DISSERTATION

CONFRONTING INPUT, PARAMETER, STRUCTURAL, AND MEASUREMENT UNCERTAINTY IN MULTI-SITE MULTIPLE-RESPONSE WATERSHED MODELING USING BAYESIAN INFERENCES

Submitted by

Haw Yen

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

Fall 2012

Doctoral Committee:

Advisor: Mazdak Arabi

Darrell G. Fontane
Dana L. Hoag
Jim C. Loftis

ABSTRACT

CONFRONTING INPUT, PARAMETER, STRUCTURAL, AND MEASUREMENT UNCERTAINTY IN MULTISITE MULTIPLE-RESPONSE WATERSHED MODELING USING BAYESIAN INFERENCES

Simulation modeling is arguably one of the most powerful scientific tools available to address questions, assess alternatives, and support decision making for environmental management. Watershed models are used to describe and understand hydrologic and water quality responses of land and water systems under prevailing and projected conditions. Since the promulgation of the Clean Water Act of 1972 in the United States, models are increasingly used to evaluate potential impacts of mitigation strategies and support policy instruments for pollution control such as the Total Maximum Daily Load (TMDL) program. Generation, fate, and transport of water and contaminants within watershed systems comprise a highly complex network of interactions. It is difficult, if not impossible, to capture all important processes within a modeling framework. Although critical natural processes and management actions can be resolved at varying spatial and temporal scales, simulation models will always remain an approximation of the real system. As a result, the use of models with limited knowledge of the system and model structure is fraught with uncertainty. Wresting environmental decisions from model applications must consider factors that could conspire against credible model outcomes.

The main goal of this study is to develop a novel Bayesian-based computational framework for characterization and incorporation of uncertainties from forcing inputs, model parameters, model structures, and measured responses in the parameter estimation process for multisite multiple-response watershed modeling. Specifically, the following objectives are defined: (i) to

evaluate the effectiveness and efficiency of different computational strategies in sampling the model parameter space; (ii) to examine the role of measured responses at various locations in the stream network as well as intra-watershed processes in enhancing the model performance credibility; (iii) to facilitate combining predictions from competing model structures; and (iv) to develop a statistically rigorous procedure for incorporation of errors from input, parameter, structural and measurement sources in the parameter estimation process. The proposed framework was applied for simulating streamflow and total nitrogen at multiple locations within a 248 square kilometer watershed in the Midwestern United States using the Soil and Water Assessment Tool (SWAT). Results underlined the importance of simultaneous treatment of all sources of uncertainty for parameter estimation. In particular, it became evident that incorporation of input uncertainties was critical for determination of model structure for runoff generation and also representation of intra-watershed processes such as denitrification rate and dominant pathways for transport of nitrate within the system.

The computational framework developed in this study can be implemented to establish credibility for modeling watershed processes. More importantly, the framework can reveal how collection of data from different responses at different locations within a watershed system of interest would enhance the predictive capability of watershed models by reducing input, parametric, structural, and measurement uncertainties.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my Lord and Savior Jesus Christ. Thank you for leading me out of the valley of shadow. Thank you for staying with me during tough times and minimizing all potential sufferings. Thank you for taking care of my parents while I was on the other side of the ocean. I would be a man of nothing without you. Thank you! I also thank my parents and my wife, Doctor Chan-Chen Yen, Hsiao-Chin Pan and Yu-Chieh Chang for their unconditional love and sacrifices throughout the journey.

For the academic work, I appreciate the great help from my advisor Dr. Mazdak Arabi and all committee members, Dr. Darrell G. Fontane, Dr. Dana L. Hoag and Dr. Jim C. Loftis. This dissertation would not be completed without their contributions.

Five years during my PhD program, I met a lot of respectable friends and professors from many different countries. Most people would give thanks to them as a whole but I insist to have their names printed right here to show my respect: Dr. Mahdi Ahmadi, Dr. Panagiotis Oikonomou, Dr. Dae-Ryong Park, Dr. Jong-Hyun Lee, Dr. Jung-Kyu Ahn. Dr. Ayman Alzraiee, Dr. Ryan T. Bailey, Dr. Masoud Negahban Azar, Dr. Stephen L. Ott, Dr. Chih-Ted Yang, Dr. Larry A. Roesner, Mr. and Mrs. Hoffman, Mr. and Mrs. Peterson, Rosemary M. Records, Zhao-Hua Huang and Adam M. Beal. In addition, old friends back home in Taiwan also gave me tremendous support when I needed the most: Dr. Kai-Cheng Hu, Dr. Tzu-Chiang Lee, Dr. Chia-Wen Wu, Dr. Wen-Hsiao Tseng, Dr. Yi-Jung Tsai, Tsung-Han Yang, Pai-Chien Wei, Chih-Yu Chang, Cheng-Yang Lee, Yi-Chun Yeh, Chiu-Wen Chen and Bei-Hua Wen. Finally, I would like to give special thanks to Dr. John W. Labadie, Dr. Frederick N.F. Chou and Dr. Chyan-Deng Jan as my lifetime mentors. May God bless you all!

TABLE OF CONTENTS

Chapter 1 Introduction.....1

1.1. Overview 1

1.2. First Part of the Framework (Chapter II, III): Development of a Framework to Evaluate the Effectiveness of Various Sampling Strategies 6

1.2.1. Statistically Valid Likelihood Function 6

1.2.2. Performance Validity by Applying Behavior Definition 7

1.2.3. Identifiability of Individual Model Parameters and Entire Parameter Hypercube....9

1.3. Second Part of the Framework (Chapter IV, V): Development of a Framework to Incorporate Input, Parameter, Structural and Measurement Uncertainty 10

1.4. Goals and Objectives.....12

1.4.1. Appraisal of Sampling Strategies for Parameter Estimation and Uncertainty Analysis by Using a Systematic Evaluation Framework in Watershed Modeling..13

1.4.2. Evaluation of Improved Parameter Estimation and Uncertainty Analysis of Watershed Models: The Role of General Information of Watershed Processes.....15

1.4.3. Evaluation of Structural Uncertainty on Hydrologic and Water Quality Predictions Using SWAT Model.....16

1.4.4. A Framework for Propagation of Input, Parameter, Structural and Measurement Uncertainty in Watershed Modeling.....16

References 18

Chapter 2 Appraisal of Sampling Techniques for Parameter Estimation and Uncertainty Analysis in Watershed Modeling 21

2.1. Overview 21

2.2. Introduction 22

2.3. Methodology 26

2.3.1. Objective Functions 26

2.3.2. Definition of Behavior Parameter Sets 29

2.3.3. Sampling Techniques 30

2.3.4. Identifiability of parameter estimation 36

2.4. Case Studies 37

2.4.1. SWAT2005 Eagle Creek Watershed Calibration 37

2.4.2. Results of All Approaches 39

2.4.3. Behavior Rate of Different Methods and Identifiability for Model Parameters 42

2.4.4. Parameter Uncertainty 46

2.5. Discussion 48

2.6. Conclusions 53

References 56

Chapter 3 Evaluation of Improved Parameter Estimation and Uncertainty Analysis of Watershed Models: The Role of General Information of Watershed Processes 59

3.1. Overview 59

3.2. Introduction 60

3.3. Methods and Materials 62

3.3.1. Site Location and Characteristics 62

3.3.2. SWAT Model 63

3.3.3. Constraints of Intra-watershed Responses 64

3.3.4. Parameter Estimation Technique 65

3.3.5. Evaluation Statistics 66

3.4. Case Studies 69

3.4.1. Eagle Creek Watershed Calibration 69

3.4.2. Results of All Approaches 69

3.5. Discussions and Conclusions 79

References 82

Chapter 4 Evaluation of Structural Uncertainty on Hydrologic and Water Quality Predictions Using SWAT Model 86

4.1. Overview 86

4.2. Introduction 87

4.3. Methods and Materials 89
4.3.1. Site Location and Characteristics 89
4.3.2. SWAT Model 90
4.3.3. SCS Curve Number Procedure 91
4.3.4. Parameter Estimation Procedure 93
4.3.5. Bayesian Model Averaging (BMA) 95
4.3.6. Brier Score 97
4.4. ECW Case Studies 97
4.4.1. Basic Settings of Case Scenarios 98
4.4.2. Results of All Approaches 101
4.5. Discussion and Conclusions 109
References 112
Chapter 5 A Framework for Propagation of Input, Parameter, Structural and Measurement Uncertainty in Watershed Modeling 116
5.1. Overview 116
5.2. Introduction 117
5.3. Methods and Materials 119
5.3.1. Site Location and Characteristics 119
5.3.2. SWAT Model 120
5.3.3. Sources of uncertainty 121

5.3.4. Evaluation Statistics 126

5.3.5. Parameter Estimation Technique.....129

5.4. Case Study.....130

5.4.1. Eagle Creek Watershed Calibration 132

5.4.2. Results of All Scenarios 133

5.5. Discussion and Conclusion 144

References 147

Chapter 6 Conclusion and Discussion.....151

6.1. Major findings.....151

6.2. Limitations and shortcomings 152

6.3. Broader impact of the proposed framework.....154

6.4. Future directions.....156

References 159

LIST OF TABLES

Table 2.1 General Performance Ratings 29

Table 2.2 Number and percentage of behavior parameter sets in 5000 model evaluations with results from TSDE calculations (Case 1). The highlighted cell indicates that 90% of behavior parameter sets (cumulative number of points = 90%) are located in 7.55% of original sampling domain (Θ = 7.55%) for method DDS in the performance of "Very Good" where the TSDE index (11.93) is calculated by equation (1). TSDE will not be calculated if there is insufficient data. 44

Table 2.3 Number and percentage of behavior parameter sets in 5000 model evaluations with results from TSDE calculations (Case 5) 46

Table 2.4 Percentage of model evaluations reaches the same performance against SCE-UA 50

Table 2.5 Inclusion rates of observed streamflow data included in 95% confidence interval 52

Table 2.6 Inclusion rates of observed NOX data included in 95% confidence interval..... 53

Table 3.1 General Performance Ratings 68

Table 3.2 Behavior Rate after the GPR is Applied..... 70

Table 3.3 Rate of Penalty in Case Studies 72

Table 3.4 Statistical Performance of All Case Studies (for Calibration Period 1997-2000) with Best Objective Function Value Disregarding the Satisfactory of Both Constraints 72

Table 3.5 Statistical Performance of All Case Studies (for Calibration Period 1997-2000) with Best Objective Function Value Considering the Satisfactory of Both Constraints 73
Table 3.6 Statistical Performance of All Case Studies (for Evaluation Period 2001-2003) with Best Objective Function Value Considering the Satisfactory of Both Constraints 73
Table 3.7 Inclusion Rate of Observed Streamflow/NOX Data Included in 95% Confidence Interval and the Corresponding Average Spread..... 73
Table 4.1 General Performance Ratings 95
Table 4.2 BMA Model Weights for Case 3 (Same Set of Weights for Entire Period)..... 100
Table 4.3 BMA Model Weights for Case 4 (Dry Season)..... 100
Table 4.4 BMA Model Weights for Case 4 (Wet Season) 100
Table 4.5 BMA Model Weights for Case 5 (Wet-Warm condition) 101
Table 4.6 BMA Model Weights for Case 5 (Dry-Warm condition)..... 101
Table 4.7 Behavior Rate of Parameter Sets of All Case Studies for Calibration and Evaluation Periods 103
Table 4.8 Results of Statistics for the Calibration and Evaluation Period in Scenario 1 104
Table 4.9 Results of Statistics for the Calibration and Evaluation Period in Scenario 2 104
Table 4.10 Results of Statistics for the Calibration and Evaluation Period in Scenario 3 .. 104
Table 4.11 Results of Statistics for the Calibration and Evaluation Period in Scenario 4 .. 104
Table 4.12 Results of Statistics for the Calibration and Evaluation Period in Scenario 5 .. 104

Table 4.13 Inclusion Rate of Observed Streamflow/NOX Data Included in 95% Confidence Interval and the Corresponding Spread for the Calibration Period (1997-2000) 107
Table 4.14 Inclusion Rate of Observed Streamflow/NOX Data Included in 95% Confidence Interval and the Corresponding Spread for the Evaluation Period (2001-2003) 107
Table 5.1 General Performance Ratings 128
Table 5.2 Implemented Scenarios and Source(s) of Uncertainty Involved in the Case Study 132
Table 5.3 Behavior Rate of Parameter Sets of All Case Studies for Calibration and Evaluation Periods 135
Table 5.4 BMA Weights for BMAI 135
Table 5.5 BMA Weights for BMAII 135
Table 5.6 Results of Statistics for the Calibration Period* 137
Table 5.7 Results of Statistics for the Evaluation Period* 137
Table 5.8 Results of Statistics for the Calibration Period** 138
Table 5.9 Results of Statistics for the Evaluation Period ** 138
Table 5.10 Results of Best Objective Function Values and the Corresponding Watershed Behavior Outputs for the Calibration Period* 139
Table 5.11 Results of Best Objective Function Values and the Corresponding Watershed Behavior Outputs for the Calibration Period** 139

Table 5.12 Inclusion Rate of Observed Streamflow/NOX Data Included within 95% Confidence Interval and the Corresponding Spread for the Calibration Period (1997-2000) 140

Table 5.13 Results that Satisfy Both Actual Watershed Behavior Constraints (%) 142

Table 5.14 Correction Factors of All Gauge Stations 144

LIST OF FIGURES

Figure 1.1 Demonstration of behavior parameter sets and the correspond subdomain spaces 8

Figure 1.2 Sources of uncertainty in watershed modeling 12

Figure 1.3 Framework of the evaluation of sampling strategies and predictive uncertainty. 14

Figure 2.1 Demonstration of behavior parameter sets and the correspond subdomain spaces 30

Figure 2.2 Case study area: Eagle Creek Watershed, Indiana 38

Figure 2.3 Results of ECW calibration: (a) Case 1 - Daily streamflow calibration by OF1 (b) Case 3 - Monthly streamflow and monthly total nitrate calibration by OF2 (c) Case 4 - Monthly streamflow and monthly total nitrate calibration (d) Case 5 - Daily streamflow and monthly total nitrate calibration. 41

Figure 2.4 Percentage of behavior parameter sets out of 5000 model evaluations: (a) Case 1 - Daily streamflow calibration by OF1 (b) Case 5 - Daily streamflow and monthly total nitrate calibration. 45

Figure 2.5 CDFs of sensitive parameters: (a) Case 1 - Daily streamflow calibration by OF1 (b) Case 5 - Daily streamflow and monthly total nitrate calibration. 48

Figure 3.1 Case study area: Eagle Creek Watershed, Indiana 63

Figure 3.2 Overall performance of objective function versus model evaluations of fours scenarios 70

Figure 3.3 Cumulative distribution functions of constraints applied in four scenarios: (A) CDFs of denitrification, (B) CDFs of SSQ ratio. (Ratio of nitrate-N losses is

calculated by the nitrate-N losses contributed from subsurface flow divided by the sum of nitrate-N losses from both surface and subsurface flow)..... 75
Figure 3.4 Cumulative distribution functions of sensitive parameter of all scenarios. 76
Figure 3.5 Quantile-Quantile plots for calibration periods for all scenarios: (A) Scenario I with no constraint applied, (B) Scenario II with constraint of denitrification, (C) Scenario III with constraint of SSQ ratio, (D) Scenario IV with both constraints applied. The dot lines in (A) and in (C) represent the best solutions without considering both constraints. 78
Figure 3.6 Quantile-Quantile plots for Evaluation periods for all Scenarios: (A) Scenario I with no constraint applied, (B) Scenario II with constraint of denitrification, (C) Scenario III with constraint of SSQ ratio, (D) Scenario IV with both constraints applied. 79
Figure 4.1 Case study area: Eagle Creek Watershed, Indiana 90
Figure 4.2 Overall performance of objective function versus model evaluations of Scenario 1 (apply SCS method by antecedent soil moisture condition) and Scenario 2 (apply SCS method by plant evapotranspiration) 102
Figure 4.3 Brier score for SCSI, SCSII and BMA associated methods: (A) Streamflow (gauge station 35), (B) Total nitrate (gauge station 32). The flow and the total nitrate threshold intervals are defined by ranking from the highest (interval = 10) to the lowest (interval = 1) quantities of observation data for specific gauge station. 105
Figure 4.4 Quantile-Quantile plots for calibration periods for all cases: (A) Scenario 1 – SCSI applied, (B) Scenario 2 – SCSII applied, (C) Scenario 3 – Same BMA

weights for entire period, (D) Scenario 4 – Different BMA weights for dry/wet seasons, (E) Scenario 5 – Different BMA weights for warm/cool and wet/dry periods respectively. 108

Figure 4.5 Quantile-Quantile plots for evaluation periods for all cases: (A) Scenario 1 – SCSI applied, (B) Scenario 2 – SCSII applied, (C) Scenario 3 – Same BMA weights for entire period, (D) Scenario 4 – Different BMA weights for dry/wet seasons, (E) Scenario 5 – Different BMA weights for warm/cool and wet/dry periods respectively. 109

Figure 5.1 Case study area: Eagle Creek Watershed, Indiana 120

Figure 5.2 Framework of uncertainty analysis incorporating four sources of uncertainty .. 131

Figure 5.3 Overall performance of objective function values versus model evaluations for all cases (Figure 5.3 is truncated at the beginning of model evaluation is because the initial results for several scenarios are multiple times higher than other scenarios.) 134

Figure 5.4 Convergence process of denitrification in Scenario I and V 141

Figure 5.5 Cumulative distribution functions of the denitrification constraint applied in all case scenarios: (A) Scenario I~IV with only one source of uncertainty included (except for parameter uncertainty), (B) Scenario V~VIII with all sources of uncertainty included 142

Figure 5.6 Cumulative distribution functions and marginal posterior distributions of latent variables in Scenario II, V, VI, VII and VIII: 6(A) CDF of latent variable nu, 6(B) CDF of latent variable sigma, 6(C) marginal posterior

distributions of latent variable nu, 6(D) marginal posterior distributions of latent variable sigma 143

Chapter 1

Introduction

1.1. Overview

Comprehensive watershed models are used to simulate hydrologic and water quality processes to evaluate the potential impact under changing land use and climatic conditions and provide information to decision makers and stakeholders. In recent years, an increasing number of parameters are included in watershed models to enhance their capability to represent hydrologic and water quality responses at various spatial and temporal scales [*Bai et al.*, 2009]. However, many model parameters cannot be measured directly and must be estimated using manual or automatic parameter estimation techniques. It has been demonstrated that multiple parameter sets may have comparable statistical solutions due to the property of parameter nonuniqueness [*Sorooshian and Gupta*, 1983]. Hence, analysis of uncertainty should be further explored during the parameter estimation process. In addition, the importance of uncertainty estimation has been previously addressed in the literature: *"Uncertainty must be explicitly acknowledged both in the models selected to develop TMDLs and in results generated by those models"* [*NRC*, 2001]. Therefore, the incorporation of parameter and uncertainty estimation is essential for applications of complex watershed models.

The advantage of manual calibration is that analysts can gain a better understanding of hydrologic/nutrients processes in a system via careful examination of individual model

parameters. However, estimation of model parameters using manual procedures can be cumbersome in highly parameterized watershed models. Therefore, a variety of automatic parameter estimation techniques using different computational strategies have been developed [*Duan et al.*, 1992; *Haario et al.*, 2006; *Klepper and Hendrix*, 1994; *Tolson and Shoemaker*, 2007; *Vrugt et al.*, 2003, 2009]. Applications of the latest parameter estimation algorithms hinge on improving the computational and convergence efficiency in finding the minimum value of an objective function. Advantages and disadvantages among algorithms can vary dramatically from case to case. Although minimization of an objective function of model errors is typically required for identifying optimal solutions, it does not guarantee that the final optimal solution adequately represents the system's behavior.

In addition to computational efficiency, the number of solutions that adequately represent observed responses for multiple variables, such as stream discharge, sediment and nutrient loads, at multiple sites along the channel network is also of interest. For example, percent bias, Nash-Sutcliffe efficiency coefficient and coefficient of determination are among the commonly applied error statistics for the evaluation of the performance validity of watershed models. The behavior of a system is defined based on pre-defined thresholds of behavior established by error statistics in matching various system behavior characteristics. Parameter sets that satisfy a behavior definition are regarded as behavior solutions. Incorporation of behavior definition in searching for optimal solutions is rarely described in the literature [*Moriasi et al.*, 2007] and the choice of the various behavior standards may be subjective.

In summary, it is necessary to have a systematic procedure for identifying the effectiveness of various sampling strategies. In the first part of this study a computational procedure is

developed and demonstrated to facilitate selection of a proper sampling strategy for analysis of uncertainty incorporating input, parameter, structural and measurement sources.

After the selection of sampling strategies is performed, the second part of this study explores the uncertainty from various sources. Estimation of model parameters can be influenced by errors in forcing inputs such as precipitation, land use and soil types; errors in measurements of hydrologic and water quality fluxes; and also from different model structures. Parameterization of parametric uncertainty has been conducted by a large number of studies over the past two decades [*Gallagher and Doherty*, 2007; *Hassan et al.*, 2009; *Kuczera and Parent*, 1998; *Loosvelt et al.*, 2011; *Osidele et al.*, 2006]. Prior to other sources of uncertainty being addressed in watershed calibration problems, parameter uncertainty has first been identified as the major source of overall uncertainty. Many efforts have been made to estimate parameter uncertainty and enhance the computational efficiency by refining optimization procedures [*Duan et al.*, 1992; *Klepper and Bedaux*, 1997; *Klepper and Hendrix*, 1994; *Tolson and Shoemaker*, 2007; *Vrugt et al.*, 2003, 2008]. In contrast, analysis of input, structural and measurement uncertainty has only been examined individually and few studies have input, parameter and structural uncertainty aggregated and explored jointly [*Ajami et al.*, 2007; *Kavetski et al.*, 2002].

Input uncertainty is rarely incorporated explicitly into watershed models. The Bayesian total error analysis (BATEA) [*Kavetski et al.*, 2002] method first includes input uncertainty into uncertainty analysis by introducing additional latent variables (input quantities at each time step are assigned a latent variable) into a systematic framework. The integrated Bayesian uncertainty estimator (IBUNE) [*Ajami et al.*, 2007] method applies a similar structure to BATEA to incorporate input uncertainty but the number of latent variables is reduced to two. Therefore, the

potential problem of considerable computational setback because of high dimensionality can be avoided.

Structural uncertainty is mostly investigated for models which use different techniques to derive the same output variables, for example, surface runoff generated from different rainfall runoff models [*Ajami et al.*, 2007; *Duan et al.*, 2007; *Wöhling and Vrugt*, 2008]).The selection of models may involve a certain amount of subjectivity especially when they are sensitive to specific case studies. To remove the subjectivity in selecting the appropriate model to solve practical case studies, outputs from different models can be aggregated by the statistical approach named the Bayesian model averaging (BMA) [*Hoeting et al.*, 1999] technique. Models with the same output quantities can be evaluated based on their ability to simulate real world phenomenon and the predictive distributions of model outputs can be combined by assigning particular weighting factors. It has been shown that structural uncertainty can be reduced by implementing the BMA technique where the superiority of each implemented model can be quantified by the BMA weights.

Compared to sources of uncertainty from forcing inputs, parameter, and model structure, measurement uncertainty has not been addressed frequently in the literature [*Harmel et al.*, 2006]. One of the most important reasons is that an insufficient amount of data is available to support the research work. The other major reason is that a scientific guidance/standards for evaluating measurement uncertainty is not available [*Harmel and Smith*, 2007]. Error statistics are calculated by deterministic observation data but measurement error should also be identified [*Harmel and Smith*, 2007; *Harmel et al.*, 2006]. Previous studies mostly focus on the incorporation of input, parameter and structural uncertainty [*Ajami et al.*, 2007; *Duan et al.*, 2007; *Kavetski et al.*, 2002; *Wöhling and Vrugt*, 2008]. The absence of involving one or more

sources of uncertainty may cause model outputs to be biased and lead to an incorrect conclusion about the source of the errors. Typically sources of uncertainty have not been examined simultaneously with measurement uncertainty. An additional issue is that the comparative importance among different sources of uncertainty is still unknown.

In this study, the influence of measurement uncertainty towards predictive uncertainty will be investigated with uncertainty from input, parameter and structural sources by recommended standards provided from previous work [*Harmel et al.*, 2006]. Therefore, the comparative importance from each source can be further explored to answer scientific questions.

In this study, Bayesian inferences are applied to incorporate input, parameter, structural and measurement uncertainty using a statistically valid likelihood function [*Ahmadi et al.*, 2012]. The framework of IBUNE which aggregates input, parameter, and structural uncertainty is used with the probability distribution (PD) [*Harmel and Smith*, 2007] method to perform the calculation for inclusion of measurement uncertainty.

The second part of this study developed a framework for the analysis of input, parameter, structural and measurement uncertainty in multi-site, multiple responses watershed modeling. The influence from each source on model outputs can be identified and analyzed for the purpose of distributing available resources (e.g. time and money) for further investigation. Case studies are used for evaluation of sampling strategies and uncertainty analyses with four sources of uncertainty were analyzed for a 248 square kilometer Eagle Creek watershed in the Midwestern United States. The soil and water assessment tool (SWAT) is adopted to simulate hydrologic and total nitrate processes.

In this dissertation, chapters are formulated in the fashion of independent but related topics as journal articles with an overall introduction in the beginning with summarized discussion and

conclusion in the end. The chapters and contents in this study are organized as follows: The first chapter has a brief view of the proposed framework with specified goals and objectives. In the second chapter, a computational procedure is implemented in a real case study to demonstrate the capabilities in evaluating sampling strategies. In the third chapter, the importance of the intra-watershed processes towards parameter estimation is substantiated by applying additional constraints. The structural uncertainty is explored by the use of two different functions in the SWAT model to calculate surface runoff in the fourth chapter. In the fifth chapter, four sources of uncertainty are aggregated using the proposed framework where corresponding impact towards predictive uncertainty is investigated. Conclusions and discussion are provided in the sixth chapter.

1.2. First Part of the Framework (Chapter II, III): Development of a Framework to Evaluate the Effectiveness of Various Sampling Strategies

In order to have a rigorous framework to compare various sampling strategies, a systematic framework is required to perform the evaluation process that does not only focus on finding the global optimal solution of an objective function. The first part of this study established a procedure for evaluating the effectiveness of sampling strategies. Several essential elements are mandatory to develop a framework for evaluating sampling strategies.

1.2.1. Statistically Valid Likelihood Function

In manual/auto calibration processes, the performance of a given model parameter set is evaluated based on the error statistics calculated from available observed and simulated outputs. For auto calibration techniques, the current best candidate parameter set(s) will be regarded as the new starting point(s) for following realization(s) (this is true in Bayesian inferences but not

for all parameter estimation techniques). In other words, the faster a parameter estimation technique can derive lower values of the objective function being minimized (in practice, the likelihood function to be maximized is multiplied by a negative sign to create a minimization problem). A method with better computational efficiency is one with superior convergence speed and the ability to achieve global or near global optimal solutions.

Previous studies mostly focus on comparing convergence speed and the ability in finding global optimal solutions instead of emphasizing the selection of the likelihood function [*Duan et al.*, 1992; *Klepper and Hendrix*, 1994; *Marshall et al.*, 2004; *Tolson and Shoemaker*, 2007]. However, the selection of the likelihood function is crucial since parameter sampling distributions of model errors will not be realistically reflected if the likelihood function is not statistically sound [*Stedinger et al.*, 2008]. Therefore, a statistically valid likelihood function [*Ahmadi et al.*, 2012] is adopted as the objective function throughout this study.

1.2.2. Performance Validity by Applying Behavior Definition

Candidate parameter sets proposed by parameter estimation techniques may have results with very good objective function value however the parameter set may not necessarily satisfy other statistics such as the Nash-Sutcliffe efficiency coefficient and percent bias which can be important as well. During the calibration process, proposed parameter sets should be selected that have reasonable behavior. In other words, various standards in evaluating behavior levels are required to identify parameter sets with significant or rational performance.

As illustrated in Figure 1.1, D is the full domain space of all parameters involved in a conceptual problem. According to the definition of behavior, one can reduce D to be the space with only behavior parameter sets BH. Then, three subdomain spaces can be found by using three commonly implemented parameter estimation techniques; Method A, B and C. Figure 1.1

illustrates several ideas. First, the behavior parameter sets should always be included in D. Second, the subdomains derived by a specific method will not necessary overlap with BH. Third, Method A is has more behavior parameter sets than Method B but it is not guaranteed to have better overall performance. The reason is because the number of behavior parameter sets represents only the parameter sets with better manually assigned statistical (it can also be defined by many other different ways, [*Beck et al.*, 2002]) values and there is still chance that the global optimal solution is located in some other part of BH.

Previous calibration work can rarely be found with applications of behavior definition. In this study, the behavior definition proposed by [*Moriasi et al.*, 2007] is implemented as the standard guideline to assess the categories of behavior for proposed parameter sets.

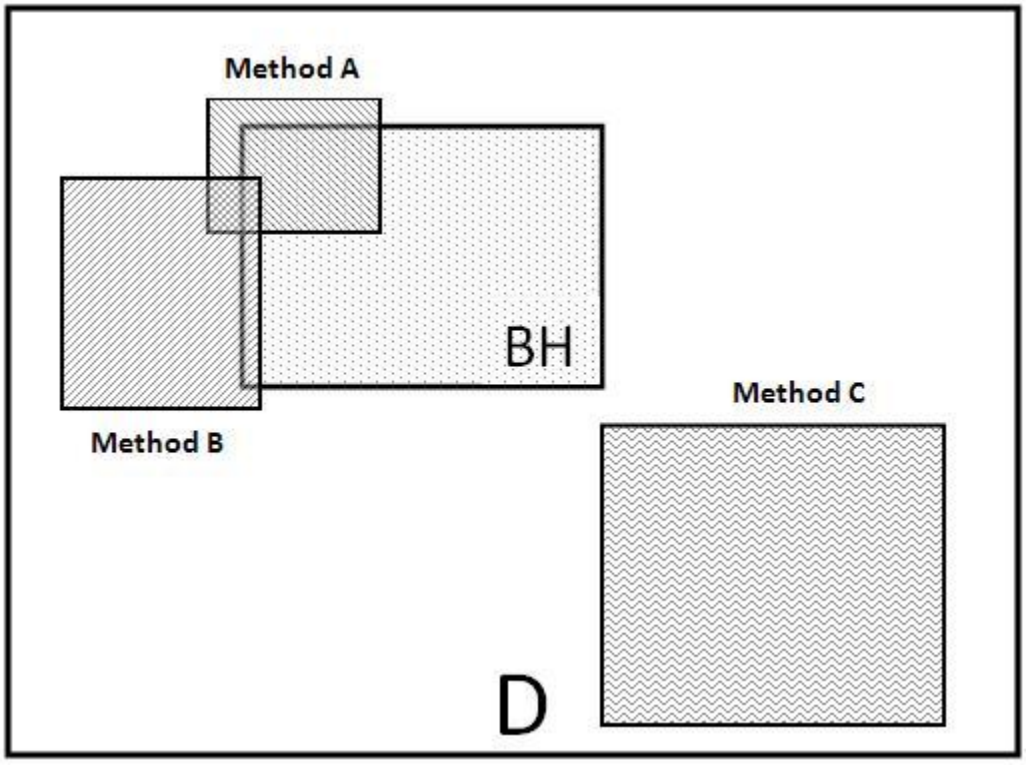

Figure 1.1 Demonstration of behavior parameter sets and the correspond subdomain spaces

1.2.3. Identifiability of Individual Model Parameters and Entire Parameter Hypercube

Computational efficiency and the application of behavior definition are used to evaluate the relative success of a parameter estimation technique. However, identifiability provides a supplementary understanding of these measures. Identifiability is a measure of the density of parameter space sampled by parameter estimation algorithms, with higher density parameters and parameter spaces corresponding to higher identifiability. Identifiability has also been applied in previous studies to assess the complexity of simulation models [*Spear et al.*, 1994]. Global optimization algorithms are better able to solve less identifiable problems (such as many-parameter, high-dimensional watershed models) and are less likely to be trapped in local solutions than algorithms that search only for a parameter set with the best objective function value. However, there is a necessary tradeoff in computational efficiency where identifiability yields insight into such tradeoffs. A parameter estimation technique that gives highly identifiable solutions may converge quickly because it draws from a relatively small area of the possible parameter space. If this parameter space corresponds to a global solution, then computational efficiency, objective functions, and behavior rates may be optimized; however, if the parameter space corresponds to a local optimum, the high identifiability may prevent the estimation algorithm from encountering better solutions such as the global optimum.

Previous studies have analyzed identifiability only for individual model parameters, but not for the entire domain space [Duan et al., 1992; Klepper and Hendrix, 1994; Marshall et al., 2004; Haario et al., 2006; Vrugt et al., 2009], giving an incomplete understanding of the full parameter space sampled by search algorithms. In this study, tree-structure density estimation (TSDE)

[*Spear et al.*, 1994] is implemented to investigate the identifiability of the entire parameter domain space.

By the proposed framework, an assessment of sampling strategies can be made not only based on computational efficiency in searching for the global optimal solution but also the effectiveness in drawing samples that adequately represent the observed behavior of the system at multiple locations within the watershed system. In the meantime, the framework provides supplementary information on the identifiability of the entire parameter space which is supportive for further inspection about various parameter estimation techniques. The innovative evaluation structure enables researchers to understand the ability of parameter estimation techniques in searching for better solutions and also have a better understanding of the parameter estimation processes in high dimensional domain space.

1.3. Second Part of the Framework (Chapter IV, V): Development of a Framework to Incorporate Input, Parameter, Structural and Measurement Uncertainty

As shown in Figure 1.2, sources of uncertainty in watershed modeling can be represented by four parts: input, parameter, structural and measurement uncertainty. Input uncertainty includes forcing inputs such as precipitation, temperature, and land use. In this study, the input uncertainty is incorporated into the watershed simulation model (Soil and Water Assessment Tool, SWAT) by applying the approach from IBUNE. Details of IBUNE applications can be found in later chapters.

Parameter uncertainty represents the uncertainty from physically/empirically based equations with various adjustable parameters such as curve number, hydraulic conductivity, and

Manning's roughness (n). During the calibration process, it is unavoidable to have system parameters involved to explore the uncertainty from other sources. In other words, comparisons among different uncertainties are guaranteed to have parameter uncertainty conditioned jointly.

Another source of uncertainty involved in the watershed simulation model comes from the model structure. For example, alternative methods in computing surface runoff, channel erosion and sediment transport may be available within a single complex watershed model. For complex watershed simulation models like SWAT which include several different functions to calculate the same output variable (e.g. two methods are available in the SWAT model for computing the amount of surface runoff) can be investigated to identify structural uncertainty. In this study, the structural uncertainty contributed from two different approaches, surface runoff calculation computed by antecedent soil moisture conditions [SCSI] and plant evapotranspiration [SCSII], are implemented in calculating surface runoff in the SWAT model. Details of the two methods can be found in literature [*Neitsch et al.*, 2011].

After executing model simulations, objective functions and error statistics are calculated using measured fluxes such as streamflow discharge and nutrient load where measurement uncertainty is involved. However, studies of measurement uncertainty can rarely be found. In this study, the influence of measurement uncertainty towards predictive uncertainty will be investigated with uncertainty from input, parameter and structural sources by recommended standards provided from previous work [*Harmel et al.*, 2006]. Therefore, the comparative importance from each source can be further explored to answer scientific questions.

After a specific sampling strategy is selected based on the framework of evaluating sampling strategies, uncertainty analysis can be then be performed by the framework which is able to incorporate input, parameter, structural and measurement uncertainty jointly. The innovative

framework of predictive uncertainty analysis provides a platform to investigate not only the importance of each uncertainty source by comparing uncertainty sources individually, but also provide comprehensive information about the predictive uncertainty affected by jointly combining different uncertainty sources. Researchers and decision makers can take advantage of this powerful tool to provide assistance for various purposes based on their own interests.

