect, the result of inter-season dependence is eliminated. This method has been used in various applications including the assessment of how seasonality influences the detection acid rain impacts (Taylor, 1989), evaluation of spatial-temporal variability of water quality (Krusche, 1997), and review of the effectiveness of water quality management strategies (Cude, 2001).

The null hypothesis of the seasonal Kendall test states that for each of p seasons the n observations (years) are randomly ordered. The alternative hypothesis states that a monotonic trend exists in one or more seasons (Hirsch, 1984). The test statistic for each season thus becomes

$$S_g = \sum_{i<j} \text{sgn}\left(X_{jg} - X_{ig}\right) \qquad g = 1,2,...,p .$$

This makes the Seasonal Kendall test statistic

$$S' = \sum_{g=1}^{p} S .$$

It is assumed that the data are independent and thus covariance terms equal zero.

B.3 Data and Methods

Water quality data have been collected intermittently within Cypress Creek since 1980.
There are three monitoring stations within the watershed (see Figure B-1) operated by the
City of Houston, Water Quality Control and Health and Human Services. Historical
water quality and streamflow data were analyzed using basic statistics and trend analysis
in order to assess the general relationship of pollutant dynamics and streamflow.
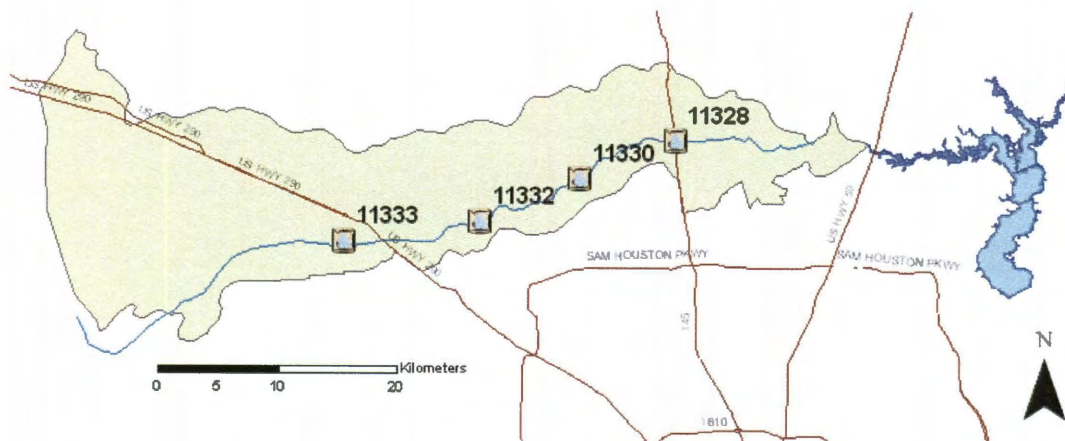

Figure B-1.Location of water quality stations in Cypress Creek Watershed

B.3.1 Basic Statistical Analysis

A flow duration analysis was performed using the streamflow at each station. Historical
streamflow data were ranked in descending order and the percent exceedance calculated
(rank/total number of points). Streamflow observations were divided into different
categories: 0 to 10% exceedance , High Flows; 10-40% exceedance , Moist Conditions;
40-60%, Mid-Range Flows; 60-90% exceedance, Dry Conditions; and 90-100%

exceedance, Low Flows (Morrison and Bonta, 2008; USEPA, 2007). Low flows and dry conditions were considered to be associated with dry weather. In contrast, moist conditions and high flows were associated with rainfall events. Mid range flows were most often associated with rainfall events but can also be during dry periods, if the point sources were active.

The water quality data was segmented according to the flow condition, so that the median value and standard deviation in each streamflow regime could be determined. These basic statistics were determined for both the concentration and loading rate. Loading rate was calculated by

$$Load = Q * C$$

where *Load* is the loading rate (mass/time), Q is the streamflow (volume/time), and C is the concentration (mass/volume). The loading rates for each station were then compared in order to determine the relative influence of rainfall-associated loading compared to dry weather loading.

B.3.2 Trend Analysis

Further statistical analysis of the water quality dataset was performed to determine if temporal trends exist within the dataset. By assessing if trends exist within the concentration and loading of the sampled constituents, conclusions can be drawn about urbanization and associated increase of pollutant sources within watershed.

