Winter 2013

Watershed-Scale Hybrid Stochastic-Deterministic Modeling Framework and Diffused Sources Superpositioning

Ruby Juvah Damalie
Old Dominion University

Recommended Citation

WATERSHED-SCALE HYBRID STOCHASTIC-DETERMINISTIC MODELING

FRAMEWORK AND DIFFUSED SOURCES SUPERPOSITIONING

by

Ruby Juvah Damalie

B.S. June 2005, Kwame Nkrumah University of Science and Technology, Ghana
M.S. June 2008, Kwame Nkrumah University of Science and Technology, Ghana

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

CIVIL ENGINEERING

OLD DOMINION UNIVERSITY
December 2013

Approved by:

_____
Jaewan Yoon (Director)

_____
Xixi Wang (Member)

_____
Thomas Chapman (Member)

UMI Number: 3579614

UMI

Dissertation Publishing

UMI 3579614

ProQuest®

ABSTRACT

WATERSHED-SCALE HYBRID STOCHASTIC-DETERMINISTIC MODELING FRAMEWORK AND DIFFUSED SOURCES SUPERPOSITIONING

Ruby Juvah Damalie
Old Dominion University, 2013
Director: Dr. Jaewan Yoon

Predicting hydrologic system behavior is imperative to planning and management of water resources. The study developed an integrated hybrid stochastic and deterministic framework to improve prediction accuracy for overland flow and diffused sources in a watershed. The methodology includes sampling input parameters at system level and contribution of nonpoint source from hydrologically disconnected areas (heretofore referred to as system-level approach and superpositioning respectively). System-level approach includes the integration of a topography-based sampling grid generalized linear model developed by the study and Monte Carlo methods. The superpositioning method adopts in-stream water quality equation for overland flow pollution estimation.

The system-level approach was applied to the Patuxent watershed to determine runoff, phosphorus and total suspended solids using continuous rainfall. For overland flow, system-level approach (p-value of 0.68) was 0.51% off the observed flow compared with -21.9% for existing method (p-value of 0.11). Similarly for phosphorus, the model prediction deviated from the observed by 7% compared to that of the existing method which deviated by -32%. The results indicate that the system-level method is a better predictor for overland flow and nonpoint sources. In the superpositioning approach, phosphorus contributions were added to the system-level approach using an event rainfall. The prediction error reduced from 4.82% to -0.29% when the system-level

method was superpositioned with nonpoint source. Data from superpositioning analysis showed that including diffused sources contribution from hydrologically disconnected areas further improves the level of accuracy.

The study demonstrates that the framework reduces prediction error and has a high accuracy in reproducing watershed response. The hybrid methodology framework is superior to existing deterministic methods. Ultimately, this dissertation shows the potential of improving prediction accuracy of hydrologic systems by incorporating the strengths of both stochastic and deterministic models. The framework serves as a background for detailed applications for the developed models.

I dedicate this work to my husband Albert for all the sacrifice.

AKNOWLEDGEMENTS

NOMENCLATURE

BASINS	Better Assessment Science Integrating Point and Nonpoint Source software
ESRI	Environmental Systems Research Institute
GIS	Geographic Information System
GLM	Generalized Linear Model
HSPF	Hydrologic Simulation Program in FORTRAN
HUC	Hydrologic Unit Code
IDF	Intensity Duration Frequency
NCDC	National Climatic Data Center
NED	National Elevation Dataset
NHD	National Hydrography Dataset
NLCD	National Landcover Database
NPS	Nonpoint Source
NRCS	Natural Resources Conservation Service
NSE	Nash and Sutcliffe Efficiency
PDF	Probability Density Function
RMSE	Root Mean Squared Error
SAS	Statistical Analysis Software
STATSGO	State Soil Geographic Database
TSS	Total Suspended Solids
USEPA	United States Environmental Protection Agency

USGS	United States Geological Survey
WASP	Water Quality Analysis Simulation Program
°F	Degree Fahrenheit
cfs	Cubic Feet per Second
ft	Feet
ft^3/yr	Cubic Feet per Year
g/day	Gram per Day
in	Inches
kg/day	Kilogram per Day
m	Meter
m^3/day	Cubic Meter per Day
mg/L	Milligram Per Liter
mi	Mile
µg/L	Microgram per Liter

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The interaction in surface water systems between land surface and atmosphere is an important piece of information for water resource engineers and managers. This is because the interaction influences the hydrologic response in the watershed. Hydrologic responses such as runoff and nonpoint source are random in nature and difficult to predict due to their constant interactions with natural and anthropogenic elements. Understanding and predicting these exchanges is important to agriculture (irrigation planning and vegetation and crop growth), natural hazards prevention and mitigation (floods, droughts, erosion, landslides), and water quality management (point and nonpoint source pollutants in catchment and stream waters) (Troch *et al.*, 2003). Efficient management of the hydrologic response of a watershed is often implemented at a watershed scale. Hence, the need to understand the interaction between the watershed and stressors or triggering elements is important to ensure system's reproducibility.

Predicting watershed system's response is often accomplished through the use of models and/or intensive monitoring to gather adequate information. Monitoring is the most desirable, as it gives observers an actual snapshot of what is physically happening in the watershed. However, monitoring (1) could be very expensive, (2) often lacks baseline data with which results can be compared, (3) is impossible to monitor all aquatic systems at all sites (Albert, 2011), and (4) could lead to deficiency in the clear understanding of physical processes underlying the data (Niemi and Niemi, 1991). Though monitoring is an excellent option when available, statistical and deterministic hydrologic models have

been developed as alternatives to look at watershed processes where extensive long term monitoring is not feasible. Modeling tools provide a better understanding of the complex relationships between land use activities and hydrologic processes that occur within a watershed (Singh *et al.*, 2005; Im *et al.*, 2007). Such modeling approach can look at historical conditions in the same watersheds, or use reference watersheds for comparative approximation to establish subjugated characteristics.

Hydrologic models are mathematical expressions used to represent the response of a physical system that are both natural and anthropogenic in nature. They estimate hydrologic conditions over time and provide tangible approximations of the real world by simulating the movement of hydrologic processes and pollutants in manageable form. Hydrologic models can be classified as either stochastic or deterministic depending on the physical process defining the input data and the modeling process. Stochastic modeling allows the introduction of probabilities and likelihood in order to simulate systems that are subject to uncertainties. Examples of frequently used stochastic methods are Monte Carlo Method (Gardner, 1983; Huang and Lee, 2009), Latin Hypercube (LHC) (Benedetti *et al.*, 2011; Yeboah-Forson, 2007; Post *et al.*, 2008) and Markov Chain Monte Carlo (MCMC) (Kleidorfer *et al.*, 2009; Kuczera *et al.*, 2006; Vrugt *et al.*, 2008). The synthesis of stochastic models is largely based on extensive data. However, predictions and extrapolations by stochastic models are impeded where data is sparse and contains high level of noise. The computation process of pure stochastic models could also be very complex.

Deterministic models, on the other hand, are built on assumptions and simplifications of complex system processes (Mix, 1994). They always produce precise

outputs through known relationships among states and events given the same set of inputs (Melone *et al.*, 2005). They are very straightforward in application but are deficient in accounting for system uncertainty. Deterministic models usually predict stable system behavior accurately where there is abundance of data. However, these models often fail to estimate the true system response and trends when there is variability in space and time within system couple with sparse data. This failure is due to the assumptions and simplifications in deterministic approaches for natural processes that vary in nature.

Evidently, combining the two models described produces a hybrid model which exploits the strengths of both deterministic and stochastic models. A hybrid model reduces computation complications associated with stochastic models and concurrently accounts for system uncertainties related to deterministic models. Vojinovic *et al.* (2003) highlights the significance of using a hybrid method for hydrologic prediction by showing its strong potentials for reducing stormwater quality prediction error and uncertainty inherent in deterministic models.

1.1 Problem Conceptualization

The hydrologic response of a watershed is erratic and random due to the continual interaction between the watershed and triggering elements or stressors. Modeling of such a watershed system without accounting for various degrees of uncertainties, often leads to erroneous predictions. Incorporating these uncertainties into models and quantifying their impact on a model's output is desirable in hydrologic modeling and prediction (Post *et al.*, 2008; Saltelli *et al.*, 2000; Zaehle *et al.*, 2005).

A vast wealth of studies has been done on uncertainty propagation on hydrologic predictions. Most of these studies focused on model parameterization which relates

uncertainty to the model and assumes the input data is accurate (Vrugt *et al.*, 2003; Moradkhani *et al.*, 2005). Although uncertainty of a model by itself is important, studies by Smith and Hebbert (1979), Kuczera *et al.* (2006) Huang and Lee (2009) and Franceschini and Tsai (2010) showed the impact of parameter inputs uncertainty on hydrologic modeling. These studies defined uncertainty on weather elements (Kuczera *et al.* 2006; Franceschini and Tsai, 2010), infiltration (Smith and Hebbert, 1979) and Mannings roughness coefficient (Huang and Lee, 2009).

In contrast, little is known about how physiographic parameters such as infiltration, lower zone soil moisture, and surface roughness impact hydrologic prediction on a basin scale. Although Smith and Hebbert (1979) and Huang and Lee (2009) successfully investigated the effect of uncertainties in physiographic parameters, these models were limited to micro catchment scale. Field scale studies that are extrapolated to a watershed scale often leads to errors (Blöschl and Sivapalan, 1995). The reason is as spatial extent increases, the models assume homogeneity and may fail to capture the increasing level of heterogeneity imposed by the increasing scale (Jetten *et al.*, 1999; Wood *et al.*, 1988).

The need for a watershed-scale analysis requires further investigation to verify the implication of uncertainty on hydrologic prediction. Besides the issue of scale, it is evident that the impact of physiographic parameter on hydrologic prediction is not fully understood and entirely explored. Thus, this study introduces a hybrid modeling framework that can be used to understand the dynamics of uncertainty propagation from physiographic parameters at a watershed scale. This approach in theory would lead to reduce prediction error and enhance the understanding of their impact on hydrologic modeling.

1.2 Research Objectives

The overall goal of this research is to develop a hybrid hydrologic modeling framework by interconnecting deterministic and stochastic methods for estimating overland flux and resultant loadings for water quality characterization. The goal would be achieved by using a hybrid model developed from stochastic methods and widely known deterministic hydrologic model. Other specific objectives of this dissertation are to:

1 Implement and verify a stochastic system-level Probability Density Function (PDF) based on sampling methodology for model parameterization with prior/posterior sensitivity analysis.

2 Develop and verify a methodology that connects stochastically estimated parameters into a deterministic computation engine.

3 Conceptualize and implement deterministic and stochastic methodologies to represent nonpoint source into the proposed hybrid hydrologic modeling framework by using method of superposition over time and space.

4 Compare hybrid hydrologic modeling framework, representing both point and nonpoint source loading influx and efflux to conventional methodologies.

1.3 Originality and Contribution

In-depth review of literature has shown that the subject of uncertainty at input data level relating to physiographical parameters has not been fully explored. The thesis bridges the gap by enhancing the understanding of uncertainty propagation from physiographic parameters and their effects on hydrologic prediction on a macro-scale. This is a novel approach for large watershed modeling. The study introduces a framework at system-level which combines both stochastic and deterministic model to investigate

the propagation of uncertainties of physiographical parameters. The core of the study is the application of stochastic concepts, supplemented by the deterministic model as a computation engine for the hybrid modeling framework. The study argues that a system level approach to hydrologic modeling could significantly improve the reproducibility of outcomes and greatly reduce prediction error. The study also introduces a method that augments current exiting nonpoint source models to account for hydrologically disconnected areas. This approach is unique to this study especially for diffused source modeling.

The major contributions of this dissertation can be found in Chapters 3, 4 and 5 which describe the modeling framework of hybrid model and its applications to Patuxent watershed in Maryland. The study enhances the understanding of uncertainty propagation from spatially variable and heterogeneous physiographical parameters. The work introduces a methodology framework on pre-evaluation of hydrologic model parameters that can significantly reduce model prediction errors. This dissertation overall, would augment the knowledge of overland hydrologic modeling and would enhance decision makers' ability to make valid and more informed logical decisions in watershed management.

1.4 Chapter Structure

The dissertation is organized into six chapters. Problem conceptualization and key questions to be addressed are contained in Chapter 1 as introduction. In Chapter 2, relevant review of literature relating to the study is presented. This includes definitions and brief description of the hydrologic system, an overview of physiographic parameters relevant to the hydrologic modeling used in this study, data sources and impact of data

inherent errors in hydrologic prediction. The key objectives of the dissertations are addressed in Chapters 3, 4 and 5. Summary and recommendations from the study are presented in Chapter 6.

Chapter 3 describes a detailed formulation of the framework including assumptions and concepts underlying the development of the model. A system-level approach which integrates stochastic and deterministic approaches to determine responses in a watershed to hydrologic processes are outline. The chapter also includes the integration of a generalized linear model and Monte Carlo method for input parameterization. The inputs are processed by computation engine (e.g. Hydrologic Simulation Program Fortran) to estimate overland flow and nonpoint source pollutants of orthophosphate (as phosphorous) and total suspended solids.

Application of the proposed hybrid modeling framework to the Patuxent watershed at a basin scale is presented in Chapter 4. The framework stochastically preprocessed some selected hydrologic data as input into a deterministic model to determine the hydrologic response to nonpoint source pollutant flux of phosphorus and total suspended solids in the watershed. Evidence from the case study indicates that prediction error reduces significantly for overland flow when compared to the conventional deterministic method. The hybrid modeling framework used in this study effectively captures the spatial and temporal characteristics inherent in natural systems.

Chapter 5 extends the methodology framework to areas within the watershed that are typically omitted in conventional modeling process presented in Chapter 4 due to delineation errors caused by segmental elevation ridges. As a result, the amount of NPS generated from these omitted areas is often unaccounted for. The method of

superpositioning was used to account for the total generation of NPS in the watershed. In this analysis the NPS loading at the outlet of the omitted area was superimposed on to the nonpoint source computed by the hybrid model to determine the overall response of the watershed to nonpoint sources propagation.

CHAPTER 2

REVIEW OF LITERATURE

A surface water system includes runoff, storage systems (ponds, lakes and reservoirs) and conveyance (rivers and streams) systems. Understanding these systems entails the knowledge of the hydrologic cycle and processes that take place within a watershed. Processes at the interface between land surface and atmosphere determine the redistribution of water on land surface and underneath (Troch *et al.*, 2003). These watershed processes are driven by both climatic conditions and physiographic characteristics. The hydrologic response of a watershed to these processes varies spatially and temporally based on the spatial scale of the watershed. Ultimately, understanding the interrelationship between physiographic parameters and the hydrologic response of a watershed is essential to hydrology professionals towards efficient management of the water resources.

2.1 Surface Runoff and Diffused Source Characterization

Surface runoff is controlled by spatial and temporal characteristics of climatic patterns and the spatial variation of physiographic parameters. While the effect of climate is an important subject for discussion, this research focuses on the variability of physiographic parameters at a macroscale level. These physiographic parameters include geology, topography (relief, slope), vegetation type, landuse and soil type (Hundecha and Bárdossy, 2004; Shi *et al.*, 2007; Fohrer *et al.*, 2005; Li *et al.*, 2012). Runoffs are effective contributors of pollution to a body of water because they serve as the main transporter of pollutants on land (Barnes *et al.*, 2001; Arnold Jr and Gibbons, 1996).

Diffused sources (nonpoint source pollution) originate as the surface runoff generated from rainfall or snowmelt washes off buildup pollutants on land and transports them into a receiving body of water. The magnitude of nonpoint source generated is directly proportional to the runoff as low amount of surface runoff leads to relatively low amount of NPS effluent and vice versa (Luo *et al.*, 2006).

Nonpoint sources pollutants have intrinsic random nature because of surface runoff as a posterior to rainfall uncertainties (Fujiwara *et al.*, 1988). They have no specific outlet or discharge point but spread out over a large area (diffuse) as they enter a water body. The NPS effluents have no definite source but are closely linked to landuse types making them difficult to locate, characterize, control, and manage (Qin *et al.*, 2010; (Ribolzi et al., 2011; Nakane and Haidary, 2010; Broussard and Turner, 2009). Agricultural lands have been identified as the major contributor to NPS pollution. They contribute approximately 50% of pollution in lakes and 60% of pollution in rivers every year (USEPA, 1996). Urban and developed areas are also known to contribute significantly to NPS pollution (Carle *et al.*, 2005; Norman *et al.*, 2008). Effluents from urban areas constitute approximately 40% of pollution to waterbodies in the United States (USEPA, 1996).

Thomann and Mueller (1987) classified nonpoint source pollution into two main categories, namely conservative and non-conservative. Conservative pollutants do not degenerate or transform but non-conservative pollutants undergo transformation during transport. Nonpoint source pollution can also be categorized into five major classes including pathogens, nutrients, toxic contaminants, debris, and sediment (USEPA, 1996). Pathogens are disease-causing organisms, and are hazardous to health. Nutrients, such as

phosphorous and nitrogen, stimulate the growth of algae, phytoplankton and other aquatic plants leading to eutrophication and depletion of oxygen in water body. Toxic contaminants, heavy metals and pesticides, are hazardous to all forms of life. Debris degrades the aesthetic quality of waterways and can also be hazardous to animals and humans. Sediments generally alter streamflow and decrease the availability of a healthy aquatic habitat, and are the leading cause of water degradation. Total suspended solids are the result of sediment erosion. They are composed of minerals and organic particles that remain suspended in water or sink slowly but easily get resuspended when agitated. They cause turbidity and cloudiness that reduce the level of light penetration in water.

2.2   Hydrologic Systems Characterization and Prediction

The hydrologic characteristics of watersheds are dynamic in nature. They exhibit spatial and temporal characteristics and phenomena due to extremities imposed by physiographic parameters within a watershed. These constraints are due to the heterogeneity and spatiotemporal characteristics of physical features found within a watershed that greatly influence the hydrologic response.

Characterizing hydrologic systems relies on the understanding of how the watershed responds to triggering elements and external stressors such as human intervention. Hydrologic systems can be characterized by (1) observing trends in data obtained from long term monitoring and or (2) using existing models. Monitoring is the most desirable, as it gives observers an actual snapshot of what is happening in the watershed at any that time. However, monitoring (1) could be very expensive, (2) could lack baseline data with which to compare results, (3) is impossible to monitor all aquatic systems at all sites (Albert, 2011), and (4) could lead to deficiency in the clear

understanding of physical processes underlying the data (Niemi and Niemi, 1991). Though monitoring is an excellent option when available, statistical and deterministic hydrologic models have been developed as alternatives to look at watershed processes where long term monitoring is not feasible. Modeling tools can provide a better understanding of the complex relationships between land use activities and hydrologic processes that occur within a watershed (Singh *et al.*, 2005; Im *et al.*, 2007).

2.3 Hydrologic Input Parameters

Predicting the hydrologic response of a watershed depends mainly on the understanding of the interrelationship that exists between physiographic parameters and hydrologic processes. The hydrologic input parameters relating to physiographic features considered in this study are topography, surface roughness, infiltration, lower zone soil moisture, evapotranspiration and interception. Topography is the backbone to hydrologic modeling (Beven and Kirkby, 1979; O'Loughlin, 1981; O'Loughlin, 1986; MacMillan *et al.*, 2004). It defines the effects of gravity on the movement of water in a watershed (Wolock and McCabe, 1995; Wu *et al.*, 2008 and Sharif *et al.*, 2010) and impacts every phase of the hydrologic system. Surface (Mannings) roughness coefficient is a function of soil conditions and land use (USEPA, 2000). In addition to topography and roughness, infiltration, a function of soil type (Dexter, 2004; Lipiec *et al.*, 2006), is crucial to determining soil moisture and the amount of runoff. Likewise, interception, a function of vegetation cover, is one of the important parameters in hydrologic processes (Savenije, 2004) resulting from the interaction of rainfall with vegetation. This parameter contributes largely to the amount of runoff as it determines how much precipitation reaches the ground (Savenije, 2004; Gerrits *et al.*, 2010; De Groen and Savenije, 2006;

Shachnovich *et al.*, 2008). The precipitation that reaches the ground influences the lower zone soil moisture. Another input factor important to watershed modeling is evapotranspiration (ET). It controls the soil moisture in the lower zone. Detail knowledge of these hydrologic input parameters in a watershed is essential to researchers, engineers and other water resources practitioner.

2.4 Classification of Hydrologic Models

Hydrologic models are mathematical expressions that are used to represent physical hydrologic conditions of a watershed over time. They help to evaluate "what if" questions for better management of the system under study. Models are classified into various categories based on the type of modeling approach employed (Melone *et al.*, 2005). They can be categorized as either stochastic or deterministic models depending on the input parameter specification and the modeling process. Other classifications include lumped or distributed, event-based or continuous, empirical, conceptual or physically-based and steady or unsteady flow.

Stochastic models use probability distributions to represent their input parameters to produce a range of output parameters within a specified confidence interval (Melone *et al.*, 2005; Obropta and Kardos, 2007). They simulate systems that are subject to uncertainties. Examples of frequently used methods includes the Monte Carlo Method (Gardner, 1983; Huang and Lee, 2009), Latin Hypercube (Benedetti *et al.*, 2011; Yeboah-Forson, 2007;) and Markov Chain Monte Carlo (Kleidorfer *et al.*, 2009; Kuczera *et al.*, 2006; Vrugt *et al.*, 2008). These methods involve random sampling from a distribution of inputs and successive model runs until a statistically significant distribution of outputs is obtained (Yeboah-Forson, 2007). The Monte Carlo method used in the development of

the modeling framework in Chapter 3, typically samples the probable values for each uncertain input parameter into ordered segments of equal probability covering each of its possible segments. The synthesis of stochastic models is largely based on extensive data. However, predictions and extrapolations are impeded when data is sparse and/or contains a high level of noise. They are also computationally complicated and exhaustive (Vojinovic *et al.*, 2003). However, their conceptual framework makes it possible to describe heterogeneity where there are limited spatial or temporal details and are able to provide decision makers with the ability to determine uncertainty associated with predictions (Melone *et al.*, 2005; (Zheng and Keller, 2008).

Deterministic models are mathematical models that always produce outputs solely through known or pre-defined relationships among states and events given the same set of inputs (Melone *et al.*, 2005). They are built on assumptions and simplifications of complex system or processes. They predict very well with sparse data but often fail when variation in space and time exceed those assumptions and simplifications. Among the widely used deterministic models in watershed studies is the Hydrologic Simulation Program in Fortran (HSPF). The HSPF has been fully validated and shows good degrees of predictive accuracy (Bicknell *et al.*, 1985; Laroche *et al.*, 1996). Hydrology, nutrients, toxics, sediment and other water quality processes have been successful estimated with HSPF (Albert, 2011). In the development of the hybrid model used in this study, the HSPF is used as the computation engine to the stochastic model due to its input hierarchical structure. The use of HSPF in the hybrid modeling platform reduces the limitations associated with deterministic models.

2.5 Hydrologic Data, Parameter Synthesis and Preprocessing

Data for hydrologic modeling are obtained from direct sources such as field data sampling and/or from governmental institutions such as the United States Geological Survey (USGS) National Resources and Conservation Service (NRCS) and the United States Environmental Protection Agency (USEPA). Obtaining the appropriate data for hydrologic modeling is a major challenge. The challenge is the result of data scarceness, non-ready-to-use data, non-usable formats and scale, and dissimilar dates. Such deficiencies often require the modeler to synthesize the required data from the available relevant data into a usable format and scale required by the model.

2.5.1 Current Techniques for Data Processing and Synthesis

Geographic information system (GIS) is an indispensable tool often used by hydrologists to process hydrologic input data. It has the ability to integrate spatial data with non-spatial attributes of a data. When applied to hydrologic systems, non-spatial, information can include description of soils, landuse, landcover, groundwater conditions as well as man-made systems (DeVantier and Feldman, 1993). The capabilities of GIS often used in hydrologic modeling include data acquisition, data processing, storage, data manipulation, and display of results (Al-Sabhan *et al.*, 2003). Applications range from characterization of hydrologic tendencies to predict the response to hydrologic events (DeVantier and Feldman, 1993), mapping and visualization for hydrologic assessment (Robayo and Maidment, 2005), data parameterization for hydrologic models (Bhaskar *et al.*, 1992; Olivera and Maidment, 1999).

2.5.2 Errors in Hydrologic Data and Impact on Prediction

A major challenge facing hydrologic modeling is the propagation of error in hydrologic predictions. The prediction error arises from multiple sources of uncertainties which has been classified into natural variability and model related uncertainties (Tung and Yen, 2006). The inherent variability of natural systems is intrinsic and characteristic of the random phenomena exhibited in the input data.

Sources of input data uncertainties include measurement anomalies and limited availability of data (Crosetto *et al.*, 2000; Franceschini and Tsai, 2010), data incoherency and disjointedness due to dissimilar dates, inadequate definition of physical processes defining the data, inherent spatial and temporal randomness of variables (Franceschini and Tsai, 2010) and excessively coarse and fine scale (Crosetto *et al.*, 2000). In addition simplification of input data during processing and synthesis also introduces data uncertainties. Uncertainties could also arise from the methods and tools used to describe and model the physical system (i.e., sampling techniques, data acquisition, data analysis and mathematical modeling). The ability to incorporate these uncertainties into a modeling process is key to reducing errors in hydrologic prediction.

Although several sources of uncertainty exist in hydrologic modeling, the context of this study is centered on spatial and temporal randomness of physiography. It must be stressed that the focus of this work will be on the impact of uncertainty inherent in input parameters. Therefore, the use of the deterministic computation engine for simulating the system behavior of overland flow and nonpoint source in this study is considered to be sufficiently accurate; its uncertainty (e.g., resulting from weak formulations or numerical impreciseness) is not subject to consideration.

2.6 Overview Hybrid Modeling on Watershed

Application of stochastic methods like Monte Carlo as a tool in hybrid hydrologic modeling has been around for decades dating back to the pioneering studies by Smith and Hebbert (1979). That study analyzed the effect of spatial variation of soil hydraulic conductivity on soil infiltration. Hybrid watershed modeling studies usually focus on physical characterization or parameterization of physical processes. Most of the current studies focused on model parameterization (Saltelli *et al.* 2004; Post *et al.* 2008) by relating uncertainty to the model inefficiency and assuming that the input data is accurate (Vrugt *et al.* 2003 and Moradkhani *et al.* 2005). However, a couple of recent hybrid modeling studies have shown that input data can have significant impact on model outcome (Kleidorfer *et al.*, 2009; Huang and Lee, 2009; Franceschini and Tsai, 2010).

Although these studies did not consider basin wide watershed hybrid modeling or the full range of input parameter in this dissertation, these studies serve as motivation in the development of the present research. For example Huang and Lee (2009) investigated the impact of the spatial variation of surface roughness on flow hydrographs in a conceptual micro catchment. The study showed that the Monte Carlo roughness generated values produced peak flows when compared to the non-Monte-Carlo scenarios. They recommended that true watersheds may produce significantly different outcomes, and future studies should include the effect of other hydrologic parameters so as to assess the true effect of the roughness parameter. Similarly, Kleidorfer *et al.* (2009) integrated Monte Carlo methods with a stormwater model to analyze the impact of input data uncertainty on stormwater model parameters for two catchment areas of 89 and 38 hectares in size. The study defined uncertainty on rainfall as random and systematic. The

authors conclude that systematic rainfall errors have significant impact on flow model parameters while pollution parameters are influenced by both systematic and random rainfall errors.

Franceschini and Tsai (2010) went a step further to combine a stochastic, modified Rosenblueth method with a deterministic model framework to quantify the overall variability of the model's estimation of total polychlorinated biphenyls (PCB) concentration in a river. Their study showed that incorporating the uncertainty inherent input data into hybrid models lead to better prediction. However, this study, like those studies discussed above was performed on a mesoscale.

The scale of a model is an important subject in hydrologic modeling (Wood et al., 1988; Blöschl and Sivapalan, 1995).The accuracy of models depends on the scale with which they are developed. Field scale models perform better at micro scale and large scale models perform better at meso and macro scale levels. This is because different set of physical laws dominates at each level of scale (Klemeš,1983). Since planning and management of nonpoint source is efficiently implemented at a watershed level (Melone et al., 2005; Yuerekli et al., 2005), extrapolating a field scale model to a basin scale could lead to error (Blöschl and Sivapalan, 1995).

The current study fills the gap by presenting a methodology framework for predicting overland flow, phosphorous and total suspended solids at a basin (macro) scale. In addition to developing a basin wide framework, this study uses the key role of topography in hydrologic modeling to develop a generalized linear model (GLM) from which a probability density function was developed for sampling (the process is described into further details in chapter 3). The use of topography to develop a GLM for

stochastic sampling to the best knowledge of the author does not exist in any past or recent study. Also unique to the study is a method that augments existing nonpoint source modeling scheme for potential sources that are unaccounted for in typical hydrologic modeling procedure (the process is described into further details in Chapter 5).

CHAPTER 3

CONCEPTUALIZATION AND DEVELOPMENT OF HYBRID MODELING FRAMEWORK

3.1 Introduction

In water resources modeling, the key challenge is the ability to accurately predict the hydrologic response of the watershed. This is usually achieved by capturing the intrinsic properties of active hydrologic systems by reducing the uncertainties and error. Uncertainties are common in hydrologic systems that are highly variable in nature due to inconsistencies and heterogeneity caused by constant human intervention, natural disruptive patterns and evolutive of sudden processes (Salas, 1980). These intrinsic properties in hydrologic environments are defined by the population of its parameters. The population of system's parameters is never truly known, however samples can be drawn from the system's parameters through measurements and observations. In watershed studies, methods of inferring system behavior include (1) intensive monitoring of system performance, (2) using readily available system information with statistical concepts for modeling and (3) using existing stochastic hydrologic models.

Statistical sampling is a widely known concept employed in various researches fields such as studies in plants (Agarwal *et al.*, 2013; Naithani *et al.*, 2013), animals (Kidd *et al.*, 2007; Porter and Dooley Jr, 1993; Wegge *et al.*, 2004) and microorganisms (Asta *et al.*, 2002; Liski, 1995; Western and Grayson, 1998) to estimate the population of these systems from sampling. The sampling process usually involves random selection of datasets relating to a parameter in a system based on statistical approach. The principal advantages of this concept are inexpensive and faster data processing.

Preprocessing input data is an important practice in hydrologic modeling as data is often acquired in the format not directly usable by a model. Although preprocessing of raw input data for hydrologic basin scale exists, the concept of statistical sampling of input data is not fully explored in hydrologic modeling. Current approaches to preprocessing input data are based on using GIS tools, expert guesses and or empirical formulae. These methods could introduce some degree of error into the preprocessed data and may lead to inaccurate representative of the system. Alternatively, statistical sampling techniques can be used to preprocess input data as a means to reduce prediction error in hydrologic models.

The statistical characteristics of a watershed can be expressed by system moments of mean, variance, skewness and kurtosis. For most probabilistic models, estimation of system's mean and variance in particular is an important step towards stochastic modeling to define the likelihood of system response to probability density functions (PDFs). One of the goals of the modeling framework is to implement a PDF stochastic system-based sampling method to express key parameters in a watershed. Implementation and verification of such sampling methods could increase our understanding of hydrologic system behaviors and processes in model parameterization with prior/posterior sensitivity analysis. The realization of this goal includes (1) developing a generalized linear model to determine grids for sampling, (2) integration of the generalized linear model with Monte Carlo methods, and (3) selection of a deterministic computation engine.

3.2 System-Level Input Data Processing

3.2.1 Generalized Linear Model (GLM)

A generalized linear model, GLM, is part of statistical models that infers system response based on dependency components and inherent errors or residuals. It is based on the method of minimizing least squares of dependent variables (or system response upon trigger elements) and is an embodiment of both systematic and random errors that are usually assumed to be normally distributed (Olsson, 2002). Generalized linear models are used to predict expected values of unknown quantities given a set of observed values. They allow response variables to connect through link functions and also describe the scale of variance for each measurement through the predicted value.

In this methodology framework, a generalized linear model (GLM) was developed for determining an optimal grid or mesh size for sampling system parameters. A GLM is often developed by relating a set of observed dependent variable to independent variables. Consider a set of n observations with dependent variable, y_i and m number of independent variables, x_m where $i = 1, 2, ..., n$ and $m = 1, 2, ...$. For all the observations, the relationship between the independent and the dependent variables can be expressed as:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m + \varepsilon \qquad (3.1)$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$x_m = \begin{bmatrix} x_{m1} \\ x_{m2} \\ \vdots \\ x_{mn} \end{bmatrix}$$

$$\beta_m = \begin{bmatrix} \beta_{m1} \\ \beta_{m2} \\ \vdots \\ \beta_{mn} \end{bmatrix}$$

$$\varepsilon \sim NID \ (0,1)$$

where Y is the dependent variable and x_1 to x_m are vectors of the observation β_0 is unknown intercept, β_1 to β_m are unknown coefficients and ε is the error term and is assumed to be independent and normally distributed (NID) with a mean of zero and a unit variance. The GLM for the observation is expressed in the form:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 \hat{X}_1 + \hat{\beta}_2 \hat{X}_2 + \cdots + \hat{\beta}_m \hat{X}_m \tag{3.2}$$

$$\hat{Y} \sim NID \ (\mu, \sigma^2) \tag{3.3}$$

where $\hat{Y}$ is the response variable and is assumed to be normally distributed, $\hat{X}_1$ to $\hat{X}_m$ are the predictor variables, $\hat{\beta}_0$ is the corrected intercept and $\hat{\beta}_1$ to $\hat{\beta}_m$ are the corrected coefficients. A GLM consists of three main components, (1) a random component (Equation 3.3) which specifies the conditional distribution of the response variable, $\hat{Y}$ given the values of the explanatory variables in the model; $\hat{Y}$ is considered to belong to the Gaussian family (Nelder and Wedderburn, 1972), (2) a linear predictor that expresses the response variable as a function of regressors (Equation 3.2), and (3) a link function $g(\cdot)$, that transforms the expectation of the response variable, $\mu \equiv E(Y)$, to the linear predictor (Fox, 2008):

$$g(\mu) = (\hat{Y}) = \hat{\beta}_0 + \hat{\beta}_1 \hat{X}_1 + \hat{\beta}_2 \hat{X}_2 + \cdots + \hat{\beta}_m \hat{X}_m \tag{3.4}$$

In this study, a GLM was developed using spatial characteristics of the watershed to determine sampling grids and to capture the inherent characteristics of a domain without over- or under-discretizing a subcatchment in the watershed. The key independent variable central to the development of the GLM is topography. Topography

is the backbone to hydrologic modeling (Beven and Kirkby, 1979; O'Loughlin, 1981; O'Loughlin, 1986; MacMillan *et al.*, 2004) since it defines the effects of gravity on the movement of water in a watershed (Wu *et al.*, 2008) that impacts every phase of the hydrologic system. For example, topography has been shown to affect the (1) overland sheet flow path that precipitation follows before it becomes streamflow (Sharif *et al.*, 2010; Wolock and McCabe, 1995; Wolock *et al.*, 1990), (2) the spatial distribution of soil moisture within a watershed (Vivoni *et al.*, 2008; Wilson *et al.*, 2004; Burt and Butcher, 1985), and (3) the chemical characteristics of streamflow (Andersson and Nyberg, 2008). Likewise, hydrology of upland forested areas (Beven and Kirkby, 1979; Agnew *et al.*, 2006; Riveros-Iregui and McGlynn, 2009), soil erosion (Finlayson *et al.*, 2002; Montgomery *et al.* (2001) and behavior of runoff (Meybeck *et al.*, 2001,Weingartner *et al.*, 2007 and Viviroli *et al.*, 2007) are significantly impacted by topography. Hence it is evident from these and other studies that topography has an enormous impact on hydrologic prediction.

Although the above-mentioned studies emphasize the significant contribution of topography and its relationship to the hydrologic response of a watershed, none of the reported studies considered topography as a contributing variable in developing a GLM for estimating system characteristics. The method proposed in this study seeks to harness the critical role of topographic to develop a grid system based on a GLM. The GLM for expressing the dependent variable, grid size, is defined as a function of topography and the surface area of the subbasins within the watershed.

$$G_{SAMPLING} = f(S_{AREA}, T) \qquad (3.5)$$

where $G_{SAMPLING}$ is the grid size, S_{AREA} is the areal extent of the subbasins in the watershed, and T is the topography. The observed data for the independent variables, topography and area of subbasins, was obtained for Patuxent watershed in Maryland (Figure 3-1), from the United States Geological Survey (USGS). The Patuxent watershed was chosen because it is often used as a model for testing environmental and management strategies without the sociopolitical complications associated with multi-state jurisdictional conflicts (Boynton *et al.*, 1995) and is also an ideal site as it has nearly all the data required by this study. The data included a 30m National Elevation Dataset (NED) (http://ned.usgs.gov/), and hydrologic boundaries defined by USGS as hydrologic unit code (HUC) 8 and 12 (http://nhd.usgs.gov/). HUC8 is the watershed boundary for Patuxent and HUC12 is a subbasins or sub-domains within the watershed.