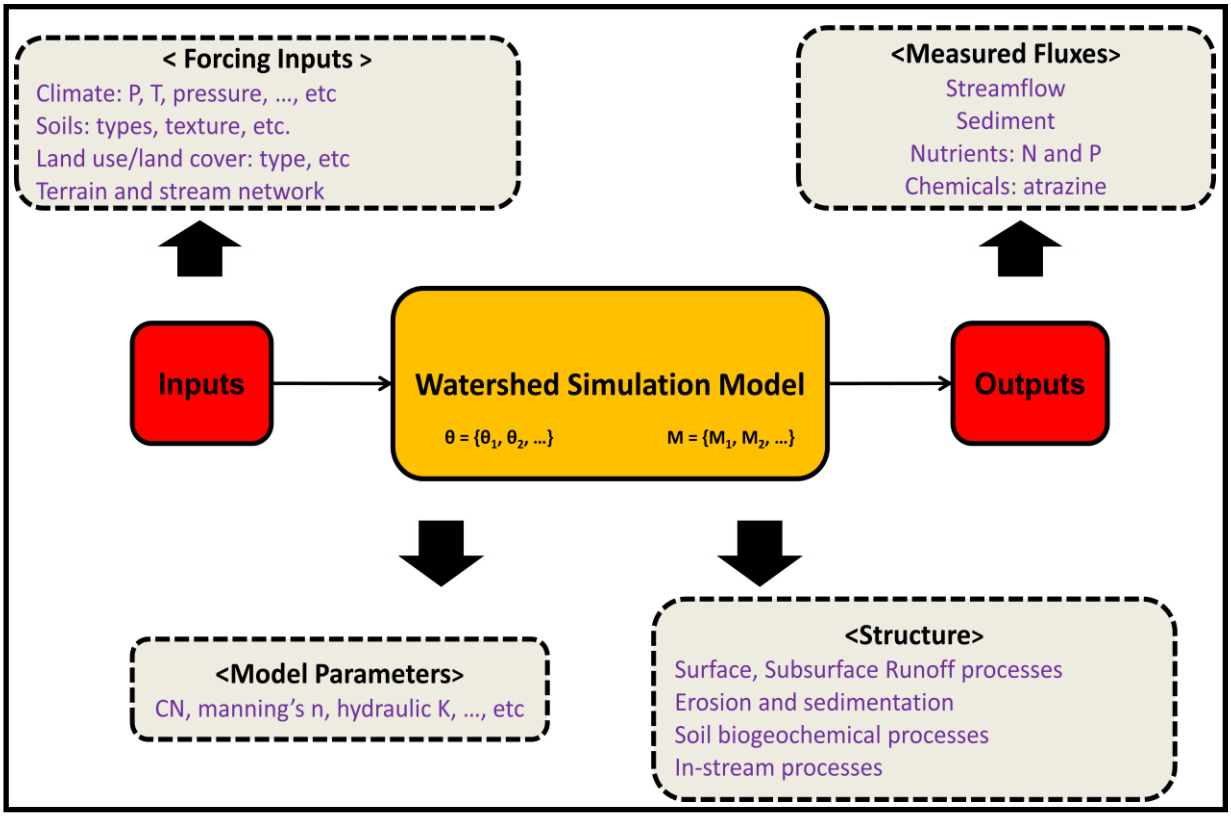

Figure 1.2 Sources of uncertainty in watershed modeling

1.4. Goals and Objectives

The framework proposed in this study can be applied in various forms according to the users' interests. As shown in Figure 1.3, each part of the framework during the watershed calibration process is managed in a general way so that different sources of uncertainty can be included or excluded. In addition, the choice of selecting sampling strategies may differ by varying

evaluation criteria. For example, the shuffled complex evolution algorithm may not be the best choice in the case that all sources of uncertainty are included simultaneously. The overall goal of this work is to develop a solid framework from the selection of sampling strategies (first part of the framework) to the investigation of predictive uncertainty analysis (second part of the framework). Specifically, four major sections (Chapter II, III, IV and V) are designed to serve the purpose of exploring challenging tasks and to respond to specific scientific questions.

1.4.1. Appraisal of Sampling Strategies for Parameter Estimation and Uncertainty Analysis by Using a Systematic Evaluation Framework in Watershed Modeling

The first part of the framework is to develop and demonstrate a systematic procedure for evaluating the performance of sampling strategies for multi-site multi-response parameter estimation and uncertainty analysis of comprehensive watershed models. The following objectives are defined:

(i) To compare the efficiency of the methods in finding better solutions, that is to compare the objective function values after convergence

(ii) To explore the efficiency of parameter estimation algorithms in finding the best solution within a given number of runs noting that the best solution may not be the same as the global optimal solution

(iii) To measure the effectiveness of the methods in realizing parameter samples that match the system behavior characterized by commonly used error statistics at various locations within the watershed

(iv) To evaluate the identifiability of the overall model parameter space using these methods

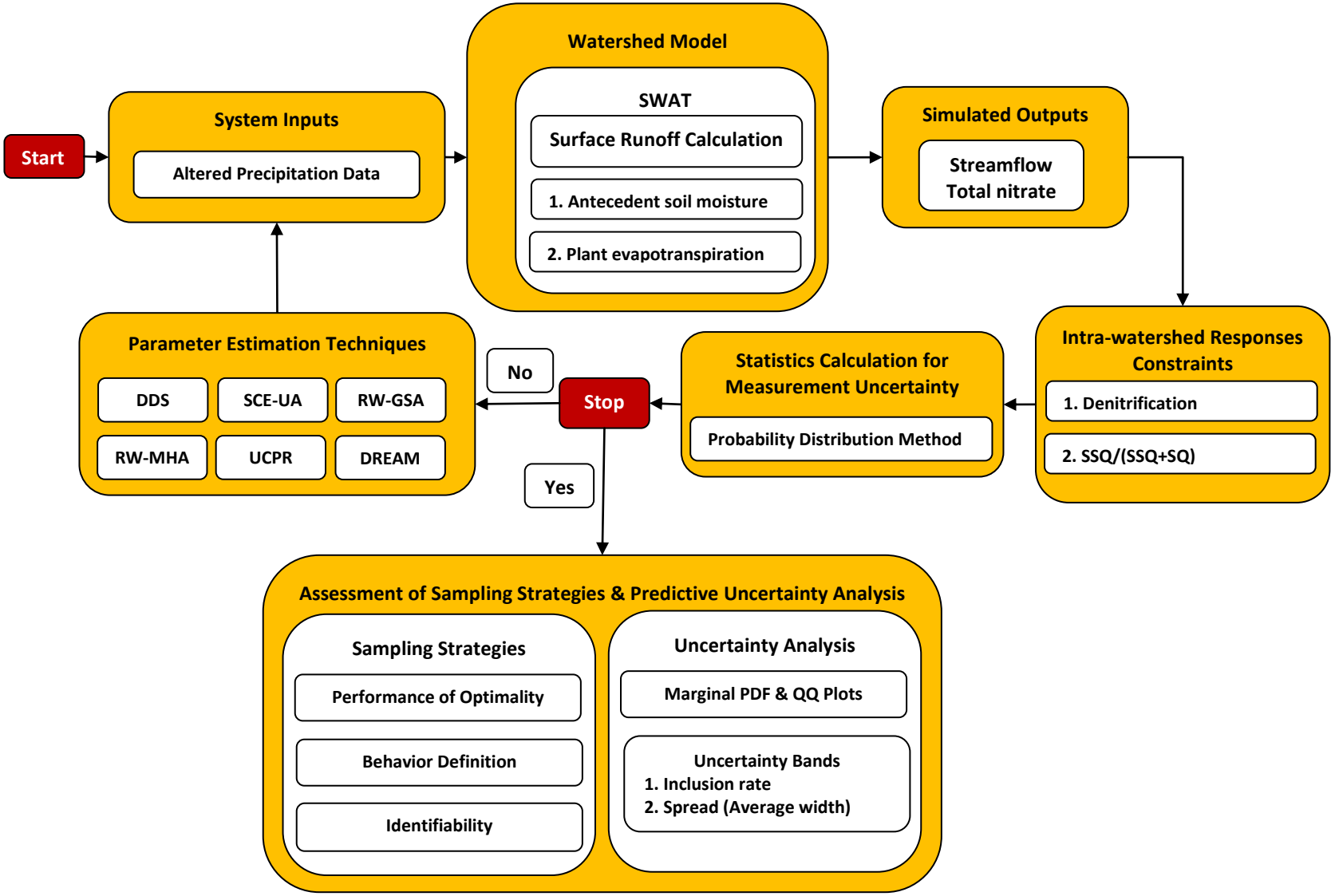

Figure 1.3 Framework of the evaluation of sampling strategies and predictive uncertainty

The development of a rigorous framework for evaluation of sampling strategies provides a tool for researchers/users to have scientific measures and comparisons of various parameter estimation techniques. It is also the groundwork before proceeding to uncertainty analysis.

1.4.2. Evaluation of Improved Parameter Estimation and Uncertainty Analysis of Watershed Models: The Role of General Information of Watershed Processes

The calibration process mostly seeks statistical coherence on the error calculated based on observation and simulation outputs such as streamflow discharge, nutrient load. General information of intra-watershed processes (e.g. aggregated system outputs such as denitrification and total nitrate yields over years) is generally not considered. However, the absence of a general understanding of a case study watershed may lead to having impressive statistical results with aggregate outputs compared against some crucial characteristics in field. In other words, the calibration process with outstanding statistical results is successfully performed for the wrong reason. To strengthen the quality of calibration results and also to evaluate the role of intra-watershed processes, the following objectives are defined:

(i) To examine how incorporation of intra-watershed processes enhances parameter estimation and calibration of the watershed model

(ii) To investigate how incorporation of intra-watershed processes reduces parameter and predictive uncertainties.

The application of intra-watershed processes represents a higher level of the calibration process which involves not only the time varying quantities such as streamflow discharge and total nitrate load, but also summative watershed outputs considered to reflect real watershed behavior processes.

1.4.3. Evaluation of Structural Uncertainty on Hydrologic and Water Quality Predictions Using SWAT Model

In the latest version of the SWAT model, users can have alternative options to compute surface runoff based on antecedent soil moisture condition (SCSI) or plant evapotranspiration (SCSII). The calculation of surface runoff is very important since many of the nutrient elements can be carried by water. Therefore, the structural uncertainty is examined by implementing two different approaches in calculating surface runoff mechanism in the SWAT model before including all sources of uncertainty simultaneously. The Bayesian model averaging (BMA) technique [*Hoeting et al.*, 1999] is implemented to combine posterior distributions of the two approaches. The following objectives are defined:

(i) To characterize improvements in hydrologic and water quality predictions by utilizing different surface runoff estimation techniques alone and in combination

(ii) To investigate how model predictive uncertainty may be affected by combining such techniques.

Applications of SCSII still can rarely be found in the literature [*Amatya and Jha*, 2011; *Jajarmizadeh et al.*, 2012; *Kannan et al.*, 2007]. Therefore, evaluation between SCSI and SCSII indicates not only the comparisons of their ability in making more realistic simulation of real watersheds, but also reveals a general idea how predictive uncertainty can be influenced by different functions in the same watershed simulation model.

1.4.4. A Framework for Propagation of Input, Parameter, Structural and Measurement Uncertainty in Watershed Modeling

The first part of the framework is developed for the selection of sampling strategies. Once a parameter estimation technique is chosen, predictive uncertainty can be investigated by the

second part of the framework. As mentioned previously, the ultimate goal of watershed calibration is to replicate watershed processes as close as possible through simulation models. Users can take advantage of the calibrated results and implement various scenarios such as management practices or human activities and compare differences among outputs. However, it is also important to have additional aspects such as uncertainty analysis for model parameters or outputs as supplementary information before making decisions. To address the lack of uncertainty analysis and predictive uncertainty during the calibration processes and also to have all sources of uncertainty involved simultaneously, the second part of the framework develops and demonstrates a systematic procedure for uncertainty analysis which incorporate uncertainty sources from input, parameter, structural and measurement jointly. The impact from each source towards predictive uncertainty can be identified and analyzed for the purposed of devoting possible resources for further investigation. The following objectives are defined:

(i) To quantify predictive uncertainty while propagating four sources of error

(ii) To understand the role and importance of four uncertainty sources on predictive uncertainty

(iii) To examine the effects of reduction of input, parameter, structural and measurement uncertainty in predictive uncertainty

The first and the second part of the framework provide a complete approach for analyzing watershed calibration problems. By using the proposed framework, users can apply various sampling strategies for watershed calibration and resolve scientific questions based on the uncertainty analysis in the field of watershed modeling.

References

Ahmadi, M., M. Arabi, J. C. Ascough II, D. G. Fontane, and B. A. Engelc (2012), Toward improved calibration of watershed models: Multisite many objective measures of information, *Submitted to: Advances in Water Resources*.

Ajami, N. K., Q. Duan, and S. Sorooshian (2007), An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction, *Water Resources Research*, *43*(1), 1–19, doi:10.1029/2005WR004745. [online] Available from: http://www.agu.org/pubs/crossref/2007/2005WR004745.shtml (Accessed 28 June 2011)

Amatya, D. M., and M. K. Jha (2011), Evaluating the SWAT model for a low-gradient forested watershed in coastal south Carolina, *Transactions of the ASABE*, *54*(6), 2151–2163.

Bai, Y., T. Wagener, and P. Reed (2009), A top-down framework for watershed model evaluation and selection under uncertainty, *Environmental Modelling & Software*, *24*(8), 901–916, doi:10.1016/j.envsoft.2008.12.012. [online] Available from: http://linkinghub.elsevier.com/retrieve/pii/S1364815208002399 (Accessed 4 May 2012)

Beck, M. B. et al. (2002), Developing a concept of adaptive community learning: Case study of a rapidly urbanizing watershed, *Integrated Assessment*, *3*(4), 299–307, doi:10.1076/iaij.3.4.299.13583. [online] Available from: http://www.tandfonline.com/doi/abs/10.1076/iaij.3.4.299.13583

Duan, Q., N. K. Ajami, X. Gao, and S. Sorooshian (2007), Multi-model ensemble hydrologic prediction using Bayesian model averaging, *Advances in Water Resources*, *30*(5), 1371–1386, doi:10.1016/j.advwatres.2006.11.014. [online] Available from: http://linkinghub.elsevier.com/retrieve/pii/S030917080600220X (Accessed 12 March 2012)

Duan, Q., S. Sorooshian, and V. Gupta (1992), Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resources Research*, *28*(4), 1015–1031.

Gallagher, M., and J. Doherty (2007), Parameter estimation and uncertainty analysis for a watershed model, *Environmental Modelling & Software*, *22*(7), 1000–1020, doi:10.1016/j.envsoft.2006.06.007. [online] Available from: http://linkinghub.elsevier.com/retrieve/pii/S1364815206001629 (Accessed 12 July 2011)

Haario, H., M. Laine, A. Mira, and E. Saksman (2006), DRAM: Efficient adaptive MCMC, *Statistics and Computing*, *16*(4), 339–354, doi:10.1007/s11222-006-9438-0. [online] Available from: http://www.springerlink.com/index/10.1007/s11222-006-9438-0 (Accessed 28 June 2011)

Harmel, R. D., R. Cooper, R. Slade, R. Haney, and J. Arnold (2006), Cumulative uncertainty in measured streamflow and water quality data for small watersheds, *Transactions of the ASABE*, *49*(3), 689. [online] Available from: http://asae.frymulti.com/abstract.asp?aid=20488&t=1 (Accessed 22 June 2011)

Harmel, R. D., and P. Smith (2007), Consideration of measurement uncertainty in the evaluation of goodness-of-fit in hydrologic and water quality modeling, *Journal of Hydrology*, *337*(3-4), 326–336, doi:10.1016/j.jhydrol.2007.01.043. [online] Available from: http://linkinghub.elsevier.com/retrieve/pii/S0022169407000741 (Accessed 10 June 2011)

Hassan, A. E., H. M. Bekhit, and J. B. Chapman (2009), Using Markov Chain Monte Carlo to quantify parameter uncertainty and its effect on predictions of a groundwater flow model, *Environmental Modelling & Software*, *24*(6), 749–763, doi:10.1016/j.envsoft.2008.11.002.

[online] Available from: http://linkinghub.elsevier.com/retrieve/pii/S1364815208002016 (Accessed 12 July 2011)

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999), Bayesian Model Averaging : A Tutorial, *Statistical Science*, *14*(4), 382–417.

Jajarmizadeh, M., S. Harun, M. H. Mokhtari, and B. Ghahraman (2012), Impact of Behaviour of New accounting curve number method using SWAT model on stream flow, *2012 International Conference on Environment Science and Engineering*, *3*, 124–128.

Kannan, N., C. Santhi, J. R. Williams, and J. G. Arnold (2007), Development of a continuous soil moisture accounting procedure for curve number methodology and its behaviour with different evapotranspiration methods, *Hydrological Processes*, *22*(13), 2114–2121, doi:10.1002/hyp.6811. [online] Available from: http://doi.wiley.com/10.1002/hyp.6811 (Accessed 7 May 2012)

Kavetski, D., S. W. Franks, and G. Kuczera (2002), Confronting input uncertainty in environmental modeling, *Water Science and Application 6*, 49–68, doi:10.1016/j.envsoft.2005.04.010. [online] Available from: http://linkinghub.elsevier.com/retrieve/pii/S1364815205000873 (Accessed 20 June 2011)

Klepper, O., and J. J. M. Bedaux (1997), A ROBUST METHOD FOR NONLINEAR PARAMETER ESTIMATION, *Science*, *30*(3), 1677–1686.

Klepper, O., and E. Hendrix (1994), A method for robust calibration of ecological models under different types of uncertainty, *Ecological Modelling*, *74*(3-4), 161–182, doi:10.1016/0304-3800(94)90118-X. [online] Available from: http://linkinghub.elsevier.com/retrieve/pii/030438009490118X

Kuczera, G., and E. Parent (1998), Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm, *Journal of Hydrology*, *211*(1-4), 69–85, doi:10.1016/S0022-1694(98)00198-X. [online] Available from: http://linkinghub.elsevier.com/retrieve/pii/S002216949800198X (Accessed 24 June 2011)

Loosvelt, L., V. R. N. Pauwels, W. M. Cornelis, G. J. M. De Lannoy, and N. E. C. Verhoest (2011), Impact of soil hydraulic parameter uncertainty on soil moisture modeling, *Water Resources Research*, *47*(3), 1–16, doi:10.1029/2010WR009204. [online] Available from: http://www.agu.org/pubs/crossref/2011/2010WR009204.shtml (Accessed 12 July 2011)

Marshall, L., D. Nott, and A. Sharma (2004), A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modeling, *Water Resources Research*, *40*(2), 1–11, doi:10.1029/2003WR002378. [online] Available from: http://www.agu.org/pubs/crossref/2004/2003WR002378.shtml (Accessed 28 June 2011)

Moriasi, D. N., J. G. Arnold, M. W. V. Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith (2007), Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, *Transactions of the ASABE*, *50*(3), 885–900.

NRC (2001), *Assessing the TMDL approach to water quality management*, National Academy Press Washington, D.C.

Neitsch, S. L., J. G. Arnold, J. R. Kiniry, and W. J. R (2011), Soil & Water Assessment Tool Theoretical Documentation Version 2009, *Texas Water Resources Institute Technical Report No. 406, Texas A&M University System*.

Osidele, O. O., W. Zeng, and M. B. Beck (2003), Coping with Uncertainty: A Case Study in Sediment Transport and Nutrient Load Analysis, *Journal of Water Resources Planning and Management*, *129*(4), 345, doi:10.1061/(ASCE)0733-9496(2003)129:4(345). [online]

Available from: http://link.aip.org/link/JWRMD5/v129/i4/p345/s1&Agg=doi (Accessed 28 June 2011)

Osidele, O. O., W. Zeng, and M. B. Beck (2006), A random search methodology for examining parametric uncertainty in water quality models, *Water Science & Technology*, *53*(1), 33, doi:10.2166/wst.2006.005. [online] Available from: http://www.iwaponline.com/wst/05301/wst053010033.htm (Accessed 28 June 2011)

Sorooshian, S., and V. K. Gupta (1983), Automatic calibration of conceptual rainfall-runoff model: The question of parameter observability and uniqueness, *Water Resources Research*, *19*(1), 260–268.

Spear, R. C., T. M. Grieb, and N. Shang (1994), Parameter uncertainty and interaction in complex environmental models, *Water Resources Research*, *30*(11), 3159, doi:10.1029/94WR01732. [online] Available from: http://www.agu.org/pubs/crossref/1994/94WR01732.shtml (Accessed 28 June 2011)

Stedinger, J. R., R. M. Vogel, S. U. Lee, and R. Batchelder (2008), Appraisal of the generalized likelihood uncertainty estimation (GLUE) method, *Water Resources Research*, *44*, 1–17, doi:10.1029/2008WR006822. [online] Available from: http://www.agu.org/pubs/crossref/2008/2008WR006822.shtml (Accessed 22 June 2011)

Tolson, B. A., and C. A. Shoemaker (2007), Dynamically dimensioned search algorithm for computationally efficient watershed model calibration, *Water Resources Research*, *43*(1), 1–16, doi:10.1029/2005WR004723. [online] Available from: http://www.agu.org/pubs/crossref/2007/2005WR004723.shtml (Accessed 28 June 2011)

Vrugt, J. A., C. J. F. Braak, C. G. H. Diks, B. A. Robinson, J. M. Hyman, and D. Higdon (2009), Accelerating Markov Chain Monte Carlo Simulation by Differential Evolution with Self-Adaptive Randomized Subspace Sampling, *International Journal of Nonlinear Sciences & Numerical Simulation*, *10*(March), 271–288.

Vrugt, J. A., C. J. F. ter Braak, M. P. Clark, J. M. Hyman, and B. A. Robinson (2008), Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation, *Water Resources Research*, *44*, 1–15, doi:10.1029/2007WR006720. [online] Available from: http://www.agu.org/pubs/crossref/2008/2007WR006720.shtml (Accessed 24 August 2011)

Vrugt, J. A., H. V. Gupta, W. Bouten, and S. Sorooshian (2003), A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resources Research*, *39*(8), doi:10.1029/2002WR001642. [online] Available from: http://www.agu.org/pubs/crossref/2003/2002WR001642.shtml (Accessed 10 June 2011)

Wöhling, T., and J. a. Vrugt (2008), Combining multiobjective optimization and Bayesian model averaging to calibrate forecast ensembles of soil hydraulic models, *Water Resources Research*, *44*(12), 1–18, doi:10.1029/2008WR007154. [online] Available from: http://www.agu.org/pubs/crossref/2008/2008WR007154.shtml (Accessed 5 March 2012)

Chapter 2

Appraisal of Sampling Techniques for Parameter Estimation and Uncertainty Analysis in Watershed Modeling

2.1. Overview

A variety of computational techniques have been developed to efficiently and effectively draw realizations from the parameter space for the purposes of parameter estimation and the analysis of uncertainty of watershed models. The main goal of this study is to develop and demonstrate a computational procedure for the appraisal of parameter sampling techniques. The analysis hinges on the evaluation of: (i) the efficiency in minimizing an objective function of weighted model errors at the lowest required realizations; (ii) the effectiveness in drawing samples that adequately represent the observed behavior of the system; and (iii) the effectiveness in reducing parametric uncertainty and enhancing the identifiability of the entire parameter space. The proposed procedure is applied to evaluate the performance of six commonly used techniques for multi-site multi-response parameter estimation and uncertainty analysis of the Soil and Water Assessment Tool (SWAT) in the Eagle Creek Watershed, Indiana, in the Midwestern United States. The objective functions and observed behavior of the system are defined according to the discharge response at the watershed outlet and nitrate responses at four sampling locations along the stream network. Four Bayesian-based approaches including random walk Metropolis-

Hastings, random walk Gibbs sampling, uniform covering by probabilistic rejection, and differential evolution adaptive Metropolis; a greedy optimization method dubbed dynamically dimensioned search; and a global optimization evolutionary method called shuffle complex evolution are evaluated. Results show that the simple structured DDS has dominant performance than all other methods in the speed of convergence, behavior rate and the parametric uncertainty is also significantly reduced. In addition, interesting finding is found and discussed when measurement error and predictive uncertainty are included as a part of consideration.

2.2. Introduction

Watershed models are increasingly embedded in decision-making processes to address a wide range of hydrologic and water quality issues. These models have evolved from lumped to distributed parameters while operating on shorter time steps so that they can utilize input data at finer temporal and spatial resolutions. The incorporation of more parameters and output variables into watershed models has increased model structural complexity (i.e., number of parameters and interactions thereof) [Yang et al., 2008; Bai et al., 2009]. In addition, watershed management programs require the assessment of many hydrologic and water quality responses (e.g., flow, sediment, nutrients, and pesticides) at multiple site locations along the channel network over long periods. The quality, frequency, and number of these observed responses may vary significantly within a watershed. Hence, multi-site, multi-response parameter estimation and uncertainty analysis of complex watershed models with a large number of parameters present a significant challenge [Vrugt et al., 2008; Thyer et al., 2009; Cassidy and Jordan, 2011].

The performance of parameter estimation and uncertainty analysis methods depends greatly upon the sampling strategy for drawing realizations from the high-dimensional parameter space of a given watershed model. A number of computational methods have been proposed to

efficiently and effectively sample the parameter space [Duan et al., 1992; Klepper and Hendrix, 1994; Haario et al., 2006; Tolson and Shoemaker, 2007; Vrugt et al., 2003, 2009]. However, their applicability for simultaneous parameter estimation and uncertainty analysis require a proper appraisal of: (i) the efficiency in finding the optimal (best) parameter solutions at a reasonable number of model evaluations; (ii) the effectiveness in drawing samples that adequately represent the observed behavior of the system at multiple site locations within the watershed system; and (iii) the effectiveness in reducing parametric uncertainty and enhancing the identifiability of the entire parameter space.

A comprehensive examination to explore the parameter domain space in guarantee of finding the global optima always demands the tradeoff of a higher number of required model evaluations. (e.g., Latin hypercube sampling or Sobol sequences). Therefore, it is important to evaluate the computational efficiency of sampling strategies in finding the best (or relatively better) solution within a minimum number of runs. Some methods aim to identify best solution within the number of runs that the users can afford (e.g., dynamically dimensioned search [Tolson and Shoemaker, 2007]) and other approaches are developed to find global optimum with comparatively longer required evaluation period (e.g. shuffle complex evolution [Duan et al., 1992]). As an additional consideration, the objective function(s) utilized may also enhance the computational efficiency of different algorithms [Stedinger et al., 2008]. In previous studies, parameter estimation techniques have mostly been investigated by comparing the global optimum achieved (as measured by a mathematical test function with known global optimum), the efficiency in searching comparably better solutions (e.g. watershed calibration problems with unknown global optimum) and the speed of convergence for discharge or water quality simulations within limited computational effort [Duan et al., 1992; Klepper and Hendrix, 1994;

Tolson and Shoemaker, 2007; Vrugt et al., 2003, 2009]. The assessment of parameter estimation algorithms are made entirely centered on the ability in achieving better objective functions based on the perspective of computational efficiency.

In addition to computational efficiency, the number of solutions that adequately represents observed responses for multiple variables (e.g. stream discharge, sediment and nutrients load) at multiple sites along the channel network is also of keen interest. Systematic quantification established to serve as guidelines (behavior definition) for model evaluation is implemented to identify the effectiveness of parameter estimation techniques [Moriasi et al., 2007]. The effectiveness of different sampling strategies is evaluated by pre-defined thresholds (error statistics) in matching various characterized system behavior. Parameter sets generated by parameter estimation techniques with statistics satisfying behavior definition will be regarded as behavior solutions. Application of behavior definition fortifies the quality of solutions generated by parameter estimation techniques. However, application of behavior definitions in literature is rare, particularly as a component of parameter estimation techniques.

As opposed to the evaluation of efficiency and effectiveness by computational performance and the rate of behavior, a complimentary analysis in present studies is the determination of the identifiability. The focus of sampling strategies has been developed to enhance the reduction of parametric uncertainty of model parameters by constructing marginal posterior distributions of model parameters. In the context of uncertainty analysis, there are several models used to construct marginal distributions of parameters and have associated comparisons [Vrugt et al., 2003; Ajami et al., 2007].

However, the marginal distributions do not represent the interactions between parameters and the overall uncertainty of the joint distribution. Therefore, some studies have used nonlinear

density estimation techniques such as the Tree Structured Density Estimation (TSDE) [Spear et al., 1994] for measuring the joint density of the entire parameter space. Applications of associated techniques are implemented to characterize the complexity of simulation models by using various parameter estimation techniques [Spear et al., 1994; Osidele et al., 2003; Osidele and Beck, 2003].

Previous assessments of parameter estimation techniques have been typically conducted for conceptual rainfall-runoff models with less than 20 parameters [Duan et al., 1992; Klepper and Hendrix, 1994; Marshall et al., 2004; Haario et al., 2006; Tolson and Shoemaker, 2007; Vrugt et al., 2009]. Comprehensive watershed models, such as the Soil and Water Assessment Tool, may contain several hundred parameters. Hence, the number of required model evaluations before convergence to the best parameter solutions may increase significantly due to the well-recognized issues of non-uniqueness (equifinality) and identifiability [Beven and Binley, 1992, Duan et al., 1993; McMillan and Clark, 2009]. Conversely, pragmatic considerations tend to favor algorithms that can find the optimal parameter values within a reasonable runtime and number of model evaluations [Marshall et al., 2004; Vrugt et al., 2009]. Apparently, three major research topics are lacking: First, parameter estimation techniques are not applied on high dimensional (multi-site, multi-variable, various time span) watershed calibration problems. Second, behavior definition is totally being ignored without any application during evaluation processes. And the last, identifiability is not implemented to evaluate the whole domain space of model parameters by applying different sampling strategies.

The overall goal of this paper is to develop and demonstrate a procedure for evaluating the performance of parameter sampling techniques for multi-site multi-response parameter estimation and uncertainty analysis of comprehensive watershed models. Specifically, the

following objectives are defined: (i) To compare the efficiency of the methods in finding better solutions (i.e., compare the objective function values after convergence); (ii) To explore the efficiency of parameter estimation algorithms in finding the best solution within a given number of runs (note that the best solution may not be the same as the global optimal solution); (iii) To measure the effectiveness of the methods in realizing parameter samples that match the system behavior characterized by commonly used error statistics at various locations within the watershed and (iv) To evaluate the identifiability of the overall model parameter space using these methods.

2.3. Methodology

Multi-criteria, multi-site calibration is performed for five test cases under various objective functions in the Eagle Creek Watershed, Indiana. Six parameter estimation algorithms are utilized to optimize objective functions during calibration. During the calibration process, proposed parameter sets are accepted only when model results (e.g. statistics calculated by comparison of simulated outputs with observed data) are within acceptable thresholds (behavior definitions). The behavior rate of all model evaluations is then calculated, and the speed of convergence under each parameter estimation technique assessed. Finally, the identifiability, parameter uncertainty, and predictive uncertainty of the various algorithms are evaluated.

There are four parts in this section; the first part defines objective functions used in this study. The second part describes behavior definition for evaluating marginal probability distributions of model parameters. The third part introduces parameter sampling techniques implemented in this study. The fourth part is the application of tree-structure density estimation (TSDE) to investigate identifiability from joint probability distribution of model outputs.

2.3.1. Objective Functions

In case studies, three objective functions are adopted for model calibration: the root mean square error (RMSE), the Nash-Sutcliffe efficiency coefficient (NS) and the global optimization criterion (GOC). Modifications are made in NS and GOC to incorporate statistics for multi-site cases.

2.3.1.1. OF1 – Root mean square error

Root mean square error (RMSE) is applied only to single-site calibration problems (in this study, to streamflow calibration). The RMSE equation may be written as:

$$RMSE_m = \sqrt{\frac{1}{N_m}\sum_{n=1}^{N_m}[y_{n,sim}^m - y_{n,obs}^m]^2} \quad (1)$$

$$OF_1 = \sum_{m=1}^{M} RMSE_m \quad (2)$$

Where n is the current time step; m is the current number of objective function; N_m is the number of observations for objective function m; $y_{n,obs}^m$ is the observed quantity of objective function m at time step n; $y_{n,sim}^m$ is the simulated quantity of objective function m at time step n; $RMSE_m$ is the root mean square error of objective function m; M is the total number of objective functions; and OF_1 is the sum of all objective functions.

2.3.1.2. OF2 – Nash-Sutcliffe efficiency coefficient

Nash-Sutcliffe efficiency coefficient (NS) has been detailed in previous work [Moriasi et al., 2007]. For calibration problems in this study which involve more than one site, OF2 is

expressed as the sum of the product of NS values and total number of observed data. OF2 may be written as:

$$NS_m = 1 - \frac{\sum_{n=1}^{N_m} [y_{n,obs}^m - y_{n,sim}^m]^2}{\sum_{n=1}^{N_m} [y_{n,obs}^m - \bar{y}_{n,obs}^m]^2} \quad (3)$$

$$OF_2 = \sum_{m=1}^{M} N_m \times NS_m \quad (4)$$

Where, NS_m is the Nash-Sutcliffe efficiency coefficient of objective function m; OF_2 is the sum of all objective functions with corresponding weight based on number of observations.

2.3.1.3. OF3 – Global optimization criterion

The global optimization criterion [van Griensven and Meixner, 2007] is a method for aggregating multi-objective functions. One disadvantage of GOC is that the minimum of the sum of square errors ($SSE_{m,\min}$) is usually unknown. In this study, the watershed model has been well calibrated in advance so it can be assumed that $SSE_{m,\min}$ is an identified constant. The equation of OF3 is written as follows:

$$SSE_m = \sum_{n=1}^{N_m} [y_{n,obs}^m - y_{n,sim}^m]^2 \quad (5)$$

$$SSE_{m,\min} = \sum_{n=1}^{N_m} [y_{n,obs}^m - y_{n,best}^m]^2 \quad (6)$$

$$OF_3 = \sum_{m=1}^{M} \frac{N_m \times SSE_m}{SSE_{m,\min}} \quad (7)$$

Where, SSE_m is the sum of squares of error for objective function m; $SSE_{m,\min}$ is the minimum sum of squares of error for objective function m and OF_3 is the sum of all objective functions.

2.3.2. Definition of Behavior Parameter Sets

During the calibration process, proposed parameter sets are accepted only when model results (e.g. statistics calculated by comparison of simulated outputs with observed data) are within acceptable thresholds (behavior definitions). This study uses the definition of Moriasi and colleagues [2007] to assess behavior of proposed parameter sets (Table 2.1). Although originally developed for calibration at a monthly time step, the general performance ratings are also applied to daily streamflow in this study. As demonstrated in **Error! Reference source not found.**2.1, is the full domain space of all parameters involved in a conceptual 2-D problem. By applying behavior definitions, one can narrow D to a smaller behavior region (BH) with only parameter sets meeting behavior definition (behavior parameter sets) (four groups of subdomain area). In practice, a specific parameter estimation method may not overlap with BH because parameter sets are located outside BH with non-behavioral solutions. In addition, a greater number of behavior parameter sets does not guarantee a method's better overall performance, because a global optimal solution may be located in some other part of the BH [Beck et al., 2002].