Water quality data was collected by the City of Houston Health and Human Services and Water Quality Control departments at three locations within Cypress Creek ranging from 1980 to 2009. Streamflow discharge was measured by the USGS at the same locations of water quality sampling. The USGS trend testing utility (Helsel, 2006) was used to test for trend in stream flow and water quality parameters including streamflow, chloride, nitrate, total phosphorus, *E. coli*, and total dissolved solids using Mann-Kendall test at a 95% confidence level. Then the same utility was used to test for seasonal trends using a two season division of data using the Seasonal Kendall test also at 95% confidence. The presence of a trend was determined and the Kendall's tau statistics computed in order to assess the magnitude and direction of the trend.

B.3.3. Low Flow and Stormwater Quality Sampling

Low flow water quality samples were collected as in-stream grab samples along the path shown in Figure B-2. Samples were taken before and after each permitted outfall on the sample path in order to establish a baseline for the constituent concentrations at low flow as well as investigate the influence of the outfalls in the downstream of Cypress Creek. The concentrations of *E. coli*, nitrate, total phosphorus, and total suspended solids (TSS) were measured by the City of Houston Water Quality Laboratory, using the techniques in table B-1. Total dissolved solids (TDS) were measured on site using a Hach hand probe.
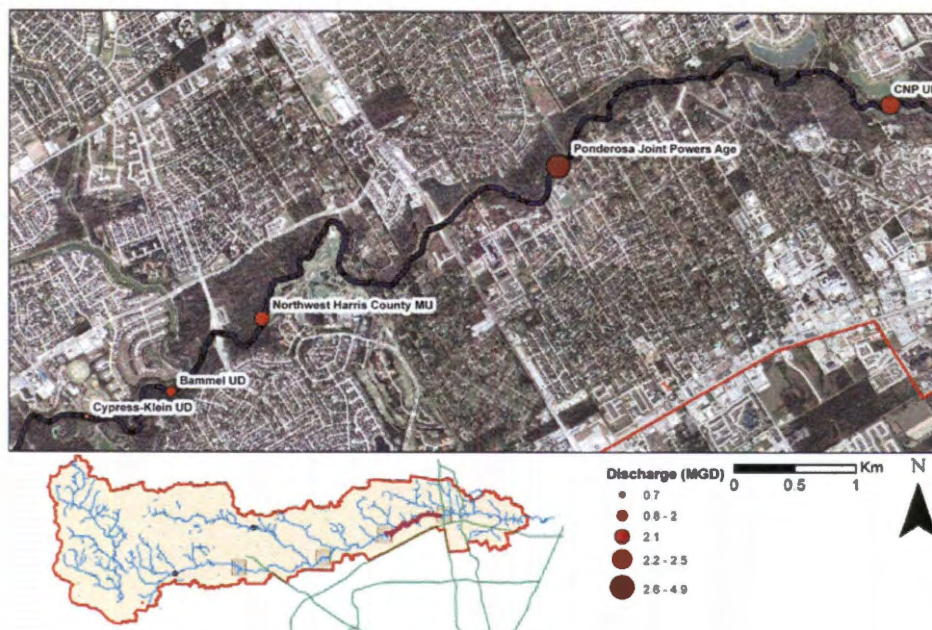
Figure B-2. Sampling path for low flow in-stream sampling

Table B - 1.Techniques to measure constituent concentrations

Sample	Storetcode	Method
E. coli	31616	SM 9223-B, IDEXX Colilert
NO3-N	620	EPA 300.0, Rev 2.0
Total Phosphorus	665	EPA 365.3
TSS	530	EPA 160.2

The percent difference from the water quality standard was calculated using the applicable Texas Commission on Environmental Quality (TCEQ) water quality standard or screening level (Table B-2). These values were then plotted versus the distance along the stream relative to the sample starting point.

Table B - 2. Water quality standards

Constituent	Standard
Total Suspended Solids (TSS)	13 mg/L
Total Dissolved Solids (TDS)	600 mg/L
Nitrate	1.95 mg/L
Total Phosphorus	0.69 mg/L
E. coli	200 mg/L

a) (TCEQ,2008), b)(H-GAC, 2010), c) (TCEQ, 2010)

Stormwater samples were collected at the downstream streamflow gage (point C in

Figure B-1) and the constituent concentrations measured as described for the low-flow

samples. Grab samples were taken from the stream during the storm event on both the

rising and falling limb of the hydrograph. Due to the resource intensive nature of

sampling throughout a storm event, NEXRAD rainfall was used to evaluate the suitability

of the rainfall event for modeling and sampling by determining the intensity, total depth,

and location of rainfall. Previous modeling with $Vflo^{TM}$, using design storms and actual

rainfall events, was used to estimate the travel time of the runoff from origin to the

downstream sampling location, which was then used to direct travel of sampling teams to

the watershed as well as the duration of sampling. Rainfall data and previous hydrologic

modelling was also used to guide the necessary time and interval of sampling. This was

done in order to balance density of data collected with the available time and resources.