A usual practice to developing a GLM is to use observed datasets to represent the dependent variable. In this study, obtaining observed data series for the grid size was not possible. Instead, the observed grid size was estimated empirically from assumptions deduced from the relationship between topography and the hydrologic response of the watershed as presented in the aforesaid studies. The assumptions include:
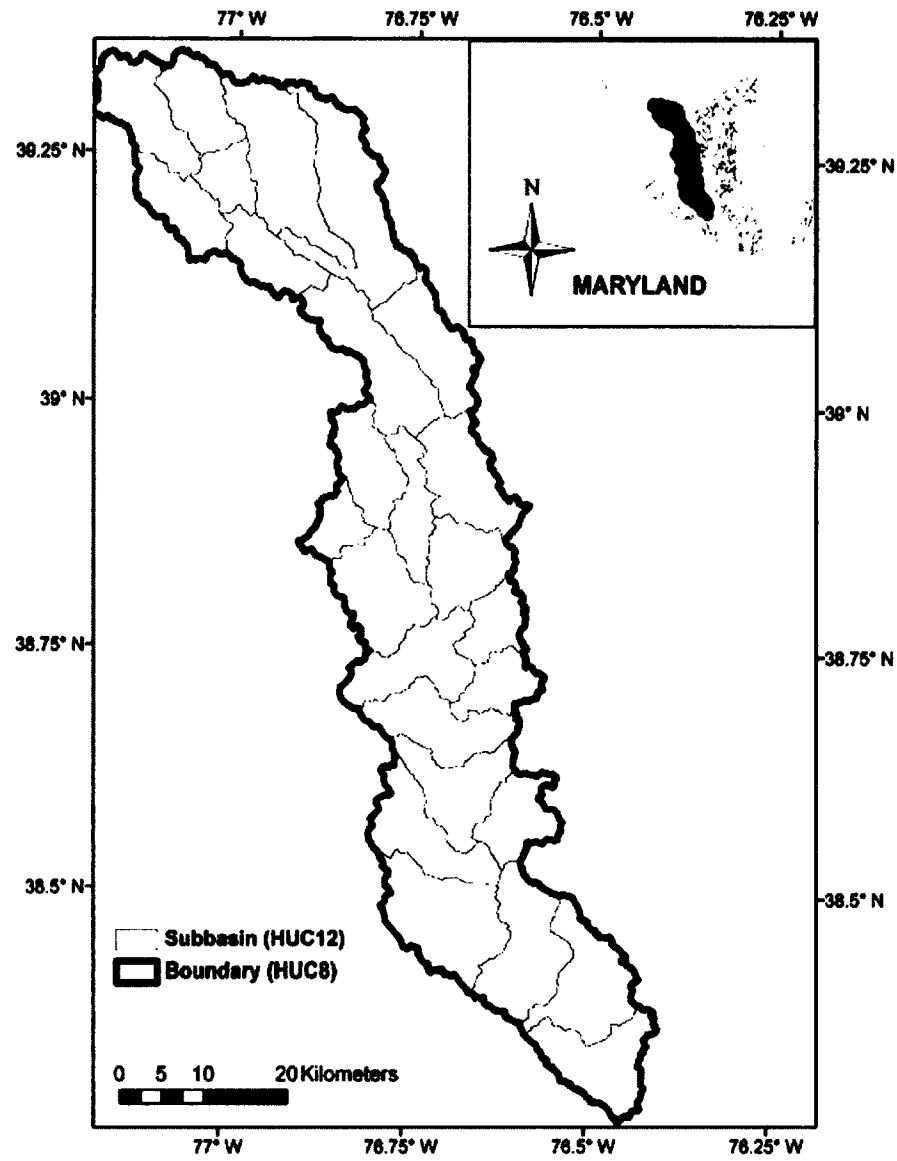
Figure 3-1: The Patuxent Watershed and Its Subbasins

(1) the topography of a domain is nonlinearly correlated with the heterogeneity and spatial variability of the domain's physiographic characteristics.

(2) the unit mesh for sampling corresponds to each physiographic parameter and is inversely proportional to the local relief and directly proportional to extent area of the hydrologic domain.

(3) the resulting mesh size is from a normally distributed population.

The concept of Chebyshev's theorem was adopted to make sure that the total number of mesh units per domain (a subbasin in the watershed) is greater or equal to 30 to ensure the normality required by the GLM model procedure. The Chebyshev's Theorem states that if a probability distribution has mean μ and standard deviation σ, the probability of obtaining a value that deviates from the mean by at least k standard deviations is at most $1/k^2$ i.e.,

$$Pr(|x - \mu| \geq k\sigma) \leq \frac{1}{k^2} \tag{3.6}$$

The empirical expression for determining the observed grid size is defined as:

$$G_{SAMPLING} = c \, S_{AREA}^{\frac{1}{ElevRange}} \tag{3.7}$$

where $G_{SAMPLING}$ is the mesh size, S_{AREA} is the area extent of the subbasins in the watershed; the coefficient c ensures the implementation of Chebyshev's Theorem (it can be any constant to ensure the total number of grid size per domain is greater or equal to 30, for this study, c has a value of 30), $ElevRange$ is the difference between the maximum and minimum elevation of the subbasin. The area for each subbasin was determined using the geometry calculating tool in ArcGIS. The field calculator tool in ArcGIS was used with Equation 3.7 to compute the observed grid size empirically for each subbasin (Figure 3-2).
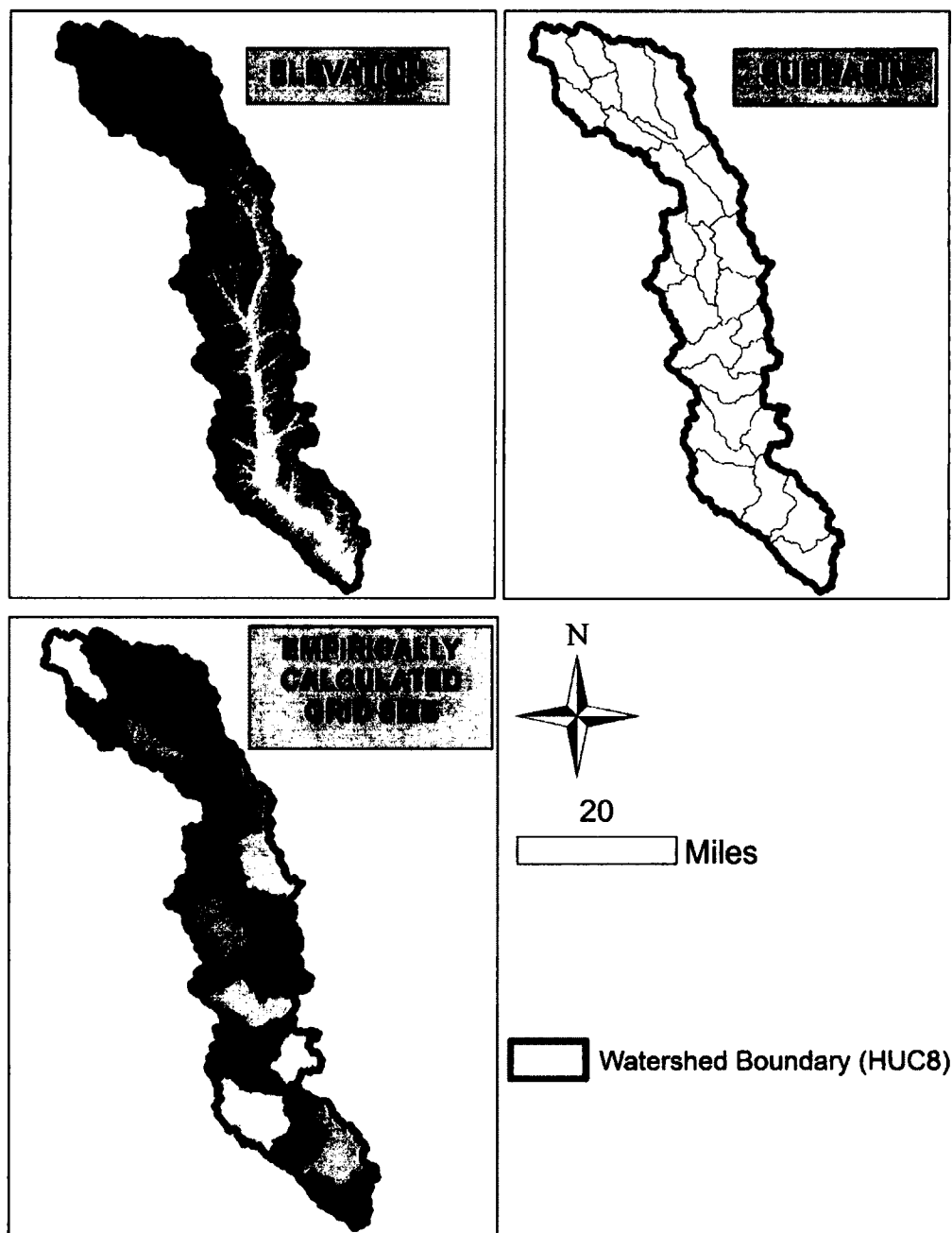
Figure 3-2: GIS Layers for Developing Generalized Linear Model

A prerequisite to developing a generalized linear model is to ensure that the

dependent variable, $G_{SAMPLING}$, is normally distributed (Equation 3.3). A normality check

for $G_{SAMPLING}$ in SAS (statistical analysis software, SAS Institute) gave a p-value of 0.38

indicating that the independent variable, the grid size, is normally distributed. Another

prerequisite, central to develop a GLM model is to establish a test of hypothesis on the model validity. Test of hypothesis (confidence interval) is a statistical deduction criterion that provides sureness of a conclusion. In this study, the hypothesis was defined at a confidence interval of 95% or a level of significance (α) of 0.05. The hypothesis is expressed as:

$$H_o: \hat{\beta}_1 = \hat{\beta}_2 = \cdots = \hat{\beta}_i = 0$$

$$H_a: at\ least\ one\ \hat{\beta}_i \neq 0$$

where $\hat{\beta}$ represents the constants and or coefficients in the equation, i represents the position of a particular constant or coefficient, H_o is the null and H_a is the alternate hypothesis. With the null hypothesis, the study assumes that all independent variables in resulting GLM do contribute to the dependent variable, $G_{SAMPLING}$, while with the alternate hypothesis, the study assumes at least one independent variable contributes toward the dependent variable. GLM process will evaluate and drop any independent variables that do not contribute. In other words, independent variables with p-values less than the level of significance are ones that actually contribute toward the dependent variable.

In developing the GLM, the topographical parameter is defined in two different forms, the elevation range (the maximum less the minimum elevation) and inverse of the elevation range. The aim is to determine which topographically defined parameter has a better goodness of fit. The resulting GLM would take the form:

$$G_{SAMPLING} = a + k\ S_{AREA} - m\ T \tag{3.8}$$

where $G_{SAMPLING}$ is the mesh size, S_{AREA} is the area extent of the hydrologic domain or a subbasin, T is a topographical parameter (elevation range or inverse of elevation range)

and a, k and m are estimated generalized linear model coefficients, $\hat{\beta}i$. Equation 3.8 produces two different equations depending on the topographical parameter used (Equation 3.9 and 3.10).

$$G_{SAMPLING} = a + k \, S_{AREA} - m \, ElevRange \qquad (3.9)$$

$$G_{SAMPLING} = a + k \, S_{AREA} - m \, /ElevRange \qquad (3.10)$$

Equation 3.9 was developed using generalized linear model option and Equation 3.10 was created with stepwise regression option in SAS (statistical analysis software, SAS Institute). The study evaluates the resulting generalized linear model to select the best equation for determining the grid size

3.2.2 Evaluation and Validation of the Generalized Linear Model

The parameters in the generalized linear model (Equation 3.9 and 3.10) are shown in Table 3-1. Table 3-1 shows quantitative statistical measures for validating the individual equations and the performance of each constant or coefficient presented in the equations based on the level of significance ($\alpha = 0.05$). Parameters that have p-values greater than the value of alpha are dropped from the final GLM models. The coefficients presented in Equation 3.9 and 3.10 meet the deduction criteria for acceptance as corresponding p-values are significant at $\alpha = 0.05$. The p-value criteria served as a basis to establish both equations as possible models for determining the size of a sampling grid.

Table 3-1: Constants and Coefficients of Generalized Linear Model

Constants/ coefficients	Equation 3.9		Equation 3.10	
	Value	P value	Value	P value
A	43.339	<.0001	28.376	<.0001
k_{ea}	2.00×10^{-8}	0.0027	68.171×10^{-10}	<.0001
m	-807.07×10^{-4}	<.0001	703.569	<.0001

Further statistical inferences of the equations are presented in Table 3-2. This includes f-statistics (F-value), p-value of f-statistics (Pr > F), RMSE, and coefficient of variation.

Table 3-2: Validation of Generalized Linear Model

Model Validation Parameters	Equation 3.9	Equation 3.10
F Statistics	106.78	6075.33
p-value for F Statistics	<.0001	<.0001
RMSE	0.923	0.129
Coefficient of Variation	0.075	0.079

F-statistical values are test statistics describing overall goodness of fit and validity of the GLM. Unlike the p-value for evaluating individual coefficients, the p-value for the f-statistics is for evaluating overall model-level validity. Thus significant p-values indicate a valid GLM model. The RMSE (root mean squared error) is the quantitative measure of the difference between the predicted and the observed value. It is defined as:

$$RMSE = \frac{1}{N} \sqrt{\sum (O_i - P_i)^2}$$

where O_i is the observed value, P_i is the predicted value and N is the number of observation. RMSE of zero indicates a perfect fit between the predicted and the observed. The coefficient of variation measures the dimensionless magnitude of variability (Abdi, 2010). Considering the quantitative statistical measures presented in Table 3-2, Equation 3.10 is considered more appropriate in representing system-level variability and heterogeneity and the physical topography of the subbasins in the watershed

The preferred GLM (referred as GLM from now onwards) was used to determine the grid size for each HUC12 to examine the distribution of the computed grid in the

watershed. The process was carried out in the ArcGIS environment using the field calculator tool. The topography in the watershed was classified into low and high topographic gradients and matched with the spatial variation of the computed (optimal) grid size. The upper portion of the watershed with higher topographic gradient is associated with smaller optimal grid sampling sizes while the lower relief areas are associated with larger grid sampling sizes (Figure 3-3).
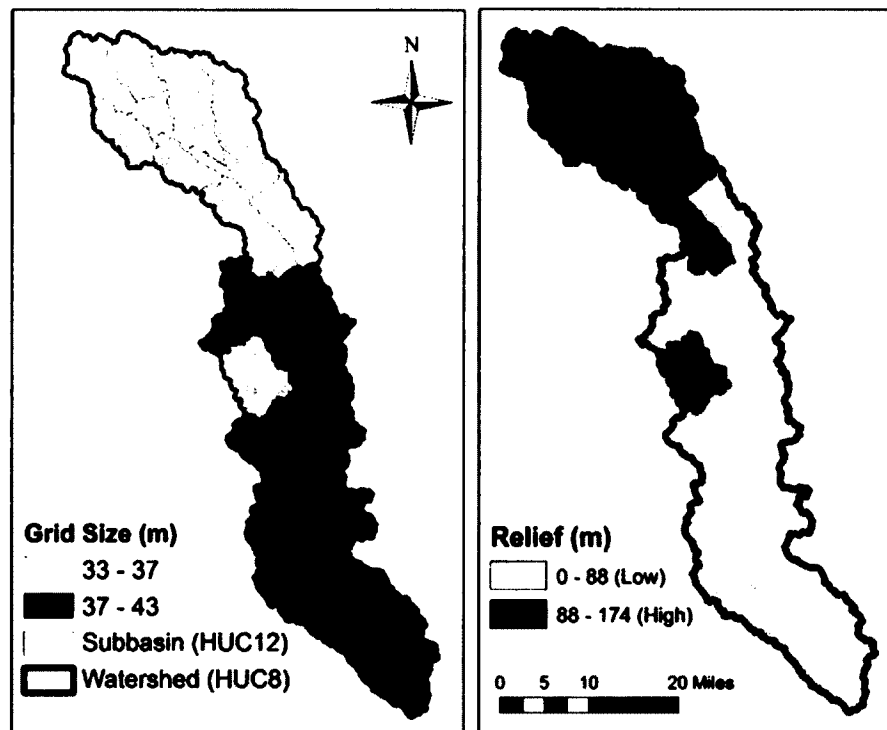
Figure 3-3: Distribution of Grid Size across the Watershed

The relationship exiting between the distribution of the grid size and relief is shown in Figure 3-4.
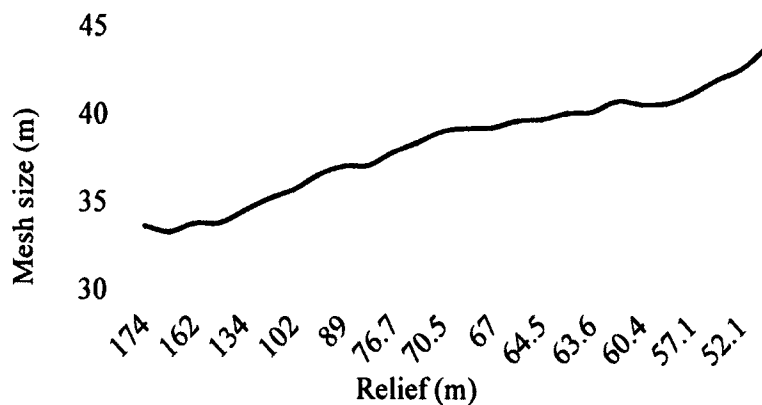
Figure 3-4: A Plot of Mesh Size versus Relief in the Patuxent Watershed

Evidence from the Figure 3-4 shows that the grid size is inversely proportional to the relief of the watershed and higher reliefs have smaller grid sizes as the GLM describes. The distribution of mesh unit sizes in the watershed according to the subbasins, HUC12 is shown in Figure 3-5.
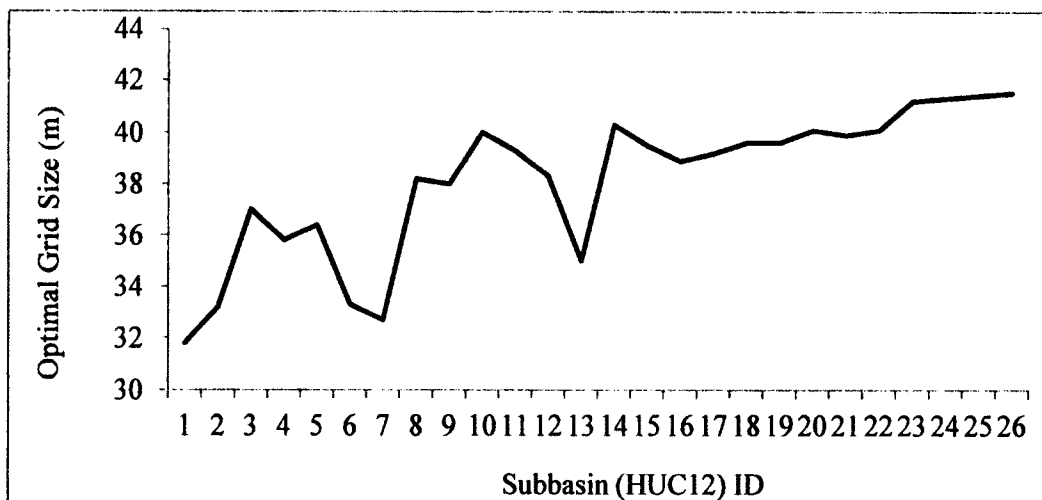
Figure 3-5: Mesh Unit Sizes for subbasins within the Patuxent Watershed

The plot shows that each subbasin has a unique mesh size. The relationship between the grid size and topography could have significant implications in hydrologic analysis. Areas with higher reliefs have complex and highly nonlinear hydrologic process

compared to gentle sloping relief areas. This conclusion is indirectly demonstrated in studies that showed that reliefs are major contributors to erosion (Finlayson *et al.*, 2002; Montgomery *et al.* (2001), a correlation exists between sediment load and elevation (Milliman and Syvitski 1992; Hay 1987) and runoff increases with steep relief (Weingartner *et al.*, 2007; Viviroli *et al.*, 2007; Meybeck *et al.*, 2001). Evidently, subbasins may exhibit localized topographical effect in the estimation of their hydrologic responses. This emphasizes the need to apply smaller grid to sample subbasins with higher reliefs as the spatial variability and heterogeneity in those areas are very prominent than subbasins with low lying areas.

3.2.3 Monte Carlo Sampling

Monte Carlo is a stochastic method and is based on probabilities and likelihood in representing system events and responses that are subject to uncertainties. It is widely used with applications ranging from engineering to zoology. Monte Carlo involves random sampling from a distribution of inputs and successive model runs until a statistically significant distribution of outputs is obtained (Yeboah-Forson, 2007). They can be used to solve problems with physical probabilistic structures, such as uncertainty propagation in models or solution of stochastic equations. Monte Carlo methods are also used in the solutions that can be modeled by a sequence of random steps that eventually converge to a desired solution (Bellin *et al.*, 1994; Isukapalli, 1999). Boyle (1977) defined the general equation of Monte Carlo methods as:

$$\bar{g} = \int_A g(x)f(x)dx \qquad (3.11)$$

$$\int_A f(x)dx = 1 \qquad (3.12)$$

where $g(x)$ is an arbitrary function and $f(x)$ is a probability density function, $\bar{g}$ is the

expected mean. The probability density functions for normal and uniform distribution:

Normal distribution:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{3.13}$$

Uniform distribution:

$$f(x) = \begin{cases} 0 & \text{for } x < a \\ \dfrac{1}{b-a} & \text{for } a \le x \le b \\ 0 & \text{for } x > b \end{cases} \tag{3.14}$$

where μ is the mean, σ is the standard deviation, σ^2 is the variance, a and b are the

minimum and maximum values respectively. For n number of samples values, x_i, the

estimate of $\bar{g}$ is obtained by:

$$\hat{g} = \frac{1}{n} \sum_{i=i}^{n} g(x_i) \tag{3.15}$$

$$\hat{s}^2 = \frac{1}{(n-1)} \sum_{i=i}^{n} [g(x_i) - \hat{g}]^2 \tag{3.16}$$

where $\hat{g}$ is the estimate of $\bar{g}$, $\hat{s}$ is the standard deviation of the estimate, x is the random

variable, and A is the range of integration. Further description of the methodology, tools,

and the applicability of the Monte Carlo methods can be found in Fishman (1996) and

Kalos and Whitlock (2008). In Monte-Carlo modeling, samples are dawn randomly from

larger sample size and repeatedly from probability distributions (e.g. uniform or normal)

through simulation to achieve convergence. The number of simulations depends on the

memory capacity of the computer and determines how extensively the parameter space is

sampled (Sarrut and Guigues, 2008; Jiang *et al.*, 2007; Landau and Binder, 2009). The variance reduces as the number of trials approaches infinity (Seydel, 2012).

The algorithm to Monte Carlo methods uses in this study follows the Post *et al.* (2008) approach shown below:

1. definition of model variables (input factors) X_i used for the analysis

2. selection of ranges and the Probability Distribution Functions (PDF) for each X_i

3. generation of samples within the PDFs (sampling)

4. evaluation of the model output for each element of the input factor sample

In this study, the model variables are defined as Mannings roughness coefficient, infiltration and lower zone soil moisture. The range of parameters was determined from the series of data sampled from the subbasin with the optimal grid. System representative values for these parameters were generated from the normal distribution defined as:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad (3.17)$$

where μ, σ and σ^2 are the mean, standard deviation and variance estimated using the sampling determined from GLM equation, and x is the model variable. The parameters were then used to determine overland flow and nonpoint source in Chapter 4.

3.2.4 Stochastic Input Data Parameterization

The system-level sampling approach (stochastic input data parameterization) includes the integration of the GLM (Equation 3.10) and standard Monte Carlo sampling to determine a lumped parameter value representative of the system. A schematic diagram of the process is shown in Figure 3-6.
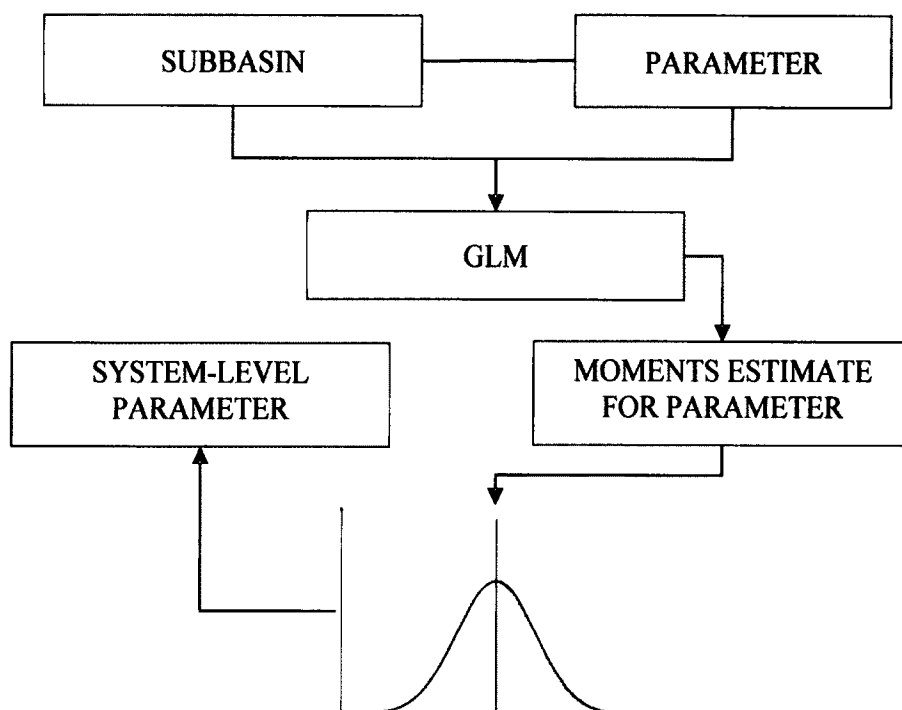
Figure 3-6: Schematic Representation of System Level Input Data Preprocessing

To determine the system-level parameter value, the GLM is first used to determine

an optimal sampling grid size for the subbasin of interest. The grid is then used to create a

sampling net or fishnet that is used to overlay the parameter of interest in the subbasin to

estimate the mean and variance of the parameter. The estimated mean and variance are

then used to initiate the sampling process in Monte Carlo. The Monte Carlo simulation

runs repetitively till a convergence to the probability distribution is reached (Bellin *et al.*,

1994). The convergence criteria proposed by Bellin *et al.* (1994) is defined as:

$$\delta^2 X_n = (X_n - X_{n-1})^2 \qquad (3.18)$$

where X_n is the target sample moment at a point, and n is the sample size (the number of

Monte Carlo Simulations). The simulation process reaches convergence when the spatial

average of $\delta^2 X_n$ over the whole computation domain vanishes (Bellin *et al.*, 1994; Ballio

and Guadagnini, 2004). The converged value is the system parameter value. The system

parameter embodies all possible variations and heterogeneity associated with the parameter in the subbasin. The sampling process accounts for possible discrepancies that may arise due to measurement anomaly, inadequate definition of physical processes defining the data, data disjoint due to dissimilar dates of collection and other possible inherent errors; consequently, reducing the level of uncertainty. The application of this approach is presented in Chapter 4.

3.2.5 Investigation of Stochastic Input Data Parameterization

The study tested the stochastic input data parameterization (system-level parameter sampling) approach on five selected subbasins in the Patuxent watershed. The subbasins were selected based on relief criteria, maximum, 1st quartile, 2nd quartile and 3rd quartile. The aim was to test the applicability of the system sampling in estimating parameters for the subbasins selected. Each of these subbasins has multiple landuses (Figure 3-7). The landuse data used was extracted from the National Landcover Database (NLCD 2001) (http://www.mrlc.gov) obtained for the watershed from the United States Geological Survey (USGS).
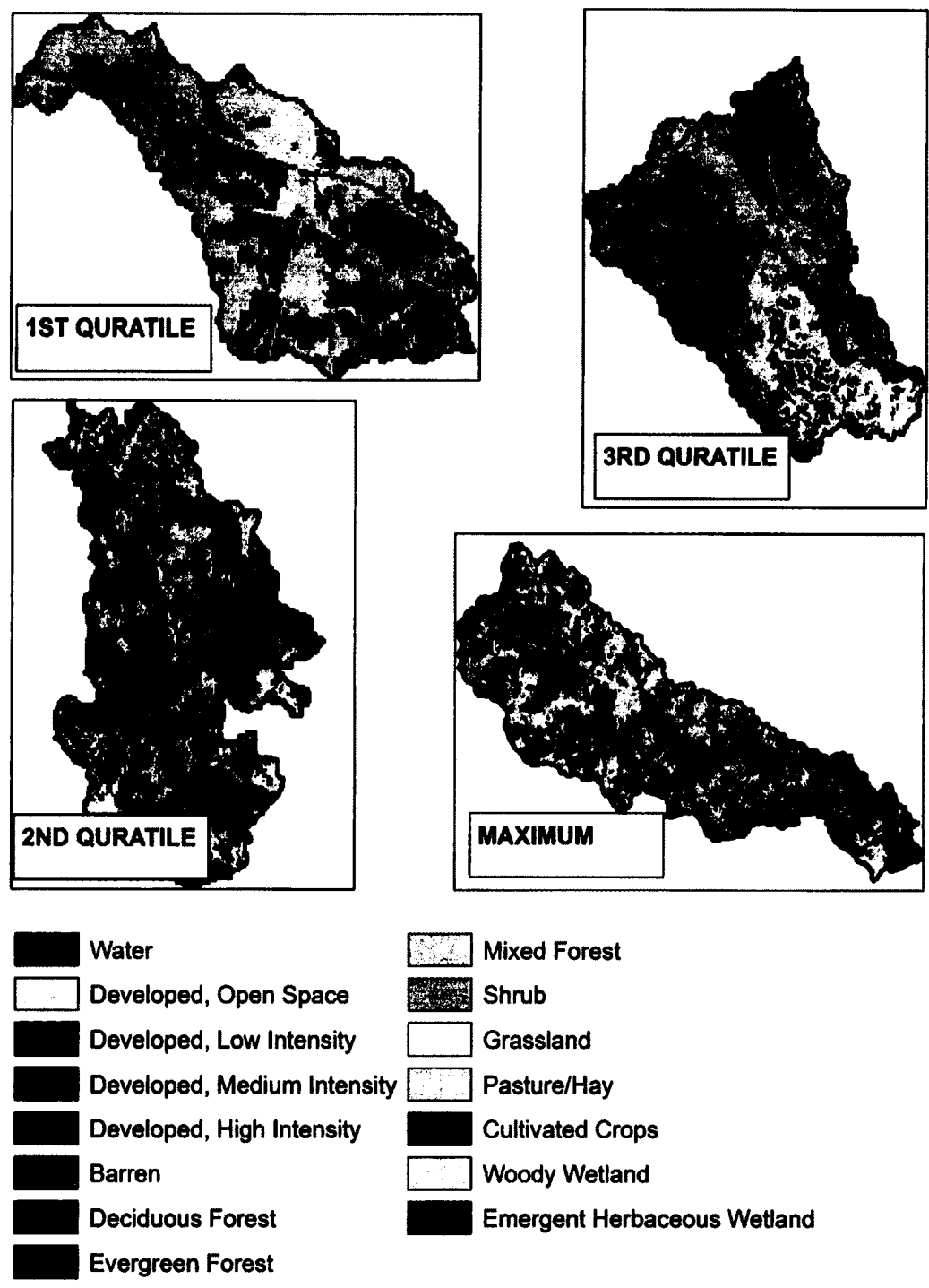
Figure 3-7: Landuse Types for the Selected Subbasins

Among the initial input parameters considered for sampling were Mannings roughness coefficient, infiltration and lower zone soil moisture. Lower zone soil moisture

defines the amount moisture content in the lower zone of the soil. To simplify the investigation process, the study selected the Mannings coefficient as the primary hydrologic input parameter. Primary hydrologic parameters are parameters that are directly linked to physiographic parameters and do not have to undergo any further processing before input into a hydrologic model. Other hydrologic input parameters that needed further processing are considered as secondary parameters, for example infiltration, interception and storage as they are often expressed in terms of other primary parameters for a basin scale modeling. Mannings roughness coefficient is one of the key hydrologic parameters used in estimating overland and channel flow. The Mannings roughness coefficient is an empirical numeric value and is a function of surface conditions (USEPA, 2000). For overland flow, Mannings roughness coefficient is the function of landuse and soil type. Mannings roughness coefficient values for this study were adopted from a published tabulation by Kalyanapu *et al.* (2010). These values were computed from NLCD (National Landcover Database 2001) landuse data based on percentage impervious and coverage of the landuse types as described by Homer *et al.* (2007).

Table 3-3: Mannings Roughness Coefficient Values for NLCD Landuse Data
(Kalyanapu *et al.*, 2010)

Code	Landuse	Published
21	developed open space	0.040
22	developed low intensity	0.068
23	developed medium intensity	0.068
24	developed high intensity	0.040
31	bareland	0.011
41	deciduous forest	0.360
42	evergreen forest	0.320
43	mixed forest	0.400
52	shrub	0.400
71	grassland/herbaceous	0.368
81	pasture/hay	0.325
90	woody wetland	0.086
95	emergent wetland herbaceous	0.183

Additional sources for Mannings roughness coefficients were obtained from published studies by Donigihan and Davis (1978), Engman (1986), Weltz *et al.* (1992), McCuen (1998) and Mays (1999).

To determine Mannings value for the selected subbasins, the spatial analyst tool was used compute the elevation range. The elevation range and surface area for each subbasin were then used as input to the GLM to compute the grid size. Each of the selected subbasins had a unique grid size. For each subbasin, the grid size was used to create a fishnet or a mesh in ArcGIS. The fishnet was then used to overlay the corresponding subbasin (Figure 3-8) to sample the Mannings roughness coefficient.
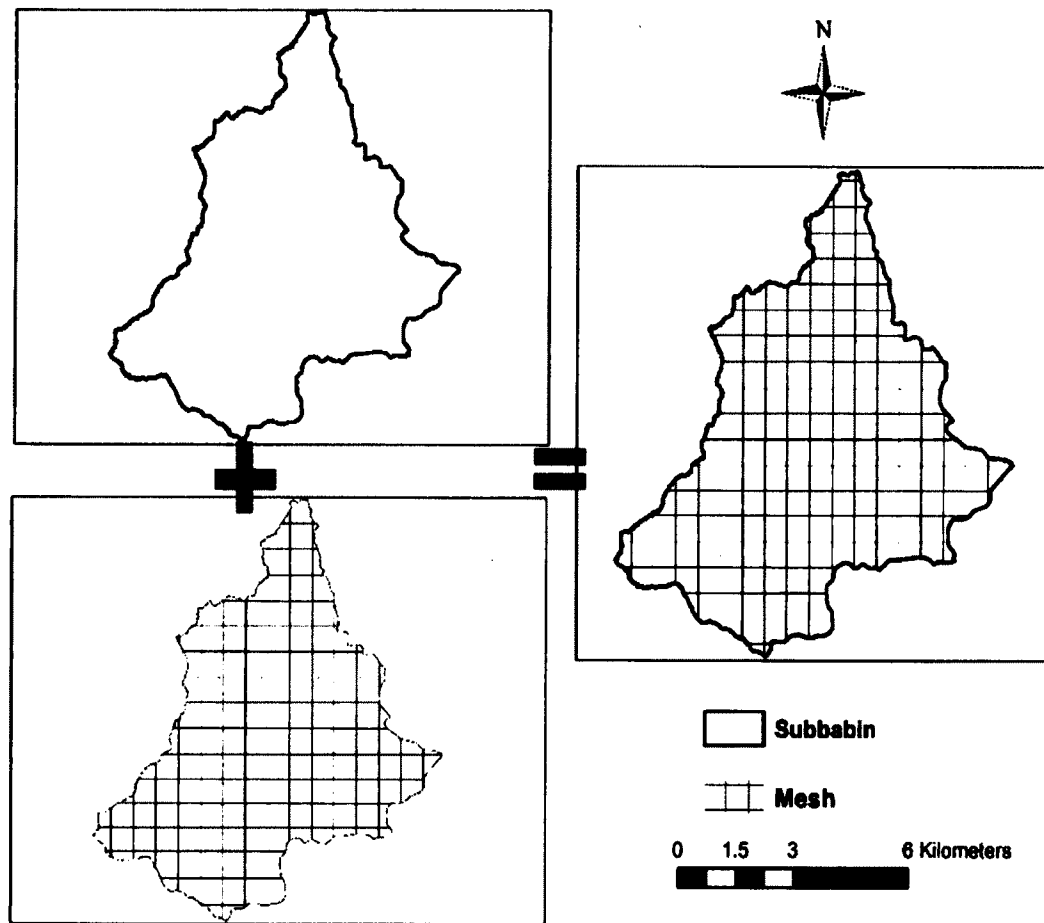
Figure 3-8: A Sample of Grid Overlay Analysis for Statistical Sampling

Each grid unit within the net was matched with the landuse type directly underneath, and the corresponding Mannings value was identified from Table 3-3 and entered into the attribute table of the grid. The spatial analyst tool is then used to estimate mean and variance of Mannings roughness coefficient for each of the subbasins (Table 3-4).

Table 3-4: Estimated System Moments for Selected Subbasins

Area Extent of Subbasin	Area (m^2)	mean	Variance
1st quartile	1080200	0.2592	0.0156
2nd quartile	16730100	0.1817	0.0191
3rd quartile	26800200	0.2153	0.0215
Maximum	80714700	0.2452	0.0212

The mean and the variance were then used with the Monte Carlo sampling method and simulated extensively to obtain the overall system parameter value. The resulting system level value for the Manning's roughness parameter is presented in Table 3-5.