Table 2.1 General Performance Ratings

Performance Rating	NSE	PBIAS (%)	
		Streamflow	NOX
Very Good	$0.75 < NSE \leq 1.00$	$PBIAS < \pm 10$	$PBIAS < \pm 25$
Good	$0.65 < NSE \leq 0.75$	$\pm 10 \leq PBIAS < \pm 15$	$\pm 25 \leq PBIAS < \pm 40$
Satisfactory	$0.50 < NSE \leq 0.65$	$\pm 15 \leq PBIAS < \pm 25$	$\pm 40 \leq PBIAS < \pm 70$
Unsatisfactory	$NSE \leq 0.50$	$PBIAS \geq \pm 25$	$PBIAS \geq \pm 70$

NSE: Nash-Sutcliffe efficiency coefficient
PBIAS: Percent bias

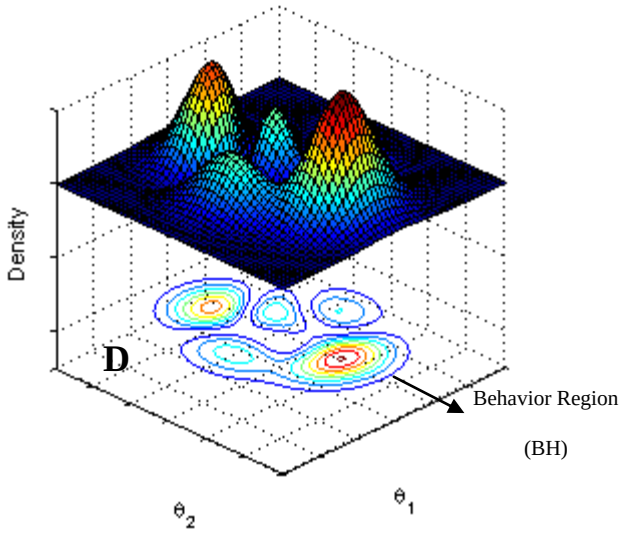

Figure 2.1 Demonstration of behavior parameter sets and the correspond subdomain spaces

2.3.3. Sampling Techniques

The general form Figure 2.2 of a hydrologic model can be written as followed:

$$Y_t^{obs} = Y_t^{sim}(\theta_t, x_t) + \varepsilon_t \tag{8}$$

Where, Y_t^{obs} is the observed hydrologic quantities (e.g. rainfall depth, sediment load, nitrate concentration, etc) at time t; θ_t is the system parameter set at time t (in most cases, θ_t remains the same throughout the simulation process); x_t is the given inputs at time t; $Y_t^{sim}(\theta_t, x_t)$ represents simulated hydrologic quantities with given θ_t and x_t at time t; ε_t is the error term which indicates the difference between observation and simulation results at time t.

Parameter estimation techniques attempt to find the best fit between observed and simulated variables by updating model parameters through minimizing the error term. This study utilizes six parameter search algorithms, including four Bayesian approaches (random walk Metropolis-Hastings algorithm, random walk Gibbs sampling algorithm, uniform covering by probabilistic rejection and differential evolution adaptive Metropolis) incorporated with Markov

chain Monte Carlo sampling (MCMC), and two additional methods, DDS and SCE-UA. The basic theories of MCMC, Bayesian methods, DDS and SCE-UA are briefly introduced here.

2.3.3.1. Bayesian methods and Markov chain Monte Carlo sampling

The Bayes' theorem can be represented in the form as the equation below:

$$p(\theta \mid Y) = \frac{p(\theta)\, p(Y \mid \theta)}{p(Y)} = \frac{p(\theta)\, p(Y \mid \theta)}{\int p(\theta)\, p(Y \mid \theta) d\theta} \tag{9}$$

Where, $p(\theta)$ is the prior distribution of parameter set θ; $p(Y \mid \theta)$ is the likelihood function; $p(Y)$ is the integral of the product of prior and likelihood functions which can be regarded as a constant and $p(\theta \mid Y)$ is the posterior distribution. In many cases, the prior distribution is assumed to be uniformly distributed because of the lack of available information (non-informative prior). In this case, the posterior distribution will be proportional to the likelihood function shown as follows.

$$p(\theta \mid Y) \propto p(Y \mid \theta) \tag{10}$$

The major impediment in implementing the Bayes' theorem is that it is very difficult to take the integral of high dimensional functions. The problem has been solved by the application of the Markov Chain Monte Carlo (MCMC) method where the posterior distribution can be estimated by Markov Chain from a great number of random variables. Broad ranges of hydrologic problems have been solved by MCMC methods recently [Marshall et al., 2004].

2.3.3.2. Random walk Metropolis-Hastings algorithm (RW-MHA)

MHA [Hastings, 1970] is a modified version of Metropolis algorithm [Metropolis et al., 1953]. In MHA, one is able to find the posterior distribution of model parameters by calculating the probability of acceptance. As shown in the equation below, $p(\theta^*)$ and $p(y \mid \theta^*)$ are the prior

distribution and the likelihood function of proposed model parameter θ^*. $p(\theta^k)$ and $p(y \mid \theta^k)$ are the prior distribution and the likelihood function of target model parameter θ^k at iteration step k.

$$\alpha = \min\left\{1, \frac{p(y \mid \theta^*)\, p(\theta^*)}{p(y \mid \theta^k)\, p(\theta^k)}\right\} \tag{11}$$

If model parameters are drawn from the same distributed function (e.g. uniform, normal distribution), the probability of target and proposal prior distribution will be the same. Therefore, the MHA acceptance rate equation can be written as follows:

$$\alpha = \min\left\{1, \frac{p(y \mid \theta^*)}{p(y \mid \theta^k)}\right\} \tag{12}$$

From the equation above, the computation process of MHA can be simplified by comparing the ratio of likelihood functions.

With the assumption of prior distribution, it is still very difficult to formulate a proper likelihood function for high dimensional problems. Therefore, MHA is revised to the random walk Metropolis-Hastings algorithm (RW-MHA) where the proposed model parameter set θ^* will be generated within a range of the current best model parameter set θ^* instead of exploring the whole domain space. In other words, the Markov chain is generated by sampling the candidate parameter set from the proposal distribution $q(\theta^* \mid \theta^{(t)}) = \Delta \pm \theta^{(t)}$ [*Haario et al.*, 1999] with Δ from a symmetric distribution.

2.3.3.3. Random walk Gibbs sampling algorithm (RW-GSA)

GSA [Geman and Geman, 1984] is a special case of MHA which has been adopted in various fields of study [Onibon et al., 2004; Emery, 2007; Kottegoda et al., 2007; Michalak, 2008;]. The major difference between GSA and MHA is that not all parameters are updated simultaneously.

In other words, the original complicated joint distributions have become a conditional distribution with relatively simple structure [Walsh, 2004]. The GSA scheme is as follows:

1. *Start at an initial parameter set*
   $\theta^{(0)} = (\theta_1^{(0)},...,\theta_n^{(0)})$

2. *Updating* $\theta_i^{(j)}$ *by* $\theta_i^{(j+1)}$ *from univariate conditional distributions*
   Generate $\theta_1^{(t+1)}$ *from* $f(\theta_1^{(t+1)} \mid \theta_2^{(t)}, \theta_3^{(t)},..., \theta_n^{(t)})$
   $\vdots$
   Generate $\theta_n^{(t+1)}$ *from* $f(\theta_n^{(t+1)} \mid \theta_1^{(t+1)}, \theta_2^{(t+1)},..., \theta_{n-1}^{(t+1)})$

3. *Repeat step* 2. *by updated parameter set until reaching the stopping criteria*

GSA is also combined with the same random walk chain as that in RW-MHA; this combination is referred to as the random walk Gibbs sampling algorithm (RW-GSA). As mentioned, GSA does not update all parameters simultaneously so the number of updated parameter(s) can be one or more. However, in this study, RW-GSA updates only one parameter at a time. GSA has been widely applied for its greater convergence speed compared to MHA, but it has more chance of becoming trapped in local solutions. Compromises between the two methods will be discussed in the latter section.

2.3.3.4. Uniform covering by probabilistic rejection (UCPR)

The uniform covering by probabilistic rejection (UCPR) algorithm [Klepper and Hendrix, 1994] is a modification of random walk method based on Bayes' theorem. Multiple parameter sets are generated to formulate a subdomain space inside the whole domain space. UCPR is initiated from two major algorithm parameters, the total number of subdomain set and the safety factor, which controls the maximum range of parameter estimation in each model evaluation. After initial parameter sets are uniformly created, the average-nearest-neighbor distance (ANND) is calculated to construct the best fitting subdomain (BFS). The BFS is then expanded by taking the

product of the safety factor and ANND, which formulates the proposal fitting subdomain (PFS). The shape and the size of BFS are varied over model iterations by substituting the current parameter set for the worst parameter set when the trial set performs better. The new generated trial parameter set will be accepted or rejected depending on whether it is located in the final approximate subdomain (FAS).

Ideally, the probability rejection process avoids computationally expensive model runs, but in practice the concept of probability rejection may carry over too much irrelevant information from the previous sampling sets. In other words, UCPR may be expected to have difficulties converging for high dimensional problems. A detailed theoretical description applications of the UCPR can be found in literature [Klepper and Hendrix, 1994; Klepper and Bedaux, 1997; Hendrix and Klepper, 2000; Osidele et al., 2003, Osidele et al., 2006].

2.3.3.5. Differential evolution adaptive Metropolis (DREAM)

The differential evolution adaptive Metropolis (DREAM) [Vrugt et al., 2009] is one of the most cited MCMC applications built upon the framework of Bayes' theorem. DREAM is the extension of DE-MC method [Braak, 2006] with several modifications. First, DREAM samples only selected parameters in each model run instead of all at once. Second, the concept of differential evolution is incorporated from genetic algorithms to generate trial parameter sets. Third, multiple chains are executed concurrently. By combining the rejection process (from Metropolis algorithm) with differential evolution (from genetic algorithm), DREAM is able to solve high-dimensional problems efficiently and outperforms random walk Metropolis (RWM), delayed rejection adaptive Metropolis (DRAM), DE-MC and SCE-UA [Vrugt et al. 2009].

Two additional parameter estimation methods which do not incorporate Bayesian approaches are also examined in this study: shuffle complex evolution and dynamically dimensioned search.

2.3.3.6. Shuffle complex evolution (SCE-UA)

The global optimization method, entitled the Shuffled Complex Evolution (SCE-UA) [Duan et al., 1992] has been extensively applied to the field of water science. SCE-UA is a mutual associated algorithm with concepts from the simplex method [Nelder and Mead, 1965], the controlled random search [Price, 1987] and the competitive evolution [Holland, 1975]. The process of shuffling among complexes minimizes the likelihood of the parameter search becoming trapped in local solutions. SCE-UA has been reviewed and compared with a great number of other optimization techniques [Duan et al., 1992; Duan et al., 1993; Sorooshian et al., 1993; Muttil and Jayawardena, 2008]. Because SCE-UA is widely used, it is included in this assessment as a standard of comparison to Bayesian-associated and other approaches.

In SCE-UA, previous arguments have been made that the system parameters of reflection (alpha) and contraction (beta) should be tuned from default values (alpha = 1.0, beta = 0.5) to the recommended ones (alpha = 0.8, beta = 0.45) [Tolson and Shoemaker, 2008]. Therefore, in this study, two versions of SCE-UA are adopted using both default and recommended values (SCEI and SCEII, respectively) to address the concern of system parameter altering.

2.3.3.7. Dynamically dimensioned search (DDS)

The dynamically dimensioned search (DDS) [Tolson and Shoemaker, 2007] is a recently proposed automatic calibration algorithm for both low and high dimensional problems in watershed simulation models. The main structure of DDS is very similar to GSA. In DDS, the

number of model parameters is updated based on an exponentially decreasing probability function P. Parameter(s) will be updated only when a uniformly distributed random number is lower than P. However, at least one parameter is assigned to be updated when P is very small where all random numbers are greater.

In contrast to SCE-UA, the primary advantage of DDS is that only one algorithm parameter (safety factor) is required in the tuning process; however, this may also be a disadvantage. The new updated parameter value is limited to a 20 % span (the value of safety factor recommended by Tolson and Shoemaker, [2007]) where it may fall into a local optimal when facing a highly nonlinear problem with great number of local solutions. To avoid this, one can generate more initial starting parameter sets and execute more initial model runs, narrow the upper and lower bound of parameters or begin from well calibrated results. This study utilizes the system parameters for DDS outlined by Tolson and Shoemaker [2007].

2.3.4. Identifiability of parameter estimation

The tree-structure density estimation (TSDE) [Spear et al., 1994] is an approach to explore the identifiability of model structure. Model parameters are partitioned into subdomains by high and low densities [Osidele et al., 2003]. As a result, a parameter with higher density is regarded to be more identifiable. The density function (TSDE index) of each subdomain i of behavior parameter sets x can be calculated as:

$$f_i(x) = \frac{n_i}{n}\frac{1}{S_i} \tag{13}$$

Where, n_i is the number of i % of behavior parameter sets; n is the total number of behavior parameter sets; and S_i is the partitioned subdomain of whole domain space S. In this study, TSDE is only applied to behavior parameter sets.

2.4. Case Studies

2.4.1. SWAT2005 Eagle Creek Watershed Calibration

A watershed model the Eagle Creek watershed (ECW), Indiana, is developed with the Soil and Water Assessment Tool (SWAT2005) (Figure 2.2). The ECW (248 km^2 in area) is a part of Upper White River watershed located in Boone, Hamilton, Hendricks and Marion counties. The average annual precipitation is 38 to 40 inches and the average annual temperature is 52°F [*Newman*, 1997]. The majority land use is agricultural activity (agriculture: 59%; rangeland: 38%; forest: 2%; urban: 1%) and the general characters are very close to of the Midwest region. Hydrologic soil type in the ECW can be primarily categorized into two major groups (group B: 51%; group C: 48%). Observation data are available for one discharge (station number 35 in Figure 2.2) and four water quality monitoring gauge stations (station number 20, 22, 27, and 32 in Figure 2.2).

The Soil and Water Assessment Tool (SWAT) is a distributed hydrologic watershed model capable of evaluating the impact of human activities on hydrology and the fate and transport of nonpoint source pollutants such as sediments, nutrients, pesticides and bacteria. A large body of literature has been published on the application of the SWAT model to diverse topics in watershed management [Gassman et al., 2007].

The ECW SWAT2005 model is delineated into 35 subbasins corresponding to 446 hydrologic response units and run for a nine-year simulation period from 1995-2003. The first two years of simulation (1995-1996) are used for model warm-up. The model is then calibrated under various cases using seven years (1997~2003) of daily streamflow observations at one gauge station and instantaneous monthly total nitrate observations at four sites (Figure 2.2).

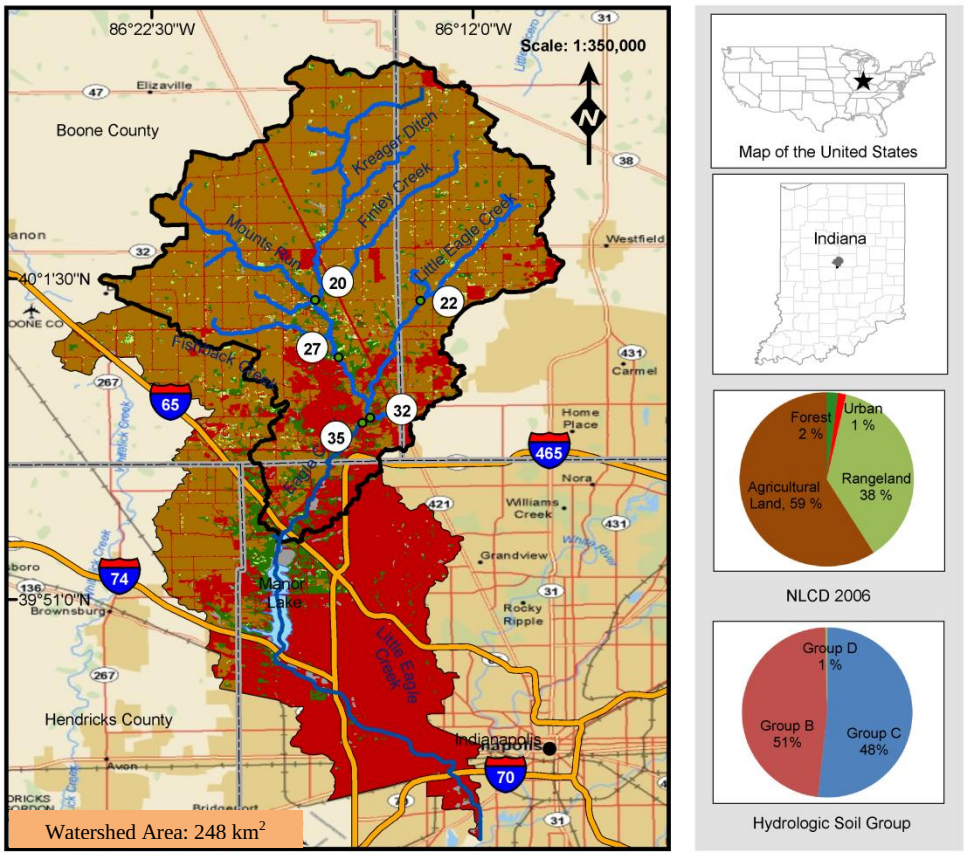

Figure 2.2 Case study area: Eagle Creek Watershed, Indiana

Five calibration cases are implemented in this study: (1) calibration of daily streamflow (30 parameters) by the objective function OF1 (RMSE); (2) calibration of daily streamflow (30 parameters) by the objective function OF2 (NS); (3) calibration of monthly streamflow (30 parameters) and monthly total nitrate (27 parameters) by the objective function OF2 (NS); (4) calibration of monthly streamflow and (30 parameters) monthly total nitrate (27 parameters) by the objective function OF3 (GOC); (5) calibration of daily streamflow (30 parameters) and monthly total nitrate (27 parameters) by the objective function OF3 (GOC). (The DREAM algorithm is only applied to Case 5, the most complicated of the five cases.)

2.4.2. Results of All Approaches

Since the total number of model evaluations affects the performance of DDS, four different runs of DDS (500, 1000, 2500 and 5000 model evaluations, hereafter referred to as DDS500, DDS1000, DDS2500 and DDS5000) are performed for Cases 1-5. All other algorithms are run under 5000 model evaluations. As mentioned in the previous section, one argument in favor of SCE-UA over DDS is that SCE-UA performs better while the default values of the system parameters (reflection parameter alpha and contraction parameter beta) are replaced from [alpha=1.0, beta=0.5] to [alpha=0.8, beta=0.45]. Both sets of system parameters of SCE-UA have been tested with 5000 runs and DREAM is applied to Case 5 with four chains executed simultaneously. All model evaluations have reached termination criteria (termination criteria is defined as no improvement more than 1% on the objective function values for 1000 model evaluations; DREAM is using the termination criteria Gelman-Rubin convergence criteria [Gelman and Rubin, 1992] for the special structure of multiple chains involved) before the preset maximum model evaluations. However, parameter estimation techniques will not stop until it reaches the maximum number model evaluations. The total time of model runs (5000 runs) ranged from 300 to 330 hours (Intel® Xeon® 2.8 GHz CPU, 24 GB Ram, 64-bit operating system, Windows Server 2008 R2 Standard). The results of all cases are described below.

2.4.2.1. Case 1 – Calibration of daily streamflow by OF1

In Case 1, 30 streamflow related parameters are calibrated by objective function OF1. Three groups of algorithm performance are apparent, Figure 2.3(a). The first group comprised of RW-GSA and DDS (5000 simulations) is the best overall in both convergence speed and objective

function achieved. However, RW-GSA converges slower than DDS during the first 1000 runs. The second group is comprised of SCE-UA methods, with no apparent difference between alpha and beta parameters (SCEI and SCEII). The last group contains RW-MHA and UCPR, both of which do not perform well compared to other algorithms, achieved better objective function values within 200 runs when compared to all other methods.

2.4.2.2. Case 2 – Calibration of daily streamflow by OF2

In Case 2, 30 streamflow related parameters are calibrated by objective function OF2. Since the equations of RMSE and NS are structurally similar, the speed of convergence and objective function achieved under various methods show a very similar pattern to Case 1 in Figure 2.3(a). Therefore, a figure of Case 2 is not shown here.

2.4.2.3. Case 3 – Calibration of monthly streamflow and monthly total nitrate by OF2

In Case 3, 30 streamflow related parameters and 27 total nitrate related parameters are calibrated for monthly streamflow and monthly total nitrate by objective function OF2. Results of objective function versus the model runs by different methods are shown in Figure 2.3(b). The relative performance of parameter estimation techniques for Case 3 is similar to Cases 1 and 2. However, in this case, DDS clearly performed the best overall in both convergence speed and objective function achieved. RW-GSA does not show the same performance as in the previous two cases and no substantial improvement in objective function after 400 runs. SCE-UA methods again show similar results for both alpha and beta parameters, with performance comparable to RW-GSA. RW-MHA converged faster than UCPR initially but it appears to be trapped in a local solution after the first 1000 runs. Both methods show similar results at the end of model evaluation.

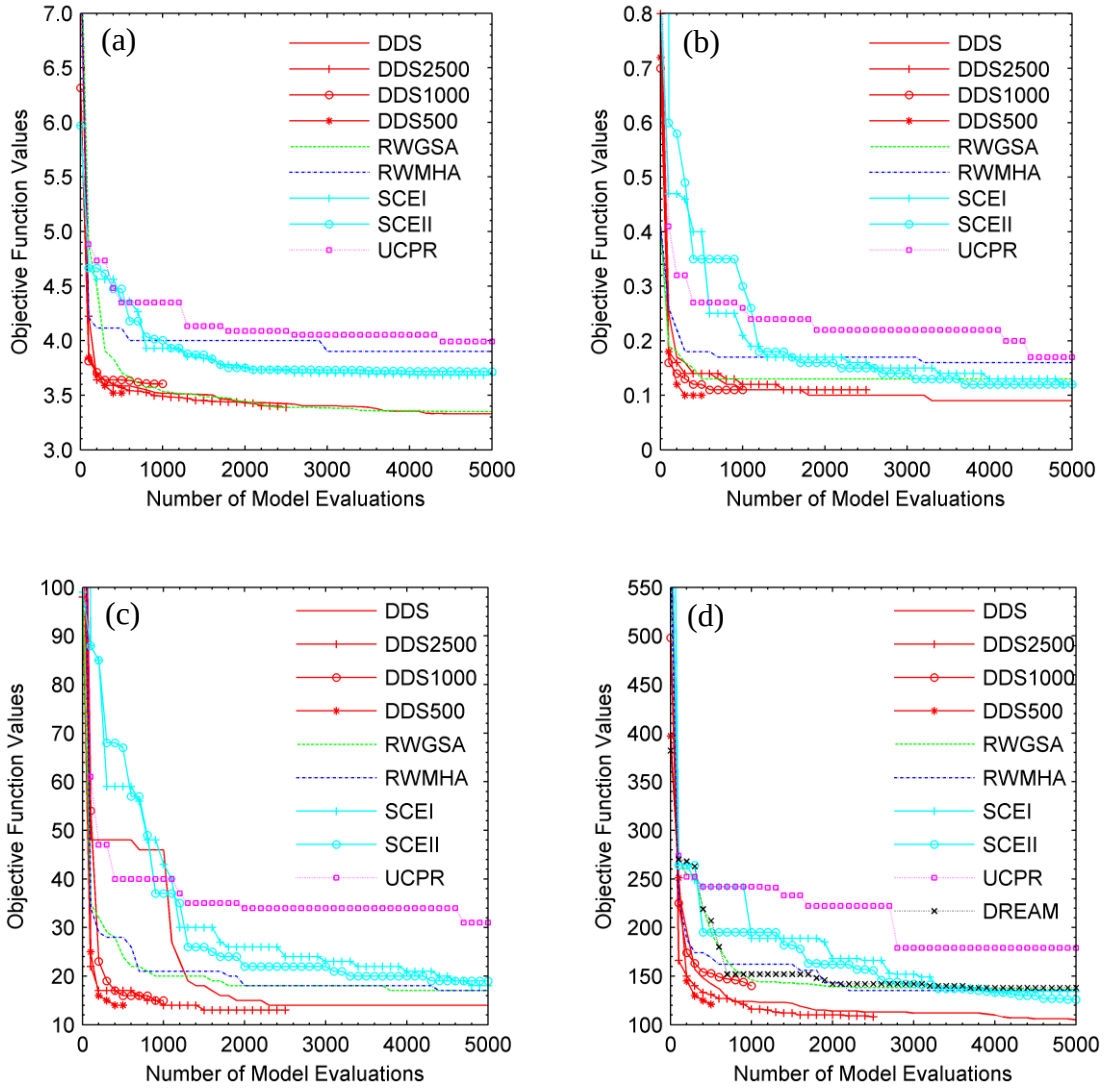

Figure 2.3 Results of ECW calibration: (a) Case 1 - Daily streamflow calibration by OF1 (b) Case 3 - Monthly streamflow and monthly total nitrate calibration by OF2 (c) Case 4 - Monthly streamflow and monthly total nitrate calibration (d) Case 5 - Daily streamflow and monthly total nitrate calibration.

2.4.2.4. Case 4 – Calibration of monthly streamflow and monthly total nitrate by OF3

In Case 4, 30 streamflow related parameters and 27 total nitrate related parameters are adjusted for monthly streamflow and monthly total nitrate calibration by objective function OF3. Results of objective function versus the model runs by different methods are shown in Figure 2.3(c). In

Case 4, DDS achieved far superior objective function values. DDS5000 does not converge as fast as RW-GSA for the first 1200 runs; however, DDS500, DDS1000, DDS2500 converge to favorable objective function values more quickly. All other methods have comparable final objective function values (except for UCPR), but RW-GSA and RW-MHA converge more quickly than SCE-UA.

2.4.2.5. Case 5 – Calibration of daily streamflow and monthly total nitrate by OF3

In Case 5, 30 streamflow related parameters and 27 total nitrate related parameters are calibrated for daily streamflow and monthly total nitrate by objective function OF3; DREAM is included as an alternative parameter estimation method. Results of objective function versus the model runs by different methods are shown in Figure 2.3(d). DDS outperforms all other algorithms in both convergence speed and objective achieved. The difference between the two SCE-UA versions remained small, and UCPR performance is the worst overall.

2.4.3. Behavior Rate of Different Methods and Identifiability for Model Parameters

All proposed parameter sets are classified into behavior and non-behavior sets according to criteria in literature [Moriasi et al., 2007], Table 2.1. Behavior (inclusion) rates in three different categories (very good, good and satisfactory) under the six algorithms are shown in Tables 2.2 and 2.3, and Figures 2.4(a) and 2.4(b), as are identifiability comparisons for DDS5000 and SCEII. Overall results for Cases 1 and 2 are similar, and results for Cases 3 and 4 are similar to Case 5. For brevity, only Cases 1 and 5 are detailed here.

DDS and RW-GSA consistently provide quality calibration results for streamflow (Case 1) (Table 2.2, Figure 2.4(a)). At least 60% (DDS: 61.6%; RW-GSA: 60.1%) of proposed parameter sets have "very good" performance, around 80% (DDS: 78.5%; RW-GSA: 80.2%) for "good" and even 90% (DDS: 88.7%; RW-GSA: 93.7%) for "satisfactory". On the other hand, the behavior rate of SCE-UA II (alpha=0.8, beta=0.45) is less than half of the rate for DDS and RW-GSA in both "very good" (SCE-UA: 23.9%) and "good" categories (SCE-UA II: 38.1%). Performance is only slightly better in the category of "satisfactory" (SCE-UA II: 48%) is barely more than half of those two methods. As mentioned previously in Figure 2.3(a), the third group (RW-MHA and UCPR are included) show the worst performance in the rate of behavior, too. There is no qualified parameter set for either method in "very good", less than 5% in "good" (RW-MHA: 3.4%; UCPR: 1.5%) and less than 30% in "satisfactory" (RW-MHA: 26.2%; UCPR: 11.5%).

From Table 2.3 and Figure 2.4(b), the case of calibrating daily streamflow and monthly total nitrate (Case 5) shows that DDS outperforms all five other methods in categories of "good" and "satisfactory". Also, DDS is the only method with behavior parameter sets in the class of "very good". In the category of "good", DDS has 8.9% more behavior sets than RW-GSA and all other algorithms are near or close to zero (RW-MHA has only 5 behavior parameter sets out of 5000). As for the "satisfactory" performance level, DDS has more than 56% of behavior parameter sets where no other methods have more than 40% (RW-GSA: 37.2%; RW-MHA: 3.7%; SCE-UA II: 0.1%; UCPR: 0%; DREAM: 7.0%).

Table 2.2 Number and percentage of behavior parameter sets in 5000 model evaluations with results from TSDE calculations (Case 1). The highlighted cell indicates that 90% of behavior parameter sets (cumulative number of points = 90%) are located in 7.55% of original sampling domain (Θ = 7.55%) for method DDS in the performance of "Very Good" where the TSDE index (11.93) is calculated by equation (1). TSDE will not be calculated if there is insufficient data.

Case 01	Performance Level	Number of Behavior Sets	% of Behavior Sets	Percentage of Cumulative Points			
				90 (%)		50 (%)	
				θ (%)	TSDE Index	θ (%)	TSDE Index
DDS5000	Very Good	3080	61.6	7.55	11.93	1.6E-05	3.1E+06
	Good	3925	78.5	7.69	11.71	1.6E-05	3.1E+06
	Satisfactory	4435	88.7	8.21	10.96	2.2E-05	2.3E+06
RW-GSA	Very Good	3005	60.1	7.06	12.74	9.2E-06	5.4E+06
	Good	4010	80.2	7.70	11.69	1.8E-05	2.7E+06
	Satisfactory	4685	93.7	7.88	11.42	2.0E-05	2.5E+06
RW-MHA	Very Good	0	-	-	-	-	-
	Good	170	3.4	17.75	5.07	0.51	98
	Satisfactory	1310	26.2	20.23	4.45	0.53	94
SCEII	Very Good	1195	23.9	74.39	1.21	3.01	17
	Good	1905	38.1	75.85	1.19	6.17	8
	Satisfactory	2400	48	82.48	1.09	12.41	4
UCPR	Very Good	0	-	-	-	-	-
	Good	75	1.5	16.18	5.56	0.27	186
	Satisfactory	575	11.5	31.59	2.85	1.00	50

Θ: Percentage of original sampling domain

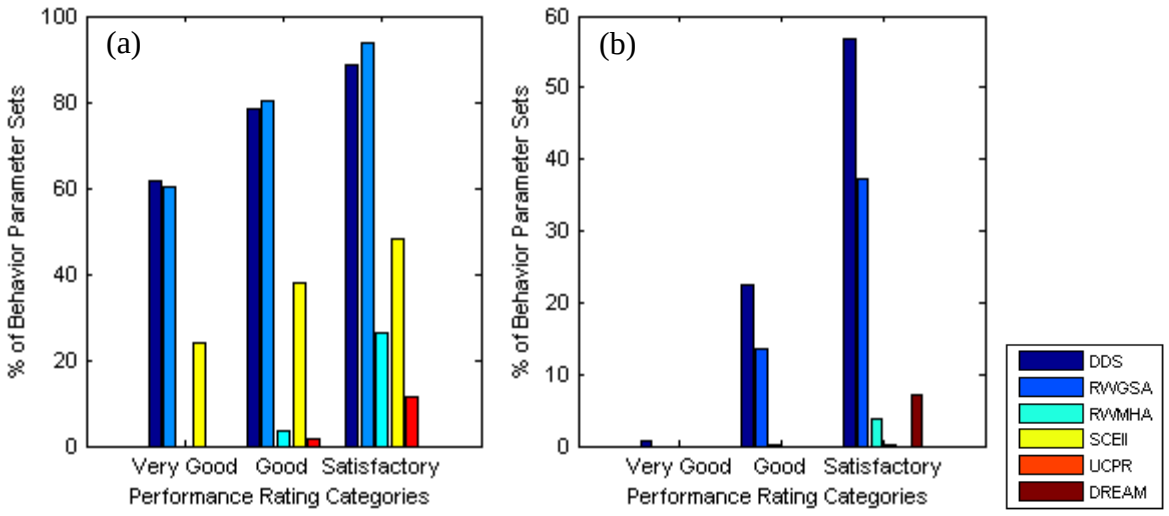

Figure 2.4 Percentage of behavior parameter sets out of 5000 model evaluations: (a) Case 1 - Daily streamflow calibration by OF1 (b) Case 5 - Daily streamflow and monthly total nitrate calibration.

The relationship between behavior parameter sets and identifiability (results from the TSDE calculation) is also of note. In Table 2.2, TSDE results for DDS and RW-GSA show that 90% of behavior parameter sets located in less than 8% of sampling domain (Θ = 7.55%). Conversely, behavior parameter sets of SCE-UA II are sparsely spread in more than 70% of the whole region. TSDE results for RW-MHA and UCPR range from 15 to 30% for all behavioral categories; however, the two methods are the worst in overall performance for all test cases, as described above. The TSDE index clearly illustrates that DDS and RW-GSA are substantially more identifiable than the other three methods. However, it is important to note that the TSDE index describes only the concentration of behavior parameter sets resulting from a given algorithm and does not necessarily guarantee better overall algorithm performance. TSDE indices for the poorly performing RW-MHA and UCPR indicate that most of the behavior parameter sets are trapped in local solutions (Figure 2.3(d)).

Table 2.3 Number and percentage of behavior parameter sets in 5000 model evaluations with results from TSDE calculations (Case 5)

Case 05	Performance Level	Number of Behavior Sets	% of Behavior Sets	Percentage of Cumulative Points			
				90 (%)		50 (%)	
				θ (%)	TSDE Index	θ (%)	TSDE Index
DDS5000	Very Good	30	0.6	-	-	-	-
	Good	1125	22.5	0.71	126.37	4.0E-07	1.2E+08
	Satisfactory	2845	56.9	3.39	26.51	5.9E-07	8.5E+07
RW-GSA	Very Good	0	-	-	-	-	-
	Good	680	13.6	0.89	101.41	< 1.0E-10	> 1.0E10
	Satisfactory	1860	37.2	3.65	24.68	3.6E-08	1.4E+09
RW-MHA	Very Good	0	-	-	-	-	-
	Good	5	0.1	-	-	-	-
	Satisfactory	185	3.7	15.37	5.86	0.32	159
SCEII	Very Good	0	-	-	-	-	-
	Good	0	-	-	-	-	-
	Satisfactory	5	0.1	-	-	-	-
UCPR	Very Good	0	-	-	-	-	-
	Good	0	-	-	-	-	-
	Satisfactory	0	-	-	-	-	-
DREAM	Very Good	0	-	-	-	-	-
	Good	0	-	-	-	-	-
	Satisfactory	350	7.0	1.13	79.71	5.1E-04	9.9E+04

2.4.4. Parameter Uncertainty

Construction of CDF for model parameters yields a better understanding of algorithm convergence patterns. CDFs for five sensitive parameters are illustrated for both Case 1 and Case 5 (Figure 2.5). DDS and RW-GSA converged to narrow ranges for these parameters, instead of searching the entire parameter value range (Figure 2.5(a)). In contrast, CDFs for UCPR are close to a uniform distribution for all five parameters. As mentioned previously, DDS and RW-GSA are the best two methods overall and show comparable results for Case 1 and 2 (Figure 2.3(a)). CDFs yield a likely explanation: In this study, DDS and RW-GSA locate optimal solutions in a narrow parameter range more quickly than other algorithms, speeding up the calibration process. However, this shows a disadvantage for RW-GSA in higher-dimensional calibration problems:

In Case 5, the dimension of the calibration problem is increased from 30 to 57 parameters and RW-GSA is no longer able to compete with DDS as in the lower-dimensional calibration problems of Case 1 and 2, because RW-GSA is trapped in local solutions corresponding to narrow parameter ranges. The convergence of DEP_IMP and DDRAIN shows that CDFs of RW-GSA are apparently skewed to the lower bounds of the parameter values. Meanwhile, from the CDF of SDNCO in Figure 2.5(b), DDS is the only method generating a large number of solutions in the upper range of the parameter value, which is another reason for its dominance over other algorithms.

