DISSERTATION

CONFRONTING THE NATURAL VARIABILITY AND MODELING UNCERTAINTY OF NONPOINT SOURCE POLLUTION IN WATER QUALITY MANAGEMENT

Submitted by

Ali Tasdighi

Department of Civil and Environmental Engineering

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2017

Doctoral Committee:

Advisor: Mazdak Arabi

Brian Bledsoe
Ryan Bailey
Dana Hoag

ABSTRACT

CONFRONTING THE NATURAL VARIABILITY AND MODELING UNCERTAINTY OF NONPOINT SOURCE POLLUTION IN WATER QUALITY MANAGEMENT

Nonpoint source pollution is the primary cause of impaired water bodies in the United States and around the world. Hence, managing the water quality is hinged mainly on controlling this type of pollution. However, characterization of nonpoint source pollution is extremely difficult due to high inherent natural variability and uncertainty. Nonpoint source pollution loads depend on climate, land use, and other environmental conditions that are highly variable by nature. On the other hand, since it is often infeasible to measure pollutant loads from nonpoint sources within a watershed using monitoring campaigns, models are increasingly used to estimate these loads. Models are simplified representations of reality. Consequently, they are subject to various sources of uncertainty including: model parameters, input data (climate, land use, etc.), model structure (conceptualization), and measurement data (streamflow, nutrient concentrations or loads, etc.).

The overarching goal of this dissertation is to characterize the natural variability and modeling uncertainty of nonpoint source pollution and probabilistically quantify the water quality benefits of conservation practices. To achieve this goal, first the relationship between land use and stream water quality was explored under various climatic conditions using multiple linear regression models. This analysis showed that strong and significant relationships exist between land use and ambient water quality. The strength and significance of these relationships changed with climatic conditions. Higher contribution of nonpoint sources in degrading water quality during the wet climate conditions was notable. Second, various sources of modeling uncertainties were characterized in simulating the hydrologic budgets specifically streamflow regimes, for various spatial scales and upstream land use

conditions. The results of this analysis highlights important implications for the selection and application of appropriate rainfall-runoff methods within complex distributed hydrologic models, particularly when simulating hydrologic responses in mixed-land use watersheds. Third, a total uncertainty estimation framework was developed to assess the effectiveness of conservation practices in reducing nonpoint source pollution. The Bayesian-based framework entails a two-stage procedure. First, various sources of modeling uncertainties are characterized during the period before implementing Best Management Practices (BMPs). Second, the effectiveness of the BMPs are probabilistically quantified during the post-BMP period. Results indicate that the modeling uncertainties in quantifying the effectiveness of BMPs vary based on hydrologic conditions. Higher errors were observed in simulating nonpoint source pollution loads during high flow events. The results of this study have important implications for decision making when models are used for water quality simulation and management.

"What man really needs is not just more knowledge, but more certainty." Bertrand Russell, 1964

ACKNOWLEDGMENTS

It is a great pleasure to thank the people who made this dissertation possible. First and foremost, I would like to express my most sincere gratitude to my advisor, Dr. Mazdak Arabi who has been a mentor for me both academically and non-academically. Without his persistent help, support, and motivation this dissertation would not have been possible. I would also like to thank my exceptional committee members, Dr. Brian Bledsoe, Dr. Ryan Bailey, and Dr. Dana Hoag, for their encouragement and insightful comments. Finally, I would like to thank my friends and colleagues at our research lab whose company, assistance, and friendship was always heartwarming. And last but not least, I have to thank my lovely wife for her patience and encouragement. Without her unconditional support, this dissertation would not have been accomplished.

DEDICATION

This dissertation is dedicated to my dear parents, brother, and wife for their unconditional love, support, and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.1. Overview and research goals

Managing water quality in watersheds is subject to many adversities. Some of these adversities are the result of difficulties in understanding hydrologic and water quality processes. Specifically, characterization of nonpoint source pollution is extremely difficult as they are prone to higher (1) natural variability, and (2) uncertainty, compared to point sources.

First, the nonpoint source pollution loads depend on climate, land use, and other environmental conditions that are highly variable by nature. Understanding the variability of ambient water quality constituents under these various conditions is an essential step in characterizing nonpoint source pollution. In this regard, exploring the relationship between land use and water quality under different climatic conditions is critical due to its significant influence on nonpoint source pollution loads. An important outcome of such analyses is the understanding of relative contributions from different nonpoint sources in the degradation of downstream water quality conditions as well as assessing the vulnerability of water bodies to various nonpoint sources of pollution.

Second, it is very challenging and often infeasible to measure pollutant loads from nonpoint sources within a watershed using monitoring campaigns. Hence, models are increasingly used to estimate nonpoint source pollutant loads. Models are mere representation of reality. Consequently, they are subject to various sources of uncertainty including: model parameters, input data (climate, land use, etc.), model structure (conceptualization), and measurement data (streamflow, nutrient concentrations or loads, etc.). The uncertain nature of models affects their predictive performance which has to be taken into account

when they are used for decision making purposes. Lack of accounting for any sources of modeling uncertainties during hydrologic and water quality simulations can lead to biased and misleading results.

The modeling uncertainties are manifested in both hydrologic and water quality simulations. While uncertainties in hydrologic simulations have been explored in the literature more or less, the effects of modeling uncertainties on water quality simulations have not been investigated sufficiently. This is while the uncertainties in simulating water quality are much higher compared to hydrologic simulations. Specifically, modeling uncertainties in assessment of water quality benefits of conservation practices have not been explored.

Implementation of nonpoint source pollution conservation practices (Best Management Practices, BMPs) is a common approach for water quality management in watersheds. The uncertainties from hydrologic and water quality simulations propagate to quantification of water quality benefits of BMPs when watershed models are employed to simulate the effectiveness of BMPs. Yet, almost all studies that investigate the effectiveness of BMPs employ a deterministic approach and disregard all or some sources of uncertainties. Such approach can be inadequate, and in many cases misleading due to lack of accounting for all sources of modeling uncertainties for making water quality management decisions in watersheds.

The overall goal of this study is to explore the variability and uncertainty of nonpoint source pollution and present new approaches for addressing these hindering issues for more realistic and robust water quality management. The results of this dissertation are significant in various aspects of water quality management including characterization of the relationship between land use and water quality under various climatic conditions, estimations of total uncertainties in simulating hydrologic and water quality components, and quantification of the water quality benefits of nonpoint source conservation practices under various sources of uncertainty.

1.2. Background, specific objectives, and propositions

Nonpoint source pollution is the leading cause of water quality problems in the United States and around the world. In the 2000 National Water Quality Inventory, states reported that agricultural nonpoint source pollution is the primary source of water quality impairments in surveyed rivers and lakes. Agricultural activities that cause nonpoint source pollution include poorly located or managed animal feeding operations, improper tillage timing and implementation, and poorly managed application of pesticides, irrigation water, and fertilizers (US EPA, 2008). In addition to agriculture, nonpoint source pollution from urban runoff also plays an important role in polluting water bodies. Section 319 of the Clean Water Act encourages states to address water pollutions by assessing nonpoint source pollution causes and problems and support nonpoint source pollution control programs within the state (US EPA, 1993). However, high levels of natural variability and uncertainty make the characterization of nonpoint source pollution very difficult.

Studies have investigated the relationship between land use and water quality to characterize the variability of nonpoint source pollution from agriculture and urban areas (Osborne and Wiley, 1998; Baker, 2003; Ahearn et al., 2005; Stutter et al., 2007; Williams et al., 2014). The majority of these studies used some statistical measure (annual median or mean) of pollution concentration (Zampella et al., 2007; Williams et al., 2014) and did not account for climatic variations (Bolstad and Swank, 1997; Basnyat et al., 1999; Arheimer and Liden, 2000). A remaining key question in characterizing the nonpoint source pollution is whether these relationships are affected by using loads or concentration of pollution and what the effects of climatic variations are. This could be particularly important in the case of nonpoint sources during wet years due to higher volumes of surface runoff.

Application of models in quantification of nonpoint sources is inevitable. Faced with myriad hydrologic and water quality models, scientists ought to select an appropriate model and demonstrate its performance validity for the desired assessments. The performance of watershed models is often

evaluated based on a deterministic approach, which entail calibration, validation, and prediction. The calibration is conducted to obtain a parameter set that provides the best fit between model responses and observations at the outlet of the watershed (Seibert, 1999; Santhi et al., 2008; Niraula et al., 2015). Such assessments can be inadequate, and in many cases misleading for selecting an appropriate distributed watershed model, particularly when the model is used to simulate interior hydrologic and water quality processes or to assess responses at various locations within the watershed (Beven, 2001; Ahmadi et al., 2014). Probabilistic approaches can address the equifinality and nonuniqueness issues in parameter estimation (Moradkhani et al., 2005), input uncertainty (Kavetski et al., 2003), and measurement errors (Harmel et al., 2007) when assessing competing model structures (Ajami et al., 2007). Application of probabilistic approaches in simulating hydrologic and water quality processes which explicitly account for various sources of modeling uncertainty can enhance the quality of simulations and help making more realistic and informed decisions.

Nonpoint source pollution conservation practices are regularly used for improving water quality in watersheds. Yet their effectiveness has been a subject of debate (Arabi et al., 2007; Park et al, 2011). Models are increasingly used along with the monitoring data to better assess the effectiveness of BMPs (Santhi et al., 2003; Arabi et al., 2006; Lin et al., 2009; Park and Roesner, 2012; Taylor et al., 2016; Jang et al., 2017). However, the majority of studies that investigated the effectiveness of BMPs using models have used a deterministic approach (Lin et al., 2009; Ullrich et al., 2009; Liu et al., 2013; Jingyuan et al., 2014; Motallebi et al., 2017). This is while propagation of uncertainties from different sources into model predictions may lead to biased and unrealistic results (Ajami et al., 2007; Arabi et al., 2007; Tasdighi et al., 2017). In any case, the impacts of modeling uncertainty when quantifying the effectiveness of BMPs have not been addressed sufficiently. Specifically, a framework that incorporates various sources of modeling uncertainty and determines uncertainty bounds around BMP effectiveness has not been developed to the best of our knowledge.

1.2.1. Objectives:

The objectives of this study were to:

O1. Characterize the natural variability of nonpoint source pollution from different sources under various land use and climatic conditions. The vulnerability of stream water to nonpoint sources of total nitrogen and phosphorus pollution as a function of land use was further assessed.

O2. Probabilistically compare the performance of a physically-based distributed watershed model in characterizing the hydrologic budget, specifically streamflow regime, under different model structures for various upstream land use conditions.

O3. Develop a total uncertainty estimation framework that accounts for various sources of modeling uncertainty for assessment of the nonpoint source pollution loads and water quality benefits of related conservation practices.

1.2.2. Propositions

Following propositions were explored en route to achieving the objectives of the study:

P1. Characterizing the variability of nonpoint source pollution requires a good understanding of relationships between water quality and natural variable conditions such as land use and climate. Multiple Linear Regression (MLR) models can be used to explore land use-water quality relationships and assess the variability of nonpoint source pollution under changing land use and climatic conditions. Analysis of Covariance (ANCOVA) can be also employed to examine the effects of climate variability on regression models. The vulnerability of water bodies to nonpoint source pollution can then be characterized using the results of the land use-water quality analysis.

P2. Bayesian based uncertainty assessment methods can be used to assess the uncertainties in modeling. Specifically, Bayesian inference coupled with a Markov Chain Monte Carlo (MCMC) sampling algorithm

can be an effective approach for characterizing various sources of modeling uncertainties in simulating the hydrologic and water quality responses. Using a statistically correct likelihood function accounting for streamflow and water quality error heteroscedasticity and autocorrelation is essential when conducting the uncertainty analysis. The probabilistic approach should explicitly account for various sources of modeling uncertainty including model parameters, input, measurements, and structure (conceptualization).

P3. Assessing the uncertainties in effectiveness of nonpoint source conservation practices should not only include uncertainties from model parameters but also parameters pertaining to modeling the BMPs. A two-stage uncertainty analysis framework can be developed for quantification of the effectiveness of nonpoint source pollution conservation practices. The framework should first characterize various sources of modeling uncertainty during a training phase. The inferences from the first stage then should be used to quantify the effectiveness of BMPs under uncertainty. The framework should benefit from an efficient algorithm to reduce the high computational burden intrinsic in analysis and assessment of BMP effectiveness.

1.3. Significance of the dissertation

This dissertation is significant in various aspects of water quality management in watersheds: (i) it presents a guideline for assessing the variability of nonpoint source pollution as a function of land use and climatic changes and provides an approach for quantifying the vulnerability of water bodies to nonpoint source pollution. The findings demonstrate the significant impacts of land use and climate on water quality. It also provides the first of its kind insights into consequences of using concentrations vs. loads of nonpoint sources when the land use water-quality relationships are investigated; (ii) the study provides a probabilistic approach for assessing the performance of watershed models in simulating the hydrologic budget and streamflow as a function of upstream land use conditions under various sources of modeling uncertainty. Application of total uncertainty estimates in characterizing the performance of the watershed models under different model structures based on upstream land use conditions is novel; (iii) it presents a

total uncertainty assessment framework for assessment of the effectiveness of conservation practices in reducing nonpoint source pollution. The framework is the pioneer in assessment of conservation practices under uncertainty and can significantly enhance the decision making capabilities by giving valuable insights into the probabilistic characteristics of conservation practices.

1.4. Organization of the dissertation

This dissertation is organized in five chapters. The first chapter provides a holistic introduction to the research study. The second chapter elaborates the land use-water quality analysis for assessment of the nonpoint source pollution variability under changing land use and climate and vulnerability of water bodies to nonpoint source pollution. The third chapter of the dissertation presents a total uncertainty assessment approach for characterizing the hydrologic budget and specifically streamflow regime under uncertainty and various upstream land use conditions. The fourth chapter demonstrates a framework for estimating the water quality benefits of nonpoint source pollution conservation practices under various sources of modeling uncertainty. The fifth and final chapter of the dissertation provides summary, conclusions, and recommendations for future work.

Bibliography

Ahearn, D.S., R.W. Sheibley, R.A. Dahlgren, M. Anderson, J. Johnson, and K.W. Tate. 2005. Land use and land cover influence on water quality in the last free-flowing river draining the western Sierra Nevada, California. J. Hydrol. 313:234–247. doi:10.1016/j.jhydrol.2005.02.038.

Ahmadi, M., M. Arabi, J. Ascough, D. G. Fontane, B. A. Engel. 2014. Toward improved calibration of watershed models: Multisite multiobjective measures of information. J. Environ. Model. Softw. 59:135-145. doi: 10.1016/j.envsoft.2014.05.012.

Ajami, N.K., Duan, Q., Sorooshian, S., 2007. An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction. Water Resources Res. 43(1), Art. No. W01403.

Arabi, M., Govindaraju, R.S., Hantush, M.M., 2007. A probabilistic approach for analysis of uncertainty in the evaluation of watershed management practices. J. Hydrol. 333, 459-471.

Arheimer, B., and R. Liden. 2000. Nitrogen and phosphorus concentrations from agricultural catchments: Influence of spatial and temporal variables. J. Hydrol. 227:140–159. doi:10.1016/S0022-1694(99)00177-8.

Arnold, J.G., D.N. Moriasi, P.W. Gassman, K.C. Abbaspour, M.J. White, R. Srinivasan, C. Santhi, R.D. Harmel, A. van Griensven, M.W. Van Liew, N. Kannan, and M.K. Jha. 2012. SWAT: Model use, calibration, and validation. Trans ASABE. 55:1491-1508. doi: 10.13031/2013.42256.

Baker, A. 2003. Land use and water quality. Hydrol. Process. 17: 2499–2501. doi: 10.1002/hyp.5140.

Basnyat, P., L.D. Teeter, K.M. Flynn, and B.G. Lockaby. 1999. Relationships between landscape characteristics and nonpoint source pollution inputs to coastal estuaries. J. Environ. Manage. 23:539–549.

Beven, K., 2001. How far can we go in distributed hydrological modelling? Hydrol and Earth Syst. Sci. 5, 1-12.

Bolstad, P.V., and W.T. Swank.1997. Cumulative impacts of land use on water quality in a southern Appalachian watershed. J. Am. Water Resour. Assoc. 33:519–533. doi: 10.1111/j.1752-1688.1997.tb03529.x.

Harmel, R.D., Smith, P.K., 2007. Consideration of measurement uncertainty in the evaluation of goodness-of-fit in hydrologic and water quality modeling. J. Hydrol. 337(3-4), 326-336.

Jang, S.S., Ahn, S.R., Kim, A.J., 2017. Evaluation of executable best management practices in Haean highland agricultural catchment of South Korea using SWAT. J. Agri. Water Manag. 180(B), 224-234.

Jingyuan, J., Shiyu, L., Hu, J., Huang, J., 2014. A modeling approach to evaluating the impacts of policy-induced land management practices on non-point source pollution: A case study of the Liuxi River watershed, China. Agri. Water Manag. 131, 1-16.

Kavetski, D., Franks, S.W., Kuczera, G., 2003. Confronting input uncertainty in environmental modeling, in: Duan, Q., Gupta, H.V., Sorooshian, S., Rousseau, A.N., Turcotte, R. (Eds.), Calibration of Watershed Models. Water Sci. Appl. Ser., AGU, Washington, D.C., pp. 49-68.

Lin, Z., Radcliffe, D.E., Risse, L.M., Romeis, J.J., Jackson, C.R.. 2009. Modeling Phosphorus in the Lake Allatoona Watershed Using SWAT: II. Effect of Land Use Change. J. Environ. Qual. 38:121-129.

Liu, R., Zhang, P., Wang, X., Chen, Y., Shen, Z., 2013. Assessment of effects of best management practices on agricultural non-point source pollution in Xiangxi River watershed. Agri. Water Manag. 117, 9-18.

Moradkhani, H., Hsu, K.L., Gupta, H.V., Sorooshian, S., 2005. Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter, Water Resources Res. 41, W05012.

Motallebi, M., Hoag, D.L., Tasdighi, A., Arabi, M., Osmond, D., 2017. An economic inquisition of water quality trading programs, with case study of Jordan Lake, NC. J. Environ. Manage. 193, 483-490.

Niraula, R., Meixner, T., Norman, L., 2015. Determining the importance of model calibration for forecasting absolute/relative changes in streamflow from LULC and climate changes. J Hydrol. 522, 439-451.

Osborne, L.L., and M.J. Wiley. 1988. Empirical relationships between land-use cover and stream water-quality in an agricultural watershed. J. Environ. Manage. 26:9–27.

Park, D. Loftis, J.C. Roesner, L.A., 2011. Modeling performance of storm water best management practices with uncertainty analysis. J. Hydrologic Engineering. 16 (4), 332–344.

Santhi, C., Kannan, N., Arnold, J.G., Di Luzio, M., 2008. Spatial calibration and temporal validation of flow for regional scale hydrologic modeling. J. Am. Water Resources Assoc. 44(4), 829-846.

Santhi, C., Srinivasan, R., Arnold, J.G., Williams, J.R., 2003. A modeling approach to evaluate the impact of water quality management plans implemented in the Big Cypress Creek Watershed. In: Second conference on watershed management to meet emerging TMDL environmental regulations, Albuquerque, NM, pp. 384-394.

Seibert, J., 1999. Regionalisation of parameters for a conceptual rainfall-runoff model. J. Agric. For. Meteorol. 98/99, 279-293.

Stutter, M.I., S.J. Langan, and B.O. Demars. 2007. River sediments provide a link between catchment pressures and ecological status in a mixed land use Scottish River system. Water Res. 41:2803–2815. doi:10.1016/j.watres.2007.03.006.

SWAT literature database. 2016. Available at: (https://www.card.iastate.edu/swat_articles/).

Tasdighi, A., Arabi, M., Osmond, D.L., 2017. The relationship between land use and vulnerability to nitrogen and phosphorus pollution in an urban watershed. J. Environ. Qual. 46, 113-122.

Taylor, S.D., He, Y., Hiscock, K.M., 2016. Modelling the impacts of agricultural management practices on river water quality in Eastern England. J. Environ. Manag. 180(15), 147-163.

Ullrich, A., Volk, M., 2009. Application of the Soil and Water Assessment Tool (SWAT) to predict the impact of alternative management practices on water quality and quantity. Agri. Water Manag. 96, 1207-1217.

U.S. EPA. 1993. Guidance specifying management measures for sources of nonpoint pollution in coastal waters, EPA 840-B-92-002, U.S. Environmental Protection Agency, Office of water. Wasgington, DC.

U.S. EPA, 2008. Handbook of developing watershed plans to restore and protect our waters, U.S. Environmental Protection Agency, Office of water. Washington, DC.

Williams, R., M. Arabi, J. Loftis, and G.K. Elmund. 2014. Monitoring design for assessing compliance with numeric nutrient standards for rivers and streams using geospatial variables. J. Environ. Qual. 43:1713–1724. doi:10.2134/jeq2013.12.0528.

Zampella, R.A., N.A. Procopio, R.G. Lathrop, and C.L. Dow. 2007. Relationship of land-use/land-cover patterns and surface-water quality in the Mullica River Basin. J. Am. Water Resour. Assoc. 43:594–604. doi: 10.1111/j.1752-1688.2007.00045.x.

Chapter 2

THE RELATIONSHIP BETWEEN LAND USE AND VULNERABILITY TO NITROGEN AND PHOSPHOROUS POLLUTION IN AN URBANIZING WATERSHED

Highlights

Characterization of the vulnerability of water bodies to pollution from natural and anthropogenic sources requires understanding the relationship between land use and water quality. This study aims to: (i) explore the influence of upstream land use on annual stream water concentrations and loads of total nitrogen (TN) and phosphorus (TP) and (ii) characterize the vulnerability of water bodies to TN and TP pollution as a function of land use under varying climatic conditions. Multiple linear regression (MLR) models were used across 23 stream locations within Jordan Lake watershed in North Carolina between 1992 and 2012 to explore land use-water quality relationships. The percentage of urban land use and wastewater treatment plants capacity were the most important factors with strong ($R^2 > 0.7$) and significant ($p < 0.01$) positive correlations with annual TN and TP concentrations and loads. Percent agricultural land was negatively correlated with TN in 18 out of 21 years of the study period. Using analysis of covariance, significant ($p < 0.01$) differences were determined between models developed for urban land use with TN and TP loads based on annual precipitation. Using concentrations instead of loads resulted in nonsignificant difference between models for average and wet years. Finally, a procedure was developed to characterize the vulnerability to TN and TP pollution computed as probability of exceeding the nutrient standard limits. Results indicated that the vulnerability to TN and TP was controlled primarily by urban land use, with higher values in dry years than normal and wet years.

2.1. Introduction

The study of relationships between land use and water quality is essential in exploring the vulnerability of water bodies to a variety of pollution sources. Investigation of land use-water quality relationships is particularly useful in the case of pollution from diffuse urban and agricultural sources (Baker, 2003). An important outcome of such analyses is the understanding of relative contributions from different nonpoint sources in the degradation of downstream water quality conditions.

Regression-based models have been used in many studies for assessing the relationship between land use types and stream water quality. The majority of these studies conclude that strong and significant correlations exist between water quality parameters and different land use types (Mehaffey et al., 2005; Stutter et al., 2007; Williams et al., 2014). Some studies concluded that percentage of agriculture land use has a strong influence on nitrogen (N) levels in stream water (Johnson et al., 1997; Smart et al., 1998). Similar results were obtained for phosphorus (P) (Hill, 1981; Arheimer and Liden, 2000) and sediments (Johnson et al., 1997; Allan et al., 1997). In contrast, some studies have demonstrated that stream water nutrient levels were a function of percentage of urban land use (Osborne and Wiley, 1998; Sliva and Williams, 2001). While these studies agree that anthropogenic activities (e.g. urban development, agriculture) have a negative impact on stream water quality, there is not a consensus about which activity (urban development or agriculture) has the dominant impact in a mixed land use watershed.

The majority of the studies that have used regression-based models for linking land use to stream water total nitrogen (TN) and total phosphorus (TP) levels have used some statistical measure (annual median or mean) of concentration (Zampella et al., 2007; Williams et al., 2014). Some studies have used flow-weighted concentrations or loads of nutrients (Jordan et al., 1997; Ahearn et al., 2005). A remaining key question in unveiling the relationship between land use and water quality is whether these relationships are affected by using load or concentration measurements/calculations. This could be particularly important in the case of nonpoint sources during wet years due to higher volumes of surface runoff.

Climate is a key factor in the variability of nutrient loadings from nonpoint sources. Hence, deriving the land use-water quality relationships for a small period of time with low variability of climatic conditions can lead to misleading and biased results. Several studies have performed their analysis only for a specific season arguing that maximum interaction between land use and stream water occurs during wet seasons (Bolstad and Swank, 1997; Basnyat et al., 1999; Arheimer and Liden, 2000). Other studies have conducted their analysis for different seasons deriving separate correlations (Osborne and Wiley, 1988; Johnson et al., 1997). While the seasonality aspect of the analysis has been addressed in some studies, the effects of inter-annual climate variations (wet/dry years) still remain largely unclear. This is particularly important in regions where climate variability is substantial.

Vulnerability to water quality pollution is a function of ambient contaminant levels (loads or concentrations) as well as the assimilative capacity of the water body. Nutrient targets are developed to maintain/restore physical, chemical and biological integrity of the water bodies. Hence, a proper characterization of vulnerability to nutrients must incorporate the relationship between land use and ambient water quality, and the likelihood of exceedance of desired standards under varying climatic conditions. Williams et al. (2014) developed a statistically rigorous procedure for the determination of sample size requirements for assessing compliance/noncompliance with desired in-stream TN and TP standards. However, the vulnerability to exceedance of nutrient targets with changing land use has not been examined.

Jordan Lake reservoir in North Carolina has been declared as hyper-eutrophic since its impoundment due to receiving high loads of N and P (NCDENR, 2007). Reduction goals were set for TN and TP delivered to the lake by North Carolina Department of Environment and Natural Resources (NCDENR; NCDENR, 2007). Various point and nonpoint sources of nutrients (urban development and agriculture) are required to reduce their loads in order to meet the reduction goals. In this regard, understanding land use-water quality relationships and vulnerability of water bodies to TN and TP as a function of land use can be beneficial in selecting TN and TP standard goals.

This study investigates the relationship between components of land use and stream water TN and TP levels in the Jordan Lake watershed under varying climatic conditions. The specific objectives are to: (1) identify the variability of annual TN and TP levels in streams in response to spatial and temporal response changes in land use and climate influences, (2) determine the effects of using TN and TP calculated loads versus concentration measurements on land use-water quality relationships, (3) explore the effects of inter-annual climate variations (i.e. precipitation) on the land use-water quality relationships, and (4) determine the vulnerability to exceeding ambient TN and TP targets as a function of land use and climate variability.

While a number of previous studies have examined the relationship between land use and water quality, the characterization of these relationships under varying climatic conditions using both loads and concentrations of TN and TP is a major contribution of this study. More importantly, a rigorous statistical procedure is used to characterize how vulnerability to TN and TP pollution changes with varying percentage of urban land use.

2.2. Materials and methods

2.2.1. Study watershed

The Jordan Lake watershed, located in the central part of North Carolina within the Piedmont region, drains an area of 4367 km^2. The watershed is comprised of three subbasins: Haw, Upper New Hope, and Lower New Hope covering 80%, 13% and 7% of the total watershed area respectively (Fig. 2.1). The Jordan Lake watershed has a land use composition of 46% forest, 21% urban/suburban and 22% agriculture, of which more than 90% is pasture (National Land Cover Database; NLCD, 2011). Upper New Hope is heavily urbanized, while Lower New Hope is being rapidly developed at suburban residential densities. Fourteen wastewater treatment plants (WWTP) with individual mean annual effluents larger than 0.05 MGD, discharge into the main streams and tributaries of Jordan Lake watershed. There are 32 permitted animal feeding operations (AFOs) within this watershed (Fig. 2.1).

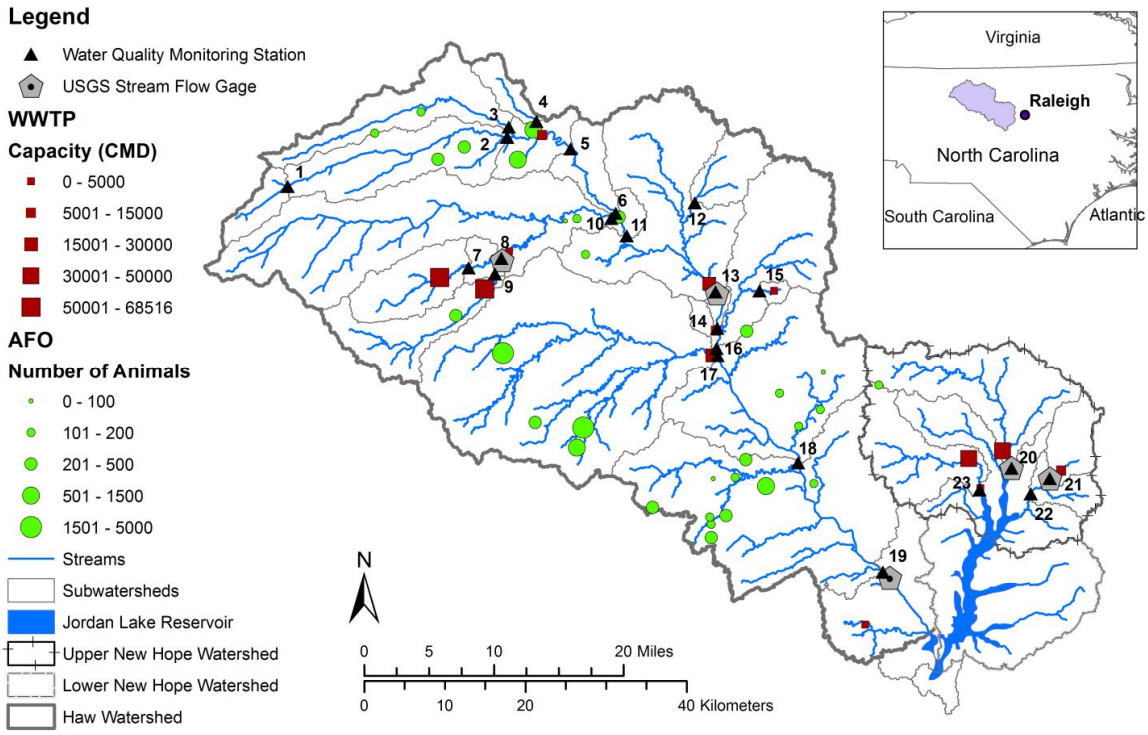

Fig. 2.1. Map of Jordan Lake watershed with names and locations of the 23 water quality monitoring stations and their corresponding subbasin boundaries, wastewater treatment plants (WWTPs), animal feeding operations (AFOs) and stream flow gages.

While there are no distinct wet or dry seasons in Jordan Lake watershed, average rainfall varies throughout the year. Summer precipitation is normally the highest and autumn is the driest season. The average yearly precipitation is 1057 mm. From 1992 through 2012, 2001 was the driest year (853 mm) and 2003 had the highest precipitation (1815 mm) (National Oceanic and Atmospheric Administration; NOAA, 2013).