Table 3-5: Mannings Values Estimated from Classical and System-Level Approach

Area Extent of Subbasin	Classical Method	System-Level Approach
	Composite Mean	System Level Mean
1st quartile	0.2594	0.2167
2nd quartile	0.1823	0.1978
3rd quartile	0.2155	0.1876
Maximum	0.2452	0.2272

The purpose of selecting the subbasins on a quartile criterion was to examine how well parameters can be estimated using the system sampling method. The system-level sampling method was able to process the Mannings roughness coefficient was successfully for the selected subbasins. It is evident from the results that the system-level approach consistently underestimated for the 1st, 3rd and the 4th quartile but overestimated for the second quartile. To explain the cause of this discrepancy, the study compared the landuse data, NLCD, to a world Imagery map obtained from Environmental Systems Research Institute (ESRI) (http://www.arcgis.com/home/group.html) to visually investigate the distribution of landuse in the subbasin. The world imagery map has a better resolution (1m) than the NLCD (30m). The comparison is based on the assumption

that the world imagery map is accurate and has no inherent uncertainty. The comparison showed that most of the landuse types that were classified as residential for the NLCD landuse were vegetation when investigated in the world imagery map (Figure 3-9). The study attributes the inability for the NLCD to identify those vegetative areas to its resolution.
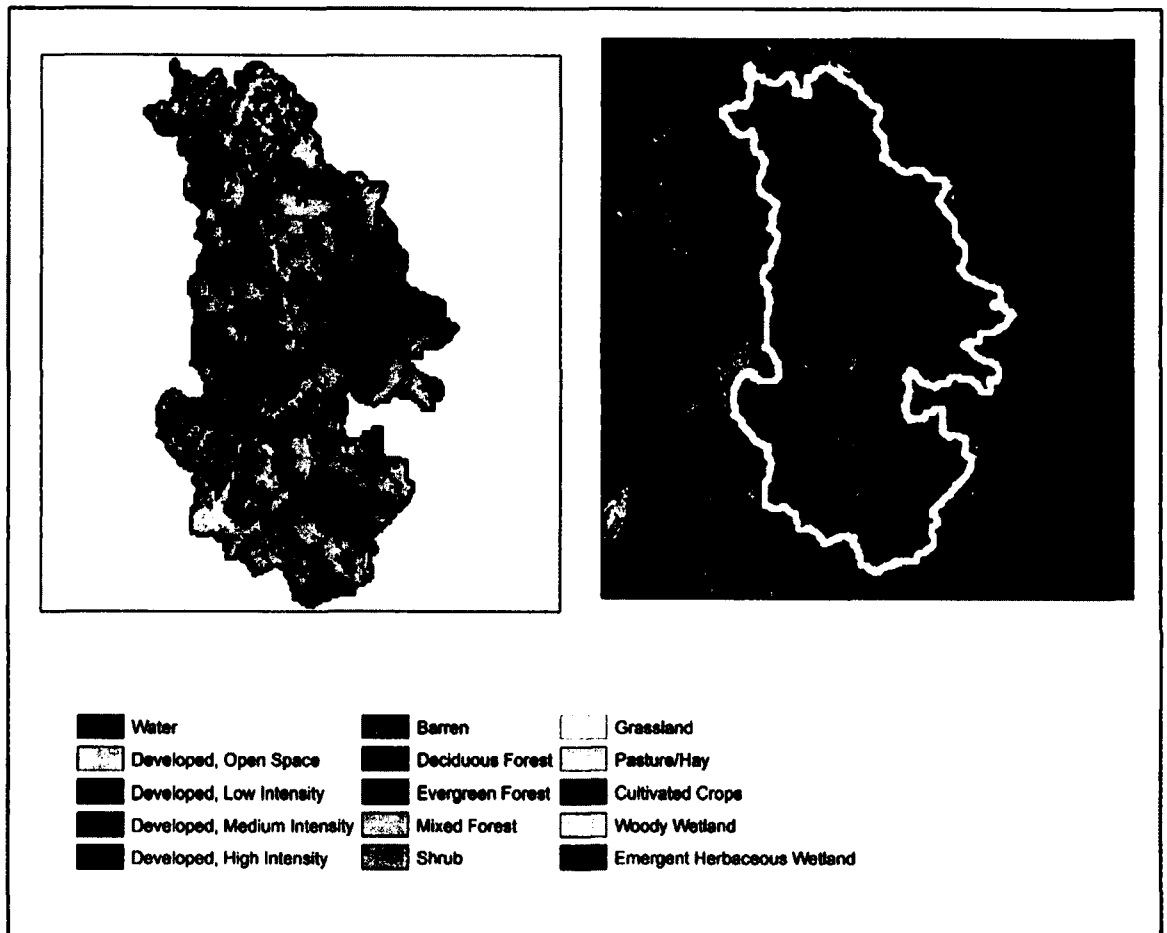
Figure 3-9: Landuse Comparison for the 2nd Quartile Subbasin

The Mannings roughness values for vegetation (e.g. forest, grassland) have higher values than residential areas (Table 3-3) thus, explaining the higher value for the 2nd quartile subbasin. The procedure was replicated to investigate the cause of underestimation for the remaining subbasins (1st, 3rd and the 4th). All the investigated

subbasins showed similar trends in the comparison between the NLCD and the world imagery map. For the sake of brevity, the study presents only the comparison for the 1st quartile subbasin. The comparison showed that the world imagery map has more built or paved areas than portrayed by the NLCD (Figure 3-10).

■ Water	■ Barren	☐ Grassland
☐ Developed, Open Space	■ Deciduous Forest	☐ Pasture/Hay
■ Developed, Low Intensity	■ Evergreen Forest	■ Cultivated Crops
■ Developed, Medium Intensity	▨ Mixed Forest	☐ Woody Wetland
■ Developed, High Intensity	▨ Shrub	■ Emergent Herbaceous Wetland
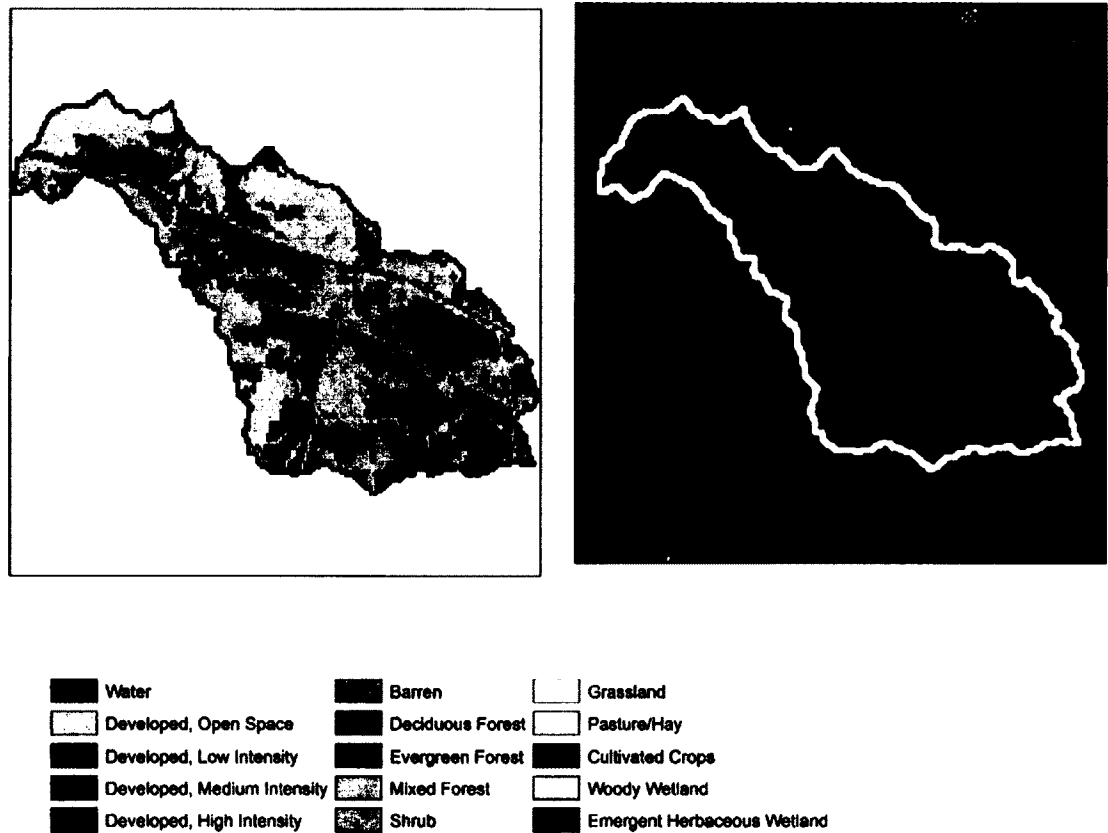
Figure 3-10: Landuse Comparison for the 1st Quartile Subbasin

Like the 2nd quartile subbasin, the NLCD was not able to capture the built areas because the areas are smaller than the resolution extent of the NLCD. Built areas usually have smaller Mannings roughness values (Table 3-3), hence explaining the cause of underestimation for the system-level value.

The illustrations (Figures 3-9 and 3-10) show that, the system-level sampling is able to capture the variability and heterogeneity associated with the Mannings parameter in each of the selected subbasins. The traditional method was not able to capture the variability due to the resolution of the NLCD data. The study infers from Figures 3-9 and 3-10 that the system-level sampling is able parameterized the Mannings roughness parameter in the selected subbasins while accounting for every possible variation errors and uncertainty inherent in the data in the NLCD data.

3.3   Deterministic Computation Engine

Unlike stochastic models, deterministic models do not account for uncertainty. This deficiency is compounded by their tendency to overestimate or underestimate hydrologic predictions due to assumptions and simplifications underlying their development. For example, deterministic models such as the Hydrologic Simulation Program in Fortran (HSPF) perform reasonably well under steady-state conditions. However, under time-variant conditions the prediction error could be high when the system's variability in space and time supersedes the assumptions and simplifications. For a calibrated HSPF model, the maximum acceptable prediction error is ±50% (Lumb *et al.*, 1994). This prediction error could be due to the inherent uncertainties present in input data sets, parameters and the assumptions underlying physical structure of the model. However, if the model is combined with stochastically preprocessed inputs, the magnitude of prediction error could be reduced. This is because stochastic methods reduce the level of uncertainty at the input and output levels of the model.

3.3.1 Description of the Selected Computation Engine

The Hydrologic Simulation Program in Fortran (HSPF) (Brian *et al.*, 1997) is a comprehensive, conceptual, dynamic watershed-scale model which simulates hydrology and water quality constituents in a watershed (Singh *et al.*, 2005). It is widely known for watershed-scale hydrologic simulations (Albek *et al.*, 2004; Hayashi *et al.*, 2004). The model has been found to be successful in handling hydrologic and water quality problems such as streamflow, loadings of sediment, nutrients, and pesticides from agricultural lands (Bicknell *et al.*, 1985; Moore *et al.*, 1988; Chew *et al.*, 1991; Laroche *et al.*, 1996). The HSPF was purposefully designed for use in mixed agricultural and urban watersheds (Borah and Bera, 2003; Obropta and Kardos, 2007). It is also incorporated as a nonpoint-source model into USEPA's Better Assessment Science Integrating Point and Nonpoint Sources (BASINS). BASINS is a comprehensive watershed tool which integrates hydrologic and hydraulic analysis tools on a GIS platform, and facilitates data input and other processes within a geographic information system (GIS) framework.

The model requires input information on landuse, soil properties, sources of nitrogen and phosphorus, stream reach characteristics, precipitation time series, and other meteorological data such temperature, solar, radiation and potential evapotranspiration to simulate the hydrologic response of a watershed. The model has three main modules that help to simulate pervious land segments (PRLND), impervious land segments (IMPLND), and free-flowing reaches or mixed reservoirs (RCHRES). HSPF uses a storage routing technique (Brian *et al.*, 1997) to route water from one reach to the next during stream processes. The model is capable of simulating flow rates, sediment loads, nutrients and pesticide concentrations. A detailed description of the HSPF model and it

capabilities are summarized in published works of Donigan *et al.* (1984) and Bicknell *et al.* (1997), and results from various applications of HSPF modeling studies are documented (Mishra *et al.*, 2007; Singh *et al.*, 2005; Lopez *et al.*, 2012; Nasr *et al.*, 2007; Johnson *et al.*, 2003; Hayashi *et al.*, 2004; Saleh and Du, 2004; Singh *et al.*, 2005; Laroche *et al.*, 1996; Benham *et al.*, 2005; Kim *et al.*, 2007).

3.3.2 Hybrid Model Set-up

The system-level framework was set-up by combining the stochastic (system sampling) approach described above and a calibrated HSPF model. The HSPF model was calibrated and calibrated for hydrology and nutrients. The system-level model receives inputs preprocessed by the stochastic method and uses the deterministic engine for the computation of the watershed hydrologic response.

3.4 Model-Level Sensitivity Analysis

A significant task performed in all modeling exercises is sensitivity analysis. Sensitivity analysis is the study of uncertainty propagation in the output of the model and the rate of change of one parameter with respect another parameter (McCuen and Snyder, 1986). Sensitivity of the model is computed as:

$$S = \frac{\partial O_o}{\partial F_i} = \frac{f\left(F_i + \Delta F_i, F_{j\,|j \neq 1}\right) - f(F_1,\ F_2, \dots F_n)}{\Delta F_i} \tag{3.19}$$

where S is the sensitivity index, O is the model's output and F is the input parameter. Equation 3.19 is expressed in two forms, absolute and relative sensitivities, depending on the application or intended use.

Absolute sensitivity

$$S = \frac{\partial O_o}{\partial F_i} = \frac{\Delta O_o}{\Delta F_i} \tag{3.20}$$

Relative sensitivity

$$S = \frac{\partial O_o}{\partial F_i} \cdot \frac{F_i}{O_o} \tag{3.21}$$

While absolute sensitivity is straightforward and easy to implement, it is neither invariant to the magnitude of the model output nor input and cannot be used to compare one input parameter to the other, it often used where hydrologic models become very complex and derivations are impossible to compute (McCuen and Snyder, 1986b). Relative sensitivity is invariant to the magnitude of the models output and input and provides valid means for comparing one input parameter to the other.

In this study model sensitivity of the system-level approach was assessed by investigating the influence of input parameters of the system-level model at macro level. Details of the application are shown in Chapters 4 and 5.

3.5 Overview of Methodology Framework

A methodology framework for predicting hydrologic responses at system level is proposed in this study. The approach is based on a hybrid stochastic method which combines a stochastic concept and a deterministic computation engine to estimate the overland flow, and nonpoint source as illustrated in Figure 3-11. Detailed methodology framework of the hybrid model can be found in appendix B.

DATA

Data
Preprocessing at
system level

Computation
Engine

Model level
Parameters

FINAL OUTPUT

Posterior
sensitivity

**Stochastic
preprocessor**
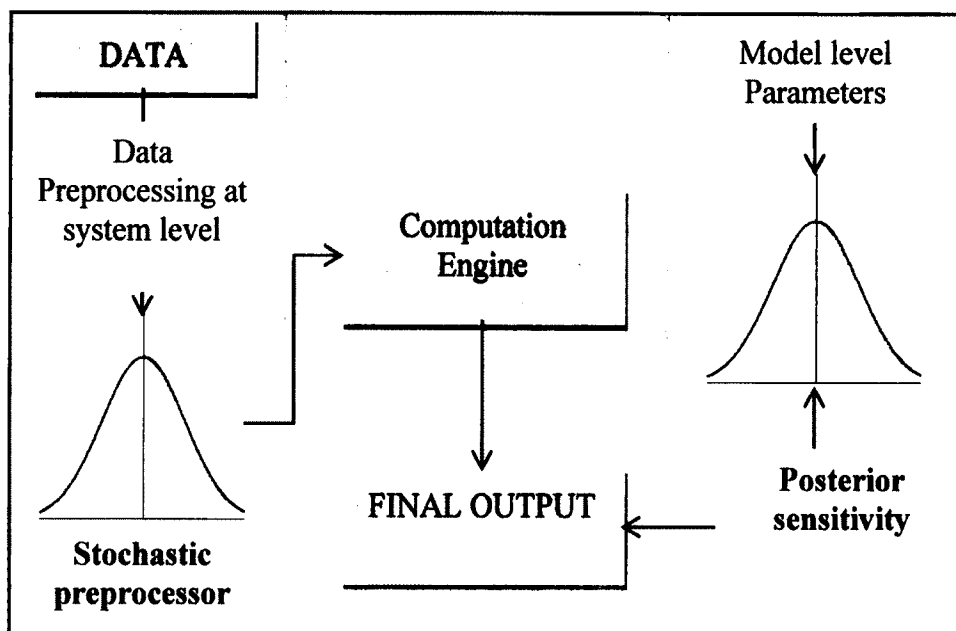
Figure 3-11: Schematic Representation of Hybrid Methodology Framework

The process involves three main phases: (1) data preprocessing at input level using stochastic concepts, (2) using deterministic model as a computation engine, and (3) model level sensitivity analysis. The data processing using stochastic concepts employs the use of the GLM (Equation 3.10) to estimate the system moments, mean and variance, of the parameter of interest in the domain being modeled. The estimated system moments are then used to initiate a Monte Carlo sampling to estimate system value for the parameter. The preprocessed values are then used as input into a deterministic computation engine to simulate the hydrologic response of the subbasin in a study watershed. In the third phase, the final output is assessed by examining the effect of other input parameters on the final outcome through sensitivity analysis. The three phase approach described here reduces the magnitude of uncertainty at the input data level prior to using the deterministic computation engine and also at the output level.

CHAPTER 4

THE HYDROLOGIC RESPONSE OF A WATERSHED: A CASE STUDY OF THE PATUXENT RIVER WATERSHED

4.1 Introduction

In this chapter, the methodology framework (system-level method) introduced in

Chapter 3 is implemented and verified to determine the hydrologic response (overland

flow, phosphorus and total suspended solids) for the Patuxent watershed. The hydrologic

input parameters applied to this study are infiltration, Mannings roughness coefficient

and lower zone soil moisture. Two other physiographic parameters, interception and

evapotranspiration, were selected to access their impact on the overall outcome of the

model. Detailed description of these parameters, their relationship with physiographic

parameters, and their importance to hydrologic modeling are discussed in detail in

Chapter 2.

4.2 Study Setting
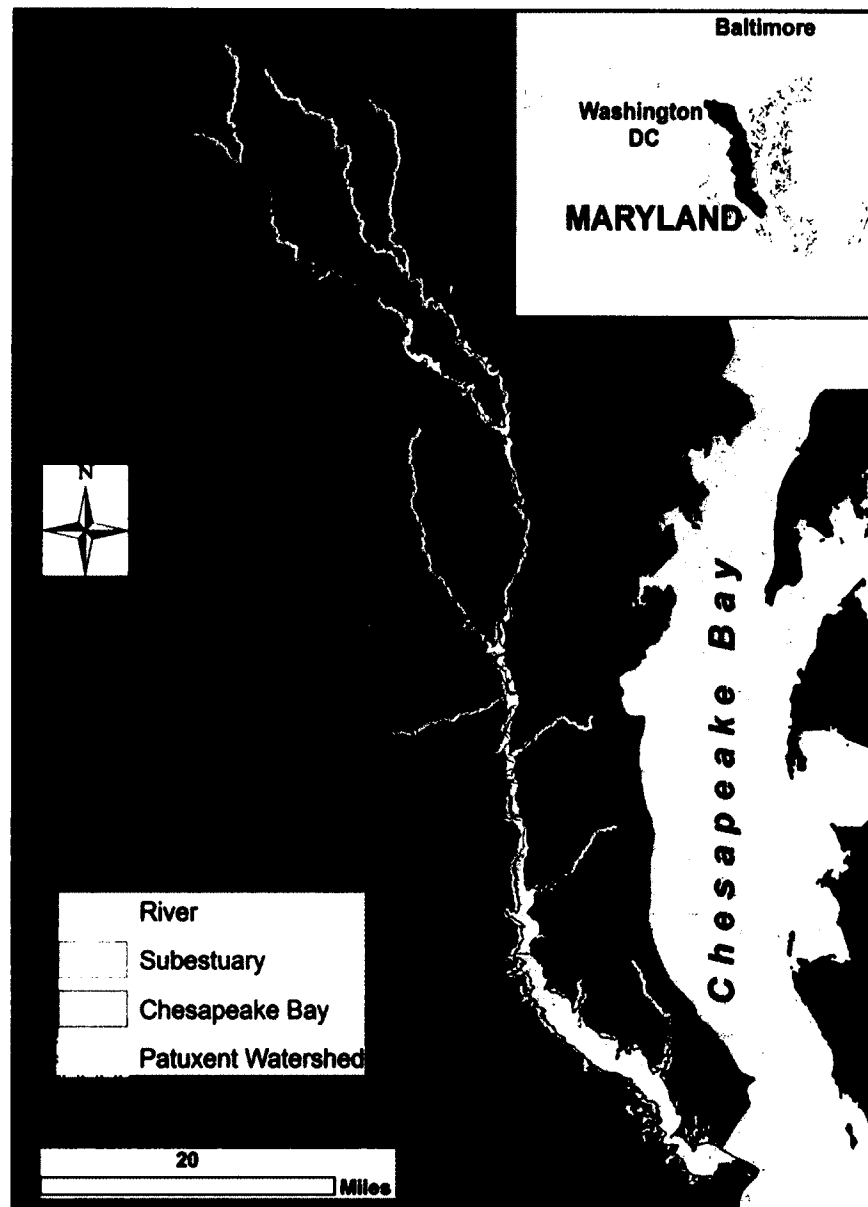
Figure 4-1: The Patuxent River Watershed

The Patuxent River Watershed (Figure 4-1) is located in the state of Maryland, geographically positioned between Baltimore, and Washington, District of Columbia metropolitan areas and runs across 9 counties (Montgomery, Howard, Carroll, Frederick, Anne Arundel, Prince Georges, Charles, St. Marys, and Calvert). It is the largest

(Harman-Fetcho *et al.*, 1999) and the longest river watershed in Maryland and cover about 10% of the state's total area (MD, 2010). The total surface area is approximately 871 mi^2 of landmass, 55 mi^2 of open tidal waters and 11 mi^2 of tidal marshes (Walter *et al.*, 2008), and drains into a 53 mi^2 sub-estuary of the Chesapeake Bay (Boynton *et al.*, 1995). The Patuxent River watershed is the sixth largest of the tributary watersheds in the Chesapeake Bay (Costanza *et al.*, 2002). The watershed is divided into two main regions, the piedmont physiographic (located in the upper portion of the watershed), which is 28% and the remaining being the coastal plain (Langland *et al.*, 1995) (Figure 4-2).
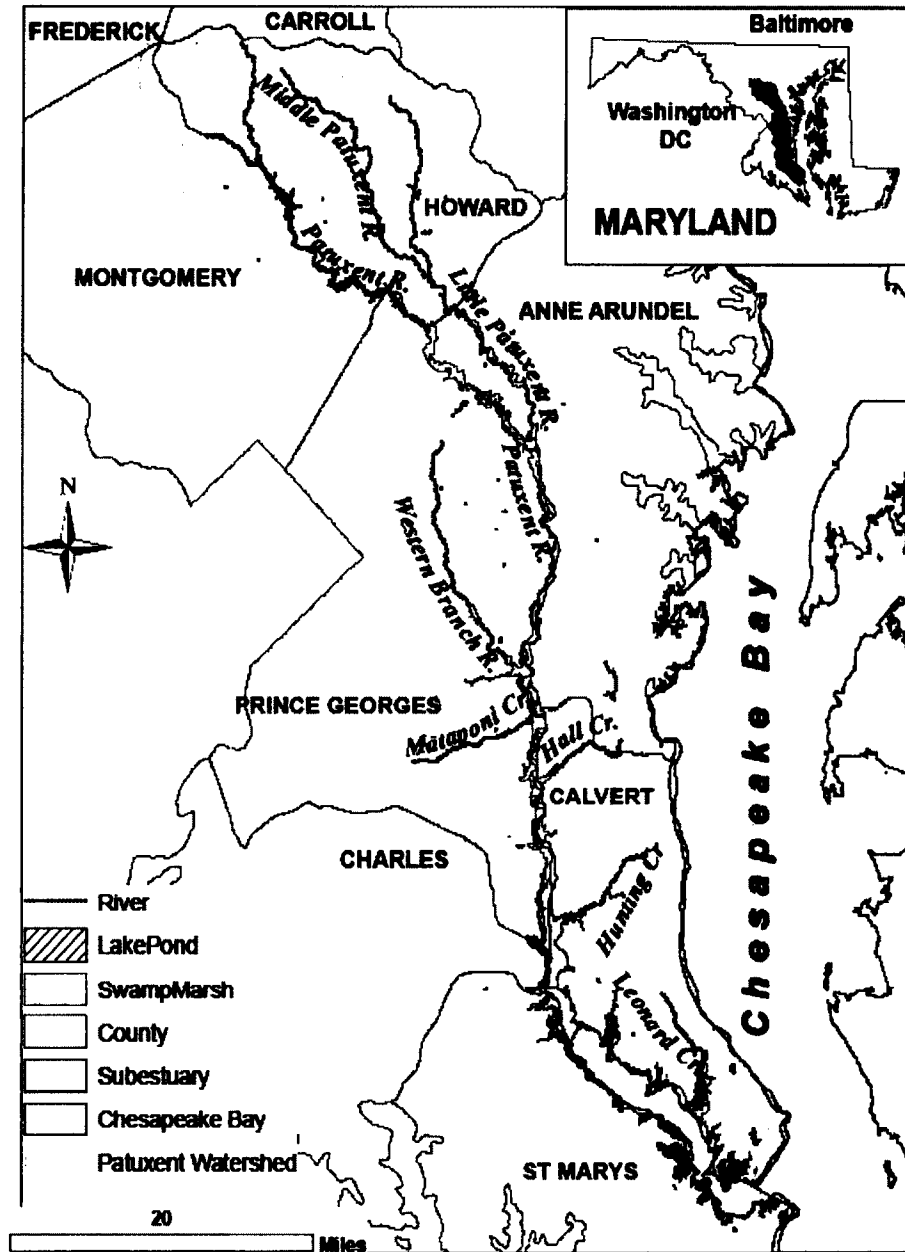
Figure 4-2: Extended Area of the Patuxent River Watershed

The Patuxent River Watershed was chosen for this case study primarily because of

the availability of nearly all the data required by this study. Also this particular watershed

is often used as a model for testing environmental and management strategies without the

sociopolitical complications associated with multi-state jurisdictional conflicts (Boynton

et al., 1995). In addition the watershed has long history of monitoring to provide clear baseline for historical events. Finally the USEPA have identified this watershed as particularly prone to NPS which is a major interest in this study.

The Patuxent River headwaters are in the piedmont region (Figure 4.2), located between 78.18°W and 77.18°W, where Carroll, Montgomery, Howard, and Frederick meet. The river has three main tributaries, the Little Patuxent, the Middle Patuxent and the Western Branch along with several minor tributaries. The Middle Patuxent flows to join the Little Patuxent River on the boundary between Anne Arundel and Prince Georges counties. The middle Patuxent flows from the west through the Prince Georges County and joins the Patuxent River at the second major confluence around the border of Anne Arundel, Prince Georges.

The Patuxent River is narrow and covers approximately 62 mi northwest from its union with the mesohaline portion at the mouth of the Chesapeake Bay (Figure 4-1) (Harman-Fetcho *et al.*, 1999), where it enters the bay. The total length of the river and the estuary is approximately 106 mi long; the lower portion is approximately 59 mi from the mouth and is constantly under the effect of tides while the upper portion is narrow, approximately 164 ft to 984 ft wide, with an average depth of 3.6 ft (Walter *et al.*, 2008).

The highest elevation in the watershed is 903 ft above the mean sea level. Annual freshwater in the watershed is approximately 3.0×10^{11} ft³/yr. Two large reservoirs (the Triadelphia and the Rocky Gorge) and a number of water storage features including ponds, lakes, swamps and marshes and reservoirs are found in the watershed. These contribute significantly to the hydrologic characteristics of the watershed. The mean annual precipitation is 44 in (Dail *et al.*, 1998) while the average annual temperature is

54°F (NOAA, 1997). The soil in the watershed ranges from clay to sandy soils. Dail *et al.* (1998) showed that the soils in the southern part of the watershed range from poorly drained silty soils along the shoreline to easily erodible sandy and loamy soils upland. However, in the northern part of the Patuxent River Watershed the soils composed of well drained silty loamy soils.

Estimates of land cover in the watershed were determined from the National Landcover Database (NLCD, 2001) landuse with a resolution of a 30m x 30m grid obtained from the United States Geological Survey (USGS) website. According to NLCD 2001, Land cover within the watershed is 39.9% forest (deciduous, evergreen, mixed), 22% developed (comprising open space, low, medium and high intensities), 21.7% agricultural lands (pasture/hay, crops), 7.9% wetlands (woody, emergent herbaceous), 5.7% water, 1.6% shrubland and 0.3% for grassland/herbaceous and barren. The main source of pollution in the Patuxent River watershed is non-point source which can be traced to development, agriculture and air deposition. The main pollutants impairing the water quality in the Patuxent watershed are phosphorous, nitrogen, sediment and bacteria level (Albert, 2011).

4.3 Data and Data Sources

Data obtained for the study includes stream network, topography soil, landuse types, water quality parameters, streamflow (Table 4-1). These were obtained from various governmental sites.

Table 4-1: Data and Data Sources

Data	Sources
Weather	the National Climatic Data Center (NCDC)
Soil - The U.S. General Soil Map was developed by the National Cooperative Soil Survey and supersedes the State Soil Geographic (STATSGO) dataset	Natural Resources Conservation Service (NRCS)
Hydrography - The National Hydrography Dataset (NHD): watershed boundaries, rivers/streams, lakes/ponds/reservoirs, marshes/swamps, estuaries	the United States Geological Survey (USGS)
Elevation - The National Elevation Dataset (NED)	the United States Geological Survey (USGS)
Landuse types - The National Land Cover Database (NLCD) 2001	the United States Geological Survey (USGS)
Percentage impervious - The National Land Cover Database (NLCD) 2001	the United States Geological Survey (USGS)
Streamflow	the United States Geological Survey (USGS)
Water quality parameters- orthophosphate as phosphorus and total suspended solids	the United States Geological Survey (USGS) and the United States Environmental Protection Agency (USEPA)

In addition to these sources, other hydrologic input parameters were obtained from published studies. Mannings roughness coefficient (Table 3-3) was taken from Kalyanapu *et al.* (2010). Kalyanapu *et al.* (2010) computed Mannings coefficient values for each NLCD landuse type using the percentage impervious and coverage specifications described by the data source. Finally, estimate for lower zone soil moisture for each landuse were extracted from published studies by Laroche *et al.* (1996) and Donigihan and Davis (1978). The lower zone soil moisture used in the study ranged from 4 to 11 inches. Infiltration values were estimated in ArcGIS by using the STATSGO soil

database, landuse types and infiltration values taken from USEPA (2000). For each landuse type in each subbasin, a value of infiltration is assigned based on guidance from USEPA (2000).

4.4 Pre-model Set-up

The Patuxent watershed was delineated into hydrologically connected subbasins and streams based on topography (National Elevation Dataset), existing stream network (National Hydrography Dataset) and watershed boundary. Delineated watershed was again subdivided to obtain outpour points that coincide with USGS streamflow and water quality gage stations that have up-to-date records (Figure 4-3). Specifically, USGS gage stations with up-to-date data in the watershed were selected for the study. This process is very important as it allows model's output to be compared with measured streamflow to assess the performance of the system-level model framework approach. The delineation gave a total of 117 subbasins and streams with two outpour points.

The delineation was performed using USEPA's Better Assessment Science Integrating Point and Nonpoint Sources (BASINS) software. BASINS is a multipurpose environmental analysis system designed to help regional, state, and local agencies perform watershed and water quality-based studies. It integrates hydrologic models on one platform and offers GIS tools to perform several GIS operations. The key advantage of BASINS is that it prepares input data files for the integrated hydrological models.
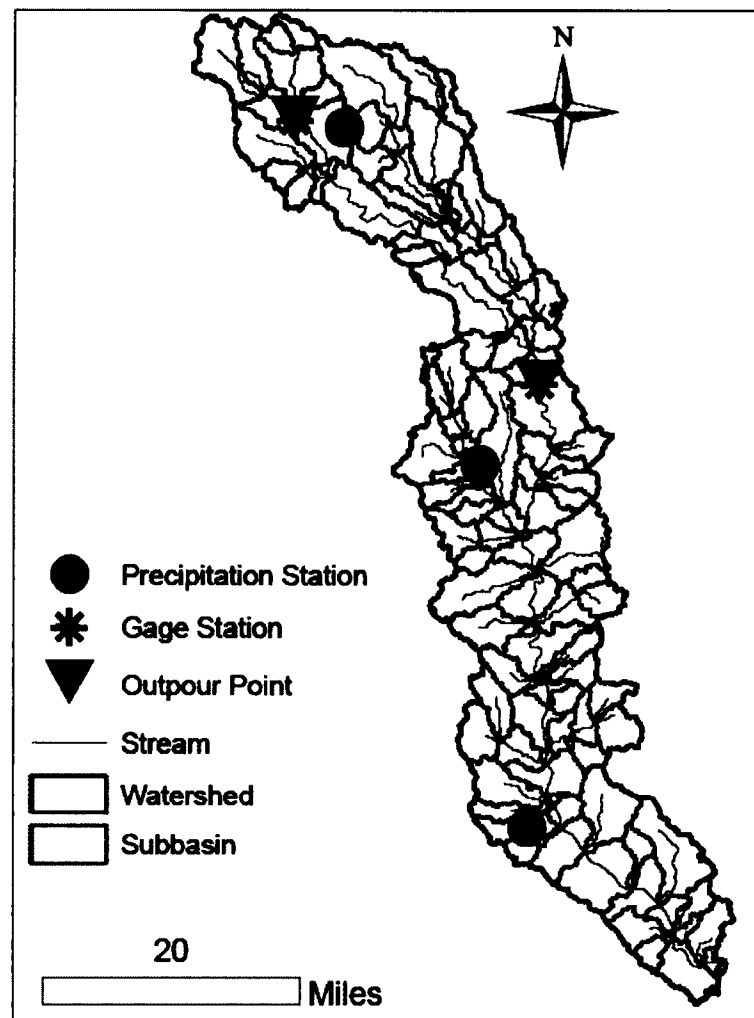
Figure 4-3: Delineated Subbasins and Streams

4.5 Methods and Procedures

Proposed hybrid modeling framework consist of three components; (1) microscale stochastic parameterization process at the subbasin-level, (2) mesoscale deterministic computational process at watershed segment-level, and (3) macroscale stochastic sensitivity process at watershed-level.

The generalized linear model (GLM) and Monte Carlo method described in Chapter 3 was employed to process input parameters. The GLM was used to estimate

system moments of the selected set of parameters within the subbasin. The Monte Carlo method was applied to draw successive samples from a PDF of each parameter to determine a system-level value for that particular parameter in the subbasin. The process was repeated for all subbasins. The resultant system-level values were collected and used as model input to HSPF for further computation of overland flow and pollutants in the watershed.

In HSPF, the watershed is modeled as homogeneous hydrologic segments. This is done by dividing the watershed based on the number of precipitation stations. In this study, three stations with up-to-date precipitation data (Figure 4.3) in the watershed was used to segment the watershed into three hydrologically homogeneous segments (Figure 4-4).
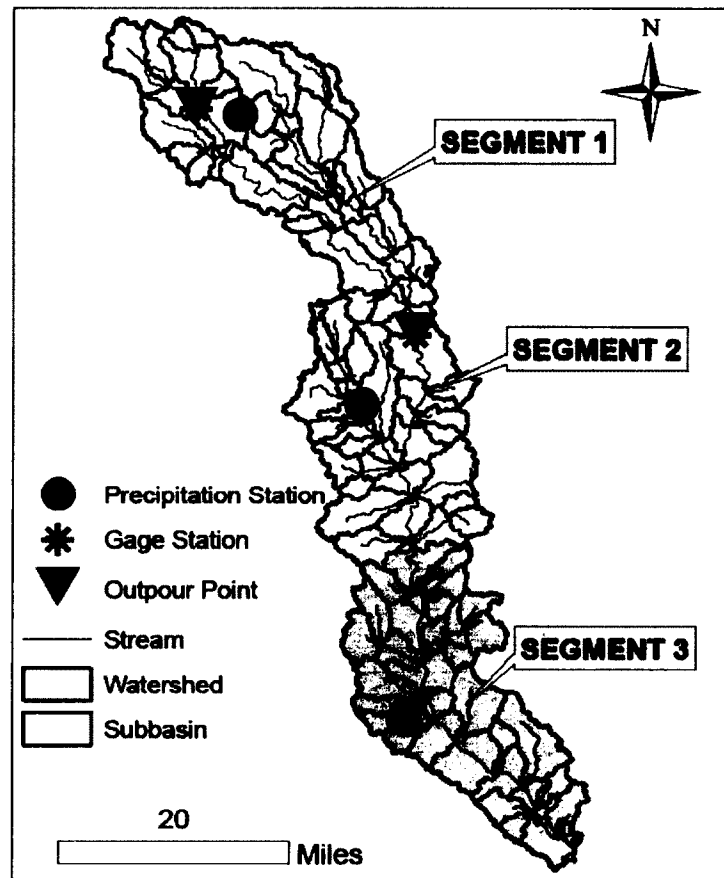
Figure 4-4: HSPF Model Segments

The stochastically processed infiltration, lower zone soil moisture and Mannings roughness coefficient were entered in HSPF to compute for overland flow phosphorus and total suspended solids. The results of the model were validated by testing the outcome at significance (α) level of 0.05 and with the Nash Sutcliffe efficiency (Nash and Sutcliffe, 1970). The coefficient of efficiency (NSE) ranges from -1 to 1. Higher values indicating better agreement; 1 indicates a perfect fit with observed data and 0 is an indication that the model is predicting better than the average of the observations. The coefficient of efficiency is mathematically defined as:

$$NSE = 1 - \frac{\sum_{i=1}^{N}(O_i - S_i)^2}{\sum_{i=1}^{N}(O_i - \bar{O})^2}$$

(4.1)

where O_i is the ith observed value, $\bar{O}$ is the mean of the observed values S_i is the ith simulated value, and N is the number of events modeled. The prediction error of the classical and the system-level method is computed as:

$$Prediction\ error\ (\%) = \frac{Simulated - Observed}{Observed} \times 100 \qquad (4.2)$$

In order to determine the macro-scale stochastic sensitivity of the whole watershed, a model-level posterior sensitivity analysis was performed on the model results using absolute model sensitivity (Equation 3.20). This was done by selecting other physiographic hydrologic input parameters that might impact the simulated results. The parameters selected were interception and evapotranspiration. Interception relates to vegetation type while landuse and evapotranspiration relates to vegetation type and root depth. The purpose was to investigate (1) the response of the system-level approach to other input parameters of the deterministic engine and to (2) determine if these parameters should be considered in future system-level preprocessing of input parameters.