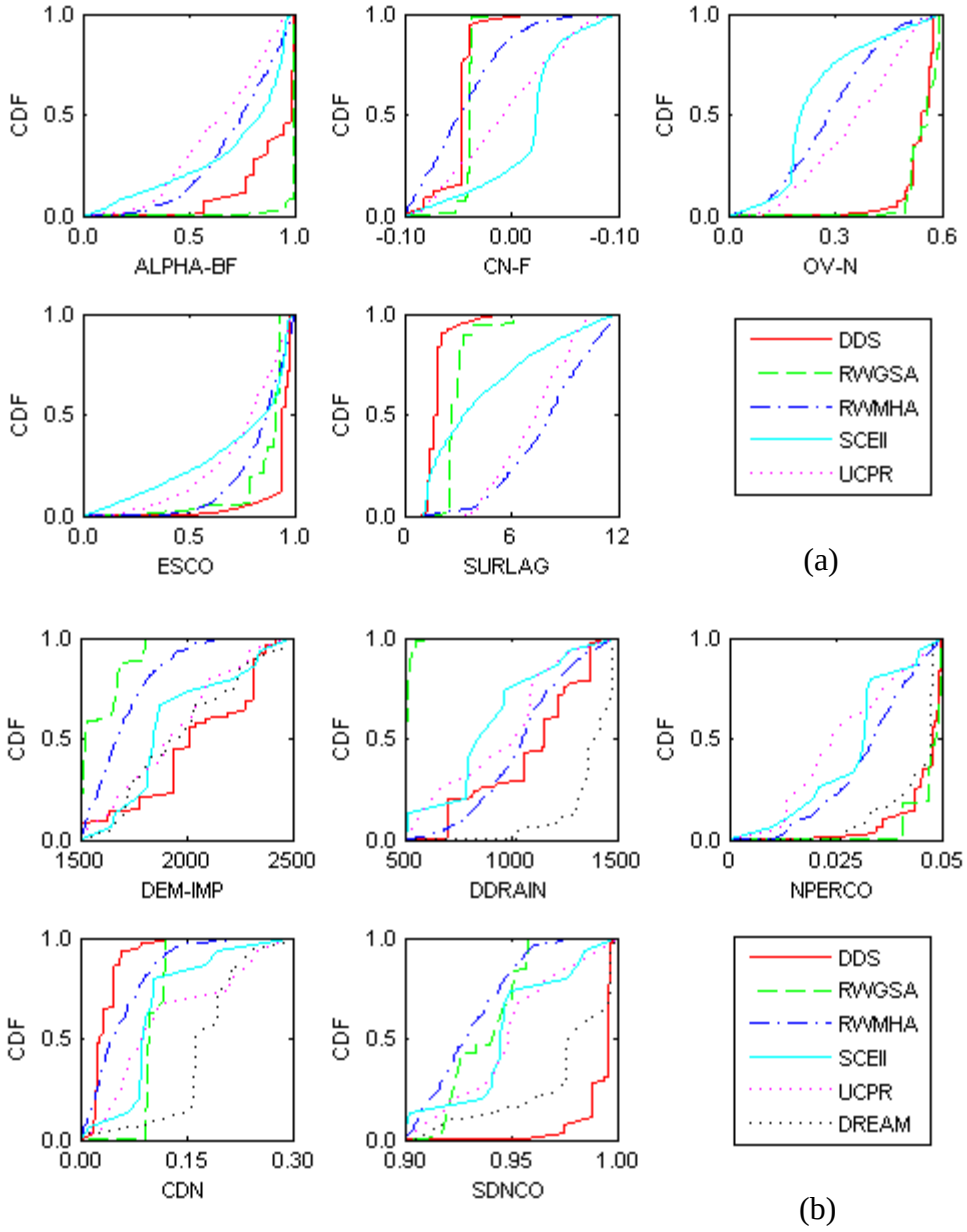

Figure 2.5 CDFs of sensitive parameters: (a) Case 1 - Daily streamflow calibration by OF1 (b) Case 5 - Daily streamflow and monthly total nitrate calibration.

2.5. Discussion

The ability in finding better objective function values and behavior rates are comparable for DDS and RW-GSA in cases which involved a smaller parameter set (30 parameters in Cases 1 and 2). However, RW-GSA appeared to become trapped in local solutions in higher dimensional

problems (Cases 3, 4 and 5). As discussed previously, both DDS and RW-GSA update only one parameter set in each model evaluation and the new proposed parameter set is condition to the current best one. However, the two search algorithms update parameters differently. RW-GSA's poorer performance for high dimensional parameter estimation problems is likely because its search algorithm updates one parameter at a time in a fixed sequence, instead of updating parameter(s) based on an exponentially decreasing equation with random patterns, as in DDS. DDS updates more parameters in the beginning stages of model evaluation, resulting in a more complete search of the whole domain space. In this study, more than 50% of total parameters are updated in the first 6% of total model runs. The number of parameters updated drops exponentially until only one parameter at a time is updated. Additionally, DDS updates parameter(s) in a random pattern, making the algorithm less likely to become trapped in local solutions. In summary, these two modifications explain DDS' superior performance relative to other search algorithms for high dimensional problems. There is no noticeable difference in results under the three distinct objective functions applied in case studies. Cases 1 and 2 yield similar objective function values (the same condition also happens to Case 3, 4, and 5), behavior rates, and convergence patterns; therefore, only Cases 1 and 5 are discussed in detail here.

One can find that the less total runs DDS has, the faster it converges, however, more model runs are required to obtain the best overall solution (objective function value). For example, DDS with 500 total runs converges faster than DDS with 5000. However, DDS with 5000 total runs has a better overall solution. This is because the shorter the total model evaluation, the faster DDS comes to the point of updating only one parameter at a time, but to reach a better final solution, it is necessary for DDS to search the whole domain space in the beginning stages of model evaluation.

Results show that the widely adopted global optimization technique SCE-UA is not comparable to DDS and RW-GSA in all categories, especially in its ability to search for behavior parameter sets. From Table 2.2 and 2.3, the percentage of behavior sets and the TSDE index indicate that most of the computation effort of SCE-UA has been consumed by parameter sets located in irrelevant domain space. As shown in Table 2.4, DDS and RW-GSA attain similar objective function values in less than 6% and 13% of total runs in Cases 1 and 2. In addition, DDS still outperforms SCE-UA in Cases 3, 4 and 5 while RW-GSA begins to be trapped in local solutions. As mentioned previously in literature [Behrangi et al., 2008], SCE-UA surpasses DDS in the cases of test functions with recommended reflection and contraction parameters. However, in the complex SWAT watershed model utilized in this study, SCEI and SCEII are surpassed by DDS. An additional advantage of DDS is that the safety factor remains at the default value with no further tuning necessary throughout the calibration process. There has been some dispute in previous studies whether DDS or SCE-UA performs better [Behrangi et al., 2008; Tolson and Shoemaker, 2008], and the two algorithms have been compared in several different cases, including test functions [Tolson and Shoemaker, 2007]. However, such comparisons may be unnecessary, as the two algorithms are oriented to different goals.

Table 2.4 Percentage of model evaluations reaches the same performance against SCE-UA

Cases	Best Objective Function Value		Percentage of Runs with Compatible Objective Function Value (%) / Method with Better Performance in 5000 Evaluations					
	SCEI	SCEII	DDS vs SCEI	DDS vs SCEII	RW-GSA vs SCEI	RW-GSA vs SCEII	DREAM vs SCEI	DREAM vs SCEII
Case 1	3.68	3.72	3.54 / DDS	3.54 / DDS	11.78 / RW-GSA	9.68 / RW-GSA	**	**
Case 2	0.23	0.23	3.54 / DDS	5.68 / DDS	10.68 / RW-GSA	12.64 / RW-GSA	**	**
Case 3	0.12	0.12	15.78 / DDS	17.66 / DDS	* / SCEI	* / SCEII	**	**
Case 4	17.89	19.01	30.12 / DDS	26.22 / DDS	72.42 / RW-GSA	33.16 / RW-GSA	**	**
Case 5	133.84	125.83	12.84 / DDS	14.28 / DDS	* / SCEI	* / SCEII	* / SCEI	* / SCEII

*: Both SCE-UA I and SCE-UA II perform better than RW-GSA or DREAM
**: Not applied.

In this study, RW-MHA, DREAM and UCPR do not show results comparable to DDS or even RW-GSA. In DREAM, parameter sets are altered by a differential evolution process like that in a genetic algorithm. DREAM behavior rates are better than SCEII and UCPR, but still far behind RW-GSA and DDS, and no significant improvement is derived by DREAM in overall performance (DDS outperforms all other algorithms). Comparisons of results under these three algorithms yield three essential points. First, the way in which an algorithm generates new parameter sets from the posterior distribution can have dramatic impact on model calibration efficiency. In this case, UCPR is obviously adopting much more information than DDS and GSA from the posterior distribution of parameter sets. Second, a complicated evolution process may not enhance the efficacy of a parameter estimation technique when applied to high-dimensional problems. Third, the technique of prior information management for UCPR is not practical for high dimensional problems. To achieve a better performance, one would like to include only sufficient information from existing data instead of including all available data at once. However, in RW-MHA, all parameters are updated in each model evaluation based on the current best solution. From Figure 2.5(a) and 2.5(b), RW-MHA evidently has difficulty making significant convergence during the model evaluation. In other words, RW-MHA cannot integrate all accessible data into a promising candidate parameter set for progressive future generations even after 5000 runs. In UCPR, the proposed parameter sets are selected (or rejected) by a physically based subdomain space. From poor performance in case studies for UCPR, it is clear that the subdomain space constructed by the average nearest neighbor distance (ANND) does not reflect a proper range from which to efficiently generate candidate parameter sets.

An additional consideration when assessing parameter estimation techniques is the predictive uncertainty in resulting model outputs (e.g. streamflow, nutrients). Inclusion rate of

observed data within the uncertainty band constructed by 95% confidence interval (without measurement uncertainty considered) for Case 5 is lower for DDS and RW-GSA than for other algorithms (Tables 2.5, 2.6). This indicates that although these parameter estimation techniques are superior to other algorithms in objective function achieved, behavior rates, and convergence speeds, they may not be implemented for predictive uncertainty. There are two reasons for this: first, the uncertainty band of DDS and RW-GSA is narrower than other algorithms, because these methods are designed to have a better ability to search for a single global solution instead of exploring the whole domain space for possible alternatives. Second, this study does not explicitly evaluate the effects of measurement error, which may have considerable impact on parameter estimation and uncertainty analysis. As shown in Table 2.5 and 2.6, the inclusion rate of observed data within the uncertainty band increases substantially when measurement uncertainty is considered. Clearly, additional research on the interaction of parameter estimation techniques in conjunction with predictive uncertainty is essential. Users should be careful in selecting the parameter estimation algorithm(s) appropriate for specific objectives (i.e. certain methods may be preferable for model calibration versus predictive uncertainty).

Table 2.5 Inclusion rates of observed streamflow data included in 95% confidence interval

Method	Inclusion rate (%) w/o MU considered	w/ MU considered by N %			
		N = 5	N = 10	N = 20	N = 40
DDS5000	25.0	35.7	45.2	67.9	75.0
RW-GSA	26.2	35.7	44.0	65.5	73.8
RW-MHA	54.8	64.3	73.8	83.3	90.5
DREAM	35.7	50.0	54.8	69.0	88.1

MU: Measurement uncertainty
Increase of measurement uncertainty based on the cumulative probable uncertainty represented by probable error range (Harmel et al., 2006)

Table 2.6 Inclusion rates of observed NOX data included in 95% confidence interval

Method	Inclusion rate (%) w/o MU considered	w/ MU considered by N %			
		N = 10	N = 50	N =100	N = 150
DDS5000	33.9	38.9	44.4	58.9	86.7
RW-GSA	33.3	37.8	42.8	56.7	86.7
RW-MHA	51.1	56.1	62.8	71.7	90.0
DREAM	32.8	37.8	44.4	66.1	86.7

2.6. Conclusions

In this study, six parameter estimation algorithms are applied to five case studies of high dimensional watershed model calibration problems. Several conclusions can be made from comparisons of algorithm performance:

(1) The selection of objective functions has no substantial influence on the performance of parameter estimation techniques.

(2) Second, in this study, the DDS algorithm is overall the best performing parameter estimation technique in terms of objective function value achieved, behavior rate, and convergence speed. The superior performance of DDS can be attributed to three characteristics of the algorithm: First, DDS uses a simple approach in which only a few parameters are updated at a time in proposed new parameter sets. Therefore, DDS needs to process only a small amount of the large quantity of information available from previous realizations, unlike other algorithms such as RW-MHA. Second, by the application of TSDE, it is apparent that DDS narrows the parameter space searched to a small region, which enhances convergence speed. Finally, DDS updates parameters in a random sequence. This improves the algorithms' ability to achieve better solutions for high dimensional calibration problems, and minimizes the risk of the search algorithm becoming trapped in a local solution. While RW-GSA performance is similar to that of DDS for lower-dimensional calibration problems, this algorithm may become trapped in

local solutions when the dimension of the problem increases (in this study, from 30 to 57 parameters).

(3) An important third conclusion is that great distinctions in efficiency among different Bayesian approaches reveal the importance of managing available information derived during calibration process for candidate populations. No significant improvement will be made if the parameter estimation technique is confounded by too much information provided from the posterior distribution, and Bayesian approaches should be adapted accordingly.

(4) Although it is widely used, SCE-UA may not be the best choice in parameter estimation techniques when dealing with high dimensional watershed calibration problems. SCE-UA performance is poorer than DDS in all five cases and RW-GSA (Case 1 and 2) in almost all categories. Even RW-MHA yields a higher behavior rate than SCE-UA.

(5) Uncertainty from measurement error plays an important role in predictive uncertainty. Low inclusion rate of DDS and RW-GSA within the uncertainty band show that an efficient and effective parameter estimation technique may not be the best option for predictive uncertainty investigations.

The proposed framework provides a general platform for future studies in evaluating parameter estimation techniques in a rigorous and systematic fashion. Various sampling strategies can be evaluated by not only the ability in achieving optimality of objective function but also can have system behavior considered in terms of implementing trustworthy statistical standards. In addition, sampling patterns of each parameter estimation technique during calibration process can be comprehended by the application of TSDE. Further modifications of

new/exist sampling strategies can be developed/refined based on this supplementary information from the character of parameter sampling.

References

Bai, Y., T. Wagener and P. Reed (2009), A top-down framework for watershed model evaluation and selection under uncertainty, Environmental Modelling & Software, 24(8), 901-916.

Bates, B. C. and E. P. Campbell (2001), A Markov Chain Monte Carlo Scheme for parameter estimation and inference in conceptual rainfall-runoff modeling, Water Resour. Res., 37(4), 937-947.

Beck, M.B., B. D. Fath, A. K. Parker, O. O. Osidele, G. M. Cowie, T. C. Rasmussen, B. C. Patten, B. G. Norton, A. Steinemann, S. R. Borrett, D. Cox, M. C. Mayhew, X. Q. Zeng and W. Zeng (2002), Developing a concept of adaptive community learning: Case study of a rapidly urbanizing watershed, Integrated Assessment, 3(4), 299-307.

Behrangi, A., B. Khakbaz, J. A. Vrugt, Q. Duan and S. Sorooshian (2008), Comment on ‘‘Dynamically dimensioned search algorithm for computationally efficient watershed model calibration’’ by Bryan A. Tolson and Christine A. Shoemaker, Water Resour. Res., 44, W12603, doi: 10.1029/2007WR006429.

Beven, K. and A. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, Hydrological Processes, 6(3), 279-298.

Cassidy, R., P. Jordan (2011), Limitations of instantaneous water quality sampling in surface-water catchments: Comparison with near-continuous phosphorus time-series data, Journal of Hydrology, 405, 182-193.

Douglas-Mankin, K. R., D. Maski, K. A. Janssen, P. Tuppad, G. M. Pierzynski (2010), Modeling Nutrient runoff yields from combined in-field crop management practices using SWAT, Transactions Of The ASABE, 53(5), 1557-1568.

Duan, Q., S. Sorooshian and V. K. Gupta (1992), Effective and efficient global optimization for conceptual rainfall-runoff models, Water Resour. Res., 28, 1015– 1031.

Duan, Q. Y., V. K. Gupta, S. Sorooshian (1993), Shuffled Complex Evolution Approach for Effective and Efficient Global Minimization, Journal of Optimization Theory and Applications, 76(3), 501-521.

Duan, Q., V. K. Gupta and S. Sorooshian (1993), Shuffled complex evolution approach for effective and efficient global minimization, J. Optim. Theory Appl., 76, 501–521.

Emery, X. (2007), Using the Gibbs sampler for conditional simulation of Gaussian-based random fields, Comput. Geosci., 33(4):522–537.

Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. Statistical Science, 7(4), 457-472. doi:10.1214/ss/1177011136

Geman, S., D. Geman (1984), Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, IEEE Transactions on Pattern Analysis and Machine Intelligence, 6(6), 721-741.

Haario, H, E. Saksman and J. Tamminen (1999), Adaptive proposal distribution for random walk Metropolis algorithm, Computational Statistics, 14(3), 375-395.

Haario, H, M. Laine, E. Saksman and J. Tamminen (2001), An adaptive Metropolis algorithm, Bernoulli, 7(2), 223-242.

Haario, H, M. Laine, A. Mira and E. Saksman (2006), DRAM: Efficient adaptive MCMC, Stat. Comput., 16, 339-356.

Hastings, W. K. (1970), Monte Carlo sampling methods using Markov chains and their applications, Biometrika, 57(1), 97-109.

Hendrix, E. M. T. and O. Klepper (2000), On uniform covering, adaptive random search and raspberries, J. Global Optim., 18, 143-163.

Holland, J. H. (1975), Adaptation in natural and artificial systems, The University of Michigan Press, Ann Arbor.

Klepper, O. and E. M. T. Hendrix (1994), A method for robust calibration of ecological models under different types of uncertainty, Ecological Modelling, 74, 161-182.

Klepper, O. and J. J. M. Bedaux (1997), A robust method for nonlinear parameter estimation illustrated on a toxicological model, Nonlinear Analysis: Theory, Methods & Applications, 30(3), 1677-1686.

Kottegoda, N.T., L. Natale, E. Raiteri (2007), Gibbs sampling of climatic trends and periodicities, J. Hydrol., 339, 54-64.

Marshall, L., D. Nott and A. Sharma (2004), A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modeling, Water Resour. Res., 40, W02501, doi: 10.1029/2003WR002378.

McMillan, H., M. Clark (2009), Rainfall- runoff model calibration using informal likelihood measures within a Markov chain Monte Carlo sampling scheme, Water Resour. Res., 45, W04418, doi: 10.1029/208WR007288.

Meng, H., A. M. Sexton, M. C. Maddox, A. Sood, C. W. Brown, R. R. Ferraro, R. Murtugudde (2010), Modelling Rappahannock river basin using SWAT – Pilot for Chesapeake Bay watershed, Applied Engineering in Agriculture, 26(5), 795-806.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller (1953), Equations of state calculations by fast computing machines, Journal of Chemical Physics, 21(6), 1087-1092.

Michalak, A. M. (2008), A Gibbs sampler for inequality-constrained geostatistical interpolation and inverse modeling, Water Resour. Res., 44, W09437, doi: 10.1029/2007WR006645.

Moriasi, D. N., J. G. Arnold, M. W. V. Liew, R. L. Bingner, R. D. Harmel and T. L. Veith (2007), Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, American Society of Agricultural and Biological Engineers, 50(3), 885-900.

Muttil, N. and A. W. Jayawardena (2008), Shuffled Complex Evolution model calibrating algorithm: enhancing its robustness and efficiency, Hydrol. Process., 22, 4628–4638.

Nelder, J. A., R. Mead (1965), A simplex method for function minimization, Comput. J., 7(4), 308-313.

Newman, J. E. (1997), Our Changing Climate, In: The Natural Heritage of Indiana, Edited M. T. Jackson.

Onibon, H., T. Lebel, A. Afouda and G. Guillot (2004), Gibbs sampling for conditional spatial disaggregation of rain fields, Water Resour. Res., 40, W08401, doi: 10.1029/2003WR002009.

Osidele, O. O., W. Zeng and M. B. Beck (2003), Coping with uncertainty: A case study in sediment transport and nutrient load analysis, J. Water Resour. Plann. Manage., 129(4), 345-355.

Osidele, O. O., W. Zeng and M. B. Beck (2006), A random search methodology for examining parametric uncertainty in water quality models, Wat. Sci. Tech., 53(1), 33-40.

Price, W. L. (1987), Global optimization algorithms for a CAD workstation, J. Optim. Theory Appl., 55(1), 133-146.

Sorooshian, S., Q. Duan and V. K. Gupta (1993), Calibration of rainfall-runoff models: Application of global optimization to the Sacramento Soil Moisture Accounting Model, Water Resour. Res., 29(4), 1185-1194.

Spear, R. C., T. M. Grieb and N. Shang (1994), Parameter uncertainty and interaction complex environmental models, Water Resour. Res., 30(11), 3159-3169.

Stedinger, J. R., R. M. Vogel, S. U. Lee and R. Batchelder (2008), Appraisal of the generalized likelihood uncertainty estimation (GLUE) method, Water Resour. Res., 44, W00B06, doi: 10.1029/2008WR006822.

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.

Tolson, B. A. and C. A. Shoemaker (2008), Reply to comment on "Dynamically dimensioned search algorithm for computationally efficient watershed model calibration" by Ali Behrangi et al., Water Resour. Res., 44, W12604, doi: 10.1029/2008WR006862.

Thyer, M., B. Renard, D. Kavetski, G. Kuczera, S. W. Franks and S. Srikanthan (2009), Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis, Water Resour. Res., 45, W00B14, doi: 10.1029/2008WR006825.

Van Griensven, A. and T. Meixner (2007), A global and efficient Multi-objective auto-calibration and uncertainty estimation method for water quality catchment models, J. Hydroinf., 9(4), 277–291.

Vrugt, J. A., H. V. Gupta, W. Bouten, and S. Sorooshian (2003), A shuffled complex evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, Water Resour. Res., 39(8), 1201, doi: 10.1029/2002WR001642.

Vrugt, J. A., C. J. F. Braak, M. P. Clark, J. M. Hyman and B. A. Robinson (2008), Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation, Water Resour. Res., 44, W00B09, doi: 10.1029/2007WR006720.

Vrugt, J. A., C. J. F. Braak, C. G. H. Diks, B. A. Robinson, J. M. Hyman, D. Higdon (2009), Accelerating Markov Chain Monte Carlo Simulation by Differential Evolution with Self-Adaptive Randomized Subspace Sampling, Int. J. Nonlinear Sci. Numer. Simul, 10(3), 271-288.

Walsh, B. (2004), Markov chain Monte Carlo and Gibbs sampling, Lecture notes for EEB 581, version 26.

Yang, J., P. Reichert, K. C. Abbaspour, J. Xia, H. Yang (2008), Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China, Journal of Hydrology, 358, 1-23.

Chapter 3

Evaluation of Improved Parameter Estimation and Uncertainty Analysis of Watershed Models: The Role of General Information of Watershed Processes

3.1. Overview

Computational efficiency in parameter estimation has been extensively improved by modifying optimization procedures. For complex watershed simulation model with large number of system parameters, comparable statistical results can be derived from a variety of candidate parameter sets by calculating the errors between simulation and observation outputs in time series. Meanwhile, model outputs can also be summative values instead of time varying hydrologic/nutrients responses such as denitrification and N-nitrate yields from surface or subsurface flow. Statistical best fit in sequential responses may violate summative outputs because the goal of optimization process does not have any information from summative quantities considered simultaneously. In other words, watershed calibration without reflecting may have results of parameter sets converging to domains which generate physically meaningless outputs with good statistics. In this study, the intra-watershed responses are applied on calibration problems of Eagle Creek watershed, Indiana. The results of watershed calibration show noteworthy improvement on overall performance in achieving global optimality, increasing

number of behavioral solutions and the quality of predictive outputs is significantly enhanced by implementing intra-watershed responses in terms of additional constraints. In summary, intra-watershed responses constraints present great potential benefits in enhancing the quality of watershed calibration, reducing parameter and predictive uncertainty, and also provides a novel direction for watershed modeling in the future.

3.2. Introduction

Modeling of large scale hydrologic systems entails simulation of complex interactions among various physical processes. To imitate real world phenomenon, sophisticated simulation models usually include a large number of parameter sets [*Vrugt et al.*, 2008; *Yang et al.*, 2008]. As manual calibration are not adequate for many-parameter, high dimensional watershed calibration problems, a number of automatic calibration methods had been proposed to solve challenging topics [*Duan et al.*, 1992; *Haario et al.*, 2006; *Klepper and Hendrix*, 1994; *Tolson and Shoemaker*, 2007; *Vrugt et al.*, 2003, 2009a].

In recent years, large efforts have been made to estimate model parameter values and to enhance computational efficiency of parameter sampling by optimization procedures. The goal of parameter estimation is to optimize the objective(s) constructed by particular mathematical equations (likelihood functions) through parameter searching processes [*Gupta et al.*, 1999; *Liew et al.*, 2007; *Marshall et al.*, 2004]. Specifically, simulated results (e.g. flow discharge, nutrient flux) are compared with corresponding observed data to calculate error statistics such as root mean square error and Nash-Sutcliffe efficiency coefficient [*Harmel et al.*, 2006]. Candidate (trial) parameter sets with better statistics indicate that the simulated outputs have a better fit with observed data. In other words, the ultimate intention of optimizing likelihood function(s) is

to search for an ideal candidate parameter set gives sufficiently accurate simulated outputs as observed quantities [*Refsgaard*, 1997].

For a complex watershed model, outputs can also be non-targeted system responses (e.g. aggregated system outputs: denitrification, total nitrate yield… etc) instead of time varying hydrologic/nutrients serial quantities (e.g. stream discharge process). Statistical best fit in sequential responses may have particular model output(s) (especially non-targeted responses) violated at the same time. As a result, hydrologic models controlled by model parameters cannot fully reflect real world phenomena because none of the existing models can perfectly simulate hydrologic processes [*Ebel and Loague*, 2006; *Sorooshian and Gupta*, 1983]. And, different parameter sets may yield the same acceptable match (it can also be statistically coherent) for observed quantities as known the issue of non-uniqueness [*Sorooshian and Gupta*, 1983]. Therefore, the watershed model gives statistically best results (e.g. flows, nutrient loads) using parameters that yield unrealistic intra-watershed responses is providing the right answer for the wrong reasons. In addition to improve present available hydrological models or develop new schemes for parameter estimation, one should also consider other perspectives as complementary supporting information.

Previous studies have mostly focused on tuning/adjusting model parameters without considering intra-watershed responses in nature [*Amatya and Jha*, 2011; *Cassidy and Jordan*, 2011; *Green et al.*, 2006]. The impact of this approach on watershed calibration problems has not been addressed adequately. Not only should the parameter values be scrutinized for credibility, but other system responses besides the standard output variables should be a focus during the calibration procedure. The lack of stressing non-targeted system responses in field could lead the

results of calibration away from reality [*Sui and Frankenberger*, 2008]. Consequently, it is essential to have important watershed characteristics properly constrained during calibration.

In this study, the watershed calibration process is performed by taking account not only the time varying quantities (e.g. discharge, total nitrate load), but also other measureable intra-watershed characteristics such as denitrification and the ratio of N-nitrate losses yield from subsurface and surface flow (lower and upper limits are implemented in terms of additional constraints). The overall goal of this paper is to evaluate the role of actual watershed behavior measures in reduction of parameter and predictive uncertainties. Specifically, the following objectives are defined: (i) To examine how incorporation of intra-watershed responses enhances parameter estimation and calibration of the watershed model; and (ii) To investigate how incorporation of intra-watershed responses reduces parameter and predictive uncertainties.

3.3. Methods and Materials

3.3.1. Site Location and Characteristics

Case studies are taken place at the Eagle Creek watershed (ECW), Indiana (Figure 3.1). The ECW (248 km^2 in area) is a part of Upper White River watershed and is located central Indiana within four counties (Boone, Hamilton, Hendricks and Marion). The average annual precipitation is 38 to 40 inches and the average annual temperature is 52°F [*Newman*, 1997]. The majority of land use is agriculture (agriculture: 59%; rangeland: 38%; forest: 2%; urban: 1%). Hydrologic soil type in the ECW can be primarily categorized into two major groups (group B: 51%; group C: 48%). Observed data used in this study are from one stream discharge station (station number 35 in Figure 3.1) and four water quality monitoring gauge stations (station number 20, 22, 27, 32 in Figure 3.1).

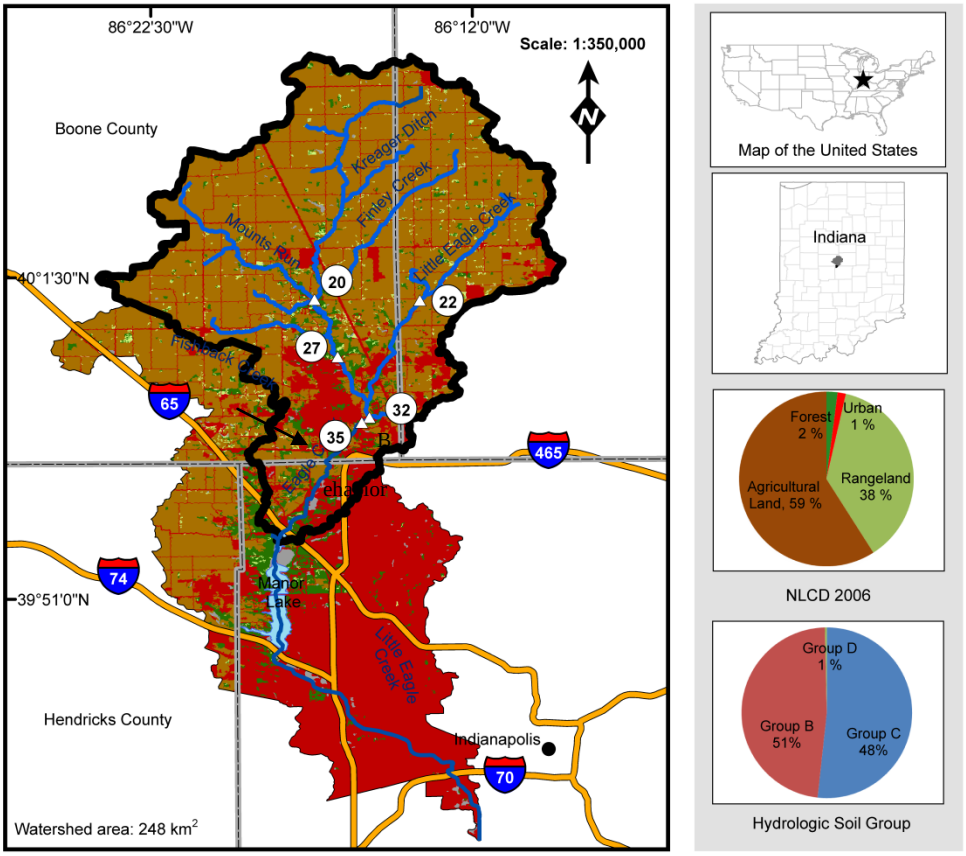

Figure 3.1 Case study area: Eagle Creek Watershed, Indiana

3.3.2. SWAT Model

The Soil and Water Assessment Tool (SWAT) [*Arnold et al.*, 1993] is a continuous-time and distributed parameter model able to simulate/predict hydrologic and water quality processes at a large-scale and watershed level. A large body of research exists on SWAT mode applications for watershed management [*Arnold et al.*, 2010; *Chiang et al.*, 2010; *Douglas-Mankin et al.*, 2010; *Du et al.*, 2005; *Ghebremichael et al.*, 2010; *Green et al.*, 2006; *Jayakrishnan et al.*, 2005; *Kim et al.*, 2010; *Meng et al.*, 2010; *Moriasi et al.*, 2009; *Srinivasan, R. X. Zhang*, 2010]. In this study, the latest version of the SWAT model (SWAT2009) is adopted as the watershed simulation tool for the evaluation of the importance of intra-watershed responses. A detailed description of the SWAT model is presented in literature [*Arnold et al.*, 1998].

3.3.3. Constraints of Intra-watershed Responses

Intra-watershed responses represent a general character of system response in terms of summative quantities. A watershed model calibrated without considering intra-watershed responses may have very good results matching simulated nutrient outputs with historical data, but incorrectly representing watershed processes (e.g. nitrate-N losses in the Midwest region should be contributed mostly by subsurface flow instead of surface runoff).

In this study, two constraints for intra-watershed responses are implemented during the calibration process. Outputs of calibrated models that violate the assigned ranges of intra-watershed responses will be automatically penalized/rejected to satisfy the actual watershed behavior in practice.

3.3.3.1. Denitrification

Denitrification is the process of nitrate reduction caused by bacterial action. Denitrification will occur when the water-filled porosity of soil is greater than 60 percent [*Aulakh et al.*, 1991; *Neitsch et al.*, 2011]. In the SWAT model, denitrification is affected by temperature, soil water content, and presence of nitrate and carbon which can be controlled by three model parameters (SOL_CBN, CDN and SDNCO) [*Neitsch et al.*, 2011]. In the Midwest region, annual denitrification is regularly no more than 50 kg/ha [*David et al.*, 2009]. Therefore, the range of denitrification is constrained from 0 to 50 kg/ha in this study. The objective function will be penalized if the simulated denitrification falls outside this range.

3.3.3.2. Ratio of nitrate-N losses contributed from subsurface and surface flow

Subsurface drainage systems are extensively applied in the Midwest region, USA [*Kalita et al.*, 2006]. Drainage systems can effectively reduce nitrate-N losses from surface runoff in this area

[*Gilliam et al.*, 1999]. Measured data of nitrate-N losses from various sources (surface runoff, lateral flow, tile drainage and groundwater) are not available in previous works for the case study watershed in this script [*Sui and Frankenberger*, 2008]. However, an assumption can be made that the majority of nitrate-N losses come from subsurface flow in the Midwest since only small amount of nitrate-N is contributed to streams by overland flow [*Jaynes et al.*, 1999]. In addition, long-term record for nitrate losses in streamflow (surface flow) and baseflow (subsurface flow) can be found in the nearby watershed (Raccoon River watershed, Iowa) [*Schilling and Zhang*, 2004] which indicates that nitrate losses contributed by baseflow is approximately 66.7% of total nitrate load in conclusion. In this study, the constraint of nitrate-N losses contributed from subsurface flow is defined to be no less than 60 percent of total nitrate-N losses (nitrate-N losses contributed from both subsurface and surface flow). The objective function will be penalized if the ratio of N-nitrate losses contributed from SSQ (SSQ Ratio) is lower than 60 percent of total N-nitrate losses.

3.3.4. Parameter Estimation Technique

Selection of a parameter sampling technique is very important for parameter estimation especially in many-parameter, high dimensional problems. The nature of high nonlinearity may cause difficulties (e.g. computational expensive, insufficient amount of behavior solutions) which make the calibration work hard to have significant progress. Well calibrated watershed with good statistical results by a parameter estimation algorithm may not be able to provide sufficient behavior solutions. A systematic procedure for evaluating advantages and disadvantages of sampling strategies can be found in literature [*Yen et al.*, 2012] which will not be discussed in this script.

3.3.4.1. Dynamically dimensioned search

Dynamically dimensioned search (DDS) [*Tolson and Shoemaker*, 2007] is an automatic calibration algorithm designed for the purpose of solving high dimensional problems. It has been proven to have outstanding performance compares to other commonly used parameter estimation techniques [*Yen et al.*, 2012] in different categories. In this study, DDS is adopted as the sampling technique to explore the role of intra-watershed responses in watershed simulation models.

3.3.5. Evaluation Statistics

The selection of a likelihood function is very important since the distribution of parameter sampling of model errors will not be convincingly reflected if it is not statistically sound [*Stedinger et al.*, 2008]. In case studies, the statistical valid likelihood function [*Ahmadi et al.*, 2012] is adopted as the objective function and the brief description is as follows.