The watershed was divided into 23 subbasins based on location of water quality monitoring stations and including a range of different characteristics such as: size, land use, WWTP and AFO capacities. Five out of the 23 water quality monitoring stations had streamflow measurements (USGS gages; Fig. 2.1).

2.2.2. Geospatial analysis

ArcSWAT 2012 (US Department of Agriculture-Agricultural Research Service, 2014) was used to delineate subbasins and stream networks in the study area using 1/3 arc-sec (10 meter) Digital Elevation Model (DEM) of the watershed and high resolution National Hydrography Dataset (NHD) from US Geological Survey (USGS). Outlets of the subbasins were selected at water quality monitoring stations. For each subbasin, land use compositions, WWTP and AFO capacities were quantified (Fig. 2.1). The NLCD 1992, 2001, 2006, and 2011 were used as land use data. The data for WWTPs and AFOs were collected from NCDENR (NCDENR, Pers. Comm., 2013). The AFOs in Jordan Lake watershed are mostly cattle and/or swine operations. For consistency in counting the animals across the subbasins, the number of swine was converted into the equivalent number of cattle based on average weight of the animal (average live cattle weight ~ 5.5 × average live swine weight; USDA ERS, 2014).

2.2.3. Water quality data

The Jordan Lake watershed has been monitored extensively since impoundment of Jordan Lake in 1980s. Water quality data is primarily collected by NCDENR (Division of Water Resources Ambient Monitoring System, AMS) and is stored on EPA STORET database available for public access (NCDENR, 2013).

Nutrient (Nitrate plus nitrite (referred to as inorganic N hereafter), total Kjeldhal N (TKN), and TP) concentrations were collected from EPA STORET for 23 sites shown in Fig. 2.1. Total nitrogen was computed by adding inorganic N and TKN. The 1992-2012 period was selected for the analysis based on highest sampling frequencies and available land use data. This time frame also covers a wide range of climatic variation including dry, average, and wet years. While the majority of the stations had at least one observation per constituent per month, some stations had bi-weekly data. The bi-weekly/monthly observations were used to generate average annual concentrations.

2.2.4. Hydrologic modeling

Five out of 23 monitoring stations included daily stream discharge data (Fig. 2.1). For the other stations, the SWAT (version 2012) model was used to simulate streamflow. SWAT is a continuous-time, semi-distributed, process-based watershed model which has been used in the literature extensively for simulating hydrologic and water quality processes (Gassman et al., 2007; Arnold et al., 2012, SWAT Literature Database, 2016). The model has the capability to run on daily time step at field to watershed scales and has a reasonable setup and run time.

The model was calibrated and tested for daily streamflow at 5 stations where measurements of streamflow were available. Dynamically Dimensioned Search (DDS; Tolson and Shoemaker, 2007) was used as the calibration engine. DDS is a simple stochastic neighborhood search algorithm used for locating best parameter sets within a user-defined maximum number of model iterations frequently used for calibrating SWAT (Tolson and Shoemaker, 2007; Yen et al., 2014; Ahmadi et al., 2014). Nash-Sutcliff coefficient of efficiency (E_{NS}; Nash and Sutcliff, 1970) and percent bias (*PBIAS*) were used to evaluate the performance of the model.

Relative error (*RE*) and coefficient of determination (R^2) were also used as additional model performance evaluation indices. The model was calibrated at the outlet of the whole watershed for 1992 to 2007 with a 2-year warm-up period (to reach a state of statistical equilibrium before using the model for simulation during target period). The calibrated model was then tested for the 2008 to 2012 period and at several other locations inside the watershed where limited observations were available. Evaluation of the model performance was conducted according to the criteria presented in Motovilov et al. (1999), Engel et al. (2007), and Moriasi et al. (2007). According to these criteria the performance of the model for simulating daily streamflow may be considered good when E_{NS} is larger than 0.75 and *PBIAS* is less than ±10%. Model performance may be considered satisfactory when E_{NS} varies between 0.36 and 0.75 while

PBIAS varies between ±10% and ±25%. The coefficient of E_{NS} less than 0.36 with *PBIAS* larger than ±25% indicates unsatisfactory performance of the model.

2.2.5. Statistical analysis

Average annual concentration of TN and TP were calculated using the bi-weekly/monthly data available for each of the 23 stream locations for the analysis period. The frequencies of non-detect values for some of the stream monitoring locations were between 5 to 30 percent. For stations with non-detect values, regression on order statistics was used to estimate the mean and median of annual concentrations (Helsel, 2005), using ProUCL 5.0 (EPA, 2013; Lockheed Martin IS&GS-CIVIL).

The USGS load estimator model (LOADEST; Runkel et al., 2004) was used to generate monthly and annual TN and TP loads. The inputs to the model are TN and TP concentrations and stream discharges. LOADEST provides options for fitting several regression models to data for estimating constituent loads (USGS, 2013). Adjusted Maximum Likelihood Estimation (AMLE) method provided the best error statistics for the fitted models and hence was selected for load estimation.

2.2.6. Analysis of land use-water quality relationships

Multiple Linear Regression (MLR) models were developed to explore the relationship between land use, and nutrient loads and concentrations:

$$\hat{Y} = \beta X + \varepsilon \quad (1)$$

where X is the matrix of independent variables, $\hat{Y}$ is the response variable, β denotes the vector of MLR coefficients, and ε represents model residuals or errors. Assuming that model residuals are normally distributed with mean zero and unknown but constant standard deviation σ_{ε}, the variance of the model response at a particular point x_0 is:

$$\sigma_{\hat{Y}}^2 = \sigma_{\varepsilon}^2 \, x_0' \, (X'X)^{-1} x_0 \quad (2)$$

In the present study, the independent variables (X) were components of land use including percent urban (% urban), agricultural (% agriculture), and forested (% forest) lands and, anthropogenic factors including capacities of WWTPs and AFOs. The capacity of WWTPs (daily discharge) in million gallons per day (MGD) and AFOs (number of animals) represent cumulative upstream characteristics and increase from upstream to downstream locations. In order to reduce the multicollinearity between anthropogenic factors, the WWTP and AFO capacities in each subbasin were normalized by the area of the subbasin. A similar approach has been used by Williams et al. (2014) in their study. They normalized the AFO and WWTP capacities by distance to reduce the cumulative characteristics and multicollinearity.

Total N and TP concentrations and loads tend to follow lognormal distributions (Kutner et al., 2005; Williams et al., 2014). A log transformation of exploratory and response variables was used prior to regression analysis to reduce deviation from normality. The MLR models were developed in MATLAB (The MathWorks, Inc.® 2012). In order to have valid MLR models several assumptions have to be met (Kutner et al., 2005). These assumptions were tested using different diagnostic statistical tests and the results are included in the results section.

2.2.7. Characterization of the vulnerability to nutrient pollution

The vulnerability to TN and TP pollution was characterized as the probability of exceeding their corresponding standard targets. Assuming the MLR response variable, i.e. log transformed concentrations or loads are normally distributed with mean and variance as in (1) and (2), the vulnerability at a given point x_0 for a desired target S can be written as:

$$Vulnerability = P\left(\hat{Y} > \log(S) \mid x_0\right) = 1 - \phi\left(\frac{\log(S) - \hat{Y}}{\sqrt{\sigma_\varepsilon^2\, x_0'(X'X)^{-1}x_0}}\right) \quad (3)$$

where $\phi(\cdot)$ is the non-exceedance probability of the standard normal variable. Note that P in (3) stands for probability.

2.2.8. The role of climatic variability on the relationship between land use and nutrient levels

Analysis of Covariance (ANCOVA) was used to examine the effects of climate (i.e. precipitation) variability on regression models. Precipitation data for the watershed were collected from NOAA National Climatic Data Center. Years of analysis were categorized into dry, average, and wet based on average annual precipitation and are referred to as climate scenarios. The years with precipitation below a threshold (1000 mm) selected based on the long-term average annual precipitation of the study watershed, were categorized as "dry". The years with precipitation between 1000 mm and 1200 mm were classified as "average" years, and any year with precipitation higher than 1200 mm was categorized as a "wet" year. For the analysis, percent urban and percent agriculture land areas were set as the continuous independent variables and dry/average/wet year as the categorical independent variables. Annual concentrations and loads of TN and TP were selected as the response variables. The ANCOVA model is written as:

$$\hat{Y}_{i,j} = (\mu_i + \hat{Y}) + (\theta_i + \beta)(X_{i,j}) + \varepsilon_{i,j} \quad (4)$$

where $\hat{Y}_{i,j}$ is the j^{th} TN or TP concentration or load for the i^{th} group (dry, average or wet year), $X_{i,j}$ is the j^{th} land use (covariate) for the i^{th} group in percent, $\hat{Y}$ and β are the model global mean and slope and μ and θ are the correction terms for the model mean and slope respectively. The constraints for terms μ and θ in the ANCOVA model are $\sum \mu_i = \sum \theta_i = 0$. The term ε is the error term which is assumed to be normally distributed with mean zero and unknown but constant standard deviation of σ_ε; i.e. $\varepsilon_{i,j} \sim N(0, \sigma_\varepsilon^2)$. After the models were developed, a multiple comparison test was performed to determine the significance of the difference between the regression models.

2.3. Results and discussion

2.3.1. Calibration and testing of the hydrologic model

The calibration of the SWAT2012 model developed for simulating daily streamflow at subbasin outlets had very good results in terms of error statistics and model performance factors such as E_{NS},

PBIAS, *RE* and R^2. The calibration of the model for 1992 to 2007 at the outlet of the watershed resulted in E_{NS} = 0.76, R^2 = 0.86, *RE* = -6.2%, *and PBIAS* = -2.13%. The model was then tested for 2008 to 2012 and at several other locations with available observations inside the watershed generating good results (E_{NS} > 0.50 and |*PBIAS* | < 15%). The model parameters used for calibration along with their calibrated value, lower and upper limits are provided in table S2.1.

2.3.2. The relationship between land use and water quality

Fig. 2.2 shows the median and variation of average annual concentrations and loads of TN and TP along with composition of land use for each subbasin. A visual trend between percent urban land use and TN and TP concentrations and loads can be observed. As illustrated in Fig. 2.2, subbasins with higher percentage of urban land use showed higher median and variability of TN and TP concentrations and loads. Land use compositions across subbasins indicates that the exchange of land use is primarily between urban and agriculture. Percent forest land indicated neither a strong nor a significant correlation with nutrient concentration or loads and it was excluded from the MLR models.

Land use and related anthropogenic parameters in the Jordan Lake watershed were strongly and significantly correlated (R^2 > 0.7, p < 0.05). Hence, to account for strong correlation between land use components, either percent urban land with normalized AFO capacities or percent agricultural land with normalized WWTP capacities were used as the independent variables in the MLR model. Twenty-one MLR models were developed corresponding to each year in the 1992-2012 period.

Strong and significant (R^2 > 0.7, p < 0.01) relationships were evident between land use types and TN and TP concentrations/loads in all cases, with the exception of TP loads and concentrations for year 2010 (R^2 = 0.25 and R^2 = 0.47 for models with percent urban-AFO and R^2 = 0.41 and R^2 = 0.53 for models with percent agriculture-WWTP). Tables 2.1 and 2.2 provide a summary of the MLR analysis along with results of various diagnostic statistical tests for assumptions in linear regression. A few of the models (<

5%) did not pass some of the diagnostic statistical tests for assumptions in linear regression and were excluded from the analysis.

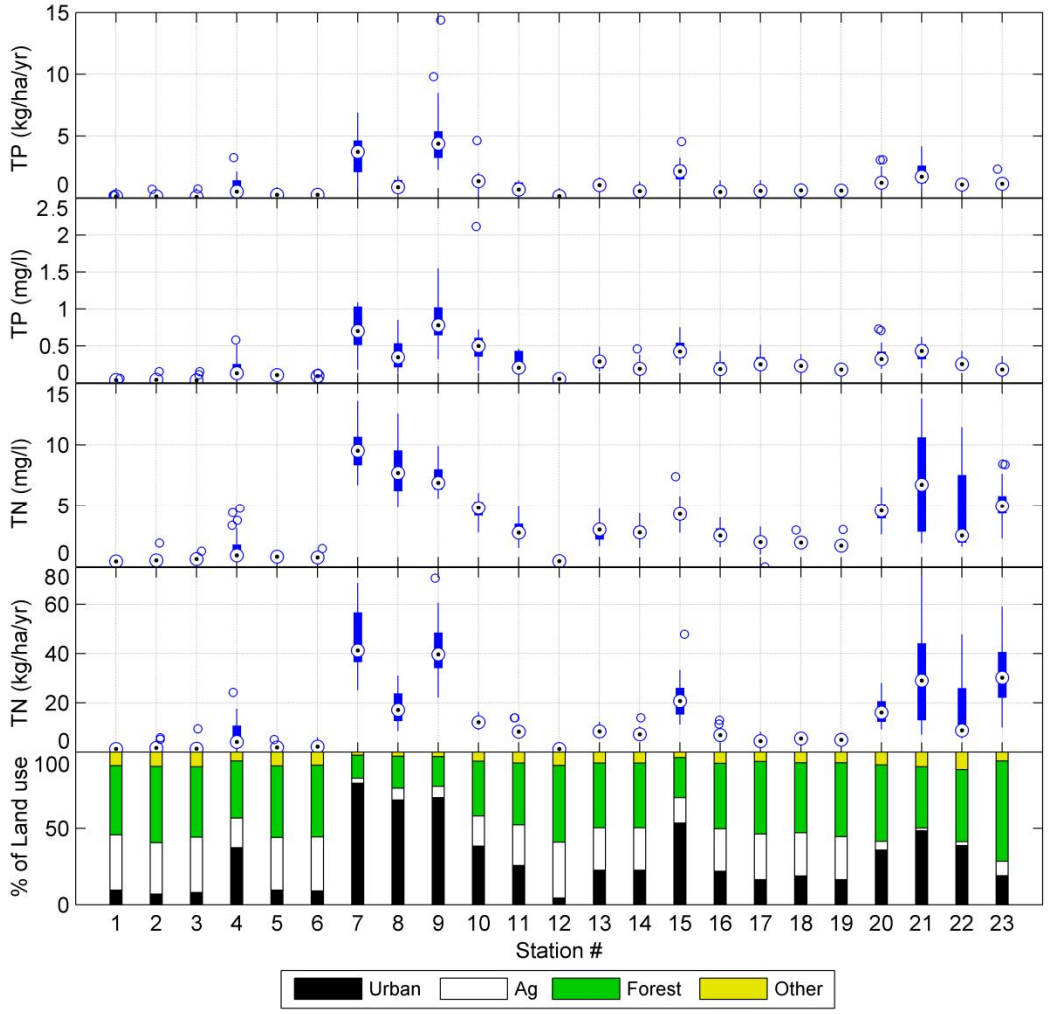

Fig. 2.2. Land use composition within each of the 23 subbasins along with boxplots of total nitrogen (TN) and total phosphorus (TP) concentrations and loads. Note the trend between percent urban land use and TN and TP concentrations and loads.

The regression model coefficients for percent urban-AFO capacity models are provided in Table 2.1. Percent urban land was the significant ($p < 0.01$) and dominant factor positively correlated with both TN and TP loads and concentrations in all years. AFO capacity was not a significant ($p > 0.01$) predictor in more than 80% of the models for TN loads and concentrations. However, it was significant ($p < 0.01$) in

more than 70% of the models for TP loads and concentrations. Similar results were reported by several studies finding urban land use to be the primary factor influencing TN and TP (Osbourne and Wiley, 1988; Tong and Chen, 2002; Schoonover et al., 2005; Mouri et al., 2011; Tu, 2011). In general, the effects of urban land on water quality were higher during dry years (table S2.2). This reasons that pollutant loadings from diffuse sources are typically insignificant due to low surface runoff during dry years.

Table 2.1. Summary of multiple linear regression (MLR) analysis for models developed with exploratory variables percent urban land (% Urb) and animal feeding operations (AFOs) capacities. % Urb. and AFO are the model coefficients for percent urban land use and capacity of AFOs respectively. P_F is the *p*-value for significance of the model, VIF is the variance inflation factor for testing multicollinearity, P_L is the *p*-value for the Lilliefors test for normality, P_{SW} is the *p*-value for Shapiro-Wilk test for normality, P_{BF} for Brown- Forsythe test for homoscedasticity, P_{DW} for Durbin-Watson test for randomness of residuals.

		% Urb.	AFO	R^2	Adj. R^2	P_F	VIF	P_L	P_{SW}	P_{BF}	P_{DW}
TN conc.	Lowest	0.82	-0.05	0.70	0.66	2.7E-11	1.1	0.01	0.01	0.01	0.01
	Median	1.28	0.06	0.83	0.81	1.8E-05	1.8	0.25	0.27	0.21	0.22
	Highest	1.61	0.11	0.96	0.95	1.4E-04	2.4	0.50	1.00	0.41	0.79
TP conc.	Lowest	0.70	-0.01	0.47	0.41	5.0E-11	1.1	0.04	0.02	0.02	0.01
	Median	1.34	0.09	0.83	0.81	1.7E-04	1.8	0.34	0.44	0.25	0.29
	Highest	1.83	0.17	0.96	0.96	3.4E-03	2.4	0.50	0.98	0.42	0.93
TN load	Lowest	0.81	-0.02	0.65	0.61	1.9E-10	1.1	0.01	0.01	0.04	0.01
	Median	1.43	0.03	0.78	0.75	2.4E-04	1.8	0.22	0.22	0.23	0.20
	Highest	1.86	0.10	0.94	0.92	2.5E-03	2.4	0.50	0.81	0.50	0.88
TP load	Lowest	0.66	0.01	0.25	0.16	2.8E-10	1.1	0.01	0.01	0.04	0.01
	Median	1.53	0.07	0.80	0.77	3.8E-03	1.8	0.30	0.47	0.25	0.42
	Highest	1.99	0.17	0.94	0.93	7.9E-02	2.4	0.50	0.97	0.48	0.92

Conversely, significant ($p < 0.01$) negative correlations were evident between percent agriculture and TN responses (concentrations or loads) in percent agriculture-WWTP capacity models (Table 2.2). However the correlation between percent agriculture and TP loads and concentrations was positive in half of the models. WWTP capacity was a significant ($p < 0.01$) factor positively correlated with both TN and TP loads and concentrations in all models. Similar results have been reported in other studies where agriculture had a nonsignificant effect on the land use-water quality regression models (Mouri et al.,

2011; Schoonover et al., 2005). Other studies indicated that agriculture significantly affects water quality (Liu et al., 2000; Jordan et al., 1997; Tu, 2011; Mehaffey et al., 2005). In all the studies where agricultural land had a strong and significant correlation with TN and TP, the percentage of the agriculture land use was much higher than the urban land use and most of the agriculture was cropland. For WWTP capacities, model coefficients were 5% higher on average during dry years compared to wet years consistent with the results obtained with percent urban (table S2.3).

Table 2.2. Summary of multiple linear regression (MLR) analysis for models developed with exploratory variables percent agricultural land (% Ag)and wastewater treatment plants (WWTPs) capacities. % Ag. and WWTP are the model coefficients for percent agricultural land use and capacity of WWTPs respectively. P_F is the *p*-value for significance of the model, VIF is the variance inflation factor for testing multicollinearity, P_L is the *p*-value for the Lilliefors test for normality, P_{SW} is the *p*-value for Shapiro-Wilk test for normality, P_{BF} for Brown- Forsythe test for homoscedasticity, P_{DW} for Durbin-Watson test for randomness of residuals.

		% Ag.	WWTP	R^2	Adj. R^2	P_F	VIF	P_L	P_{SW}	P_{BF}	P_{DW}
TN conc.	Lowest	-0.63	0.09	0.68	0.64	2.1E-10	1.4	0.01	0.01	0.08	0.10
	Median	-0.35	0.18	0.84	0.82	1.8E-05	1.9	0.30	0.52	0.25	0.46
	Highest	-0.16	0.23	0.96	0.95	2.0E-04	2.5	0.50	0.97	0.49	0.91
TP conc.	Lowest	-0.29	0.09	0.53	0.47	8.2E-11	1.4	0.01	0.01	0.10	0.02
	Median	-0.11	0.21	0.82	0.79	1.2E-04	1.9	0.29	0.31	0.30	0.36
	Highest	0.36	0.28	0.92	0.91	1.2E-03	2.5	0.50	1.00	0.47	0.94
TN load	Lowest	-0.72	0.07	0.76	0.72	1.9E-13	1.4	0.03	0.03	0.05	0.01
	Median	-0.40	0.19	0.87	0.86	3.0E-05	1.9	0.36	0.49	0.32	0.34
	Highest	-0.17	0.25	0.97	0.97	3.8E-04	2.5	0.50	0.98	0.47	0.72
TP load	Lowest	-0.40	0.11	0.41	0.34	2.5E-15	1.4	0.01	0.01	0.05	0.02
	Median	-0.28	0.23	0.85	0.83	4.5E-04	1.9	0.24	0.33	0.21	0.40
	Highest	0.22	0.32	0.98	0.97	9.1E-03	2.5	0.50	0.98	0.43	0.85

A Primary transport mechanism for P is erosion and sediment transport (Emsley, 1980). In Jordan Lake watershed, the monthly TP loads were significantly ($p < 0.01$) correlated with total suspended solids (TSS) loads (table S2.4). Interestingly, the strength of the correlations (in terms of R^2) between monthly TP and TSS loads increased as percent agricultural land increased. This finding indicates that in Jordan Lake watershed, soil erosion and sediment conservation practices could serve benefits for TP pollution control specifically in areas with higher percentage of agricultural land use.

2.3.3. The effects of the inter-annual variability of precipitation on land use-water quality relationships

The ANCOVA results indicated a significant ($p < 0.01$) difference between the slopes of regression models with percent urban land using TN and TP loads based on climate scenarios. Models for dry and wet years had the steepest and mildest slopes respectively, while the models for average years had intermediary slopes (Fig. 2.3). Thus, an equal increase in percent urban land use resulted in higher increases in loads during dry years compared to wet or average years (Fig. 2.3).

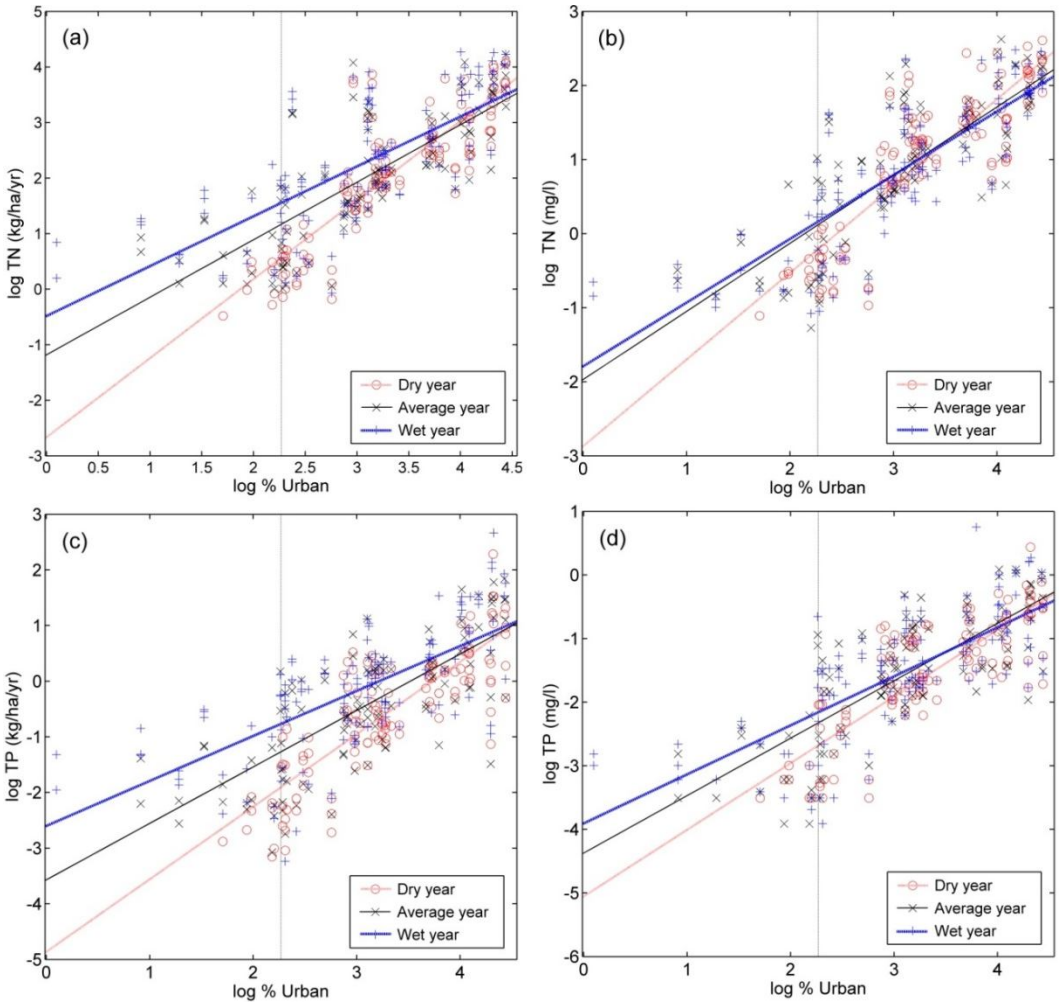

Fig. 2.3. Analysis of covariance (ANCOVA) results for percent urban land use versus total nitrogen (TN) (a and b) and total phosphorus (TP) (c and d) concentrations and loads.

For models developed for percent urban land with TN and TP concentrations, the difference between the model slopes for average and wet years was not significant ($p > 0.01$). Also, the difference between the slopes for wet and dry years is diminished and less than the case with loads (Fig. 2.3). This is an important outcome since it accentuates the importance of differentiating between loads and concentrations when investigating nonpoint source pollution under variable climatic (i.e. precipitation) conditions. Table 2.3 summarizes the results for ANCOVA models.

Table 2.3. Intercepts and slopes of ANCOVA models, $\hat{Y}$ and β for global average and slope respectively, (μ_{dry}, $\mu_{average}$ and μ_{wet}) and (θ_{dry}, $\theta_{average}$ and θ_{wet}) for correction of global mean and slope respectively for dry, average and wet years.

	$\hat{Y}$	μ_{dry}	$\mu_{average}$	μ_{wet}	β	θ_{dry}	$\theta_{average}$	θ_{wet}
Urban-TN (Load)	-1.44	-1.22	0.26	0.96	1.12	0.30	-0.08	-0.22
Urban-TN (Conc.)	-2.21	-0.66	0.24	0.42	0.98	0.19	-0.06	-0.12
Urban-TP (Load)	-3.68	-1.18	0.11	1.08	1.04	0.26	-0.03	-0.23
Urban-TP (Conc.)	-4.44	-0.61	0.07	0.54	0.91	0.14	0.00	-0.14
Ag-TN (Load)	3.76	-0.03	-0.09	0.12	-0.08	-0.0012	0.0012	0.0000
Ag-TN (Conc.)	2.26	0.06	-0.04	-0.01	-0.07	0.0024	0.0015	-0.0039
Ag-TP (Load)	1.04	-0.14	-0.01	0.15	-0.07	-0.0018	-0.0016	0.0034
Ag-TP (Conc.)	-0.46	-0.05	0.04	0.01	-0.05	0.0020	-0.0015	-0.0004

The difference between the slopes for models with percent agriculture was not significant for different climate scenarios (Fig. 2.4). However, the models developed based on loads produced better results in terms of correlation strength, especially for TP loads. This is congruent with the findings in previous sections that agriculture has a small impact on the water quality conditions in Jordan Lake watershed.

Ahearn et al. (2005) found stronger (higher R^2 values) correlations between human population densities and nitrate-N loads during dry years compared to wet years, accompanied with lower impact of percent agriculture. Osborne and Wiley (1988) found the nitrate concentrations in stream water to be correlated with urban land use during dry seasons and conversely correlated with agriculture land use during high-flow seasons.

With regard to loads and concentrations it should be noted that, while using loads can help distinguish between wet and dry hydrologic years, it also introduces higher errors inherent in the process of calculating the loads. Regression-based load estimation methods like LOADEST can improve the accuracy of load estimation and reduce the bias involved by incorporating the full streamflow time series and concentrations compared to the case where individual observations of streamflow are simply multiplied by related concentration to compute the loads.

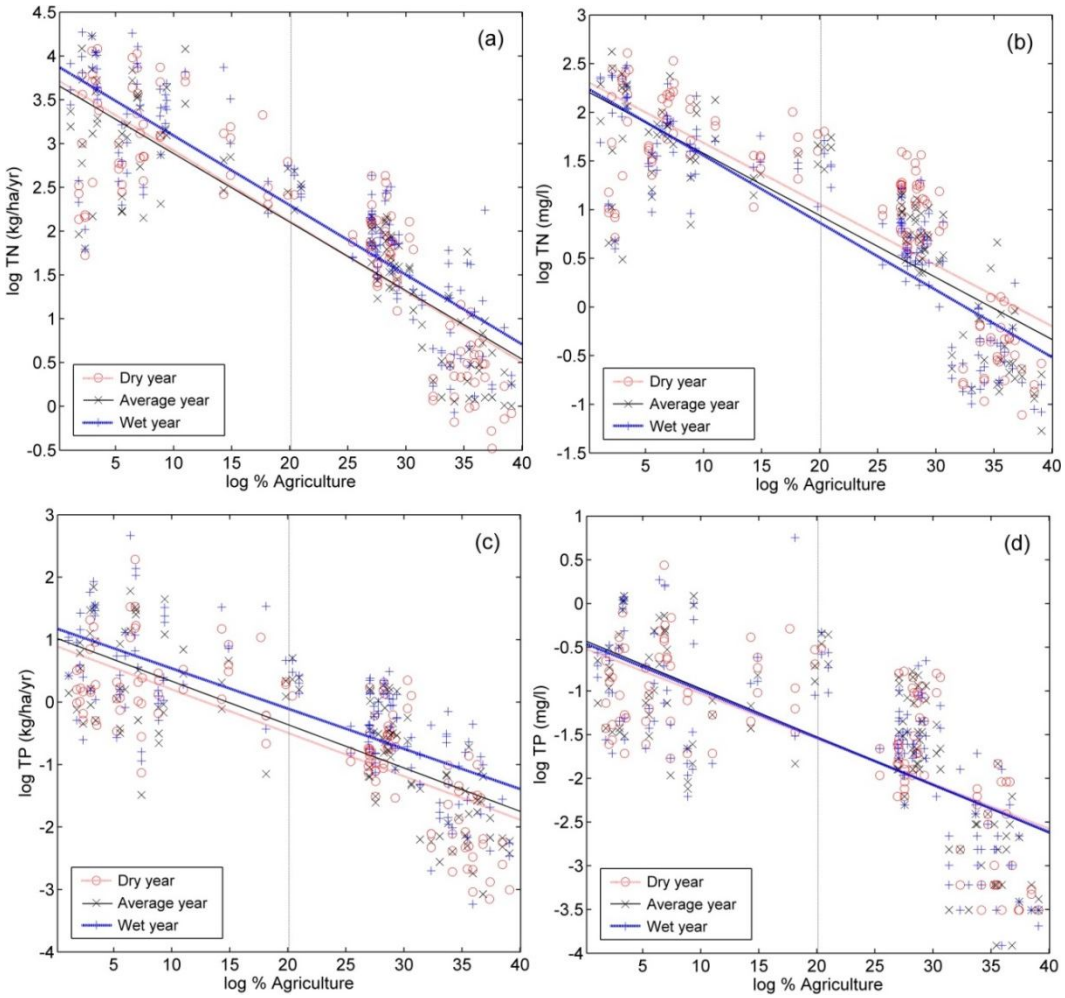

Fig. 2.4. Analysis of covariance (ANCOVA) results for percent agriculture land use vs. total nitrogen (TN) (a and b) and total phosphorus (TP) (c and d) concentrations and loads.