The water quality observations for the different constituents were then plotted against

each other to assess the relationships between the parameters. Linear regression was

performed and the coefficient of determination (R^2) value calculated. Each of the

relationships was examined to determine the general agreement between the constituents

and thus characterize the similarity of pollutant source types. If there was a linear

relationship between the constituents, as determined by regression analysis, then the sources were considered to be similar.

B.4 Results

B.4.1 Basic Statistics

The median and standard deviation of water quality parameters at each station are reported in Appendix A. Streamflow from each gauging and water quality station was analyzed via flow duration analysis to determine the breakdown of the flow regimes to determine the low, dry, mid-range, wet, and high flows. With these general conditions, the median concentration and loading rate for rainfall and dry weather related streamflows. A comparison of the loading rates is illustrated by Figure B-3.
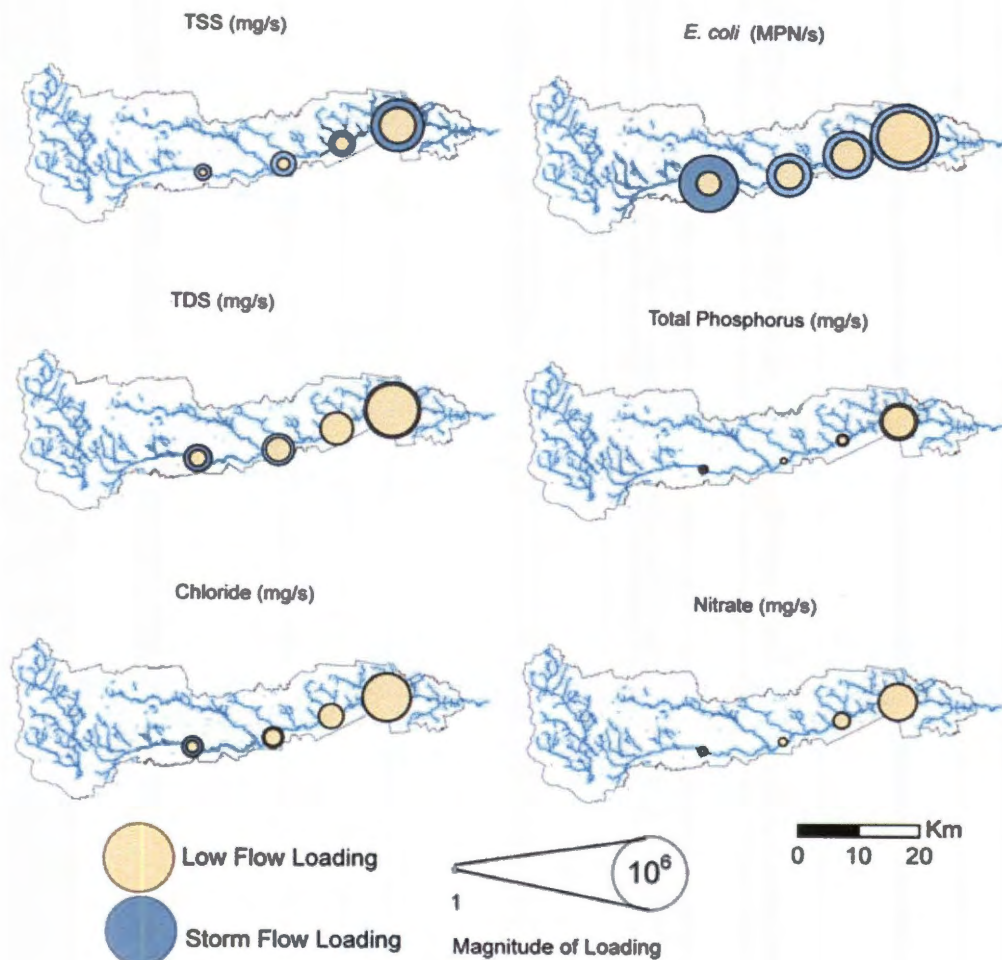
Figure B-3. Comparison of low flow and storm flow loading rates