Upper and lower bounds for the parameters were selected from the deterministic model to carry out the model level sensitivity. The lower and the upper bound selected from the model values were set at 25th and 75th percentiles. The purpose was to prevent any possible form of extreme skewness in the generated data. For each parameter considered, a series of values were stochastically generated within the specified bounds and at different positions from Monte Carlo (minimum, 25 percentile, 50 percentile, 75 percentile and maximum). The model runs were then evaluated for overland flow and nonpoint source pollutants.

4.6 Results

4.6.1 Overland Flow

The overland flow from the Patuxent watershed over a four-year period (1998 – 2001) was modeled for segments 1 and 2 using stochastically parameterized input data at system-level. The results of the simulation are shown Figure 4-5 for model segment 1 which is the upper headwater portion of the watershed.
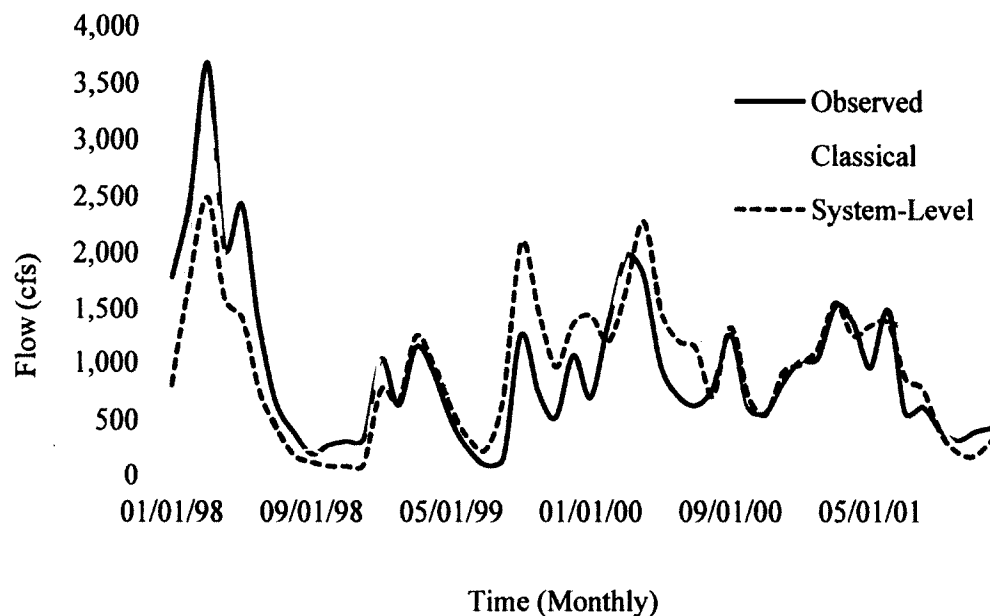
Figure 4-5: Overland Flow Prediction for Segment 1

The observed flow for segment 1 ranged from a low of 10 cfs in May 1999 to a high of 3600 cfs in June 1998. Evidence from the Figure 4-5 indicates that flow varies significantly depending on the season by exhibiting peak values mostly in summer and low values in the winter. Both the system-level and the classical (or conventional) methods generally predicted the behavior of the observed flow. For example, in January 2001, the observed flow was 980 cfs while the system-level and the classical methods were 982 and 990, respectively. Also in May 1999, the observed value was 745 cfs while

that of system-level and classical methods were 760 and 775 respectively. The results generally show that, the classical approach over-predicts for most months (where peak observed flow is recorded) when compared to the system-level approach which underestimated for most of the months. For example, in October 1999, the simulated flow for the classical method was 3250 cfs, while that of the system-level and the observed were 2200 cfs and 1500 cfs, respectively. Similarly, in April 2001, the simulated flow for the classical approach was 1755 cfs, while that for the system-level and the observed was 1550 cfs and 1510 cfs, respectively. The results also show that, the system-level method coincides with the observed flow in most of the months. This is evident for months like September 2000, February 1999, and July 2001. The results show that the system-level approach is a better predictor of the overland flow than the classical method.
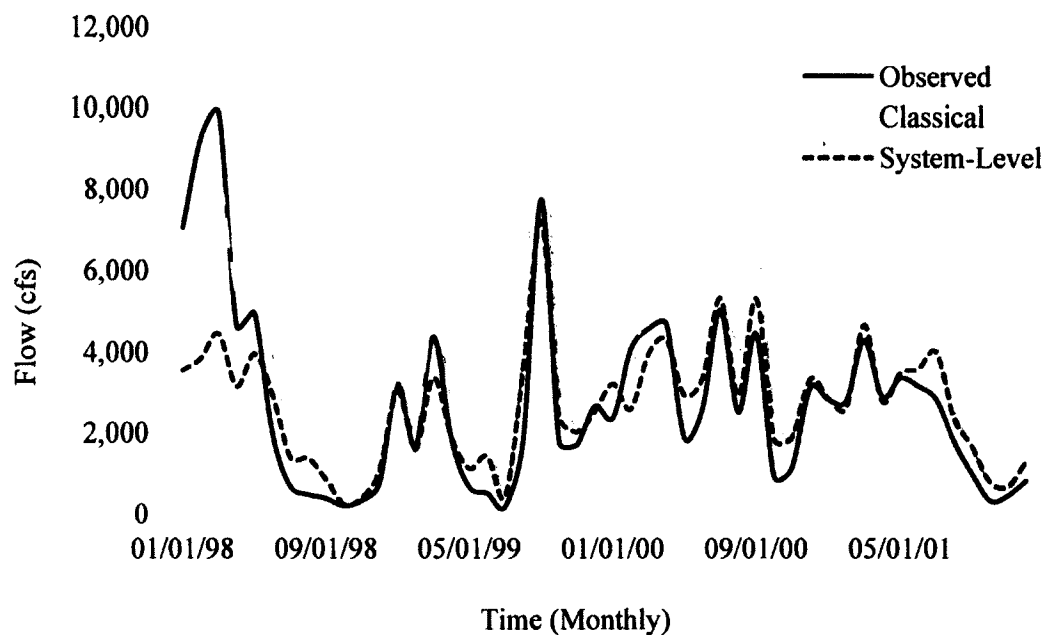
Figure 4-6: Overland Flow Prediction for Segment 2

The observed flow for segment 2 was generally higher than the observed flow for segment1. The observed flow for segment 2 for the four year period ranged from a low of

1 cfs in May of 1999 to maximum peak of 10000 cfs in July, 1998. Evidence from the Figure 4-6 indicate that flow varies significantly with seasonal variations. The flow is at its peak in the summer and lowest in the winter. Like segment 1 of the watershed, both the system-level and the classical predicted the behavior of the observed flow pattern. The classical method consistently over predicted the flow for most months, except those months in 1998. For example, in May 1999, the observed flow was 1000 cfs, the system-level was 1600 cfs while that for the classical was 2000 cfs. Also, in April 2000, the observed flow was 4700 cfs, the system-level was 4500 cfs while that for the classical was 6100 cfs. Unlike the classical method, the system-level approach evens out the amount of overestimation and underestimation throughout the simulation period. For instance, there was underestimation in February 2001 and overestimation in April 2001. Overall, results for segment 2 show that the system-level approach is a better predictor for overland flow than the classical approach. Model estimates from both methods were quantified statistically for high flow, low flows, summer and total seasonal flows for both model segments (Table 4-2).

Table 4-2: Statistical Validation for Predicted Overland Flow. Negative values imply under prediction and non-negative values indicate over prediction

	PREDICTION ERROR	
	Classical (%)	System-Level (%)
SEGMENT 1		
total runoff	21.9	-0.51
Seasonal flows	22	-0.45
Summer flows	68.6	28.3
10% high flows	-17.7	-8.3
50% low flows	30.3	0.6
SEGMENT 2		
total runoff	-27.9	-1.98
Seasonal flows	28.3	-2.34
Summer flow	59	26.01
10% high flows	7	-0.18
50% low flows	23.4	1.75

Table 4-2 compares the prediction error (Equation 4.2) of the classical approach to the system-level method. Evidence from the table shows that the classical method in segment 1 and 2 consistently overestimated for all the flow components except for low flows in segment 1 and summer and high flows in segment 2. Where the classical method over- or under-predicts, the system-level approach showed a reduction in prediction error. For example, in segment 1, the classical approach over-estimated high flows by 30.3% while the system-level approach over-estimated by 0.6%. Similarly, in segment 2, the classical approach under-estimated total runoff by 27.9% while system-level approach underestimated by 1.98%. The results show that the system-level approach reduced the prediction error significantly for overland flow in the study area.

The model results for the classical and the system-level approach were further validated using p-value and the Nash and Sutcliffe efficiency values (Equation 4.1). In segment 1, the p-value and Nash and Sutcliffe efficiency for the system-level approach

are 0.68 and 0.33 while that for the classical method are 0.11 and 0.12 respectively. In segment 2, the p-value and Nash and Sutcliffe efficiency are 0.45 and 0.25 for system-level approach while that for the classical method are 0.03 and 0.15 respectively. The statistical validation shows that the system-level method is robust and is more effective in representing the system characteristics.

4.6.2 Phosphorus

Results presented here are based on a two and half years of observed data (01/1998 to 08/2000) as data from 08/2000 to 12/2001 were missing Phosphorus as orthophosphate concentration was modeled for segments 1 and 2. Estimated orthophosphate concentration indicated that the system-level approach performed exceedingly robust compared to classical model of segments 1 and segment 2 (Figure 4-7).
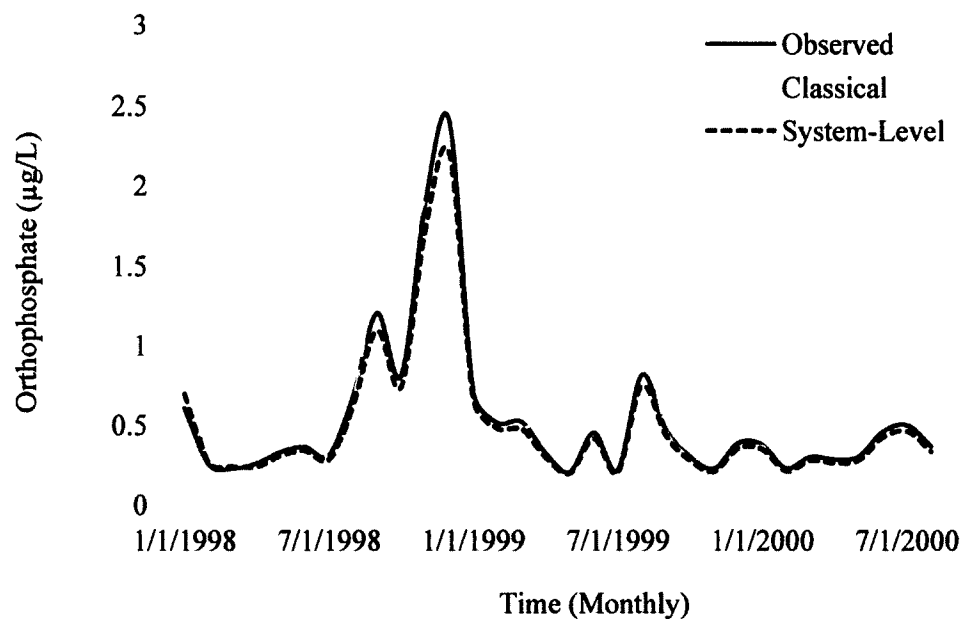
Figure 4-7: Orthophosphate Prediction for Segment 1

In the segment 1 of the watershed, the observed orthophosphate for the two and half year period ranged from a low of 0.19 μg/L in July of 1999 to a maximum peak of 2.45 μg/L in December, 1998. Orthophosphate production from the watershed was generally nominal except in the summer and the fall seasons. Evidence from Figure 4-7 shows that the system-level predicted closely to the observed value. The classical method registered severe localized deviations since the spikes for the simulated orthophosphate do not coincide with observed data. This is typified in June 1998 and 1999 where peaked prediction of classical approach was 0.95 μg/L and 1.59 μg/L respectively compared with approximately 0.35 μg/L and 0.2 μg/L predicted by the system-level and the observed. The results generally show that, the system-level approach predicts the system response to orthophosphate production for segment 1 better than the classical approach as the system-level was able to reproduce the peaks and lows and other variations exhibited by the observed data. Interestingly both methods consistently underestimated the simulated orthophosphate in segment 2 (Figure 4-8).
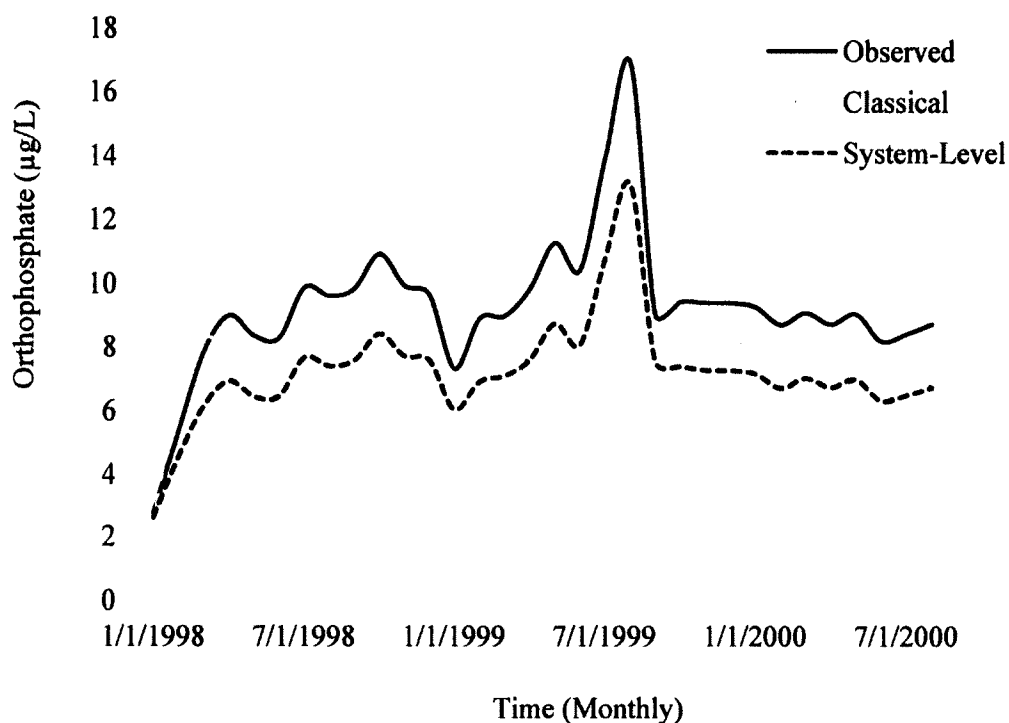
Figure 4-8: Orthophosphate Prediction for Segment 2

Like segment 1, the system-level approach produced strong consistency in the trend and showed the same patterns for localized spikes and depression values similar to the observed. This trend is evident in months like October 1998, September 1999, and June 2000. In general, the system-level is able to capture and reproduce system trend and variation in both peaks and lows. Statistical evaluation for classical and system-level model performance including the p-value, Nash and Sutcliffe efficiency (Equation 4.1) and prediction error are summarized in Table 4-3.

Table 4-3: Statistical Validation for Predicted Orthophosphate

Parameter	Classical	System-level
SEGMENT1		
NSE	-0.182	0.989
Prediction Error (%)	32	-7
p-value	0.02	0.53
SEGMENT2		
NSE	0.33	0.64
Prediction Error (%)	-33	-21
p-value	0.04	0.052

In segment 1, the system-level approach has a Nash and Sutcliffe efficiency of 0.989, under predicted by 7% and is significant at a p-value of 0.53. The classical approach, however, has a low Nash and Sutcliffe efficiency value of -0.182, over estimated by 32% and is has insignificant at a p-value of 0.02. Similarly in segment 2, the system-level approach has a Nash and Sutcliffe efficiency value of 0.64, underestimated by 21% and is significant at a p-value of 0.052. The classical approach has a Nash and Sutcliffe efficiency value of 0.33, over estimates by 33% and is insignificant at a p-value of 0.04. Nash and Sutcliffe efficiency values range from minus infinity to 1; values close to one indicate a better fit (Nash and Sutcliffe, 1970). The results show that the system-level approach is superior to the classical method for predicting orthophosphate in the watershed.

4.6.3 Total Suspended Solids

Results presented here are based on a two and half years of observed data (01/1998 to 08/2000) as data from 08/2000 to 12/2001 were missing. The observed total suspended

(TSS) for segment 1 for the two and half year period ranged from a low of 49 mg/L in June 1999 to maximum peak of 1120 mg/L in August1999 (Figure 4-9).
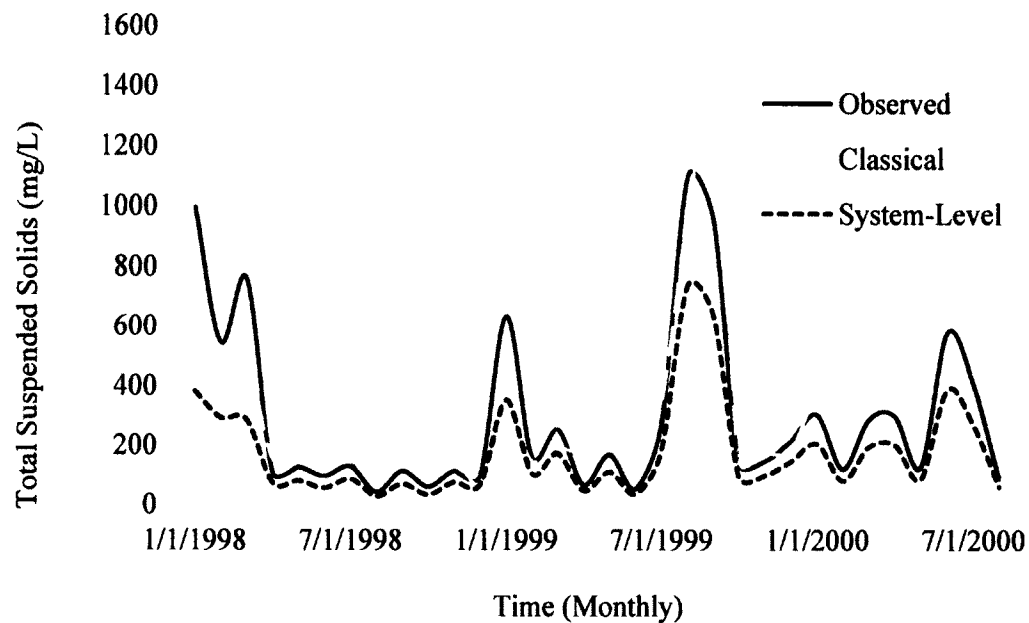
Figure 4-9: Total Suspended Solids Prediction for Segment 1

Evidence from the Figure 4-9 indicates that the observed TSS is generally nominal but shows seasonal variations throughout the year. Both the system-level and the classical approach predicted a similar pattern as that of the observed TSS. The system-level consistently under predicted the TSS concentration through the simulation period while the classical method over predicted for most months. For example, in July 2000, the observed was 590 mg/L, the system-level approach was 400 mg/L while that for the classical was 760 mg/L. Similarly, for in September 1998, the observed was 110mg/L, the system-level approach was 60 mg/L while that for the classical approach was 220 mg/L. Also, in September, 1999, the observed 1000 mg/L, the system-level approach was 750 mg/L while that for the classical 1375 mg/L. Although, the system-level consistently under-predicted, it is able to capture all the system patterns of peaks and lows and other

variations shown by the observed TSS (Figure 4-9) better than the classical approach. Evidence can be seen from April to July 1999.

Segment 2 follows a similar trend of prediction observed in the estimates for segment 1 (Figure 4-9). The observed TSS for segment 2 ranged from a low of 25 mg/L in October 1998 to high peak of 1350 mg/L in August 1999 (Figure 4-10).
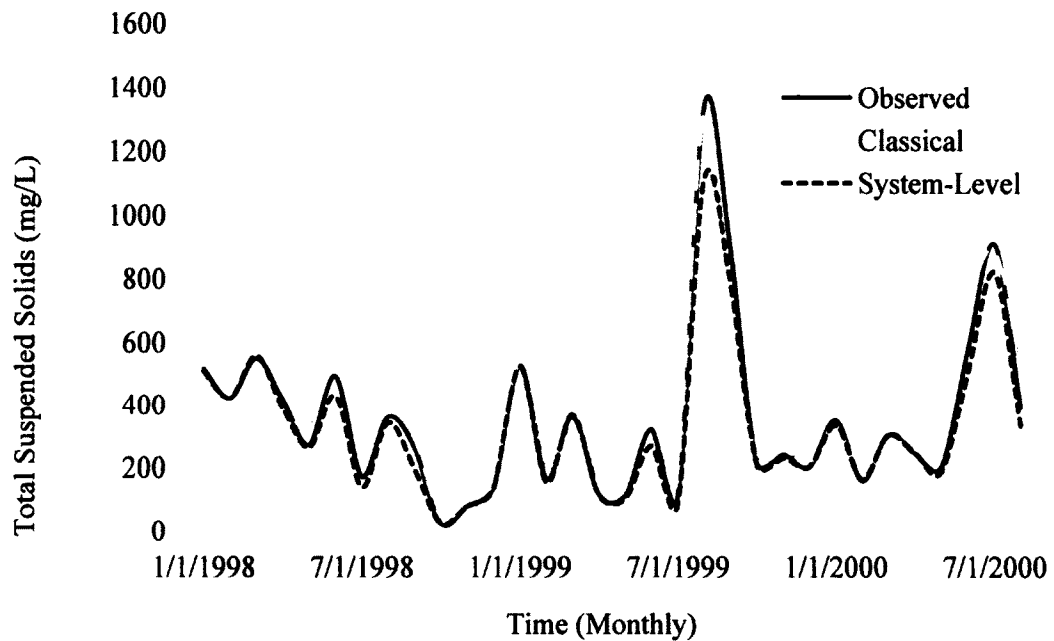
Figure 4-10: Total Suspended Solids Prediction for Segment 2

Unlike segment 1, the system-level estimated close to the observed in most months throughout the year. For example, in January 1999, the system-level and classical simulated values coincided with the observed values. In other months like September 1999, the system-level approach under predicted (observed value = 1380 mg/L, system-level =1150 mg/L, classical approach = 1378 mg/L). Interestingly, the character trait of prediction exhibited by the system-level method (under-estimates and precise-estimates) is also detected with the classical method as it over predicted and predicted well in most

months throughout the simulation period. Hence, Figure 4-10 alone does not provide good evidence to conclude that both conventional and system-level approaches are robust in case of predicting TSS characteristics. Comparative statistics for TSS including the p-value, Nash and Sutcliffe efficiency and prediction error were used to investigate the performance of the two methods (Table 4-4).

Table 4-4: Statistical Validation for Predicted Total Suspended Solids

Parameter	Classical	System-level
SEGMENT1		
NSE	0.65	0.77
Prediction Error (%)	2	-40
p-value	0.051	0.15
SEGMENT2		
NSE	0.95	0.96
Prediction Error (%)	5	9
p-value	0.69	0.59

Evidence in Table 4-4 shows that the system-level approach consistently underestimated while the classical method overestimated for TSS. The p-values for the system-level approach for all the segments are significant (0.15 and 0.59) at the level of significance ($\alpha = 0.05$) while that for the classical method are significant for segment 1 (0.043) and significant for segment 2 (0.69). The Nash and Sutcliffe efficiency values are marginally different for both methods (Table 4-4); however, the values to the system-level approach are consistently higher that for the classical method. The results show that the system-level approach is robust that the classical method as the approach showed consistency in prediction throughout the watershed

4.6.4 Macro-Level Model Sensitivity

Macro-level stochastic model sensitivity for the system-level approach was analyzed to evaluate the magnitude of changes in system response and reproducibility upon posterior variation of additional model parameters. Parameters considered interception and evapotranspiration are known to have significant impact on overland flow. The general definition for sensitivity is described in Chapter 3 (section 3.3).

The model level sensitivity for the system-level approach is presented for overland flow for model segments in Figures 4-11 and 4-12. The figures show that the sensitivity of the overland flow increases with increasing parameter values. The sensitivity to interception varies between -1 and 1 while that for evapotranspiration is between -5 and 2 for segment 1 and -9 and 2 for segment 2. Evidence from both Figures 4-11 and 4-12 shows that a mild gradient in interception does not incur a significant change in the simulated overland flow. The sensitivity varied between -1 and 1 for both model segments. However, a smaller increment in evapotranspiration could cause significant change in the resulting runoff.
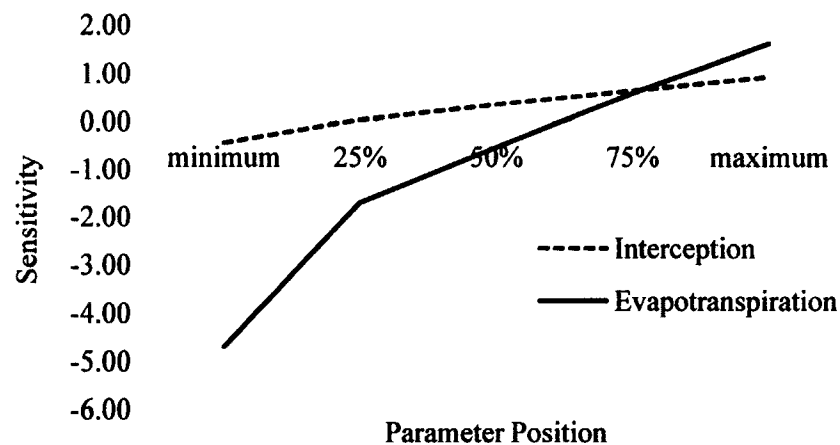
Figure 4-11: Model Level Sensitivity for Overland Flow in Segment 1
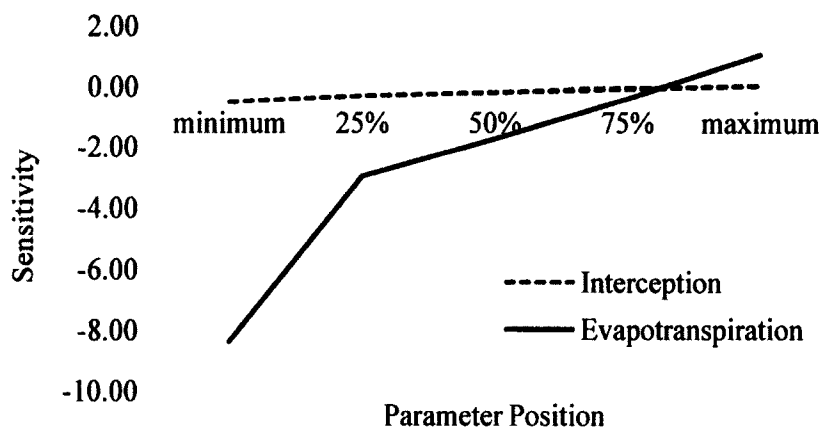
Figure 4-12: Model Level Sensitivity for Overland Flow in Segment 2

The sensitivity of orthophosphate to changes in evapotranspiration and interception is presented in Figures 4-13 and 4-14 for segment 1 and segment 2 respectively.
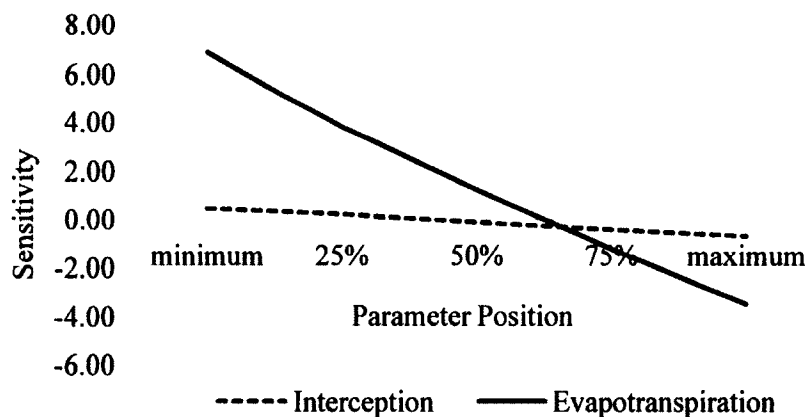
Figure 4-13: Model Level Sensitivity for Orthophosphate in Segment 1
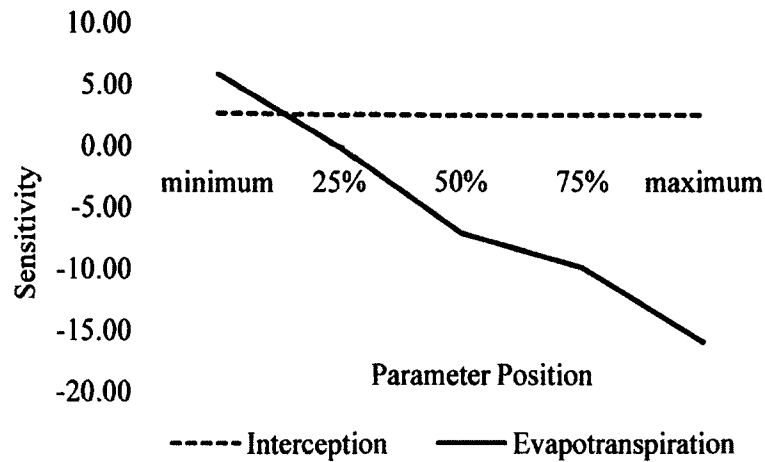
Figure 4-14: Model Level Sensitivity for Orthophosphate in Segment 2

The Figures show the sensitivity of the simulated orthophosphate relates inversely to increasing parameter values. The behavior of orthophosphate is similar the sensitivity of overland flow. For example a small increase in interception does not incur a substantial change in the simulated orthophosphate. However, a minor rise in evapotranspiration could cause significant change in the resulting orthophosphate. This trend can be seen in the total suspended solids in Figures 4-15 and 4-16 for segment 1 and segment 2 respectively.
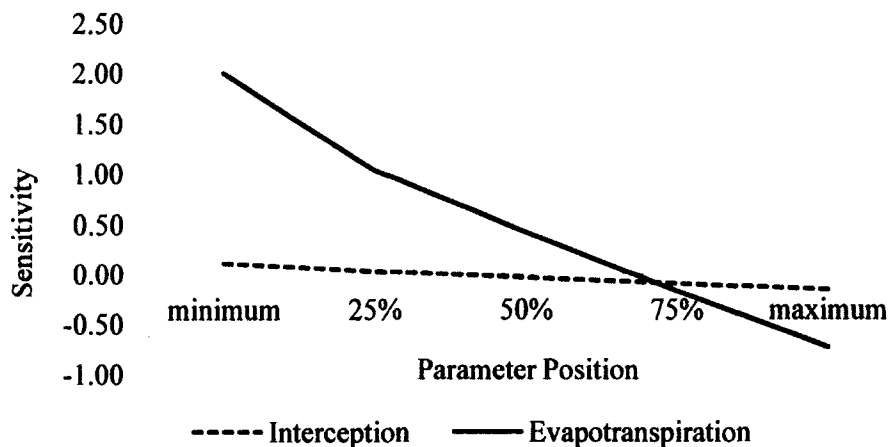
Figure 4-15: Model Level Sensitivity for Total Suspended Solids in Segment 1
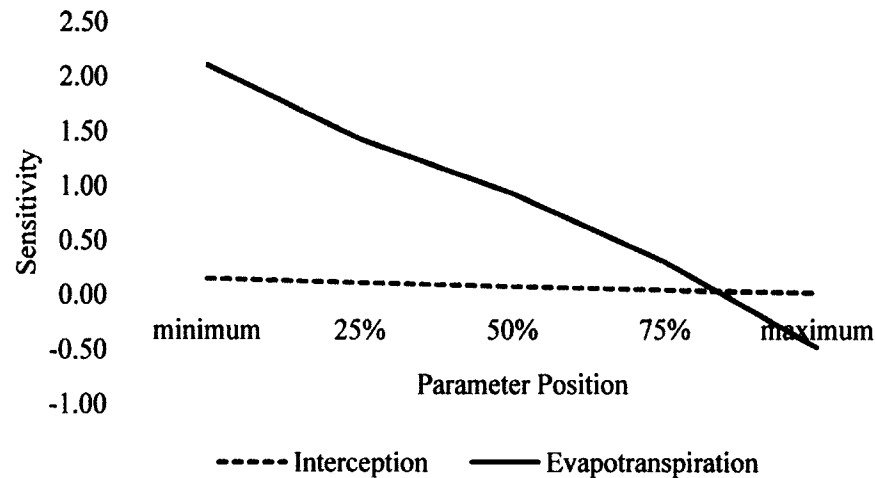
Figure 4-16: Model Level Sensitivity for Total Suspended Solids in Segment 2

The data indicate that overflow, orthophosphate and TSS prediction the sensitivity response to evapotranspiration is more acute than that of interception parameter.

4.7 Discussion

The key objective for this chapter is to implement and verify the methodology framework in the determination of overland and nonpoint source for the Patuxent watershed. The system-level approach was successfully implemented and verified. Overall, the system-level approach proved to be a better estimator of the hydrologic response of the watershed than the classical method.

4.7.1 Overland Flow

The system-level approach predicted the overland flow for both model segments with accuracy and significantly reduced the errors associated with the classical model. The prediction error ranged from -8.3% to 28.3% for segment 1 and -2.34% to 26.01% for segment 2 compared to the classical method which ranged from -17.7% to 68.6% for segment 1 and -27.9% to 59% for segment 2. The highest prediction error for both model

segments was observed during the summer months. This is attributed to the temporal and ephemeral characteristics inherent in the watershed. The p-values for segment 1 were 0.68 and 0.11 for the system-level and the classical approach respectively while that for segment 2 were 0.45 and 0.03 for the system-level and the classical approach correspondingly. The Nash and Sutcliffe efficiency values for segment 1 were 0.33 and 0.12 the system-level and the classical approach respectively while that for segment 2 were 0.25 and 0.15 for the system-level and the classical approach, correspondingly. The significance of the overland flow at p-values greater the level of significance ($\alpha = 0.05$) indicate that the system-level approach is better to predicting overland flow in the Patuxent watershed.

Such significant improvement in prediction accuracy is attributed to the approach employed by system-level method, the hybrid stochastic-deterministic modeling framework, to estimate hydrologic input parameters that truly represent the watershed at its system level. The hybrid stochastic-deterministic modeling framework parameterizes input data from the system's own inherent characteristics that also reduce uncertainty in the prediction processes. The findings (Table 4-2) support and confirm that the system-level method truly accounts for uncertainty and reduces prediction error as evident in the overland flow estimate results.

4.7.2 Nonpoint Source

Overall, the system-level approach successfully predicted orthophosphate concentration for both model segments representing the watershed as illustrated in Figures 4-7, 4-8 and Table 4-3. In comparison with the observed data, the system-level approach proved to be a better estimator for orthophosphate concentration in the

watershed than the classical method. In model segment 1 the prediction error for the system-level method was -7% significant at a p-value of 0.53 while that for the classical method was 32% with a p-vale of 0.02. The system-level approach also showed a higher Nash and Sutcliffe efficiency value (0.989) than the classical method (-0.18). The prediction error in segment 2 was -21% significant at a p-value of 0.052 and -33% with a p-value of 0.04 for the system-level and the classical approach respectively. Segment 2 also showed higher Nash and Sutcliffe values for the system-level approach (0.64) and a lower value for the classical approach (0.33). Values of Nash and Sutcliffe efficiency range from minus infinity to 1; values close to one indicate a good fit between the observed and the model while negative values show a very poor model (Nash and Sutcliffe, 1970). The higher values for the Nash and Sutcliffe efficiency confirm that the system-level approach is a better predictor of orthophosphate for the watershed than the classical approach. Evidence from Figures 4-7 and 4-8 confirms that the system-level approach is more suitable to reproduce the patterns in amplitude of the observed data than the classical method. The study attributes this finding to the inherent abilities of the systems-level approach to estimating input data that represents true system behavior rather than using moments from sample-level variations often associated with classical approach. Another plausible attribution could relate to the prediction accuracy of overland flow. Luo et al (2006) and Jiang *et al.* (2010) pointed out that the amount of nonpoint source generated directly relates to magnitude of surface runoff. Hence better prediction of overland flow could lead to improved prediction of nonpoint sources.