In general, in the Jordan Lake watershed impact of agriculture on nutrient levels seems to be spurious. An explanation can be the relatively low agricultural activities in the watershed that resulted in low impact from this section. Baker (2003) states in his commentary that in general, as soon as urban land use increases beyond a small percentage of total cover (~ 5%), the impact of urban land use on water quality dominates agricultural impacts. Based on land use analysis in Jordan Lake watershed, more than 90% of the agriculture land use is pasture where no significant amounts of fertilizers are applied compared to cropping systems (Osmond et al., 2014). Hence, low yields of TN and TP from agriculture can be another cause of low contribution from agriculture in Jordan Lake watershed. Furthermore, a review of land use compositions across subbasins reveals that the exchange of land use is primarily between urban and agriculture (Fig. 2.2). In other words, higher percentages of agricultural land use is often accompanied by lower percentages of urban land use and since urban land use had a significantly higher impact on water TN and TP levels, shifting from urban to agriculture will result in lower levels of these nutrients.

2.3.4. Effects of land use and climate conditions on vulnerability to exceeding the nutrient targets

Vulnerability to exceeding the TN and TP limits was influenced substantially by both percent urban land use and climate variability. Since agriculture land use was negatively correlated with TN and TP loads and concentrations, it was excluded from the vulnerability analysis. Fig. 2.5 illustrates the vulnerability curves for TN and TP as a function of percent urban land use for different climatic conditions. In general, it can be observed that even small percentages (< 10%) of urban land use can result in high vulnerability (> 0.75) to TN and TP (Fig. 2.5). As TN and TP targets get more stringent, smaller percentages of land use can cause higher levels of vulnerability. With more relaxed targets, higher increases in percent urban land are required to increase vulnerability to TN and TP.

During dry years, even for the less stringent nutrient targets, increases as small as 5% in urban land use percentages result in substantial increases (> 50%) in the vulnerability to TN and TP. This is in

accordance with findings in previous results indicating urban land use had the highest impact during dry years. For average and wet years, higher increases (10% - 40%) in urban land use are required to increase the vulnerability by 50%, especially for more relaxed targets (Fig. 2.5).

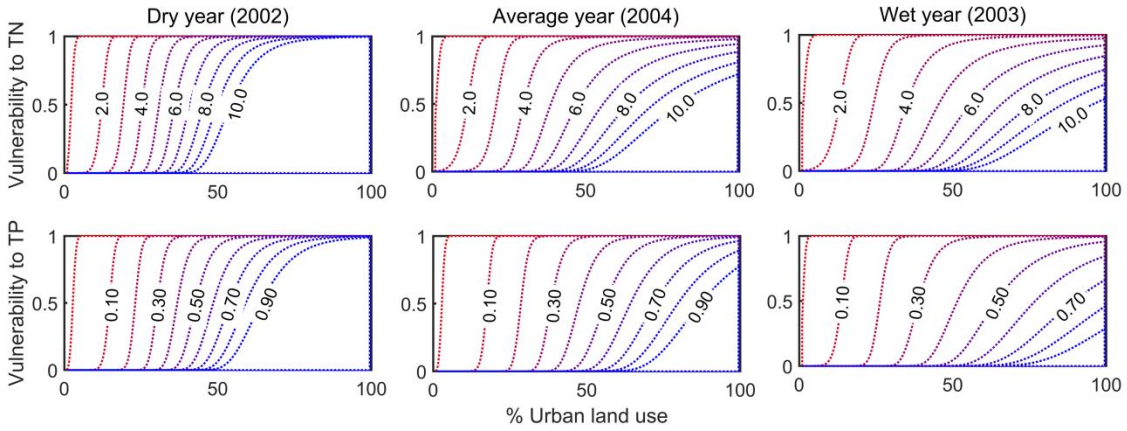

Fig. 2.5. Vulnerability to violating a range of total nitrogen (TN) and total phosphorus (TP) targets (in mg/l) as a function of percent urban land use for dry, average and wet years. The numbers on the curves are nutrient targets (mg/l).

2.4. Conclusions

This study revealed that, strong (R^2 > 0.7) and significant (p < 0.01) positive correlations exist between percent urban land use and annual TN and TP concentrations/loads in Jordan Lake watershed. Agricultural land use was determined to be negatively correlated with TN and TP in most years. The negative correlation of agricultural land use with TN and TP is explained by the relatively low agricultural activities compared to urban development and exchange of land use between these two sectors in Jordan Lake watershed. Further analysis showed significant correlation between TP and TSS loads which tend to become stronger as percent agriculture land increased. This finding indicates that in Jordan Lake watershed, TP conservation practices that are based on sediment reduction will probably be more effective in areas with higher percentage of agricultural land use.

Inter-annual precipitation variations had an important effect on land use-water quality analysis. Using TN and TP loads, significant ($p < 0.01$) differences were observed between regression models based on climatic conditions (determined by the amount of annual precipitation). Using concentrations resulted in nonsignificant differences between average and wet years. This outcome has important implications for selecting between loads and concentrations as well as a time frame for land use-water quality analyses.

Percent urban land use and climate variability had a profound effect on the vulnerability of stream water to exceeding the TN and TP targets. While urban land use was substantially influencing the vulnerability to exceeding the targets, climate conditions played an important role in determining the extent of the impact. During dry years, 5% increase in urban land use percentage results in substantial vulnerability to TN and TP while for average and wet years, 10 to 40 percent increases in percent urban land use is required to reach the same level of vulnerability.

Bibliography

Ahearn, D.S., R.W. Sheibley, R.A. Dahlgren, M. Anderson, J. Johnson, and K.W. Tate. 2005. Land use and land cover influence on water quality in the last free-flowing river draining the western Sierra Nevada, California. J. Hydrol. 313:234–247. doi:10.1016/j.jhydrol.2005.02.038.

Ahmadi, M., M. Arabi, J. Ascough, D. G. Fontane, B. A. Engel. 2014. Toward improved calibration of watershed models: Multisite multiobjective measures of information. J. Environ. Model. Softw. 59:135-145. doi: 10.1016/j.envsoft.2014.05.012.

Allan, J.D., D.L. Erickson, and J. Fay. 1997. The influence of catchment land use on stream integrity across multiple spatial scales. Freshwater Biol. 37:149–161. doi:10.1046/j.1365-2427.1997.d01-546.x.

Arheimer, B., and R. Liden. 2000. Nitrogen and phosphorus concentrations from agricultural catchments: Influence of spatial and temporal variables. J. Hydrol. 227:140–159. doi:10.1016/S0022-1694(99)00177-8.

Arnold, J. G., R. Srinivasan, R. S. Muttiah, and J. R. Williams. 1998. Large area hydrologic modeling and assessment, Part I: Model development. J. Am. Water Resour. Assoc. 34: 73-89. doi:10.1111/j.1752-1688.1998.tb05961.x.

Arnold, J.G., D.N. Moriasi, P.W. Gassman, K.C. Abbaspour, M.J. White, R. Srinivasan, C. Santhi, R.D. Harmel, A. van Griensven, M.W. Van Liew, N. Kannan, and M.K. Jha. 2012. SWAT: Model use, calibration, and validation. Trans ASABE. 55:1491-1508. doi: 10.13031/2013.42256.

Baker, A. 2003. Land use and water quality. Hydrol. Process. 17: 2499–2501. doi: 10.1002/hyp.5140.

Basnyat, P., L.D. Teeter, K.M. Flynn, and B.G. Lockaby. 1999. Relationships between landscape characteristics and nonpoint source pollution inputs to coastal estuaries. J. Environ. Manage. 23:539–549.

Bolstad, P.V., and W.T. Swank.1997. Cumulative impacts of land use on water quality in a southern Appalachian watershed. J. Am. Water Resour. Assoc. 33:519–533. doi: 10.1111/j.1752-1688.1997.tb03529.x.

Brown, M.B., and A.B. Forsythe. 1974. Robust tests for equality of variances. J. Am. Stat. Assoc. 69:364–367. doi:10.1080/01621459.1974.10482955.

Durbin, J., and G.S. Watson. 1950. Testing for serial correlation in least squares regression: I. Biometrika 37:409–428. doi:10.1093/biomet/37.3-4.409.

Durbin, J., and G.S. Watson. 1951. Testing for serial correlation in least squares regression: II. Biometrika 38:159–179. doi:10.1093/biomet/38.1-2.159.

Emsley, J. 1980. The phosphorus cycle. in: The handbook of environmental chemistry. Springer Berlin Heidelberg, 147-167. doi: 10.1007/978-3-662-24940-6_7.

Engel, B., D. Storm, M. White, J. Arnold, and M. Arabi. 2007. A hydrologic/water quality model application protocol. J. Am. Water Resour. Assoc. 43(5):1223–1236. doi:10.1111/j.1752-1688.2007.00105.x.

Environmental Protection Agency (EPA). 2013. ProUCL Version 5.0.00 Technical Guide: Statistical software for environmental applications for data sets with and without nondetect observations, Washington, DC.

Gassman, P. W., M. R. Reyes, C. H. Green, and J. G. Arnold. 2007. The soil and water assessment tool: historical development, applications, and future research directions. Trans. ASABE. 50(4): 1211–1250.

Helsel, D.R. 2005. More than obvious: Better methods for interpreting nondetect data. Environ. Sci. Technol. 39:419A–423A. doi:10.1021/es053368a

Hill, A.R. 1981. Stream phosphorus exports from watersheds with contrasting land uses in southern Ontario. Water Resour. Bull. 17:627–634.

Johnson, L.B., C. Richards, G.E. Host, and J.W. Arthur. 1997. Landscape influences on water chemistry in Midwestern stream ecosystems. Freshwater Biol. 37:193–208.

Jordan, T.E., D.L. Correll, and D.E. Weller. 1997. Relating nutrient discharges from watersheds to land use and streamflow variability. Water Resour. Res. 33:2579–2590, doi:10.1029/97WR02005.

Kutner, M.H., C.J. Nachtsheim, J. Neter, and L. William. 2005. Applied linear statistical models. 5th ed. McGraw-Hill/Irwin, New York.

Lilliefors, H.W. 1967. On the Kolmogorov–Smirnov test for normality with mean and variance unknown. J. Am. Stat. Assoc. 62:399–402. doi:10.1080/01621459.1967.10482916.

Liu, Z.J., D.E. Weller, D.L. Correll, and T.E. Jordan. 2000. Effect of land cover and geology on stream chemistry in watershed of Chesapeake Bay. J. AM. Water Resour. Assoc. 36:1349–1365. doi: 10.1111/j.1752-1688.2000.tb05731.x.

Mehaffey, M.H., M.S. Nash, T.G. Wade, D.W. Ebert, K.B. Jones, and A. Rager. 2005. Linking land cover and water quality in New York City's water supply watersheds. Environ. Monit. Assess. 107:29–44. doi:10.1007/s10661-005-2018-5.

Moriasi, D.N., J.G. Arnold, M.W. Van Liew, R.L. Bingner, R.D. Harmel, and T.L. Veith. 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE. 50:885–900. doi:10.13031/2013.23153.

Motovilov, Y., Gottschalk, L., Engeland, K., Rodhe, A. 1999. Validation of a distributed hydrological model against spatial observations. Agric. For. Meteorol. 257–277. doi:10.1016/S0168-1923(99)00102-1.

Mouri, G., S. Takizawa, and T. Oki. 2011. Spatial and temporal variation in nutrient parameters in stream water in a rural-urban catchment, Shikoku, Japan: effects of land cover and human impact. J. Environ. Manage. 92:1837–1848. doi:10.1016/j.jenvman.2011.03.005.

Nash, J. E., J. V. Sutcliff. 1970. River flow forecasting through conceptual models part I- A discussion of principles. J. Hydrol. 10:282-290. doi: 10.1016/0022-1694(70)90255-6.

National Oceanic and Atmospheric Administration (NOAA). 2013. National climatic data center (NCDC). Available online at: (http://www.ncdc.noaa.gov/).

North Carolina Department of Environment and Natural Resource (NCDENR). 2007. B. Everett Jordan Reservoir, North Carolina phase I total maximum daily load (Final Report). Available online at: (http://portal.ncdenr.org/c/document_library/get_file?uuid=bc043b19-0787-466f-aa7b-779717e55201&groupId=38364).

North Carolina Department of Environment and Natural Resource (NCDENR). 2013. Water Quality Data Assessment. Available online at: (http://portal.ncdenr.org/web/wq/ps/mtu/assessment#4).

Osborne, L.L., and M.J. Wiley. 1988. Empirical relationships between land-use cover and stream water-quality in an agricultural watershed. J. Environ. Manage. 26:9–27.

Osmond, D.L., D.LK. Hoag, A.E. Luloff, D.W. Meals, K. Neas. 2014. Farmers Use of Nutrient Management: Lessons from Watershed Case Studies. J. Environ. Qual. doi:10.2134/jeq2014.02.0091.

Runkel, R.L., Crawford, C.G., and Cohn, T.A. 2004. Load Estimator (LOADEST): A FORTRAN Program for Estimating Constituent Loads in Streams and Rivers: U.S. Geological Survey Techniques and Methods Book 4, Chapter A5, 69 p.

Schoonover, J. E., Lockaby, B., G., and Pan, S. 2005. Changes in chemical and physical properties of stream water across an urban-rural gradient in western Georgia. J Urb. Ecosys. 8:27-43.

Shapiro, S.S., and M.B. Wilk. 1965. An analysis of variance test for normality (complete samples). Biometrika 52:591–611. doi:10.1093/biomet/52.3-4.591.

Sliva, L., and D.D. Williams. 2001. Buffer zone versus whole catchment approaches to studying land use impact on river water quality. Water Res. 35:3462–3472. doi:10.1016/S0043-1354(01)00062-8.

Smart, R.P., C. Soulsby, C. Neal, A. Wade, M.S. Cresser, M.F. Billett, S.J. Langan, A.C. Edwards, H.P. Jarvie, and R. Owen. 1998. Factors regulating the spatial and temporal distribution of solute concentrations in a major river system in NE Scotland. Sci. Total Environ. 221:93–110. doi:10.1016/S0048-9697(98)00196-X.

Stutter, M.I., S.J. Langan, and B.O. Demars. 2007. River sediments provide a link between catchment pressures and ecological status in a mixed land use Scottish River system. Water Res. 41:2803–2815. doi:10.1016/j.watres.2007.03.006.

SWAT literature database. 2016. Available at: (https://www.card.iastate.edu/swat_articles/)

Tolson, B.A., and C.A. Shoemaker. 2007. Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. Water Resour. Res. 43:W01413. doi:10.1029/2005WR004723.

Tong, S. T.Y., and W. Chen. 2002. Modeling the relationship between land use and surface water quality. J. Environ. Manage. 66:377-393. doi:10.1006/jema.2002.0593.

Tu, J. 2011. Spatially varying relationships between land use and water quality across an urbanization gradient explored by geographically weighted regression. J. Appl. Geog. 31:376-392. doi:10.1016/j.apgeog.2010.08.001.

U.S. Department of Agriculture, Economic Research Service. 2014. Livestock and Meat Domestic Data, Livestock and poultry live and dressed weights. Available online at: (http://www.ers.usda.gov/data-products/livestock-meat-domestic-data.aspx).

U.S. Department of Agriculture Agricultural Research Service. 2014. ArcSWAT 2012.10_1.13. Available online at: (http://swat.tamu.edu/).

U.S. Geological Survey. 2013. Load Estimator (LOADEST): A Program for Estimating Constituent Loads in Streams and Rivers. Available online at: (http://water.usgs.gov/software/loadest/).

Williams, R., M. Arabi, J. Loftis, and G.K. Elmund. 2014. Monitoring design for assessing compliance with numeric nutrient standards for rivers and streams using geospatial variables. J. Environ. Qual. 43:1713–1724. doi:10.2134/jeq2013.12.0528.

Yen, H., X. Wang, D. G. Fontane, R. D. Harmel, M. Arabi. 2014 A framework for propagation of uncertainty contributed by parameterization, input data, model structure, and calibration/validation data in watershed modeling. J Environ. Model. Softw. 54:211-221. doi: 10.1016/j.envsoft.2014.01.004.

Zampella, R.A., N.A. Procopio, R.G. Lathrop, and C.L. Dow. 2007. Relationship of land-use/land-cover patterns and surface-water quality in the Mullica River Basin. J. Am. Water Resour. Assoc. 43:594–604. doi: 10.1111/j.1752-1688.2007.00045.x.

Chapter 3

A PROBABILISTIC APPRAISAL OF RAINFALL-RUNOFF MODELING APPROACHES WITHIN SWAT IN MIXED LAND USE WATERSHEDS

Highlights

A probabilistic approach is presented to assess the performance validity of the empirical Curve Number (CN) and physically-based Green and Ampt (G&A) rainfall-runoff methods in the SWAT model. Specifically, the effects of modeling uncertainties on characterization of the hydrologic budgets and streamflow regimes at various spatial scales and upstream land use conditions are investigated. A Bayesian total uncertainty assessment framework, which explicitly accounts for uncertainties from model parameters, inputs, structure, and measurements, was employed to explore uncertainties in streamflow simulation using SWAT with different rainfall-runoff methods in a mixed-land use watershed. While the models were trained for streamflow estimation only at the watershed outlet, the performances of the models were compared at different stream locations within the watershed. At the watershed outlet, the CN method had a slightly better, but not significant, performance in terms of streamflow error statistics. Similar results were obtained for the predominantly forested and agricultural tributaries. However, in tributaries with higher percentage of developed land, G&A outperformed the CN method in simulating streamflow based on various performance metrics. In general, the 95% prediction intervals from the models with G&A method covered a higher percentage of observed streamflow especially during the high flow events. However, they were approximately 20-45% wider than the corresponding 95% prediction intervals from the CN methods. Using 95% prediction interval for estimated flow duration curves, results indicated that the models with CN-based models underestimated high flow events especially in tributaries with highly developed land use. However, models using CN generated higher water yields to streams compared to models using G&A resulting from overestimated surface runoff. The results of this study

have important implications for the selection and application of appropriate rainfall-runoff methods within complex distributed hydrologic models particularly when simulating hydrologic responses in mixed-land use watersheds. In the present study, while CN and G&A methods in the SWAT model performed similarly at the outlet of a mixed-land use watershed, G&A captured the internal processes more realistically. The subsequent effects on the representation of internal hydrological processes and budgets are discussed.

3.1. Introduction

Watershed models are increasingly used to assess hydrologic responses of watersheds to changes in land use, climate variability and change, and other alterations of system characteristics. Faced with myriad hydrologic models, hydrologists ought to select an appropriate model and demonstrate its performance validity for the desired assessments. The performance of watershed models is often evaluated based on a deterministic approach, i.e. calibrate → validate → predict, where calibration is conducted to obtain a parameter set that provides the best fit between model responses and observations at the outlet of the watershed (Seibert, 1999; Santhi et al., 2008; Niraula et al., 2015). Such assessments can be inadequate, and in many cases misleading for selecting an appropriate distributed hydrologic model, particularly when the model is used to simulate interior hydrologic processes or to assess responses at various locations within the watershed (Beven, 2001; Ahmadi et al., 2014). Probabilistic approaches can address the equifinality and nonuniqueness issues in parameter estimation (Moradkhani et al., 2005), input uncertainty (Kavetski et al., 2003), and measurement errors (Harmel et al., 2007) when assessing competing model structures (Ajami et al., 2007).

Literature is replete with studies where a model calibrated at the watershed outlet is used to predict streamflows at interior locations within the watershed under varying upstream land use conditions (Ahearn et al., 2005; Baker and Miller, 2013; Yan et al., 2013). Similarly, studies continue to use the deterministic approach to explore the hydrologic effects of land use change (Zhou et al., 2013; Sunde et

al., 2016; Zuo et al., 2016), climate change (Xu et al., 2013; Meaurio et al., 2017), or implementation of management practices (Taylor et al., 2016; Jang et al., 2017; Motallebi et al., 2017). Interestingly, calibrated model responses at the watershed outlet are also used to compare the performance of competing models (Seibert, 1999; Santhi et al., 2008; Niraula et al., 2015). The validity of these studies remains unclear, mainly because a model could produce responses at the watershed outlet that adequately match observations (i.e. right answer) while representing the internal processes and responses incorrectly (i.e. wrong reasons) (Beven, 2006).

Two common methods for rainfall-runoff modeling are the Curve Number method (CN; USDA-NRCS, 2004) and the Green and Ampt method (G&A; Green and Ampt, 1911). The CN is an empirical method that provides estimates of runoff under varying land use and soil types using total volume of rainfall. Conversely, the G&A is a physically-based method that uses rainfall intensity and duration along with soil physical characteristics such as hydraulic conductivity to simulate infiltration. A major limitation of the CN method is that it only uses the total volume of rainfall and does not account for rainfall intensity and duration. Hence, the applicability of the method is limited to simulations at daily to annual time steps, and cannot be extended to resolve processes at sub-daily time steps.

Such limitations can result in erroneous simulations of runoff processes in areas with inherently quick hydrologic responses to rainfall events, such as small catchments, areas with relatively low soil permeability, and developed areas with large impervious surfaces (Miller et al, 2014). Yet, many studies continue to use the CN method to quantify runoff in mixed-land use watersheds with considerable areas of developed land (Zhou et al., 2013; Yan et al., 2013). Similarly, several studies have used the CN method calibrated in a dominantly agricultural or forested watershed to quantify changes in runoff or streamflow under projections of urban growth (Du et al., 2012; Zhou et al., 2013; Niraula et al., 2015, Wagner et al., 2016).

The Soil and Water Assessment Tool (SWAT; Arnold et al., 1998, 2012) is a semi-distributed watershed model, which includes both the CN and G&A methods to simulate rainfall-runoff processes. The large majority of studies conducted with SWAT have used the CN method for hydrologic and water quality assessments (Gassman et al., 2014; Bauwe et al., 2016). A few recent studies have compared the performance of the CN or G&A methods. The results from these studies are often inconsistent and in some cases contradictory. Some studies concluded better performance of the CN method (Wilcox et al., 1990; Kannan et al., 2007; Bauwe et al., 2016; Cheng et al., 2016), while other studies demonstrated better performance of the G&A (King et al., 1999; Ficklin and Zhang, 2013; Yang et al., 2016). However, the generalizability of the conclusions from these studies remains limited because they neglected three important considerations. First, comparison of models based on a single calibrated parameter set is subject to biases in model selection and modelers' expertise. Second, a model calibrated for responses at the outlet of the watershed may not produce reliable simulations of interior locations with different geospatial characteristics such as land use. Third, modeling uncertainties must be incorporated in comparison and selection of models. Modeling uncertainty includes the uncertainty in parameters, algorithms, input data, and measurements (Beven and Binley, 1992; Vrugt et al., 2003; Ajami et al, 2007; Harmel et al., 2014; Yen et al., 2014). Not accounting for any of these sources of uncertainty when comparing the performance of models can mask real differences in model performance.

The overall goal of this study is to probabilistically investigate the performance validity of the CN and G&A methods within SWAT for simulating hydrologic responses under varying land use conditions. The specific objectives are to: (i) evaluate the uncertainties from different sources (parameters, input data, and model structure) under the CN and G&A methods when simulating streamflow; (ii) compare the streamflow prediction uncertainty for SWAT with different rainfall-runoff methods; and (iii) quantify the total hydrologic regime and components of streamflow simulated using CN and G&A methods at locations with various dominant upstream land use. While a number of previous studies have compared the performance of CN and G&A methods, applying a total uncertainty estimation framework and

accounting for upstream land use variations to assess the performance of the methods is novel. Considering that different climate inputs (daily vs. subdaily) are used for CN vs. G&A, the study reveals the importance of accounting for input data uncertainty when comparing the performance of the methods in simulating hydrologic responses. Also, using full flow statistics via predictive flow duration curve uncertainty for assessing the performance of the methods sheds light on the benefits and deficiencies of each method. This study provides useful insight into the benefits and limitations of different runoff simulation methods in the widely applied SWAT model.

3.2. Materials and methods

Three separate SWAT models were developed. The models were identical except for the rainfall-runoff mechanism and the precipitation time step accommodating the mechanism. The analysis period was 2002 to 2012 of which 2002 to 2008 was used for training (2000-2001 was used for model warmup), and 2009 to 2012 was used for testing the models. The uncertainty assessment framework developed was used with the likelihood function set to consider the errors only at the outlet of the watershed. At each model realization, the error statistics including likelihood, sum of squared errors (SSE), and Nash-Sutcliff (NS) as well as time series of simulated streamflow were stored for the outlet of the watershed and five other stream locations (USGS gauges) inside the watershed. The stream locations inside the watershed were selected such that they included a variety of sizes and dominant land use types (agriculture, developed, forest) of upstream subwatersheds. By comparing model performance at these locations inside the watershed, we assessed how models with CN and G&A mechanisms perform at subwatersheds with different dominant land use types while training the model at the watershed outlet.

3.2.1. Study watershed

The Haw watershed in central North Carolina within the Piedmont region (Fig. 3.1) drains 3280 km^2. The land use within the watershed is: 43% forest, 20% urban/suburban, and 27% agriculture, of which more than 90% is pasture (National Land Cover Database; NLCD2011; USGS TNM, 2016). Average

annual precipitation is 1060 millimeter (mm) with summer precipitation being the highest and autumn the driest. From 2002 to 2012, the driest year was 2002 with 780 mm precipitation, and 2003 had the highest precipitation (1815 mm) (National Oceanic and Atmospheric Administration; NOAA, 2016).

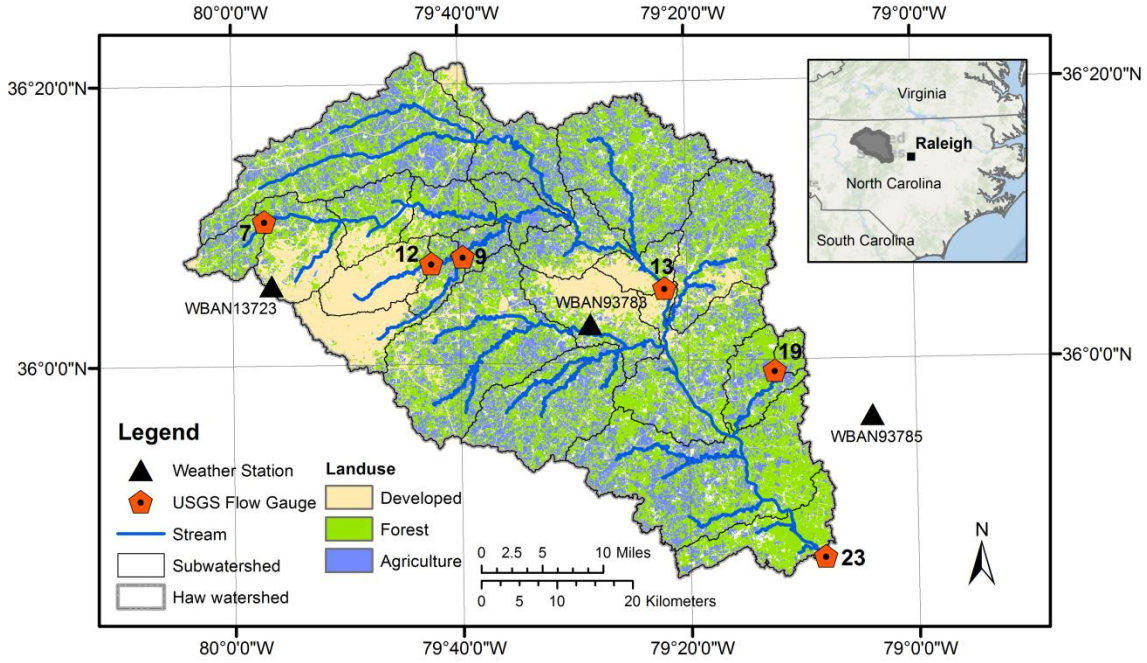

Fig. 3.1. Location Map of the study watershed, streamflow gaging and meteorological stations.

3.2.2. Hydrologic model

SWAT is a continuous-time, distributed-parameter, process-based watershed model, which has been used extensively for hydrologic and water quality assessments under varying climatic, land use, and management conditions in small watersheds to large river basins (Gassman et al., 2007; Arnold et al., 2012, CARD Staff, 2016). The model has the capability to run on daily or smaller time steps. In SWAT, the watershed is split into smaller subwatersheds, which are further discretized into Hydrologic Response Units (HRUs). HRUs are the smallest spatial units in SWAT, and are defined as areas within each subwatershed with unique combinations of land use, soil, and slope class.

Climate inputs drive hydrologic responses and provide moisture and energy inputs in SWAT. Hydrologic processes simulated in the model include canopy storage, surface runoff, infiltration, evapotranspiration, lateral flow, tile drainage, redistribution of water within the soil profile, return flow, and recharge (Arnold et al., 2012). Surface runoff is simulated using either the modified G&A (Mein and Larson, 1973) with subdaily rainfall or CN method with daily rainfall.

The Green and Ampt and CN methods are available runoff estimation mechanisms (i.e. model structures) in SWAT. In the CN method, surface runoff is estimated with daily rainfall depth (R_{day}) and retention parameter (S):

$$Q_{surf} = \frac{(R_{day}-0.2S)^2}{(R_{day}+0.8S)} \quad (1)$$

where Q_{surf} is the depth of the surface runoff. All parameters are values for the day in millimeter (mm). The retention parameter for CN in SWAT is often estimated based on antecedent soil moisture:

$$S = S_{max}\left(1 - \frac{SW}{[SW+\exp(w_1-w_2*SW)]}\right) \quad (2)$$

where SW is the soil water content of the entire profile excluding the amount of water (mm) held in profile at wilting point, w_1 and w_2 are shape coefficients (explained in SWAT documentation, Neitsch et al., 2011), and S_{max} denotes the maximum value of retention parameter (mm) calculated as:

$$S_{max} = 25.4\left(\frac{1000}{CN} - 10\right) \quad (3)$$

where CN is the curve number for soil moisture condition I (explained in SWAT documentation, Neitsch et al., 2011). This approach tends to overestimate the runoff in shallow soils (Kanan et al., 2007). Hence, another option can be used to compute the retention parameter at a given time step t based on plant evapotranspiration (Neitsch et al., 2011):

$$S_t = S_{t-1} + E_0 * \exp\left(\frac{-CNCOEF * S_{t-1}}{S_{max}}\right) - R_{day} + Q_{surf} \quad (4)$$

where S_{t-1} is the retention parameter from the previous time step (i.e. day), E_0 is the potential evapotranspiration for the day, and $CNCOEF$ is the weighting coefficient determined by calibration. At the beginning of the simulation (first day), the retention is defined as $S = 0.9 * S_{max}$. Calculating the CN based on plant evapotranspiration instead of soil moisture makes its value more dependent on climate instead of soil storage (Neitsch et al., 2011). The CN method based on soil moisture and evapotranspiration will be referred to as CN I and CN II, respectively, hereafter.