Point sources are associated with low flow loading. In contrast, both point sources and the diffuse, non-point sources are associated with storm flow loading. By comparing the median loading rates of low flow and storm flow loading, the relative influence of the different sources can be assessed. Figure B-3 illustrates that TSS and *E .coli* have large differences between the low flow and storm flow loading rates. For TDS, total phosphorus, nitrate, and chloride, the difference in the median loading rates is approximately a single order of magnitude or less. For most constituents, the loading

increases from upstream to downstream of Cypress Creek due to the cumulative effects of runoff through the watershed as well as the large number of point sources in the down stream portion of the watershed. However, the storm flow loading of *E. coli* does not follow this spatial pattern. Instead, the storm loading rate at the most upstream gage is greater than the two middle gages. This suggests that the non-point sources from the upstream agricultural area heavily influence the stream at this gauge. For each of the other stations, the dry loading is similar in magnitude to the storm flow loading, suggesting that point sources heavily influence the total loading at these gages.

Overall, the comparison of storm and low flow loading shows the importance of rainfall-runoff modeling to address non-point sources of TSS and *E. coli* as part of a comprehensive water resource management plan. On the other hand, plans to address nutrient loading to Cypress Creek, should include further investigation of point sources.

B.4.2 Trend Analysis

The results of trend testing using the Mann-Kendall test are summarized in Table B-3, and the trend testing using the Seasonal Kendall test are in Table B-4. The tests were assessed at a 95% confidence level, and the p-values, and Kendall's tau reported.
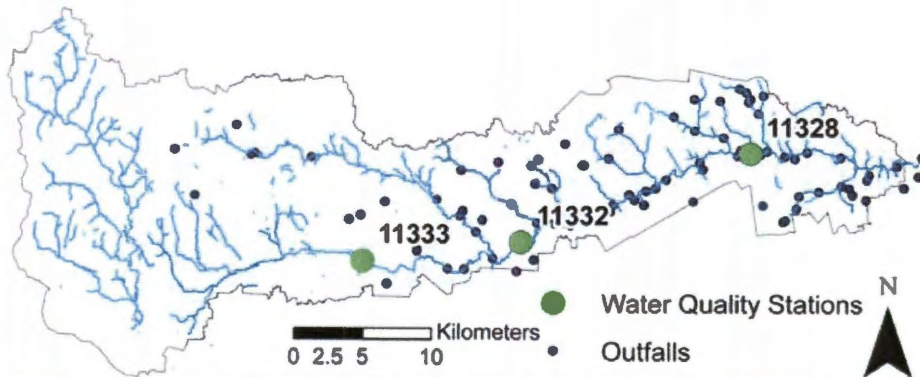
Table B - 3. Mann Kendall trend analysis

Station	Constituent	Concentration				Loading			
		# Years of Data	Trend at 0.05	p-value	Kendall's tau	# Years of Data	Trend at 0.05	p-value	Kendall's tau
113328	Q	29	Yes	0.001	0.137	10	Yes	0.0053	0.159
	Cl	29	No			29	Yes	<0.0000	0.413
	EC	9	Yes	0.0197	0.141	9	Yes	0.0421	0.123
	NO3	10	No			10	Yes	<0.0000	0.412
	TP	26	Yes	<0.0000	-0.39	26	No		
	TDS	14	Yes	0.0254	-0.118	14	Yes	<0.0000	0.224
11332	Q	10	No						
	Cl	9	Yes	0.0049	0.154	9	Yes	<0.0000	0.3
	EC	9	No			9	No		
	TP	20	No			Insufficient Data			
	TDS	15	No			9	No		
11333	Q	10	No						
	TDS	9	No			9	No		

Table B - 4. Seasonal Kendall trend analysis

Station	Constituent	Concentration				Loading			
		# Years of Data	Trend at 0.05	p-value	Kendall's tau	# Years of Data	Trend at 0.05	p-value	Kendall's tau
113328	Q	29	Yes	0.0399	0.211	10	No		
	Cl	29	No			29	Yes	0.0002	0.436
	EC	9	No			9	No		
	NO3	10	No			10	Yes	0.0369	0.378
	TP	26	Yes	0.0018	-0.393	26	No		
	TDS	14	No			14	No		
11332	Q	10	No			Insufficient Data			
	Cl	19	Yes	0.0394	0.286	Insufficient Data			
	EC	Insufficient Data				Insufficient Data			
	TP	18	No			Insufficient Data			
	TDS	9	No			Insufficient Data			
11333	Q	10	No						
	TDS	9	No			9	No		