3.3.5.1. Likelihood function

Let's consider a watershed model M with a vector of p parameters ($\boldsymbol{\theta}$) within the feasible parameter space ($\boldsymbol{\Theta}$) that simulates the response vector of the watershed ($\hat{y}$) as follows:

$$\hat{y} = M(\boldsymbol{\theta}) \quad \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \Re^n \tag{1}$$

The discrete stochastic time-series vector of model residuals is:

$$\varepsilon(\boldsymbol{\theta}) = \hat{y} - y = M(\boldsymbol{\theta}) - y \tag{2}$$

Where y is the vector of observed output response. The goal of calibration procedures is to estimate $\boldsymbol{\theta}$ such that the residuals are as close to zero as possible. Joint posterior probability distribution of the parameters conditioned on the observed response can be expressed using Bayesian statistics as [*Box and Tiao*, 1992]:

$$P(\boldsymbol{\theta}|y) = c \cdot P(\boldsymbol{\theta}) \cdot \ell(\varepsilon|\boldsymbol{\theta}) \tag{3}$$

Where c is a normalization constant, ℓ is the likelihood function and represents the likelihood of producing model residuals (ε) for a given set of model parameters ($\boldsymbol{\theta}$), while $P(\boldsymbol{\theta})$ denotes the prior probability density function of $\boldsymbol{\theta}$ that is assigned before assimilation of any observed data. Assuming that residuals are normally and independently distributed (NID) with mean equal to zero and unknown but constant standard deviation σ_e, the likelihood function ℓ will then take the following form [*Box and Tiao*, 1992]:

$$\ell(\varepsilon|\boldsymbol{\theta}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left[-\frac{(\hat{y}_i(\boldsymbol{\theta})-y_i)^2}{2\sigma_e^2}\right] \tag{4}$$

Where n is the number of observations. Since in most cases the errors in hydrological and water quality modeling are not normally distributed, independent, and homoscedastic, suitable transformations must be applied to account for error characteristics that are not consistent with assumptions made for deriving Eq. (4). The first-order autoregressive (AR-1) transformation of the residuals can be used to account for correlated errors:

$$\varepsilon_i = \rho\varepsilon_{i-1} + \vartheta_i \qquad i = 1, \dots, n \tag{5}$$

Where ρ is the lag-1 serial correlation coefficient for the residuals ε, and $\vartheta \sim N\left(0\,, \sigma_\vartheta^2\right)$ is the innovation term with zero mean and constant variance σ_ϑ^2. Sorooshian and Darcup [*Sorooshian and Dracup*, 1980] showed that substituting the AR-1 transformation into the likelihood function of (4) after taking the logarithm (log-likelihood function) yields:

$$\begin{aligned}\ell^*(\varepsilon|\boldsymbol{\theta}) = &-\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln\frac{\sigma_\vartheta^{2n}}{1-\rho^2} - \frac{1}{2}(1-\rho^2).\sigma_\vartheta^{-2}[\hat{y}_1(\boldsymbol{\theta}) - y_1]^2 - \\ &\frac{1}{2}\sigma_\vartheta^{-2}.\sum_{i=2}^{n}\{(y_i - \rho y_{i-1}) - [\hat{y}_i(\boldsymbol{\theta}) - \rho\hat{y}_{i-1}(\boldsymbol{\theta})]\}^2\end{aligned} \tag{6}$$

The terms σ_ϑ^2 and ρ can be estimated using the Bayesian approach [*Vrugt et al.*, 2009b] or can be assigned based on prior knowledge. A proper likelihood function for multiple variables

similar to the single output case is as follows:

$$\ell^*(\varepsilon|\boldsymbol{\theta}) = \sum_{j=1}^{m}\left\{-\frac{n_j}{2}\ln(2\pi) - \frac{1}{2}\ln\frac{\sigma_{\vartheta,j}^{2n_j}}{1-\rho_j^2} - \frac{1}{2}\left(1-\rho_j^2\right)\times\sigma_{\vartheta,j}^{-2}\right.$$
$$\times\left[\hat{y}_{1,j}(\boldsymbol{\theta}) - y_{1,j}\right]^2 - \frac{1}{2}\sigma_{\vartheta,j}^{-2} \qquad (7)$$
$$\left.\times\sum_{i=2}^{n_j}\left\{\left(y_{i,j}-\rho_j y_{i-1,j}\right) - \left[\hat{y}_{i,j}(\boldsymbol{\theta}) - \rho_j\hat{y}_{i-1,j}(\boldsymbol{\theta})\right]\right\}^2\right\}$$

3.3.5.2. Behavior definition

In the calibration process, proposed parameter sets are accepted only when results (e.g. statistics calculated by comparing simulated outputs with observed data) are within a certain degree of reasonable behavior. Various standards for behavior ratings may be used to classify parameter sets with significant or rational performances (e.g. very good, good, and satisfactory). However, a method that can generate more behavior parameter sets is not guaranteed to have better overall performance. The reason for that is because the number of behavior parameter sets represents only the parameter sets with better manually assigned statistical (it can also be defined by many other different ways, [*Beck et al.*, 2002]) values and there is still chance that global optimal solution is located in some other part of behavioral region.

Table 3.1 General Performance Ratings

Performance Rating	NSE	PBIAS (%)	
		Streamflow	NOX
Very Good	0.75 < NSE ≤ 1.00	PBIAS < ±10	PBIAS < ±25
Good	0.65 < NSE ≤ 0.75	±10 ≤ PBIAS < ±15	±25 ≤ PBIAS < ±40
Satisfactory	0.50 < NSE ≤ 0.65	±15 ≤ PBIAS < ±25	±40 ≤ PBIAS < ±70
Unsatisfactory	NSE ≤ 0.50	PBIAS ≥ ±25	PBIAS ≥ ±70

NSE: Nash-Sutcliffe efficiency coefficient
PBIAS: Percent bias

In this study, the basis of assessing parameter behavior is by the general performance ratings (GPR) [*Moriasi et al.*, 2007] shown in Table 3.1. The general performance ratings are designed for monthly time steps but the calibration of daily streamflow is also applying the same sets of

rating groups in case studies.

3.4. Case Studies

3.4.1. Eagle Creek Watershed Calibration

In this study, the SWAT2009 is applied to be the watershed simulation model for ECW. Four years (1997~2000) of observed streamflow (one gauge station) and total nitrate (four gauge stations) data are available within 35 subbasins correspond to 446 hydrologic response units. Each run includes six-year SWAT simulation with a two years warm up (1995 and 1996) period.

Four approaches are implemented to calibrate monthly streamflow and monthly total nitrate (28 parameters) in this study: (I) calibration without any constraint, (II) calibration with including denitrification constraint, (III) calibration with including constraint applied on the ratio of nitrate-N losses contributed from subsurface and surface flow, (IV) calibration with including both constraints included.

3.4.2. Results of All Approaches

3.4.2.1. Results of objective function versus model evaluations

From Figure 3.2, results of overall performance in achieving objective function show all four scenarios converge to similar values and all scenarios have converged approximately before 2000 runs (convergence criterion in this study is defined as: no improvement on the objective function value for more than 1% for 1000 model evaluations). The best solutions of objective function for each cases show that Scenario IV is the best of all (330.7), Scenario III (336.7) is in the second place which is very close to that in Scenario I (337.2) and Scenario II is overall the worst. The convergence speed of Scenario II and III is slightly slower than that of Scenario I. However, it is clear that Scenario IV is converging slower than all other cases.

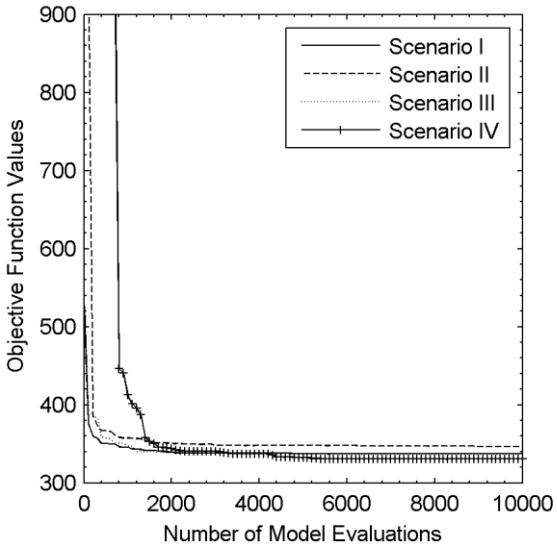

Figure 3.2 Overall performance of objective function versus model evaluations of fours scenarios

3.4.2.2. Behavior rate

By applying the GPR, the rate of behavior can be shown in Table 3.2. The behavior rate (percentage of model evaluations meeting behavior definitions) is significantly decreased when constraint of denitrification is involved. Nevertheless, the second constraint in Scenario III does not show an influence comparable to Scenario II. By applying both constraints simultaneously, behavior rate declines dramatically from the GPR category "Satisfactory" to "Good" which means the denitrification-only constraint has the greatest impact on the behavior rate of model results.

Table 3.2 Behavior Rate after the GPR is Applied

Case Studies	Behavior Rate (%)		
	Satisfactory	Good	Very Good
Scenario I	78.73	71.39	19.97
Scenario II	1.91	0.03	0
Scenario III	64.4	29.5	0.07
Scenario IV	60.78	1.22	0

3.4.2.3. Rate of penalty

The GPR is a statistical evaluation standard designed for results which have already been generated. However, penalization of the objective function by additional constraints (two constraints are applied in this study) is implemented during the parameter estimation process. If the model outputs violate the predetermined constraints (e.g. denitrification is larger than 50 kg/ha), the objective function value will be penalized with a large number (since a minimized objective function correspond to a better fit between simulated and observed data). As shown in Table 3.3, the penalty rate for Scenario I is zero but almost all model evaluations are violating at least one constraint (99.96%). In Scenario II, the application of constraint for denitrification effectively reduces percentage of penalized parameter sets. However, the overall (in three GPR categories) behavioral percentage (from Table 3.2) is relatively low which means the available behavioral solutions are short in number. In Scenario III, by applying constraint 2, the overall rate of violation is reduced but the rate of violation for constraint 1 (92.99%) is as similar to that for Scenario I. More than 93% of all model evaluations violate at least one constraint. In Scenario IV, the rate of violation for at least one constraint drops to 59% and the rate of individual constraint being violated is ranged between Scenario II and Scenario III.

The most interesting finding here is that penalty rate of violating the second constraint decreases evidently with the application of constraining denitrification even though the second constraint is actually not applied. More details will be discussed in the latter section.

Table 3.3 Rate of Penalty in Case Studies

Case Studies	Rate of Penalty	Constraint 1 violated	Constraint 2 violated	At least one constraint violated	Both constraints violated
Scenario I	0	99.83	99.59	99.96	99.46
Scenario II	24.5	24.5	7.6	26.71	5.39
Scenario III	48.48	92.99	48.48	93.11	48.34
Scenario IV	59.24	56.08	22.69	59.24	19.55

* Rate of penalty: Total number of model runs with objective function being penalized because of the violation of constraint(s) applied out of total model evaluations
** All values in the unit of %

3.4.2.4. Statistical performances before and after considering both constraints

Statistics of four scenarios (for calibration period 1997-2000) with best objective function value disregarding the satisfactory of both constraints are shown in Table 3.4. From Table 3.4, it is clear that results of best objective function in Scenario I and III obviously violate at least one constraint. Therefore, best solutions from case studies are reselected by considering the satisfactory of both constraints are as shown in Table 3.5. NSE and PBIAS values in Table 3.5 are evidently worse than results in Table 3.4 for Scenario I and III. In Table 3.6, statistics of candidate parameter sets selected by considering both constraints show that Scenario IV dominant three other cases in all categories. As mentioned previously that SSQ ratio may be carried to the proper range by only applying denitrification constraint. However, it does not change the fact that both constraints are required to attain good solutions in both calibration and Evaluation periods.

Table 3.4 Statistical Performance of All Case Studies (for Calibration Period 1997-2000) with Best Objective Function Value Disregarding the Satisfactory of Both Constraints

Case Studies	35		32		27		22		20		Denitrification	SSQ Ratio
	RE	NS	RE	NS	RE	NS	RE	NS	RE	NS		
Scenario I	-0.37	0.91	9.18	0.95	20.17	0.91	6.86	0.85	30.61	0.84	257.3	0.13
Scenario II	19.57	0.87	-41.5	0.52	-21.47	0.78	-41.9	0.49	-17.61	0.87	21.40	0.68
Scenario III	12.29	0.89	-2.06	0.89	16.09	0.88	-10.4	0.75	23.35	0.86	175.3	0.60
Scenario IV	19.57	0.87	-26.6	0.72	-8.45	0.88	-26.9	0.66	-5.63	0.94	33.10	0.67

Table 3.5 Statistical Performance of All Case Studies (for Calibration Period 1997-2000) with Best Objective Function Value Considering the Satisfactory of Both Constraints

Case Studies	35		32		27		22		20		Denitrification	SSQ Ratio
	RE	NS	RE	NS	RE	NS	RE	NS	RE	NS		
Scenario I	52.84	0.49	31.89	0.83	41.52	0.74	41.18	0.62	43.54	0.70	30.37	0.93
Scenario II	19.57	0.87	-41.5	0.52	-21.47	0.78	-41.9	0.49	-17.61	0.87	21.40	0.68
Scenario III	11.40	0.89	-131.4	-1.7	-90.55	-0.42	-164.6	-2.9	-69.7	0.39	25.02	0.76
Scenario IV	19.57	0.87	-26.6	0.72	-8.45	0.88	-26.9	0.66	-5.63	0.94	33.10	0.67

Table 3.6 Statistical Performance of All Case Studies (for Evaluation Period 2001-2003) with Best Objective Function Value Considering the Satisfactory of Both Constraints

Case Studies	35		32		27		22		20		Denitrification	SSQ Ratio
	RE	NS	RE	NS	RE	NS	RE	NS	RE	NS		
Scenario I	62.82	0.11	39.88	0.60	49.35	0.54	40.97	0.61	59.71	0.52	28.94	0.90
Scenario II	38.00	0.61	-32.34	0.23	-15.55	0.47	-49.56	-0.13	-14.86	0.67	25.10	0.66
Scenario III	30.34	0.69	-136.8	-2.87	-106.0	-1.67	-207.5	-7.37	-84.56	-0.30	28.94	0.71
Scenario IV	37.99	0.60	-21.47	0.43	-6.12	0.58	-38.98	0.15	-6.01	0.72	39.04	0.64

Table 3.7 Inclusion Rate of Observed Streamflow/NOX Data Included in 95% Confidence Interval and the Corresponding Average Spread

Case Studies		Inclusion Rate (%)				
		35	32	27	22	20
Scenario I	IR (%)	16.67	44.44	48.15	48.15	55.56
	Spread	0.42	0.54	0.56	0.52	0.53
Scenario II	IR (%)	20.83	40.74	48.15	40.74	48.15
	Spread	0.43	4.87	4.78	6.40	3.42
Scenario III	IR (%)	18.75	37.04	37.04	37.04	48.15
	Spread	0.39	8.00	9.51	4.45	5.83
Scenario IV	IR (%)	35.42	51.85	59.26	55.56	62.96
	Spread	0.47	11.75	19.18	9.99	58.06

Streamflow station number: 35
Total nitrate (NOX) station number: 35, 27, 22, 20
Measurement error included based on the cumulative probable uncertainty represented by probable error range [*Harmel et al.*, 2006]

3.4.2.5. Inclusion rate

As shown in Table 3.7, inclusion rate of observed streamflow data (gauge station number 35) within the uncertainty band constructed by a 95% confidence interval increases when an additional constraint(s) is included. The rate of inclusion for total nitrate stations (gauge station

number 32, 37, 22, 20) decreases when only one constraint is applied, but increases in Scenario IV, which has both constraints implemented.

3.4.2.6. Cumulative distribution functions of constraints

Cumulative distribution functions of constraints applied in four cases are shown in Figure 3.3(A) and 3.3(B). In Scenario I with no constraints, denitrification converges to 230~260 kg/ha and the SSQ ratio converges to 0.13~0.15 which are both unrealistic for the case study area (99.96% of outputs violate at least one constraint). In Scenario II, the inclusion of the denitrification constraint forces model denitrification rate converging to a reasonable (0~50 kg/ha) and the SSQ ratio has also converged to 0.89~0.92 with no constraint involved. Only 26.7% of model outputs violate at least one constraint. In Scenario III, the constraint of SSQ ratio holds half of the model evaluations larger than 0.6 but more than 92% of outputs violate the denitrification constraint. In the last case, around 40% of outputs satisfy both constraints and the violation rate of denitrification is still higher than SSQ ratio. From the patterns of Figure 3.3(A) and 3.3(B), an important conclusion as that in the section of penalty rate can be found which will be discussed in latter section.

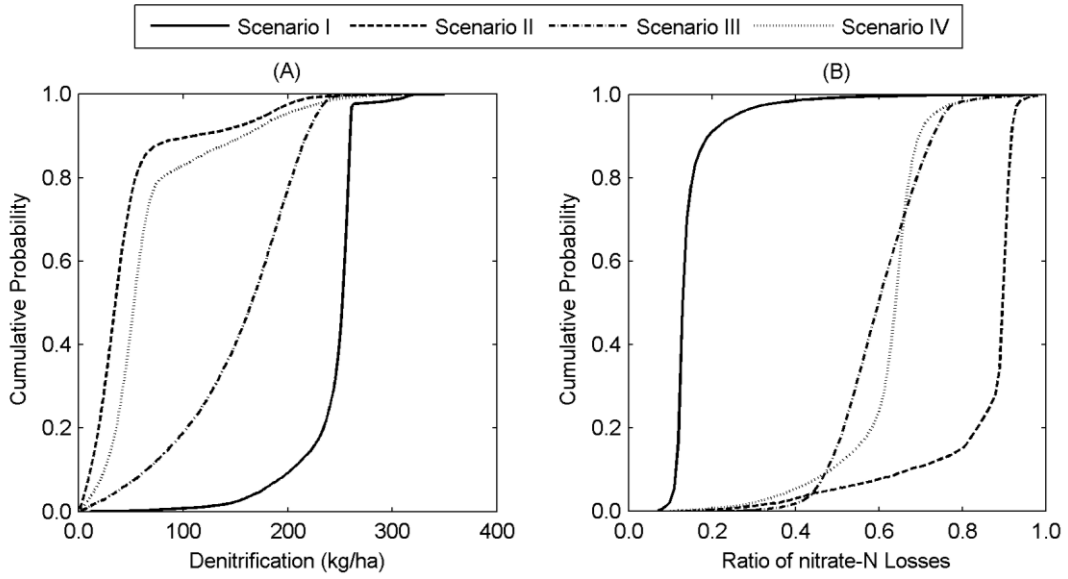

Figure 3.3 Cumulative distribution functions of constraints applied in four scenarios: (A) CDFs of denitrification, (B) CDFs of SSQ ratio. (Ratio of nitrate-N losses is calculated by the nitrate-N losses contributed from subsurface flow divided by the sum of nitrate-N losses from both surface and subsurface flow)

3.4.2.7. Cumulative distribution functions of sensitive parameters

CDFs of comparatively sensitive parameters are shown in Figure 3.4 where NPERCO, DEM_IMP and DDRAIN are sensitive to denitrification process and CDN, SDNCO are sensitive to SSQ ratio. From Figure 3.4, one can see that parameters in Scenario I are converging to ranges which are not realistic (e.g. the values of NPERCO is close to 1 represents N-nitrate losses are mostly contributed from surface runoff [*Neitsch et al.*, 2011]; values of DEM_IMP and DDRAIN are too high [*Green et al.*, 2006]; CDN should be lower than 0.5 and SDNCO is too low [*Birr et al.*, 2007; *Sui and Frankenberger*, 2008]). In Scenario II, CDFs of sensitive parameters are close to that in Scenario IV except for NPERCO (which means SSQ ratio sensitive parameter are also converging to rational a region even without assigning constraint). In Scenario III, CDN and SDNCO converge to reasonable ranges but not for denitrification sensitive parameters. In Scenario IV, the two water quality processes (denitrification and nitrate-

N losses) are being balanced by both constraints and all sensitive parameters are converging to moderately appropriate ranges.

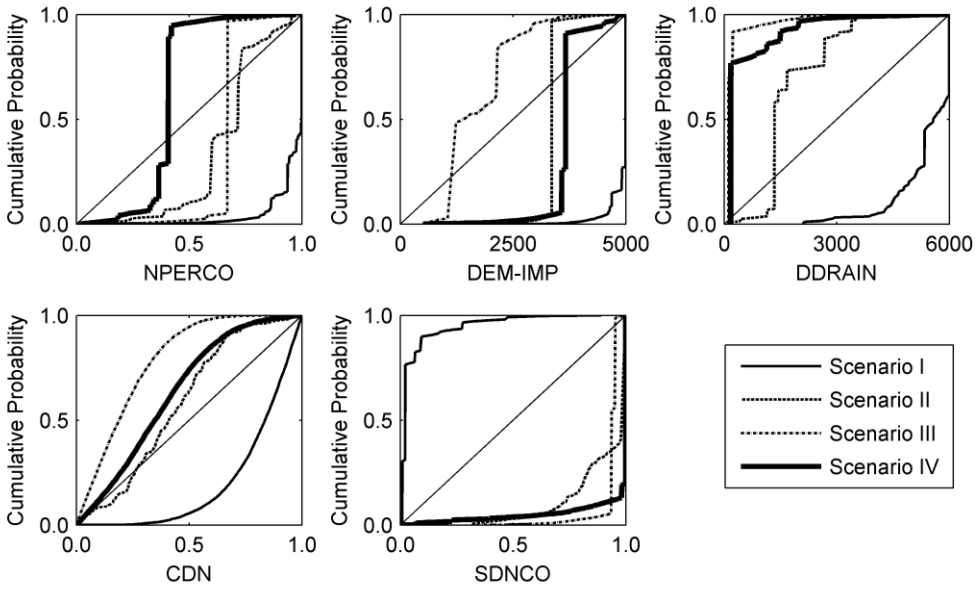

Figure 3.4 Cumulative distribution functions of sensitive parameter of all scenarios.

3.4.2.8. Predictive uncertainty of quantile-quantile plot

The predictive uncertainty in this study is adopting the quantile-quantile plot (QQ plot) as the guidance in evaluating the consistency between predicted and observed discharge and total nitrate processes [*Thyer et al.*, 2009]. The QQ plot for calibration (1997-2000) and Evaluation (2001-2003) periods are shown in Figure 3.5 (the solid lines are solutions based on the best solutions considering both constraints and the dot lines in (A) and (C) represent the best solutions without considering both constraints) and 6 (based on the best solutions considering both constraints). From Figure 3.5, it is clear that streamflow is being properly simulated in Scenario II, III and IV. For total nitrate stations, Scenario II and IV are showing similar patterns which are consistent with the findings in previous sections that the denitrification constraint will automatically guide the calibration results to have SSQ ratio converged to the proper range

simultaneously. On the other hand, QQ plots of Scenario III show significant overestimation in total nitrate station 22 and 20. In Figure 3.6, the case of applying both constraints demonstrates significant advantage while Scenario I, II and III do not have comparable performance in making discharge and total nitrate predictions. Results from the standpoint of statistical evaluations and uncertainty feature both validate the importance to include both constraints simultaneously instead of one or less.

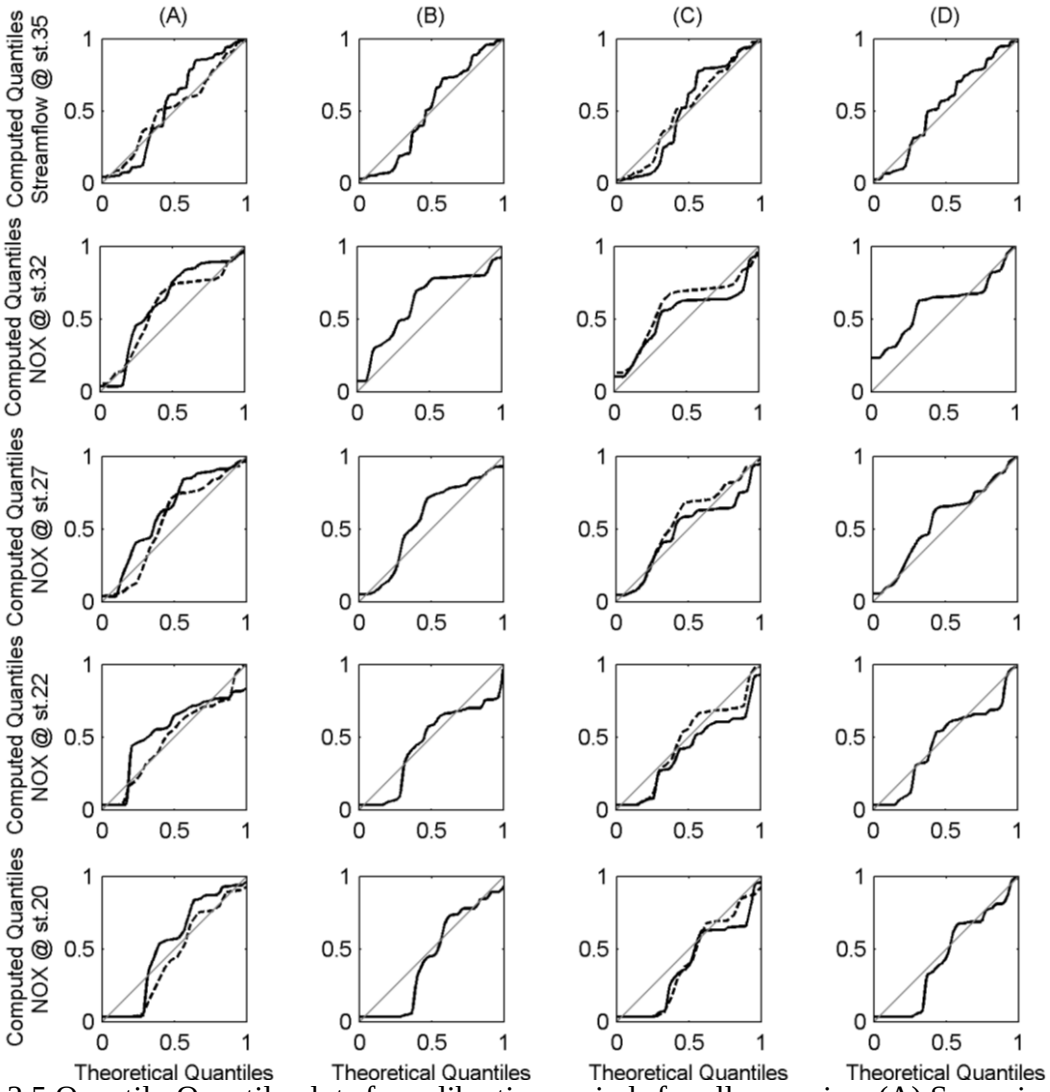

Figure 3.5 Quantile-Quantile plots for calibration periods for all scenarios: (A) Scenario I with no constraint applied, (B) Scenario II with constraint of denitrification, (C) Scenario III with constraint of SSQ ratio, (D) Scenario IV with both constraints applied. The dot lines in (A) and in (C) represent the best solutions without considering both constraints.

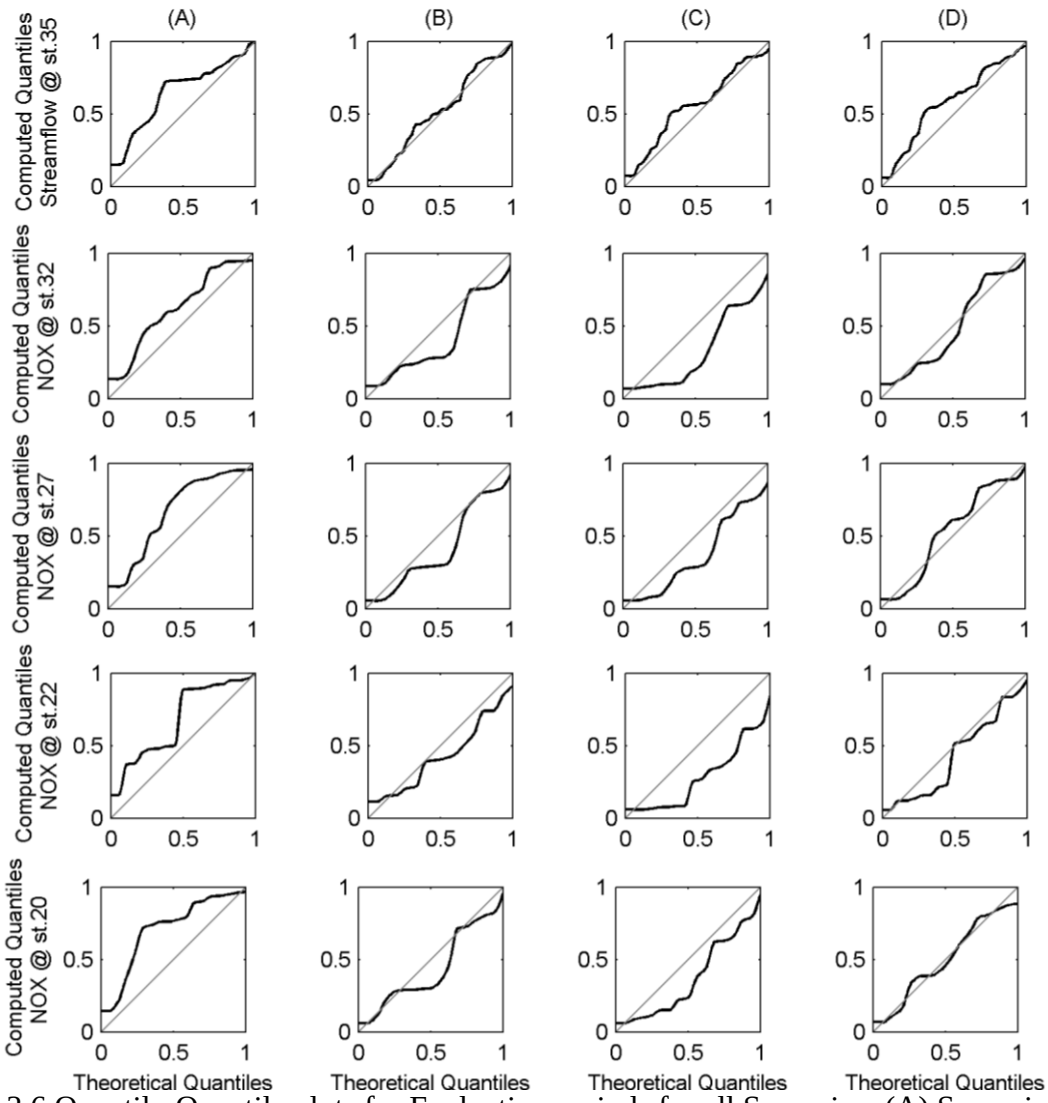

Figure 3.6 Quantile-Quantile plots for Evaluation periods for all Scenarios: (A) Scenario I with no constraint applied, (B) Scenario II with constraint of denitrification, (C) Scenario III with constraint of SSQ ratio, (D) Scenario IV with both constraints applied.

3.5. Discussions and Conclusions

From results in the previous section, significant advantage has been found with applying constraints based on intra-watershed responses. Results of watershed calibration may not satisfy realistic water quality responses without applying additional constraint(s) correspond to intra-watershed responses. The selection of a statistically sound likelihood function and the application of behavior definition alone do not provide effective support in watershed calibration. In other words, execution of watershed calibration without implementing intra-

watershed responses may result in model parameters converging to domains which generate physically meaningless outputs with good statistics.

Moreover, application of a denitrification constraint can also direct SSQ ratio to a proper range (Scenario II) but denitrification is not affected by only applying the constraint of SSQ ratio. This implies that the interaction between associated nitrate processes may have dominant influence over the other. Total nitrate may be well calibrated by constraining only some of the major model nutrient processes while still reducing model uncertainty.

In this study, the role of intra-watershed responses is explored by the applications of additional constraints. Several conclusions can be made from comparisons of different scenarios.

1. Application of intra-watershed responses shows noteworthy improvement on overall performance in achieving global optimality, increasing number of behavioral solutions and the quality of predictive outputs is significantly enhanced.
2. Results with good statistics can be derived without any constraint applied. However, large proportion of generated parameter sets violates intra-watershed responses. In other words, statistically well performed realizations are giving physically wrong outputs in real world case studies (disobey watershed characteristics in nature).
3. Both the denitrification and the nitrate-N losses ratio constraints are required to attain good statistical solutions in both calibration and Evaluation periods.
4. Constraint of denitrification shows greater impact on the behavior rate of model results. The penalty rate of violating the nitrate-N losses ratio constraint decreases evidently with the application of denitrification constraint when the nitrate-N losses ratio constraint is actually not applied.

5. The inclusion rate of all outputs (streamflow and total nitrate) stations and the average spread of uncertainty band have increased when both constraints are applied (which means the increase of uncertainty enhances the ability to include more observation data into the uncertainty band). However, the inclusion rate decreases but the average spread increases when only one constraint is applied (which means the increase of uncertainty does not enhance but having the ability of including more observation data into the uncertainty band declined).

For complex large-scale watershed simulation models, it is essential to apply intra-watershed responses because behavior definition based on statistical thresholds and the choice of objective function may not result in representative model outputs of realistic watershed hydrologic/nutrients processes. The application of intra-watershed responses by additional constraints successfully enhances the quality of calibration results and also reduces parameter and predictive uncertainty. Watershed model calibration in the future should have more emphasis on the inclusion of physically essential information by including realistic field data in watershed simulation processes. In summary, intra-watershed responses constraints present great potential benefits in enhancing the quality of watershed calibration, reducing parameter and predictive uncertainty, and also provides a novel direction for watershed modeling in the future.

References

Ahmadi, M., M. Arabi, J. C. Ascough II, D. G. Fontane, and B. A. Engelc (2012), Toward improved calibration of watershed models: Multisite many objective measures of information, *Submitted to: Advances in Water Resources*.

Amatya, D. M., and M. K. Jha (2011), Evaluating the SWAT model for a low-gradient forested watershed in coastal south Carolina, *Transactions of the ASABE*, *54*(6), 2151–2163.

Arnold, J. G., P. M. Allen, and G. Bernhardt (1993), A comprehensive surface-groundwater flow model, *Journal of Hydrology*, *142*, 47–69.

Arnold, J. G., P. M. Allen, M. Volk, J. R. Williams, and D. D. Bosch (2010), Assessment of Different Representations of Spatial Variability on SWAT Model Performance, *Transactions of the ASABE*, *53*(5), 1433–1443.

Arnold, J. G., R. Srinivasan, R. S. Muttiah, and J. R. Williams (1998), Large area hydrologic modeling and assessment. I. Model development, *Journal of the American Water Resources Association*, *34*(1), 73–89.

Aulakh, M. S., J. W. Doran, D. T. Walters, and J. F. Power (1991), Legume residue and soil water effects on denitrification in soil of different textures, *Soil Biology & Biochemistry*, *23*(12), 1161–1167.

Beck, M. B. et al. (2002), Developing a concept of adaptive community learning: Case study of a rapidly urbanizing watershed, *Integrated Assessment*, *3*(4), 299–307, doi:10.1076/iaij.3.4.299.13583. [online] Available from: http://www.tandfonline.com/doi/abs/10.1076/iaij.3.4.299.13583

Birr, A., S. Folle, B. Dalzell, and D. Mulla (2007), *Evaluation of best management practices (BMPs) in impaired watersheds using the SWAT model*.

Box, G. E. P., and G. C. Tiao (1992), *Bayesian inference in statistical analysis*, John Wiley & Sons.

Cassidy, R., and P. Jordan (2011), Limitations of instantaneous water quality sampling in surface-water catchments: comparison with near-continuous phosphorus time-series data, *Journal of Hydrology*, *405*(1-2), 182–193, doi:10.1016/j.jhydrol.2011.05.020. [online] Available from: http://linkinghub.elsevier.com/retrieve/pii/S0022169411003301 (Accessed 22 June 2011)

Chiang, L., I. Chaubey, M. W. Gitau, and J. G. Arnold (2010), Differentiating impacts of land use changes from pasture management in a CEAP watershed using the SWAT model, *Transactions of the ASABE*, *53*(2), 1569–1584.

David, M. B., S. J. Del Grosso, X. Hu, E. P. Marshall, G. F. McIsaac, W. J. Parton, C. Tonitto, and M. Youssef (2009), Modeling denitrification in a tile-drained, corn and soybean agroecosystem of Illinois, USA, *Biogeochemistry*, *93*, 7–30, doi:10.1007/s10533-008-9273-9.

Douglas-Mankin, K. R., R. Srinivasan, and J. G. Arnold (2010), Soil and Water Assessment Tool (SWAT) model: current developments and applications, *Transactions of the ASABE*, *53*(5), 1423–1431.

Du, B., J. G. Arnold, A. Saleh, and D. B. Jaynes (2005), Development and application of SWAT to landscapes with tiles and potholes, *Transactions of the ASABE*, *48*(3).

Duan, Q., S. Sorooshian, and V. Gupta (1992), Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resources Research*, *28*(4), 1015–1031.

Ebel, B. a., and K. Loague (2006), Physics-based hydrologic-response simulation: Seeing through the fog of equifinality, *Hydrological Processes*, *20*(13), 2887–2900, doi:10.1002/hyp.6388. [online] Available from: http://doi.wiley.com/10.1002/hyp.6388 (Accessed 8 March 2012)

Ghebremichael, L. T., T. L. Veith, and M. C. Watzin (2010), Determination of critical source areas for phosphorus loss: Lake Champlain Basin, Vermont, *Transactions of the ASABE*, *53*(5), 1595–1604.