System-level approach performed in a robust manner for estimating total suspended solids (TSS). Table 4-4 shows that the Nash and Sutcliffe efficiency for the system-level

method (0.77) was marginally higher than the classical approach (0.65) for model segment 1 but for segment 2, the values for both methods, system-level (0.96) and classical method (0.95), were practically equivalent. However, comparing the performance of the classical method to the system-level approach, the system-level is significant at p-values greater that the level of significance ($\alpha = 0.05$) and is consistent throughout the watershed. The consistency in the p-values and the Nash and Sutcliffe efficiency values confirms the system-level method as a robust for the prediction of total suspended solids in the watershed

In general, the system-level approach exhibited a robust reproducibility for system's inherent characteristics in estimating the nonpoint source pollutants. The system-level approach estimated the watershed's response to orthophosphate (as phosphorus) concentration very closely and for total suspended solids.

4.7.3 Macro–Level Model Sensitivity

Model level sensitivity of overland flow, TSS and orthophosphate is presented for model segment 1 and model segment 2 for the hydrologic input parameters, interception and evapotranspiration. In both segment, small changes in interception parameter would not have any significant change on the resulting flow, orthophosphate and TSS. However, a slight increase in the evapotranspiration parameter would cause significant change in the simulated overland flow, orthophosphate and total suspended solids. It is recommended to include these parameters in hydrologic input parameterization prior to hydrologic modeling in future system-level approach applications.

4.8 Summary and Conclusion

The chapter successfully implemented and verified the methodology framework presented in Chapter 3. The methodology was applied to the Patuxent watershed to reproduce the hydrologic response of the watershed in terms of overland flow and nonpoint source (phosphorus and total suspended solids).

The results for the overland flow, phosphorus and total suspended solids were compared to that of the classical method (current method) to assess the prediction accuracy of the system-level approach. The findings of the study show that the system-level method significantly reduces prediction error and truly replicates the hydrologic response of the watershed for overland flow and orthophosphate (as phosphorus) and total suspended solids. Based on the findings, the study deduces that the system-level approach is capable of reproducing system's responses for flow and nonpoint source pollutants. The study draws on the fact that evapotranspiration is very sensitive to the system-level input data processing and should be included in future system-level processing prior to hydrologic modeling.

The system-level approach is accurate in predicting the hydrologic response of the watershed. This method can be easily replicated in other watersheds. The study shows that the proposed hybrid methodology framework reduces uncertainties inherent in input datasets and improves prediction by using system's own intrinsic characteristics. Ultimately, the study concludes that the methodology framework can be used to parameterized input datasets to account for uncertainty and to reduce error associated with hydrologic predictions.

CHAPTER 5

NONPOINT SOURCE SUPERPOSITIONING

5.1 Introduction

An essential phenomenon underlying the total response of a hydrologic domain is the connectivity and contribution of every spatial unit located within the watershed. Omission of a spatial unit as a result of inherent data inaccuracy could lead to erroneous prediction of a watershed true response to hydrologic processes. Therefore including every spatial unit of a watershed in the modeling process involves the identification and consistent classification of model boundaries. This ensures that the entire spatial extent of a hydrologic domain is truly represented.

Delineation of a watershed is based on topographic gradients which define the direction of flow. During delineation, digital terrain analysis algorithms often require minimal extent of topographical heterogeneity (Krause and Bronstert, 2005). Where topographic gradients are too small, a topographical discrepancy may occur. This may lead to hydrologic disconnection of spatial units that are located in the topographic inconsistent areas. The resolution of the elevation data is a key factor to determining minimum gradient (Wise, 2000). Thus higher resolutions are associated with better delineation accuracy and vice versa. This phenomenon is very common with floodplains, low lying environment and flat areas where topography is almost homogeneous (Krause and Bronstert, 2005). The effect is more pronounced on basin-wide scale as such areas require high resolution data. A common resulting consequence of this delineation

anomaly includes misrepresentation or misalignment of flow paths and spatial extent of the delineated boundaries.

Hydrologic models respond only to spatial units that are hydrologically connected as they are designed to follow the path of water flow. This phenomenon could result in the approximation of total response to nonpoint sources of the watershed being represented by the model. Current approach to hydrologic modeling overlooks possible existence of topographical discrepancies and assumes delineation is accurate. Although such an oversight and assumption could simplify the modeling process, the outcome of the simulation could be heavily plagued with error if the delineated watershed contains multiple counts of the aforementioned discrepancies.

This chapter proposes a framework that accounts for nonpoint sources contribution from the hydrologically disconnected spatial units of the watershed. The framework would help to accurately determine the total response of a watershed to pollution arising from diffused sources. The goal is to conceptualize a methodology that incorporates nonpoint source into the proposed hybrid hydrologic modeling framework by using method of superposition over time and space. The chapter also examines the response of nonpoint (diffused) sources contributed from the hydrologically disconnected areas to storm events. The intent is to define criteria on when to ignore/include nonpoint source contribution from the hydrologically disconnected areas. The realization of the above goals include to (1) develop and modify existing mathematical expressions for estimating overland diffused sources, (2) identify hydrologically disconnected areas (disconnected areas) within the watershed, (3) estimate nonpoint source contribution from the disconnected areas and superimpose on the nonpoint source estimated from the system-

level approach in Chapter 4, and (5) characterize the response of the nonpoint source from disconnected areas to storm events.

5.2 Mathematical Formulation

Mass loading relating to nonpoint sources varies with respect to location and time; hence, are classified as continuous or instantaneous depending on the duration. Instantaneous loadings are over a very short time period (Thomann and Mueller, 1987). Continuous loadings introduce pollutants to a receiving body of water for extended period of time (Runkel and Bencala, 1995). In this study the concept of continuous loading is extended to the characteristics of estimating nonpoint source contributions coming from the disconnected area. The analogy is based on the ephemeral features and the extended period of loading during the duration of overland flow. The concept of pollutant mass loading, overland flow routing and advection and diffusion are combined to derive mathematical equations that estimate the amount of nonpoint source generated from hydrologically disconnected areas. Details of the interrelationship between this parameters and their derivation from first principles are presented in appendix A. In general, pollution mass loading is the product of the flow rate and the concentration of the pollutant. It is expressed mathematically as:

$$W = QC \qquad (5.1)$$

where W mass loading, Q is the flow rate and C is the concentration of the effluent. For overland nonpoint (diffused) sources, Q is the overland flow rate and C is the concentration resulting from the washoff of pollutants during runoff.

Consider the elemental volume in Figure 5-1 as a catchment in a watershed. The mass of pollutant loading leaving and entering can be described using the conservation of

mass or mass balance. The mass balance accounts for all fluxes entering and leaving the control volume.
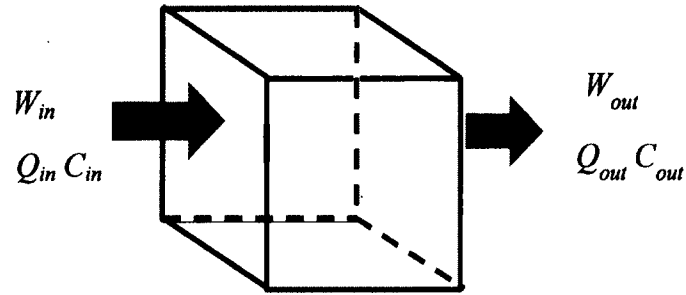
Figure 5-1: Mass Balance for an Elemental Volume

$$\Delta QC = Q_{in}C_{in} - Q_{out}C_{out} \qquad (5.2)$$

The classical advection-dispersion for defining the transport and fate of pollutant in Figure 5-1 can be described by Equation 5.3.

$$\frac{\partial C}{\partial t} = -U\frac{\partial C}{\partial x} + E\frac{\partial^2 C}{\partial x^2} - kC \qquad (5.3)$$

Equation 5.3 is the classical advection-dispersion equation used in quality modeling. Equation 5.3 is often used for in-stream pollutant routing where the variable C is the concentration for point sources. Nonpoint sources pollutants are subsequently added as contributing sources. In this context, Equation 5.3 is adapted to estimate nonpoint source pollutants overland. The Equation describes the spatial and temporal variation of overland nonpoint sources pollution transport over time and space. In this study, the durational characteristic of continuous loading is extended to the temporal characteristics of surface runoff. For a storm duration (τ) overland flow begins at time, $t = 0$ and ends at time $t > \tau$. Consequently, the mass loading begins at time, $t = 0$ and ends at time $t > \tau$. The solution to Equation 5.3 for time variable nonpoint source loading can be obtained by using the boundary conditions:

$$C(x,0) = 0 \quad for \quad x \geq 0$$

$$C(0,t) = C_0 \quad for \quad \tau \geq t \geq 0$$

$$C(x,t) = 0 \quad for \quad t \geq \tau$$

$$C(\infty,t) = 0 \quad for \quad t > 0$$

The solution to Equation 5.3 in the final form is given by O'Loughlin and Bowmer (1975) and Runkel (1996).

O'Loughlin and Bowmer (1975):

$$C(x,t) = \frac{C_0}{2}\left[e^{\frac{Ux}{2E}(1-\Gamma)}efrc\left(\frac{x - Ut\Gamma}{2\sqrt{Et}}\right) + e^{\frac{Ux}{2E}(1-\Gamma)}efrc\left(\frac{x + Ut\Gamma}{2\sqrt{Et}}\right)\right] \qquad (5.4)$$

Runkel (1996):

$$C(x,t) = \frac{C_0}{2}\left\{e^{\frac{Ux}{2E}(1-\Gamma)}\left[efrc\left(\frac{x - Ut\Gamma}{2\sqrt{Et}}\right) - efrc\left(\frac{x - U(t-\tau)\Gamma}{2\sqrt{E(t-\tau)}}\right)\right]\right.$$

$$\left. + e^{\frac{Ux}{2E}(1+\Gamma)}\left[efrc\left(\frac{x + Ut\Gamma}{2\sqrt{Et}}\right) - efrc\left(\frac{x + U(t-\tau)\Gamma}{2\sqrt{E(t-\tau)}}\right)\right]\right\} \qquad (5.5)$$

$$\Gamma = \sqrt{1 + 4\frac{kE}{U^2}}$$

$$C_0 = \frac{W}{Q} \qquad (5.6)$$

where C is the concentration of the pollutant, x is distance traveled, τ is the rainfall duration, t is the time, C_0 is the initial concentration, U is the advection coefficient (mean velocity) and E is the longitudinal dispersion coefficient. The solution during the rainfall, $t < \tau$ is given in Equation 5.4 and that after the rainfall, $t > \tau$ is given by Equation 5.5. A combination of Equations 5.4 and 5.5 superpositions nonpoint loading over time and space during and after a rainfall event.

5.3 Methods and Procedures

The equations 5.1 – 5.6 are applied in this section to estimate mass load generated during overland flow from areas that are hydrologically disconnected from the modeled watershed.

5.3.1 Hydrologically Disconnected Areas

This section describes the process of identifying areas that are hydrologically disconnected in the hydrologic model. The delineated watershed in Chapter 4 was examined to identify the areas that are hydrologically disconnected as a result of topographical discrepancy. Two of such areas were identified in the watershed. The landuse types associated with these areas were identified to be emergent herbaceous wetlands or marshes (Figure 5.2).
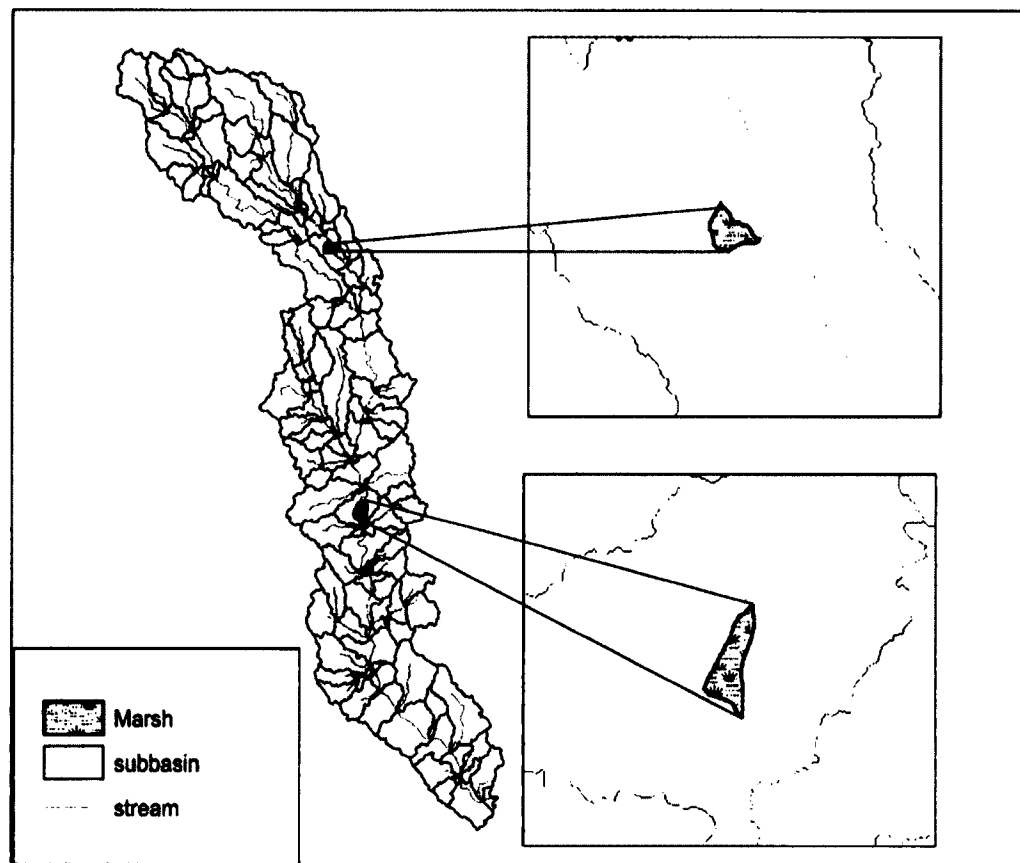
Figure 5-2: Hydrologically Disconnected Areas

Geographic Information System techniques were used for the identification process. The study performed an overlay analysis using multiple layers consisting of hillshade, slope, aspect, subbasins, streams and NHD stream to locate areas within the watershed that have discontinuous flow paths. The hillshade served as a terrain or relief feature, the slope was used to identify areas with successive low gradient towards downstream while aspect was used to detect the pattern of flow.

5.3.2 Estimation of Nonpoint Source from Disconnected Areas

This section describes the process of applying the equations 5.4 and 5.5 to estimate amount of mass load generated during overland flow for the disconnected areas identified in section 5.3.1. These areas will be superpositioned onto the system-level approach to

predict the overall response to nonpoint source in the watershed. Equation 5.5 was used to determine the concentration during the storm while Equation 5.4 was used to compute the lingering concentration after the storm. The equations require inputs such as initial concentration, duration of rainfall, simulation time period, decay rate, longitudinal dispersion coefficient, mean velocity and distance traveled by the diffused source. Input variables such as the length, area, slope and surface roughness was determined using GIS tools. The initial concentration was obtained by using published areal loading (mass per unit area) for phosphorus for the Patuxent (Boynton *et al.*, 1995) and the Equation 5.6. The volumetric flow (Q) rate in Equation 5.6 was determined using the rational method:

$$Q = \alpha i A \tag{5.7}$$

where α is the runoff coefficient, i is the rainfall intensity (precipitation) and A is the surficial of the marsh. The runoff coefficient for the marsh areas was obtained by using a relationship between runoff coefficient and imperviousness (Schueler and Yousef, 1994, Wright *et al.*, 2006). A six day precipitation event was obtained from a nearby station in the watershed. The precipitation days were selected such that they have zero precipitation days preceding and succeeding the rainfall events. Other required inputs such as reaction rate constant for phosphorus was obtained from Bowie *et al.* (1985) while dispersion coefficient for wetlands was determined from the Kadlec (1994). The dispersion coefficient was expressed as:

$$\frac{E}{UL} = 0.264 \tag{5.8}$$

where E is longitudinal dispersion coefficient; L is length of wetland and U is velocity. Velocity was calculated by using Mannings equation and assuming that the overland flow

is a rectangular channel with infinite width and shallow depth (Chapra, 1997; Akan and

Houghtalen, 2003; Ponce, 1989).

$$U = \frac{1}{n} R^{2/3} \sqrt{S_o}$$

$$R = \frac{A_c}{P} = \frac{b.h}{b + 2h}$$

As $b \to \infty$

$$R \cong h$$

$$U = \frac{1}{n} h^{2/3} \sqrt{S_o} \qquad (5.9)$$

where U is the velocity, R is the hydraulic radius, n is the Mannings coefficient, S_o is the

slope of the marsh, A_c is the cross-sectional area, P is the wetted perimeter, b is the flow

width and h is the flow depth. The assumptions governing the computation process

described above include: (1) the landuses found within the disconnected areas are

homogeneous; (2) the rainfall duration is equal to the time of concentration resulting in

peak discharge, and (3) internal reactions within the marsh are not considered.

Assumptions 1 and 2 meet the criteria for using the rational method

5.3.3 Nonpoint Source Superpositioning

The temporal variation of phosphorus concentration at the exit of the marsh was

determined for simulation ten days with precipitation beginning on day 2 and ending on

day 8 (Figure 5-3). Day 1 and days 8-10 had no precipitation while days 2 to 7 had

precipitation continuously. The computed concentration (Equations 5.4 and 5.5) at the

outlet of the disconnected area was superpositioned onto the system-level model in

Chapter 4. Details of the methodology framework for nonpoint source superpositioning can be found in appendix B.
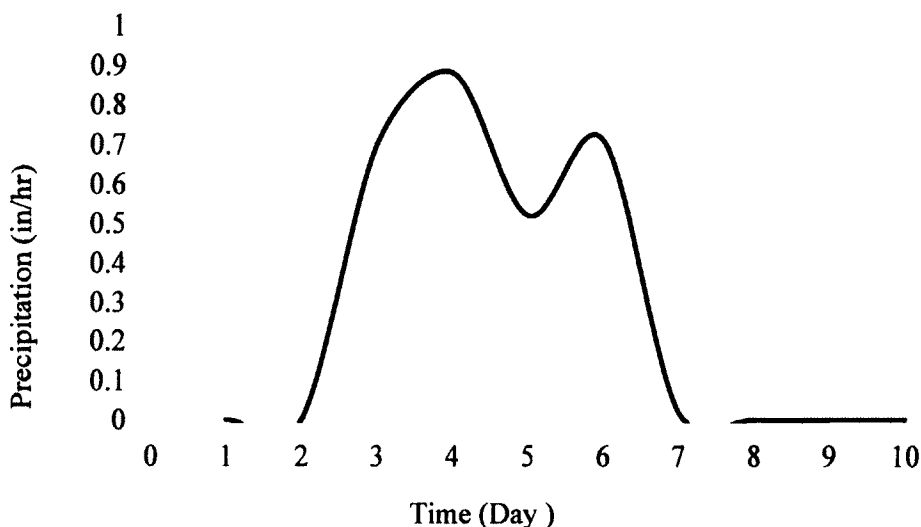
Figure 5-3: A Six-Day Precipitation Event

5.3.4 Nonpoint Source Characterization

Nonpoint sources are known to exhibit spatial and temporal characteristics (Thomann and Mueller, 1987; Fujiwara et al., 1988). Jiang et al. (2010) and Luo et al. (2006) showed that the amount of nonpoint source generated directly relates to the magnitude of overland flow. This study examines the response of the nonpoint from the hydrologically-disconnected area to different storm events. The characteristics are defined in terms of magnitude and spatial and temporal characteristics.

The rainfall intensity duration frequency (IDF) was used for this purpose. Rainfall intensity duration frequency curves are precipitation patterns used as input into hydrologic models to estimate runoff from a watershed (Chow et al., 1988) or for design purposes in hydrologic engineering (Levy and McCuen, 1999). The storm events include different storms durations (6, 12, 24 and 48 hours) at different return periods (2, 5, 10,

and 25 years). The IDFs were determined using the rainfall frequency atlas map, Technical Paper No. 40 (Hershfield, 1961) and Technical Paper No. 49 (Miller, 1964). The atlas maps are isohyet charts covering different parts of the United States. To determine rainfall intensity, rainfall depths (Figure 5-4) for the study area were determined from the atlas at the specified return periods and storm durations and then converted into rainfall intensities values (Figure 5-5) by dividing the rainfall depth by the corresponding storm duration.
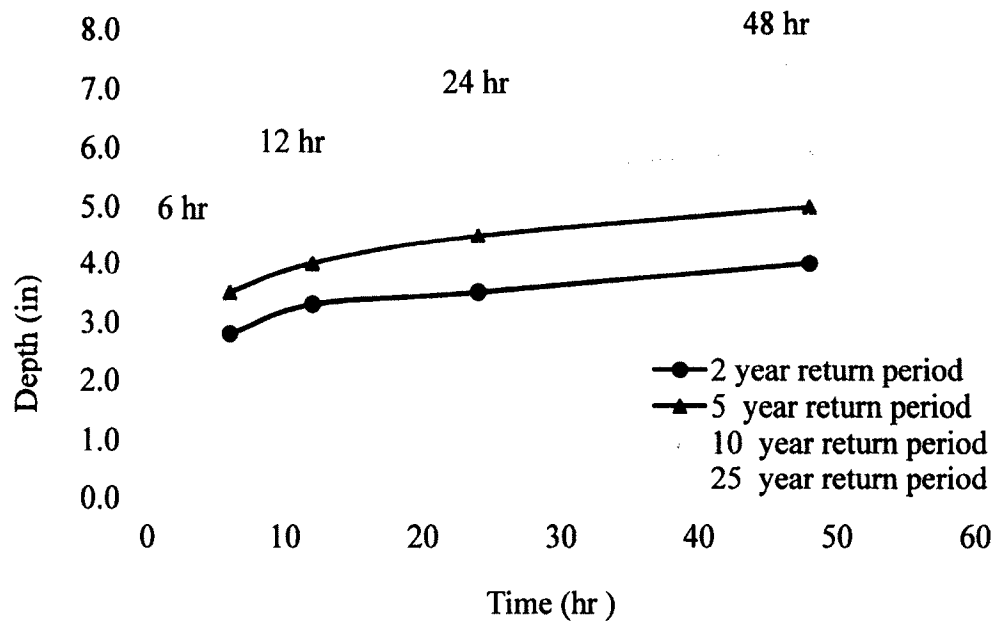
Figure 5-4: Rainfall Depths Extracted from Atlas Maps
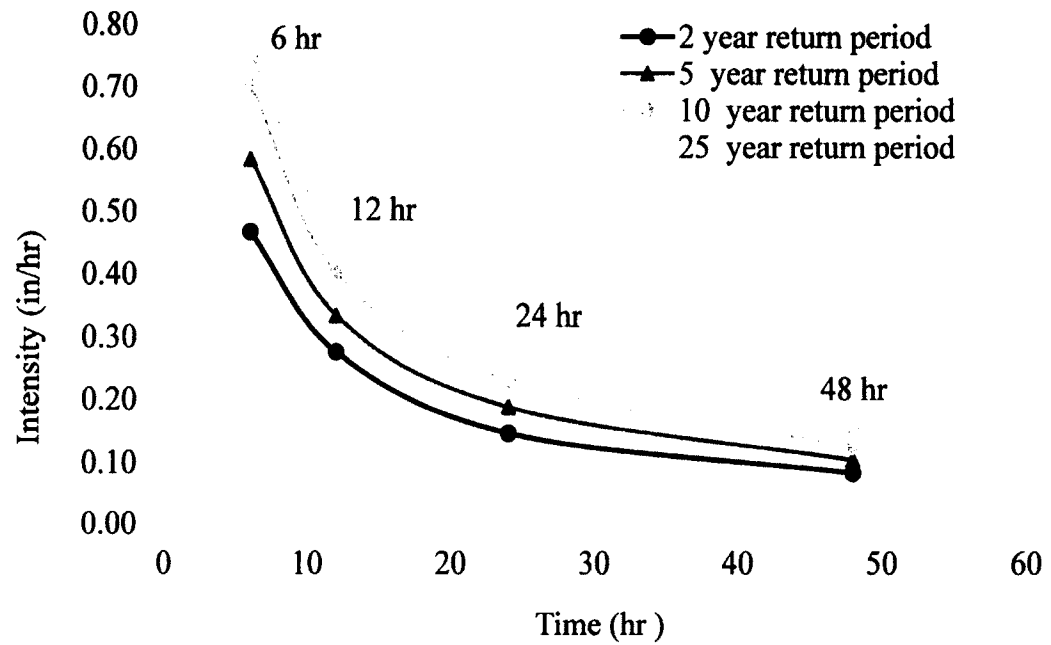
Figure 5-5: Rainfall Intensity Duration Frequency Curve

Characterization of the diffused source in terms of magnitude was defined as overland flow rate, concentration and mass loading. For each return period and storm duration, the corresponding rainfall intensity was used with the rational method (Equation 5.7) to determine the volumetric flow rate. The concentration of phosphorus was then determined using Equation 5.6. Equations 5.4 and 5.5 were used to determine the concentration. The mass loading was determined using Equation 5.1.

The spatial and temporal variation of the diffused source was investigated for 12 hour storm duration for a 2 year return period. To determine the spatial variation, the study divided the volumetric flow rate by the total length of the disconnected area to obtain a unit flow rate (q_o). The length of the marsh was then divided into a n number of segments (S) along the length of the marsh. A schematic representation is shown in Figure 5.6.
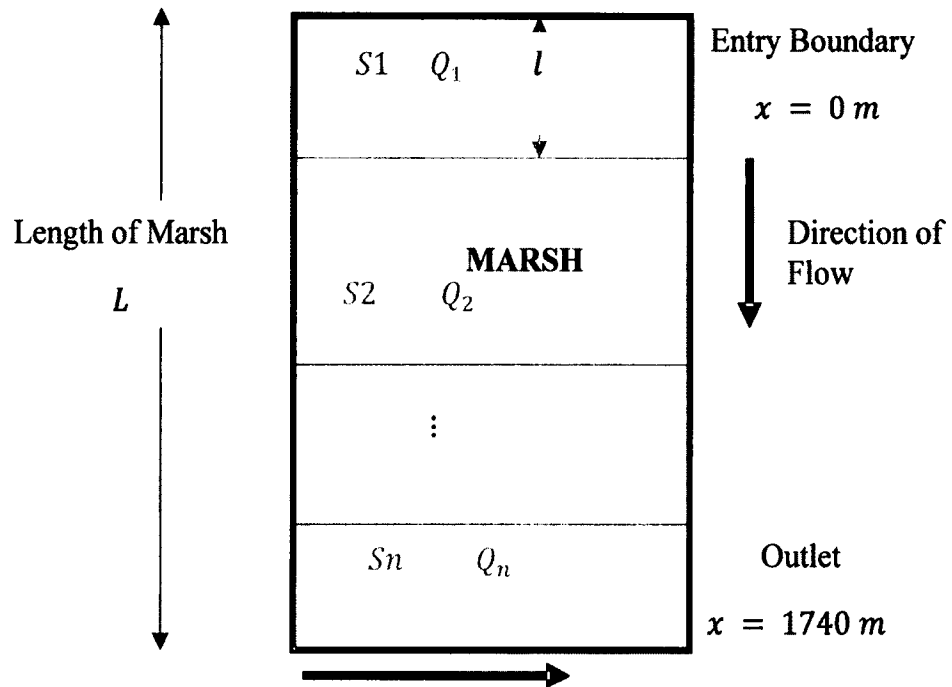
Figure 5-6: Schematic Diagram for Computing Spatial Distribution of Runoff

For each segment, the volumetric flow rate is defined as

$$Q_n = q_0 \, n \, l$$

$$q_0 = \frac{Q}{L}$$

where l is the length from the entry boundary, n is the position of the segment form upstream and q_0 is the unit flow rate, Q_l is the flow rate at a certain length from the entry boundary and Q is the flow rate computed with Equation 5.7. Equations 5.4 and 5.5 are then used to compute the concentration and Equation 5.1 was used to determine the mass load for each segment.

5.4   Model Sensitivity

In this study two different sensitivity analyses were performed: (1) sensitivity of the superpositioned system-level to physiographic related input parameters and (2)

sensitivity of the nonpoint source generated from the disconnected areas to storm events. Absolute sensitivity (Equation 3.20) was used to assess response of the system-level approach with superpositioned nonpoint due to the level of complexity of interconnected equations. Relative sensitivity (Equation 3.21) was used for the nonpoint source generated from the disconnected areas.

5.4.1 Sensitivity of System-Level Superpositioned Model

This section uses sensitivity analysis to gain insight into the dynamics of the system-level approach superpositioned with nonpoint by investigating its response to other physiographic input parameters that were not processed at system-level. The parameters considered were interception and evapotranspiration. To measure the sensitivity, bounds for the parameters mentioned above were selected from their respective range of values represented in the deterministic computation engine. The bounds were set at 25 and 75 percentile and a standard Monte Carlo was used to generate series of input data sets for the analysis. The purpose for selecting bounds at 25 and 75 percentile was to prevent any possible forms of bias that may result from high or low values. The Monte Carlo method considers each model input parameter to be a random variable with a probability density function (PDF) (Loague and Corwin, 1996). The extent of sensitivity was carried out by changing separately each model parameter and measuring the response of the model (Costanza *et al.*, 2002; Arhonditsis *et al.*, 2000).

5.4.2 Nonpoint Source Sensitivity to Storm Events

This section uses relative sensitivity approach to investigate the response of the nonpoint source generated for the disconnected area to various storm events. Studies such

as Jiang *et al.* (2010); Luo *et al.* (2006) have shown that, the generation of nonpoint sources relates directly to the magnitude of overland flow. The study investigates the response of nonpoint source generated from the hydrologically-disconnected areas in the watershed to different rainfall events for different storm durations (6, 12, 24 and 48 hours) at different return periods (2, 5, 10 and 25 years). The sensitivity was determined using basic equations governing the generation of NPS in the disconnected area and expressing them in terms of the sensitivity index:

Flow rate:

$$Q = \alpha i A \tag{5.13}$$

Loading:

$$W = QC \tag{5.14}$$

Intensity:

$$i = \frac{h}{t} \tag{5.15}$$

where Q is the volumetric flow rate obtained from the rational method, A is the surficial area, α is the runoff coefficient, i is the intensity, W is the mass load, C is the concentration, h is the flow depth, and t is the duration. Relative sensitivity with respect to intensity can be expressed using Equation 3.21 as

$$R_{intensity} = \frac{\partial W}{\partial i} \cdot \frac{i}{w} \tag{5.16}$$

Using the equations 5.13 to 5.15

$$\frac{\partial W}{\partial i} = \frac{\partial W}{\partial Q} \cdot \frac{\partial Q}{\partial i}$$

$$\frac{\partial W}{\partial i} = C\alpha A$$

The relative sensitivity with respect to the intensity is described as

$$R_{intensity} = C\alpha A . \frac{i}{w}$$
(5.17)

Relative sensitivity with respect to duration can be expressed using Equation 3.21

as

$$R_{intensity} = \frac{\partial W}{\partial t} . \frac{t}{w}$$
(5.18)

Using the equations 5.13 to 5.15

$$\frac{\partial W}{\partial t} = \frac{\partial W}{\partial Q} . \frac{\partial Q}{\partial i} . \frac{\partial i}{\partial t}$$

$$\frac{\partial W}{\partial t} = C\alpha A . - \frac{h}{t^2}$$

The relative sensitivity with respect to the duration is described as

$$R_{duration} = - \frac{C\alpha Ah}{t^2} . \frac{t}{w}$$
(5.19)

$$R_{duration} = - \frac{C\alpha Ah}{t^2} . \frac{t}{w}$$
(5.20)

5.5 Results

5.5.1 Nonpoint Source Superpositioning

The amount of nonpoint source generated was successfully determined for the disconnected area for ten simulation days for a six-day precipitation event.
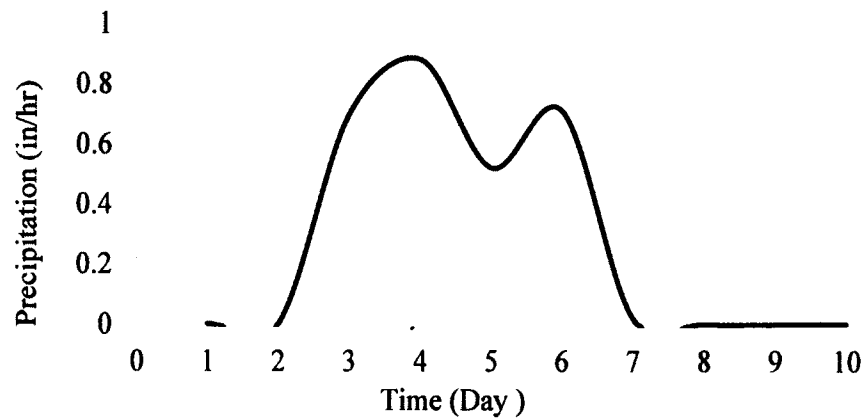
Figure 5-7: A Six-Day Precipitation Event

Loading began on day 2 and ended on day 8 (Figure 5-8) of the precipitation days. Day 2 corresponds with the day precipitation began and day 8 corresponds with a day after the precipitation ends (Figure 5-7).
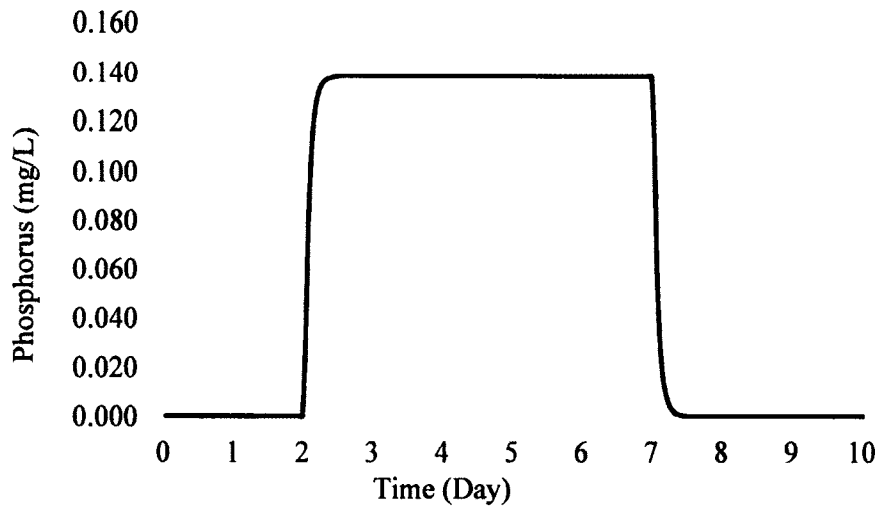
Figure 5-8: A Ten-Day Phosphorus Simulation

Figure 5-8 shows concentration of the phosphorus at different times scales at the outlet of the disconnected area. Evidence from the graph shows the limb of the pollutograph (Figure 5-8) rises with the rising limb of the hyetograph (Figure 5-7) to

peak, remains in equilibrium and then recedes as the hyetograph recedes. Consequently, the concentration increased from 0 mg/L to a peak of 0.14 mg/L on day 2, remained constant throughout the duration of the rainfall and reduces to 0 mg/L at the end of the rainfall. The transport process began on day 2 and ended on day 8, 12 hours after the storm duration. The study attributes the continuous contribution after the storm duration to the continuous outflow of the runoff from the watershed as the transport of the phosphorus depends on the overland flow. Overall, the result demonstrates that nonpoint source exhibits ephemeral characteristics and varies over time and space.

The phosphorus concentration at the outlet of the disconnected area was to superposition the system-level model (from Chapter 4) in order to determine the overall response of the segment to phosphorus loading. Results are shown in Figure 5-9.
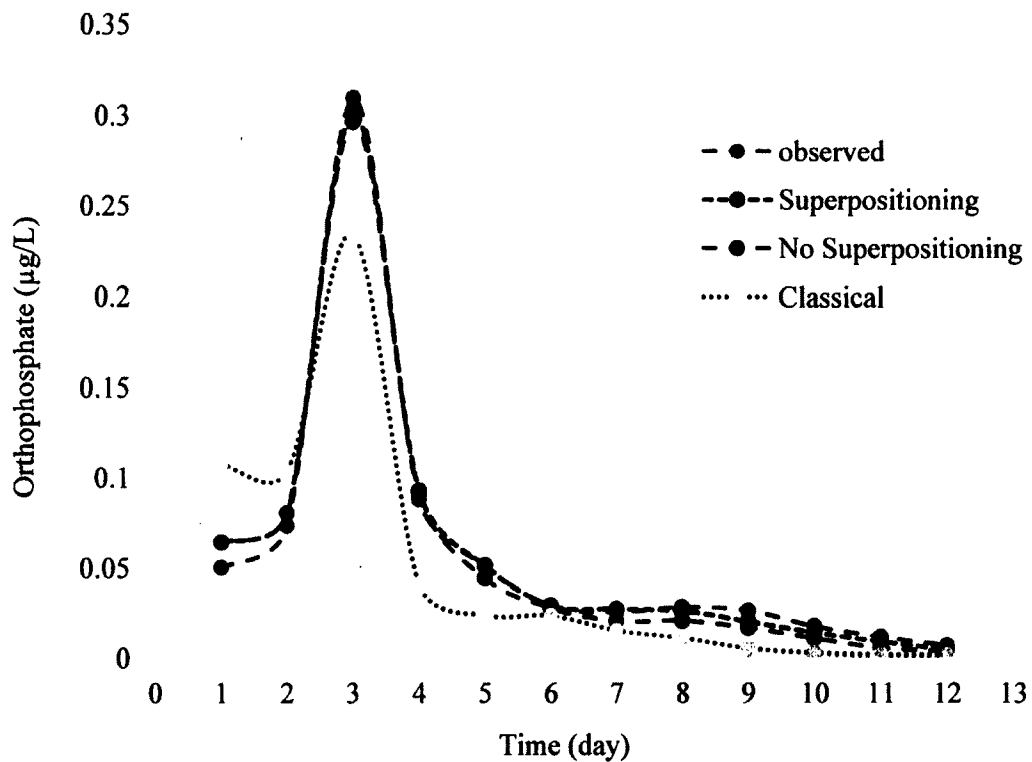
Figure 5-9: Nonpoint Source Superpositioning

Evidence from the graph show that the observed orthophosphate increased from 0.05 µg/L on day 1 to a peak of 0.3 µg/L on day 3 and receded to 0 µg/L on day 11. Similarly, the simulated phosphorus exhibited similar characteristics on the same days. However, the system-level approach without superpositioning (no superpositioning) and with superpositioning (superpositioning) overestimated on days 1 and 2 but underestimated consistently after day 2 onwards. Similar characteristics were shown by the classical approach. For example, on day 8, the observed phosphorus was 0.027 µg/L while the system-level method with super-positioning was 0.027 µg/L, the system-level approach without superpositioning was 0.020 µg/L and that for the classical approach was 0.011 µg/L on the same day. Although the system-level with superpositioning, the system-level without superpositioning and the classical approach appears to follow similar pattern, the system-level with superpositioning and the system-level without superpositioning matches the observed closely better than the classical approach. The results emphasize the conclusion drawn in the preceding chapter (Chapter 4) that the system level approach is a better replicator of system's response than the classical approach.