There were insufficient data to perform the trend analysis for all the constituents at each station. For the most upstream station (11333), where there has been little land use change and there were few permitted discharges (See Figure B-4), there was no trend detected in the streamflow or total suspended solids (Table B-4 and B-5). In the center of the watershed, the only trend that was identified was an increasing trend in both chloride concentration and loading. The most downstream station had increasing trends in streamflow over the last 10 and 20 year periods. While there was no trend in the concentration of chloride and nitrate, there was an increasing trend in the loading of these constituents. On the other hand, a decreasing trend in the concentration of total

phosphorus was identified. Both *E. coli* and TDS showed increasing trends for both
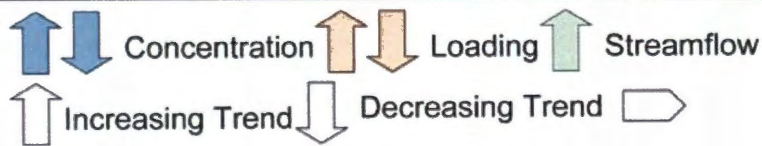
concentration and loading.

Figure B-4 .Trend analysis of water quality data

B.4.3 Storm and Low Flow Sampling

Additional low flow sampling along the lower 10 km of Cypress Creek was performed in order to establish a baseline concentration of the sampled constituents as well as to observe the impact of the different permitted discharges in this region. The percent difference of the measured concentration from the water quality standard for each constituent was plotted against the sampling path distance in Figure B-5 . The vertical lines denote the locations of the permitted discharges, or in other words the point sources in the lower portion of the watershed.
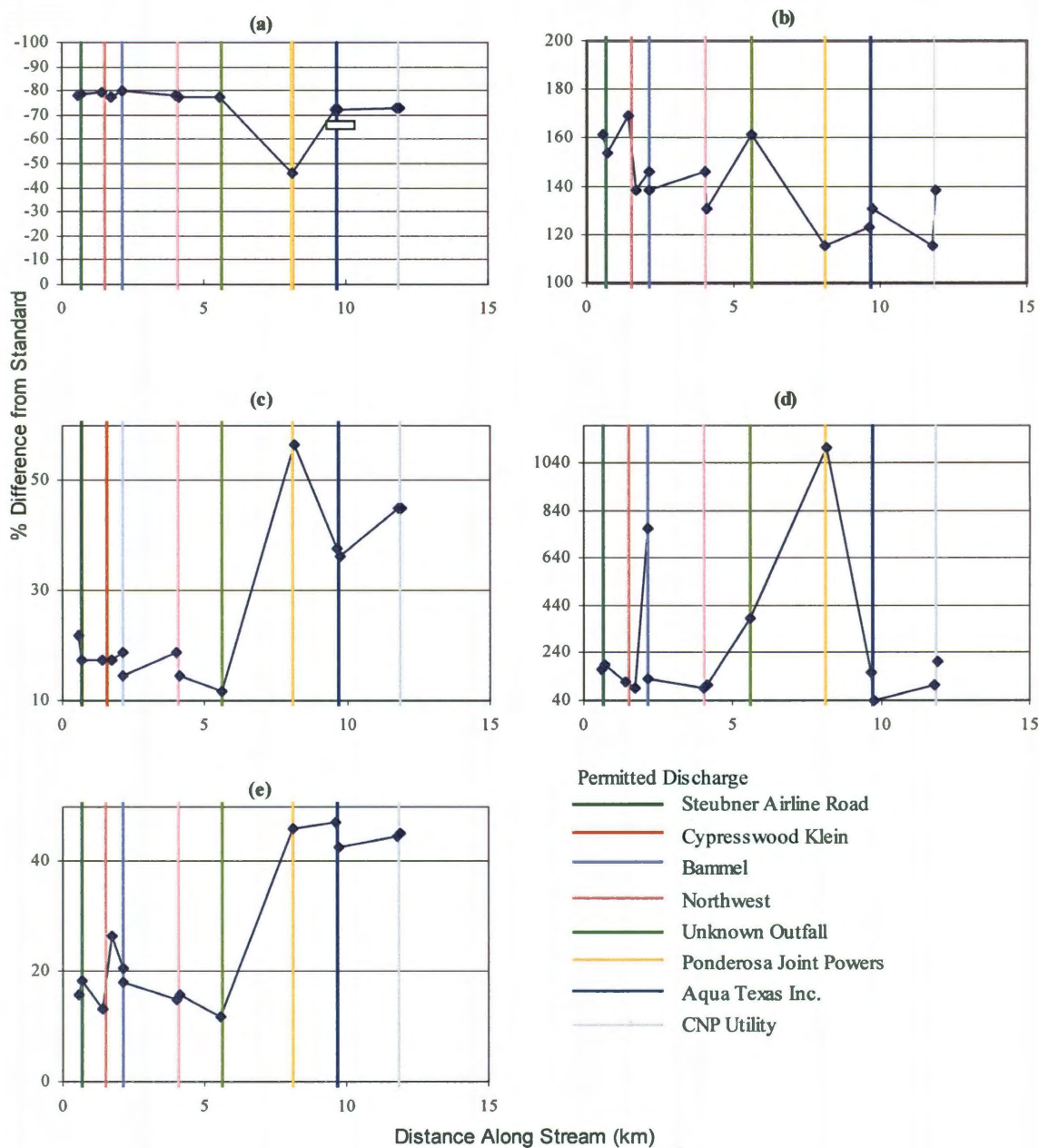
Figure B-5. Low flow water quality profiles of (a) TDS, (b) TSS, (c) nitrate, (d) E. coli and (e) total phosphate