Gilliam, J. W., J. L. Baker, and K. R. Reddy (1999), Water quality effects of drainage in humid regions, *Agricultural Drainage*, *SSSA, Madison WI*, 801–830.

Green, C. H., M. D. Tomer, M. D. Luzio, and J. G. Arnold (2006), Hydrologic Evaluation if the Soil and Water Assessment Tool for a Large Tile-Drianed Watershed in IOWA, *Transactions of the ASABE*, *49*(2), 413–422.

Gupta, H. V., S. Sorooshian, and P. O. Yapo (1999), Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration, *Journal of Hydrologic Engineering*, *4*(2), 135–143.

Haario, H., M. Laine, A. Mira, and E. Saksman (2006), DRAM: Efficient adaptive MCMC, *Statistics and Computing*, *16*(4), 339–354, doi:10.1007/s11222-006-9438-0. [online] Available from: http://www.springerlink.com/index/10.1007/s11222-006-9438-0 (Accessed 28 June 2011)

Harmel, R. D., R. Cooper, R. Slade, R. Haney, and J. Arnold (2006), Cumulative uncertainty in measured streamflow and water quality data for small watersheds, *Transactions of the ASABE*, *49*(3), 689. [online] Available from: http://asae.frymulti.com/abstract.asp?aid=20488&t=1 (Accessed 22 June 2011)

Jayakrishnan, R., R. Srinivasan, C. Santhi, and J. G. Arnold (2005), Advances in the application of the SWAT model for water resources management, *Hydrological Processes*, *19*(3), 749–762, doi:10.1002/hyp.5624. [online] Available from: http://doi.wiley.com/10.1002/hyp.5624 (Accessed 18 March 2012)

Jaynes, D. B., J. L. Hatfield, and D. W. Meek (1999), Water Quality in Walnut Creek Watershed: Herbicides and Nitrate in Surface Waters, *Journal of Environmental Quality*, *28*, 45–59.

Kalita, P. K., A. S. Algoazany, J. K. Mitchell, R. a. C. Cooke, and M. C. Hirschi (2006), Subsurface water quality from a flat tile-drained watershed in Illinois, USA, *Agriculture, Ecosystems & Environment*, *115*(1-4), 183–193, doi:10.1016/j.agee.2006.01.006. [online] Available from: http://linkinghub.elsevier.com/retrieve/pii/S0167880906000338 (Accessed 25 April 2012)

Kim, N. W., A. H. Shin, and J. Lee (2010), Effects of streamflow routing schemes on water quality with SWAT, *Transactions of the ASABE*, *53*(5), 1457–1468.

Klepper, O., and E. Hendrix (1994), A method for robust calibration of ecological models under different types of uncertainty, *Ecological Modelling*, *74*(3-4), 161–182, doi:10.1016/0304-3800(94)90118-X. [online] Available from: http://linkinghub.elsevier.com/retrieve/pii/030438009490118X

Liew, M. W. V., T. L. Veith, D. D. Bosch, and J. G. Arnold (2007), Suitability of SWAT for the Conservation Effects Assessment Project : Comparison on USDA Agricultural Research Service Watersheds, *Journal of Hydrologic Engineering*, (APRIL), 173–189.

Marshall, L., D. Nott, and A. Sharma (2004), A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modeling, *Water Resources Research*, *40*(2), 1–11,

doi:10.1029/2003WR002378. [online] Available from: http://www.agu.org/pubs/crossref/2004/2003WR002378.shtml (Accessed 28 June 2011)

Meng, H., A. M. Sexton, M. C. Maddox, A. Sood, C. W. Brown, R. R. Ferraro, and R. Murtugudde (2010), Modelling Rappahannock river basin using SWAT – Pilot for Chesapeake Bay watershed, *Applied Engineering in Agriculture*, *26*(5), 795–806.

Moriasi, D. N., J. G. Arnold, B. A. Engel, and C. G. Rossi (2009), Incorporation of a new shallow water table depth algorithm into SWAT2005, *Transactions of the ASABE*, *52*(3), 771–784.

Moriasi, D. N., J. G. Arnold, M. W. V. Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith (2007), Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, *Transactions of the ASABE*, *50*(3), 885–900.

Neitsch, S. L., J. G. Arnold, J. R. Kiniry, and W. J. R (2011), Soil & Water Assessment Tool Theoretical Documentation Version 2009, *Texas Water Resources Institute Technical Report No. 406, Texas A&M University System*.

Newman, J. E. (1997), *Our Changing Climate, In: The Natural Heritage of Indiana, Edited M. T. Jackson*.

Refsgaard, J. C. (1997), Parameterisation, calibration and Evaluation of distributed hydrological models, *Journal of Hydrology*, *198*(1-4), 69–97, doi:10.1016/S0022-1694(96)03329-X. [online] Available from: http://linkinghub.elsevier.com/retrieve/pii/S002216949603329X (Accessed 11 March 2012)

Schilling, K., and Y.-K. Zhang (2004), Baseflow contribution to nitrate-nitrogen export from a large, agricultural watershed, USA, *Journal of Hydrology*, *295*(1-4), 305–316, doi:10.1016/j.jhydrol.2004.03.010. [online] Available from: http://linkinghub.elsevier.com/retrieve/pii/S0022169404001738 (Accessed 31 July 2012)

Sorooshian, S., and J. A. Dracup (1980), Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: Correlated and heteroscedastic error cases, *Water Resources Research*, *16*(2), 430, doi:10.1029/WR016i002p00430.

Sorooshian, S., and V. K. Gupta (1983), Automatic calibration of conceptual rainfall-runoff model: The question of parameter observability and uniqueness, *Water Resources Research*, *19*(1), 260–268.

Srinivasan, R. X. Zhang, J. A. (2010), Swat Ungaged: Hydrological Budget and Crop Yield Predictions in the Upper Mississippi River Basin, *Transactions of the ASABE*, *53*(5), 1533–1546.

Stedinger, J. R., R. M. Vogel, S. U. Lee, and R. Batchelder (2008), Appraisal of the generalized likelihood uncertainty estimation (GLUE) method, *Water Resources Research*, *44*, 1–17, doi:10.1029/2008WR006822. [online] Available from: http://www.agu.org/pubs/crossref/2008/2008WR006822.shtml (Accessed 22 June 2011)

Sui, Y., and J. R. Frankenberger (2008), Nitrate loss from subsurface drains in an agricultural watershed using SWAT2005, *Transactions of the ASABE*, *51*(4), 1263–1272.

Thyer, M., B. Renard, D. Kavetski, G. Kuczera, S. W. Franks, and S. Srikanthan (2009), Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis, *Water Resources Research*, *45*(3), 1–22, doi:10.1029/2008WR006825. [online] Available from: http://www.agu.org/pubs/crossref/2009/2008WR006825.shtml (Accessed 19 July 2011)

Tolson, B. A., and C. A. Shoemaker (2007), Dynamically dimensioned search algorithm for computationally efficient watershed model calibration, *Water Resources Research*, *43*(1),

1–16, doi:10.1029/2005WR004723. [online] Available from: http://www.agu.org/pubs/crossref/2007/2005WR004723.shtml (Accessed 28 June 2011)

Vrugt, J. A., C. J. F. Braak, C. G. H. Diks, B. A. Robinson, J. M. Hyman, and D. Higdon (2009a), Accelerating Markov Chain Monte Carlo Simulation by Differential Evolution with Self-Adaptive Randomized Subspace Sampling, *International Journal of Nonlinear Sciences & Numerical Simulation*, *10*(March), 271–288.

Vrugt, J. A., C. J. F. Braak, V. K. Gupta, and B. A. Robinson (2009b), Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?, *Stochastic Environmental Research and Risk Assessment*, *23*(7), 1011–1026, doi:10.1007/s00477-008-0274-y.

Vrugt, J. A., C. J. F. ter Braak, M. P. Clark, J. M. Hyman, and B. A. Robinson (2008), Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation, *Water Resources Research*, *44*, 1–15, doi:10.1029/2007WR006720. [online] Available from: http://www.agu.org/pubs/crossref/2008/2007WR006720.shtml (Accessed 24 August 2011)

Vrugt, J. A., H. V. Gupta, W. Bouten, and S. Sorooshian (2003), A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resources Research*, *39*(8), doi:10.1029/2002WR001642. [online] Available from: http://www.agu.org/pubs/crossref/2003/2002WR001642.shtml (Accessed 10 June 2011)

Yang, J., P. Reichert, K. Abbaspour, J. Xia, and H. Yang (2008), Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China, *Journal of Hydrology*, *358*(1-2), 1–23, doi:10.1016/j.jhydrol.2008.05.012. [online] Available from: http://linkinghub.elsevier.com/retrieve/pii/S0022169408002370 (Accessed 10 June 2011)

Yen, H., R. M. Records, M. Arabi, and B. A. Tolson (2012), Appraisal of sampling techniques for parameter estimation and uncertainty analysis in watershed modeling, *Under internal review*.

Chapter 4

Evaluation of Structural Uncertainty on Hydrologic and Water Quality Predictions Using SWAT Model

4.1. Overview

For surface runoff estimation in the Soil and Water Assessment Tool (SWAT) model, the curve number (CN) procedure is commonly adopted to calculate surface runoff by utilizing antecedent soil moisture condition (SCSI) in field. In the recent version of SWAT (SWAT2005), an alternative approach is available to apply CN method by implementing information from plant evapotranspiration (SCSII). Improvement of representing surface runoff process by SCSII has been shown in previous studies. However, few quantitative comparison of model performance in simulating hydrologic processes has been made between the two CN approaches alone or in combination. In addition, the effect of SCSII on water quality responses (e.g. total nitrate, pesticide) in SWAT has not been evaluated. The main goal of this study is to evaluate the role of structural uncertainty of the SWAT model on hydrologic and water quality predictions by two different methods. The analysis hinges on the evaluation of: (i) To characterize improvements in hydrologic and water quality predictions may be made by utilizing different surface runoff estimation techniques alone and in combination; and (ii) To investigate how model predictive uncertainty may be affected by combining such techniques. Two approaches are combined by the

Bayesian model averaging (BMA) method in multi-site, multiple-response case studies at the Eagle Creek watershed, Indiana. Results show that SCSII and BMA associated approaches have outstanding performance in both discharge and total nitrate predictions compare to SCSI. In addition, applications of BMA have positive effect on the increase of inclusion rate but the predictive uncertainty is not evidently reduced/enhanced.

4.2. Introduction

Hydrological models are used extensively for water resources planning and management [*Arnold et al.*, 1993; *Beven and Kirkby*, 1979; *Boyle*, 2001; *Burnash et al.*, 1973; *Schaake et al.*, 1996]. Among models for various purposes, the Soil and Water Assessment Tool (SWAT) is a widely used, semi-distributed river basin scale model developed to quantify responses in water quality and quantity to land use changes and other anthropogenic activities which has been applied to a variety of water resources investigations [*Arnold et al.*, 1993; *Borah et al.*, 2006; *Gassman et al.*, 2007].

In the SWAT model, some hydrologic and water quality processes can be represented with several modules. For surface runoff estimation, two methods are available: the SCS curve number (CN) procedure [*USDA-SCS*, 1972], and the Green & Ampt infiltration method [*Green and Ampt*, 1911]. The Green & Ampt method is not commonly applied for it requires frequent climate observations (Green & Ampt method will not be discussed in this script). The CN method accounts for short-term losses (canopy interception, depression storage, and infiltration), but may not properly reflect long-term predictions during simulation (e.g. evapotranspiration and evaporation) [*Ponce and Hawkins*, 1996]. In the past, the retention parameter of CN method has typically been calculated using antecedent soil moisture conditions [*Kannan et al.*, 2007]. While this approach has been widely applied in previous versions of the SWAT models and relevant

studies, however, it tends to overestimate surface runoff in shallow soils when storage condition is low [*Amatya and Jha*, 2011; *Kannan et al.*, 2007; *Setegn et al.*, 2010].

An alternate approach which applies plant evapotranspiration to calculate the CN retention parameter may be used to improve representation of surface runoff under such conditions, and has been made available in recent versions of SWAT (SWAT2005). Previous hydrologic assessments using SWAT have generally calculated surface runoff with the CN method either utilizing antecedent soil moisture condition [*King et al.*, 1999; *Narasimhan et al.*, 2005; *Nie et al.*, 2012; *Schuol et al.*, 2008; *Setegn et al.*, 2010] or plant evapotranspiration [*Amatya and Jha*, 2011; *Jajarmizadeh et al.*, 2012]. However, few quantitative comparisons of model performance in simulating hydrologic processes has been made between the two CN approaches alone or in combination [*Kannan et al.*, 2007]. In addition, the effect of the plant evapotranspiration method on water quality responses (e.g. total nitrate, pesticide) in SWAT has not been evaluated. Potential differences in predicted surface runoff characteristics may be crucial for simulated water quality fluxes.

The overall goal of this study is to evaluate the role of structural uncertainty of the SWAT model on hydrologic and water quality predictions. Specifically, the following objectives are defined: (i) To characterize improvements in hydrologic and water quality predictions may be made by utilizing different surface runoff estimation techniques alone and in combination; and (ii) To investigate how model predictive uncertainty may be affected by combining such techniques. This study implements an innovative Bayesian model averaging (BMA) technique [*Hoeting et al.*, 1999]to combine hydrologic and water quality predictions by the two surface runoff calculation approaches in watershed calibration and to characterize structural uncertainty

within a single hydrologic/nutrient simulation model. The Eagle Creek Watershed of central Indiana is utilized as the case study for different scenarios.

4.3. Methods and Materials

4.3.1. Site Location and Characteristics

Eagle Creek watershed (ECW) in central Indiana (Figure 4.1) is selected as case study watershed. The ECW (248 km^2 in area) is a part of Upper White River watershed located in Boone, Hamilton, Hendricks and Marion counties. The average annual precipitation is 38 to 40 inches and the average annual temperature is 52°F [*Newman*, 1997]. The dominant land use is agricultural (59%), followed by rangeland (38%), forest (2%), and urban (1%). There are two major hydrologic soil groups in the ECW: (group B: 51%, and group C: 48%). As for available observation data, one discharge (station number 35 in Figure 4.1) and monthly instantaneous observations of nitrate at four water quality monitoring gauge stations (station number 20, 22, 27, and 32 in Figure 4.1). A hydrologic model for the ECW was constructed in SWAT2005 utilizing input data and delineated into 35 subbasins corresponding to 446 hydrologic response units.

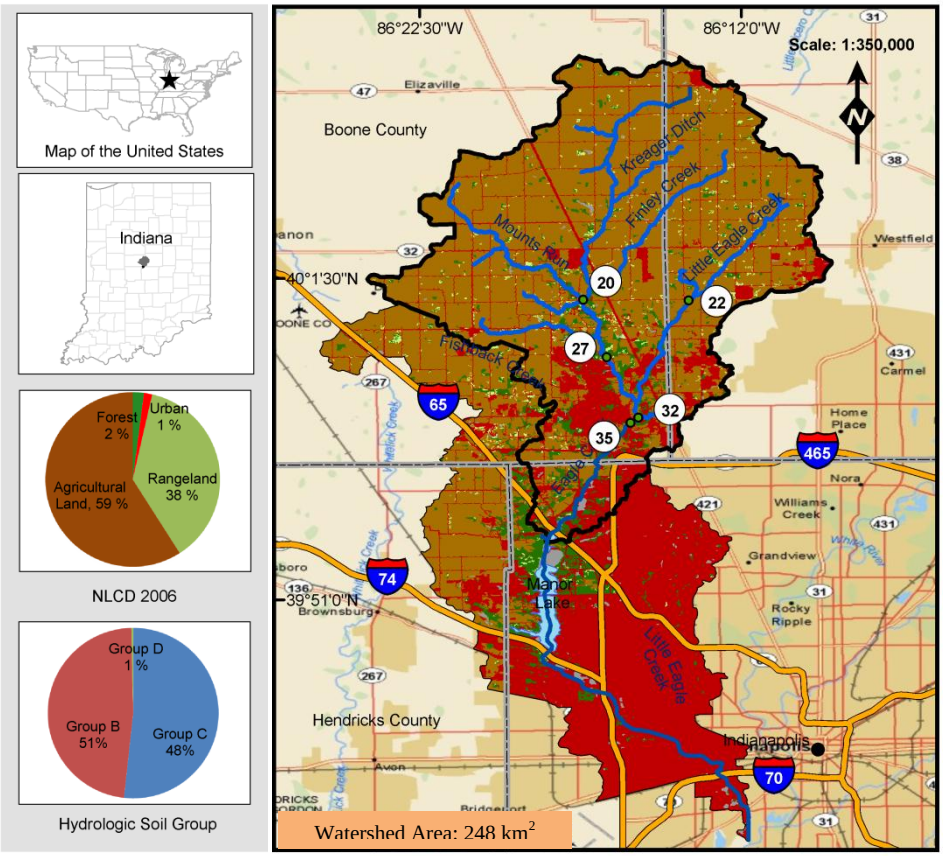

Figure 4.1 Case study area: Eagle Creek Watershed, Indiana

4.3.2. SWAT Model

The Soil and Water Assessment Tool (SWAT) [*Arnold et al.*, 1993] is a continuous-time and semi-distributed watershed simulation model which is able to simulate and predict hydrologic and water quality processes at the watershed level, and has been applied extensively in the field of watershed management [*Arnold et al.*, 2010; *Chiang et al.*, 2010; *Douglas-Mankin et al.*, 2010; *Du et al.*, 2005; *Ghebremichael et al.*, 2010; *Green et al.*, 2006; *Jayakrishnan et al.*, 2005; *Kim et al.*, 2010; *Meng et al.*, 2010; *Moriasi et al.*, 2009; *Srinivasan, R. X. Zhang*, 2010]. A detailed description of the SWAT model is presented in literature [*Arnold et al.*, 1998].

The cycle of hydrologic processes in the SWAT model is simulated based on the water balance equation as follows:

$$S_t = S_0 + \sum_{i=1}^{t}(P_i - Q_i^{surf} - E_i - W_i^{seep} - Q_i^{gw}) \quad (1)$$

Where, S_t is the final soil water content (mm H_2O); S_0 is the initial soil water content at time step i (mm H_2O); P_i is the amount of precipitation at time step i (mm H_2O); Q_i^{surf} is the amount of surface runoff at time step i (mm H_2O); E_i is the amount of evapotranspiration at time step i (mm H_2O); W_i^{seep} is the amount of water entering vadose zone from soil profile at time step i (mm H_2O); Q_i^{gw} is the amount of return flow at time step i (mm H_2O). In the latest version of SWAT, three optional functions (Green-Ampt infiltration method and two approaches using CN method) are available to calculate the amount of Q_i^{surf} discussed in the latter section.

4.3.3. SCS Curve Number Procedure

In the SWAT model, the surface runoff can be calculated by either the Green-Ampt method or the SCS curve number procedure. From Equation (2), the SCS curve number procedure [*USDA-SCS*, 1972] is designed to estimates a time series of cumulative surface runoff based on land uses, hydrologic conditions, and soil types [*Rallison and Miller*, 1981]. Cumulative surface runoff is calculated as:

$$Q = \frac{(P - I_a)^2}{(P - I_a + S)} \quad (2)$$

Where Q is cumulative surface runoff (mm); P is cumulative precipitation (mm); I_a is cumulative initial abstraction, that is, canopy interception (mm) and depression storage (mm) and S is the retention parameter (mm). The initial abstraction is assumed to be 20% of the retention parameter ($I_a = 0.2S$). Thus, Equation (2) can be re-written as:

$$Q = \frac{(P - 0.2S)^2}{(P + 0.8S)} \quad (3)$$

From Equation (3), the value of surface runoff is controlled by precipitation and the retention parameter. The retention parameter can be calculated as:

$$S = 25.4\left(\frac{1000}{CN} - 10\right) \quad (4)$$

Where CN is the empirically determined curve number associated with a given land use, slope, and soil type, and may be adjusted according to antecedent soil moisture condition or, in newer methods, the retention parameter can be adjusted by incorporating the information from plant evapotranspiration. Calculation of the retention parameter is an optional function in the SWAT model initiated in version SWAT2005 [*Kannan et al.*, 2007].

4.3.3.1. Antecedent Soil Moisture Condition (SCSI)

The original SCS approach assumes an antecedent soil moisture condition (AMC) of II (average conditions). Curve numbers for antecedent soil moisture conditions I (dry) or III (wet) can be adjusted from reported AMC values as follows:

$$CN_1 = CN_2 - \frac{20(100 - CN_2)}{(100 - CN_2 + \exp[2.533 - 0.0636(100 - CN_2)])} \quad (5)$$

$$CN_3 = CN_2 \exp[0.00673(100 - CN_2)] \quad (6)$$

Where CN_1 is the curve number of soil moisture condition I; CN_2 is the curve number of soil moisture condition II; and CN_3 is the curve number of soil moisture condition III. Previous research indicates that use of the AMC technique to calculate surface runoff results in overestimation for shallow soils and for soils with low storage [*Kannan et al.*, 2007]. Therefore, a new approach is proposed to address this problem by including a simple structure one-parameter module which will be discussed in this following section.

4.3.3.2. Plant Evapotranspiration (SCSII)

An alternative method for calculating surface runoff utilizes plant evapotranspiration in place of antecedent soil moisture condition to adjust the CN value. This method better reflects previous climatic conditions, is less reliant on soil water storage, and has been shown to improve

simulation of hydrologic processes in shallow soils [*Amatya and Jha*, 2011; *Kannan et al.*, 2007]. As shown in Equation (7), $S_t$is the retention parameter at time t; $S_{t-1}$is the retention parameter at time $t-1$; $S_{\max}$ is the maximum value of the retention parameter of all S_t; PET_t is the potential evapotranspiration for the given time t; $CNCOEF$ is the depletion coefficient; P_t is the precipitation depth for the given time t and Q_t is the surface runoff.

$$S_t = S_{t-1} + PET_t \times \exp\left[\frac{-CNCOEF \times S_{t-1}}{S_{max}}\right] - P_t + Q_t \quad (7)$$

To apply SCSII, the model parameter CNCOEF is required to be included during simulation process. Users have to predefine or calibrate the value for CNCOEF before implementing SCSII.

4.3.4. Parameter Estimation Procedure

Performance of the model for calibration and evaluation periods was assessed with the statistically valid likelihood function [*Ahmadi et al.*, 2012] as objective function and behavior definitions [*Moriasi et al.*, 2007]. The selection of likelihood function is critical for statistically sound evaluation of model performance [*Stedinger et al.*, 2008] and the use of behavior definitions allows for a broader, multi-criteria assessment of model performance utilizing several objective functions (NSE and PBIAS) [Moriasi et al. 2007].

4.3.4.1. Likelihood Function

For a watershed model M with a vector of p parameters ($\boldsymbol{\theta}$) within the feasible parameter space ($\boldsymbol{\Theta}$) that simulates the response vector of the watershed ($\hat{y}$), the discrete stochastic time-series vector of model residuals is:

$$\varepsilon(\boldsymbol{\theta}) = \hat{y} - y = M(\boldsymbol{\theta}) - y \qquad \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \Re^n \quad (8)$$

By the application of the first-order autoregressive (AR-1) transformation of the residuals can be used to account for correlated errors:

$$\varepsilon_i = \rho\varepsilon_{i-1} + \vartheta_i \qquad i = 1, \dots, n \tag{9}$$

Where ρ is the lag-1 serial correlation coefficient for the residuals ε; $\vartheta \sim N\left(0\,, \sigma_\vartheta^2\right)$ is the innovation term with zero mean and constant variance σ_ϑ^2 and n is the total number of observed data. A proper likelihood function for multiple variables case can be expressed as follows:

$$\ell\ \ (\varepsilon|\boldsymbol{\theta}) = \sum_{j=1}^{m}\left\{-\frac{n_j}{2}\ln(2\pi) - \frac{1}{2}\ln\frac{\sigma_{\vartheta,j}^{2n_j}}{1-\rho_j^2} - \frac{1}{2}\left(1-\rho_j^2\right)\times\sigma_{\vartheta,j}^{-2}\times\left[\hat{y}_{1,j}(\boldsymbol{\theta}) - y_{1,j}\right]^2 - \right.$$
$$\left.\frac{1}{2}\sigma_{\vartheta,j}^{-2}\times\sum_{i=2}^{n_j}\left\{\left(y_{i,j}-\rho_j y_{i-1,j}\right) - \left[\hat{y}_{i,j}(\boldsymbol{\theta}) - \rho_j\hat{y}_{i-1,j}(\boldsymbol{\theta})\right]\right\}^2\right\} \tag{10}$$

Where j is the number of variables; m is the total number of variables; n_j is the total number of observed data for variable j; $\sigma_{\vartheta,j}$ is the standard deviation of residuals for variable j; ρ_j is the lag-1 serial correlation coefficient for the residuals for variable j; The terms $\sigma_{\vartheta,j}$ and ρ_j can be estimated using the Bayesian approach [*Vrugt et al.*, 2009] or can be assigned based on prior knowledge. The value of the likelihood function above will be minimized for all approaches in case studies.

4.3.4.2. Dynamically Dimensioned Search

Dynamically dimensioned search (DDS) [*Tolson and Shoemaker*, 2007] is an automatic calibration algorithm designed for the purpose of solving high dimensional problems. It has been shown that DDS has outstanding performance compares to other commonly used parameter estimation techniques [*Yen et al.*, 2012] in different categories. In this study, DDS is adopted as the sampling technique to explore the role of structural uncertainty in the SWAT model.

4.3.4.3. Model Performance Evaluation Criteria

During the calibration process, proposed parameter sets (candidate parameter sets) are accepted with results (e.g. statistics calculated by simulated output variables with observation data) only in certain level of reasonable behavior. Therefore, various standards in evaluating behavior

conditions are required to identify parameter sets with significant or rational performances (e.g. very good, good, and satisfactory). However, methods which can generate more behavior parameter sets do not guarantee to have better overall performance. The number of behavior parameter sets represents only the parameter sets with better manually assigned statistical (it can also be defined by many other different ways, [*Beck et al.*, 2002]) thresholds and there is still chance that global optimal solution is located outside the behavior region.

Table 4.1 General Performance Ratings

Performance Rating	NSE	PBIAS (%)	
		Streamflow	NOX
Very Good	0.75 < NSE ≤ 1.00	PBIAS < ±10	PBIAS < ±25
Good	0.65 < NSE ≤ 0.75	±10 ≤ PBIAS < ±15	±25 ≤ PBIAS < ±40
Satisfactory	0.50 < NSE ≤ 0.65	±15 ≤ PBIAS < ±25	±40 ≤ PBIAS < ±70
Unsatisfactory	NSE ≤ 0.50	PBIAS ≥ ±25	PBIAS ≥ ±70

NSE: Nash-Sutcliffe efficiency coefficient
PBIAS: Percent bias
NOX: Total nitrate

In this study, the statistical basis of assessing parameter behavior is by the general performance ratings (GPR) [*Moriasi et al.*, 2007] shown in Table 4.1. The GPR is designed for monthly time steps but the calibration of daily streamflow is also applying the same sets of rating groups in case studies.

4.3.5. Bayesian Model Averaging (BMA)

The BMA [*Hoeting et al.*, 1999; *Kass and Raftery*, 1995; *Leamer*, 1978] technique is a standard framework developed to combine models and predictive distributions [*Raftery et al.*, 2005; *Wöhling and Vrugt*, 2008]. According to the law of total probability, the posterior distribution of N different models with given data of observation Y can be written as: [*Hoeting et al.*, 1999]:

$$P(\Delta|Y) = \sum_{n=1}^{N} P(\Delta|M_n, Y)P(M_n|Y) \quad (11)$$

Where Δ is the quantity of prediction, M_n $(n = 1\sim N)$ is the ensemble of implemented model predictions, $P(\Delta|M_n, Y)$ is the posterior probability of M_n (assume it is correct for the

training data [*Raftery et al.*, 2003]) which reveals the manner how M_n fits the training data, $P(\Delta|M_n, Y)$ is the forecast posterior distribution of Δ given prediction quantities from model M_n and observation data Y. As shown in Equation (12), the term of posterior probability $P(M_n|Y)$ sums to one:

$$\sum_{n=1}^{N} P(M_n|Y) = \sum_{n=1}^{N} w_n = 1 \quad (12)$$

Where w_n is the posterior probability of prediction (the one with best solution). Therefore, the posterior probability of prediction can be regarded as weights which represent the contribution of each in favor of predictions. Recently, the BMA has been extended to ensembles of dynamic models where the forecast f_n is associated with a conditional pdf [*Raftery et al.*, 2005]. The BMA predictive model can be expressed as equation (13) where $g_n(\Delta|f_n)$ is the conditional pdf of Δ given f_n.

$$P(\Delta|f_n, \dots, f_N) = \sum_{n=1}^{N} w_n g_n(\Delta|f_n) \quad (13)$$

The assumption of the original form of BMA [*Raftery et al.*, 2005] suggests that the conditional pdf can be approximated by a linear function which is normal distribution centered. As shown in equation (14), the mean of a normally distributed pdf is $a_n + b_n f_n$ with standard deviation σ.

$$\Delta|f_n \sim N(a_n + b_n f_n, \sigma^2) \quad (14)$$

From above, the BMA mean and variance can be written as follows [*Raftery et al.*, 2005]:

$$E(\Delta|f_n, \dots, f_N) = \sum_{n=1}^{N} w_n (a_n + b_n f_n) \quad (15)$$

$$Var(\Delta|f_n, \dots, f_N) = \sum_{n=1}^{N} w_n \left[(a_n + b_n f_n) - \sum_{i=1}^{N} w_i [(a_i + b_i f_i)]\right]^2 + \sum_{n=1}^{N} w_n \sigma_n^2 \quad (16)$$

In Equation (13), it is concerned in cases of discharge and water quality (both observation and simulation error are non-Gaussian) calibration that the BMA method is assumed to have conditional probability distribution to be Gaussian. Data transformation is required to properly

perform the BMA procedure. Therefore, the log-likelihood function of Equation (13) is as follows (assume independence of forecast errors in time and space):

$$L(w_1, \dots, w_N, \sigma^2) = log(\sum_{n=1}^{N} w_n \, g_n(\Delta_{st}|f_{nst})) \quad (17)$$

The values of the weights and variance can be derived by applying maximum likelihood estimation through various optimization techniques such as the Expectation-Maximization (EM) algorithm and the Shuffle Complex Evolution Metropolis (SCEM-UA) algorithm [*Duan et al.*, 2007; *Raftery et al.*, 2005; *Wöhling and Vrugt*, 2008]. In this study, equation (17) will be solved by the EM algorithm.

4.3.6. Brier Score

In this study, the Brier score (BS) is implemented to compare simulation skill of different scenarios [*Ajami et al.*, 2007; *Georgakakos et al.*, 2004]. The Brier score is a quantified scalar measure of model simulation/forecast and has been widely applied in multi-model topics. The base function of BS is written as follows [*Georgakakos et al.*, 2004]:

$$BS = 1 - \frac{1}{N}\sum_{i=1}^{N}(f(i) - o(i))^2 \quad (18)$$

Where $f(i)$ is the frequency of the simulated target event at time step i estimated by the fraction of model simulations which satisfy (larger than) predefined threshold Q^*; $o(i)$ is equals to 1 if observed quantities at time step i is larger than Q^* . Otherwise, $o(i)$ is 0. The higher the BS is the better since BS is a positive oriented measure.

4.4. ECW Case Studies

The Eagle Creek Watershed (ECW) of central Indiana is utilized as the case study watershed. The SWAT model is calibrated for streamflow and total nitrate (NOX) using five test cases with CN method utilizing antecedent soil moisture condition (SCSI), plant evapotranspiration (SCSII), and the two techniques in three different combinations according to BMA weights for

posterior model outputs. The automated calibration technique (DDS) employs an objective function (statistically valid likelihood function) as well as behavioral definitions to search for optimal solutions.

The effectiveness of the five approaches to the CN method was evaluated by the speed of convergence of model simulations to an objective function; the percentage of model simulations meeting multi-criteria threshold values (behavioral definition) for a statistically valid likelihood function; and, for BMA-weighted test cases, the BMA weights assigned to the CN calculation technique. Structural uncertainty between methods is assessed by the inclusion rate of observed data within predictive uncertainty intervals (95% confidence interval) and average width of uncertainty band (spread), and predictive uncertainty under the five approaches for calibration and evaluation periods are visually compared using quantile-quantile (QQ) plots.

4.4.1. Basic Settings of Case Scenarios

In this study, five case studies for the ECW SWAT2005 model are implemented in calibration of daily streamflow and monthly total nitrate to assess model performance under the two methods for calculation of the curve number retention parameter. Streamflow is calibrated on a daily basis for one gauge station (station number 35, Figure 4.1) and a monthly basis for nitrate at four water quality monitoring locations (station numbers 20, 22, 27, 32, Figure 4.1). Calibration and evaluation periods are from 1997~2000 and 2001~2003 respectively. Computational time for each model evaluation (10,000 runs) ranged from 450 to 500 hours (Intel® Core™ 2 Duo CPU E8400 @ 3.00 GHz, 32-bit operating system, Microsoft Windows XP). Model simulations are assessed using an objective function (statistically valid likelihood function) and multi-criteria behavioral definitions.

The first two scenarios calibrate the model (1) utilizing antecedent soil moisture condition [SCSI] and (2) utilizing plant evapotranspiration [SCSII]. Cases 3 4 and 5 subsequently utilize the BMA technique and assign BMA weights to the ECW models calibrated under case (1) and (2): (3) application of the same set of BMA weights for all time series; (4) application of different sets of BMA weights for wet/dry seasons respectively; and (5) application of different sets of BMA weights for warm/cool and wet/dry periods respectively. Calibration and evaluation periods, criteria, and statistics are described in detail below.

4.4.1.1. Scenario 1 – Calibration by applying SCSI

In Scenario 1, the SWAT model calibration is executed by adopting the approach of SCSI (total number of parameters under calibration: 28) where the CN number is calculated by embracing information from antecedent soil moisture conditions. In the SWAT settings, the system parameter ICN (selection of daily curve number calculation method) should be set to the value of 0.

4.4.1.2. Scenario 2 – Calibration by applying SCSII

In Scenario 2, the SWAT model calibration is using the approach of SCSII for surface runoff calculation where the plant evapotranspiration is applied (total number of parameters under calibration: 29). In the SWAT settings, the system parameter ICN should be tuned to the value of 1 (this method is available from the version of SWAT2005) with an extra model parameter CNCOEF included.

4.4.1.3. Scenario 3 – Application of the same BMA weights for all time series (BMAI)

The BMA will be applied on Scenario 3, 4 and 5 where results Scenario 1 and 2 are aggregated by the BMA weights. In Scenario 3, the same set of BMA weights are assigned for all time series

for each gauge station. As shown in Table 4.2, BMA weights for SCSI and SCSII are applied for posterior outputs of discharge and total nitrate for entire calibration period.

Table 4.2 BMA Model Weights for Case 3 (Same Set of Weights for Entire Period)

BMAI	ω @ st.35	ω @ st.32	ω @ st.27	ω @ st.22	ω @ st.20
SCSI	0.5237	0.0204	0.0001	0.0004	0.2086
SCSII	0.4763	0.9796	0.9999	0.9996	0.7914

ω: BMA weight (BMA weights in each gauge station is summed to one)
st.: gauge station number on Figure 4.1

4.4.1.4. Scenario 4 – Application of BMA weights for wet/dry seasons respectively (BMAII)

In Scenario 4, twelve months of a year has been grouped into two categories (wet/dry season) according to the information from National Climatic Data Center [*NOAA*, 2012]. Months with relatively more precipitation will be grouped as wet season (from March to August) and the rest of the year is defined as dry season (from September to the next February). As shown in Table 4.3 and 4.4, two sets of BMA weights are assigned for wet/dry seasons. This approach is designed to enable the BMA having better ability in capturing the advantages from SCSI and SCSII in wet/dry seasons.