Figure 5-9 also compared the system-level with superpositioning to system-level without superpositioning. Results show that both methods matched each other from day 1 to day 2 but differed from day 3 to the end of the simulation period. The difference is due to the inclusion of the nonpoint source from the disconnected area which began on day 2 (Figure 5-9). The contribution of the external nonpoint source can be clearly seen on day 3 to 10 (Figure 5-9). Evidence show that the system-level with superpositioning estimates closer to the observed than the system-level without superpositioning. Table 5-1 shows

the prediction error and the Wilcoxon Rank p-value associated with each of the methods presented in Figure 5-9.

Table 5-1: Statistical Validation for Predicted Nonpoint Source

	System-level with Superpositioning	System-level without Superpositioning	Classical
Prediction error (%)	0.29	-4.82	-20.55
p-value	0.954	0.583	0.194

Results in the Table above shows that the error of prediction for system-level with superpositioning approach is 0.29% while that of system-level without superpositioning and classical method are -4.82% and -20.55% respectively. Although, the system-level with superpositioning approach overestimates, the degree of error associated with estimation is less compared with that of the system-level without superpositioning. A Wilcoxon Rank Sum (Mann Whitney) test showed that all the three methods are significant at p-values (Table 5-1) greater than the level of significance (alpha = 0.05) None of the three methods could be rejected as each of the methods are significant at the required level of significance; however, the superpositioned-nonpoint-source-system-level model is superior to estimating system-level approach considering the prediction error and the performance shown in Figure 5-9.

5.5.2 Nonpoint Source Characterization

The nonpoint source characterization in terms of magnitude is presented for storm events at different return periods (2, 5, 10 and 25 years) and durations (6, 12, 24 and 48 hours). The hydrologic response in terms of flow rate is shown below in Figure 5-10.
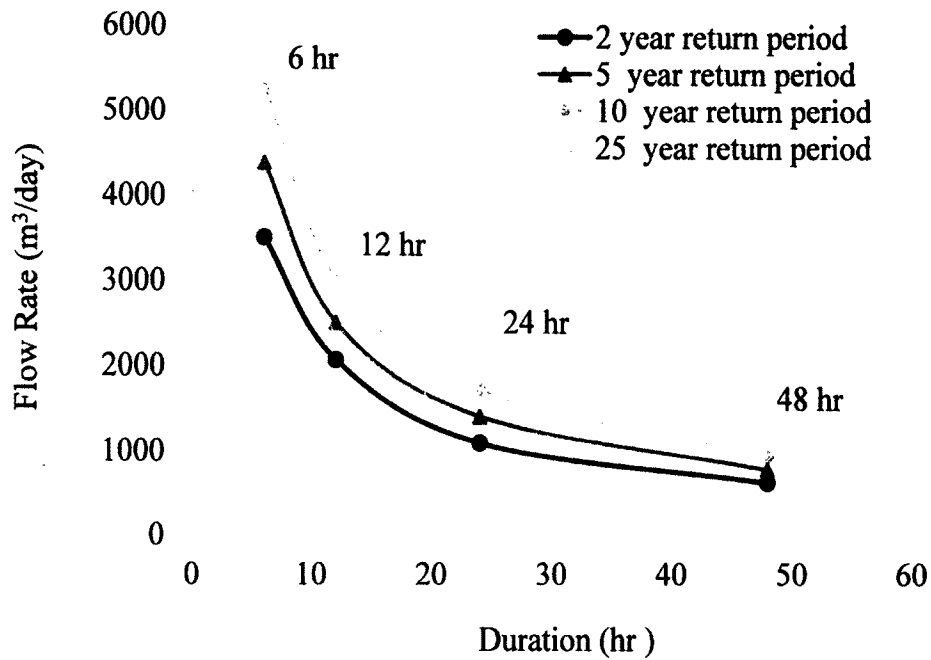
Figure 5-10: Hydrologic Response to Overland Flow

Results from Figure 5-10 shows that the flow rate increases for all return periods with decreasing duration. For the same return period and different durations, flow rate is maximum at lowest duration but reduces to minimum as the duration increases. For example, for a 25 year return period, the flow rate was approximately 6000 m^3/day but as the duration increases the flow rate reduces gradually to 1000 m^3/day in 48 hours. For different return periods but the same duration, the flow rate increases with increasing return period. In general, the lower the storm duration, the higher the flow rate for same return period, and the higher the return period the higher the peak flow rate for the same storm duration. This phenomenon is attributed to the relationship between flow depth, rainfall intensity and duration as storms with long durations have higher flow depths and smaller intensities than those with shorter durations.

The corresponding concentration at different return periods for different durations is present in Figure 5-11. The concentration presented here is the maximum concentration for each return period occurring at the outlet of the marsh.
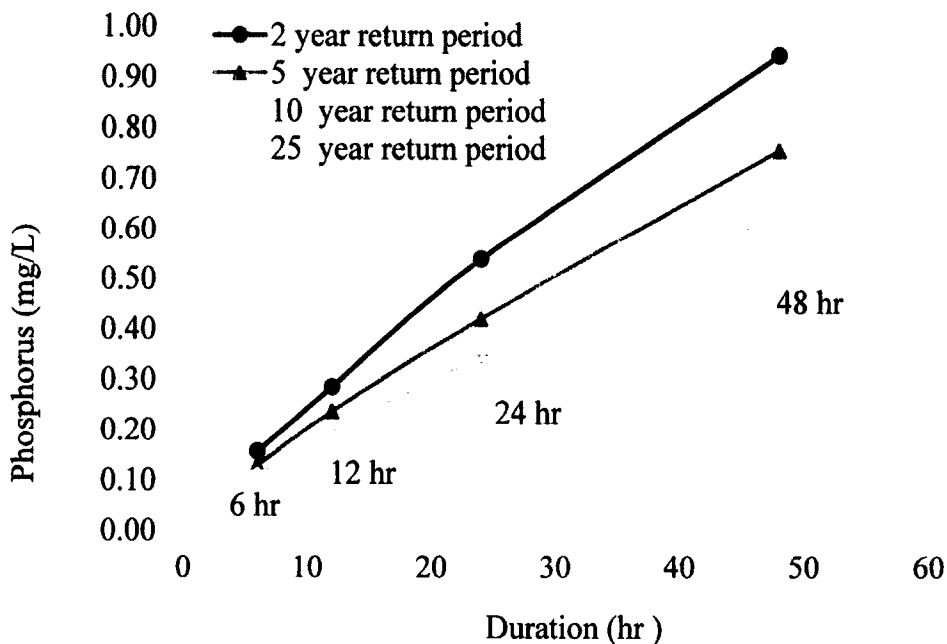
Figure 5-11: Hydrologic Response to Phosphorus Concentration

The results in Figure 5-11 show that, concentration is at its lowest for smaller durations but increases for all return periods as the duration of the storm increases. For example, at a 6-hour duration of rainfall, the concentration was low at 0.1 mg/L and 0.15 mg/L for 25 and 2 year return periods respectively; this increases as the storm duration increases to 48 hours at 0.5 mg/L and 0.9 mg/L for 25 and 2 year respectively. For the same return period but different storm durations, the concentration increases progressively with increase in storm duration. For different return periods but the same duration, concentration reduces with increasing return period; for example, for a 12-hour duration rainfall for different return periods, the concentration recorded was 0.3, 0.25,

0.17 and 0.15 mg/L for a 2, 5, 10 and 25 year return periods respectively. In general, the higher the storm duration, the higher the concentration for the same return periods, and the higher the return period the lower the concentration for the same storm duration.

The mass loading for the hydrologically disconnected area is shown in Figure 5-12.
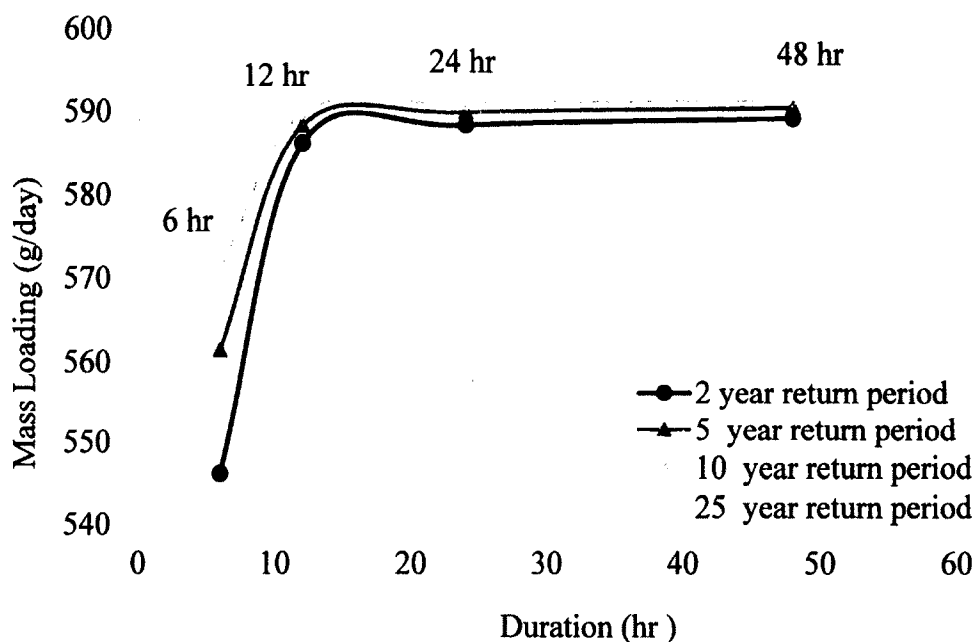
Figure 5-12: Hydrologic Response to Phosphorus Mass Loading

Figure 5-12 shows that the mass load increases for all return periods up to 12 hours and then maintains equilibrium after 12 hours. For example, for a 5-yr return period, the mass load spikes from 560 g/day at a 6 hour duration to 587 g/day at 12 hours and remained constant at 585 g/day even though the duration kept increasing. All the return periods exhibit similar trends. Unlike the concentration, the mass loading increases as the return period increase. For example, for a 12-hour storm duration, the mass load generated was 585, 587, 589, and 590 g/day for 2, 5, 10 and 25 year return periods.

However, as the return period increases, the difference in magnitude between the mass loads produced becomes very insignificant. Also, smaller return periods are quicker to spike at durations up to 12 hours than higher return periods.

The spatial variation of the diffused source along the length of the marsh (location 0 m to 1740 m, Figure 5.6) for a 12- hour duration is presented in Figure 5-13.
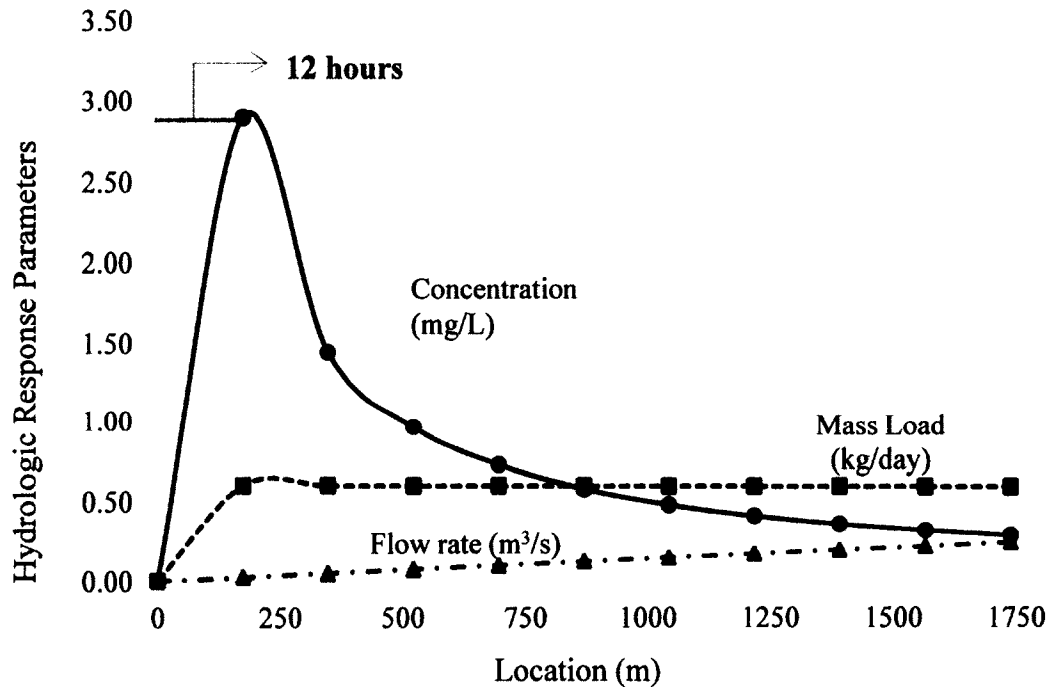
Figure 5-13: Spatial Variation of Diffused Sources

The results show the spatial characteristics of the phosphorus pollutant as it travels from location 0 m (entry boundary of the marsh) to location 1740 m (outlet). The Figure shows the relationship between the flow rate, the concentration and the mass load. Evidence shows that at 0 m, phosphorus concentration was at 0 mg/L the concentration surges to a peak of 3.0 mg/L at 174 m downstream. The increasing trend reverses gradually with increasing distance from the peak value (3.0 mg/L) to 0.3 mg/L at the outlet, 1740m downstream from location 0 m. The mass load for the phosphorus also

began the transport process at location 0m with an initial value at 0 kg/day. Like the

concentration, it also experiences a sharp rise from 0 kg/day to a peak value of 0.5 kg/day

at location 174 m downstream, it then stays in equilibrium with increasing distance until

it exits the domain outlet, located 1740 m downstream. The Figure also shows the spatial

and temporal variation of the overland flow with location. The overland flow is at its

lowest at location 0 m and maximum at the outlet (location 1740 m). The progressive of

the flow rate along the length of the disconnected area (Figure 5.13) is attributed to the

cumulative of spatial locations in the area.

The temporal variation for the phosphorus concentration at location 500 m and

1740 m (outlet) (corresponding to the locations in Figure 5-13) is presented in Figure 5-

14.
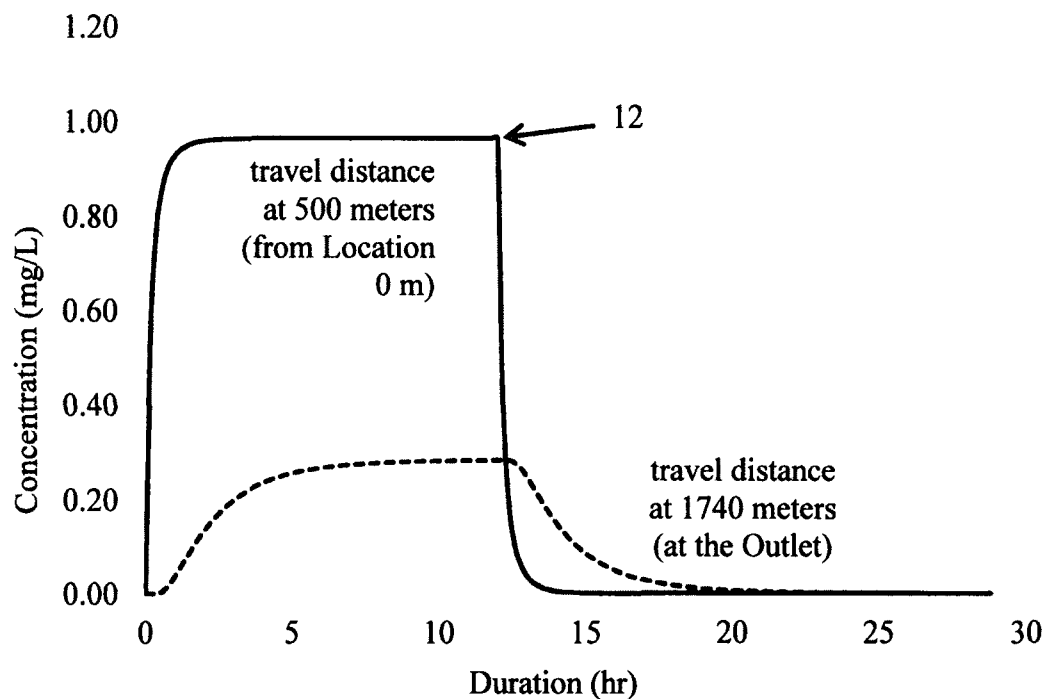
Figure 5-14: Temporal Variation of Diffused Sources

The results show the ephemeral or temporal characteristics of the phosphorus pollutant at two locations, 500m and 1740 m, downstream of location 0 m (Figure 5-13). Evidence from the graph shows both spatial locations exhibit similar ephemeral characteristics as limb of the pollutograph for both locations rises and falls. At location 500 m, the concentration rose sharply from 0 mg/L at time 0 hours to 0.98 mg/L at time 3 hours into the storm duration, maintains equilibrium up to 12 hours (end of storm duration) and then drops sharply to 0 mg/L at 15 hours (3 hours after the storm). At location 1740 m (outlet), concentration was at 0 mg/L at time 0 hours; it then rises gently to a peak of 0.3 mg/L at time 6 hours into the storm, remains constant up to the end of the storm and then recedes gently from 12 hours (the time at which the rain stops) to a value of 0 mg/L at time 21 hours (9 hours after the rain).

5.5.3 Sensitivity of System-Level Superpositioned Model

The runs for each series of input parameter generated from the Monte Carlo runs are presented in Figure 5-15 for interception and Figure 5-16 for evapotranspiration.
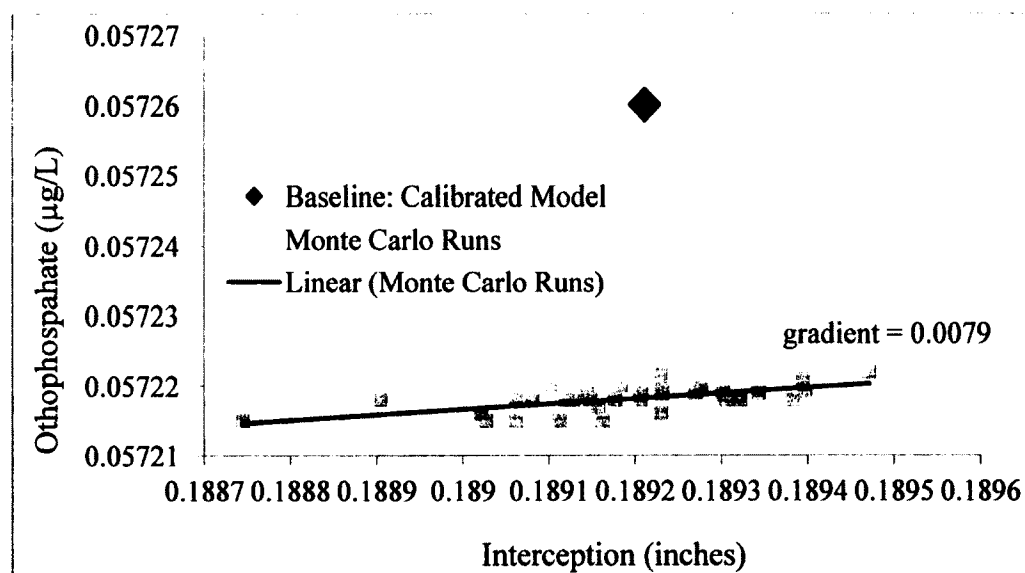
Figure 5-15: Response of Superpositioned Model to Variations in Interception
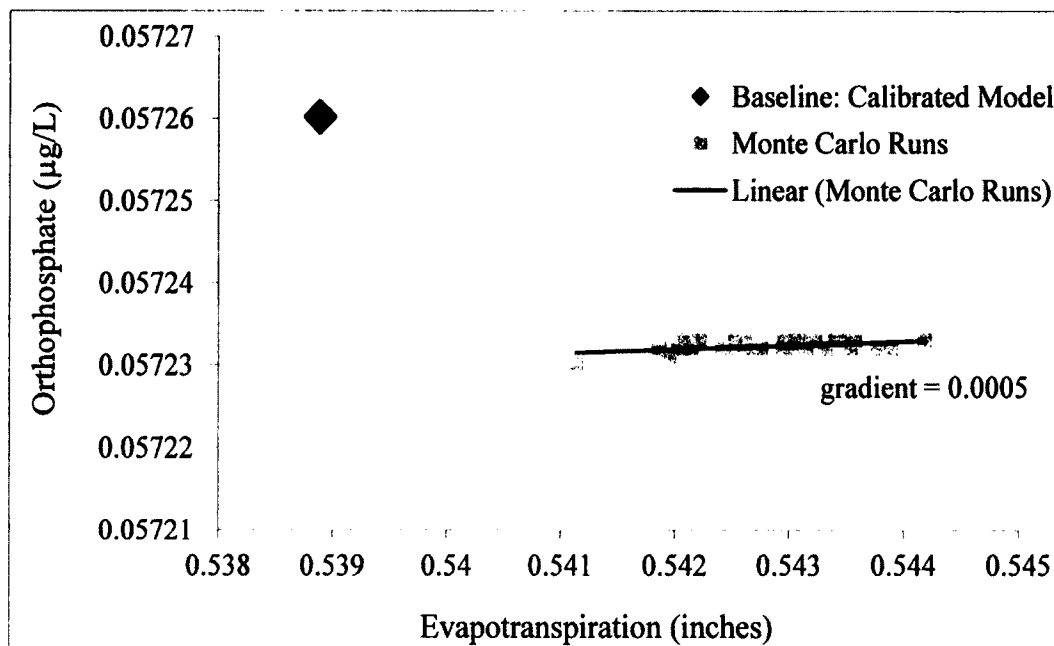
Figure 5-16: Response of Superpositioned Model to Variations in
Evapotranspiration

Figures 5-15 and 5-16 demonstrate how the inherent uncertainty in the interception

and the evapotranspiration individually affects the model's outcome. The results compare

the baseline model with outputs obtained from the series of input data generated from

Monte Carlo simulation. The trend line describes the steepness or gradient of the

distribution of the outputs from Monte Carlo runs. The response of the model to variation

in the interception parameter shows a steeper gradient (0.079) than that of the

evapotranspiration parameter (0.0005). The corresponding absolute sensitivity to the

results presented in Figures 5-15 and 5-16 is shown in Figure 5-17
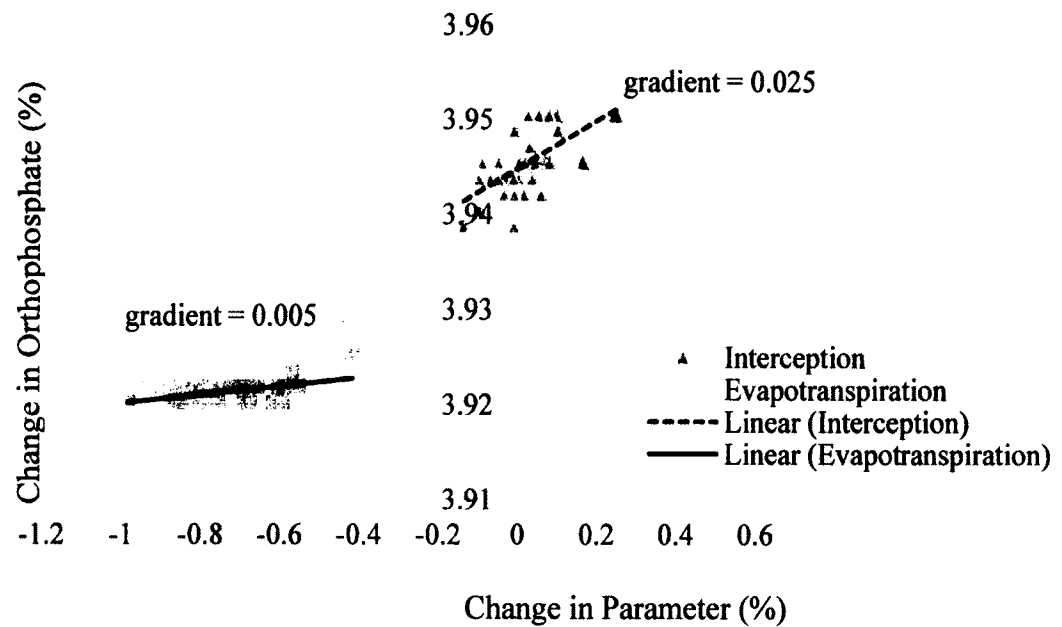
Figure 5-17: Sensitivity of Superpositioned Model to Interception and Evapotranspiration

The absolute sensitivity of the model to variations in the interception and the evapotranspiration parameter is presented in percentages so as to enable comparison between the two input parameters. The result shows the percentage variation for each of the parameters and the corresponding percentage change in the model's outcome. The gradient of model's response to the interception parameter was higher than that for the evapotranspiration parameter. The model is more responsive to smaller changes in the interception parameter than the evapotranspiration parameter. Smaller increment in the interception parameter would cause significant change in the model's outcome.

5.5.4 Sensitivity of Nonpoint source to Storm Events

The relative sensitivity of the mass loading to different intensities for all the storm events (2, 5, 10, 25 years) at 6, 12, 24 and 48 hour duration is shown in Figure 5-18
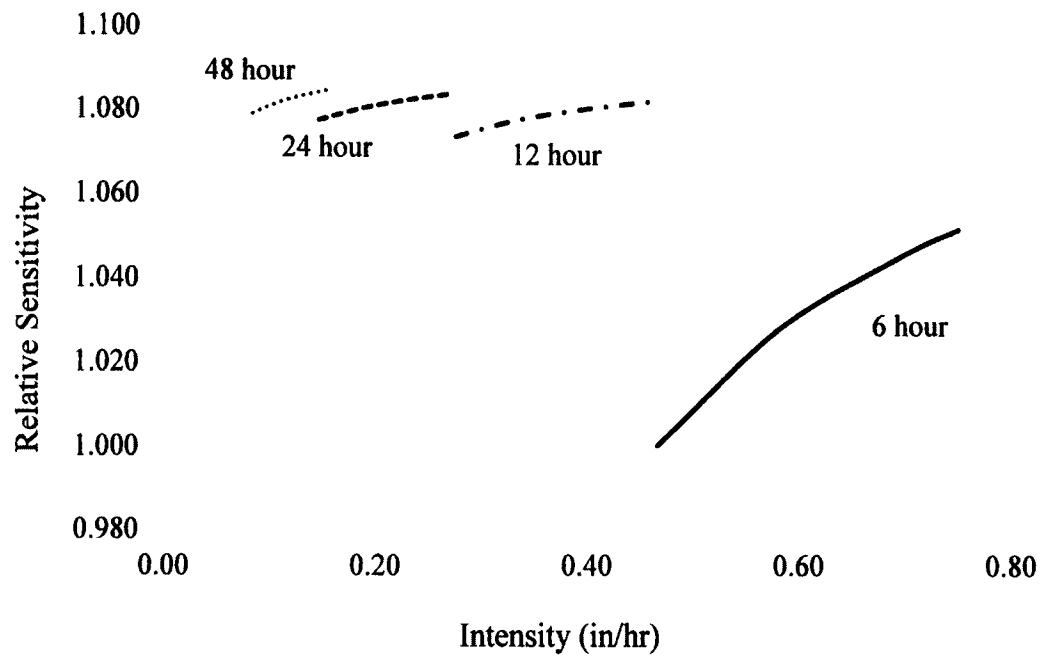
Figure 5-18: Sensitivity of Diffused Sources to Rainfall Intensity

The Figure 5-18 shows the sensitivity of nonpoint source mass loading in the disconnected area to variations in intensity for different storm durations. The curves represent relative sensitivity at 6 hours, 12 hours, 24 hours and 48 hours. Relative sensitivity was between 0.987 and 1.1 with intensity varying between 0.1 to 0.8 inches for all storm durations. The storm duration with the greatest variation in sensitivity was 6 hours, followed by 12 hours with 48 hours having the least variation in sensitivity. The curves suggest that a 6 hour duration of rainfall with intensities has the most significant impact on mass loading.

The sensitivity of the mass load to variations in storm duration for return periods is shown in Figure 5-19.
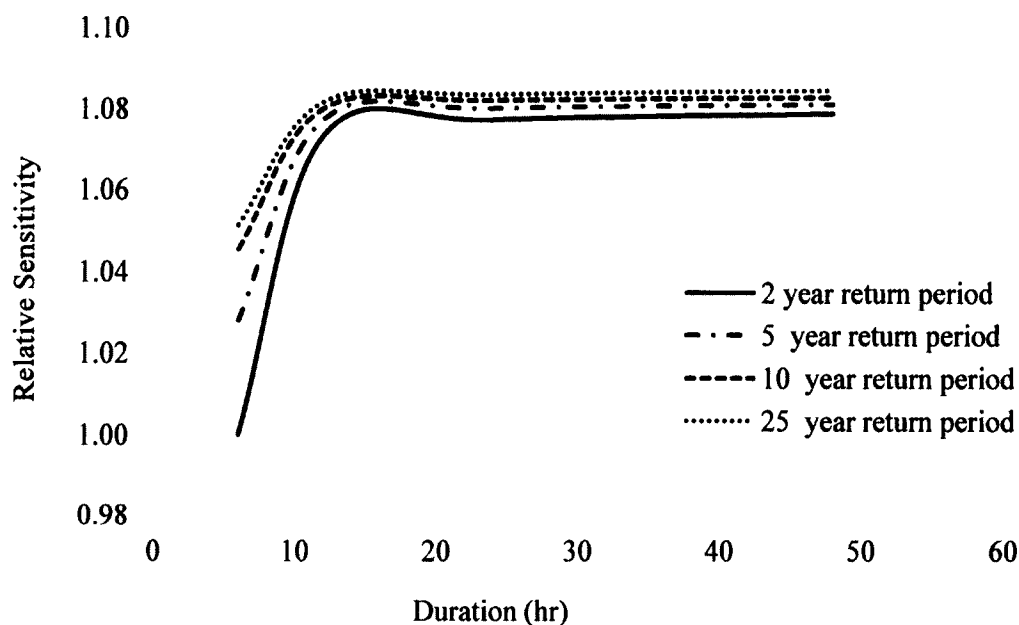
Figure 5-19: Sensitivity of Diffused Sources to Storm Duration

The curves represent relative sensitivity at 2 year return period, 5 year return period, 10 year return period and 25 year return period. The sensitivity curve for the 2 year return period shows the greatest variation, succeeded by 5 year return period with the 48 year return period having the least variation. The curves suggest that a 2 year return period with duration of rainfall between 5 hours and 15 hours has the most significant impact on mass loading.

5.6   Discussion

The key objective for this chapter is to develop an approach to estimate overall response of a watershed to nonpoint source through the superpositioning of nonpoint source contribution from hydrologically disconnected areas in the Patuxent watershed. The method was successfully applied to predict orthophosphate pollution in the Patuxent watershed. Overall, the approach proved to be a better estimator of the response to

orthophosphate pollution in the watershed than the system-level approach only and the classical method. The study also examined the response of nonpoint (diffused) source from the disconnected area to storm events and investigated the spatiotemporal characteristics of the diffused source.

5.6.1 Nonpoint Source Superpositioning

Concentration for phosphorus was estimated during a six day rainfall event. Concentration was found to increase from 0 mg/L on day 1 (precipitation day 0) to peak on day 2 (precipitation day 1), remained in equilibrium and receded to 0 mg/L on day 8 (12 hours after the storm). Overall, the range of phosphorus concentration estimated for the marsh was 0 mg/L to 0.14 mg/L. The range is in agreement with that recorded by Makepeace et al. (1995) and Lee and Bang (2000) for storm water quality. The study attributes the rise to peak and staying in equilibrium to the ephemeral characteristics of the overland flow (Fujiwara et al., 1988; Luo et al., 2006; Jiang et al., 2010) as the generation and the transport of diffused sources relies solely on rainfall runoff. Overall, the findings show that, nonpoint source inherits and exhibits temporal characteristic of rainfall runoff (Fujiwara et al., 1988).

The system-level approach was super-positioned with the estimated overland phosphorous concentration to simulate orthophosphate for the watershed. The resulting orthophosphate concentration was predicted within reasonable accuracy. It ranged from 0 µg/L on the days of no precipitation to 0.3 µg/L within days of precipitation (Figure 5-9). The simulated result was compared with results from the system level method without superpositioning and the classical method (Figure 5-9). The outcome of the comparison showed that the system-level with nonpoint source superpositioning matches the

observed better than the other methods. The prediction error for the system-level approach with the super-positioned nonpoint sources was 0.29% (significant at a p-value of 0.954) while prediction error for the system-level approach without superpositioning and the classical method were -4.82% and -20.55 with p-values 0.58 and 0.19 respectively. None of the three methods could be rejected; however, based on the low prediction error for the system-level with superpositioning, the study concludes that the system-level approach with nonpoint source superpositioning is superior and a better approach to estimating the overall system's response to nonpoint source pollution.

5.6.2 Nonpoint Source Characterization

The hydrologic response to diffused sources in terms of magnitude was successfully characterized for the hydrologically-disconnected area. The response variables determined were flow rate, concentration and mass loading. The variables were determined for storm event at 2, 5, 10, and 25 year return periods at 6, 12, 24 and 48 hour storm durations.

For the overland flow, the results generally showed that (1) flow rate increases with decreasing storm durations for the same return period, and (2) flow rate increase with increasing return period for the same storm duration. This phenomenon is attributed to the relationship between the flow depth, rainfall intensity and duration and the assumptions underlying the rational method. These are explained in details in Akan and Houghtalen (2003). The higher flow rates for smaller storm durations are explained by the decreasing of rainfall intensity for increasing storm durations. Thus for the same return period, smaller durations have higher rainfall intensity than higher durations. The phenomenon is also explained by effect of the time of concentration. Time of

concentration is the time required for a stormwater to flow from the hydrologically most remote point in a basin to the outlet. It represents the hydrologic response time of the basin. All storm durations are assumed to have the same time of concentration; hence, a storm event whose duration equals the time of concentration would produce a higher flow rate.

For concentration, the results showed that (1) increasing the storm duration produces higher concentration for the same return period and (2) increasing the return periods decreases the concentration for the same storm duration. Case 1 agrees with Luo *et al.* (2006) and Jiang *et al.* (2010). These studies pointed out that the amount of nonpoint depends greatly on the magnitude of runoff. In case 2 where the concentration decreases with increasing return period is attributed to the effect of "the first flush of stormwater runoff" and runoff volume. The first flush of stormwater runoff assumes the first part of runoff is most polluted (Deletic, 1998). This concept is often applied in stormwater sampling as stormwater samples are usually measured within 1 to 2 hours into the rainfall duration (Zou and Christensen, 2010). Studies have shown that, during the first flush, a significant amount of pollutant is washed off into an initial percentage of generated runoff volume. For example, 80% of pollutant is transferred to the first 30% of runoff volume (Saget *et al.*, 1996; Kim *et al*, 2005) and 80% of pollutant is transferred to the first 25% of runoff volume (Vorreiter and Hickey, 1994). For different return periods but the same duration, runoff volume increases with increasing return period. Hence smaller return periods will have higher concentration than storms with higher return periods.

The estimated concentration values ranged from 0 mg/L to 0.95 mg/L. The range agrees with the range values found in literature (Maristany and Bartel, 1989; Athayde, 1984; Line *et al.*, 2002). The normal values are within the range of 0 to 0.7 mg/L (Taebi and Droste, 2004; McLeod *et al.*, 2006). The higher values for concentration (greater than 0.7 mg/L) are related to storm durations greater than 24 hours.

The results presented for the mass load generally shows that (1) smaller return periods generate smaller mass load and higher return period higher mass load, (2) smaller return periods rises to peak faster for shorter durations, and (3) all return periods reach a state of equilibrium after 12 hours. The study attributes the first finding to the magnitude of volumetric flow (Luo *et al.*, 2006; Jiang *et al.*, 2010) as these studies demonstrated the amount nonpoint source generated relates directly the magnitude of runoff. The second finding is attributed to the described effect of first flush of stormwater runoff. Higher return periods will have higher runoff volumes hence lower concentration while smaller return period will have smaller runoff volumes hence higher concentration. This phenomenon could be the likely cause of the quick rise in the mass load observed for the 2 and the 5 year return periods. For the third finding, the study could not find any plausible cause or other explanation other than assuming that the loading capacity of the marsh considered in this study has reached a state of saturation for all storm durations and return periods after 12 hours. However, the cause of the equilibrium and the plausible explanation given is subject to further investigation, which is beyond the scope of this study.