As illustrated in Figure B-5, the only constituent for which the stream was in compliance

with the standard is TDS (a). In contrast, the TSS concentration (b) exceeded the water

quality standard for Cypress Creek by over 100%. There were only screening standards

available to assess nitrate and total phosphorus (c and e). The stream was generally 20%

higher than the screening level. The largest difference from the water quality standard

was the *E. coli* concentration (d). The general pattern for each constituent was that the

largest change in constituent concentration is observed at the Ponderosa Joint Powers

discharge point. This would suggest that additional investigation and observation needs

to be performed to assess the impact of the point sources in both low flow and storm flow

conditions. Furthermore, future sampling at low flow conditions, should be performed at

a greater spatial density to further investigate the impact of point source loading into the

stream.

Sampling during storm flow conditions was performed at the downstream gauging station

during four rainfall events. The goal was to observe the constituent concentrations on

both the rising and falling limbs of the hydrograph. The relationships between the

different constituents are shown by the regression analysis of the stormwater observations

in Figure B-6. It can be seen that there is a linear relationship amongst TDS, total

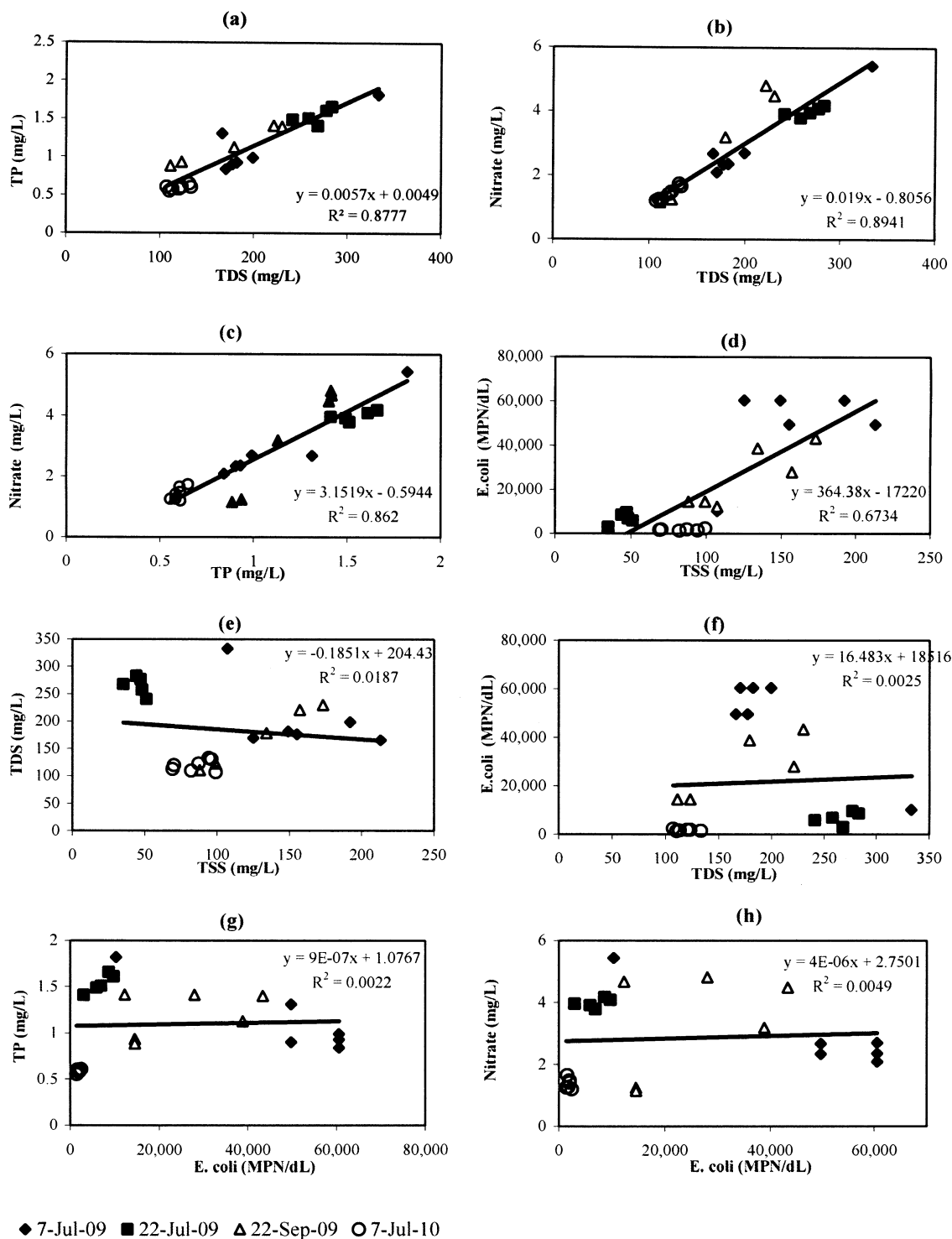phosphorus, and nitrate, and between TSS and *E. coli*.