The modified G&A method is the alternative model structure in SWAT for runoff estimation that incorporates rainfall duration and intensity to compute the infiltration rate on subdaily time steps:

$$f_{inf,t} = K_e \left(\frac{\Psi_{wf} * \Delta\theta_v}{F_{inf,t}} \right) \quad (5)$$

where $f_{inf,t}$ is the infiltration rate at time t (mm per hour), and $F_{inf,t}$ is the cumulative infiltration at time t (mm), K_e denotes the effective hydraulic conductivity (mm per hour), Ψ_{wf} denotes the wetting front potential (mm), and $\Delta\theta_v$ represents the change in volumetric moisture content across the wetting front (mm/mm). Equations for $F_{inf,t}$, K_e, Ψ_{wf}, and $\Delta\theta_v$ are explained in SWAT documentation (Neitsch et al., 2011).

3.2.3. Model inputs: Terrain, Soils, Land use, Climate, and Hydrography

The elevation data for building the SWAT model was the 1/3 arc-second (~10 m) resolution digital elevation model (DEM) of the Haw watershed obtained from United States Geological Survey The National Map (USGS TNM, 2016). The Soil Survey Geographic (SSURGO) database from United States Department of Agriculture Natural Resources Conservation Services (USDA-NRCS, 2016) was used to represent soil characteristics and variability in the watershed. The SSURGO soil data were not available for a small area in the north-eastern part of the Haw watershed (~ 3% of total watershed area). The State Soil Geographic (STATSGO; USDA-NRCS, 2016) data were used to cover the area with missing SSURGO information. The National Land Cover Database (NLCD) for year 2006 was obtained from the

USGS TNM and used as land use for building the model. The resolution of the SSURGO, STATSGO, and NLCD data was 1 arc-second (~30 m).

Stream flowlines, subwatershed boundaries, and other hydrography information were obtained from USGS National Hydrography Dataset (USGS-NHD, 2016). The stream flowlines were used for more accurate stream delineation in SWAT. By superimposing the NHD flowlines on the DEM in the process of watershed delineation, the hydrographic segmentation and subwatershed boundary delineation was improved especially in locations where the DEM does not provide enough accuracy (Winchell et al., 2007).

Observed climate for three meteorological stations (Fig. 3.1) were obtained from the National Climatic Data Center (NCDC) Quality Controlled Local Climatological Data (QCLCD; NOAA, 2016) database. Daily and hourly precipitation, minimum and maximum temperature were collected for the period of 2000 to 2012. Wind speed, solar radiation, and relative humidity were simulated by SWAT.

3.2.4. Measurements: Stream discharge

Daily stream discharge data from USGS National Information System (USGS-NWIS, 2016) were obtained for six stations with daily measurements for the analysis period (2002 to 2012) (Fig. 3.1). The locations of gauges (Table 3.1) were carefully selected to enable adequate characterization of the predictive skill of different model structures under varying land use conditions, particularly developed (i.e. urban) versus undeveloped areas.

Table 3.1. Description of selected subwatersheds.

USGS Gauge ID	Subwatershed Outlet#	Area (km^2)	% Developed	% Forest	% Agriculture
2093800	7	53	16	44	33
209553650	9	230	74	17	7
2095500	12	96	85	11	3
2096500	13	606	24	40	29
2096846	19	19	4	73	17
2096960	23	1275	18	45	29

3.2.5. The SWAT model setup

The DEM was used along with the NHD flowlines to delineate the watershed and streams. The watershed was divided into 23 subwatersheds. The HRU definition was done using the land use and soil data with 10% threshold for delineation. Since the topographic variability was small, a single class slope was assumed within each subwatershed. Using these settings, 122 HRUs were defined for the watershed.

Two separate SWAT models were developed using ArcSWAT 2012 (USDA-ARS, 2014). The models were completely identical except for the runoff estimation method and precipitation time step. One of the models was developed with daily precipitation and CN method using the soil moisture (CN I) and plant evapotranspiration (CN II), and the other model was developed using the hourly precipitation and G&A method for runoff simulation. Therefore, three model setups were prepared for the analyses.

3.2.6. Probabilistic model assessment framework

A probabilistic approach was used to assess the predictive skill and performance validity of competing model structures under varying land use conditions. The Bayesian-based approach explicitly accounts for uncertainties from model parameterization, climate input data (i.e. precipitation), model structure (CN I, CN II, or G&A), and measurement data (i.e. daily streamflow). The framework was developed in MATLAB (The MathWorks, Inc.®). A Markov Chain Monte Carlo (MCMC) sampling scheme, the DREAM method, was used to sample the parameter space and derive the posterior distributions. A statistically-correct likelihood function, which explicitly accounts for streamflow error heteroscedasticity and autocorrelation, was employed to ensure the reliability of the search algorithm. Input data uncertainty was incorporated by using precipitation multipliers drawn from a Gaussian distribution with an uncertain mean and standard deviation for each meteorological station. Inferences on posterior distributions of precipitation multipliers were obtained simultaneously along with SWAT model parameters. Finally, Bayesian Model Averaging (BMA) was used to quantify model structural uncertainty and to assess the predictive skill of the competing model structures.

3.2.6.1. Model parameter uncertainty

In a hydrologic model (M), streamflow is simulated ($\widehat{Q}$) as a function of climatic inputs (i.e. precipitation, R), and a vector of model parameters (θ):

$$\widehat{Q} = M(R, \theta) \tag{6}$$

The simulated streamflow is subject to a variety of errors stemming from measured climate inputs, model parameters, and insufficiency of the model conceptualization (model structural error). The streamflow error residual then becomes:

$$e(R, \theta, M) = Q - M(R, \theta) \tag{7}$$

where Q is the observed streamflow, and $e(R, \theta, M)$ denotes the streamflow error residuals due to errors from observed precipitation, model parameters, and model structure. Applying the Bayes theorem, the parameter set θ is assigned posterior probability distribution, $p(\theta|Q)$, which is proportional to the product of the parameter prior probability distribution, $p(\theta)$, and a likelihood function, $L(\theta|Q)$.

The likelihood function assuming normally and independently distributed model residuals (e) with mean zero and variable standard deviation at each observation time step (σ_t), can be expressed as (Vrugt, 2016):

$$L(\theta|Q, \sigma) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma_t^2}} exp\left[-\frac{1}{2\sigma_t^2}\left(Q_t - \hat{Q}_t(\theta)\right)^2\right] \tag{8}$$

where n is the number of time steps. Using a variable standard deviation at each observation time step (σ_t) allows accounting for error heteroscedasticity which often exists in streamflow simulations. However, in most cases it is infeasible to determine σ at each time step due to lack of repeated streamflow measurements. Different approaches have been proposed to circumvent this issue. Some studies have used different transformations including natural log or Box-Cox to stabilize σ and reduce

heteroscedasticity (Sorooshian and Dracup, 1980). Others have proposed alternative forms for the likelihood function (Schoups and Vrugt, 2010). In this study, we applied a natural log transformation on measured and observed streamflow to reduce heteroscedasticity. Often it is easier to maximize the logarithm of the likelihood function due to numerical stability and algebraic simplicity (Ajami, 2007; Vrugt, 2016). Hence, the natural log of the likelihood function was adopted for optimization.

The streamflow error residuals are usually not independently distributed and in most cases temporal autocorrelation exists in the residuals specifically when errors are estimated at daily or smaller time steps. A common approach to reduce the autocorrelation in model residuals is applying an auto-regressive scheme on the error residuals (Sorooshian and Dracup, 1980):

$$e_t = \rho e_{t-1} + v_t \tag{9}$$

where ρ is the lag serial correlation coefficient for the error residuals and v_t is a vector of random components, $v_t \in N(0, \sigma_v)$. In this study we used a first-order autoregressive transformation (AR-1). Applying the natural log and AR-1 ($|\rho| < 1$), the log-likelihood function becomes (Vrugt, 2016):

$$l(\theta|Q, \sigma, \rho) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln\left(\frac{\sigma_v^2}{1-\rho^2}\right) - \frac{1}{2\sigma_v^2}\left((1-\rho^2)e_1^2 + \sum_{t=2}^{n}(e_t - \rho e_{t-1})^2\right) \tag{10}$$

Parameters ρ and σ are determined along with model parameters at each model realization during the MCMC sampling algorithm.

The parameters for the uncertainty analysis were selected based on experience and sensitivity analysis performed previously (Arabi et al., 2007; Arnold et al., 2012; Tasdighi et al., 2017). Table 3.2 lists the parameters selected for uncertainty analysis along with their ranges. In this study, uniform (noninformative) prior distributions were assumed for parameters within predefined ranges. The same assumption has been used in many other hydrological modeling studies since prior knowledge of the model parameters is often not available and is case-specific (Ajami et al., 2007). The ranges for

parameters were selected based on the SWAT user manual and experience from previous study (Tasdighi et al., 2017; Arnold et al., 2012).

Table 3.2. Parameters of SWAT chosen for uncertainty analysis in this study.

Parameter	Input file	Description	Lower bound	Upper bound
SURLAG	.bsn	Surface runoff lag coefficient	0.001	15
CNCOEF	.bsn	Plant ET curve number coefficient	0.5	2
ALPHA_BF	.gw	Base flow alpha factor for recession constant	0.001	1
GW_DELAY	.gw	Groundwater delay time	0.001	100
GW_REVAP	.gw	Groundwater revap coefficient	0.02	0.2
GWQMN	.gw	Threshold depth of water for return flow to occur	0.01	5000
RCHRG_DP	.gw	Deep aquifer percolation fraction	0	1
CANMX	.hru	Maximum canopy storage	0	10
ESCO	.hru	Soil evaporation compensation factor	0.5	1
OV_N	.hru	Manning's n value for overland flow	0.01	0.6
SLSUBBSN	.hru	Average slope length	10	150
CN_F	.mgt	Fraction change in SCS runoff curve number	-0.2	0.2
CH_KII	.rte	Effective hydraulic conductivity in the main channel	0.025	150
CH_NII	.rte	Manning's n value for the main channels	0.01	0.3
SOL_AWC	.sol	Fraction change in available soil water capacity	-0.5	1
SOL_K	.sol	Fraction change in saturated hydraulic conductivity	-0.5	2
SOL_Z	.sol	Fraction change in soil depth	-0.5	1
SOL_BD	.sol	Fraction change in soil moist bulk density	-0.5	0.5
CLAY	.sol	Fraction change in % Clay	-0.5	0.5
SAND	.sol	Fraction change in % Sand	-0.5	0.5
CH_KI	.sub	Effective hydraulic conductivity in tributary channels	0.025	150
CH_NI	.sub	Manning's n value for the tributary channels	0.01	0.3

3.2.6.2. Model input uncertainty

While uncertainty from model parametrization has been examined in many hydrologic modeling studies, there are only a few cases where input uncertainty is explicitly accounted for (Kavetski et al., 2003; Ajami et al., 2007). The majority of these studies incorporate input uncertainty by applying latent variables, which are basically multipliers for precipitation events drawn randomly from a predefined distribution along with model parameters (Kavetski et al., 2003; Leta et al., 2015). This approach can lead to dimensionality problems as the number of precipitation events increase. In this study a method

proposed by Ajami et al. (2007) is implemented to account for precipitation uncertainty. Using this method, instead of iterating on each single multiplier, the iteration is performed on the mean and standard deviation of a random Gaussian distribution from which the multipliers are randomly drawn at each time step:

$$\hat{R}_t = \phi_t R_t; \qquad \phi_t \sim N(\mu_\phi, \sigma_\phi^2) \tag{11}$$

where $\hat{R}_t$ and R_t are the corrected and observed precipitation depths respectively, ϕ_t is the random multiplier drawn from a normal distribution with a random mean μ_ϕ, $\mu_\phi \in [0.9,1.1]$ and variance, σ_ϕ^2, $\sigma_\phi^2 \in [1e-5,1e-3]$ (Ajami et al., 2007). Incorporating precipitation multipliers using this approach reduces the dimensionality issue.

3.2.6.3. Model structural uncertainty

Evaluation of model structural uncertainty has been investigated in hydrologic modeling (Raftery et al., 2005; Duan et al., 2007; Yen et al., 2014). Different methods have been proposed to account for model structural uncertainty (Hoeting et al., 1999; Georgakakos et al., 2004). One such approach is the Bayesian Model Averaging (BMA; Hoeting et al., 1999). The BMA is a probabilistic algorithm for combining competing models based on their predictive skills (Ajami et al., 2007; Madadgar and Moradkhani, 2014). In this study, the three rainfall-runoff model structures in SWAT ($M_1: CN\ I, M_2: CN\ II, M_3: G\&A$) were used to explore the effects of model structural uncertainty. Under the BMA theory, the posterior distribution of the BMA prediction ($\hat{Q}_{BMA}$) is:

$$p\left(\hat{Q}_{BMA} \middle| M_1, M_2, M_3, Q\right) = \sum_{i=1}^{3} \left[p(M_i|Q) \times p_i(\hat{Q}_i|M_i, Q)\right] \tag{12}$$

where $p(M_i|Q)$ is the posterior probability of the model M_i. This term can be assumed as a probabilistic weight (w_i) for model M_i in the BMA prediction ($\hat{Q}_{BMA}$). The constraint for BMA weights is: $\sum_{i=1}^{3} w_i = 1$. Higher values of w_i can be interpreted as higher predictive skill for a given model structure.

The model weights can be determined using different optimization techniques. The expectation-maximization (EM) algorithm (Dempster et al., 1977) is one such technique to estimate model weights used in several studies (Yen et al., 2014; Ajami et al., 2007). In this study the EM method was used to determine the model weights.

The Brier scores were also employed to compare the performance of the three models while incorporating parameter and input data uncertainties. The Brier score (BS) is a measure of the accuracy of the prediction and has been frequently used in the probabilistic forecast analysis (Georgakakos et al., 2004). BS is defined as:

$$BS = \frac{1}{n}\sum_{t=1}^{n}(f(t) - o(t))^2 \quad (13)$$

where n is the number of time steps in the record, $f(t)$ is estimated by the fraction of model simulations larger than the predefined streamflow threshold, and $o(t)$ is a binary value equal to 1 if the observation at time step t is larger than the predefined threshold and equal to zero in all other cases. In this form (Eq. 13), the lower the value of the BS the better the prediction skill of the model.

3.2.6.4. The DREAM algorithm for MCMC analyses

Several Bayesian algorithms are available which have been widely used for uncertainty assessment in hydrologic modeling including the Generalized Likelihood Uncertainty Estimation (GLUE; Beven and Binley, 1992), the Shuffled Complex Evolution Metropolis (SCEM-UA; Vrugt et al., 2003), and the DiffeRential Evolution Adaptive Metropolis (DREAM; Vrugt et al., 2009). DREAM is a multi-chain MCMC method that randomly samples the parameter space and automatically tunes the scale and orientation of the sampling distribution to move toward the target distribution by maximizing the value of the likelihood function. The method has been used extensively for parameter estimation of complex environmental models (Vrugt, 2016). The convergence of the algorithm can be monitored using the procedure proposed by Gelman and Rubin (1992). In this procedure, a scale reduction score (R) is

monitored to check whether each parameter has reached a stationary distribution (Gelman and Rubin, 1992). The common convergence criterion of $R \leq 1.2$ was used in this study as well. DREAM is specifically beneficial in the optimization of complex high dimensional problems. In this study, DREAM was employed to sample the parameter space and derive the posterior distributions.

3.2.7. The strategy for assessment of model performances

Several criteria were used to assess the performance of the models including: error statistics (likelihood, SSE, and NS) determined based on simulated and observed hydrographs, width of the band of uncertainty (spread), inclusion rate (coverage) determined based on the streamflow observations and 95% confidence interval of simulation ensembles, flow duration curves along with bands of uncertainty, Brier scores, and BMA weights. The value of utilizing multiple performance indicators is commonly recommended because it produces a more comprehensive evaluation of model performance (Legates and McCabe, 1999; Harmel et al., 2010). Finally, the water budget in the watershed was analyzed using the results obtained from different models at various locations.

3.3. Results and discussion

3.3.1. Evaluation of model performances

The purpose of this step of analyses was to monitor the variation of streamflow error statistics at different stream locations in the watershed during the training of the model at the watershed outlet using the Bayesian total uncertainty analysis framework. Fig. 3.2 illustrates the variation of error statistics at different stream locations. It is important to note that the training of the models were done only at the watershed outlet (outlet 23) using the values of likelihood (Eq. 10) as the objective function.

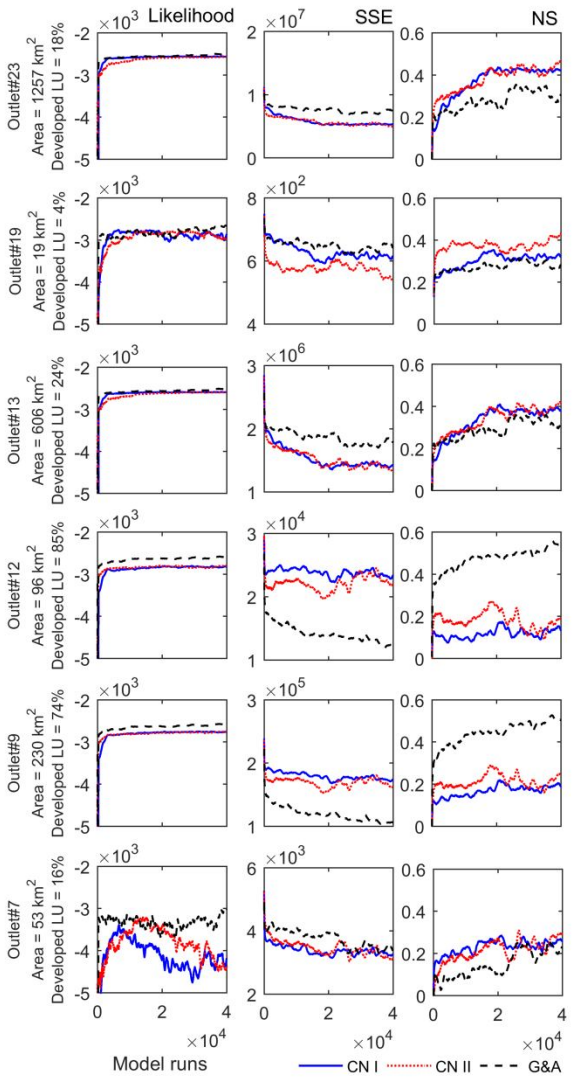

Fig. 3.2. The variation of error statistics at different locations during the model training using the Bayesian total uncertainty analysis framework.

The performances of the three models were identical in terms of likelihood at the outlet of the watershed (outlet 23). However, CN I and CN II had a slightly better performance in terms of SSE and NS. At outlets 13 and 19 with mainly agricultural and forested land use, CN models were slightly better in terms of SSE and NS. While CN I and CN II had very close performance at different locations, CN II had a relatively better performance in the highly forested subwatershed (outlet 19) showing lower values of SSE and higher values of NS. This was foreseeable with CN II as the curve number is determined

based on plant evapotranspiration instead of soil moisture. Similar results were reported for better performance of the CN method over the G&A in other agricultural watersheds (Kannan et al., 2007; Cheng et al, 2016). In contrast, comparing the performance of the CN and G&A methods in an intensive agricultural watershed, Ficklin and Zhang (2013) concluded that the G&A model is more likely to generate better daily simulations. It should be noted that these studies used a deterministic approach, and none compared the performance of the CN and G&A with regard to upstream land use variations.

In highly developed subwatersheds (outlets 9 and 12), G&A significantly outperformed the CN methods with all error statistics (Fig. 3.2). This is an important finding toward the objectives of this study as it demonstrates that while trained at the outlet of the watershed, the G&A model had a much better performance in urbanized subwatersheds inside the watershed. Outlet 7 had an erratic behavior in terms of likelihood which can indicate fundamental deficiency of the models in simulating streamflow for that subwatershed.

3.3.2. Model parameter uncertainty

The posterior Cumulative Distribution Functions (CDF) of parameters for the three models are illustrated in Fig. 3.3. The posterior distributions were generated using 10,000 parameter sets sampled after the convergence of the MCMC algorithm. It can be observed that using different rainfall-runoff methods, different posterior distributions were inferred for parameters. In general, most posterior parameter distributions showed some level of skewness which indicates deficiency in identifiability (Ajami et al., 2007). While for most of the parameters the rainfall-runoff methods determined the degree of the skewness, for some parameters (CH-NII, GW-REVAP, GWQMN, and RCHRG-DP) the skewness changed from positive to negative using different methods.

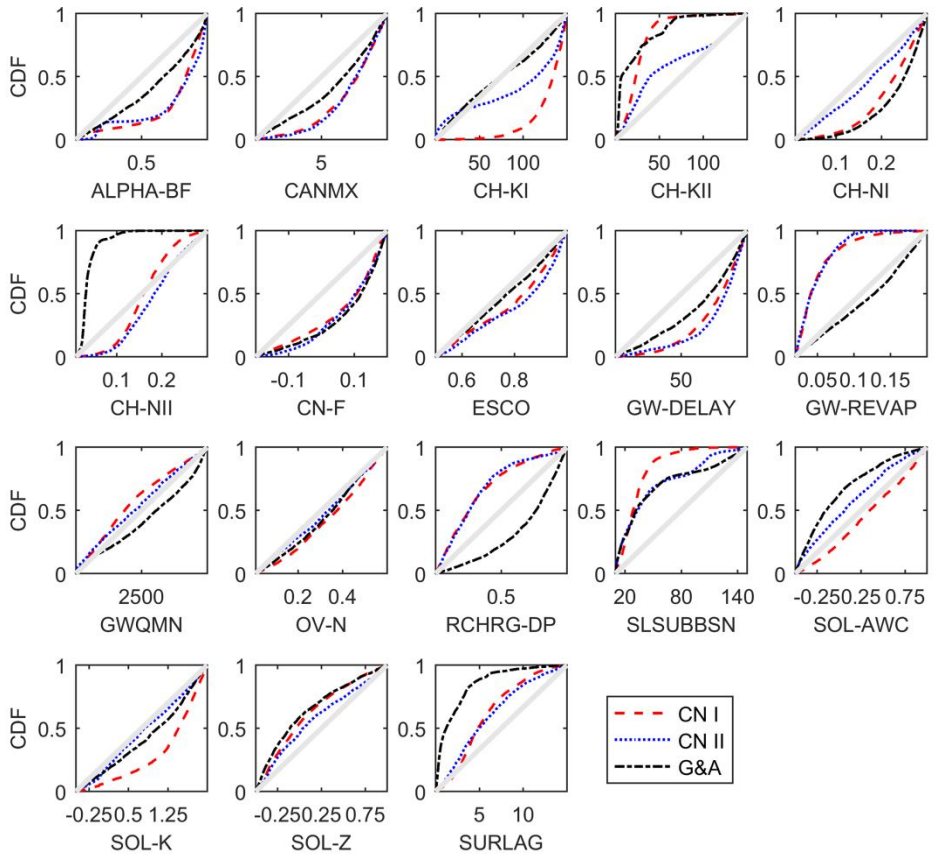

Fig. 3.3. The cumulative posterior distribution (CDF) of parameters for the three models.

The CN I and CN II resulted in similar posterior distributions for most parameters. However, for CH-KI, SOL-AWC, and SOL-K different distributions were inferred which conforms to intuition as CN I uses the soil water content for determining the curve number in runoff estimation. Parameter distributions that show great deviation from normality indicate some sort of deficiency in the combination of model structure, input, and training data.

These findings have important implications with regard to parameterization of SWAT when using different rainfall-runoff mechanisms. The majority of studies that have compared the performance of the rainfall-runoff mechanisms within SWAT have used a deterministic approach for setting the model parameters using the CN or G&A methods (King et al., 1999; Ficklin and Zhang, 2013). The results

obtained here suggest that this approach may mask the differences between the methods and result in misleading inferences regarding the performance of the models under the CN or G&A methods.

3.3.3. Model input uncertainty

Precipitation multipliers were drawn from normal distributions with random mean and standard deviation sampled during each model run within the Bayesian total uncertainty analysis framework. Fig. 3.4 shows the CDF of the values of the mean (μ_ϕ) sampled for input error models under different rainfall-runoff mechanisms for the three climate stations. Since the posterior distributions of standard deviation (σ_ϕ) for input error models were close to uniform, they were excluded from the figure. In general, the mean of input error model under the G&A method showed a distribution closer to normal with mean around one. In contrast, the CN methods produced posterior distribution skewed toward the higher bound, especially at stations WBAN93783 and WBAN 93785. Skewness toward the higher bound indicates the sampling algorithm's attempt to increase the magnitude of the precipitation events by using multipliers larger than one. This can be explained by the structural deficiency of CN methods in simulating the peak streamflows, which caused higher values of precipitation multipliers to be drawn to augment the runoff volume to capture the high flow events. Under these circumstances, it can be hypothesized that the CN methods will lead to systematic overestimation of streamflow compared to G&A (assessed subsequently).

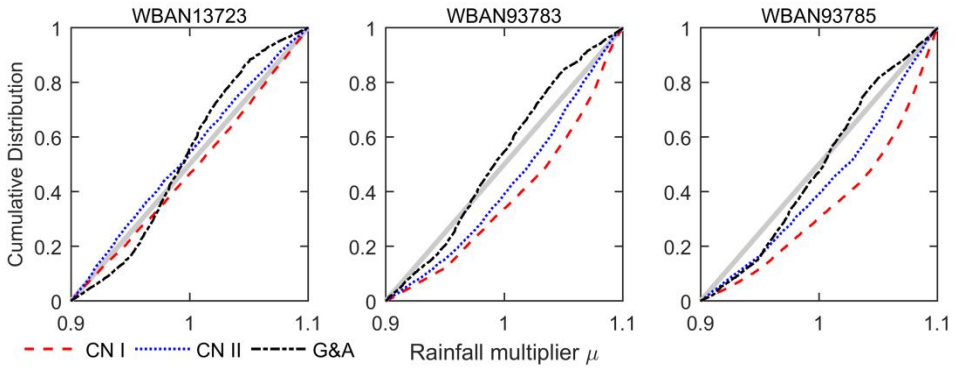

Fig. 3.4. The cumulative posterior distribution (CDF) of the mean (μ_ϕ) for the random Gaussian distributions from which precipitation multipliers were randomly drawn.

3.3.4. Model structural uncertainty

Bayesian model averaging was used at each model realization to combine the three rainfall-runoff models. BMA weights were determined using EM optimization method. Fig. 3.5 shows the boxplots of BMA weights at different subwatershed outlets. Mean of the weights and the weight for the optimized solution (max likelihood) are also marked on the boxplots. Model performance in the agricultural and forested subwatersheds was relatively close in terms of BMA weights. It is evident that in the highly urbanized subwatersheds, G&A significantly outperformed the CN methods. However, at the outlet of the watershed, the CN methods showed a slightly better performance. These results are in accordance with findings in section 3.3.1 where G&A had a better performance in the subwatersheds with dominant urban land in terms of various error statistics.

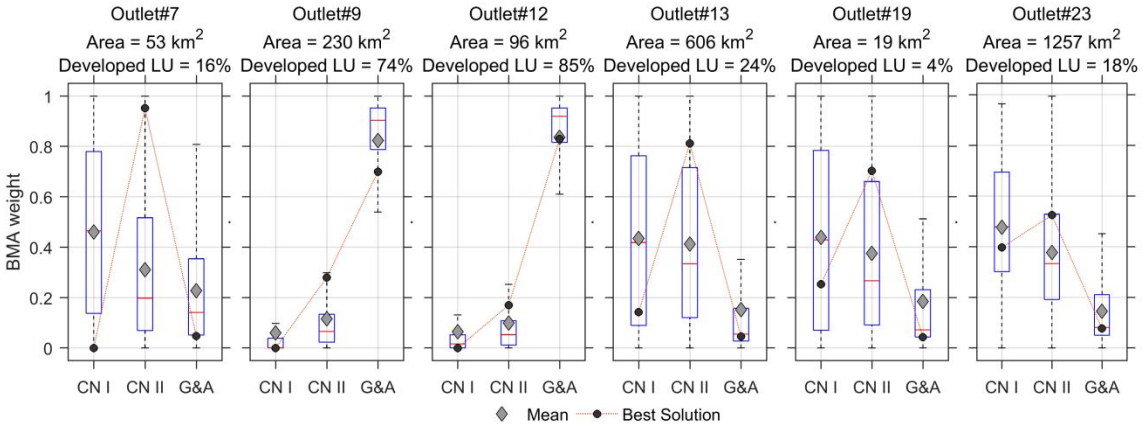

Fig. 3.5. BMA weights for the 3 models at different locations (Solid horizontal lines on the boxes show the median; the boxes show the range of values between 25th and 75th percentile; the whiskers show the 0.5 and 99.5 percentile).

In addition to BMA, Bier scores were also employed to assess the skill of the models in streamflow simulation. In original form (Eq. 13), BS varies between 0 and 1; the lower the value of the BS the better the prediction skill of the model. For illustration, we used $BS' = 1 - BS$ to compare the performance of the models in Fig. 3.6. In this figure, higher values of BS' indicate better performance of the model. At the

outlet of the watershed, the CN models showed a slightly better performance compared to the G&A model specifically at low flow events. With the agricultural and forested watershed, the performances of the models were very close in terms of BS. However, at the highly urbanized subwatersheds, the G&A mechanism clearly outperformed the CN methods during low flow as well as high flow events resulting in higher values of BS'. These findings are also congruent with results from the BMA and error statistics.

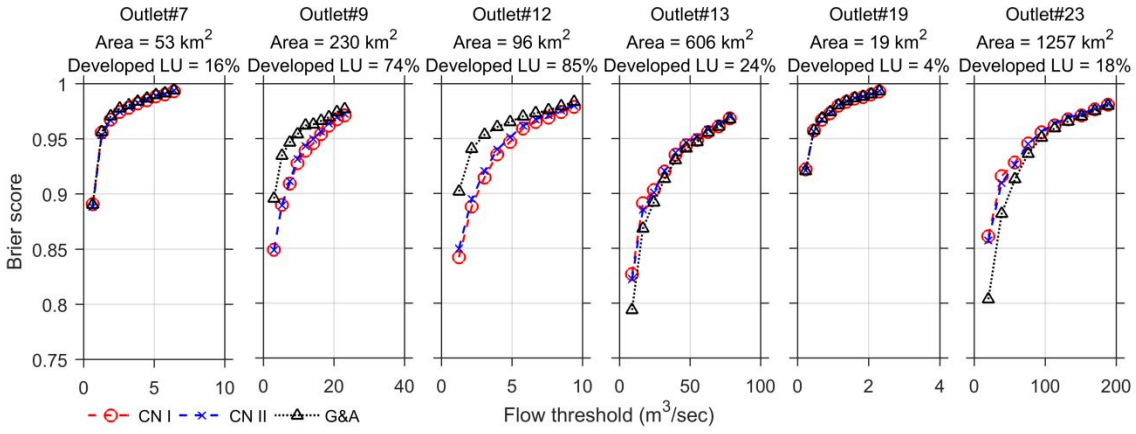

Fig. 3.6. Brier scores for the 3 models at different locations. Note that we used $BS' = 1 - BS$ for illustration of model performances in this figure. In this form, higher values of BS' indicate better performance of the model.