Table 4.3 BMA Model Weights for Case 4 (Dry Season)

BMAII Dry Season	ω @ st.35	ω @ st.32	ω @ st.27	ω @ st.22	ω @ st.20
SCSI	0.8636	<0.0001	0.0002	<0.0001	0.0007
SCSII	0.1364	>0.9999	0.9998	1.0000	0.9993

ω: BMA weight (BMA weights in each gauge station is summed to one)
st.: gauge station number on Figure 4.1

Table 4.4 BMA Model Weights for Case 4 (Wet Season)

BMAII Wet Season	ω @ st.35	ω @ st.32	ω @ st.27	ω @ st.22	ω @ st.20
SCSI	0.2688	<0.0001	0.0006	0.0007	0.1879
SCSII	0.7312	>0.9999	0.9994	0.9993	0.8121

ω: BMA weight (BMA weights in each gauge station is summed to one)
st.: gauge station number on Figure 4.1

4.4.1.5. Scenario 5 – Application of BMA weights for wet warm/cool and wet/dry periods respectively (BMAIII)

Instead of using comparative quantity of precipitation to categorize different characters of dry/wet seasons, another approach is also implemented to classify weather conditions. The historical median of the temperature and precipitation data (1950~2004) is set as thresholds to evaluate if it is under warm/cool or wet/dry conditions. In Scenario 5, only wet-warm and dry-warm conditions can be found because the temperature for the calibration period is above historical median constantly. The BMA weights for Scenario 5 are shown in Table 4.5 and 4.6.

Table 4.5 BMA Model Weights for Case 5 (Wet-Warm condition)

BMAIII Wet-Warm	ω @ st.35	ω @ st.32	ω @ st.27	ω @ st.22	ω @ st.20
SCSI	0.5204	0.0002	0.0004	0.0006	0.1029
SCSII	0.4796	0.9998	0.9996	0.9994	0.8971

ω: BMA weight (BMA weights in each gauge station is summed to one)
st.: gauge station number on Figure 4.1

Table 4.6 BMA Model Weights for Case 5 (Dry-Warm condition)

BMAIII Dry-Warm	ω @ st.35	ω @ st.32	ω @ st.27	ω @ st.22	ω @ st.20
SCSI	0.5629	0.2415	0.3381	0.3174	0.3526
SCSII	0.4371	0.7585	0.6619	0.6826	0.6474

ω: BMA weight (BMA weights in each gauge station is summed to one)
st.: gauge station number on Figure 4.1

4.4.2. Results of All Approaches

The first step to characterize improvements in hydrologic and water quality predictions is to investigate the ability in achieving better objective function value (in this study is tempting to minimize the objective function) by applying different rainfall runoff approaches. Second, the behavior definition is implemented to compensate the disadvantage that the objective function may not be able to determine representative watershed behavior according to only one aggregated term. Third, the BMA technique is implemented on Scenario 1 and 2 on both

calibration and evaluation periods with corresponding statistical results and the Brier score is applied to evaluate the performance among different methods. And the last, uncertainty analysis is performed to investigate prediction uncertainty for all case scenarios.

4.4.2.1. Results of objective function versus model evaluations

The overall performance in finding better objective function values by applying SCSI and SCSII is shown in Figure 4.2. In general, better performance was achieved using SCSI method. Additionally, an assessment of convergence speed illustrated that calibration was slightly more computationally efficient under SCSI. In Figure 4.2, SCSII converges slightly slower than SCSI and the two cases are converging to similar values (SCSI: 342.1; SCSII: 343.1) where no significant progress can be found after 4000 model evaluations.

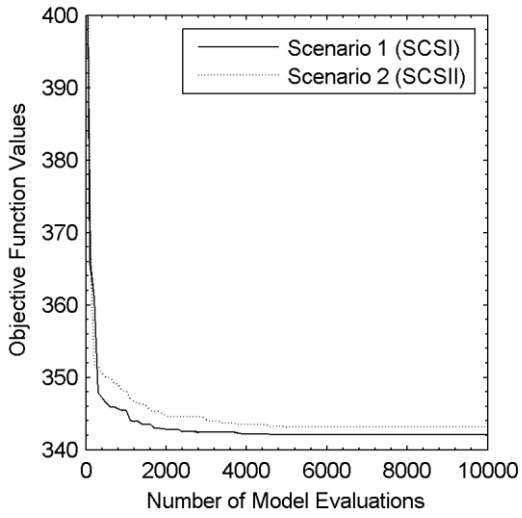

Figure 4.2 Overall performance of objective function versus model evaluations of Scenario 1 (apply SCS method by antecedent soil moisture condition) and Scenario 2 (apply SCS method by plant evapotranspiration)

4.4.2.2. Application of behavior definition

The rate of behavior in the category of "Satisfactory" defined in GPR is shown in Table 4.7. All cases studies are giving high behavior rate (percentage of parameter sets satisfies the predefined

behavior definition) in calibration period (SCSI is the lowest: 67.19% and BMAIII gives the highest: 76.37%). On the other hand, SCSII and BMA associated methods outperform SCSI (only 0.3%) in the Evaluation period but the behavior rate of all cases has significantly declined (behavior rate in Evaluation period for SCSII associated methods varies from 1 to 3%). One of the reasons is that a severe drought has happened in the first year of Evaluation (2001) which makes the simulated data can barely follow the historical patterns. The Second reason is that the SCSI has been known to have poor performance in catching low flow conditions where SCSII is developed to overcome the problem initially.

Table 4.7 Behavior Rate of Parameter Sets of All Case Studies for Calibration and Evaluation Periods

Behavior Rate (%)	SCSI	SCSII	BMAI	BMAII	BMAIII
Calibration	67.19	75.18	75.3	75.93	76.37
Evaluation	0.30	3.01	2.56	1.31	1.04

Behavior rate in Table 4.1 is applying the "Satisfactory" category in GPR

4.4.2.3. Results of statistics and BMA applications

Table 4.8, 4.9, 4.10, 4.11, 4.12 are the statistical results based on NSE and PBIAS (definition of NSE and PBIAS can be found in [*Moriasi et al.*, 2007]). From the results of statistics, one can see that Scenario 2, 3, 4 and 5 are consistently presenting better performance compare to the SCSI case in both calibration and evaluation periods. BMA assigns higher weights to methods resulting in a better match of simulated data with historical observations (see Table 4.2, 4.3, 4.4, 4.5 and 4.6). Since the SCSII method results in a better match of simulated and observed datasets, methods utilizing BMA (BMAI, BMAII and BMAIII) show similar statistics with the exception of streamflow.

Table 4.8 Results of Statistics for the Calibration and Evaluation Period in Scenario 1

SCSI	Stats	st.35	st.32	st.27	st.22	st.20
Calibration (1997 - 2000)	PBIAS (%)	11.94	-6.27	7.69	0.22	21.08
	NSE	0.92	0.54	0.78	0.62	0.93
Evaluation (2001 - 2003)	PBIAS (%)	14.18	-23.09	-1.60	-32.02	20.54
	NSE	0.13	0.30	1.00	-0.35	0.58

Table 4.9 Results of Statistics for the Calibration and Evaluation Period in Scenario 2

SCSII	Stats	st.35	st.32	st.27	st.22	st.20
Calibration (1997 - 2000)	PBIAS (%)	3.39	2.09	13.75	-5.37	21.43
	NSE	0.91	0.85	0.93	0.76	0.95
Evaluation (2001 - 2003)	PBIAS (%)	4.26	-23.73	-5.76	-55.94	5.91
	NSE	0.92	0.26	0.96	-3.14	0.97

Table 4.10 Results of Statistics for the Calibration and Evaluation Period in Scenario 3

BMAI	Stats	st.35	st.32	st.27	st.22	st.20
Calibration (1997 - 2000)	PBIAS (%)	8.16	2.87	14.6	-3.8	22.2
	NSE	0.94	0.85	0.93	0.76	0.95
Evaluation (2001 - 2003)	PBIAS (%)	9.69	-23.8	-5.58	-54.6	8.1
	NSE	0.59	0.26	0.96	-2.93	0.94

Table 4.11 Results of Statistics for the Calibration and Evaluation Period in Scenario 4

BMAII	Stats	st.35	st.32	st.27	st.22	st.20
Calibration (1997 - 2000)	PBIAS (%)	7.85	1.90	13.75	-5.37	21.20
	NSE	0.94	0.85	0.93	0.76	0.95
Evaluation (2001 - 2003)	PBIAS (%)	9.40	-23.69	-5.72	-55.94	8.92
	NSE	0.62	0.27	0.96	-3.14	0.92

Table 4.12 Results of Statistics for the Calibration and Evaluation Period in Scenario 5

BMAIII	Stats	st.35	st.32	st.27	st.22	st.20
Calibration (1997 - 2000)	PBIAS (%)	7.99	2.09	13.75	-5.37	21.20
	NSE	0.94	0.85	0.93	0.76	0.95
Evaluation (2001 - 2003)	PBIAS (%)	9.16	-23.73	-5.76	-55.94	6.77
	NSE	0.64	0.26	0.96	-3.14	0.95

4.4.2.4. Comparisons of model performance by Brier score

Brier score for SCSI, SCSII and BMA associated methods is shown in Figure 4.3(A) and 4.3(B). In Figure 4.3(A), all methods are not showing dominant streamflow predictive over the other

(BMAII is performs relatively better in high flow condition). However, SCSII and BMA associated methods present superior forecast on total nitrate process in Figure 4.3(B). Brier score for other total nitrate stations demonstrate similar results as that in Figure 4.3(B) which are not shown in this script (SCSII and BMA associated methods present superior forecast in simulating total nitrate process).

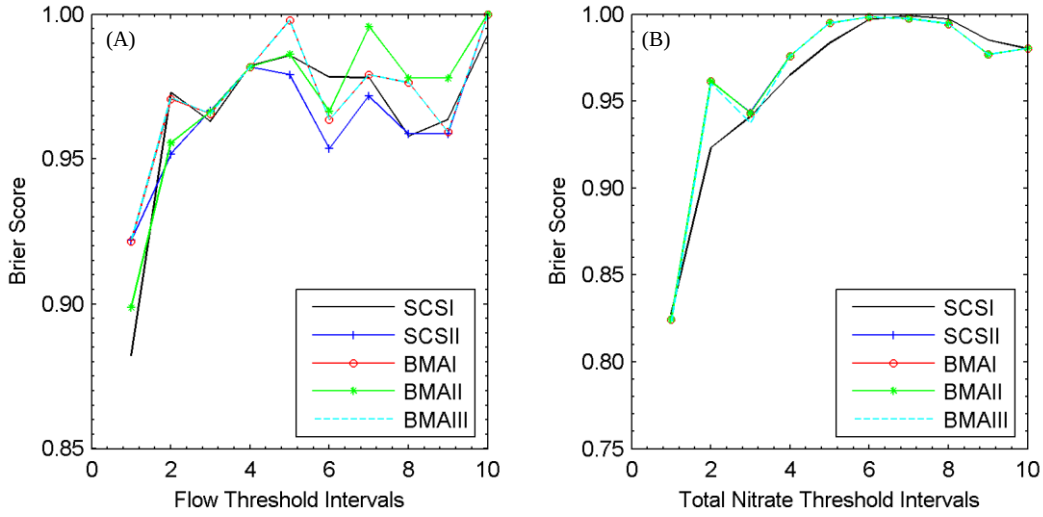

Figure 4.3 Brier score for SCSI, SCSII and BMA associated methods: (A) Streamflow (gauge station 35), (B) Total nitrate (gauge station 32). The flow and the total nitrate threshold intervals are defined by ranking from the highest (interval = 10) to the lowest (interval = 1) quantities of observation data for specific gauge station.

4.4.2.5. Structural uncertainty and predictive quantile-quantile plot

To verify the impact caused by integrating different model structures through the BMA technique, structural uncertainty is explored by constructing/comparing characters of predictive uncertainty bands and other supplemental tool for evaluation purposes such as the quantile-quantile plots is demonstrated.

Table 4.13, 4.14 summarize the effect of cases on model structural uncertainty. Inclusion rate is calculated by counting observation data points locate within 95% of predictive uncertainty intervals and the spread is the average width of uncertainty band.

From Table 4.13 and 4.14, case studies utilizing BMA (Scenario 3, 4 and 5) result in increased uncertainty and a higher inclusion rate in the calibration period (including all streamflow and total nitrate stations). In the evaluation period, the SCSII and BMA associated approaches (Scenario 2) result in higher spread and inclusion rate for streamflow and total nitrate (Scenario 1, 3, 4 and 5 are having similar results). In addition, the average spread of total nitrate stations from Scenario 2, 3, 4 and 5 is narrower than that in Scenario 1 but with higher inclusion rate. In general, the inclusion rate for total nitrate increases with the application of BMA in calibration and Evaluation periods.

Quantile-quantile (QQ) plots are a graphical tool to assess model predictive uncertainty. Instructions as guidance in interpreting QQ plot can be found in literature [*Thyer et al.*, 2009] where the QQ plots for calibration and Evaluation periods are shown in Figure 4.4 and 4.5. As shown in Figure 4.4 for calibration period, most cases are not having significant differences in matching with observation data but slightly overestimating streamflow and total nitrate. SCSI is the only case gives distinct performance compare to other four cases. In Figure 4.5, SCSII and BMA associated cases are showing better match with observation data compare to Scenario 1 except for station 22 and 20. In addition, SCSII and BMA associated methods are generating QQ plots with smoother shape which means more consistent results (less fluctuated CDFs) can be expected from these approaches and the state of over- or under-estimated predictive uncertainty is not as noticeable as SCSI.

Table 4.13 Inclusion Rate of Observed Streamflow/NOX Data Included in 95% Confidence Interval and the Corresponding Spread for the Calibration Period (1997-2000)

Calibration		st.35	st.32	st.27	st.22	st.20
Scenario 1	IR (%)	33.33	18.52	33.33	14.81	33.33
	Spread	0.50	1.00	1.00	0.66	0.68
Scenario 2	IR (%)	29.17	37.04	48.15	29.63	48.15
	Spread	0.48	0.82	0.82	0.86	0.80
Scenario 3	IR (%)	41.67	44.44	51.85	40.74	55.56
	Spread	0.92	1.55	1.57	1.62	1.26
Scenario 4	IR (%)	45.83	48.15	51.85	40.74	55.56
	Spread	0.94	1.57	1.57	1.62	1.36
Scenario 5	IR (%)	41.67	44.44	40.74	29.63	48.15
	Spread	0.92	1.55	1.40	1.24	1.15

Table 4.14 Inclusion Rate of Observed Streamflow/NOX Data Included in 95% Confidence Interval and the Corresponding Spread for the Evaluation Period (2001-2003)

Evaluation		st.35	st.32	st.27	st.22	st.20
Scenario 1	IR (%)	38.89	33.33	27.78	27.78	38.89
	Spread	0.74	2.52	3.01	1.77	4.02
Scenario 2	IR (%)	44.44	61.11	66.67	50.00	50.00
	Spread	1.30	1.80	1.85	1.68	1.78
Scenario 3	IR (%)	44.44	61.11	66.67	50.00	55.56
	Spread	0.81	1.78	1.85	1.68	1.99
Scenario 4	IR (%)	41.67	61.11	66.67	50.00	55.56
	Spread	0.83	1.80	1.85	1.68	1.62
Scenario 5	IR (%)	41.67	55.56	55.56	44.44	61.11
	Spread	0.79	1.57	1.54	1.39	1.66

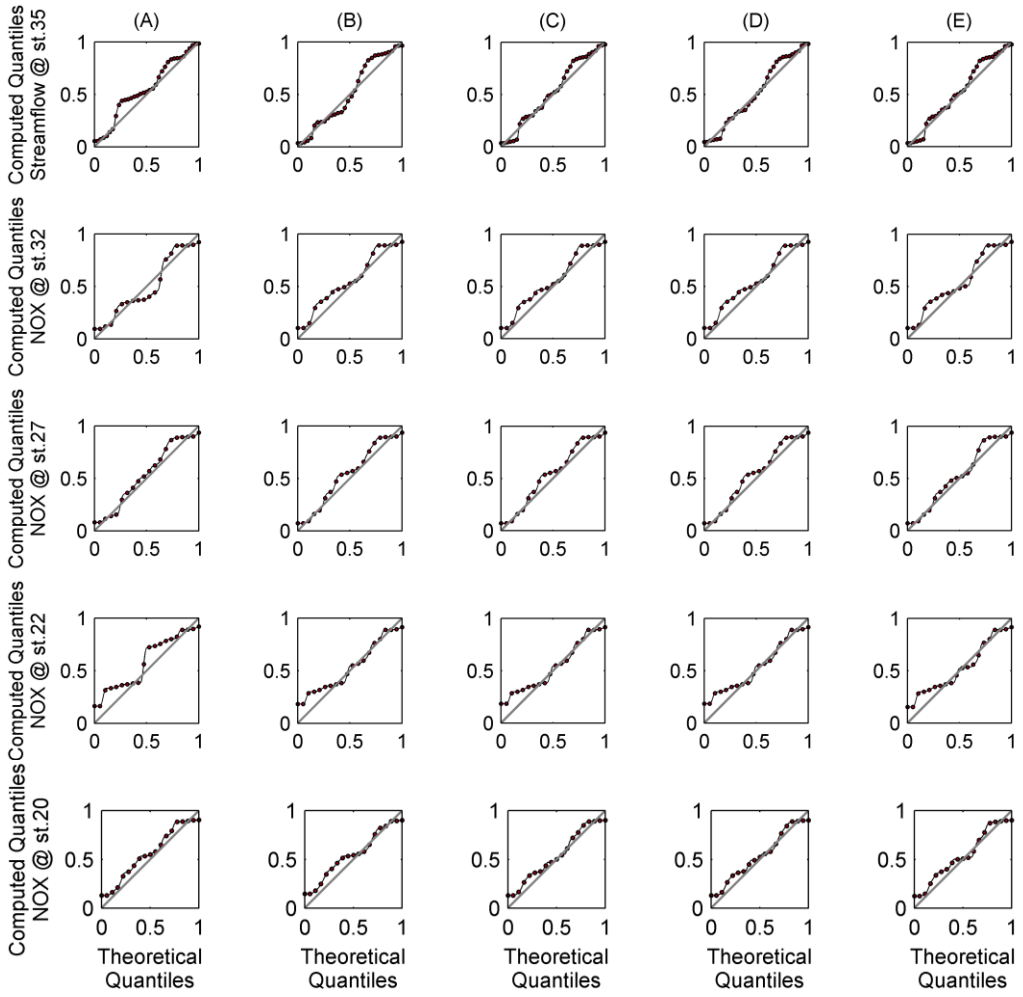

Figure 4.4 Quantile-Quantile plots for calibration periods for all cases: (A) Scenario 1 – SCSI applied, (B) Scenario 2 – SCSII applied, (C) Scenario 3 – Same BMA weights for entire period, (D) Scenario 4 – Different BMA weights for dry/wet seasons, (E) Scenario 5 – Different BMA weights for warm/cool and wet/dry periods respectively.

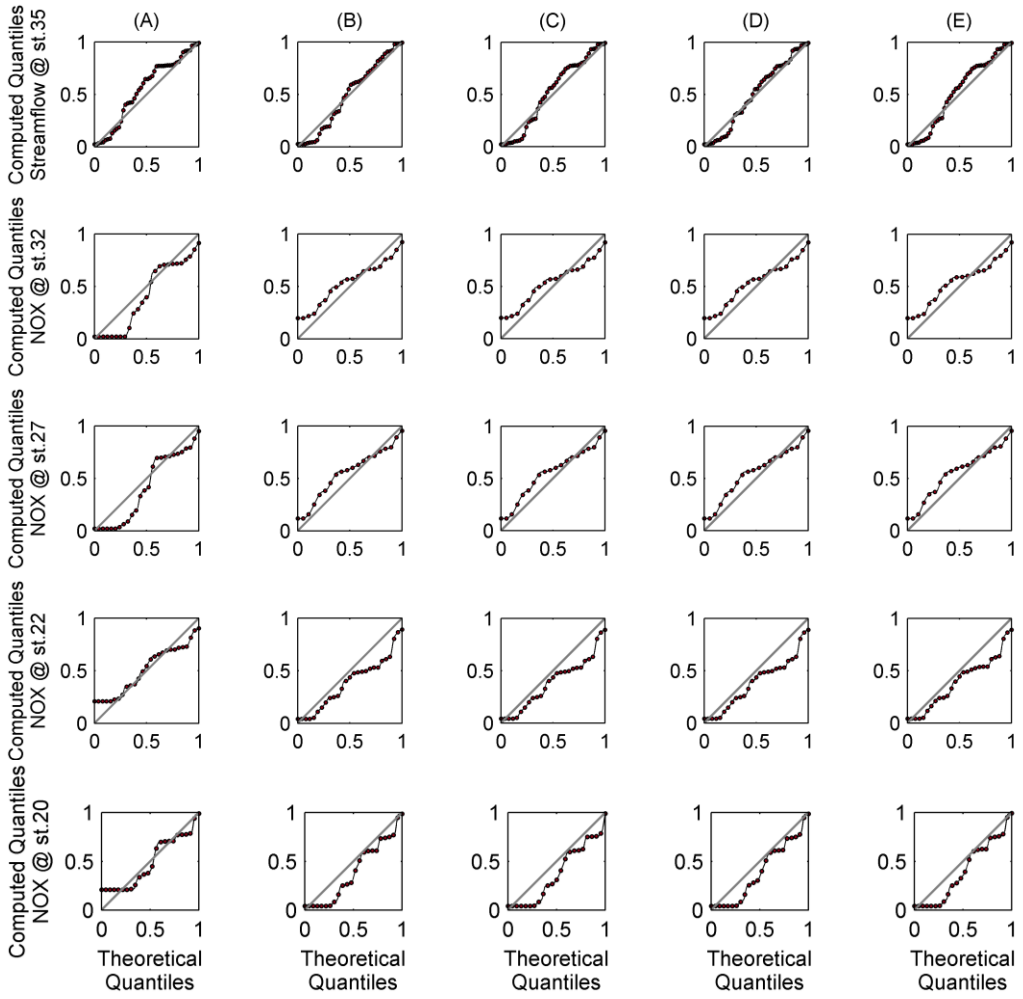

Figure 4.5 Quantile-Quantile plots for evaluation periods for all cases: (A) Scenario 1 – SCSI applied, (B) Scenario 2 – SCSII applied, (C) Scenario 3 – Same BMA weights for entire period, (D) Scenario 4 – Different BMA weights for dry/wet seasons, (E) Scenario 5 – Different BMA weights for warm/cool and wet/dry periods respectively.

4.5. Discussion and Conclusions

From the stand point of view in finding better solutions and the speed of convergence of the objective function, SCSI and SCSII are showing compatible results though the SCSII converges a little slower. The application of a formal likelihood function sustains high inclusion rate in the "Satisfactory" level for calibration but it decreases dramatically under evaluation period. SCSII and BMA associated methods derive more behavioral solutions (by the application of the behavior definition) in both calibration and evaluation periods. It has been explained previously

that a severe draught event happened in the first year of evaluation period where SCSII and BMA associated methods are showing better ability in giving quality solutions under this situation. In other words, SCSI is less capable in properly simulating low flow conditions. Moreover, SCSII and BMA associated approaches are constantly generating solutions with better statistics in both calibration and evaluation periods.

On the other hand, the inclusion rate for total nitrate increases with the incorporation of the BMA technique tells another major benefit of aggregating SCSI and SCSII is to improve better coverage (including more observation data point within uncertainty band) on total nitrate simulation. From Table 4.2, 4.3, 4.4, 4.5 and 4.6, SCSII is assigned dominant weights for total nitrate gauge stations which mean SCSI is not able to simulate total nitrate processes as well as SCSII in practice. In previous studies, SCSII has not yet been applied on water quality (in this study: total nitrate) topics [*Amatya and Jha*, 2011; *Jajarmizadeh et al.*, 2012; *Kannan et al.*, 2007]. Therefore, it is innovative to discover that SCSII (and also BMA applications along with SCSII) has better performance not only on streamflow simulation but also shows significant improvement on one of the water quality issues.

In this study, the role of structural uncertainty on hydrologic and water quality (referencing total nitrate here) predictions from two different approaches of surface runoff calculation is explored by the application BMA. Several conclusions are made by comparisons of different scenarios.

1. SCSI and SCSII do not give significant difference in the speed of convergence and the ability in achieving optimality of the objective function. However, SCSII and BMA associated methods are making superior model predictions with statistical better solutions and higher behavior rate compare to SCSI.

2. SCSI has relatively higher BMA weights on streamflow outputs over SCSII (except for the wet season in Scenario 4) but the ability in simulating streamflow is not perceptibly superior. On the other hand, SCSII outperforms SCSI in simulating total nitrate processes based on error statistics, Brier score and QQ plots. Markedly high BMA weights are assigned to SCSII for total nitrate outputs.
3. Error statistics, predictive uncertainty and patterns of total nitrate processes of BMA associated methods are particularly close to that in SCSII because dominant weights are assigned to SCSII on total nitrate outputs.
4. The approach of classifying seasons into different categories (three different methods are applied in this study) does not show substantial influence on either statistical results or predictive uncertainty. Applications of BMA have positive effect on the increase of inclusion rate but the predictive uncertainty is not evidently reduced/enhanced.

For a complex large-scale watershed simulation model as SWAT, optional alternatives can be applied on different functions for specific hydrologic/nutrient processes (e.g. surface runoff in this study). Advantages and disadvantages ensue with the application of each algorithm can be demonstrably affected especially when methods are designed to catch dissimilar characters or purposes. In this study, the lately proposed method in calculating surface runoff by using information from plant evapotranspiration has been shown to have outstanding performance in both discharge and total nitrate predictions. SCSII originally designed for improving hydrologic simulation can also provide exceptional performance on one of the most important issues in water quality topics. In summary, SWAT model calibration/evaluation work in the future should have more emphasis on the impact from the application of SCSII toward various topics (e.g. pesticide, phosphorus).

References

Ahmadi, M., M. Arabi, J. C. Ascough II, D. G. Fontane, and B. A. Engelc (2012), Toward improved calibration of watershed models: Multisite many objective measures of information, *Submitted to: Advances in Water Resources*.

Ajami, N. K., Q. Duan, and S. Sorooshian (2007), An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction, *Water Resources Research*, *43*(1), 1–19, doi:10.1029/2005WR004745. [online] Available from: http://www.agu.org/pubs/crossref/2007/2005WR004745.shtml (Accessed 28 June 2011)

Amatya, D. M., and M. K. Jha (2011), Evaluating the SWAT model for a low-gradient forested watershed in coastal south Carolina, *Transactions of the ASABE*, *54*(6), 2151–2163.

Arnold, J. G., P. M. Allen, and G. Bernhardt (1993), A comprehensive surface-groundwater flow model, *Journal of Hydrology*, *142*, 47–69.

Arnold, J. G., P. M. Allen, M. Volk, J. R. Williams, and D. D. Bosch (2010), Assessment of Different Representations of Spatial Variability on SWAT Model Performance, *Transactions of the ASABE*, *53*(5), 1433–1443.

Arnold, J. G., R. Srinivasan, R. S. Muttiah, and J. R. Williams (1998), Large area hydrologic modeling and assessment. I. Model development, *Journal of the American Water Resources Association*, *34*(1), 73–89.

Beck, M. B. et al. (2002), Developing a concept of adaptive community learning: Case study of a rapidly urbanizing watershed, *Integrated Assessment*, *3*(4), 299–307, doi:10.1076/iaij.3.4.299.13583. [online] Available from: http://www.tandfonline.com/doi/abs/10.1076/iaij.3.4.299.13583

Beven, K. J., and M. J. Kirkby (1979), A physically based , variable contributing area model of basin hydrology, *Hydrology Sciences Bulletin*, *24*, 43–69.

Borah, D. K., G. Yagow, A. Saleh, P. L. Barnes, W. Rosenthal, E. C. Krug, and L. M. Hauck (2006), Sediment and nutrient modeling for TMDL development and implementation, *Transactions of The ASABE*, *49*(4), 967–986.

Boyle, D. (2001), Multicriteria calibration of hydrological models, *Ph.D. dissertation, Univ. of Ariz., Tucson.*

Burnash, R. J., R. L. Ferral, and R. A. McGuire (1973), A generalized streamflow simulation system conceptual modeling for digital computers, *report, U. S. Dep. of Commer. Natl. Weather Serv. and State of Calif. Dep. of Water Resour.*

Chiang, L., I. Chaubey, M. W. Gitau, and J. G. Arnold (2010), Differentiating impacts of land use changes from pasture management in a CEAP watershed using the SWAT model, *Transactions of the ASABE*, *53*(2), 1569–1584.

Douglas-Mankin, K. R., R. Srinivasan, and J. G. Arnold (2010), Soil and Water Assessment Tool (SWAT) model: current developments and applications, *Transactions of the ASABE*, *53*(5), 1423–1431.

Du, B., J. G. Arnold, A. Saleh, and D. B. Jaynes (2005), Development and application of SWAT to landscapes with tiles and potholes, *Transactions of the ASABE*, *48*(3).

Duan, Q., N. K. Ajami, X. Gao, and S. Sorooshian (2007), Multi-model ensemble hydrologic prediction using Bayesian model averaging, *Advances in Water Resources*, *30*(5), 1371–1386, doi:10.1016/j.advwatres.2006.11.014. [online] Available from:

http://linkinghub.elsevier.com/retrieve/pii/S030917080600220X (Accessed 18 August 2012)

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, *Transactions Of The Asabe*, *50*(4), 1211–1250.

Georgakakos, K., D. Seo, H. Gupta, J. Schaake, and M. Butts (2004), Towards the characterization of streamflow simulation uncertainty through multimodel ensembles, *Journal of Hydrology*, *298*(1-4), 222–241, doi:10.1016/j.jhydrol.2004.03.037. [online] Available from: http://linkinghub.elsevier.com/retrieve/pii/S002216940400246X (Accessed 10 June 2011)

Ghebremichael, L. T., T. L. Veith, and M. C. Watzin (2010), Determination of critical source areas for phosphorus loss: Lake Champlain Basin, Vermont, *Transactions of the ASABE*, *53*(5), 1595–1604.

Green, C. H., M. D. Tomer, M. D. Luzio, and J. G. Arnold (2006), Hydrologic Evaluation if the Soil and Water Assessment Tool for a Large Tile-Drianed Watershed in IOWA, *Transactions of the ASABE*, *49*(2), 413–422.

Green, W. H., and G. A. Ampt (1911), Studies on soil physics, part I, the flow of air and water through soils, *Journal of Agricultural Sciences*, *4*(1), 1–24.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999), Bayesian Model Averaging : A Tutorial, *Statistical Science*, *14*(4), 382–417.

Jajarmizadeh, M., S. Harun, M. H. Mokhtari, and B. Ghahraman (2012), Impact of Behaviour of New accounting curve number method using SWAT model on stream flow, *2012 International Conference on Environment Science and Engineering*, *3*, 124–128.

Jayakrishnan, R., R. Srinivasan, C. Santhi, and J. G. Arnold (2005), Advances in the application of the SWAT model for water resources management, *Hydrological Processes*, *19*(3), 749–762, doi:10.1002/hyp.5624. [online] Available from: http://doi.wiley.com/10.1002/hyp.5624 (Accessed 18 March 2012)

Kannan, N., C. Santhi, J. R. Williams, and J. G. Arnold (2007), Development of a continuous soil moisture accounting procedure for curve number methodology and its behaviour with different evapotranspiration methods, *Hydrological Processes*, *22*(13), 2114–2121, doi:10.1002/hyp.6811. [online] Available from: http://doi.wiley.com/10.1002/hyp.6811 (Accessed 7 May 2012)

Kass, R. E., and A. E. Raftery (1995), Bayes factors, *Journal of the American Statistical Association*, *90*, 773–795.

Kim, N. W., A. H. Shin, and J. Lee (2010), Effects of streamflow routing schemes on water quality with SWAT, *Transactions of the ASABE*, *53*(5), 1457–1468.

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

Leamer, E. E. (1978), Specification Search, *Wiley*, 370 pp.

Meng, H., A. M. Sexton, M. C. Maddox, A. Sood, C. W. Brown, R. R. Ferraro, and R. Murtugudde (2010), Modelling Rappahannock river basin using SWAT – Pilot for Chesapeake Bay watershed, *Applied Engineering in Agriculture*, *26*(5), 795–806.

Moriasi, D. N., J. G. Arnold, B. A. Engel, and C. G. Rossi (2009), Incorporation of a new shallow water table depth algorithm into SWAT2005, *Transactions of the ASABE*, *52*(3), 771–784.

Moriasi, D. N., J. G. Arnold, M. W. V. Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith (2007), Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, *Transactions of the ASABE*, *50*(3), 885–900.

NOAA (2012), Local weather observation station record, NOAA Satellite and Information Service and National Climatic Data Center, *Available at: http://www.ncdc.noaa.gov/oa/ncdc.html*.

Narasimhan, B., R. Srinivasan, J. G. Arnold, and M. D. Luzio (2005), Estimation of long-term soil moisture using a distributed parameter hydrologic model and verification using remotely sensed data, *Transactions of the ASABE*, *48*(3), 1101–1113.

Newman, J. E. (1997), *Our Changing Climate, In: The Natural Heritage of Indiana, Edited M. T. Jackson*.

Nie, W., Y. Yuan, W. Kepner, C. Erickson, and M. Jackson (2012), Hydrological impacts of mesquite encroachment in the upper San Pedro watershed, *Journal of Arid Environments*, *82*, 147–155, doi:10.1016/j.jaridenv.2012.02.008. [online] Available from: http://linkinghub.elsevier.com/retrieve/pii/S0140196312000729 (Accessed 7 May 2012)

Ponce, V. M., and R. H. Hawkins (1996), Runoff Curve Number: Has It Reached Maturity?, *Journal of Hydrologic Engineering*, *1*, 11–19.

Raftery, A. E., F. Balabdaoui, T. Gneiting, and M. Polakowski (2003), Using Bayesian model averaging to calibrate forecast ensembles, *Tech. Rep. 440, Dep. of Stat., Univ. of Wash., Seattle*.

Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski (2005), Using Bayesian Model Averaging to Calibrate Forecast Ensembles, *American Meteorological Society*, 1155–1174.

Rallison, R. E., and N. Miller (1981), Past, present, and future SCS runoff procedure, *In Rainfall Runoff Relationship, 353-364, ed. V. P. Singh. Littleton, Colo.: Water Resources Publication.*

Schaake, J. C., V. Koren, Q. Y. Duan, and F. Chen (1996), Simple water balance model for estimating runoff at different spatial and temporal scales, *Journal of Geophysical Research*, *101(D3)*, 7461–7475.

Schuol, J., K. C. Abbaspour, R. Srinivasan, and H. Yang (2008), Estimation of freshwater availability in the West African sub-continent using the SWAT hydrologic model, *Journal of Hydrology*, *352*(1-2), 30–49, doi:10.1016/j.jhydrol.2007.12.025. [online] Available from: http://linkinghub.elsevier.com/retrieve/pii/S0022169407007755 (Accessed 11 April 2012)

Setegn, S. G., R. Srinivasan, A. M. Melesse, and B. Dargahi (2010), SWAT model application and prediction uncertainty analysis in the Lake Tana Basin, Ethiopia, *Hydrological Processes*, *24*, 357–367, doi:10.1002/hyp.7457. [online] Available from: http://doi.wiley.com/10.1002/hyp.7457 (Accessed 7 May 2012)

Srinivasan, R. X. Zhang, J. A. (2010), Swat Ungaged: Hydrological Budget and Crop Yield Predictions in the Upper Mississippi River Basin, *Transactions of the ASABE*, *53*(5), 1533–1546.

Stedinger, J. R., R. M. Vogel, S. U. Lee, and R. Batchelder (2008), Appraisal of the generalized likelihood uncertainty estimation (GLUE) method, *Water Resources Research*, *44*, 1–17, doi:10.1029/2008WR006822. [online] Available from: http://www.agu.org/pubs/crossref/2008/2008WR006822.shtml (Accessed 22 June 2011)

Thyer, M., B. Renard, D. Kavetski, G. Kuczera, S. W. Franks, and S. Srikanthan (2009), Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis, *Water Resources Research*, *45*(3), 1–22,

doi:10.1029/2008WR006825. [online] Available from: http://www.agu.org/pubs/crossref/2009/2008WR006825.shtml (Accessed 19 July 2011)

Tolson, B. A., and C. A. Shoemaker (2007), Dynamically dimensioned search algorithm for computationally efficient watershed model calibration, *Water Resources Research*, *43*(1), 1–16, doi:10.1029/2005WR004723. [online] Available from: http://www.agu.org/pubs/crossref/2007/2005WR004723.shtml (Accessed 28 June 2011)

USDA-SCS (1972), National Engineering Handbook, Hydrology, Section 4, Chapter 4-10. Washington, D. C.: Soil Conservation Service,

Vrugt, J. A., C. J. F. Braak, V. K. Gupta, and B. A. Robinson (2009), Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?, *Stochastic Environmental Research and Risk Assessment*, *23*(7), 1011–1026, doi:10.1007/s00477-008-0274-y.