(a) $y = 0.0057x + 0.0049$ $R^2 = 0.8777$

(b) $y = 0.019x - 0.8056$ $R^2 = 0.8941$

(c) $y = 3.1519x - 0.5944$ $R^2 = 0.862$

(d) $y = 364.38x - 17220$ $R^2 = 0.6734$

(e) $y = -0.1851x + 204.43$ $R^2 = 0.0187$

(f) $y = 16.483x + 18516$ $R^2 = 0.0025$

(g) $y = 9E\text{-}07x + 1.0767$ $R^2 = 0.0022$

(h) $y = 4E\text{-}06x + 2.7501$ $R^2 = 0.0049$

◆ 7-Jul-09   ■ 22-Jul-09   △ 22-Sep-09   ○ 7-Jul-10

Figure B-6 . Realtionships between water quality constituents observed during stormwater quality stampling

B.5 Discussion

The present analysis, with its limited dataset was restricted in its utility for

characterization of the watershed and its pollutant sources. Given the severity of water

quality violations and the city of Houston's dependence on the water for domestic

consumption, it is important that more frequent sampling be performed through a wider

range of flow conditions. This will provide an appropriate breadth of data in order to

study the nature of water quality violations and characterize the sources and behavior of

pollutant sources. With further study of the stream, water quality trends will provide

valuable information for appropriate watershed management.

Within the past 10 years, urban development has resulted in an increase in both point

sources and urban runoff, primarily in the downstream portion of Cypress Creek. During

this same time period there was an increasing trend in streamflow for the downstream

gauge and a lack of trend in the upstream station. While the mid-watershed station

lacked a trend for streamflow, the increasing trend in both the concentration and loading

of chloride shows that this region was also influenced by point source discharges,

including waste water treatment plants. Chloride is generally attributed to waste water

treatment plant effluent (Sawyer et al, 2006). While there was no trend in chloride

concentration at the down stream station, there was an increasing trend in chloride

loading. This difference is attributed to the increasing streamflow. This same pattern

was displayed by nitrate. This would suggest a common source for both nitrate and

chloride.

A large number of point source outfalls were located in the downstream portion of the watershed (Figure B-4). The similar behavior of nitrate and chloride suggests that these outfalls are the source of nitrate loading to the stream. In contrast, the decreasing trend in total phosphorus concentration combined with a lack of trend in loading indicates that the source of phosphorus is not associated with point sources but rather non-point sources. Increasing trends in both the concentration and loading of *E. coli* signify complex sources and transport dynamics.

When the low flow water quality observations are examined (Figure B-5), a profile of the stream can be developed. The percent difference from the water quality standard or screening level was used in order to normalize the different magnitudes of the various constituents. For each constituent, a significant change can be observed corresponding to the Ponderosa Joint Powers permitted discharge. The TDS concentration decreased, where as the TSS, nitrate, total phosphorus, *E. coli* concentration increased. The average TSS concentration for low flows as assessed through this sampling and the historical water quality sampling, is approximately 20 to40 mg/L.