3.3.5. Streamflow prediction uncertainty

Hydrographs with bands of uncertainty have been used frequently for visual assessment of streamflow prediction uncertainty. While visual inspection of these graphs can provide useful information about the quality of simulation, in cases where a relatively long record of streamflow is to be inspected, this method can result in graphs that are difficult to read. Also, hydrographs may not be very effective when comparing the predictive performance of several models specifically when bands of uncertainty are involved. In this situation, one has to either use a smaller time period or select some specific events for illustration of the hydrograph or use other measures for visualizing streamflow. We, therefore, resorted to flow duration curves (FDC) to compare the performance of the models. Graphs of FDC along with bands

of uncertainty provide an easily-readable informative measure which can be used to assess the quality of streamflow prediction (Vogel et al., 1994). More importantly, such graphs are much easier for comparing the predictive performance of the competing models. It should be noted that FDCs can also be misleading since the serial structure and autocorrelation of the sequence of the streamflow record is removed in them (Vogel et al., 1994). Bearing that in mind, while we used FDCs to compare the predictive performance of the models, the coverage (percent of observations lying inside the 95% confidence interval of simulation ensembles) and spread (average width of the 95% confidence interval uncertainty band) were determined based on hydrographs instead of FDCs. Moreover, regular hydrographs for shorter periods of time were used to visually compare the performance of the models side by side. Since the performance of the models was similar in the smaller agricultural and forested subwatersheds (outlets 7, 13, and 19), they were excluded from further analysis. Fig. 3.7 shows the hydrographs for outlet#9 for 50 day period during 2004 containing low, medium, and high flow events. Outlet#9 drains a highly urbanized subwatershed. Since the hydrographs for outlet of the watershed and agricultural subwatersheds were fairly similar we did not include them in this figure. Better performance of the G&A can be readily observed in this graph as it better captures the peak flows. The CNI and CNII models showed very close performances in terms of capturing the variations of streamflow during this period.

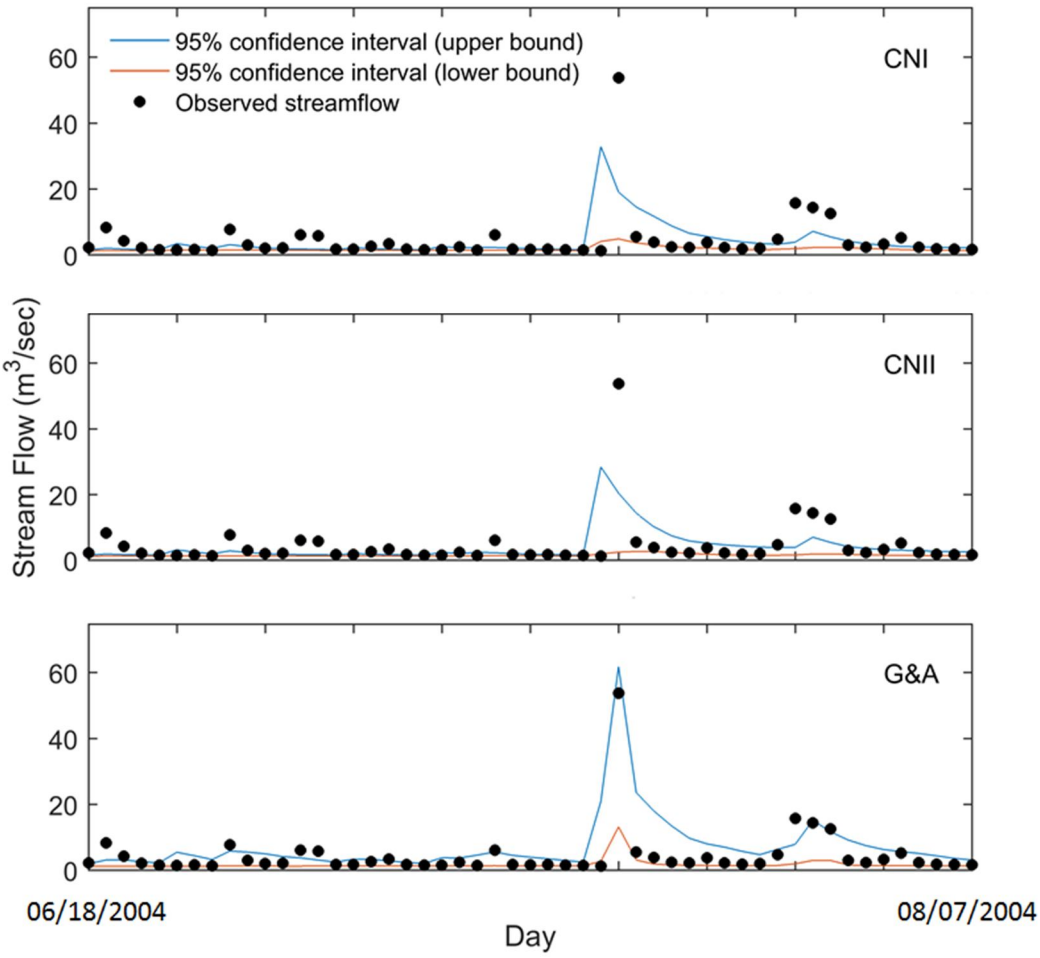

Fig. 3.7. Hydrographs for outlet#9 (area = 230 km^2; 74% developed land) under different model structures.

The G&A method resulted in wider bands of uncertainty (wider spread) and higher coverage rates compared to the CN methods (Table 3.3) during both training (2002-2008) and testing (2009-2012) periods except for outlet 23 during testing where CN II resulted in in higher spread. At the outlet of the watershed (outlet 23), the difference between the methods is small in terms of coverage and spread. However, at stations 9 and 12 (urban-dominant subwatersheds), G&A generated higher coverage rates compared to the CN models.

Table 3.3. Coverage rates and spread at three outlets under the three models. The coverage rates and spread are calculated using the 95% confidence interval of simulated ensemble hydrographs and corresponding observations.

				Training		Testing	
Outlet#	Area (km^2)	% Developed	Model	Coverage (%)	Spread (m^3/sec)	Coverage (%)	Spread (m^3/sec)
9	230	74	CN I	45	1.8	46	1.5
			CN II	52	2.3	56	2.2
			G&A	71	3.9	67	2.9
12	96	85	CN I	46	0.65	54	0.8
			CN II	53	0.87	61	1
			G&A	67	1.4	73	1.1
23	1257	18	CN I	64	27.6	55	18
			CN II	70	31.1	56	23.9
			G&A	76	36.5	54	19

Fig. 3.7 shows the flow duration curves for the observed streamflow and 95% confidence interval bounds for the simulated streamflows for the three models at the three stream locations during the training period. Comparing the performance of the models at the outlet of the watershed, the CN method resulted in narrower bands of uncertainty compared to the G&A method. However, the G&A model showed a slightly better performance in capturing the higher streamflow values. At highly developed subwatersheds 9 and 12, the CN methods were unable to capture the high flow events. Also for medium/low flow events, the observations lay on the upper bound of the uncertainty band. In contrast, for the G&A model, the high flow events were mostly captured and the observations lay somewhere close to the center of the uncertainty band.

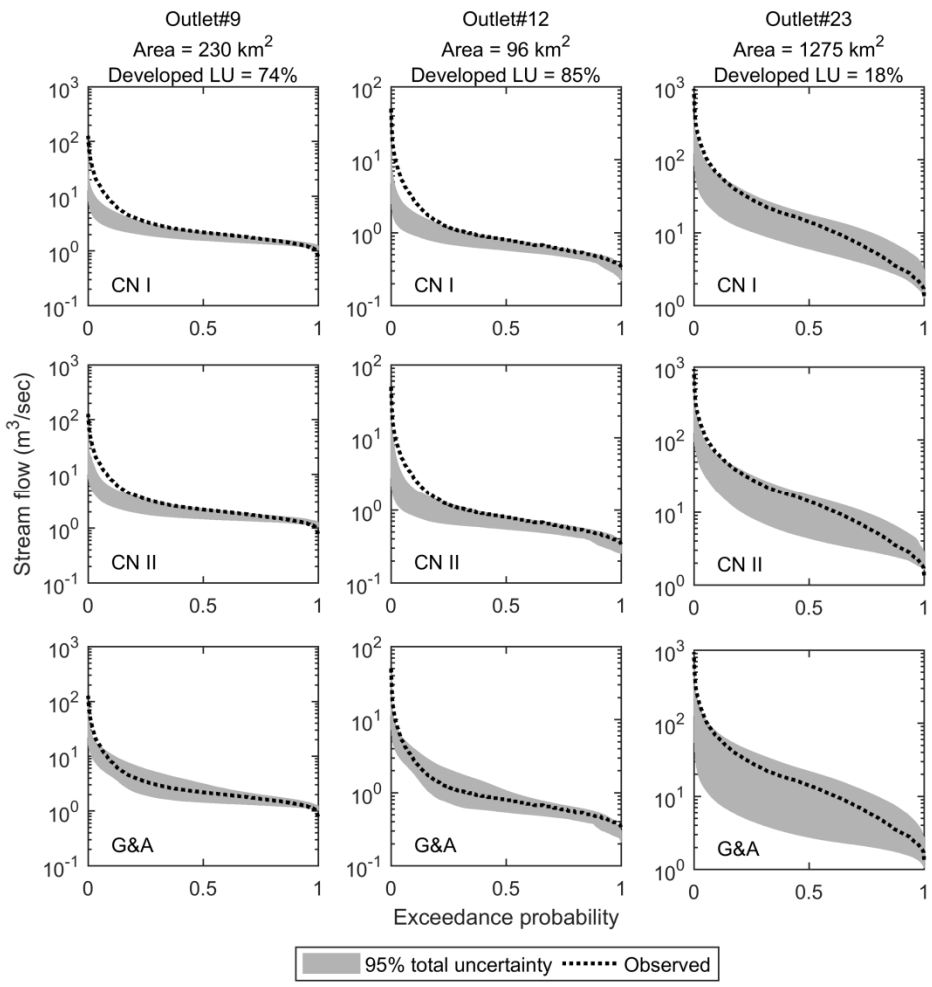

Fig. 3.8. Flow duration curves for the three models at different locations for training period (2002-2008).

Similar results were obtained for the testing period (Fig. 3.8). Slightly narrower bands of uncertainty were determined during the testing period specifically at the outlet of the watershed (outlet 23). Table 3.4 presents the summary of different error statistics during training and testing periods.

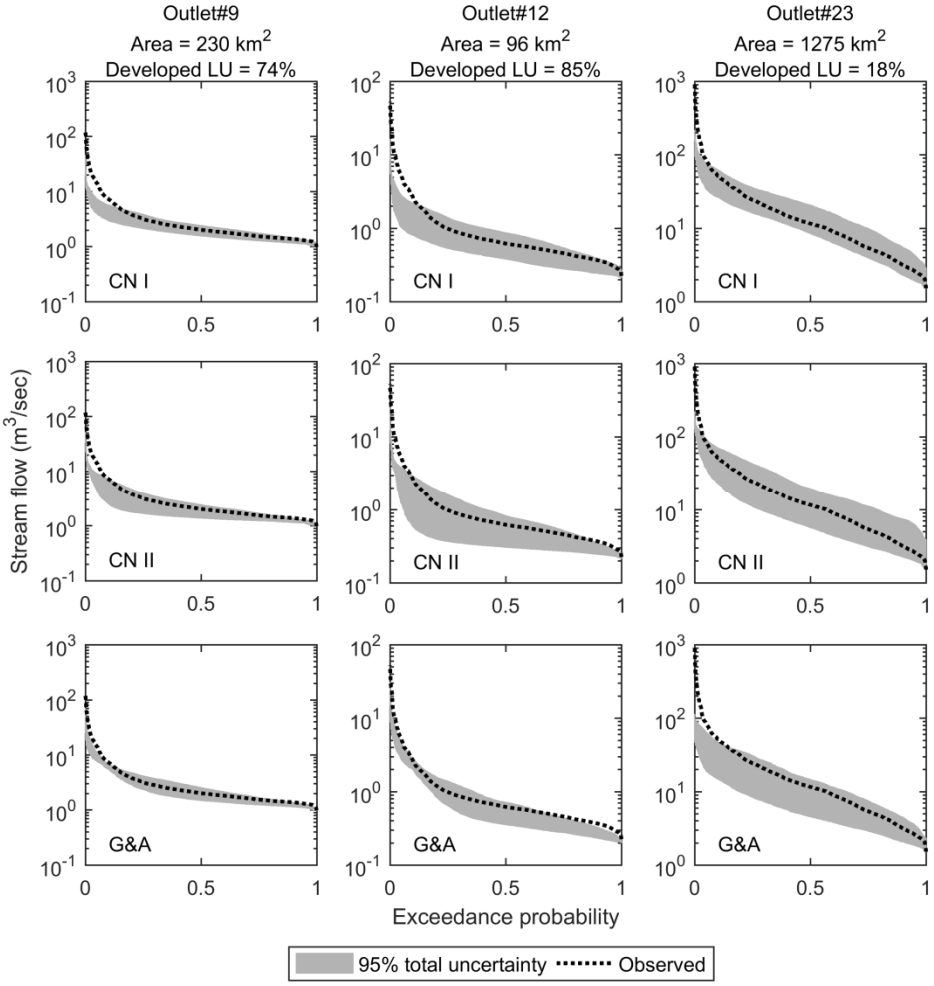

Fig. 3.9. Flow duration curves for the three models at different locations for testing period (2009-2012).

The assumptions of the likelihood function used were assessed using different diagnostics (Fig. 3.9). Homoscedasticity, normality, and autocorrelation of streamflow residuals at watershed outlet (outlet 23) where model training is performed were assessed and illustrated in Fig. 3.9. The figure reveals that the AR-1 and log-transformation were successful in fulfilling the assumptions of the likelihood function in

terms of normality, independence, and homoscedasticity of residuals. It can be observed that the G&A method resulted in narrower normal distribution around zero for residuals.

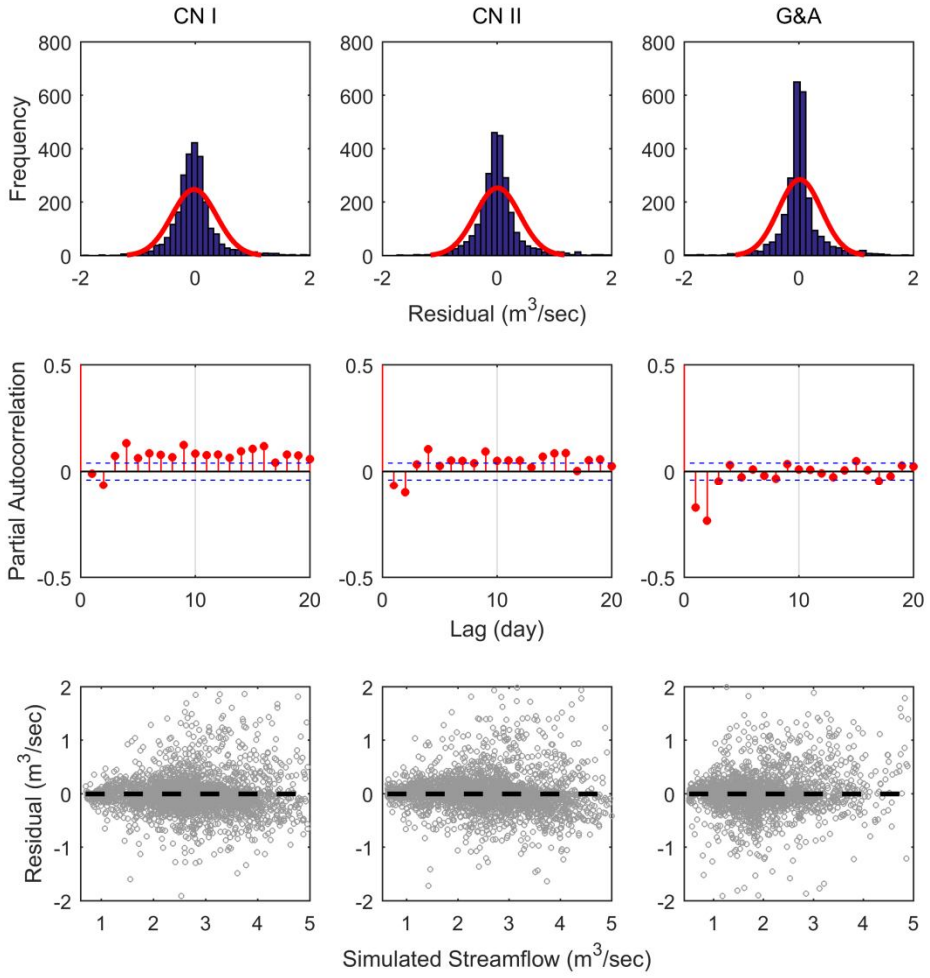

Fig. 3.10. Distribution of residuals (top row), partial autocorrelation coefficients of residuals with 95% confidence intervals (middle row), and residuals as a function of simulated streamflows (bottom row) at training station (outlet 23).

Table 3.4. Summary of different error statistics for training and testing periods

					Training			Testing		
Outlet #	Area (km^2)	% Developed	Model	Statistic	Min	Median	Max	Min	Median	Max
9	230	74	CN I	Likelihood	-2.5E6	-2770	-2662	-2.9E6	-2873	-2736
				RMSE	6.8	8.3	9.7	7	7.5	8.3
				NS	-0.11	0.17	0.45	0.03	0.21	0.32
			CN II	Likelihood	-3.7E4	-2776	-2666	-1.5E5	-2931	-2815
				RMSE	6.3	8.1	9.7	6.9	7.1	7.8
				NS	-0.11	0.22	0.53	0.15	0.3	0.35
			G&A	Likelihood	-8.3E4	-2588	-2534	-1.1E5	-2711	-2645
				RMSE	5.4	6.5	9.4	7.1	7.5	8.4
				NS	-0.05	0.5	0.66	0.03	0.39	0.45
12	96	85	CN I	Likelihood	-2.5E6	-2834	-2663	-6.5E5	-3555	-2939
				RMSE	2.4	3.1	3.4	3	3.2	3.5
				NS	-0.1	0.11	0.45	0.01	0.16	0.27
			CN II	Likelihood	-3.9E4	-2825	-2657	-2.8E4	-9784	-8591
				RMSE	2.3	2.9	3.4	2.8	2.9	3.3
				NS	-0.11	0.18	0.49	0.1	0.3	0.37
			G&A	Likelihood	-7.9E4	-2596	-2530	-1.1E5	-2945	-2704
				RMSE	1.9	2.2	3.3	2.9	3.1	3.5
				NS	-0.02	0.53	0.65	0.04	0.4	0.49
23	1257	18	CN I	Likelihood	-8.7E6	-2561	-2540	-1.0E4	-3451	-2767
				RMSE	39.4	46.3	66.7	41.8	45.1	53
				NS	-0.2	0.41	0.6	0.08	0.41	0.45
			CN II	Likelihood	-7.4E4	-2586	-2544	-1.1E5	-3121	-2609
				RMSE	38.6	45.6	72.3	43.3	45	50.5
				NS	-0.44	0.43	0.6	0.16	0.34	0.4
			G&A	Likelihood	-1.2E4	-2513	-2479	-1.8E5	-5487	-3550
				RMSE	37.8	54.6	67.2	50	53.5	55
				NS	-0.25	0.18	0.6	0.01	0.17	0.33

3.3.6. Assessment of hydrologic budget and streamflow components

The mean annual total water budget and components of streamflow for the three subwatersheds were quantified to gain insight into the differences among the hydrologic processes generating the outcomes. Fig. 3.10 shows the cumulative bar plots for the overall water budget (a-c) and more detailed components

of streamflow (e-f). The G&A method generated lower water yield (total amount of water contributing to streamflow) and higher Evapotranspiration (ET) compared to the CN at all locations. In all models, ET is the major component of water loss (between 60-75%) while about 10-15% and 20-25% of water turns into water yield for the G&A and CN methods respectively.

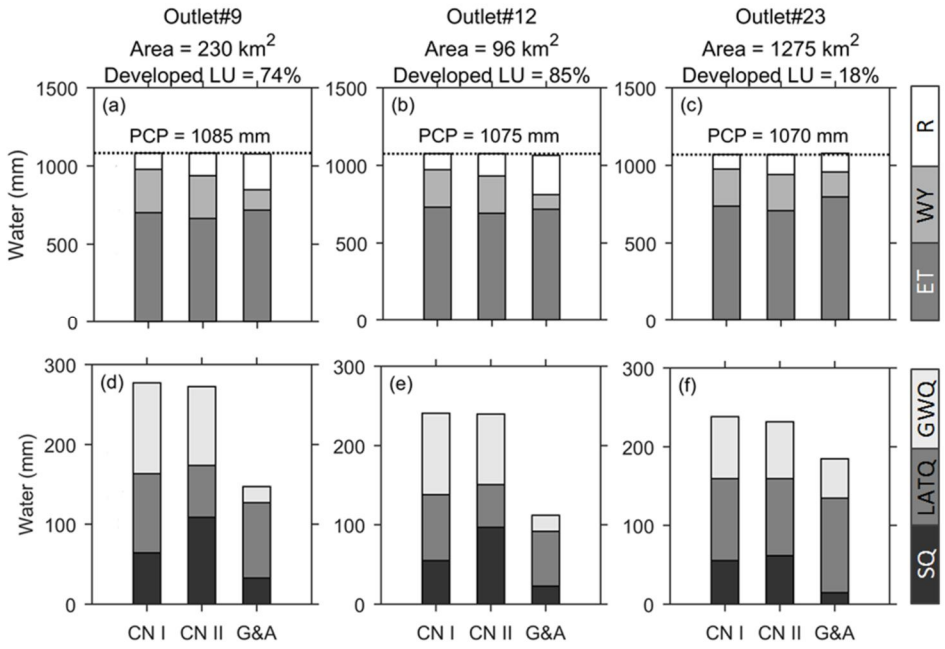

Fig. 3.11. Water budget for the watershed under different models: Panels (a)-(c) show the hydrologic budgets (ET: evapotranspiration; WY: total water yield to streams; R: deep groundwater recharge) for Outlets 9, 12, and 23, respectively; Panels (d)-(f) show surface runoff (SQ), lateral flow (LATQ), and groundwater (GWQ) contributions to water yield (WY).

The G&A method predicted lower surface runoff (SQ) and groundwater flow (GWQ) contributing to streamflow compared to the CN method at all locations. However, it simulated higher lateral flow (LATQ) contribution to streamflow. With CN I and CN II, about 35% of the water yield was derived from the lateral flow, while in the G&A method lateral flow contributed up to 75% of the streamflow. Other studies have shown higher baseflow contribution from the G&A method compared to the CN (Bauwe et al., 2016; Kannan et al., 2007). Considering all components of subsurface flow, the G&A model infiltrated more precipitation. More water in soil profile resulted in higher ET for the G&A model as well.

3.4. Conclusions

SWAT has been used extensively in the literature for hydrologic and water quality simulations. While the model has the capability to employ either CN or G&A method for runoff estimation, almost all studies have employed the CN method. This may be partly due to lack of a rigorous comparison study to justify merits of using one method over the other along with higher simplicity of the CN method. This study attempted to address this shortcoming. Regarding the extreme popularity of the SWAT model, the findings of this study can shed light on selecting the rainfall-runoff method within SWAT that can potentially lead to more realistic streamflow simulations in mixed-land use watersheds.

In this study, a Bayesian total uncertainty assessment framework was implemented to compare the performance of the three runoff generation mechanisms within SWAT under different upstream land use conditions. Using the uncertainty assessment framework at the watershed outlet, models' performances were assessed at several stream locations inside the watershed. At the watershed outlet and subwatersheds with dominant agricultural or forest land use, CN models performed slightly better. However, at the two subwatersheds with highly developed land use, models with G&A method had a much better performance in simulating the streamflow compared to the CN models.

Overall, the streamflow prediction intervals from models with G&A method covered more observations. However, they were slightly wider indicating higher uncertainty for streamflow prediction. The CN models were unable to capture the high flow events specifically in developed subwatersheds. Posterior distribution of mean for Gaussian distributions from which precipitation multipliers were randomly drawn were closer to normal using the hourly precipitation data and G&A method while using daily precipitation with CN methods resulted in substantial negative skewness. The deficiency of models with CN methods in simulating the peak streamflows caused higher values of precipitation multipliers to be sampled to augment the runoff volume.

The CN method simulated higher volume of water yield specifically at the urban-dominated subwatersheds while G&A method simulated higher ET values. The higher volume of water yield by CN in the highly urbanized subwatersheds can be explained by CN attempt to simulate the high flow events which results in overall overestimation of water yield. The G&A model resulted in lower surface runoff at all locations compared to the CN models; however, it simulated higher infiltration and subsurface flows.

The results of this study have important implications for determining which rainfall-runoff method performs better in simulating the hydrologic regime. The evaluation is specifically relevant for applying a distributed hydrologic model such as SWAT in a mixed-land use watershed where model training will be performed only at the watershed outlet but the model is to be used for simulating hydrologic responses at different locations inside the watershed. In summary, the results suggest that while trained at watershed outlet, the SWAT model with G&A method can potentially perform better in areas inside the watershed with higher percentage of developed land. The SWAT models with CN methods proved to have similar or slightly better performance in areas with agriculture or forest dominant land use.

Bibliography

Ahearn, D.S., Sheibley, R.W., Dahlgren, R.A., Anderson, M., Johnson, J., Tate, K.W., 2005. Land use and land cover influence on water quality in the last free-flowing river draining the western Sierra Nevada, California. J. Hydrol. 313, 234-247.

Ahmadi, M., Arabi, M., Ascough, J., Fontane, D.G., Engel, B.A., 2014. Toward improved calibration of watershed models: Multisite multiobjective measures of information. Environ. Modell. Software (59), 135-145.

Ajami, N.K., Duan, Q., Sorooshian, S., 2007. An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction. Water Resources Res. 43(1), Art. No. W01403.

Arabi, M., Govindaraju, R.S., Hantush, M.M., 2007. A probabilistic approach for analysis of uncertainty in the evaluation of watershed management practices. J. Hydrol. 333, 459-471.

Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R., 1998. Large-area hydrologic modeling and assessment: Part I. Model development. J. American Water Resources Assoc. 34(1), 73-89.

Arnold, J.G., Moriasi, D.N., Gassman, P.W., Abbaspour, K.C., White, M.J., Srinivasan, R., Santhi, C., Harmel, R.D., van Griensven, A., Van Liew, M.W., Kannan, N., Jha, M.K., 2012. SWAT: model use, calibration, and validation. Trans. ASABE. 55(4), 1491-1508.

Baker, T.J., Miller, S.N., 2013. Using the Soil and Water Assessment Tool (SWAT) to assess land use impact on water resources in an East African watershed. J. Hydrol. 486, 100-111.

Bauwe, A., Kahle, P., Lennartz, B., 2016. Hydrologic evaluation of the curve number and Green and Ampt infiltration methods by applying Hooghoudt and Kirkham tile drain equations using SWAT. J. Hydrol. 537, 311-321.

Beven, K.J., Binley, A.M., 1992. The future of distributed models: model calibration and uncertainty prediction. Hydrol. Process. 6, 279-298.

Beven, K., 2001. How far can we go in distributed hydrological modelling? Hydrol and Earth Syst. Sci. 5, 1-12.

Beven, K., 2006. A Manifesto for the equifinality thesis. J. Hydrol. 320(1-2), 18-36.

CARD Staff, 2016. SWAT literature database for peer-reviewed journal articles. Center for Agricultural and Rural Development, Ames, IA. https://www.card.iastate.edu/swat_articles/ (accessed 1 Aug. 2016).

Cheng, Q.B., Reinhardt-Imjela, C., Chen, X., Schulte, A., Ji, X., Li, F., 2016. Improvement and comparison of the rainfall–runoff methods in SWAT at the monsoonal watershed of Baocun, Eastern China. Hydrol. Sci. J. 61(8), 1460-1476.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. Series B. 39, 1-39.

Du, J., Qian, L., Rui, H., Zuo, T., Zheng, D., Xu, Y., Xu, C.Y., 2012. Assessing the effects of urbanization on annual runoff and flood events using an integrated hydrological modeling system for Qinhuai River basin, China. J Hydrol. 464, 127-139.

Duan, Q., Ajami, N.K., Gao, X., Sorooshian, S., 2007. Multi-model ensemble hydrologic prediction using Bayesian model averaging. Adv. Water Resources. 30, 1371-1386.

Ficklin, D.L., Zhang, M., 2013. A comparison of the curve number and Green-Ampt models in an agricultural watershed. Trans. ASABE. 56(1), 61-69.

Gassman, P.W., Reyes, M.R., Green, C.H., Arnold, J.G., 2007. The soil and water assessment tool: historical development, application, and future research directions. Trans. ASABE. 50(4), 1211-1250.

Gassman, P.W., Sadeghi, A.M., Srinivasan, R., 2014. Application of the SWAT model special section: overview and insights. J. Environ. Qual. 43, 1-8.

Gelman, A.G., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Stat. Sci. 7, 457-472.

Georgakakos, K.P., Seo, D.J., Gupta, H., Schaake, J., Butts, M.B., 2004. Towards the characterization of streamflow uncertainty through multimodel ensembles. J. Hydrol. 298, 222-241.

Green, W.H., Ampt, G.A., 1911. Studies on soil physics, 1. The flow of air and water through soils. J. Agric. Sci. 4, 11-24.

Harmel, R.D., Smith, P.K., 2007. Consideration of measurement uncertainty in the evaluation of goodness-of-fit in hydrologic and water quality modeling. J. Hydrol. 337(3-4), 326-336.

Harmel, R.D., Smith, P.K., Migliaccio, K.L., 2010. Modifying goodness-of-fit indicators to incorporate both measurement and model uncertainty in model calibration and validation. Trans. ASABE. 53(1), 55-63.

Harmel, R.D., Smith, P.K., Migliaccio, K.L., Chaubey, I., Douglas-Mankin K., Benham, B., Shukla, S., Muñoz-Carpena, R., and Robson, B.J., 2014. Evaluating, interpreting, and communicating performance of hydrologic/water quality models considering intended use: A review and recommendations. Environ. Modeling Software. 57, 40-51.

Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T., 1999. Bayesian model averaging: A tutorial. Stat. Sci. 14(4), 382-417.

Jang, S.S., Ahn, S.R., Kim, A.J., 2017. Evaluation of executable best management practices in Haean highland agricultural catchment of South Korea using SWAT. J. Agri. Water Manag. 180(B), 224-234.

Kannan, N., White, S.M., Worrall, F., Whelan, M.J., 2007. Sensitivity analysis and identification of the best evapotranspiration and runoff options for hydrological modelling in SWAT-2000. J. Hydrol. 332, 456-466.

Kavetski, D., Franks, S.W., Kuczera, G., 2003. Confronting input uncertainty in environmental modeling, in: Duan, Q., Gupta, H.V., Sorooshian, S., Rousseau, A.N., Turcotte, R. (Eds.), Calibration of Watershed Models. Water Sci. Appl. Ser., AGU, Washington, D.C., pp. 49-68.