The spatial characteristics of the nonpoint source were shown to vary for all locations along the length of disconnected area. The temporal variation was also shown

for two locations along the length of the disconnected area. The results showed that nonpoint source vary in space and time, thus confirming Thomann and Mueller (1987) and Fujiwara *et al.* (1988) that nonpoint sources exhibit ephemeral characteristics.

5.6.3 Sensitivity of System-Level Superpositioned Model

The response of the system-level superpositioned with nonpoint source- to variations in interception and evapotranspiration was successfully analyzed using sensitivity analysis. The results suggest that the system-level superpositioned- with nonpoint- source is more sensitive to variations in interception than the evapotranspiration parameter. The findings contradict most published sensitivity studies such as Rouse (2000), Domec *et al.* (2012), Bryant *et al.* (2005) and Dadaser-Celik *et al.* (2006) as these studies have established that wetlands are most sensitive to evapotranspiration. Only a few studies like Miller (2002) mentioned that wetlands are moderately sensitive to the interception parameter. This finding for a marsh area is unusual as wetland lands in hydrology are often identified with evapotranspiration (or evaporation).

The study attributes this anomaly to the sensitivity of the hydrologic section of the system-level model being carried over and argues although the magnitude of diffuse depend on the magnitude of runoff, diffused do not directly relate to the physiographic parameters. The physiographical parameters play key roles in determining the amount of runoff. When the runoff is created, these parameters have little or no effect on the transport process which relates directly to the generation and the driving of the nonpoint source. This finding stresses the need to precede every nonpoint source modeling with a hydrologic (overland flow) that has already accounted for possible uncertainty that might

be associated with its candidate parameters. It is recommended to include the interception parameter in hydrologic input parameterization (described in Chapter 3) prior to hydrologic modeling in the future for nonpoint source superpositioning applications.

5.6.4 Nonpoint Source Sensitivity to Storm Events

The relative sensitivity analysis for the nonpoint source was successfully implemented. Results show that the nonpoint source generated from the disconnected area is sensitive to both intensity and duration. For different intensities, the nonpoint source was found to be more sensitive to 6 hours for all return periods with intensity variation between 0.4 and 0.8 inches. The 6 hours was succeeded by 12 hours which had intensities varying between 0.2 and 0.4 inches. The duration with least or insignificant impact was the 48 hour duration. For different storm durations, the nonpoint source generation was much sensitive to shorter return period than with longer return periods. A 2 year return period has greater impact on the generation of nonpoint source than a 25 year return period. Overall, nonpoint source is sensitive lower return period and shorter durations.

5.7 Implications to Water Quality Modeling

Water quality is a mathematical representation of real world phenomenon. A model that incorporates every possible variation inherent in a physical system to predict hydrologic response variables for a watershed or any domain of interest is desirable. This chapter provided the basis for augmenting existing nonpoint source models with potential sources that are unaccounted for. Equations 5.4 and 5.5 were successfully applied in this chapter to account for sources unaccounted for in the system-level method in Chapter 4.

The approach performed reasonably well as it was able further improve prediction accuracy. Based on the findings of this study, equations 5.4 and 5.5 can be used to augment existing nonpoint source models to determine the overall response to nonpoint source pollution.

The findings on the sensitivity analysis to storm events imply that whenever a storm event similar to 2 and 5 year return periods at durations up to 12 occurs in a watershed, this methodology in this chapter can be used to estimate nonpoint source from hydrologically disconnected areas. The estimated nonpoint source can be superpositioned onto an existing hydrologic model to determine the overall response of the watershed. Neglecting this fact may result in a significant error in prediction if multiple disconnected areas exist in the watershed.

5.8 Summary and Conclusion

Equations to predict the spatial and temporal variation was successfully developed and applied to augment the model in Chapter 4 in order to determine the overall response of the watershed to nonpoint source. Areas defined by the study as hydrologically disconnected from the model in Chapter 4 were identified and the equations were used to determine the amount of nonpoint source generated for superpositioning. Results show that superpositioning nonpoint source contribution from areas not accounted for by hydrological models reduces prediction error.

The equations were also used to investigate the response of the nonpoint generated from the disconnected areas to different storm events. The finding of the study indicates that flow rate is higher for higher return periods for the same storm duration and also higher smaller durations for the same return period. Concentration was found to increase

with storm duration for all return periods. However, smaller return periods tend be much higher in concentration than higher return periods. The mass load was also found to be higher for higher return periods for the same duration. However, it was found that smaller return period with duration less than 12 hours have the tendency of spiking up faster than higher return periods. The study also investigated the spatial and temporal characteristics of the nonpoint source from the disconnected area. The result confirmed that diffused (nonpoint) sources vary in space and time.

A sensitivity analysis showed that the system-level model superpositioned with nonpoint source is sensitive to the interception parameter. The study classified this finding as unusual as marshes (emergent wetlands) are expected to be sensitive the evapotranspiration parameter. The finding was attributed to the sensitivity of the hydrologic section of the system-level model (Chapter 4) being carried over. Also a sensitivity analysis on the response of nonpoint source to different storm events show that, smaller return periods and smaller durations have the greatest impact on nonpoint source generation. Hence, when storm events similar to the 2 and 5 year return periods at durations up to 12 occur at a watershed, the methodology described in this chapter can be used. The methodology will help estimate nonpoint source from hydrologically disconnected areas which can be superpositioned onto an existing hydrological model to determine the overall response of the watershed. Overlooking this finding could result in a significant prediction error.

CHAPTER 6

SUMMARY, CONTRIBUTIONS, AND FUTURE WORK

Predicting system's response to hydrologic processes is fundamental to hydrologic modeling. However, the accuracy of prediction is often plagued by error due to various forms of inherent uncertainty associated with input data of the physical system being modeled. These uncertainties arise from sources such as mis-interpretation of system characteristics, measurement error, data unavailability and processing or synthesis technique. This study presents a hybrid methodology framework which integrates stochastic and deterministic methods at system-level to improve prediction accuracy of system's responses. Improving system's respond accuracy would help in the overall protection and management of these vital water resources.

This chapter reviews the methodology framework set-up in Chapter 3 and highlights the findings for the applications found in Chapter 4 and 5. The main findings and conclusions for each of the objectives mentioned in section 1.2 of the dissertation are individually summarized (section 6.1), and later integrated within a broader context (Section 6.2) of the overall study. Also, a brief discussion on the significance of the study to predicting systems response, contribution to the field of hydrologic modeling and recommendations for future research are presented in Section 6.3, Section 6.4, and Section 6.5 respectively.

6.1 Summary of Conclusions

In Chapter 3, the study presented a methodology that integrates stochastic and deterministic methods to reproduce the true response of systems to hydrologic processes.

Proposed hybrid modeling framework consists of three components; (1) microscale stochastic parameterization process at the subbasin-level, (2) mesoscale deterministic computational process at watershed segment-level, and (3) macroscale stochastic sensitivity process at whole watershed scale.

Microscale stochastic portion involves the use of a GLM and Monte Carlo sampling to preprocess hydrologic input parameters at system-level. The GLM developed in this study requires topography and surficial area as input to determine sampling grid resolution and estimating system moments. The system moments were used to stochastically draw series of samples in the form of probability density functions (PDFs) from the population of the parameter being preprocessed. An overall lumped parameter value representation of the watershed is then estimated from generated samples. This system-level preprocessing of parameter is underlain by the assumption that, the resulting lumped parameter value embodies every possible variation that can occur in the watershed.

The mesoscale deterministic computational process includes the selection of a computation engine to serve as a carrier for the stochastically prepared input data. For this study, a calibrated Hydrologic Simulation Program in Fortran (HSPF) was selected for computation. The final stage of the methodology framework is a macroscale stochastic posterior sensitivity process. The process consisted of a model level sensitivity where other class of system parameters that were not processed during the microscale stochastic pre-parameterization stage were investigated to assess their impact on the model outcome.

Chapter 4 verifies and implements the methodology framework presented in Chapter 3 to the study a watershed. The Patuxent watershed in Maryland was chosen for analysis of overland flow and diffused (nonpoint) sources – orthophosphate and total suspended solids. Comparative analysis on prediction effectiveness and efficiency between the proposed (system-level approach) and the conventional (classical approach) were also made. The input parameters preprocessed were focused on physiographic parameters of the physical system being modeled. This includes surface roughness, infiltration and lower zone soil moisture. Model level stochastic posterior sensitivity analysis was performed on interception and the evapotranspiration parameters. Compared to the observed field data, the system-level approach predicted with improved accuracy than the classical approach for overland flow. In both the high flowing northern segment and the low flowing southern segment of the watershed the system-level approach prediction was more accurate for overland flow. The prediction error in general for the system-level approach ranged from -28.3% to 8.3% while that for the classical approach varied from of -68% to 17%.

Unlike the classical approach, the system-level approach showed systematic similarity in the prediction trend throughout the watershed, an indication of better reproducibility of natural system behavior. For system-level approach predicted the orthophosphate with a robustness evidence by high Nash and Sutcliffe values ranging from 0.64 to 0.989 while that by the classical approach was -0.18 to 0.33. For the total suspended solids, the magnitude Nash and Sutcliffe for both methods were approximately the same. The range of Nash and Sutcliffe efficiency values ranged from 0.77 to 0.96 for the system-level approach while that for the classical was from 0.65 to 0.95. However,

the results for the system-level approach was significant at p-values greater than the level of significance ($\alpha = 0.05$) while that for the classical approach was insignificant for the watershed through the watershed The trend in nonpoint source prediction was consistent throughout the watershed for orthophosphate and the total suspended solids.

The model level sensitivity showed that the overland flow in the watershed is very sensitive to small variations in the evapotranspiration parameter. For the nonpoint sources, the orthophosphate and total suspended solids also showed similar trend of sensitivity as that of the overland flow throughout the watershed. The system-level approach facilitates significant improvements in predictive accuracy and reproducing system's responses in overland flow and dissolved pollutants compared to the classical approach used commonly hydrologic modeling practices.

In Chapter 5, a methodology for improving the prediction of diffused (nonpoint) sources through the method superpositioning was presented. The method augments existing nonpoint source models to account for potential sources coming from hydrologically disconnected areas. Superpositioning method in theory should lead to improvement in overall system's response to nonpoint sources. However this approach is not employed in hydrologic models due to its complexity.

The study adapted and modified existing concepts of mathematical models for water bodies to estimate overland pollutant efflux from areas that are hydrologically disconnected, i.e., island, as a result topographic segmentation. The methodology was successfully applied to hydrologically disconnected areas found in the watershed through superpositioning onto the system-level method. The prediction error for the system-level method reduced from 4.82% to -0.29% for nonpoint source when the method

superposition was employed. The chapter also determined the watershed response for various storm events in the hydrologically disconnected catchment. Among the key findings is that flow rate decreased with increasing storm duration for the same return period. Also, concentration increased proportionally with increasing duration for all storms but decreased with increasing return period for the same duration, indicating that smaller return periods produce higher concentrations than higher return period. The mass loading increased with duration and with return period. Results for the mass loading also showed that shorter durations and shorter return periods are quicker to arrive at the peak concentration. Findings also confirmed that nonpoint sources vary in space and time

A sensitivity analysis was performed for the system-level approach with superpositioned nonpoint source. The model was found to be very sensitive to small variations in the interception parameter. Also, the nonpoint source from the hydrologically disconnected area was found to be very sensitive to shorter storm durations and shorter return periods. The study results demonstrated that incorporating those hydrologically disconnected areas into modeling improves the overall response of the watershed as these areas may contribute significantly to nonpoint source pollution.

6.2 General Conclusions

Considering the current levels of error associated with most hydrologic predictions that are often the consequence of uncertainties inherent in hydrologic input data. The study argued that reproducing a system based on its true characteristics could improve prediction. A methodology framework which integrates stochastic and deterministic concepts based on system's principles was conceptualized and superpositions nonpoint source contributions from hydrologically disconnected areas was

verified and implemented. The methodology was successfully applied to predict a system's response to hydrologic processes in the Patuxent watershed. The focus of the study employing this approach was to reduce prediction errors. The findings presented in the preceding sections clearly demonstrated that the approach of using system-level characteristics significantly reduces model prediction errors associated with current methods and improves reproducibility described in the hydrologic response of the system being modeled.

6.3 Significance and Contribution to Hydrology

This dissertation provides baseline for hydrologic modeling at system-level by presenting a methodology framework that integrates stochastic and deterministic methods to improve prediction accuracy. The method estimates input parameters at system-level and superpositions diffused sources from hydrologically disconnected areas to account for overall systems response. The results of the study (Chapters 4 and 5) confirm the pivotal role of system characteristics in determining prediction accuracy especially for response variables such as overland flow and diffused sources. Chapter 4 which verified and implemented the system-level approach showed that the methodology framework reduces prediction error for overland flow and nonpoint source pollutants.

Chapter 5 further provided an approach and a basis for augmenting existing nonpoint source modeling scheme for potential sources that are unaccounted for in typical hydrologic modeling processes. The chapter also provided criteria on when it is important to ignore potential nonpoint sources for hydrologically disconnected areas. This information would be useful for water resources engineers to improve prediction accuracy and lead to effective management of watershed.

The successful integration of stochastic and deterministic methods to improve hydrologic prediction is the ultimate dream in the field of hydrology. This dissertation makes significant contribution to the field by estimating hydrologic input parameters at system-level and superpositioning of contributions from hydrologically disconnected areas often ignored by models. The study modifies the concept of in stream pollutant transport for application to diffused sources (Chapter 5). The dissertation shows that system characteristics can be used to accurately and quantitatively improve prediction accuracy. In general, the undertaken research improves our knowledge and understanding of hydrologic prediction at system-level.

6.4 Future Research

Modeling of a coastal basin requires the integration of a watershed model to in-stream model. The methodology framework in this study was developed for determining the hydrologic response of a watershed to overland processes. However, controlling pollution at a basin scale would require the integration of both overland and in-stream processes for efficient planning and management. Hence additional studies on in-stream processes need to be conducted. The study recommends the integration of overland and in-stream model like the Water Quality Analysis Simulation Program (WASP) to understand and capture the full extent of system's response in a watershed.

REFERENCES

Abdi, H. (2010). Coefficient of variation. *Encyclopedia of Research Design. SAGE Publications, Inc., Thousand Oaks, CA*, 169–171.

Agarwal, S., Vailshery, L. S., Jaganmohan, M., and Nagendra, H. (2013). Mapping Urban Tree Species Using Very High Resolution Satellite Imagery: Comparing Pixel-Based and Object-Based Approaches. *ISPRS International Journal of Geo-Information, 2*(1), 220–236.

Agnew, L. J., Lyon, S., Gérard-Marchant, P., Collins, V. B., Lembo, A. J., Steenhuis, T. S., and Walter, M. T. (2006). Identifying hydrologically sensitive areas: Bridging the gap between science and application. *Journal of environmental management, 78*(1), 63–76.

Akan, A. O., and Houghtalen, R. J. (2003). *Urban hydrology, hydraulics and stormwater quality*. Wiley Hoboken, NJ.

Al-Sabhan, W., Mulligan, M., and Blackburn, G. A. (2003). A real-time hydrological model for flood prediction using GIS and the WWW. *Computers, Environment and Urban Systems, 27*(1), 9–32.

Albek, M., Bakır Ö\ugütveren, Ü. and Albek, E. (2004). Hydrological modeling of Seydi Suyu watershed (Turkey) with HSPF. *Journal of Hydrology, 285*(1), 260–271.

Albert, R. (2011). The influence of past and future urbanization on watershed nitrogen export and hydrology dynamics in two mid-Atlantic watersheds in Fairfax, Virginia. Retrieved from http://digilib.gmu.edu:8080/xmlui/handle/1920/6638

Andersson, J.-O., and Nyberg, L. (2008). Relations between topography, wetlands, vegetation cover and stream water chemistry in boreal headwater catchments in Sweden. *Hydrology and Earth System Sciences Discussions Discussions, 5*(3), 1191–1226.

Arhonditsis, G., Tsirtsis, G., Angelidis, M. O., and Karydis, M. (2000). Quantification of the effects of nonpoint nutrient sources to coastal marine eutrophication: applications to a semi-enclosed gulf in the Mediterranean Sea. *Ecological Modelling, 129*(2), 209–227.

Arnold Jr, C. L., and Gibbons, C. J. (1996). Impervious surface coverage: the emergence of a key environmental indicator. *Journal of the American Planning Association, 62*(2), 243–258.

Asta, J., Erhardt, W., Ferretti, M., Fornasier, F., Kirschbaum, U., Nimis, P. L., ... Van Haluwyn, C. (2002). Mapping lichen diversity as an indicator of environmental quality. In *Monitoring with lichens—Monitoring lichens* (pp. 273–279). Springer.

Athayde, D. N. (1984). *Results of the nationwide urban runoff program*. National Technical Information Service.

Ballio, F., and Guadagnini, A. (2004). Convergence assessment of numerical Monte Carlo simulations in groundwater hydrology. *Water Resources Research, 40*(4), n/a–n/a. doi:10.1029/2003WR002876

Barnes, K. B., Morgan III, J. M., and Roberge, M. C. (2001). *Impervious surfaces and the quality of natural and built environments*. Project to map impervious cover for the entire Chesapeake Bay and Maryland Coastal Bays watersheds, asp. Retrieved from http://pages.towson.edu/morgan/files/Impervious_Surfaces.pdf

Bellin, A., Rubin, Y., and Rinaldo, A. (1994). Eulerian-Lagrangian approach for modeling of flow and transport in heterogeneous geological formations. *Water Resources Research, 30*(11), 2913–2924.

Benedetti, L., Claeys, F., Nopens, I., and Vanrolleghem, P. A. (2011). Assessing the convergence of LHS Monte Carlo simulations of wastewater treatment models. *Water Science and Technology, 63*(10), 2219.

Benham, B. L., Brannan, K. M., Yagow, G., Zeckoski, R. W., Dillaha, T. A., Mostaghimi, S., and Wynn, J. W. (2005). Development of Bacteria and Benthic Total Maximum Daily Loads. *Journal of environmental quality, 34*(5), 1860–1872.

Beven, K. J., and Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrological Sciences Journal, 24*(1), 43–69.

Bhaskar, N. R., James, W. P., and Devulapalli, R. S. (1992). Hydrologic parameter estimation using geographic information system. *Journal of Water Resources Planning and Management, 118*(5), 492–512.

Bicknell, B. R., Donigian Jr, A. S., and Barnwell, T. A. (1985). Modeling water quality and the effects of agricultural best management practices in the Iowa River Basin. *Water Science and Technology, 17*(6-7), 1141–1153.

Bicknell, Brian R., Imhoff, J. C., Kittle Jr, J. L., Donigian Jr, A. S., and Johanson, R. C. (1997). *Hydrological simulation program–Fortran: User's manual for version 11*. US Environmental Protection Agency, National Exposure Research Laboratory Athens, GA.

Blöschl, G., and Sivapalan, M. (1995). Scale issues in hydrological modelling: a review. *Hydrological processes, 9*(3-4), 251–290.

Borah, D. K., and Bera, M. (2003). Watershed-scale hydrologic and nonpoint-source pollution models: Review of mathematical bases. *Transactions of the ASAE, 46*(6), 1553–1566.

Bowie, G. L., Mills, W. B., Porcella, D. B., Campbell, C. L., Pagenkopf, J. R., Rupp, G. L., Chamberlin, C. E. (1985). Rates, constants, and kinetics formulations in surface water quality modeling. *EPA, 600*, 3–85.

Boyle, P. P. (1977). Options: A monte carlo approach. *Journal of financial economics, 4*(3), 323–338.

Boynton, W. R., Garber, J. H., Summers, R., and Kemp, W. M. (1995). Inputs, transformations, and transport of nitrogen and phosphorus in Chesapeake Bay and selected tributaries. *Estuaries, 18*(1), 285–314.

Boynton, Walter Raymond, Hagy, J. D., Cornwell, J. C., Kemp, W. M., Greene, S. M., Owens, M. S., ... Larsen, R. K. (2008). Nutrient budgets and management actions in the Patuxent River estuary, Maryland. *Estuaries and Coasts, 31*(4), 623–651.

Broussard, W., and Turner, R. E. (2009). A century of changing land-use and water-quality relationships in the continental US. *Frontiers in Ecology and the Environment, 7*(6), 302–307.

Bryant, M. L., Bhat, S., and Jacobs, J. M. (2005). Measurements and modeling of throughfall variability for five forest communities in the southeastern US. *Journal of Hydrology, 312*(1), 95–108.

Burt, T. P., and Butcher, D. P. (1985). Topographic controls of soil moisture distributions. *Journal of Soil Science, 36*(3), 469–486.

Carle, M. V., Halpin, P. N., and Stow, C. A. (2005). Patterns Of Watershed Urbanization And Impacts On Water Quality1. *JAWRA Journal of the American Water Resources Association, 41*(3), 693–708.

Chapra, S. C. (1997). *Surface water-quality modeling* (Vol. 1). McGraw-Hill New York.

Chew, C. Y., Moore, L. W., and Smith, R. H. (1991). Hydrological simulation of Tennessee's north Reelfoot Creek watershed. *Research Journal of the Water Pollution Control Federation*, 10–16.

Costanza, R., Voinov, A., Boumans, R., Maxwell, T., Villa, F., Wainger, L., and Voinov, H. (2002). Integrated ecological economic modeling of the Patuxent River watershed, Maryland. *Ecological Monographs, 72*(2), 203–231.

Crosetto, M., Tarantola, S., and Saltelli, A. (2000). Sensitivity and uncertainty analysis in spatial modelling based on GIS. *Agriculture, ecosystems and environment, 81*(1), 71–79.

Dadaser-Celik, F., Stefan, H. G., and Brezonik, P. L. (2006). Dynamic hydrologic model of the Örtülüakar Marsh in Turkey. *Wetlands, 26*(4), 1089–1102.

Dail, H. M., Kazyak, P. F., Boward, D. M., and Stranko, S. A. (1998). *Environmental assessment of stream conditions* (CBWP-MANTA-EA-98-7). NOAA, Maryland Deparment of Natural resources.

De Groen, M. M., and Savenije, H. H. (2006). A monthly interception equation based on the statistical characteristics of daily rainfall. *Water Resources Research, 42*(12).

Deletic, A. (1998). The first flush load of urban surface runoff. *Water Research, 32*(8), 2462–2470. doi:10.1016/S0043-1354(97)00470-3

DeVantier, B. A., and Feldman, A. D. (1993). Review of GIS applications in hydrologic modeling. *Journal of Water Resources Planning and Management, 119*(2), 246–261.

Dexter, A. R. (2004). Soil physical quality: Part I. Theory, effects of soil texture, density, and organic matter, and effects on root growth. *Geoderma, 120*(3), 201–214.

Domec, J.-C., Sun, G., Noormets, A., Gavazzi, M. J., Treasure, E. A., Cohen, E., ... King, J. S. (2012). A comparison of three methods to estimate evapotranspiration in two contrasting loblolly pine plantations: age-related changes in water use and drought sensitivity of evapotranspiration components. *Forest Science, 58*(5), 497–512.

Donigihan, A. S., and Davis, H. H. (1978). User's manual for agricultural runoff management (ARM) model. *Ecological research series*.

Engman, E. T. (1986). Roughness coefficients for routing surface runoff. *Journal of Irrigation and Drainage Engineering, 112*(1), 39–53.

Finlayson, D. P., Montgomery, D. R., and Hallet, B. (2002). Spatial coincidence of rapid inferred erosion with young metamorphic massifs in the Himalayas. *Geology, 30*(3), 219–222.

Fishman, G. S. (1996). *Monte Carlo: concepts, algorithms, and applications* (Vol. 1196). Springer New York.

Fohrer, N., Haverkamp, S., and Frede, H.-G. (2005). Assessment of the effects of land use patterns on hydrologic landscape functions: development of sustainable land use concepts for low mountain range areas. *Hydrological Processes, 19*(3), 659–672. x, J. (2008). *Applied regression analysis and generalized linear models* (Second.). California, USA: Sage.

Franceschini, S., and Tsai, C. W. (2010). Assessment of uncertainty sources in water quality modeling in the Niagara River. *Advances in Water Resources, 33*(4), 493–503.

Fujiwara, O., Puangmaha, W., and Hanaki, K. (1988). River basin water quality management in stochastic environment. *Journal of Environmental Engineering*, *114*(4), 864–877.

Gardner, C. W. (1983). *Handbook of stochastic methods for physics, chemistry and the natural sciences*. Berlin: Springer-Verlag.

Gerrits, A. M. J., Pfister, L., and Savenije, H. H. G. (2010). Spatial and temporal variability of canopy and forest floor interception in a beech forest. *Hydrological Processes*, *24*(21), 3011–3025.

Harman-Fetcho, J. A., McConnell, L. L., and Baker, J. E. (1999). Agricultural pesticides in the Patuxent River, a tributary of the Chesapeake Bay. *Journal of Environmental Quality*, *28*(3), 928–938.

Hayashi, S., Murakami, S., Watanabe, M., and Bao-Hua, X. (2004). HSPF simulation of runoff and sediment loads in the upper Changjiang River basin, China. *Journal of Environmental Engineering*, *130*(7), 801–815.

Hershfield, D. M. (1961). Technical Paper No. 40, Rainfall Frequency Atlas of the United States. *Cooperative Studies Section, Hydrologic Services Division for Engineering Division, Soil Conservation Service US Department of Agriculture, Washington*.

Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., Wickham, J. (2007). Completion of the 2001 National Land Cover Database for the Counterminous United States. *Photogrammetric Engineering and Remote Sensing*, *73*(4), 337.

Huang, J.-K., and Lee, K. T. (2009). Influences of spatially heterogeneous roughness on flow hydrographs. *Advances in water resources*, *32*(11), 1580–1587.

Hundecha, Y., and Bárdossy, A. (2004). Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model. *Journal of Hydrology*, *292*(1–4), 281–295. doi:10.1016/j.jhydrol.2004.01.002

Im, S., Brannan, K. M., Mostaghimi, S., and Kim, S. M. (2007). Comparison of HSPF and SWAT models performance for runoff and sediment yield prediction. *Journal of Environmental Science and Health Part A*, *42*(11), 1561–1570.

Isukapalli, S. S. (1999). *Uncertainty analysis of transport-transformation models*. Rutgers, The State University of New Jersey. Retrieved from http://www.ccl.rutgers.edu/~ssi/thesis-screen.pdf

Jetten, V., de Roo, A., and Favis-Mortlock, D. (1999). Evaluation of field-scale and catchment-scale soil erosion models. *CATENA*, *37*(3–4), 521–541. doi:10.1016/S0341-8162(99)00037-5

Jiang, H., Seco, J., and Paganetti, H. (2007). Effects of Hounsfield number conversion on CT based proton Monte Carlo dose calculations. *Medical physics*, *34*, 1439.

Jiang, T., Teng, L., Wei, S., Deng, L., Luo, Z., Chen, Y., and Flanagan, D. C. (2010). Application of polyacrylamide to reduce phosphorus losses from a Chinese purple soil: A laboratory and field investigation. *Journal of environmental management*, *91*(7), 1437–1445.

Johnson, M. S., Coon, W. F., Mehta, V. K., Steenhuis, T. S., Brooks, E. S., and Boll, J. (2003). Application of two hydrologic models with different runoff mechanisms to a hillslope dominated watershed in the northeastern US: a comparison of HSPF and SMR. *Journal of Hydrology*, *284*(1), 57–76.

Kadlec, R. H. (1994). Detention and mixing in free water wetlands. *Ecological Engineering*, *3*(4), 345–380.

Kalos, M. H., and Whitlock, P. A. (2008). *Monte carlo methods*. John Wiley and Sons.

Kalyanapu, A. J., Burian, S. J., and McPherson, T. N. (2010). Effect of land use-based surface roughness on hydrologic model output. *Journal of Spatial Hydrology*, *9*(2).

Kidd, S. E., Chow, Y., Mak, S., Bach, P. J., Chen, H., Hingston, A. O., ... Bartlett, K. H. (2007). Characterization of environmental sources of the human and animal pathogen Cryptococcus gattii in British Columbia, Canada, and the Pacific Northwest of the United States. *Applied and environmental microbiology*, *73*(5), 1433–1443.

Kim, L.-H., Kayhanian, M., Lau, S.-L., and Stenstrom, M. K. (2005). A new modeling approach for estimating first flush metal mass loading. *Water Science and Technology*, *51*(3-4), 159–167.

Kim, S. M., Benham, B. L., Brannan, K. M., Zeckoski, R. W., and Doherty, J. (2007). Comparison of hydrologic calibration of HSPF using automatic and manual methods. *Water Resources Research*, *43*(1), n/a–n/a. doi:10.1029/2006WR004883

Kleidorfer, M., Deletic, A., Fletcher, T. D., and Rauch, W. (2009). Impact of input data uncertainties on stormwater model parameters. *Water Sci. Technol*, *60*(6), 1545–1554.

Klemeš, V. (1983). Conceptualization and scale in hydrology. *Journal of Hydrology*, *65*(1-3), 1–23. doi:10.1016/0022-1694(83)90208-1

Krause, S., and Bronstert, A. (2005). An advanced approach for catchment delineation and water balance modelling within wetlands and floodplains. *Advances in Geosciences*, *5*(5), 1–5.

Kuczera, G., Kavetski, D., Franks, S., and Thyer, M. (2006). Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters. *Journal of Hydrology, 331*(1), 161–177.

Landau, D. P., and Binder, K. (2009). *A guide to Monte Carlo simulations in statistical physics*. Cambridge university press.

Langland, M. J., Lietman, P. L., and Hoffman, S. (1995). *Synthesis of nutrient and sediment data for watersheds within the Chesapeake Bay drainage basin*. US Geological Survey Lemoyne, Pennsylvania. Retrieved from http://www.getcited.org/pub/100175945

Laroche, A.-M., Gallichand, J., Lagace, R., and Pesant, A. (1996). Simulating atrazine transport with HSPF in an agricultural watershed. *Journal of Environmental Engineering, 122*(7), 622–630.

Lee, J. H., and Bang, K. W. (2000). Characterization of urban stormwater runoff. *Water Research, 34*(6), 1773–1780. doi:10.1016/S0043-1354(99)00325-5

Levy, B., and McCuen, R. (1999). Assessment of storm duration for hydrologic design. *Journal of Hydrologic Engineering, 4*(3), 209–213.

Li, H., Sivapalan, M., and Tian, F. (2012). Comparative diagnostic analysis of runoff generation processes in Oklahoma DMIP2 basins: The Blue River and the Illinois River. *Journal of Hydrology, 418*, 90–109.

Line, D. E., White, N. M., Osmond, D. L., Jennings, G. D., and Mojonnier, C. B. (2002). Pollutant export from various land uses in the Upper Neuse River Basin. *Water Environment Research*, 100–108.

Lipiec, J., Kuś, J., S\lowińska-Jurkiewicz, A., and Nosalewicz, A. (2006). Soil porosity and water infiltration as influenced by tillage methods. *Soil and Tillage research, 89*(2), 210–220.

Liski, J. (1995). Variation in soil organic carbon and thickness of soil horizons within a boreal forest stand–effect of trees and implications for sampling. Retrieved from http://helda.helsinki.fi/handle/1975/9212

Liu, Y. B., Gebremeskel, S., De Smedt, F., Hoffmann, L., and Pfister, L. (2003). A diffusive transport approach for flow routing in GIS-based flood modeling. *Journal of Hydrology, 283*(1), 91–106.

Loague, K., and Corwin, D. L. (1996). Uncertainty in regional-scale assessments of non-point source pollutants. *Applications of GIS to the Modeling of Non-point Source Pollutants in the Vadose Zone*, (applicationsofg), 131–152.

Lopez, S. R., Hogue, T. S., and Stein, E. D. (2012). A framework for evaluating regional hydrologic sensitivity to climate change using archetypal watershed modeling. *Hydrology and Earth System Sciences Discussions, 9*(12), 13729–13771.

Lumb, A. M., McCammon, R. B., and Kittle, J. L. (1994). *Users Manual for an Expert System (HSPEXP) for Calibration of the Hydrological Simulation Program–Fortran.* US Geological Survey Reston, VA. Retrieved from http://pubs.usgs.gov/wri/1994/4168/report.pdf

Luo, B., Li, J. B., Huang, G. H., and Li, H. L. (2006). A simulation-based interval two-stage stochastic model for agricultural nonpoint source pollution control through land retirement. *Science of the total environment, 361*(1), 38–56.

MacMillan, R. ., Jones, R. K., and McNabb, D. H. (2004). Defining a hierarchy of spatial entities for environmental analysis and modeling using digital elevation models (DEMs). *Computers, Environment and Urban Systems, 28*(3), 175–200. doi:10.1016/S0198-9715(03)00019-X

Makepeace, D. K., Smith, D. W., and Stanley, S. J. (1995). Urban stormwater quality: summary of contaminant data. *Critical Reviews in Environmental Science and Technology, 25*(2), 93–139.

Maristany, A. E., and Bartel, R. L. (1989). Wetlands and stormwater management: a case study of Lake Munson. Part I: long-term treatment efficiencies. In *Wetlands: Concerns and Successes. Proceedings of a Symposium held September 17-22 1989, Tampa, Florida. American Water Resources Association, Bethesda, Maryland. 1989. p 215-229, 7 fig, 6 tab, 8 ref.*

Mays, L. W. (1999). *Hydraulic Design Handbook: Larry W. Mays, Editor-in-chief.* McGraw-Hill. Retrieved from http://www.getcited.org/pub/100418665

McCuen, R. H. (1998). Hydrologic design and analysis. *Pren ce Hall, New Jersey,* 814.

McCuen, Richard H. (1973). The role of sensitivity analysis in hydrologic modeling. *Journal of Hydrology, 18*(1), 37–53.

McCuen, Richard H., and Snyder, W. M. (1986a). *Hydrologic modeling: Statistical methods and applications.* Prentice-Hall Englewood Cliffs, NJ.

McCuen, Richard H., and Snyder, W. M. (1986b). *Hydrologic modeling: Statistical methods and applications.* Prentice-Hall Englewood Cliffs, NJ.

McLeod, S., Kells, J., and Putz, G. (2006). Urban Runoff Quality Characterization and Load Estimation in Saskatoon, Canada. *Journal of Environmental Engineering, 132*(11), 1470–1481. doi:10.1061/(ASCE)0733-9372(2006)132:11(1470)

MD. (2010). *Chesapeake Bay Watershed Environmental Audit.*

Melone, F., Barbetta, S., Diomede, T., Peruccacci, S., Rossi, M., and Tessarolo, A. (2005). Review and selection of hydrological models–integration of hydrological models and meteorological inputs. *Resulting from Work package, 1*, 34.

Meybeck, M., Green, P., and Vörösmarty, C. (2001). A new typology for mountains and other relief classes. *Mountain Research and Development, 21*(1), 34–45.

Miller, J. F. (1964). *Two-to ten-day precipitation for return periods of 2 to 100 years in the contiguous United States*. US Department of Commerce, Weather Bureau. Retrieved from http://www.getcited.org/pub/102195477

Miller, S. N. (2002). Scale effects of geometric complexity, misclassification error and land cover change in distributed hydrologic modeling. Retrieved from http://arizona.openrepository.com/arizona/handle/10150/280103

Mishra, A., Kar, S., and Singh, V. P. (2007). Determination of runoff and sediment yield from a small watershed in sub-humid subtropics using the HSPF model. *Hydrological Processes, 21*(22), 3035–3045.

Mix, S. J. (1994). *Urban stormwater modeling and simulation*. CRC Press.

Moore, L. W., Matheny, H., Tyree, T., Sabatini, D., and Klaine, S. J. (1988). Agricultural runoff modeling in a small west Tennessee watershed. *Journal (Water Pollution Control Federation)*, 242–249.

Moradkhani, H., Sorooshian, S., Gupta, H. V., and Houser, P. R. (2005). Dual state-parameter estimation of hydrological models using ensemble Kalman filter. *Advances in Water Resources, 28*(2), 135–147.

Naithani, K. J., Baldwin, D. C., Gaines, K. P., Lin, H., and Eissenstat, D. M. (2013). Spatial Distribution of Tree Species Governs the Spatio-Temporal Interaction of Leaf Area Index and Soil Moisture across a Forested Landscape. *PloS one, 8*(3), e58704.

Nakane, K., and Haidary, A. (2010). Sensitivity analysis of stream water quality and land cover linkage models using Monte Carlo method. *Int. J. Environ. Res, 4*(1), 121–130.

Nash, Je., and Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of hydrology, 10*(3), 282–290.

Nasr, A., Bruen, M., Jordan, P., Moles, R., Kiely, G., and Byrne, P. (2007). A comparison of SWAT, HSPF and SHETRAN/GOPC for modelling phosphorus export from three catchments in Ireland. *Water Research, 41*(5), 1065–1073.

Nelder, J. A., and Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, 370–384.

Niemi, R. M., and Niemi, J. S. (1991). Bacterial pollution of waters in pristine and agricultural lands. *Journal of Environmental Quality, 20*(3), 620–627.

NOAA (National Oceanic and Atmospheric Administration. (1997). Climatological Data Annual Summary; maryland and Delaware, *120*(13).

Norman, L. M., Guertin, D. P., and Feller, M. (2008). A coupled model approach to reduce nonpoint-source pollution resulting from predicted urban growth: a case study in the Ambos Nogales watershed. *Urban Geography, 29*(5), 496–516.

O'loughlin, E. M. (1981). Saturation regions in catchments and their relations to soil and topographic properties. *Journal of Hydrology, 53*(3), 229–246.

O'loughlin, E. M. (1986). Prediction of surface saturation zones in natural catchments by topographic analysis. *Water Resources Research, 22*(5), 794–804.