The storm flow observations were analyzed to identify the relationships between different water quality constituents. The positive linear relationship between TDS, nitrate, and total phosphorus indicates a similar source as well as transport dynamics. The linear relationship between TSS and *E. coli*, is expected because it is assumed that *E.coli* will be attached to the suspended solids. As such, the concentration of TSS is an acceptable surrogate for storm water modeling of *E. coli*. Likewise the correlation of nitrate and

total phosphorus with TDS indicates that TDS could be used as a surrogate for these constituents.

Statistical analysis of both historical and custom water quality sampling data has provided valuable information regarding the differences between low flow and storm flow pollutant loading to the stream. The conclusions from this information were used to guide future water quality sampling, parameter assumptions, and modeling.

B.6 Conclusions

Historical water quality collected from Cypress Creek watershed was analyzed in order to compare low flow and storm flow median concentration and loading rates. While storm flow loading of *E. coli* and TSS were significantly larger than the low flow loading, this was not the case for chloride or nutrients. This lead to the conclusion that storm flow loading of suspended solids and bacteria was significant. Consequently, addressing this pollutant export requires detailed stormwater sampling. Trend analysis was then performed using the Mann Kendall and Seasonal Kendall trend tests to ascertain the presence of trends in the water quality constituent concentration and loading. Trends existed in both the concentration and loading of the constituents in the downstream of the watershed, but not the upstream of the watershed. The increasing trends occurred during a period of intense urbanization. Specialized low flow and storm flow sampling were performed to acquire additional information regarding the relationships between constituents during storm flows and to establish a baseline of constituent concentration during low flows. Additionally, a single point source was observed to have significant

impact on the spatial profile of pollutant concentrations along the creek during low flow sampling.

By analyzing the basic statistics and trends for both low flow and storm flow events a number of conclusions were drawn regarding the relative influence of dry and rainy weather on pollutant loading to the stream. These conclusions were used to guide future water quality modeling efforts.

B.7 References

Boeder, M., and H. Chang. 2008. Multi-scale Analysis of Oxygen Demand Trends in an Urbanizing Oregon Watershed, USA. *Journal of Environmental Management*. 87:567-581.

Burn, D., and M. Hag Elner. 2002. Detection of Hydrologic Trends and Variability". *Journal of Hydrology*. 255:107-122.

Cude,C. 2001. Oregon Water Quality Index: A Tool for Evaluation Water Quality Management Effectiveness. *Journal of the American Water Resources Association*. 37(1):125-137.

Haan, s. 2002. "*Statistical Methods in Hydrology*". 2nd Edition. Iowa State Press. Ames, Iowa: 340-346.

Hamed, K. and A. Rao. 1998. A Modified Mann-Kendall Trend Test for Autocorrelated Data. *Journal of Hydrology*. 204: 182-196.

Helsel, D., D. Mueller, and J. Slack. 2006. Computer Program for the Kendall Family of Trend Tests. US Geological Survey Scientific Investigations Report 2005-5275.

H-GAC. 2005. Measurement Performance Specification Tables: A7 Quality Objectives and Criteria". Houston-Galveston Area Council. Acessed online at http://www.h-gac.com/rds/environmental/water/documents/mps_tables_fy04-05.pdf on January 10, 2009.

Hirsch, R.M., Slack, J.R., 1984. A Nonparametric Trend Test for Seasonal Data with Serial Dependence. *Water Resources Research* . 20 (6), 727–732.

Hirsch, R.M., Slack, J.R., Smith, R.A., 1982. Techniques of Trend Analysis for Monthly Water Quality Data. *Water Resources Research* 18 (1), 107–121.

Kahya, E. and S. Kalayci. 2004. Trend Analysis of Streamflow in Turkey. *Journal of Hydrology*. 289:128-144.

Kruesche, A. R. Carvalho, J. Moraes. P. Camargo, M. Ballester, S. Hornink, L. Martinellli, and R. Victoria.1997 Spatial and Temproal Water Quality Variability in the Piracicaba River Basin, Brazil. *Journal of the American Water Resources Association*. 33(5):1117-1123.

Larsen, S. B. Kronvang, J. Windolf, and M. Svendsen. 1999. Trends in Diffuse Nutrient Concentrations and Loading in Denmark: Statistical Trend Analysis of Stream Monitoring Data. *Water Science Technology*. 39(12):197-205.

Taylor, C., and J. Loftis. 1989. Testing for Trend in Lake and GroundWater Quality Time Series. *Water Resources Bulletin*. 25(4):715-726.

Yu, Y., S. Zou and D. Whittemore. 1993. Non-parametric Trend Analysis of Water Quality Data of Rivers in Kansas. *Journal of Hydrology*. 150:61-80.