King, K.W., Arnold, J.G., Bingner, R.L., 1999. Comparison of Green-Ampt and curve number methods on Goodwin Creek watershed using SWAT. Trans. ASABE. 42(4), 919-925.

Legates, D.R., McCabe, G.J., Jr., 1999. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resources Res. 35(1), 233-241.

Leta, O.T., Nossent, J., Velez, C., Shrestha, N.K., van Grinsven, A., Bauwens, W., 2015. Assessment of the different sources of uncertainty in a SWAT model of the River Senne (Belgium). Environ. Modell. Software. 68, 129-146.

Madadgar, S., Moradkhani, H., 2014. Improved Bayesian multimodeling: Integration of copulas and Bayesian model averaging. Water Resources Res. 50, 9586-9603.

Meaurio, M., Zabaleta, A., Boithias, L., Epelde, A.M., Sauvage, S., Sanchez-Perez, J.M., Srinivasan, R., Antiguedad, I., 2017. Assessing the hydrological response from an ensemble of CMIP5 climate projections in the transition zone of the Atlantic region (Bay of Biscay). J. Hydrol. 548, 46-62.

Mein, R.G., Larson, C.L., 1973. Modeling infiltration during a steady rain. Water Resources Res. 9(2), 384-394.

Miller, J.D., Kim, H., Kjeldsen, T.R., Packman, J., Grebby, S., Dearden, R., 2014. Assessing the impact of urbanization on storm runoff in a peri-urban catchment using historical change in impervious cover. J. Hydrol. 515, 59-70.

Moradkhani, H., Hsu, K.L., Gupta, H.V., Sorooshian, S., 2005. Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter, Water Resources Res. 41, W05012.

Motallebi, M., Hoag, D.L., Tasdighi, A., Arabi, M., Osmond, D., 2017. An economic inquisition of water quality trading programs, with case study of Jordan Lake, NC. J. Environ. Manage. 193, 483-490.

Neitsch, S.L., Arnold, J.G., Kiniry, J.R., Williams, J.R., 2011. Soil and Water Assessment Tool Theoretical Documentation Version 2009. Texas Water Resources Institute. Available online at: http://hdl.handle.net/1969.1/128050.

Niraula, R., Meixner, T., Norman, L., 2015. Determining the importance of model calibration for forecasting absolute/relative changes in streamflow from LULC and climate changes. J Hydrol. 522, 439-451.

NOAA, 2016. National climatic data center (NCDC). National Oceanic and Atmospheric Administration, Asheville, NC. http://www.ncdc.noaa.gov/ (accessed 1 Feb. 2016).

Raftery, A.E., Gneiting, T., Balabdaoui, F., Polakowski, M., 2005. Using Bayesian model averaging to calibrate forecast ensembles. Mon. Weather Rev. 133, 1155-1174.

Santhi, C., Kannan, N., Arnold, J.G., Di Luzio, M., 2008. Spatial calibration and temporal validation of flow for regional scale hydrologic modeling. J. Am. Water Resources Assoc. 44(4), 829-846.

Schoups, G., Vrugt, J.A., 2010. A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. Water Resources Res. 46(10), W10531.

Seibert, J., 1999. Regionalisation of parameters for a conceptual rainfall-runoff model. J. Agric. For. Meteorol. 98/99, 279-293.

Sorooshian, S., Dracup, J.A., 1980. Stochastic parameter estimation procedures for hydrologie rainfall-runoff models: Correlated and heteroscedastic error cases. Water Resources Res. 16(2), 430-442.

Sunde, M., He, H., Hubbart, J., Scoggins, C., 2016. Forecasting streamflow response to increased imperviousness in an urbanizing Midwestern watershed using a coupled modeling approach. Appl. Geogr. 72, 14-25.

Tasdighi, A., Arabi, M., Osmond, D.L., 2017. The relationship between land use and vulnerability to nitrogen and phosphorus pollution in an urban watershed. J. Environ. Qual. 46, 113-122.

Taylor, S.D., He, Y., Hiscock, K.M., 2016. Modelling the impacts of agricultural management practices on river water quality in Eastern England. J. Environ. Manag. 180(15), 147-163.

USDA-NRCS, 2004. Chapter 10: Estimation of direct runoff from storm rainfall, in: NRCS National Engineering Handbook, Part 630: Hydrology. U.S. Department of Agriculture, Natural Resources Conservation Services, Washington. D.C. http://www.wcc.nrcs.usda.gov/ftpref/wntsc/H&H/NEHhydrology/ch10.pdf/ (accessed 1 Dec. 2015).

USDA-ARS, 2014. ArcSWAT 2012.10_1.13. U.S. Department of Agriculture, Agricultural Research Service, Washington, D.C. http://swat.tamu.edu/ (accessed 1 Dec. 2015).

USDA-NRCS, 2016. USDA-NRCS Geospatial data gateway. Available at: https://datagateway.nrcs.usda.gov/ (accessed 1 Dec 2015).

USGS-NHD, 2016. National hydrography dataset. Available at: https://nhd.usgs.gov/ (accessed 1 Dec. 2015).

USGS-NWIS, 2016. National Water Information System, Available at: https://waterdata.usgs.gov/nwis (accessed 1 Dec. 2015).

USGS-TNM, 2016. TNM Map viewer, available at: https://viewer.nationalmap.gov/viewer/ (accessed 1 Dec 2015).

Vogel, R.M., Fennessey, N.M., 1994. Flow-Duration Curves. I: New Interpretation and Confidence Intervals. J. Water Resources Plan. Manag. 120(4), 485-504.

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

Vrugt, J.A., ter Braak, C.J.F., Diks, C.G.H., Robinson, B.A., Hyman, J.M., Higdon, D., 2009. Accelerating Markov Chain Monte Carlo Simulation by Differential Evolution with Self-Adaptive Randomized Subspace Sampling. Int. J. Nonlinear Sci. 10, 273-290.

Vrugt, J.A., 2016. Markov chain Monte Carlo simulation using the DREAM software package: theory, concepts, and MATLAB implementation. Environ. Modell. Software. 75, 273-316.

Wagner, P.D., Bhallamudi, S.M., Narasimhan, B., Kantakumar, L.N., Sudheer, K.P., Kumar, S., Schneider, K., Fiener, P., 2016. Dynamic integration of land use changes in a hydrologic assessment of a rapidly developing Indian catchment. Sci. Total Environ. 539, 153-164.

Wilcox, B.P., Rawls, W.J., Brakensiek, D.L., Ross Wight, J., 1990. Predicting runoff from rangeland catchments: a comparison of two models. Water Resources Res. 26(10), 2401-2410.

Winchell, M.R., Srinivasan, R., Di Luzio, M., Arnold, J.G., 2007. ArcSWAT interface for SWAT2005: user's guide. USDA-ARS Blackland Research Center.

Xu, Y.-P., Zhang, X., Ran, Q., Tian, Y., 2013. Impact of climate change on hydrology of upper reaches of Qiantang River Basin, East China. J. Hydrol. 484, 51-60.

Yan, B., Fang, N.F., Zhang, P.C., Shi, Z.H., 2013. Impacts of land use change on watershed streamflow and sediment yield: An assessment using hydrologic modelling and partial least squares regression. J. Hydrol. 484, 26-37.

Yang, X., Liu, Q., He, Y., Luo, X., Zhang, X., 2016. Comparison of daily and sub-daily SWAT models for daily streamflow simulation in the upper Huai river basin of China. Stoch. Environ. Res. Risk Assess. 30, 959-972.

Yen, H., Wang, X., Fontane, D.G., Harmel, R.D., Arabi, M., 2014. A framework for propagation of uncertainty contributed by parameterization, input data, model structure, and calibration/validation data in watershed modeling. Environ. Modell. Software. 54, 211-221.

Zhou, F., Xu, Y. Chen, Y., Xu, C.Y., Gao, Y., Du, J., 2013. Hydrological response to urbanization at different spatio-temporal scales simulated by coupling of CLUE-S and the SWAT model in the Yangtze River Delta region. J. Hydrol. 485, 113-125.

Zuo, D., Xu, Z., Yao, W., Jin, S., Xiao, P., Ran, D., 2016. Assessing the effects of changes in land use and climate on runoff and sediment yields from a watershed in the Loess Plateau of China. Sci. Total Environ. 544, 238-250.

Chapter 4

A BAYESIAN TOTAL UNCERTAINTY ESTIMATION FRAMEWORK FOR ASSESSMENT OF MANAGEMENT PRACTICES USING WATERSHED MODELS

Highlights

A Bayesian total uncertainty assessment framework is presented to assess the effectiveness of the watershed management practices in reducing nonpoint source water pollutants. The framework entails a two-stage procedure. First, various sources of modeling uncertainties are characterized during the period before implementing Best Management Practices (BMPs). Second, the effectiveness of the BMPs are probabilistically quantified during the post-BMP period. Various sources of modeling uncertainties are accounted for including uncertainties from model parameters, inputs, structure, and measurements. The framework was used to assess the uncertainties in effectiveness of two BMPs, nutrient management and cattle exclusion fencing, in reducing daily total nitrogen (TN) loads in a 54 hectare agricultural watershed in North Carolina using the SWAT model. The results indicate that the modeling uncertainties in quantifying the effectiveness of BMPs in reducing TN loads are relatively large. However, most of the uncertainty stems from the model parameters and not BMP parameters. In general, the 95% prediction intervals from the model with curve number method based on plant evapotranspiration covered a higher percentage of observed daily TN loads. Assessment of measurement uncertainty revealed that higher errors are observed in simulating TN loads during high flow events. Between the two BMPs, nutrient management had the highest impact on the TN load reductions. The parameters pertaining to cattle exclusion fencing showed low sensitivity when quantifying daily TN loads. The results of this study have important implications for decision making when models are used for water quality simulation. The framework presented in this study is deemed a pioneer in incorporation of robust uncertainty estimates in decision making for water quality management.

4.1. Introduction

Watershed management practices are used to abate nonpoint source pollution at field to watershed scales. Yet their effectiveness has been a subject of debate due to challenges in monitoring and modeling these practices (Arabi et al., 2007; Park et al, 2011). Studies have quantified the effectiveness of nonpoint source conservation practices using monitoring campaigns (Clausen et al., 1996; Bishop et al., 2005; Byers et al., 2005; Line et al., 2016). Most of these studies employ a paired watershed approach for assessing the water quality benefits of conservation practices. However, this approach can be costly and the results obtained are often site-specific and inconsistent between studies (Harmel et al., 2014). Hence, models are increasingly used along with the monitoring data to better assess the effectiveness of best management practices (BMPs; Santhi et al., 2003; Arabi et al., 2006; Lin et al., 2009; Park and Roesner, 2012; Taylor et al., 2016; Jang et al., 2017).

The majority of studies that investigated the effectiveness of BMPs using models have used a deterministic approach which often entails calibrating a watershed model by changing the parameters in order to get a good fit between model simulations and measured observations (Lin et al., 2009; Ullrich et al., 2009; Liu et al., 2013; Jingyuan et al., 2014; Motallebi et al., 2017). The calibrated model is then tested against a record of observations from a different time period to examine the performance validity of the model. The model is then used to quantify the effectiveness of a specific BMP or a combination of BMPs by changing some parameters that represent the effects on hydrologic and water quality processes and responses (Arabi et al., 2008; Liu et al., 2013; Jang et al., 2017). This approach can be inadequate, and in many cases misleading due to lack of accounting for different sources of modeling uncertainties (Ajami et al., 2007; Arabi et al., 2007; Tasdighi et al., 2017). Propagation of errors from different sources into model predictions during any modeling practice may result in biased and unrealistic decisions.

Simulation models are mere representation of reality with assumptions about natural and human processes that invariably result in uncertainty in model predictions. The uncertainty in any modeling

practice stems from different sources including: model parameters, input data (climate, land use, etc.), model structure (conceptualization), and measurement data (streamflow, nutrient concentrations or loads, etc.). While literature is replete with studies on modeling uncertainties of hydrologic and water quality processes (Beven and Binley, 1992; Vrugt et al., 2003; Ajami et al, 2007; Harmel et al., 2014; Yen et al., 2014; Vrugt, 2016), propagation of modeling uncertainties forward into the assessment of the effectiveness of BMPs has not been addressed sufficiently (Arabi et al., 2007). Specifically, a framework that incorporates various sources of modeling uncertainty to determine prediction intervals for BMP effectiveness has not been fully developed.

An important source of uncertainty when assessing the effectiveness of BMPs using a model is the measured data used for training and testing the model. Measurements of streamflow and water quality concentrations include substantial errors due to monitoring design, instrumentation, data processing, storage, and operator (human) errors (Harmel et al., 2007). Often, water quality loads are used for training and testing the models. Loads are usually derived either by multiplying measurements of streamflow and corresponding concentration of water quality constituent or using statistical load estimation techniques such as LOADEST. In any case, computing loads entails large errors which must be taken into account when simulating water quality. Propagation of measurement uncertainty in modeling the effects of conservation practices has been neglected in previous studies (Arabi et al., 2007).

A key challenge in estimation of uncertainties associated with BMP effectiveness is discerning among the uncertainty in general model parameters and the uncertainty in specific parameters that are used to represent the BMPs. Assessing the effectiveness of nonpoint source BMPs often entails a two stage approach. First, a model should be developed for quantifying the nonpoint source pollution. Second, by changing some parameters in the model which resemble the operation of BMPs, effectiveness of BMPs is quantified. This two-stage approach makes the uncertainty analysis of BMPs cumbersome and time consuming. Within an uncertainty assessment framework, at the first stage for quantifying nonpoint source pollution, instead of one optimal set of parameters (deterministic approach), each parameter has a

probability distribution. Some of these parameters and/or other additional parameters should be changed in the second stage to reflect the operation of BMPs. Hence, adopting a probabilistic approach for estimating these BMP parameters can quickly result in a large number of model runs. In this regard, a technique that can be used to efficiently conduct the BMP uncertainty analysis without compromising the statistical inferences gained during any step of the analysis is essential.

The overall goal of this study is to develop a probabilistic approach to assess the effectiveness of nonpoint source conservation practices in reducing nutrient loads. The specific objectives were to: 1) quantify the nutrient loads under different sources of modeling uncertainty, 2) assess the effects of measurement uncertainties on simulating nutrient loads, and 3) determine total uncertainty bounds around the effectiveness of BMPs. While previous studies have determined the effectiveness of nonpoint source BMPs through monitoring campaigns or deterministic modeling approach, using a total uncertainty assessment framework to probabilistically assess the effectiveness of BMPs is novel. Specifically, incorporation of measurement uncertainty which plays an important role in realistic simulation of nonpoint source nutrient loads, can bridge the gap in the literature for assessing the water quality benefits of nonpoint source conservation practices under uncertainty.

4.2. Material and methods

The SWAT model was used to represent hydrological and water quality processes in the study area. Three separate SWAT models were developed with different rainfall-runoff separation mechanisms to account for model structure uncertainty. A two-stage Bayesian total uncertainty assessment framework was used to assess the effectiveness of the BMPs in reducing daily total nitrogen (TN) loads at the outlet of the study watershed. First, hydrological and water quality data from 2008 to 2011, representing the pre-BMP conditions, was used to characterize and modeling uncertainties. Subsequently, data from 2012 to 2015, representing the post-BMP period, was used along with posterior distributions from the pre-BMP conditions to quantify uncertainty in uncertainties in parametric representation of BMPs as well as

prediction uncertainty in estimated reduction of nutrient loads due to implementation of BMPs. Since training the model during the pre-BMP period results in large number of parameters, the sizes of the posterior distributions of model parameters from pre-BMP analysis were reduced while maintaining their statistical characteristics using a uniform random sampling algorithm combined with a linear interpolation technique. The DREAM Markov Chain Monte Carlo (MCMC) sampling scheme (Vrugt et al., 2004; Vrugt, 2016) with a statistically correct likelihood function (Sorooshian and Dracup, 1980; Vrugt, 2016) were used to implement the total uncertainty assessment approach. The measurement uncertainty was incorporated by application of correction factors in the likelihood function. Since the observed and simulated data were log-transformed, Taylor series expansion for the second moment of a function of random variable was used to approximate the standard deviation required to compute correction factors.

4.2.1. Study watershed

The study watershed was a 54 hectare (ha) pasture-dominated watershed located in central North Carolina (Fig. 4.1). The land use composition within the watershed is 78% pasture, 14% forest, 6% developed, and 2% cultivated crops (National Land Cover Database; NLCD2011). The watershed was selected due to availability of comprehensive monitoring data (streamflow, nutrients, and sediments) from a published paired watershed study (Line et al., 2016). More importantly, the monitoring dataset provides a record of event-based measurements for 4 years before installing BMPs, and another 4 years after BMPs were installed.

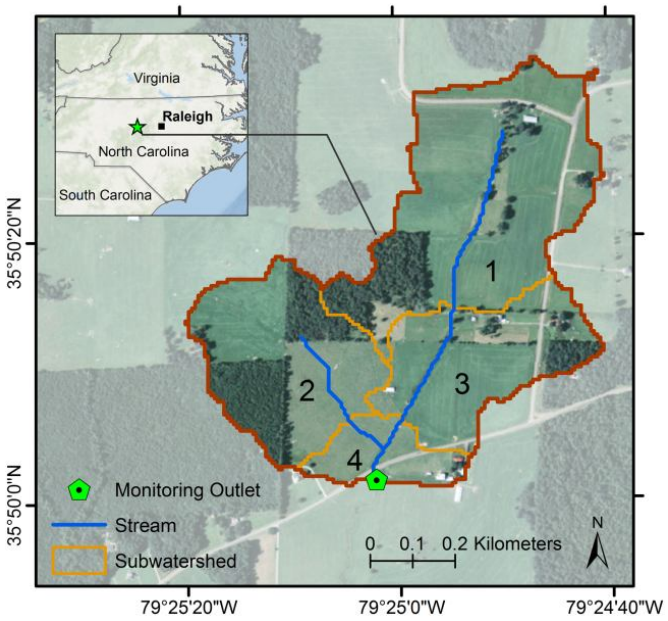

Fig. 4.1. Map of the study watershed. The numbers on the map are subwatershed numbers.

The main stream in the watershed known as the Mud Lick Creek, flows much of the year except some periods during late summer and early fall where it dries up. In general, the streams in this region are known for low baseflow (Line et al., 2016).

4.2.2. Watershed model

SWAT is a continuous-time, distributed-parameter, process-based watershed model, which has been used extensively for hydrologic and water quality assessments under varying climatic, land use, and management conditions in small watersheds to large river basins (Gassman et al., 2007; Arnold et al., 2012, CARD Staff, 2016). The model has the capability to run on daily or smaller time steps. In SWAT, the watershed is split into smaller subwatersheds, which are further discretized into Hydrologic Response Units (HRUs). HRUs are the smallest spatial units in SWAT, and are defined as areas within each subwatershed with unique combinations of land use, soil, and slope class.

Climate inputs drive hydrologic responses and provide moisture and energy inputs in SWAT. Hydrologic processes simulated in the model include canopy storage, surface runoff, infiltration, evapotranspiration, lateral flow, tile drainage, redistribution of water within the soil profile, return flow, and recharge (Arnold et al., 2012). Surface runoff is simulated using either the modified Green and Ampt (G&A; Green and Ampt, 1911) with subdaily rainfall or Curve Number method (CN; USDA-NRCS, 2004) with daily rainfall.

4.2.3. Model inputs: Terrain, Soils, Land use, Climate, and Hydrography

The elevation data for building the SWAT model was the 1/3 arc-second (~10 m) resolution digital elevation model (DEM) obtained from United States Geological Survey The National Map (USGS TNM, 2016). The Soil Survey Geographic (SSURGO) database from United States Department of Agriculture Natural Resources Conservation Services (USDA-NRCS, 2016) was used to represent soil characteristics and variability in the watershed. The National Agricultural Statistics Service (NASS) land use data for year 2011 was obtained from the USDA (USDA-CropScape, 2016). The resolution of the SSURGO and NASS data was 1 arc-second (~30 m).

Stream flowlines were obtained from USGS National Hydrography Dataset (USGS-NHD, 2016). The stream flowlines were used for more accurate stream delineation in SWAT. By superimposing the NHD flowlines on the DEM in the process of watershed and stream delineation, the hydrographic segmentation, subwatershed boundary, and stream delineation is improved especially in smaller scales such as the study watershed or locations where the DEM does not provide enough accuracy (Winchell et al., 2007).

Observed climate data for the closest meteorological station were obtained from the National Climatic Data Center (NCDC) Quality Controlled Local Climatological Data (QCLCD) database (NOAA, 2016). Daily and hourly precipitation, minimum and maximum temperature were collected for 2005 to 2015.

4.2.4. Measurements: Stream discharge and nutrient concentrations

The monitoring station was located at the outlet of the watershed (Fig. 4.1). Streamflow and nutrient measurements were available from 2008 to 2015. This period includes pre-BMP (1 Jan. 2008 to 5 Oct. 2011) and post-BMP (6 Oct. 2011 to 31 Dec. 2015) periods. Stream discharge and nutrient concentrations were sampled during storm events. The discharge measurement and nutrient sampling procedures (collecting, storing, and analyzing) are explained in detail in Line et al. (2016). Nutrient loads were computed by multiplying the mean storm discharge and corresponding nutrient or sediment concentration.

4.2.5. Cattle grazing and manure deposition

There were 75 beef cows grazing on various parts of the pasture within the watershed. Based on the observations and communication with land owners, Line et al. (2016) reported a density of 1.2 cows per hectare which was used in this study for computing grazing and manure deposition rates. Manure deposition rate for beef cows was set at 29.5 kg/day/500-kg-animal (USDA-NRCS, 2016). Assuming an average weight of a beef cow equal to 600 kg, the average manure deposition rate was computed as 35.4 kg/ha/day. The cattle deposit their manure either on land or in streams. On average, cattle spend 7.0% of their time in the streams (Byes et al., 2005). Therefore it was assumed that 2.5 kg/ha/day of manure was directly deposited into streams while the remaining fraction (32.9 kg/ha/day) was deposited on the land. Each beef cow consumes on average 45 kg/ha/day of grass (USDA-NRCS, 2016) which along with the 1.2 cows per hectare density gives about 54 kg/ha/day of grazing rate.

4.2.6. The SWAT model setup

ArcSWAT 2012 (USDA-ARS, 2014) was used to develop three SWAT models. The models were completely identical except for the runoff estimation method and precipitation time step. One of the models was developed with daily precipitation and CN method using the soil moisture (CN I) and plant evapotranspiration (CN II), and the other model was developed using the hourly precipitation and G&A method for runoff simulation. Therefore, three model setups were prepared for the analyses. A 2-year

warm up period was used when running the models to ensure that initial conditions for hydrological responses are adequately represented.

The DEM was used along with the NHD flowlines to delineate the watershed and streams. The watershed was divided into 4 subwatersheds. The HRU definition was done using the land use and soil data. Since the topographic variability was small, a single class slope was assumed within each subwatershed. Using these settings, 29 HRUs were defined for the watershed.

4.2.7. Representing BMPs in the SWAT model

The BMPs implemented in the watershed were nutrient management and cattle exclusion fencing. While nutrient management can be readily modeled by changing the rate of fertilizer application (Arabi et al., 2008; Ahmadi et al., 2014), incorporating cattle exclusion fencing in the model is more complicated. Few studies have discussed methods for representing cattle exclusion fencing in SWAT. One such method is to use point sources to represent the presence of cattle in streams (Lin et al., 2009). This approach provides the capability to directly change the rate of nutrients or sediments introduced into the streams during the implementation of the exclusion fencing resembling the operation of the BMP (limiting the access of the cattle to streams).

During the pre-BMP period, 168 kg/ha of 15-15-15 fertilizer was applied uniformly on the pasture. Biosolids were also applied on the pasture. Analysis of biosolids showed that 130 and 116 kg/ha of N and P were applied at each application. Nutrient management was implemented by replacing the biosolids and fertilizer with a granular N only fertilizer at a rate of 70 kg/ha during the post-BMP period. This BMP was directly implemented in the SWAT model by replacing and changing the application rate of the fertilizers.

The exclusion fencing was installed in October 2011. The fences were installed along approximately 520 m of the mainstream of the Mud Lick Creek in the watershed. About half of the upper part of the mainstream was not fenced. So, the cattle were still able to have access to streams in the upper half part of

the watershed. We modeled the cattle exclusion fencing by changing the rate of organic N introducing into the streams via point sources in subbasins 3 and 4 where stream fencing was installed on streams.

4.2.8. The BMP uncertainty assessment framework

A probabilistic framework was developed for assessing the effectiveness of nonpoint source conservation practices. The Bayesian-based approach explicitly accounts for uncertainties from model parameterization, climate input data (i.e. precipitation), model structure (CN I, CN II, or G&A), and measurement data (i.e. streamflow and nutrient loads).

The framework entails a two-stage procedure. First the modeling uncertainties in simulating nutrient loads are characterized during the pre-BMP period (2008-2011). During this stage, a Markov Chain Monte Carlo (MCMC) sampling scheme, the DREAM method (Vrugt et al., 2009), along with a statistically correct likelihood function (Sorooshian and Dracup, 1980; Vrugt, 2016) is used to sample the parameter space and derive the posterior distributions. Input data uncertainty is incorporated by using precipitation multipliers drawn from a Gaussian distribution with an uncertain mean and standard deviation sampled along with model parameters during the MCMC procedure. The measurement uncertainty is also incorporated by applying correction factors on model residuals ($observed - simulated$) computed based on a probable error range for each measurement (Harmel et al., 2007). Finally, Bayesian Model Averaging (BMA) is used to account for model structural uncertainty (Hoeting et al., 1999). Having characterized different sources of modeling uncertainties in the first stage, the model can be used for simulation of nutrient loads during the post-BMP period (2012-2015) without incorporating the BMPs in models. The outputs provide prediction intervals for nutrient loads during the post-BMP period assuming no BMPs were implemented.

The second stage entails iterating on BMP parameters for each model parameter set (from stage 1) to characterize the performance of the BMPs under various sources of modeling uncertainty during the post-BMP period (2012-2015). The first stage often results in too many parameter sets due to large number of

iterations required for the MCMC algorithm to converge especially in highly parametrized complex models such as SWAT. A method was proposed to reduce the number of model parameter sets without compromising the statistical characteristics of the inferred posterior distributions. The method involves a uniform random sampling scheme along with a linear interpolation algorithm during which model parameter sets are drawn randomly from the posterior distributions while maintaining statistical characteristics of the distributions. The BMP parameters were then sampled for each set of the reduced model parameter sets using a uniform sampling scheme. The model is then run using each new joint parameter set (model parameters and BMP parameters) and the posterior distribution of BMP parameters are derived with the new likelihood function during the post-BMP period. The outputs from this stage provide prediction intervals for simulated nutrient loads during the post-BMP period after BMPs are implemented which along with nutrient loads from stage 1 can be used to compute prediction intervals for nutrient load reductions. Availability of monitoring data during the post-BMP conditions provides a unique opportunity to compute values of a new likelihood function in contrast to previous studies where the same likelihood function was used during both pre-BMP and post-BMP period due to lack of measurements during post-BMP conditions (Arabi et al., 2007).

4.2.8.1. Model and BMP parameters uncertainty

The generic equation for quantifying the effectiveness of a specific BMP using simulations from a watershed model is:

$$\eta_{BMP} = \hat{Q}_{pre-BMP} - \hat{Q}_{post-BMP} \quad (1)$$

where η_{BMP} denotes the efficiency or effectiveness of the BMP, $\hat{Q}_{pre-BMP}$ and $\hat{Q}_{post-BMP}$ are the simulated response variables before and after application of the BMP respectively. The parameter uncertainty in quantifying the simulated response variables in Eq. 1 is due to (i) model parameters (θ_M), and (ii) BMP parameters (θ_{BMP}). It should be noted that the uncertainty from model parameters propagates to predictions during both pre-BMP and post-BMP periods. However, the BMP parameter

uncertainty is only manifested in the post-BMP conditions. The simulated response variables are also subject to other errors stemming from measured climate inputs (R), insufficiency of the model conceptualization (M; model structural error), and measurement data used for training and testing the model ($Q_{pre-BMP}$ and $Q_{post-BMP}$). The simulation error residuals (e) then take the form:

$$e(R, \theta, M, Q) = Q - \hat{Q} = Q - M(R, \theta) \quad (2)$$

During the post-BMP conditions, the error residuals are a function of joint probability density of model and BMP parameters, $e = f(\theta_M, \theta_{BMP})$, with expected value of:

$$E(e) = \iint f(\theta_M)\, f(\theta_{BMP}) d_{\theta_{BMP}} d_{\theta_M} \quad (3)$$

where $E(e)$ denotes the expected value of the error residual as a function of both model and BMP parameters. The Eq. 3 is conditioned on the independence of model and BMP parameters. Analytical solution to this equation is often infeasible in case of complex models. Hence, we resort to MCMC methods. Applying the Bayes theorem, the parameter set θ is assigned posterior probability distribution, $p(\theta|Q)$, which is proportional to the product of the parameter prior probability distribution, $p(\theta)$, and a likelihood function, $L(\theta|Q)$.

The likelihood function assuming normally and independently distributed model residuals (e) with mean zero and variable standard deviation at each observation time step (σ_t), can be expressed as (Vrugt, 2016):

$$L(\theta|Q, \sigma) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma_t^2}} exp\left[-\frac{1}{2\sigma_t^2}\left(Q_t - \hat{Q}_t(\theta)\right)^2\right] \quad (4)$$

where n is the number of time steps. Often it is easier to maximize the logarithm of the likelihood function due to numerical stability and algebraic simplicity (Ajami, 2007; Vrugt, 2016). Hence, the natural log of the likelihood function was adopted for optimization.

Streamflow and nutrient error residuals are often not independently distributed and in most cases temporal autocorrelation exists in the residuals. A first order auto-regressive scheme (AR-1) was employed to reduce autocorrelation. Applying the natural log and AR-1 transformation, the log-likelihood function takes the form:

$$l(\theta|Q,\sigma,\rho) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln\left(\frac{\sigma_v^2}{1-\rho^2}\right) - \frac{1}{2\sigma_v^2}\left((1-\rho^2)e_1^2 + \sum_{t=2}^{n}(e_t - \rho e_{t-1})^2\right) \quad (5)$$

Parameters ρ and σ are determined along with model parameters at each model realization during the MCMC sampling algorithm.

The model parameters for the uncertainty analysis were selected based on experience and sensitivity analysis performed previously (Arabi et al., 2007; Arnold et al., 2012; Tasdighi et al., 2017). Table 4.1 lists the model parameters selected for uncertainty analysis along with their ranges. The ranges for parameters were also selected based on the SWAT user manual and experience from previous study (Tasdighi et al., 2017; Arnold et al., 2012). In this study, uniform (noninformative) prior distributions were assumed for parameters within predefined ranges. The same assumption has been used in many other hydrological modeling studies since prior knowledge of the model parameters is often not available and is case-specific (Ajami et al., 2007).

Table 4.1. Model parameters chosen for uncertainty analysis in this study.