O'Loughlin, E. M., and Bowmer, K. H. (1975). Dilution and decay of aquatic herbicides in flowing channels. *Journal of Hydrology, 26*(3), 217–235.

Obropta, C. C., and Kardos, J. S. (2007). Review of Urban Stormwater Quality Models: Deterministic, Stochastic, and Hybrid Approaches1. *JAWRA Journal of the American Water Resources Association, 43*(6), 1508–1523.

Olivera, F., and Maidment, D. (1999). Geographic information systems (GIS)-based spatially distributed model for runoff routing. *Water Resources Research, 35*(4), 1155–1164.

Olsson, U. (2002). Generalized linear models. *An applied approach. Studentlitteratur, Lund, 18.* Retrieved from http://www.moodle.ufba.br/file.php/10752/material_2012/Generalized_Linear_M odels.pdf

Ponce, V. M. (1989). *Engineering hydrology: Principles and practices* (Vol. 640). Prentice Hall Englewood Cliffs. Retrieved from http://www.getcited.org/pub/102711448

Porter, J. H., and Dooley Jr, J. L. (1993). Animal dispersal patterns: a reassessment of simple mathematical models. *Ecology*, 2436–2443.

Post, J., Hattermann, F. F., Krysanova, V., and Suckow, F. (2008). Parameter and input data uncertainty estimation for the assessment of long-term soil organic carbon dynamics. *Environmental Modelling and Software, 23*(2), 125–138. doi:10.1016/j.envsoft.2007.05.010

Qin, H.-P., Khu, S.-T., and Yu, X.-Y. (2010). Spatial variations of storm runoff pollution and their correlation with land-use in a rapidly urbanizing catchment in China. *Science of the Total Environment, 408*(20), 4613–4623.

Ribolzi, O., Cuny, J., Sengsoulichanh, P., Mousquès, C., Soulileuth, B., Pierret, A., ... Sengtaheuanghoung, O. (2011). Land use and water quality along a Mekong Tributary in Northern Lao PDR. *Environmental management*, *47*(2), 291–302.

Riveros-Iregui, D. A., and McGlynn, B. L. (2009). Landscape structure control on soil CO_2 efflux variability in complex terrain: Scaling from point observations to watershed scale fluxes. *Journal of Geophysical Research*, *114*(G2), G02010.

Robayo, O., and Maidment, D. R. (2005). *Map to map: converting a NEXRAD rainfall map into a flood inundation map*. Center for Research in Water Resources, University of Texas at Austin. Retrieved from http://www.idswater.com/Common/Paper/Paper_176/rtp05-01[2].pdf

Rouse, W. R. (2000). The energy and water balance of high-latitude wetlands: controls and extrapolation. *Global Change Biology*, *6*(S1), 59–68.

Runkel, R. L. (1996). Solution of the advection-dispersion equation: Continuous load of finite duration. *Journal of Environmental Engineering*, *122*(9), 830.

Runkel, R. L., and Bencala, K. E. (1995). Transport of reacting solutes in rivers and streams. In *Environmental hydrology* (pp. 137–164). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-94-017-1439-6_5

Saget, A., Chebbo, G., and Bertrand-Krajewski, J.-L. (1996). The first flush in sewer systems. *Water Science and Technology*, *33*(9), 101–108.

Salas, J. D. (1980). *Applied modeling of hydrologic times series*. Water Resources Publication.

Saleh, A., and Du, B. (2004). Evaluation of SWAT and HSPF within BASINS program for the upper North Bosque River watershed in central Texas. *Transactions of the ASAE*, *47*(4), 1039–1049.

Saltelli, A., Chan, K., and Scott, E. M. (2000). *Sensitivity analysis* (Vol. 134). Wiley New York.

Saltelli, A., Tarantola, S., Campolongo, F., and Ratto, M. (2004). *Sensitivity analysis in practice: a guide to assessing scientific models*. John Wiley and Sons.

Sarrut, D., and Guigues, L. (2008). Region-oriented CT image representation for reducing computing time of Monte Carlo simulations. *Medical physics*, *35*, 1452.

Savenije, H. H. G. (2004). The importance of interception and why we should delete the term evapotranspiration from our vocabulary. *Hydrological Processes*, *18*(8), 1507–1511. doi:10.1002/hyp.5563

Schueler, T., and Yousef, Y. L. (1994). Pollutant dynamics of pond muck. *Watershed Protection Techniques*, *1*(2), 39–46.

Seydel, R. (2012). *Tools for computational finance* (5th ed.). Springer.

Shachnovich, Y., Berliner, P. R., and Bar, P. (2008). Rainfall interception and spatial distribution of throughfall in a pine forest planted in an arid zone. *Journal of Hydrology, 349*(1), 168–177.

Sharif, H. O., Sparks, L., Hassan, A. A., Zeitler, J., and Xie, H. (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.

Shi, P.-J., Yuan, Y., Zheng, J., Wang, J.-A., Ge, Y., and Qiu, G.-Y. (2007). The effect of land use/cover change on surface runoff in Shenzhen region, China. *Catena, 69*(1), 31–35.

Singh, J., Knapp, H. V., Arnold, J. G., and Demissie, M. (2005). Hydrological modeling of the iroquois river watershed using HSPF and SWAT1. *JAWRA Journal of the American Water Resources Association, 41*(2), 343–360.

Smith, R. E., and Hebbert, R. H. B. (1979). A Monte Carlo analysis of the hydrologic effects of spatial variability of infiltration. *Water Resources Research, 15*(2), 419–429.

Taebi, A., and Droste, R. L. (2004). Pollution loads in urban runoff and sanitary wastewater. *Science of The Total Environment, 327*(1–3), 175–184. doi:10.1016/j.scitotenv.2003.11.015

Te Chow, V., Maidment, D. R., and Mays, L. W. (1988). *Applied hydrology.* Tata McGraw-Hill Education.

Thomann, R. V., and Mueller, J. A. (1987). *Principles of surface water quality modeling and control.* Harper and Row, Publishers.

Troch, P. A., Paniconi, C., and McLaughlin, D. (2003). Catchment-scale hydrological modeling and data assimilation. *Advances in Water Resources, 26*(2), 131–135. doi:10.1016/S0309-1708(02)00087-8

Tung, Y.-K., and Yen, B.-C. (2006). Hydrosystems engineering uncertainty analysis.

USEPA. (2000). *BASINS Technical Note 6: Estimating Hydrology and Hydraulic Parameters for HSPF* (No. EPA-823-R-99-013). USEPA, Office of Water.

USEPA, (U.S. Environmental Protection Agency). (1996). *Nonpoint Source Pollution: The Nation's Largest Pollution Problem* (No. EPA/841/F-96/004a). Washington, D.C.

Viviroli, D., Dürr, H. H., Messerli, B., Meybeck, M., and Weingartner, R. (2007). Mountains of the world, water towers for humanity: Typology, mapping, and

global significance. *Water Resources Research*, *43*(7), W07447. doi:10.1029/2006WR005653

Vivoni, E. R., Gebremichael, M., Watts, C. J., Bindlish, R., and Jackson, T. J. (2008). Comparison of ground-based and remotely-sensed surface soil moisture estimates over complex terrain during SMEX04. *Remote Sensing of Environment*, *112*(2), 314–325.

Vojinovic, Z., Kecman, V., and Babovic, V. (2003). Hybrid approach for modeling wet weather response in wastewater systems. *Journal of water resources planning and management*, *129*(6), 511–521.

Vorreiter, L., and Hickey, C. (1994). Incidence of the first flush phenomenon in catchments of the Sydney region. *Water Down Under 94: Surface Hydrology and Water Resources Papers; Preprints of Papers*, 359.

Vrugt, J. A., Gupta, H. V., Bouten, W., and Sorooshian, S. (2003). A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research*, *39*(8), 1201.

Vrugt, J. A., Ter Braak, C. J., Clark, M. P., Hyman, J. M., and Robinson, B. A. (2008). Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. *Water Resources Research*, *44*(12).

Wegge, P., Pokheral, C. P., and Jnawali, S. R. (2004). Effects of trapping effort and trap shyness on estimates of tiger abundance from camera trap studies. *Animal Conservation*, *7*(3), 251–256.

Weingartner, R., Viviroli, D., and Schädler, B. (2007). Water resources in mountain regions: a methodological approach to assess the water balance in a highland-lowland-system. *Hydrological Processes*, *21*(5), 578–585. doi:10.1002/hyp.6268

Weltz, M. A., Arslan, A. B., and Lane, L. J. (1992). Hydraulic roughness coefficients for native rangelands. *Journal of Irrigation and Drainage Engineering*, *118*(5), 776–790.

Western, A. W., and Grayson, R. B. (1998). The Tarrawarra data set: Soil moisture patterns, soil characteristics, and hydrological flux measurements. *Water Resources Research*, *34*(10), 2765–2768.

Wilson, D. J., Western, A. W., and Grayson, R. B. (2004). Identifying and quantifying sources of variability in temporal and spatial soil moisture observations. *Water Resources Research*, *40*(2).

Wise, S. (2000). Assessing the quality for hydrological applications of digital elevation models derived from contours. *Hydrological processes*, *14*(11-12), 1909–1929.

Wolock, D. M., Homberger, G. M., and Musgrove, T. J. (1990). Topographic effects on flow path and surface water chemistry of the Llyn Brianne catchments in Wales. *Journal of Hydrology, 115*(1), 243–259.

Wolock, David M., and McCabe, G. J. (1995). Comparison of single and multiple flow direction algorithms for computing topographic parameters in TOPMODEL. *Water Resources Research, 31*(5), 1315–1324.

Wood, E. F., Sivapalan, M., Beven, K., and Band, L. (1988). Effects of spatial variability and scale with implications to hydrologic modeling. *Journal of Hydrology, 102*(1), 29–47.

Wright, T., Tomlinson, J., Schueler, T., Cappiella, K., Kitchell, A., and Hirschman, D. (2006). Direct and indirect impacts of urbanization on wetland quality. *Wetlands and watersheds. Center for Watershed Protection, Elliot City, MD.*

Wu, S., Li, J., and Huang, G. H. (2008). A study on DEM-derived primary topographic attributes for hydrologic applications: sensitivity to elevation data resolution. *Applied Geography, 28*(3), 210–223.

Yeboah-Forson, A. (2007). *Impact of Uncertainty Estimation on Early Assessment of Technology Options in Contaminated Land Management* (Masters Thesis). Eberhard - Karls - Universität Tübingen Germany.

Yuerekli, K., Kurunc, A., and Oeztuerk, F. (2005). Testing the residuals of an ARIMA model on the Cekerek Stream Watershed in Turkey. *Turkish Journal of Engineering and Environmental Sciences, 29*(2), 61–74.

Zaehle, S., Sitch, S., Smith, B., and Hatterman, F. (2005). Effects of parameter uncertainties on the modeling of terrestrial biosphere dynamics. *Global Biogeochemical Cycles, 19*(3).

Zheng, Y., and Keller, A. A. (2008). Stochastic Watershed Water Quality Simulation for TMDL Development–A Case Study in the Newport Bay Watershed1. *JAWRA Journal of the American Water Resources Association, 44*(6), 1397–1410.

Zou, Y., and Christensen, E. R. (2010). Phosphorus speciation and loads in stormwater and CSOs of the MMSD service area, 2000–2008. *Final Rep. to Milwaukee Metropolitan Sewerage District*, 11–036.

APPENDIXES

APPENDIX A

MATHEMATICAL FORMULATION FOR ESTIMATING NONPOINT POLLUTION OVERLAND

Mass loading is the product of the flow rate and the concentration of the pollutant. It is expressed as:

$$\text{Mass Loading, } W = QC \qquad \text{(Eq. 1)}$$

where W [MT^{-1}], Q is the flow rate [L^3T^{-1}], C is the concentration of the effluent [ML^{-3}], T is time and L is length. For overland nonpoint (diffused) sources Q is the overland flow rate [L^3T^{-1}], C is the concentration [ML^{-3}] resulting from the washoff of pollutants during runoff.

Overland flow Routing

In hydrologic modeling, the movement of water on land is often represented by using the saint Venant kinematic equation. The Saint Venant equation the resultant of continuity (conservation of mass) and momentum equations that assumes flow to be one-dimensional.

Continuity equation:

$$\frac{\partial Q}{\partial x} + \frac{\partial A}{\partial t} = q_{LAT} \qquad \text{(Eq. 2)}$$
$$q_{LAT} = i - f$$

Momentum equation:

$$\frac{1}{A} \cdot \frac{\partial Q}{\partial t} + \frac{1}{A} \cdot \frac{\partial}{\partial x}\left(\frac{Q^2}{A}\right) + g\frac{\partial y}{\partial x} - g(S_0 - S_f) = 0 \qquad \text{(Eq. 3)}$$

where A is the cross-sectional area of the flow, y is the flow depth, g is acceleration of gravity, x is distance, S_0 is bottom slope, S_f is friction slope, f is infiltration, i is rainfall intensity, and t is time. The first term in equation 1 describes the behavior of flow in the x direction, the 2nd describes the rate change of the flow's cross-sectional area with respect to time and the last term refers to the net inflow. In equation 3, the first, second and third are respectively known as the local acceleration, convective acceleration and pressure force respectively while the fourth term is the product of the force of gravity and the friction force ($S_0 = S_f$). Simplifying equation 3 by dropping the pressure force and the acceleration terms and equation 2 by neglecting lateral inflows leads to the saint Venant kinematic wave equation, expressed as:

$$\frac{\partial Q}{\partial x} + \frac{\partial A}{\partial t} = 0 \qquad \text{(Eq. 4)}$$

$$S_0 = S_f \tag{Eq. 5}$$

The solution to the pair of equations, equation 4 and 5, is often approximated using the Mannings equation for channel flow.

Mannings equation:

$$Q = \frac{k}{n} AR^{2/3} S_0^{0.5} \tag{Eq. 6}$$

$$R = \frac{A}{P}$$

$$A = by \quad \text{(for rectangular channels)}$$

$$P = b + 2y$$

where n is the Mannings coefficient, R is the hydraulic radius, P is the wetted perimeter, A is the cross-sectional area of the channel, b is the width of the channel, y is the flow depth, and k is a conversion factor: 1 for SI units and 1.489 for English units. Equation 6 can be re-written as:

$$AR^{2/3} = \frac{nQ}{kS_0^{0.5}}$$

$$\frac{A^{5/3}}{P^{2/3}} = \frac{nQ}{kS_0^{0.5}}$$

$$A = \left(\frac{n}{kS_0^{0.5}} \cdot P^{2/3} \right)^{3/5} \cdot Q^{3/5}$$

which can be re-written as:

$$A = \alpha Q^m \tag{Eq. 7}$$

where

$$\alpha = \left(\frac{n}{kS_0^{0.5}} \cdot P^{2/3} \right)^{3/5} \tag{Eq. 8}$$

$$m = \frac{3}{5} \tag{Eq. 9}$$

Overland flows are characterized with shallow flow depths and infinite widths, hence can be considered as wide rectangular channels with a flat bottom (Akan and Houghtalen, 2003). For wide rectangular channel with shallow depths and much wider widths, the hydraulic radius, R is equal to the flow depth, y and equation 8 can be simplified as:

$$\alpha = \left(\frac{n}{kS_0^{0.5}} \cdot b^{2/3} \right)^{3/5} \tag{Eq. 10}$$

Assuming α is constant, differentiating equation 7 becomes:

$$\frac{\partial A}{\partial t} = m\alpha Q^{m-1} \frac{\partial Q}{\partial t} \tag{Eq. 11}$$

Substituting equation 11 into equation 4 yields:

$$\frac{\partial Q}{\partial x} + m\alpha\, Q^{m-1}\frac{\partial Q}{\partial t} = 0 \qquad\qquad \text{(Eq. 12)}$$

The only unknown variable in equation 11 is the overland flow rate, Q. It can be easily solved using numerical and other available methods.

Pollutant Mass Loading

Consider a catchment as an elemental volume (Figure 5-1). The mass of pollutant loading leaving and entering can be described using the conservation of mass or mass balance. The mass balance takes account for all fluxes entering and leaving the control volume.
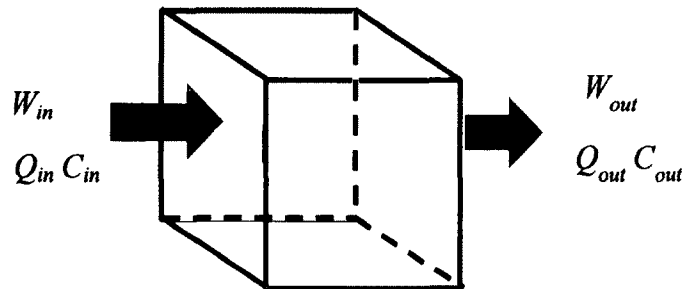
$$\Delta QC = Q_{in}C_{in} - Q_{out}C_{out} \qquad\qquad \text{(Eq. 13)}$$

where ΔW is accumulation, defined as the rate of change of mass with respect to time within the unit volume, Q is the overland flow rate and C is the concentration of the nonpoint source pollutant mass.

Equation 13 can be written as:

$$V\frac{\Delta C}{\Delta t} = Q_{in}C_{in} - Q_{out}C_{out} \qquad\qquad \text{(Eq. 14)}$$

$$\Delta Q = \frac{V}{\Delta t}$$

where V is volume. The mass loading can also be defined in terms of flux. Flux is the amount pollutant mass generated per unit area. Flux is expresses as:

$$J = \frac{W}{A} = \frac{QC}{A} = UC \qquad\qquad \text{(Eq. 15)}$$

$$J = \frac{QC}{A} = UC \qquad\qquad \text{(Eq. 16)}$$

Expressing equation 14 can be expressed in terms flux:

$$V\frac{\Delta C}{\Delta t} = V\frac{\partial C}{\partial t} = J_1 A - J_2 A \qquad\qquad \text{(Eq. 17)]}$$

$$J_2 = J_{1+\Delta x} = J_1 + \frac{\partial J}{\partial x} \Delta x$$

where J_1 and J_2 are fluxes entering and leaving the elemental volume (Figure 5-1). During the course of overland flow, the physical processes that influence the transport and fate of the pollutant mass are advection and dispersion. Advection describes the downstream movement of the pollutant at mean flow velocity (Runkel and Bencala, 1995) while dispersion defines the turbulence of the flow, the mixing and spread of the pollutant as it travels downstream (Runkel and Bencala, 1995; Liu et al., 2003). The fluxes, J_1 and J_2 expressed in terms of the transport mechanisms, advection and dispersion.

Advection:

Fluxes due to advection are defined as the product of the velocity the concentration of the pollutant. From equation 15,

The advective flux entering the unit volume is defined as:

$$J_{1_{advection}} = UC_1 \tag{Eq. 18}$$

The advective flux leaving the unit volume is defined as:

$$J_{2_{advection}} = J_{1+\Delta x} = J_1 + \frac{\partial J}{\partial x} \Delta x$$

$$J_{2_{advection}} = UC_1 + U \frac{\partial C}{\partial x} \Delta x \tag{Eq. 19}$$

Dispersion:

Fluxes due to dispersion in water quality modeling are defined based on the Fick's Law of diffusion. The Fick's Law of diffusion states that the mass flux due to molecular diffusion is proportional to the concentration gradient. The Fick's Law of diffusion:

$$\text{dispersive flux} = -E \frac{\partial C}{\partial x}$$

where E is the longitudinal dispersion coefficient.

The dispersive flux entering the unit volume is expressed as:

$$J_{1_{dispersion}} = -E \frac{\partial C_1}{\partial x} \tag{Eq. 20}$$

The dispersive flux leaving the unit volume is expressed as:

$$J_{2_{dispersion}} = J_{1+\Delta x} = J_1 + \frac{\partial J}{\partial x} \Delta x$$

$$J_{2_{dispersion}} = -E \frac{\partial C_1}{\partial x} - E \frac{\partial^2 C_1}{\partial x^2} \Delta x \tag{Eq. 21}$$

Substituting equation 18, 19, 20 and 21 into equation 17 reads:

$$V \frac{\partial C}{\partial t} = \left[J_{1_{advection}} A + J_{1_{dispersion}} A \right]$$

$$- \left[J_{2_{advection}} A + J_{2_{dispersion}} A \right] \tag{Eq. 22}$$

$$V \frac{\partial C}{\partial t} = \left[UC_1 - E \frac{\partial C_1}{\partial x} \right] A$$

$$- \left[UC_1 + U \frac{\partial C}{\partial x} \Delta x - E \frac{\partial C_1}{\partial x} - E \frac{\partial^2 C}{\partial x^2} \Delta x \right] A \qquad \text{(Eq. 23)}$$

$$V = A\Delta x$$

Dividing each term by $A\Delta x$ yields:

$$\frac{\partial C}{\partial t} = -U \frac{\partial C}{\partial x} + E \frac{\partial^2 C}{\partial x^2} \qquad \text{(Eq. 24)}$$

During the overland flow and the pollutant transport process, pollutants may undergo decay due to chemical reactions or degradation. Such phenomena can be incorporated into equation 24 to account for the fate of pollutant loading overland. The assumption often employed in water quality modeling for such phenomena is the first-order reaction law. It states that, the rate of loss of a substance is directly proportional to the concentration at any time, t. It is expressed as:

$$\frac{\partial C}{\partial t} = -kC \qquad \text{(Eq. 25)}$$

where k is the reaction constant or decay rate $[T^{-1}]$. For conservative pollutants, $k = 0$

Using the boundary conditions,

$$C = C_0 \quad \text{at time } t = 0$$

Equation 25 can be written as:

$$\ln C - \ln C_0 = -kt$$
$$C = C_0 e^{-kt}$$

Equation defines the exponential depletion or decay of the pollutant.
Substituting equation 25 into equation 24 yields:

$$\frac{\partial C}{\partial t} = -U \frac{\partial C}{\partial x} + E \frac{\partial^2 C}{\partial x^2} - kC \qquad \text{(Eq. 26)}$$

Equation 26 is the classical advection-dispersion equation used in quality modeling. Equation 26 is often used for in-stream pollutant routing where the variable C is the concentration for point sources. Nonpoint sources pollutants are subsequently added on as contributing sources. In this context, equation 26 is adapted to estimate nonpoint source pollutants overland. The equation describes the spatial and temporal variation of overland nonpoint sources pollution transport over time and space.

The spatial and temporal characteristics of diffused sources can be classified as continuous or instantaneous loading depending on the duration. Instantaneous loadings are over a very short time period (Thomann and Mueller, 1987) while continuous loading introduce pollutants to a receiving body of water for extended period of time (Runkel and Bencala, 1995) an example is continuous loading from waste water treatment plant into a body of water.

In this study, the durational characteristics of continuous loading are extended to the temporal characteristics of surface runoff. For a storm duration, τ, overland flow begins at time, $t = 0$ and ends at time $t > \tau$. Consequently, the mass loading begins at

time, $t = 0$ and ends at time $t > \tau$. The solution to equation 26 for time variable nonpoint source loading can be obtained by using the boundary conditions:

$$C(x, 0) = 0 \quad \text{for} \quad x \geq 0$$
$$C(0, t) = C_0 \quad \text{for} \quad \tau \geq t \geq 0$$
$$C(x, t) = 0 \quad \text{for} \quad t \geq \tau$$
$$C(\infty, t) = 0 \quad \text{for} \quad t > 0$$

The solution in the final form is given by O'Loughlin and Bowmer (1975) and Runkel (1996).

O'Loughlin and Bowmer (1975):

$$C(x, t) = \frac{C_0}{2} \left[e^{\frac{Ux}{2E}(1-\Gamma)} \text{efrc}\left(\frac{x - Ut\Gamma}{2\sqrt{Et}}\right) + e^{\frac{Ux}{2E}(1-\Gamma)} \text{efrc}\left(\frac{x + Ut\Gamma}{2\sqrt{Et}}\right) \right] \qquad \text{(Eq. 27)}$$

Runkel (1996):

$$C(x, t) = \frac{C_0}{2} \left\{ e^{\frac{Ux}{2E}(1-\Gamma)} \left[\text{efrc}\left(\frac{x - Ut\Gamma}{2\sqrt{Et}}\right) - \text{efrc}\left(\frac{x - U(t-\tau)\Gamma}{2\sqrt{E(t-\tau)}}\right) \right] \right.$$

$$+ e^{\frac{Ux}{2E}(1+\Gamma)} \left[\text{efrc}\left(\frac{x + Ut\Gamma}{2\sqrt{Et}}\right) \right.$$

$$\left. \left. - \text{efrc}\left(\frac{x + U(t-\tau)\Gamma}{2\sqrt{E(t-\tau)}}\right) \right] \right\} \qquad \text{(Eq. 28)}$$

$$\Gamma = \sqrt{1 + 4\frac{kE}{U^2}}$$

$$C_0 = \frac{W}{Q} \qquad \text{(Eq. 29)}$$

where C is the concentration of the pollutant, x is distance traveled, τ is the rainfall duration, t is the time, C_0 is the initial concentration, U is the advection coefficient (mean velocity) and E is the longitudinal dispersion coefficient. The solution during the rainfall, $t < \tau$ is given in equation 27 and after the rainfall, $t > \tau$ is given by equation 28. A combination of equation 27 and 28 superpositions nonpoint loading over time and space during and after a rainfall event.

APPENDIX A

MAIN MODELING FRAMEWORK
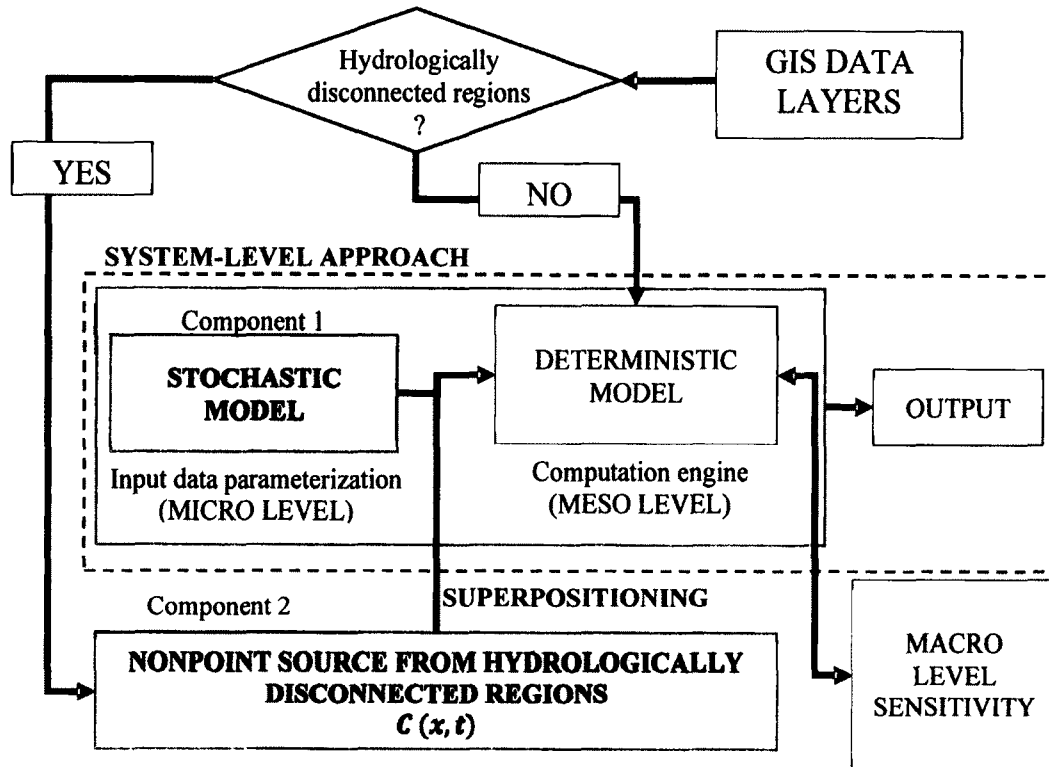
Component 1: Stochastic Model
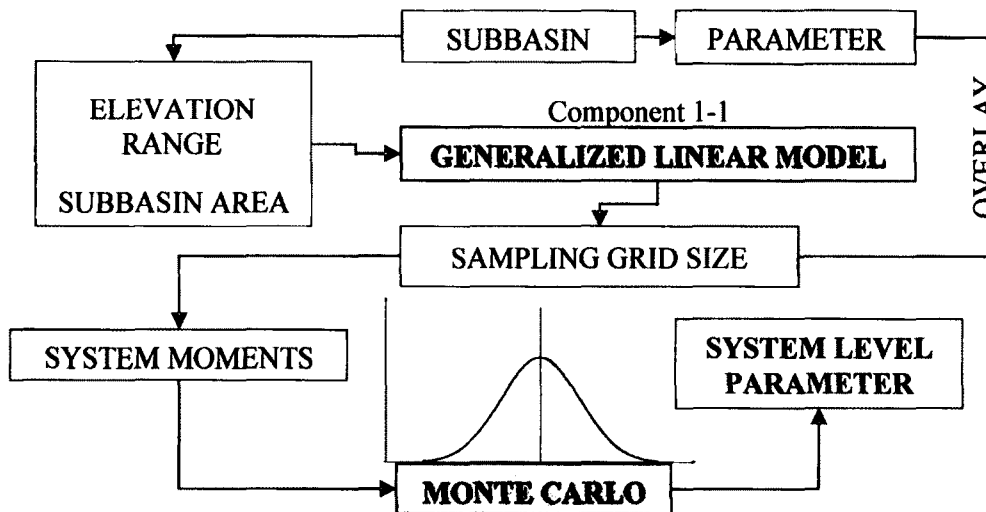
Component 1-1: Generalized Linear Model (GLM) Development

SURFACE AREA OF SUBBASINS S_{AREA}	ELEVATION RANGE *ElevRange*

GLM

INDEPENDENT VARIABLE

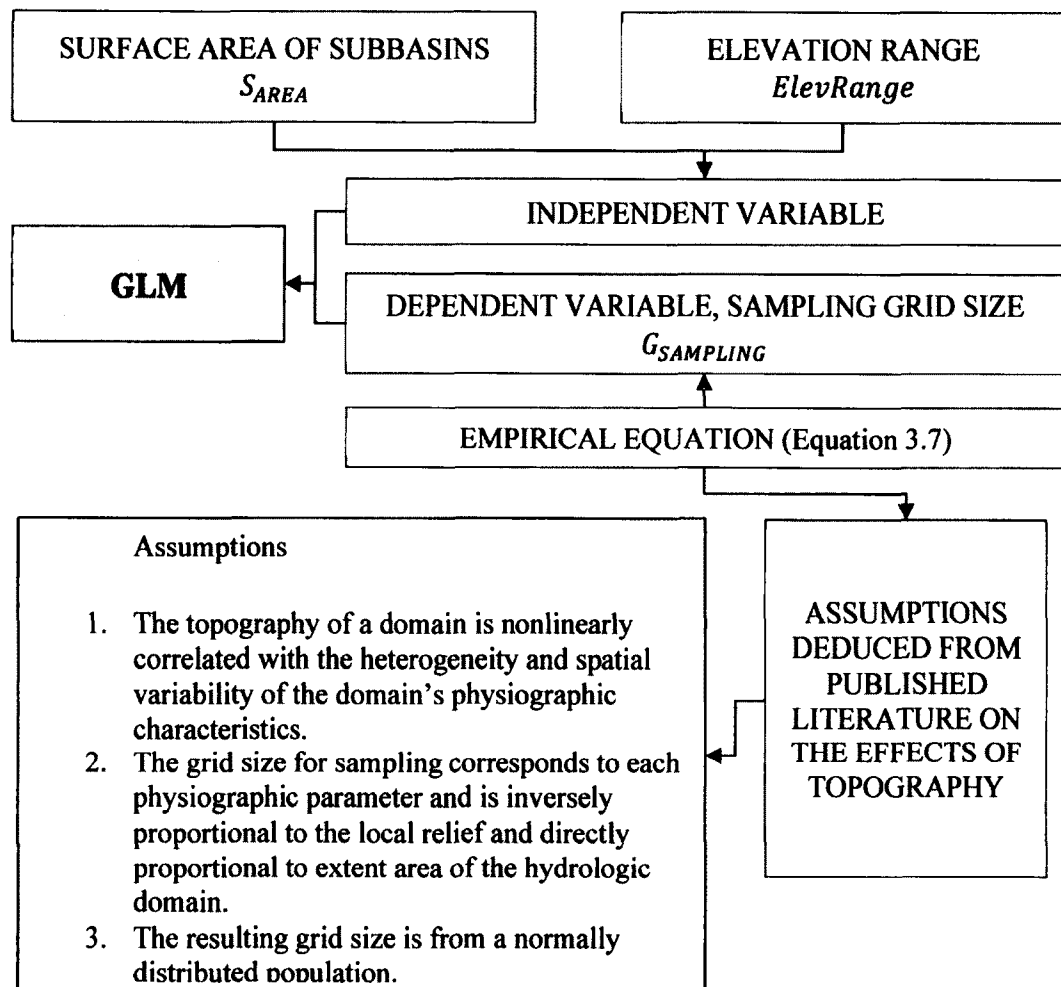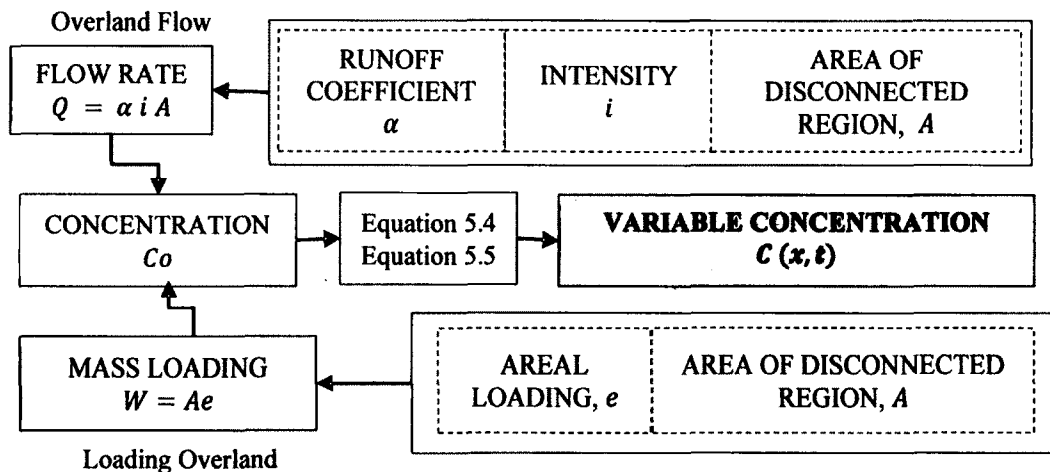DEPENDENT VARIABLE, SAMPLING GRID SIZE $G_{SAMPLING}$

EMPIRICAL EQUATION (Equation 3.7)

Assumptions

1. The topography of a domain is nonlinearly correlated with the heterogeneity and spatial variability of the domain's physiographic characteristics.
2. The grid size for sampling corresponds to each physiographic parameter and is inversely proportional to the local relief and directly proportional to extent area of the hydrologic domain.
3. The resulting grid size is from a normally distributed population.

ASSUMPTIONS DEDUCED FROM PUBLISHED LITERATURE ON THE EFFECTS OF TOPOGRAPHY

Component 2: Nonpoint Source from Hydrologically Disconnected Regions

Overland Flow

FLOW RATE $Q = \alpha i A$	RUNOFF COEFFICIENT α	INTENSITY i	AREA OF DISCONNECTED REGION, A

CONCENTRATION Co	Equation 5.4 Equation 5.5	**VARIABLE CONCENTRATION** $C(x,t)$

MASS LOADING $W = Ae$	AREAL LOADING, e	AREA OF DISCONNECTED REGION, A

Loading Overland

VITA

RUBY JUVAH DAMALIE

EDUCATION

2013	Doctoral Candidate, Civil Engineering Old Dominion University Norfolk, Virginia *Research: Watershed-Scale Hybrid Stochastic-Deterministic Modeling Framework and Diffused Sources Superpositioning*
2008	M.S Water Resources Engineering and Management Kwame Nkrumah Univ. of Science and Technology Ghana *Research: Capacity Estimation of Small Reservoirs Using Remotely Sensed Data: A Case Study of Upper East Region of Ghana*
2005	B.S. Geomatic Engineering Kwame Nkrumah Univ. of Science and Technology, Ghana

PUBLICATIONS PRESENTATIONS AND ABSTRACTS

Damalie, R. and Yoon, J. (2013). Stability of Stochastic Estimation of Spatial Parameters in Extreme Value Distribution. Abstract H33B-11. Meeting of the Americas 2013 AGU, Cancun, Mexico. 3–7 Dec.

Damalie, R. and Yoon, J. (2012). Optimum Grid Size for Hydrological Modeling. Abstract H41G-1254 presented at 2012 Fall Meeting, AGU, San Francisco, Calif., 3–7 Dec.

Shahvari, A. **Damalie, R.** and Yoon, J. (2011). Numerical Dispersive Flux Modeling Study for RO Brine Waste Discharge under Tidal Currents, Water Conference 2011, Virginia Lake and Watershed Association (VLWA), Richmond, March 9-11, 2011.

Coles, K., Shahvari A., **Damalie, R.** and Yoon, J. (2010). Modeling a Dissolved Oxygen-Total Dissolved Nitrogen Relationship in the lower Reach of the Elizabeth River using Finite Segment Method, Water Conference (2010). Virginia Lake and Watershed Association (VLWA), Session 1D: Water Quality, Richmond, March 8-9, 2010.

Damalie, R. J., Odai, S. N., and Annor, F. O. (2008). Capacity Estimation of Small Reservoirs Using Remotely Sensed Data: A Case Study of Upper East Region of Ghana. In AGU Fall Meeting Abstracts (Vol. 1, p. 0875).