Parameter	Input file	Description	Lower bound	Upper bound
ALPHA-BF	.gw	Base flow alpha factor for recession constant	0.001	1
GWHT	.gw	Initial groundwater height	0	25
CANMX	.hru	Maximum canopy storage	0	10
ESCO	.hru	Soil evaporation compensation factor	0.5	1
OV-N	.hru	Manning's n value for overland flow	0.01	0.6
SLSUBBSN	.hru	Average slope length	10	150
RSDIN	.hru	Initial residue cover	0	10000
CH-KII	.rte	Effective hydraulic conductivity in the main channel	0.025	150
CH-NII	.rte	Manning's n value for the main channels	0.01	0.3
EPCO	.bsn	Plant uptake compensation factor	0.001	1
SURLAG	.bsn	Surface runoff lag coefficient	0.001	15
CNCOEF	.bsn	Plant ET curve number coefficient	0.5	2
CDN	.bsn	Exponential rate of denitrification	0	1
CMN	.bsn	Rate factor for mineralization of organic nutrients	0.001	0.003
NPERCO	.bsn	Nitrogen percolation coefficient	0.01	1
RCN	.bsn	Concentration of nitrogen in rainfall	0	15
RSDCO	.bsn	Residue decomposition coefficient	0.02	1
SDNCO	.bsn	Denitrification threshold water content	0	1
SOL-ALB	.sol	Moist soil albedo	-0.5	1
SOL-AWC	.sol	Fraction change in available soil water capacity	-0.5	1
SOL-K	.sol	Fraction change in saturated hydraulic conductivity	-0.5	2
SOL-Z	.sol	Fraction change in soil depth	-0.5	1
BIOMIN	.mgt	Minimum biomass for grazing	0	5000
BIOMIX	.mgt	Biological mixing efficiency	0	1
CN-F	.mgt	Fraction change in SCS runoff curve number	-0.2	0.2
ORGN	.chm	Initial organic N in soils	1	10000
SOLN	.chm	Initial NO_3 in soils	0.1	5

The BMP parameters for uncertainty analysis were selected based on the SWAT capabilities in representing BMPs and studies where similar BMPs were implemented in SWAT (Arabi et al., 2008; Lin et al., 2009; Ahmadi et al., 2014). Table 4.2 lists the parameters selected for uncertainty analysis of BMPs. Similar to model parameters, uniform (noninformative) prior distributions were assumed for BMP parameters as well. The rate of fertilizer application was selected to represent nutrient management practices. The rate of organic N introducing streams via point sources in subbasins 3 and 4 where cattle

exclusion fencing was implemented were selected as parameters for simulating exclusion fencing in SWAT.

Table 4.2. BMP parameters chosen for uncertainty analysis in this study.

Practice	Parameter	Input file	Description	Lower bound	Upper bound
Fertilizer management	FRT-KG	.mgt	Fertilizer application (kg/ha)	10	155
Cattle exclusion fencing	ORGNCNST (sub3)	ptsrc3.dat	Direct manure deposition as N in streams (kg/d)	0.08	0.25
	ORGNCNST (sub4)	ptsrc4.dat	Direct manure deposition as N in streams (kg/d)	0.01	0.12

4.2.8.2. Model input uncertainty

A method proposed by Ajami et al. (2007) was implemented to account for precipitation uncertainty through application of multipliers on precipitation events. Using this method, instead of iterating on each single multiplier, the iteration is performed on the mean and standard deviation of a random Gaussian distribution from which the multipliers are randomly drawn at each time step:

$$\hat{R}_t = \phi_t R_t; \qquad \phi_t \sim N(\mu_\phi, \sigma_\phi^2) \tag{6}$$

where $\hat{R}_t$ and R_t are the corrected and observed precipitation depths respectively, ϕ_t is the random multiplier drawn from a normal distribution with a random mean μ_ϕ, $\mu_\phi \in [0.9,1.1]$ and variance, σ_ϕ^2, $\sigma_\phi^2 \in [1e-5,1e-3]$ (Ajami et al., 2007). Incorporating precipitation multipliers using this approach instead of iterating on each precipitation multiplier reduces the dimensionality issue and improves the identifiability.

4.2.8.3. Model structural uncertainty

Bayesian Model Averaging (BMA) was used to account for model structural uncertainty (Hoeting et al., 1999; Georgakakos et al., 2004). The BMA is a probabilistic algorithm for combining competing models based on their predictive skills (Ajami et al., 2007; Madadgar and Moradkhani, 2014). In this study, the three rainfall-runoff model structures in SWAT (M_1: $CN\ I$, M_2: $CN\ II$, M_3: $G\&A$) were used to explore the effects of model structural uncertainty. Using the BMA, the three model structures were combined using probabilistic weights to reduce the model structural uncertainty. The posterior distribution of the BMA prediction ($\hat{Q}_{BMA}$) is:

$$p\left(\hat{Q}_{BMA}\middle|M_1, M_2, M_3, Q\right) = \sum_{i=1}^{3}\left[p(M_i|Q) \times p_i(\hat{Q}_i|M_i, Q)\right] \quad (7)$$

where $p(M_i|Q)$ is the posterior probability of the model M_i. This term can be assumed as a probabilistic weight (w_i) for model M_i in the BMA prediction ($\hat{Q}_{BMA}$). The constraint for BMA weights is: $\sum_{i=1}^{3} w_i = 1$. Higher values of w_i can be interpreted as higher predictive skill for a given model structure. The model weights can be determined using different optimization techniques. The expectation-maximization (EM) algorithm (Dempster et al., 1977) was used in this study to estimate model weights.

4.2.8.4. Incorporating the measurement uncertainty

The uncertainty inherent in measured data used for training and testing of models often stems from errors in monitoring design, instrumentation, data processing, storage, and operator (human) errors. This type of uncertainty is rarely accounted for in evaluation of model performance (Harmel et al., 2007; Yen et al., 2014). The measurement data uncertainty is often manifested in heteroscedasticity of error residuals (variable error variance). Some observations may be less reliable than others which results in their errors to have different variances. Different approaches have been proposed to circumvent this issue including application of different transformations such as natural log or Box-Cox to stabilize the error variances (Sorooshian and Dracup, 1980). Others have proposed alternative forms for the likelihood function

(Schoups and Vrugt, 2010) which accounts for error heteroscedasticity by assuming a variable variance for each model residual. The main challenge in this approach is determination of the variance for each residual as it requires having repeats of each measurement which most often are not available. As a result, in such studies, variance was subject to inference along with model parameters at each model realization during the sampling procedure (Schoups and Vrugt, 2010; Vrugt, 2016). This approach can lead to dimensionality issues in case of highly parameterized models or when a relatively long record of measurements is used for model training.

In this study, we employed a method based on the study by Harmel et al. (2007). In this method, each error residual is modified using a correction factor computed based on the properties of the probability distribution of each measured value. Previous experience and expert's opinion can be used to make informed assumptions about probability distribution for each measured value in the record. Assuming a normal distribution for each measured value (Q_i), mean and median of the distribution is represented by Q_i. The variance can then be computed based on a probable error range (PER) which can be assumed based on literature or professional judgment. PER can be constant or variable for all measurements depending on the experience and level of knowledge about the monitoring design. Once the PERs are determined, the variance (σ_i^2) for each record of measurement can be computed as:

$$\sigma_i^2 = \left(\frac{PER_i \times Q_i}{3.9 \times 100}\right)^2 \quad (8)$$

Eq. 8 is useful when the measured values are not transformed. In this study, a log transformation on simulated and measured streamflow and nutrient loads was used before computing the likelihood function as explained in section 4.2.8.1. To circumvent this issue, the Taylor series expansion for the second moment of a function of random variable ($h(X)$) was used to estimate the variance ($Var[h(X)]$) as:

$$Var[h(X)] \approx \left(\left.\frac{\partial h}{\partial X}\right|_{E(X)}\right)^2 Var[X] = (h'(\mu_x))^2 \sigma_X^2 \quad (9)$$

Applying this on the natural log-transform function gives:

$$Var[ln(X)] \approx \left(\frac{\sigma_X}{x}\right)^2 \quad (10)$$

$$SD[ln(X)] \approx \frac{\sigma_X}{x} = COV = PER \quad (11)$$

where SD stands for the standard deviation and COV denotes the coefficient of variation. Using this approach, σ for log transformed observations, can be estimated with PER. It should be noted that this assumption holds for smaller PERs (< 0.3) and higher approximation errors are introduced in higher values of PER. We used expert's opinion to determine PERs for each streamflow or nutrient load measurement (Q_i). The correction factors were then computed as the area under the standard normal distribution:

$$CF_i = F(X_i|\mu_i, \sigma_i) = \frac{1}{\sigma_i\sqrt{2\pi}} \int_{-\infty}^{X} e^{\frac{-(x-\mu)^2}{2\sigma_i^2}} dx - 0.5 \quad \text{if } \mu \leq X \quad (12a)$$

$$CF_i = F(X_i|\mu_i, \sigma_i) = -\frac{1}{\sigma_i\sqrt{2\pi}} \int_{-\infty}^{X} e^{\frac{-(x-\mu)^2}{2\sigma_i^2}} dx + 0.5 \quad \text{if } \mu > X \quad (12b)$$

where $F(.)$ denotes the normal cumulative distribution function, X denotes the simulated data ($\hat{Q}_i$) and μ represents the observed data (Q_i) and σ is determined using Eq. 11. Using the correction factors, the residuals are adjusted based on the estimated measurement uncertainty. It should be noted that this method requires some level of knowledge about the characteristics of the measurement errors. These characteristics are often case-specific. Experience form previous studies can be used to enhance the assumptions and generate better estimates.

4.2.8.5. The DREAM algorithm for MCMC analyses

Several Bayesian algorithms are available which have been widely used for uncertainty assessment in hydrologic modeling including the Generalized Likelihood Uncertainty Estimation (GLUE; Beven and Binley, 1992), the Shuffled Complex Evolution Metropolis (SCEM-UA; Vrugt et al., 2003), and the

DiffeRential Evolution Adaptive Metropolis (DREAM; Vrugt et al., 2009). DREAM is a multi-chain MCMC method that randomly samples the parameter space and automatically tunes the scale and orientation of the sampling distribution to move toward the target distribution by maximizing the value of the likelihood function. The method has been used extensively for parameter estimation of complex environmental models (Vrugt, 2016). The convergence of the algorithm can be monitored using the procedure proposed by Gelman and Rubin (1992). In this procedure, a scale reduction score (R) is monitored to check whether each parameter has reached a stationary distribution (Gelman and Rubin, 1992). The common convergence criterion of $R \leq 1.2$ was used in this study as well. DREAM is specifically beneficial in the optimization of complex high dimensional problems. In this study, DREAM was employed to sample the parameter space and derive the posterior distributions.

4.3. Results and discussion

4.3.1. Evaluation of models during the pre and post-BMP periods

The performance of the models in simulating daily TN loads under the total uncertainty assessment framework was evaluated using different error statistics computed at each model realization. During the pre-BMP period (2008-2011), the uncertainty analysis was performed using the values of the likelihood function (Eq. 5) computed for daily TN loads as the objective function at the monitoring station. The error statistics during the post-BMP (2012-2015) period were calculated using the joint distribution of reduced model parameters and BMP parameters. Table 4.3 summarizes the error statistics for models during the pre and post-BMP conditions.

Compared to models with CN, the model with G&A method had a poor performance in terms of various error statistics during the pre-BMP period and it was excluded from further analysis for that reason. Similar results were reported on better performance of the CN method over the G&A in other agricultural watersheds (Kannan et al., 2007; Cheng et al., 2016). In contrast, Ficklin and Zhang (2013)

concluded that models with G&A are more likely to generate better daily simulations in agricultural watersheds. It should be noted that these studies have used a deterministic approach.

Table 4.3. Summary of error statistics for pre and post-BMP conditions.

		Pre-BMP (2008-2011)			Post-BMP (2012-2015)		
Model	Statistic	Min	Med	Max	Min	Med	Max
CN I	Likelihood	-3.72E+05	-196	-183	-2.32E+04	-1081	-161
	RMSE	0.04	0.08	1.19	0.06	5.51	18.2
	NS	-0.05	0.06	0.62	-9.65	-0.08	0.36
CN II	Likelihood	-3.97E+04	-198	-185	-9.98E+04	-252	-167
	RMSE	0.06	0.08	0.73	0.05	4.3	10.2
	NS	0.03	0.12	0.66	-8.36	0.02	0.47
G&A	Likelihood	-7.14E+05	-574	-352			
	RMSE	0.06	0.09	6.6	-	-	-
	NS	-6.8	-0.22	0.25			

Tasdighi et al. (2017) compared the performance of the CN and G&A methods based on upstream land use conditions using a probabilistic approach. They concluded that the G&A method had a better performance in highly developed subwatersheds while the CN method had a slightly better performance in agricultural watersheds. Between models with CNI and CNII methods, CNII models had a slightly better performance during the pre-BMP period. During the post-BMP period the superiority of the CNII model was more accentuated generating better error statistics.

4.3.2. Characterizing the modeling uncertainties during the pre-BMP period (stage 1)

4.3.2.1. Model parameters uncertainty

The posterior Cumulative Distribution Functions (CDF) of parameters under the CNI and CNII model structures are illustrated in Fig. 4.2. The distributions are derived after the pre-BMP uncertainty analysis (2008-2011). Note that only the most sensitive parameters are included in this figure. As observed in Fig. 4.2, models with CNI and CNII resulted in different distributions for the same

parameters. While the model structure determined the level of skewness for most parameters, for some parameters (ESCO, SOLZ, CDN) the skewness changed from positive to negative under different model structures. In general, CNI model showed higher sensitivity to parameters pertaining to soil characteristics (SOL-ALB, SOL-AWC, and SOL-Z) which conforms to intuition as the CNI method uses the soil water content for determining the curve number. High skewness from normality indicates deficiency in identifiability which is often present in the uncertainty analysis of highly parametrized complex models (Ajami et al., 2007).

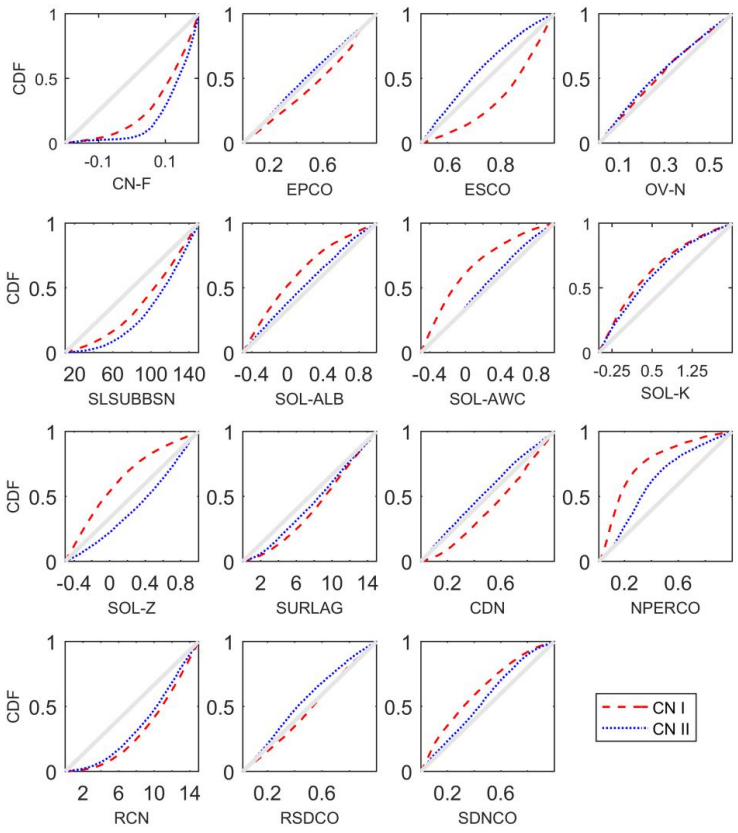

Fig.4.2. The cumulative posterior distribution (CDF) of model parameters during the pre-BMP period.

The distributions were derived using 20000 samples after the convergence of the DREAM algorithm. The uniform random sampling algorithm along with the linear interpolation technique discussed in section 4.2.8 was then used to reduce the number of parameter sets while maintaining the characteristics

of the distributions to be used for post-BMP analysis. The algorithm was set to draw 1000 parameter sets from the posterior distributions. The reduced parameter sets act as priors for the post-BMP period. In other words, the prior distributions of model parameters for the post-BMP period are inferred from the posterior distribution of parameters from the pre-BMP period. For each of the new parameter sets, the uniform random sampling technique discussed in section 4.2.8 was used to generate BMP parameter sets. Since only 3 parameters were used for BMPs, 30 random BMP parameter samples were assumed to adequately represent the BMP parameter space. The BMP parameter sets combined with the model parameter sets resulted in 30000 joint parameter sets for the second stage of the analysis (post-BMP).

4.3.2.2. Model input uncertainty

The cumulative distribution of mean and standard deviation of normal distributions from which precipitation multipliers were drawn were almost similar and close to uniform. The uniformity of the distributions indicates that precipitation multipliers did not have a major impact on TN load simulations from each model. The similarity of the distributions on the other hand, could be predicted as the precipitation data for models with CNI and CNII was identical daily time step measurements. Hence, any difference in distributions resulted from the model structural difference. Another explanation in this regard is the large number of sensitive model parameters which can diminish the impact of precipitation multipliers in the range assumed. Using a wider range for mean and standard deviation of the normal distributions from which precipitation multipliers were drawn could result in higher impact from multipliers and probably better performance of the input uncertainty estimation routine.

4.3.2.3. Model structural uncertainty

Bayesian model averaging was used at each model realization to probabilistically combine the models and reduce the model structural uncertainty. BMA weights were determined using EM optimization technique explained in section 4.2.8.3. Fig. 4.3 shows the boxplots of BMA weights generated during the MCMC procedure. Based on BMA weights, CNII had a better performance in

simulating daily TN loads. These results are congruent to the results in section 4.3.1 where CNII model showed superior performance in terms of various error statistics. One explanation for these results is that in CNII method, the curve number is determined based on plant evapotranspiration and since the study watershed is a pasture-dominated agricultural watershed, the CNII has a better performance. Similar results were obtained in other studies (Tasdighi et al., 2017; Yen et al., 2014).

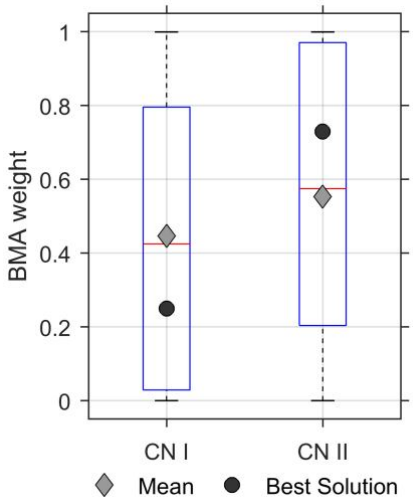

Fig. 4.3. BMA weights for the models (Solid horizontal lines in the boxes show the median; the boxes show the range of values between 25th and 75th percentile; the whiskers show the 0.5 and 99.5 percentile).

4.3.2.4. Measurement data uncertainty

Measurement data uncertainty was incorporated using correction factors. The correction factors were applied on the daily TN load residuals during the computation of the likelihood function at each model realization. Fig. 4.4 illustrates the probability distribution (PDF) of correction factors for each model during the pre-BMP period.

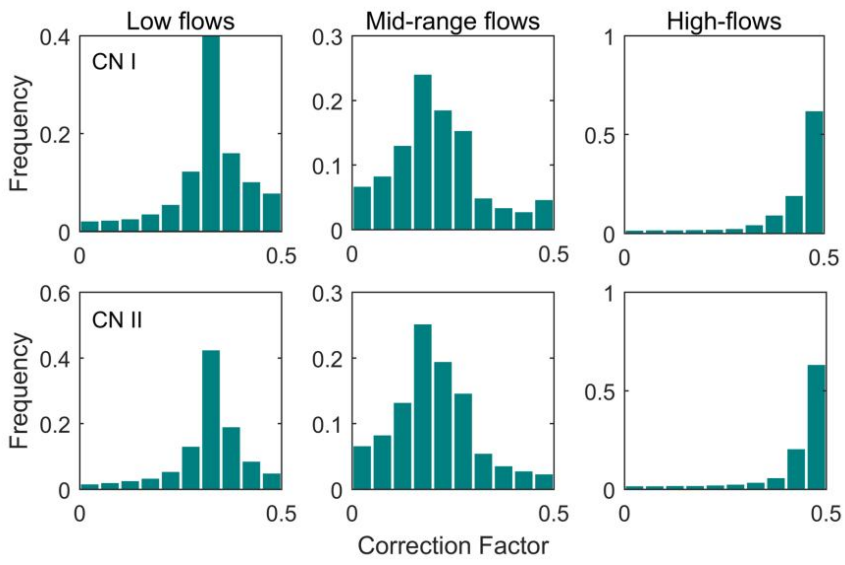

Fig. 4.4. PDF of TN load correction factors for CNI and CNII models under low, medium, and high flows conditions.

The correction factors were categorized based on flow conditions (low, medium, and high flows) before generating the PDFs to assess the effects of flow regime on measurement errors. The highest values of correction factors were observed during high flow events. Medium and low flow events resulted in relatively lower values for correction factors. This is an important finding as it indicates higher errors during high flow events compared to medium and low flows. The phenomenon of higher errors during high flow events is often described as the heteroscedasticity of error residuals which is deemed to be attenuated by applying correction factors. Other studies that have investigated the uncertainties in measurement data have reported similar behavior (Sorooshian and Dracup, 1980; Harmel et al, 2007). The results also conform to intuition as monitoring during high flow events often entails higher errors due to difficulties in measurements.

4.3.3. Assessing the effectiveness of BMPs under various sources of modeling uncertainty (stage 2)

4.3.3.1. BMP parameters uncertainty

The posterior distribution of BMP parameters are illustrated in Fig. 4.5. Interestingly, the parameter pertaining to nutrient management BMP (FRT-KG) showed higher sensitivity when quantifying the TN loads. This is while the posterior distributions of parameters pertaining to the cattle exclusion fencing were close to uniform which indicates lower sensitivity. This outcome also indicates that the nutrient management had a higher impact on nutrient load reductions. A possible explanation for this can be the higher uncertainty in the nature of representing cattle exclusion fencing in the SWAT model. Several assumptions were used when representing the exclusion fencing such as changing the nutrient introducing to the streams via point sources and the rate adjustments for nutrients introducing streams. A more rigorous approach for representing cattle exclusion fencing may enhance the performance of this BMP and result in more meaningful inferences. All BMP parameter posterior distributions showed high deviations from normality which indicates deficiencies in identifiability. Assuming wider prior distributions and larger number of iterations on BMP parameters may be effective in reducing these effects.

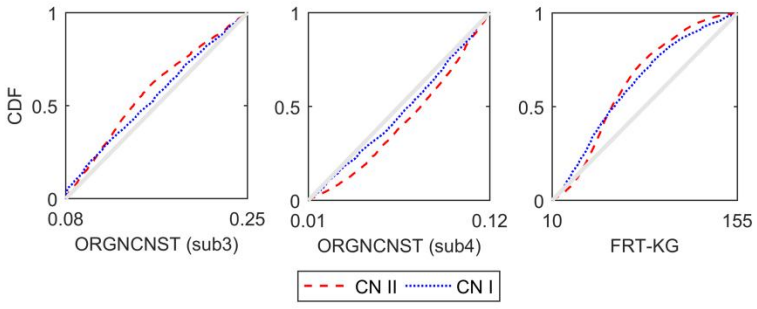

Fig. 4.5. The posterior distribution of BMP parameters. ORGCNST (sub3) and ORGCNST (sub4) are the parameters pertaining to cattle exclusion fencing in subwatersheds 3 and 4 respectively. FRT-KG is the BMP parameter for nutrient management.

4.3.3.2. Estimating TN load prediction intervals before and after implementation of BMPs

The cumulative exceedance probability curves for daily TN loads were developed. These curves along with bands of uncertainty before and after implementing BMPs provide an easily-readable informative measure for assessing the effectiveness of BMPs in reducing TN loads under uncertainty. Fig. 4.6 illustrates the 95% confidence interval for cumulative exceedance probability curves for TN loads.

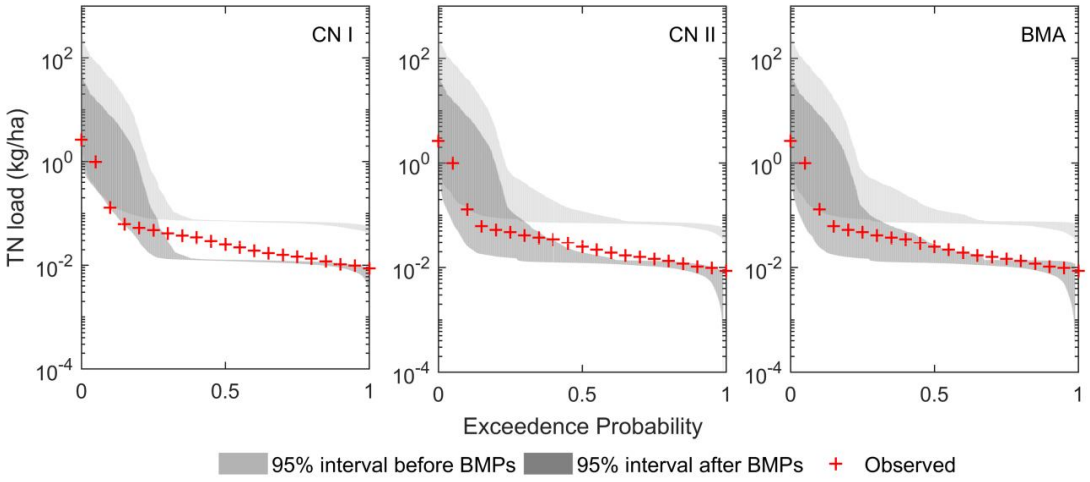

Fig. 4.6. 95% confidence intervals for cumulative exceedance probability of TN loads.

Comparing the prediction intervals before and after implementation of BMPs, it is evident that the combination of BMPs (nutrient management and cattle exclusion fencing) was successful in reducing the TN loads in the watershed. Compared to CNI, the model with CNII had a better performance in capturing TN loads at all ranges especially the high and medium loads. However, they showed wider bands of uncertainty. The BMA had a close performance to the CNII model which was expected as the CNII method better simulated the TN loads resulting in higher BMA weight.

While the cumulative exceedance probability curves provide a valuable insight into the statistical characteristic of BMPs' effectiveness, they can be misleading as the serial structure and autocorrelation of the sequence of the simulation record is removed in them (Vogel et al., 1994). In this regard, the fraction

of observations lying within the prediction intervals should be determined using the time series of simulations and observations. The coverage (percent of observations lying inside the 95% confidence interval of simulation ensembles) and spread (average width of the 95% confidence interval uncertainty band) were determined based on the record of simulation and observations of TN loads. Table 4.4 summarizes the coverage and spread for the models during the pre and post-BMP conditions.

Table 4.4. Coverage rates and spread for models. The coverage rates and spread are calculated using the 95% confidence interval of simulated TN loads and corresponding observations.

	Pre-BMP		Post-BMP	
Model	Coverage (%)	Spread (kg/ha)	Coverage (%)	Spread (kg/ha)
CN I	47	1.05	18	2.6
CN II	41	1.30	33	2.7
BMA	51	1.45	35	3.0

4.3.3.3. Quantifying the effectiveness of BMPs in terms of TN load reductions

The effectiveness of BMPs in reducing the TN loads was computed by subtracting the TN loads from models before and after implementation of BMPs during the common post-BMP period (2012-2015). Fig. 4.7 depicts the cumulative exceedance probability curves for TN load reductions under different models. The highest reductions were observed for higher loads. In general the results demonstrate the high level of uncertainty in simulating the daily TN load reductions. For example for CNII model in Fig. 4.7, at exceedance probability of 25%, the TN load reduction from BMPs can be any number between 0.1 to 1 kg/ha. This outcome indicates the importance of accounting for various sources of uncertainty in modeling the performance of the BMPs as they directly affect the decision making process. In general, modeling pollution loads from nonpoint sources is subject to high levels of uncertainty. However, most often this uncertainty is ignored and models are used deterministically to compute pollution loads which can result in unrealistic and biased decisions.

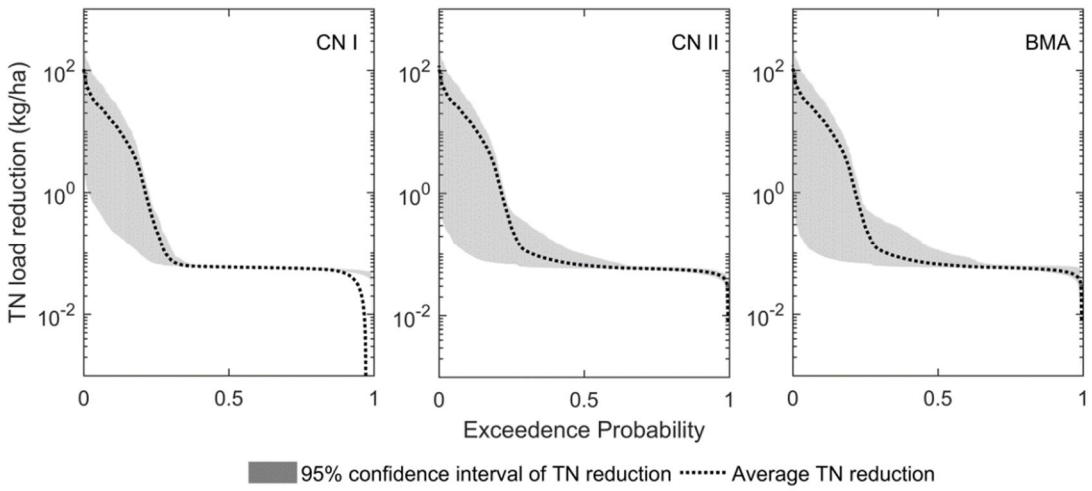

Fig. 4.7. 95% confidence intervals for cumulative exceedance probability of TN load reductions.

The average annual TN load reduction from the ensemble of CNI, CNII models, and BMA were 1.5, 1.7, and 1.8 kg/ha respectively. In terms of percentage, the average annual reductions were 58%, 65%, and 69% respectively. Fig. 4.8 shows the PDFs of average annual load reductions for each model. These values conform to the reductions determined by the paired watershed study on this watershed conducted by Line et al. (2016). They found statistically significant reductions in total Kjeldhal nitrogen (34%), and ammonia (54%) while changes in nitrate loads were not significant.

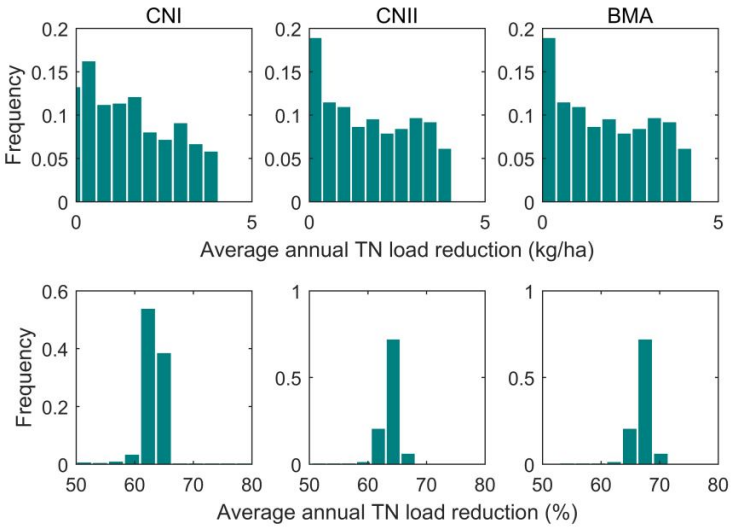

Fig. 4.8. The PDFs of average annual TN load reductions.

The TN load reductions were resulting from the performance of the nutrient management and cattle exclusion fencing combined. While it was not feasible to decompose the total TN load reductions between each BMP to determine how each BMP performed, the nutrient management was determined to be more effective as the related BMP parameters showed higher sensitivities when quantifying TN loads.

4.4. Conclusions

A total uncertainty estimation framework was developed for assessing the water quality benefits of management practices using watershed models. The two-stage framework first characterizes various sources of modeling uncertainties during the pre-BMP period. The second stage of the framework uses the inferences on modeling uncertainties from the first stage and quantifies the effectiveness of BMPs using a probabilistic approach.

In general, the modeling uncertainties were large resulting in wide bands of uncertainty around both TN loads and load reductions. The framework however was successful in capturing the effects of different sources of modeling uncertainties on simulations and propagating them to the BMP effectiveness assessment stage. Between the three model structures (CNI, CNII, and G&A), CNII showed the best performance in terms of various performance measures including error statistics and BMA weights. This was attributed to the intensive agricultural land use in the watershed. The G&A method had an unsatisfactory performance in simulating the TN loads. As shown in previous chapter, the G&A method excelled in simulating streamflow and hydrologic regime in highly developed subwatersheds. Application of the BMA slightly enhanced the quality of simulating TN load reductions under uncertainty as it resulted in higher coverage 51% and 35% during the pre and post-BMP period respectively. The distribution of the correction factors for measurement uncertainty indicated higher uncertainty for high flow events correctly capturing the heteroscedasticity of error residuals for daily TN load simulations.

Between the two BMPs, nutrient management had the highest impact on the TN load reductions. The parameters pertaining to cattle exclusion fencing did not show much sensitivity when quantifying daily

TN loads. A possible explanation for this is the higher uncertainty in the nature of representing cattle exclusion fencing in SWAT. Several assumptions were used when representing the exclusion fencing such as changing the nutrient introducing to the streams via point sources and the rate adjustments for nutrients introducing streams. It should be noted that the intent of this study was to develop a framework for probabilistic assessment of BMP effectiveness in reducing pollutants in streams and not to develop a method to represent specific BMPs in models.

The results of this study have important implications for decision making when models are used for water quality simulation. While numerous uncertainty analysis frameworks have been developed to explore modeling uncertainties in quantification of streamflow and water quality components, the lack of pragmatic applications of such methods to tackle decision making challenges is a major shortcoming. The framework presented in this study is deemed a pioneer attempt to fill this gap in the literature and add to the pragmatic aspects of the uncertainty analysis in hydrologic and water quality simulations.

Bibliography

Ahmadi, M., Arabi, M., Ascough, J., Fontane, D.G., Engel, B.A., 2014. Toward improved calibration of watershed models: Multisite multiobjective measures of information. Environ. Modell. Software (59), 135-145.

Ajami, N.K., Duan, Q., Sorooshian, S., 2007. An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction. Water Resources Res. 43(1), Art. No. W01403.

Arabi, M. Govindaraju, R.S., Hantush, M.M., Engel, B.A., 2006. Role of watershed subdivision on modeling the effectiveness of best management practices with SWAT. J. American Water Resources Association. 42(2), 513-528.

Arabi, M., Govindaraju, R.S., Hantush, M.M., 2007. A probabilistic approach for analysis of uncertainty in the evaluation of watershed management practices. J. Hydrol. 333, 459-471.

Arabi, M., Frankenberger, J.R., Engel, B.A., Arnold, J.G., 2008. Representation of agricultural conservation practices with SWAT. J. Hydrol. Process. 22(16), 3042-3055.

Arnold, J.G., Moriasi, D.N., Gassman, P.W., Abbaspour, K.C., White, M.J., Srinivasan, R., Santhi, C., Harmel, R.D., van Griensven, A., Van Liew, M.W., Kannan, N., Jha, M.K., 2012. SWAT: model use, calibration, and validation. Trans. ASABE. 55(4), 1491-1508.

Beven, K.J., Binley, A.M., 1992. The future of distributed models: model calibration and uncertainty prediction. Hydrol. Process. 6, 279-298.

Bishop, P.L., Hively, W.D., Stedinger, J.R., Rafferty, M.R., Lojpersberger, J.L, Bloomfield., J.A., 2005. Multivariate Analysis of Paired Watershed Data to Evaluate Agricultural Best Management Practice Effects on Stream Water Phosphorus. J. Environ. Qual. 34:1087-1101.

Byers, H.L., Cabrera , M.L., Matthews , M.K., Franklin , D.H., Andrae , J.G., Radcliffe , D.E., McCann, M.A., Kuykendall , H.A., Hoveland , C.S., Calvert , V.H. 2005. Phosphorus, Sediment, and Escherichia coli Loads in Unfenced Streams of the Georgia Piedmont, USA. J. Environ. Qual. 34:2293-2300.

CARD Staff, 2016. SWAT literature database for peer-reviewed journal articles. Center for Agricultural and Rural Development, Ames, IA. https://www.card.iastate.edu/swat_articles/ (accessed 1 Aug. 2016).

Cheng, Q.B., Reinhardt-Imjela, C., Chen, X., Schulte, A., Ji, X., Li, F., 2016. Improvement and comparison of the rainfall–runoff methods in SWAT at the monsoonal watershed of Baocun, Eastern China. Hydrol. Sci. J. 61(8), 1460-1476.

Clausen, J.C., Jokela, W.E., Potter, F.I., Williams, J.W., 1996. Paired Watershed Comparison of Tillage Effects on Runoff, Sediment, and Pesticide Losses. J. Environ. Qual. 25, 1000-1007.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. Series B. 39, 1-39.

Ficklin, D.L., Zhang, M., 2013. A comparison of the curve number and Green-Ampt models in an agricultural watershed. Trans. ASABE. 56(1), 61-69.

Gassman, P.W., Reyes, M.R., Green, C.H., Arnold, J.G., 2007. The soil and water assessment tool: historical development, application, and future research directions. Trans. ASABE. 50(4), 1211-1250.

Gelman, A.G., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Stat. Sci. 7, 457-472.

Georgakakos, K.P., Seo, D.J., Gupta, H., Schaake, J., Butts, M.B., 2004. Towards the characterization of streamflow uncertainty through multimodel ensembles. J. Hydrol. 298, 222-241.

Green, W.H., Ampt, G.A., 1911. Studies on soil physics, 1. The flow of air and water through soils. J. Agric. Sci. 4, 11-24.

Harmel, R.D., Smith, P.K., 2007. Consideration of measurement uncertainty in the evaluation of goodness-of-fit in hydrologic and water quality modeling. J. Hydrol. 337(3-4), 326-336.

Harmel, R.D., Smith, P.K., Migliaccio, K.L., Chaubey, I., Douglas-Mankin K., Benham, B., Shukla, S., Muñoz-Carpena, R., and Robson, B.J., 2014. Evaluating, interpreting, and communicating performance of hydrologic/water quality models considering intended use: A review and recommendations. Environ. Modeling Software. 57, 40-51.

Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T., 1999. Bayesian model averaging: A tutorial. Stat. Sci. 14(4), 382-417.

Jang, S.S., Ahn, S.R., Kim, A.J., 2017. Evaluation of executable best management practices in Haean highland agricultural catchment of South Korea using SWAT. J. Agri. Water Manag. 180(B), 224-234.

Jingyuan, J., Shiyu, L., Hu, J., Huang, J., 2014. A modeling approach to evaluating the impacts of policy-induced land management practices on non-point source pollution: A case study of the Liuxi River watershed, China. Agri. Water Manag. 131, 1-16.

Kannan, N., White, S.M., Worrall, F., Whelan, M.J., 2007. Sensitivity analysis and identification of the best evapotranspiration and runoff options for hydrological modelling in SWAT-2000. J. Hydrol. 332, 456-466.

Lin, Z., Radcliffe, D.E., Risse, L.M., Romeis, J.J., Jackson, C.R.. 2009. Modeling Phosphorus in the Lake Allatoona Watershed Using SWAT: II. Effect of Land Use Change. J. Environ. Qual. 38:121-129.

Line, D.E., Osmond, D.L., Childres, W. 2016. Effectiveness of Livestock Exclusion in a Pasture of Central North Carolina. J. Environ. Qual. 45:1926-1932.

Liu, R., Zhang, P., Wang, X., Chen, Y., Shen, Z., 2013. Assessment of effects of best management practices on agricultural non-point source pollution in Xiangxi River watershed. Agri. Water Manag. 117, 9-18.

Madadgar, S., Moradkhani, H., 2014. Improved Bayesian multimodeling: Integration of copulas and Bayesian model averaging. Water Resources Res. 50, 9586-9603.

Moradkhani, H., Hsu, K.L., Gupta, H.V., Sorooshian, S., 2005. Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter, Water Resources Res. 41, W05012.

Motallebi, M., Hoag, D.L., Tasdighi, A., Arabi, M., Osmond, D., 2017. An economic inquisition of water quality trading programs, with case study of Jordan Lake, NC. J. Environ. Manage. 193, 483-490.

NOAA, 2016. National climatic data center (NCDC). National Oceanic and Atmospheric Administration, Asheville, NC. http://www.ncdc.noaa.gov/ (accessed 1 Feb. 2016).

Park, D. Loftis, J.C. Roesner, L.A., 2011. Modeling performance of storm water best management practices with uncertainty analysis. J. Hydrologic Engineering. 16 (4), 332–344.

Park, D., L.A. Roesner, 2012. Evaluation of pollutant loads from stormwater BMPs to receiving water using load frequency curves with uncertainty analysis. Water Res. 46(20), 6881-6890.

Santhi, C., Srinivasan, R., Arnold, J.G., Williams, J.R., 2003. A modeling approach to evaluate the impact of water quality management plans implemented in the Big Cypress Creek Watershed. In: Second conference on watershed management to meet emerging TMDL environmental regulations, Albuquerque, NM, pp. 384-394.

Schoups, G., Vrugt, J.A., 2010. A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. Water Resources Res. 46(10), W10531.

Sorooshian, S., Dracup, J.A., 1980. Stochastic parameter estimation procedures for hydrologie rainfall-runoff models: Correlated and heteroscedastic error cases. Water Resources Res. 16(2), 430-442.

Tasdighi, A., Arabi, M., Osmond, D.L., 2017. The relationship between land use and vulnerability to nitrogen and phosphorus pollution in an urban watershed. J. Environ. Qual. 46, 113-122.

Tasdighi, A., Arabi, M., Harmel, D., 2017. A probabilistic appraisal of rainfall-runoff modeling approaches within SWAT in mixed land use watersheds. Under review in J. Hydrol.

Taylor, S.D., He, Y., Hiscock, K.M., 2016. Modelling the impacts of agricultural management practices on river water quality in Eastern England. J. Environ. Manag. 180(15), 147-163.

Ullrich, A., Volk, M., 2009. Application of the Soil and Water Assessment Tool (SWAT) to predict the impact of alternative management practices on water quality and quantity. Agri. Water Manag. 96, 1207-1217.

USDA-NRCS, 2004. Chapter 10: Estimation of direct runoff from storm rainfall, in: NRCS National Engineering Handbook, Part 630: Hydrology. U.S. Department of Agriculture, Natural Resources

Conservation Services, Washington. D.C. http://www.wcc.nrcs.usda.gov/ftpref/wntsc/H&H/NEHhydrology/ch10.pdf/ (accessed 1 Dec. 2015).

USDA-ARS, 2014. ArcSWAT 2012.10_1.13. U.S. Department of Agriculture, Agricultural Research Service, Washington, D.C. http://swat.tamu.edu/ (accessed 1 Dec. 2015).

USDA-NRCS, 2016. USDA-NRCS Geospatial data gateway. Available at: https://datagateway.nrcs.usda.gov/ (accessed 1 Dec 2015).

USGS-NHD, 2016. National hydrography dataset. Available at: https://nhd.usgs.gov/ (accessed 1 Dec. 2015).

USGS-TNM, 2016. TNM Map viewer, available at: https://viewer.nationalmap.gov/viewer/ (accessed 1 Dec 2015).

Vogel, R.M., Fennessey, N.M., 1994. Flow-Duration Curves. I: New Interpretation and Confidence Intervals. J. Water Resources Plan. Manag. 120(4), 485-504.

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

Vrugt, J.A., ter Braak, C.J.F., Diks, C.G.H., Robinson, B.A., Hyman, J.M., Higdon, D., 2009. Accelerating Markov Chain Monte Carlo Simulation by Differential Evolution with Self-Adaptive Randomized Subspace Sampling. Int. J. Nonlinear Sci. 10, 273-290.

Vrugt, J.A., 2016. Markov chain Monte Carlo simulation using the DREAM software package: theory, concepts, and MATLAB implementation. Environ. Modell. Software. 75, 273-316.

Winchell, M.R., Srinivasan, R., Di Luzio, M., Arnold, J.G., 2007. ArcSWAT interface for SWAT2005: user's guide. USDA-ARS Blackland Research Center.

Yen, H., Wang, X., Fontane, D.G., Harmel, R.D., Arabi, M., 2014. A framework for propagation of uncertainty contributed by parameterization, input data, model structure, and calibration/validation data in watershed modeling. Environ. Modell. Software. 54, 211-221.

Chapter 5

SUMMARY AND CONCLUSIONS

Nonpoint source pollution is the primary cause of impaired water bodies in the United States. Control of this type of pollution requires deep understanding of the processes that characterize them. However, characterizing the nonpoint source pollution is extremely difficult due to natural variability due to land use, climatic conditions, and uncertainties from various sources during the application of models. This study was an attempt to confront the variability and uncertainty of nonpoint sources and develop a probabilistic framework for assessing the water quality benefits of conservation practices in reducing nonpoint sources.

5.1. The relationship between land use and vulnerability to nitrogen and phosphorus pollution in an urban watershed

Multi Linear Regression models were used to explore the relationship between land use and water quality under various climatic conditions. Such analyses are specifically important in assessing the variability of nonpoint source pollution. The results of this study revealed that, strong ($R^2 > 0.7$) and significant ($p < 0.01$) positive correlations exist between percent urban land use and annual TN and TP concentrations/loads in Jordan Lake watershed. Agricultural land use was determined to be negatively correlated with TN and TP in most years. The negative correlation of agricultural land use with TN and TP is explained by the relatively low agricultural activities compared to urban development and exchange of land use between these two sectors in Jordan Lake watershed. Further analysis showed significant correlation between TP and TSS loads which tend to become stronger as percent agriculture land increased. This finding indicates that in Jordan Lake watershed, TP conservation practices that are based on sediment reduction will probably be more effective in areas with higher percentage of agricultural land use.

Inter-annual precipitation variations had an important effect on land use-water quality analysis. Using TN and TP loads, significant ($p < 0.01$) differences were observed between regression models based on climatic conditions (determined by the amount of annual precipitation). Using concentrations resulted in nonsignificant differences between average and wet years. This outcome has important implications for selecting between loads and concentrations as well as a time frame for land use-water quality analyses.

Percent urban land use and climate variability had a profound effect on the vulnerability of stream water to exceeding the TN and TP targets. While urban land use was substantially influencing the vulnerability to exceeding the targets, climate conditions played an important role in determining the extent of the impact. During dry years, 5% increase in urban land use percentage results in substantial vulnerability to TN and TP while for average and wet years, 10 to 40 percent increases in percent urban land use is required to reach the same level of vulnerability.

5.2. A probabilistic appraisal of rainfall-runoff modeling approaches within SWAT in mixed land use watersheds

SWAT has been used extensively in the literature for hydrologic and water quality simulations. While the model has the capability to employ either CN or G&A method for runoff estimation, almost all studies have employed the CN method. This may be partly due to lack of a rigorous comparison study to justify merits of using one method over the other along with higher simplicity of the CN method. This study attempted to address this shortcoming. Regarding the extreme popularity of the SWAT model, the findings of this study can shed light on selecting the rainfall-runoff method within SWAT that can potentially lead to more realistic streamflow simulations in mixed-land use watersheds.

In this study, a Bayesian total uncertainty assessment framework was implemented to compare the performance of the three runoff generation mechanisms within SWAT under different upstream land use conditions. Using the uncertainty assessment framework at the watershed outlet, models' performances

were assessed at several stream locations inside the watershed. At the watershed outlet and subwatersheds with dominant agricultural or forest land use, CN models performed slightly better. However, at the two subwatersheds with highly developed land use, models with G&A method had a much better performance in simulating the streamflow compared to the CN models.

Overall, the streamflow prediction intervals from models with G&A method covered more observations. However, they were slightly wider indicating higher uncertainty for streamflow prediction. The CN models were unable to capture the high flow events specifically in developed subwatersheds. Posterior distribution of mean for Gaussian distributions from which precipitation multipliers were randomly drawn were closer to normal using the hourly precipitation data and G&A method while using daily precipitation with CN methods resulted in substantial negative skewness. The deficiency of models with CN methods in simulating the peak streamflows caused higher values of precipitation multipliers to be sampled to augment the runoff volume.

The CN method simulated higher volume of water yield specifically at the urban-dominated subwatersheds while G&A method simulated higher ET values. The higher volume of water yield by CN in the highly urbanized subwatersheds can be explained by CN attempt to simulate the high flow events which results in overall overestimation of water yield. The G&A model resulted in lower surface runoff at all locations compared to the CN models; however, it simulated higher infiltration and subsurface flows.

The results of this study have important implications for determining which rainfall-runoff method performs better in simulating the hydrologic regime. The evaluation is specifically relevant for applying a distributed hydrologic model such as SWAT in a mixed-land use watershed where model training will be performed only at the watershed outlet but the model is to be used for simulating hydrologic responses at different locations inside the watershed. In summary, the results suggest that while trained at watershed outlet, the SWAT model with G&A method can potentially perform better in areas inside the watershed

with higher percentage of developed land. The SWAT models with CN methods proved to have similar or slightly better performance in areas with agriculture or forest dominant land use.

5.3. A Bayesian total uncertainty estimation framework for assessment of management practices using watershed models

A total uncertainty estimation framework was developed for assessing the water quality benefits of management practices using watershed models. The two-stage framework first characterizes various sources of modeling uncertainties during the pre-BMP period. The second stage of the framework uses the inferences on modeling uncertainties from the first stage and quantifies the effectiveness of BMPs using a probabilistic approach.

In general, the modeling uncertainties were large resulting in wide bands of uncertainty around both TN loads and load reductions. The framework however was successful in capturing the effects of different sources of modeling uncertainties on simulations and propagating them to the BMP effectiveness assessment stage. Between the three model structures (CNI, CNII, and G&A), CNII showed the best performance in terms of various performance measures including error statistics and BMA weights. This was attributed to the intensive agricultural land use in the watershed. The G&A method had an unsatisfactory performance in simulating the TN loads. As shown in previous chapter, the G&A method excelled in simulating streamflow and hydrologic regime in highly developed subwatersheds. Application of the BMA slightly enhanced the quality of simulating TN load reductions under uncertainty as it resulted in higher coverage 51% and 35% during the pre and post-BMP period respectively. The distribution of the correction factors for measurement uncertainty indicated higher uncertainty for high flow events correctly capturing the heteroscedasticity of error residuals for daily TN load simulations.

Between the two BMPs, nutrient management had the highest impact on the TN load reductions. The parameters pertaining to cattle exclusion fencing did not show much sensitivity when quantifying daily TN loads. A possible explanation for this is the higher uncertainty in the nature of representing cattle

exclusion fencing in SWAT. Several assumptions were used when representing the exclusion fencing such as changing the nutrient introducing to the streams via point sources and the rate adjustments for nutrients introducing streams. It should be noted that the intent of this study was to develop a framework for probabilistic assessment of BMP effectiveness in reducing pollutants in streams and not to develop a method to represent specific BMPs in models.

The results of this study have important implications for decision making when models are used for water quality simulation. While numerous uncertainty analysis frameworks have been developed to explore modeling uncertainties in quantification of streamflow and water quality components, the lack of pragmatic applications of such methods to tackle decision making challenges is a major shortcoming. The framework presented in this study is deemed a pioneer attempt to fill this gap in the literature and add to the pragmatic aspects of the uncertainty analysis in hydrologic and water quality simulations.

5.4. Future work

This study investigated the variability and uncertainty of nonpoint source pollution. A total uncertainty estimation framework for assessing the effectiveness of conservation practices in reducing nonpoint source pollution was developed. The inferences from this study on variability and uncertainty of nonpoint source pollution and the framework developed will be used for quantifying trading ratios in water quality trading programs. Specifically the framework will be used for quantifying trading ratios for a nutrient trading program in Jordan Lake Watershed in North Carolina.

Nutrient trading programs are market-based programs that involve the exchange of nutrient pollution allowances between sources. Trades can be point to point, point to nonpoint, or nonpoint to nonpoint sources. Conceptually, nutrient trading is appealing as a cost-effective and flexible way to achieve and maintain water quality goals. An important motivation for these programs is the potential to achieve the required nutrient reductions at a lower cost. In addition, trading programs offer the potential to increase

conservation on nonpoint sources. Applications of water quality trading have been promoted by various agencies and have grown in popularity in recent years due to economic and pollution reduction incentives.

However, implementation of nutrient trading programs may be plagued by difficulties in determining the amount of credits that can be subject to trade. Particularly, the credit transfer may be more challenging in cases where one side of the trade is nonpoint for two reasons: (1) higher natural "variability" in pollutant loads from nonpoint sources compared to point source effluents, and (2) the "uncertainty" in quantifying the loads from nonpoint sources.

First, the variability of pollutant loads from point source effluents is often less than the variability of nonpoint source pollutant loads. The contribution from nonpoint sources depends on climate, land use, and other environmental conditions that are highly variable by nature. Understanding the variability of ambient water quality constituents under these various conditions is an essential step in developing water quality trading programs. In this regard, exploring the relation between land use and water quality due to its significant influence on the nutrient loads is of crucial importance. One important aim of understanding the land use-water quality relationship would be to determine the relative significance of contribution from different land use types in polluting waterbodies. Such analyses will provide valuable information for developing hypotheses in designing nutrient trading programs.

Second, it is infeasible to measure pollutant loads from nonpoint sources within a watershed using monitoring campaigns. Hence, simulation models are commonly used to estimate nonpoint source pollutant loads and assess benefits of conservation practices. Models are mere representation of reality. Consequently, they bear an inherent stochasticity in practice. The uncertain nature of models affects their predictive performance which has to be taken into account when they are used for decision making purposes.

Currently, "trading ratios" are used within trading programs to account for variability of nonpoint source loads. Using trading ratios, credit buyers are required to purchase more reductions than they need to meet their regulatory obligation. Uncertainties involved in modeling nonpoint sources and the

effectiveness of conservation practices also have to be reflected in trading ratios although currently there are no cases of any rigorous scientific approach to account for these types of uncertainties in trading ratios. Trading ratios serve as a margin of safety which aims for reaching nutrient reduction goals set in trading programs. Underestimating trading ratios can result in failure to achieve the reduction goals while using overly conservative values could unnecessarily reduce trading volumes and expected efficiency gains. Application of the framework developed in this study can reduce the subjectivity in adopting trading ratios and help in making more informed decisions in water quality trading.

Appendix

Table S2.1. SWAT 2012 model parameters after auto-calibration along with their corresponding units and ranges

Parameter	Input file	Units	Calibrated value	Lower limit	Upper limit
ALPHA_BF	.gw	days	0.9507	0	1
CH_SI	.sub	%	-0.02959	-0.05	0.05
CN_F	.mgt	%	-0.0419	-0.1	0.1
SOL_ALB	.sol	%	0.795	-0.5	1
SURLAG	.bsn	day	1.523	1	12
OV_N	.hru	-	0.3212	0.01	0.6
CH_NI	.sub	-	0.1836	0.008	0.3
CANMX	.hru	mm	3.202	0	10
CH_NII	.rte	-	0.03806	0.01	0.3
SOL_AWC	.sol	%	-0.09386	-0.1	2
CH_SII	.rte	%	0.02295	-0.05	0.05
CH_KII	.rte	mm/hr	81	-0.01	500
ESCO	.hru	-	0.9573	0.01	1
EVRCH	.bsn	-	0.4113	0	1
GWQMN	.gw	mm	414.9	0	5000
CH_COV	.rte	-	0.2242	-0.001	1
SLSUBBSN	.hru	m	18.92	10	150
SLOPE	.hru	%	-0.08796	-0.1	0.1
CH_KI	.sub	mm/hr	1.295	0	300
GW_REVAP	.gw	-	0.11	0.02	0.2
DEP_IMP	.hru	mm	2470	1500	2500
USLE_K	.sol	%	0.4448	-0.5	1
GW_DELAY	.gw	day	57.82	0	60
EPCO	.bsn	-	0.6926	0.01	1
RES_K	.res	mm/h	0.6491	0	2
EVRSV	.res	-	1.486	0	2
RCHRG_DP	.gw	-	0.888	0	1
SOL_K	.sol	%	2.844	1	3

Table S2.2. Summary of regression model coefficients for percent urban-animal feeding operations (AFOs) capacity models.

	TN (Concentration)		TN (Load)		TP (Concentration)		TP (Load)	
	% Urban	AFO cap	% Urban	AFO cap	% Urban	AFO cap	% Urban	AFO cap
1992[a]	1.02	0.03*	0.95	-0.02*	1.23	0.11	1.16	0.06
1993[w]	1.00	0.03*	0.85	-0.02*	1.12	0.10	0.96	0.05
1994[a]	1.01	0.01*	0.98	0.03*	1.33	0.10	1.30	0.11
1995[w]	0.83	-0.05*	0.81	-0.02*	0.86	-0.01*	0.84	0.01*
1996[w]	0.82	-0.03*	0.88	0.00*	0.79	-0.01*	0.85	0.03*
1997[a]	1.24	0.02*	1.40	0.02*	1.27	0.01*	1.41	0.01*
1998[w]	1.28	0.06*	1.43	0.00*	1.37	0.07*	1.53	0.02*
1999[w]	1.36	-0.01*	1.59	-0.02*	1.45	0.08*	1.68	0.06*
2000[a]	1.20	0.02*	1.31	-0.02*	1.34	0.06*	1.44	0.02*
2001[d]	1.43	0.06	1.69	0.06*	1.53	0.15	1.79	0.15
2002[d]	1.42	0.08	1.64	0.07*	1.44	0.17	1.67	0.17
2003[w]	1.06	-0.02*	1.06	-0.02*	1.32	0.08	1.32	0.07
2004[a]	1.09	0.02*	1.28	0.03*	1.00	0.07	1.19	0.07
2005[d]	1.33	0.08	1.55	0.10*	1.34	0.09	1.58	0.11
2006[a]	1.33	0.06*	1.50	0.03*	1.62	0.14	1.79	0.11
2007[d]	1.61	0.11	1.78	0.10*	1.83	0.16	1.99	0.15
2008[w]	1.46	0.09	1.74	0.08*	1.52	0.07*	1.80	0.06*
2009[d]	1.49	0.11	1.64	0.09*	1.51	0.14	1.66	0.13
2010[w]	1.35	0.06*	1.32	0.05*	0.70	0.08	0.66	0.07*
2011[d]	1.53	0.09	1.86	0.07*	1.42	0.14	1.75	0.12
2012[a]	1.25	0.09	1.61	0.09*	1.17	0.09	1.53	0.09*

The * sign indicates nonsignificanct coefficient ($p > 0.01$); *a*, *w* and *d* correspond to average, wet and dry years respectively; 0.00 represents a value less than 0.01

Table S2.3. Summary of regression model coefficients for percent agriculture-wastewater treatment plants (WWTPs) capacity models.

	TN (Concentration)		TN (Load)		TP (Concentration)		TP (Load)	
	% Ag	WWTP cap	% Ag	WWTP cap	% Ag	WWTP cap	% Ag	WWTP cap
1992[a]	-0.18	0.19	-0.30	0.18	0.16*	0.24	0.04*	0.23
1993[w]	-0.32	0.17	-0.42	0.15	0.06*	0.21	-0.04*	0.19
1994[a]	-0.24	0.19	-0.28	0.18	0.10*	0.26	0.07*	0.25
1995[w]	-0.63	0.10	-0.72	0.07	-0.22	0.18	-0.31	0.15
1996[w]	-0.51	0.11	-0.62	0.09	-0.29	0.14	-0.40	0.12
1997[a]	-0.46	0.13	-0.65	0.13	-0.14*	0.21	-0.33	0.20
1998[w]	-0.03*	0.20	-0.40	0.19	0.25	0.26	-0.10*	0.25
1999[w]	-0.38	0.17	-0.67	0.18	0.11*	0.24	-0.18*	0.25
2000[a]	-0.35	0.15	-0.54	0.16	0.09*	0.23	-0.11*	0.23
2001[d]	-0.23	0.20	-0.34	0.25	0.28	0.28	0.17	0.32
2002[d]	-0.22	0.18	-0.36	0.22	0.36	0.24	0.22	0.28
2003[w]	-0.63	0.09	-0.51	0.12	-0.11*	0.18	0.02*	0.11
2004[a]	-0.51	0.10	-0.61	0.13	-0.23	0.11	-0.33	0.14
2005[d]	-0.16	0.18	-0.23	0.22	-0.20	0.16	-0.26	0.20
2006[a]	0.02*	0.23	-0.24	0.24	0.14*	0.26	-0.12*	0.27
2007[d]	0.08*	0.23	-0.02*	0.25	-0.11	0.21	-0.20	0.23
2008[w]	-0.03*	0.19	-0.17	0.24	0.01*	0.22	-0.12*	0.27
2009[d]	-0.06*	0.19	-0.12*	0.23	0.00*	0.19	-0.05*	0.22
2010[w]	-0.05*	0.19	0.00*	0.21	0.06*	0.09	0.11*	0.12
2011[d]	-0.06*	0.19	-0.21	0.24	0.15	0.21	0.00*	0.26
2012[a]	0.07*	0.17	-0.03*	0.23	0.04*	0.17	-0.05*	0.23

The * sign indicates nonsignificanct coefficient ($p > 0.01$); *a*, *w* and *d* correspond to average, wet and dry years respectively; 0.00 represents a value less than 0.01

Table S2.4. Summary of linear regression models developed between total phosphorus (TP) and total suspended solids (TSS). % Urban and % Agriculture represent percent urban land use and percent agricultural land use. R^2 is the coefficient of determination. P_F is the p-value for significance of the model.

Station	R^2	P_F	% Urban	% Agriculture
8	0.42	< 0.01	85	3
13	0.86	< 0.01	26	27
19	0.9	< 0.01	20	28
20	0.54	< 0.01	41	5
21	0.62	< 0.01	59	2