Wöhling, T., and J. a. Vrugt (2008), Combining multiobjective optimization and Bayesian model averaging to calibrate forecast ensembles of soil hydraulic models, *Water Resources Research*, *44*(12), 1–18, doi:10.1029/2008WR007154. [online] Available from: http://www.agu.org/pubs/crossref/2008/2008WR007154.shtml (Accessed 5 March 2012)

Yen, H., R. M. Records, M. Arabi, and B. A. Tolson (2012), Appraisal of sampling techniques for parameter estimation and uncertainty analysis in watershed modeling, *Under internal review*.

Chapter 5

A Framework for Propagation of Input, Parameter, Structural and Measurement Uncertainty in Watershed Modeling

5.1. Overview

As a result of the progressive improvement of computer science and the development of auto-calibration techniques, computationally intensive calibration processes of watershed simulation models are no longer a major challenge for watershed planning and management. Scientists are more interested in the decisions made before performing model calibration/evaluation, such as, the possible influence on predictive uncertainty from various sources of error. It has been shown that the assumption that all sources of uncertainty are contributed from model parameters is inadequate. The absence of considering one or more sources of uncertainty may cause biased results and infer the prejudiced conclusion to other sources mistakenly. In this study, a framework is proposed to incorporate uncertainty from input, parameter, structural and measurement sources jointly where the influence from each can be identified and analyzed for the purpose of devoting available resources for further investigation. Results show that input uncertainty reveals relatively greater impact compared to other sources. In addition, predictive uncertainty is significantly enhanced with the inclusion of all four sources of uncertainty

simultaneously. The best statistical solutions tend to satisfy intra-watershed responses in terms of supplementary constraints with the inclusion of uncertainty sources simultaneously. The proposed framework is an innovative tool to investigate and explore the significance of different uncertainty sources individually and jointly. Watershed calibration problems can take advantage of the proposed framework for relevant studies in the future.

5.2. Introduction

Improvements in watershed simulation models enrich their ability to imitate real world phenomenon. Meanwhile, the number of model parameters has also noticeably increased corresponding to empirically or physically based functions which reflect specific interactions in the natural environment [*Bai et al.*, 2009; *Yang et al.*, 2008].

Due to progressive improvements in computer science and the development of auto-calibration techniques, computationally intensive model calibration/evaluation processes (solving high dimensional problems) are no longer a major challenge [*Duan et al.*, 1992; *Tolson and Shoemaker*, 2007; *Vrugt et al.*, 2009a; *Yen et al.*, 2012b]. Scientists now are more interested in whether the assumptions made to perform model simulation/calibration are appropriate or not. For example, in a watershed calibration problem, it is not adequate to assume all sources of uncertainty are contributed from model parameters only [*Ajami et al.*, 2007]. Even if the uncertainty from parameters model inputs, and model structure is considered, error statistics are still being calculated using observation data which has been assumed to be without error. Actually measurement error should be identified and is as important as the other sources of error [*Harmel and Smith*, 2007; *Harmel et al.*, 2006]. The absence of considering one or more sources of uncertainty may cause the final results of a watershed calibration to be biased and yield incorrect conclusions.

Parameter uncertainty is the most studied and a large amount of research can be found in the literature [*Gallagher and Doherty*, 2007; *Hassan et al.*, 2009; *Kuczera and Parent*, 1998; *Loosvelt et al.*, 2011; *Osidele et al.*, 2006]. The parameter-calibration approach [*Ajami et al.*, 2007] assumes that parameter errors are the ultimate attribution of all possible sources. Beyond that, errors due to model inputs have been explored and shown to have significant impact on model calibration [*Ajami et al.*, 2007; *Kavetski et al.*, 2002]. The uncertainty from forcing inputs should not be regarded as a part of error contributed from the model parameters. Studies of model structure error demonstrate the importance of structural uncertainty in model predictions [*Clark et al.*, 2008; *Refsgaard et al.*, 2006]. Structural uncertainty can be explored by aggregating different models using the Bayesian Model Averaging (BMA) technique [*Ajami et al.*, 2007; *Duan et al.*, 2007; *Kavetski et al.*, 2006a, 2006b]. In addition, structural uncertainty may be reduced by the BMA technique where the significance of each implemented model can be represented by BMA weights.

Above all, the role of measurement uncertainty has not dawn as much attention as uncertainty due to input, parameters and model structure. The major reason is because it is difficult to have sufficient data to estimate measurement uncertainty such that it can be evaluated [*Harmel et al.*, 2006]. Some research work can be found [*Harmel and Smith*, 2007; *Harmel et al.*, 2006, 2010] where measurement uncertainty was estimated; however, it was not incorporated with the three other sources of uncertainty.

The main contribution of this study is to incorporate uncertainty from input, parameter, structural and measurement jointly such that the impact from each source can be identified and analyzed. Specifically, the following objectives are: (i) To quantify predictive uncertainty while propagating four sources of error; (ii) To understand the role and importance of four uncertainty

sources on predictive uncertainty, and (iii) To examine the effects of reduction of input, parameter, structural and measurement uncertainty in predictive uncertainty.

5.3. Methods and Materials

5.3.1. Site Location and Characteristics

As shown in Figure 5.1, case studies will be taken place at the Eagle Creek watershed (ECW), Indiana. The ECW (248 km^2 in area) located at the central region of Indiana with four counties (Boone, Hamilton, Hendricks and Marion) is a part of Upper White River watershed. The average annual precipitation is 38 to 40 inches and the average annual temperature is 52°F [*Newman*, 1997]. The majority land use is agricultural activity (agriculture: 59%; rangeland: 38%; forest: 2%; urban: 1%) and the general characters are very close to of the Midwest region. Hydrologic soil type in the ECW can be primarily categorized into two major groups (group B: 51%; group C: 48%). Observation data are available for one discharge (station number 35 in Figure 5.1) and four water quality monitoring gauge stations (station number 20, 22, 27, and 32 in Figure 5.1). The EWC is adopted as the research case study area for two major reasons. First, a sufficient amount of historical data is available for scientists to do relevant research such as the impact on crop production in field from different sources of uncertainty. Second, the ECW is located within a highly developed area for agricultural activities with a large population. Results of model calibration and uncertainty analysis may impact decision makers and stakeholders.

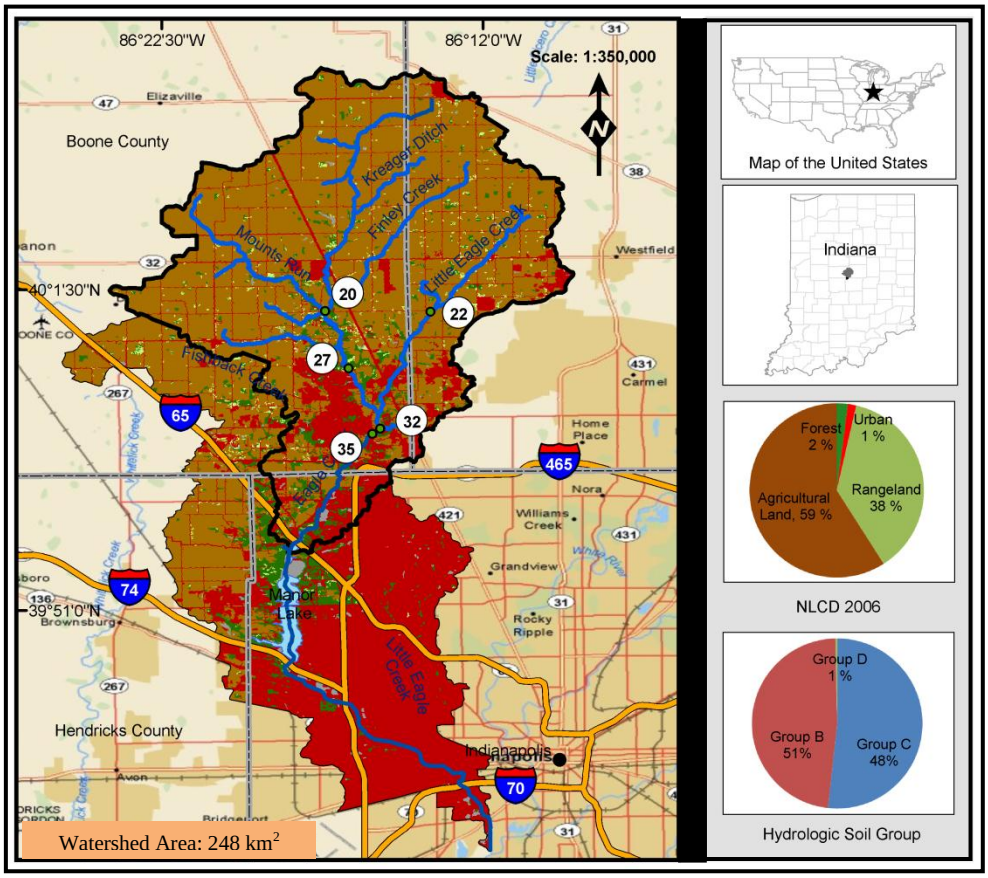

Figure 5.1 Case study area: Eagle Creek Watershed, Indiana

5.3.2. SWAT Model

The Soil and Water Assessment Tool (SWAT) [*Arnold et al.*, 1993] is a continuous-time and semi-distributed parameter model which is developed to simulate/predict hydrologic and water quality processes on a large-scale watershed level. A large amount of research can be found in the field of water resources planning and management regarding the application of the SWAT model [*Arnold et al.*, 2010; *Chiang et al.*, 2010; *Douglas-Mankin et al.*, 2010; *Du et al.*, 2005; *Ghebremichael et al.*, 2010; *Green et al.*, 2006; *Jayakrishnan et al.*, 2005; *Kim et al.*, 2010; *Meng et al.*, 2010; *Moriasi et al.*, 2009; *Srinivasan, R. X. Zhang*, 2010]. In this study, the latest version of the SWAT model (SWAT2009) is used as the watershed simulation model to evaluate

the impact from input, parameter, structural and measurement uncertainty. A detailed description of the SWAT model is presented by *Arnold et al.*, 1998.

5.3.3. Sources of uncertainty

Sources of uncertainty in watershed modeling can be categorized into input, parameter, structural and measurement uncertainty. Input uncertainty includes forcing inputs such as precipitation, temperature, and land use types. Parameter uncertainty represents the uncertainty from physically or empirically based equations with various adjustable model parameters such as curve number, hydraulic conductivity and Manning's roughness (n). Another source of uncertainty in the watershed simulation model comes from the model structure. For example, alternative methods for computing surface runoff, channel erosion and sediment transport may be available within a single watershed simulation model. After performing model simulations, error statistics are calculated by comparing simulated and measured fluxes such as streamflow discharge and nutrient loads. The existence of measurement uncertainty is undeniable but rarely has been studied.

5.3.3.1. Parameter Uncertainty

A hydrologic model can be written as a general form as follows:

$$Y_t^{obs} = Y_t^{sim}(\theta_t\ , x_t\ ) + \varepsilon_t \qquad (1)$$

Where, Y_t^{obs} is the observed data for hydrologic quantities (e.g. rainfall, nitrate concentration, sediment load… etc) at time t; θ_t is the model parameter set at time t(θ_t remains the same during the entire simulation period in most cases); x_t is the given model inputs at time t; $Y_t^{sim}(\theta_t\ , x_t\ )$ is the simulated hydrologic (or water quality) quantities with given θ_t and x_t at time t; the error term ε_t represents the difference between observation and simulation quantities at time t.

Since the model has model parameters (e.g. watershed model calibration), parameter uncertainty is always included as a part of the problem. In cases that all errors are attributed to model parameters, parameter uncertainty is the only and ultimate source. It has been shown that the parameter-calibration approach is not appropriate to use by simply adjusting model parameters [*Ajami et al.*, 2007]. Therefore, other sources of errors should also be considered simultaneously.

5.3.3.2. Input Uncertainty

Few studies can be found that explicitly incorporate input uncertainty in the field of hydrologic modeling [*Ajami et al.*, 2007]. Two exceptional approaches are the Bayesian total error analysis (BATEA) [*Kavetski et al.*, 2002] and the integrated Bayesian uncertainty estimator (IBUNE) [*Ajami et al.*, 2007]. As shown in Equation (2), both methods apply latent variables as explicit terms (random noise from a normal distribution with mean and variance) into the likelihood function.

$$R_i^{adj} = kR_i^{obs} \qquad k \sim N(\theta, \sigma^2) \tag{2}$$

Where, $R_i^{obs}$is the observed quantity at time t ; k is normally distributed random noise with θ mean and σ^2 variance; and $R_i^{adj}$is the adjusted quantity at time t. In BATEA, the value of θ is assumed to be 1 and the value of σ should be predefined for each observed data point which may result in considerable difficulty due to the dimensionality of the problem. In IBUNE, the problem has been solved by assigning θ as another predefined value and applying the same set of θ and σ throughout the model evaluation. Therefore, the number of latent variables decreases to two regardless of the number of observed variables. In this study, the input uncertainty is incorporated into the watershed simulation model (SWAT) by applying the approach from IBUNE. The range of θ is set from 0.9 to 1.1 and $10^{-5}$to $10^{-3}$for σ^2 .

5.3.3.3. Structural Uncertainty

Structural uncertainty has generally been explored by comparing performance among different models [*Ajami et al.*, 2007; *Duan et al.*, 2007; *Wöhling and Vrugt*, 2008]. In recent years, the Bayesian model averaging method [*Hoeting et al.*, 1999] has been widely implemented as a framework to combine models and predictive distributions [*Raftery et al.*, 2005; *Wöhling and Vrugt*, 2008]. A brief introduction of BMA is described as follows.

By the law of total probability, the posterior distribution of N models with given observation Y can be written as follows [*Hoeting et al.*, 1999]:

$$P(\Delta|Y) = \sum_{n=1}^{N} P(\Delta|M_n, Y)P(M_n|Y) \quad (3)$$

Where Δ is the prediction quantity, M_n $(n = 1{\sim}N)$ is the ensemble of implemented model predictions, $P(\Delta|M_n, Y)$ is the posterior probability of M_n (it is assumed to be correct for training data [*Raftery et al.*, 2003]) which reveals how M_n fits the training data, $P(\Delta|M_n, Y)$ is the forecast posterior distribution of Δ given prediction quantities from model M_n and observation data Y. In Equation (4), the posterior probability $P(M_n|Y)$ can be summed to one.

$$\sum_{n=1}^{N} P(M_n|Y) = \sum_{n=1}^{N} w_n = 1 \quad (4)$$

Where, w_n is the posterior probability of the best prediction. The posterior probability of prediction can be considered as weights. The BMA is implemented to be ensembles of dynamic models where the forecast f_n is associated with a conditional pdf [*Raftery et al.*, 2005]. The BMA predictive model can be expressed as Equation (5). In Equation (5), $g_n(\Delta|f_n)$ is the conditional pdf of Δ given f_n.

$$P(\Delta|f_n, \dots, f_N) = \sum_{n=1}^{N} w_n g_n(\Delta|f_n) \quad (5)$$

The assumption of the original version of BMA [*Raftery et al.*, 2005] suggests that the conditional pdf can be estimated by a normal distribution centered at a linear function (from the

original forecast). As shown in Equation (6), the mean of a normally distributed pdf is $a_n + b_n f_n$ with standard deviation σ .

$$\Delta | f_n \sim N(a_n + b_n f_n, \sigma^2) \quad (6)$$

In Equation (5), both observation and simulation error are non-Gaussian for time series of streamflow and water quality. However, the conditional probability distribution of the BMA method is assumed to be Gaussian. Data transformation is required to use the BMA in practice. Therefore, the log-likelihood function of Equation (5) can be written as Equation (7) (assume that forecast errors are independent in time and space):

$$L(w_1, \dots, w_N, \sigma^2) = log(\sum_{n=1}^{N} w_n \, g_n(\Delta_{st} | f_{nst})) \quad (7)$$

The weights can be estimated by different optimization techniques (e.g. the Expectation-Maximization (EM) algorithm and the Shuffle Complex Evolution Metropolis algorithm [*Duan et al.*, 2007; *Raftery et al.*, 2005; *Wöhling and Vrugt*, 2008]). In this study, Equation (7) is solved by the EM algorithm.

In addition, complex watershed simulation models like SWAT include various functions to calculate the same output quantity (e.g. different methods for computing the amount of surface runoff) and the different functions can be investigated to determine structural uncertainty. In this study, the structural uncertainty contributed from two approaches (by antecedent soil moisture condition [SCSI] and plant evapotranspiration [SCSII]) is applied in calculating surface runoff in the SWAT model. Details of the two methods can be found in *Neitsch et al.*, 2011.

5.3.3.4. Measurement Uncertainty

The role of measurement uncertainty is frequently stated in literature but it not often considered as part of model evaluations in the field of hydrologic and water quality modeling [*Harmel and Smith*, 2007]. One of the most important reasons is that an insufficient amount of data is

available to support relevant research and the other is the absence of scientific guidance for evaluating measurement uncertainty [*Harmel and Smith*, 2007]. In this study, the impact of measurement uncertainty will be investigated by recommended thresholds provided from previous research work [*Harmel et al.*, 2006]. The probability distribution method (PD) [*Harmel and Smith*, 2007] is adopted to perform calculations involving measurement uncertainty. The core of the PD method is to assign a correction factor to error statistics for each observed data point. Taking the error between observed and simulated values for an illustration, the adjusted error can be calculated using Equation (8).

$$E_k = \frac{CF_k}{0.5}(Q_k^{obs} - Q_k^{sim}) \tag{8}$$

Where, E_k is the adjusted error between the observed (Q_k^{obs}) and the simulated data (Q_k^{sim}) at time k; and CF_k is the correction factor at time k. The reason why CF_k is divided by 0.5 is because that is the maximum probability of one half of the pdf. Therefore, the maximum value for CF_k is 0.5 as well. Let's assume the measured data is normally distributed. A normal cumulative distribution function can be calculated with known mean and variance. The variance can be computed as follows:

$$\sigma^2 = \begin{cases} \left(\frac{Q_k^{obs} - UB_k^{lower}}{\mu}\right)^2 \\ \left(\frac{UB_k^{upper} - Q_k^{obs}}{\mu}\right)^2 \end{cases} \tag{9}$$

Where, σ is the variance of Q_k^{obs}; μ is the mean which contains a certain amount of the normal probability distribution (e.g. $\mu = \pm 3.9$ represents standard deviation which includes > 99.99% of a normal probability distribution); UB_k^{lower} and UB_k^{upper} are the upper and lower boundaries of every measured data point at time k which can be calculated as follows:

$$UB_k^{upper} = Q_k^{obs} + \frac{PER_k \cdot Q_k^{obs}}{100} \quad (10\text{-}1)$$

$$UB_k^{lower} = Q_k^{obs} - \frac{PER_k \cdot Q_k^{obs}}{100} \quad (10\text{-}2)$$

Where, PER_k is the probable error range recommended by Harmel and Smith (2007) at time k. Details of the PD method can be found in the literature [*Harmel and Smith*, 2007; *Harmel et al.*, 2006, 2010].

5.3.4. Evaluation Statistics

The selection of likelihood function is very important since the distribution of parameter sampling of model errors will not be correctly reflected if the likelihood function is not statistically sound [*Stedinger et al.*, 2008]. The statistical valid likelihood function described in *Ahmadi et al.*, 2012 is adopted as the objective function for the case study in this chapter

5.3.4.1. Objective function

Assume a watershed model M with a vector of p parameters ($\boldsymbol{\theta}$) within the feasible parameter space ($\boldsymbol{\Theta}$) that simulates the response vector of the watershed ($\hat{y}$) as follows:

$$\hat{y} = M(\boldsymbol{\theta}) \quad \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \Re^n \quad (11)$$

The discrete stochastic time-series vector of model residuals can be written as:

$$\varepsilon(\boldsymbol{\theta}) = \hat{y} - y = M(\boldsymbol{\theta}) - y \quad (12)$$

Where y is the vector of observed output response. The goal of calibration procedures is to estimate $\boldsymbol{\theta}$ such that the residuals are as close to zero as possible. The joint posterior probability distribution of the parameters conditioned on the observed response can be expressed using Bayesian statistics as [*Box and Tiao*, 1992]:

$$P(\boldsymbol{\theta}|\boldsymbol{y}) = c \cdot P(\boldsymbol{\theta}) \cdot \ell(\varepsilon|\boldsymbol{\theta}) \quad (13)$$

Where c is a normalization constant, ℓ is called the likelihood function and represents the likelihood of producing model residuals (ε) for a given set of model parameters ($\boldsymbol{\theta}$), while $P(\boldsymbol{\theta})$ denotes the prior probability density function of $\boldsymbol{\theta}$ that is assigned before assimilation of any observed data. Assuming that residuals are normally and independently distributed with mean equal to zero and unknown but constant standard deviation σ_e, the likelihood function ℓ will then take the following form [*Box and Tiao*, 1992]:

$$\ell(\varepsilon|\boldsymbol{\theta}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left[-\frac{(\hat{y}_i(\boldsymbol{\theta})-y_i)^2}{2\sigma_e^2}\right] \tag{14}$$

Where n is the number of observations. Since in most cases the errors in hydrological and water quality modeling are not normally distributed, independent, and homoscedastic, suitable transformations must be applied to account for error characteristics that are not consistent with assumptions made for deriving Equation (14). The first-order autoregressive (AR-1) transformation of the residuals can be used to account for correlated errors:

$$\varepsilon_i = \rho\varepsilon_{i-1} + \vartheta_i \qquad i = 1, \dots, n \tag{15}$$

Where ρ is the lag-1 serial correlation coefficient for the residuals ε, and $\vartheta \sim N\left(0\,, \sigma_\vartheta^2\right)$ is the innovation term with zero mean and constant variance σ_ϑ^2. Sorooshian and Darcup [1980] showed that substituting the AR-1 transformation into the likelihood function of (14) after taking the logarithm (log-likelihood function) yields:

$$\ell^*(\varepsilon|\boldsymbol{\theta}) = -\frac{n}{2}ln(2\pi) - \frac{1}{2}ln\frac{\sigma_\vartheta^{2n}}{1-\rho^2} - \frac{1}{2}(1-\rho^2).\sigma_\vartheta^{-2}[\hat{y}_1(\boldsymbol{\theta}) - y_1]^2 - \frac{1}{2}\sigma_\vartheta^{-2}.\sum_{i=2}^{n}\{(y_i - \rho y_{i-1}) - [\hat{y}_i(\boldsymbol{\theta}) - \rho\hat{y}_{i-1}(\boldsymbol{\theta})]\}^2 \tag{16}$$

The terms σ_ϑ^2 and ρ can be estimated using the Bayesian approach [*Vrugt et al.*, 2009b] or can be assigned based on prior knowledge. A proper likelihood function for multiple variables similar to the single output case is as follows:

$$\ell^*(\varepsilon|\boldsymbol{\theta}) = \sum_{j=1}^{m}\left\{-\frac{n_j}{2}\ln(2\pi) - \frac{1}{2}\ln\frac{\sigma_{\vartheta,j}^{2n_j}}{1-\rho_j^2} - \frac{1}{2}\left(1-\rho_j^2\right)\times\sigma_{\vartheta,j}^{-2}\times\left[\hat{y}_{1,j}(\boldsymbol{\theta}) - y_{1,j}\right]^2 - \right.$$
$$\left.\frac{1}{2}\sigma_{\vartheta,j}^{-2}\times\sum_{i=2}^{n_j}\left\{\left(y_{i,j}-\rho_j y_{i-1,j}\right) - \left[\hat{y}_{i,j}(\boldsymbol{\theta}) - \rho_j\hat{y}_{i-1,j}(\boldsymbol{\theta})\right]\right\}^2\right\} \quad (17)$$

5.3.4.2. Behavior definition for parameter estimation process

During the calibration process, proposed parameter sets are accepted only with results (e.g. statistics calculated by comparing simulated outputs with observed data) within a certain degree of reasonable behavior. A variety of standards in evaluating behavior levels are required to identify parameter sets with significant or rational performance (e.g. very good, good, and satisfactory). However, a method which is able to generate more behavioral parameter sets does not guarantee better overall performance. The number of behavioral parameter sets represents only the parameter sets with better manually assigned statistical values (these can also be defined by many other different ways, [*Beck et al.*, 2002]) and there is still chance that the global optimal solution is located in some other part of the behavior region.

In this study, parameter behavior is assessed by the general performance ratings [*Moriasi et al.*, 2007] shown in Table 5.1. The general performance ratings are designed for monthly time steps but the calibration of daily streamflow also implements the same sets of rating standards in the case study.

Table 5.1 General Performance Ratings

Performance Rating	NSE	PBIAS (%)	
		Streamflow	NOX
Very Good	0.75 < NSE ≤ 1.00	PBIAS < ±10	PBIAS < ±25
Good	0.65 < NSE ≤ 0.75	±10 ≤ PBIAS < ±15	±25 ≤ PBIAS < ±40
Satisfactory	0.50 < NSE ≤ 0.65	±15 ≤ PBIAS < ±25	±40 ≤ PBIAS < ±70
Unsatisfactory	NSE ≤ 0.50	PBIAS ≥ ±25	PBIAS ≥ ±70

NSE: Nash-Sutcliffe efficiency coefficient
PBIAS: Percent bias

5.3.4.3. Intra-watershed responses constraints

In addition to time varying hydrologic/nutrients responses (daily streamflow, monthly nitrate concentration), model outputs can also be summative values (e.g. denitrification and N-nitrate yield from surface or subsurface flow). Statistical best fit in time varying responses may violate summative outputs because the original purpose of optimization process does not consider information about summative quantities. In other words, watershed calibration without reflecting intra-watershed responses may result in parameter sets converging to domains with good statistical values; however with physically meaningless outputs.

In this study, the intra-watershed responses are considered within the calibration process by implementing additional constraints for denitrification and the ratio of nitrate-N losses contributed from subsurface and surface flow. Denitrification will be restricted within the range between 0 to 50 kg/ha and the ratio of N-nitrate losses contributed from SSQ (SSQ Ratio) is constrained to be larger than 60 percent of total N-nitrate losses. The objective function is penalized if the results violate these ranges. Details on the applications of intra-watershed responses constraints can be found in literature [*Yen et al.*, 2012a].

5.3.5. Parameter Estimation Technique

Selection of sampling technique is very important especially in high dimensional problems. The nature of high nonlinearity may cause difficulties (e.g. computational expensive, insufficient amount of behavior solutions) which make the calibration work hard to have significant progress. A well calibrated watershed with good statistical results using a parameter estimation algorithm may not be able to provide sufficient behavior solutions corresponding to watershed behavior. Systematic procedure in evaluating advantage/disadvantage of sampling strategies can be found in Chapter 2.

5.3.5.1. Dynamically dimensioned search

Dynamically dimensioned search (DDS) [*Tolson and Shoemaker*, 2007] is an automatic calibration algorithm designed for the purpose of solving high dimensional problems. It has been shown to have outstanding performance compared to other commonly used parameter estimation techniques. In this study, DDS is adopted as the sampling technique to explore the role of comprehensive watershed behavior measures in watershed simulation calibration.

5.4. Case Study

The framework incorporates four sources of uncertainty implemented in this study is shown in Figure 5.2. Details of each segment have been introduced previously where comparisons among uncertainty sources will be identified in different scenarios. The calibration work for the ECW is described as follows:

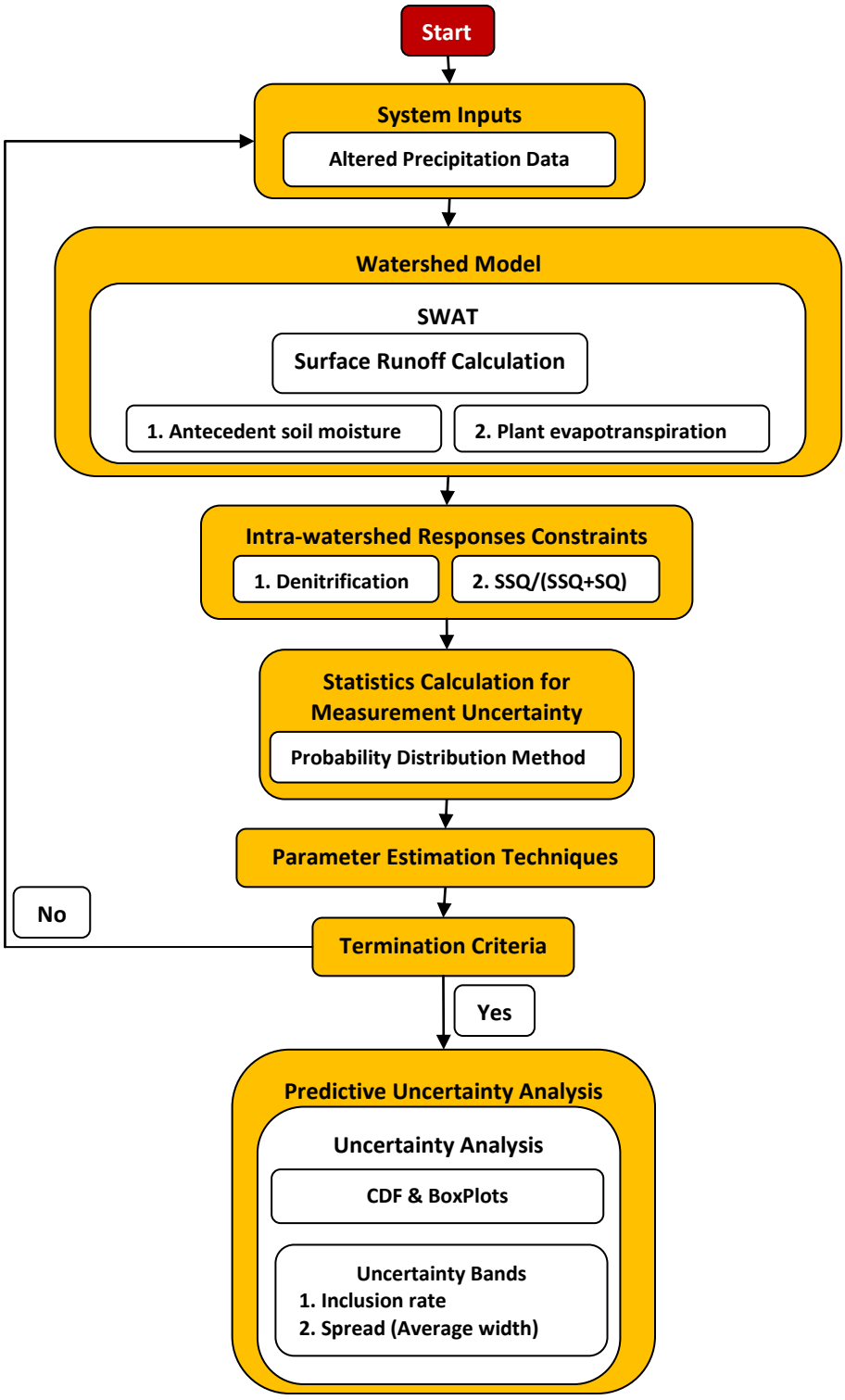

Figure 5.2 Framework of uncertainty analysis incorporating four sources of uncertainty

5.4.1.

5.4.2. Eagle Creek Watershed Calibration

In this study, the SWAT2009 is the watershed simulation model used to simulate the Eagle Creek Watershed. Seven years (1997~2003) of observation data of streamflow (one gauge station) and total nitrate (four gauge stations) are available within 35 subbasins correspond to 446 hydrologic response units. Calibration and evaluation periods are from 1997~2000 and 2001~2003 respectively. Computational time for each model evaluation (10,000 runs) ranged from 450 to 500 hours (Intel® Core™ 2 Duo CPU E8400 @ 3.00 GHz, 32-bit operating system, Microsoft Windows XP).

Table 5.2 Implemented Scenarios and Source(s) of Uncertainty Involved in the Case Study

Scenarios	Source(s) of Uncertainty Involved
I	PU (SCSI)
II	PU (SCSI) + IU
III	PU (SCSII)
IV	PU (SCSI) + MU
V	PU (SCSI) + IU + MU
VI	PU (SCSII) + IU + MU
VII	PU (SCSI) + IU + MU + IWB
VIII	PU (SCSII) + IU + MU + IWB
BMAI	V + VI
BMAII	VII + VIII

PU: Parameter uncertainty
IU: Input uncertainty
MU: Measurement uncertainty
IWB: Intra-watershed responses constraints applied

As shown in Table 5.2, ten scenarios are implemented to calibrate monthly streamflow and monthly total nitrate in this study: (1) calibration by including parameter uncertainty using SCSI in calculating surface runoff [Scenario I], (2) calibration by including parameters and input uncertainty using SCSI in calculating surface runoff [Scenario II], (3) calibration by including parameter uncertainty using SCSII in calculating surface runoff [Scenario III], (4) calibration by including parameter and measurement uncertainty using SCSI in calculating surface runoff

[Scenario IV], (5) calibration by including parameter, input, measurement uncertainty using SCSI in calculating surface runoff [Scenario V], (6) calibration by including parameter, input, measurement uncertainty using SCSII in calculating surface runoff [Scenario VI], (7) calibration using the same settings as in Scenario V with the application of intra-watershed responses constraints [Scenario VII], (8) calibration using the same settings as in Scenario VI with the application of intra-watershed responses constraints [Scenario VIII], (9) application of BMA to Scenario V and VI [Scenario IX], (10) application of BMA to Scenario VII and VIII [Scenario X]. A total of 28 model parameters are included in the calibration process. However, scenarios with input uncertainty have two more parameters (two latent variables), and scenarios with SCSII have one additional model parameter (CNCOEF).

5.4.3. Results of All Scenarios

Watershed calibration begins by finding optimal solutions for parameters with respect to a specific objective function (a minimizing objective function in this study) by comparing simulation and observation outputs. The second step is to further investigate the statistical results by applying a behavior definition to evaluate the quality of optimal solutions. Intra-watershed responses are controlled by additional constraints in Scenario VII and VIII. The final step is to explore the influence of predictive uncertainty caused by different sources of uncertainty.

5.4.3.1. Results of objective function versus model evaluations

The overall performance in achieving better objective function values of all scenarios is shown in Figure 5.3. Scenario II had the poorest performance and Scenario V are had the best performance. All case scenarios converged within 5000 model evaluations and the patterns of convergence are similar in general.

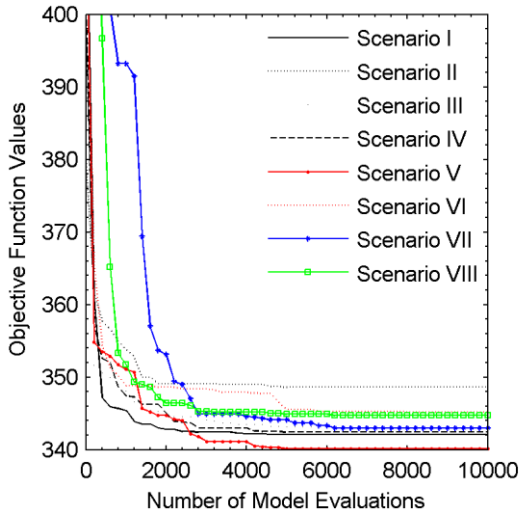

Figure 5.3 Overall performance of objective function values versus model evaluations for all cases (Figure 5.3 is truncated at the beginning of model evaluation is because the initial results for several scenarios are multiple times higher than other scenarios.)

5.4.3.2. Application of behavior definition

The behavior rate of the scenarios using the category of "Satisfactory" defined in the GPR is shown in Table 5.3. Case studies with only one source of uncertainty included (except for parameter uncertainty) show a relatively higher behavior rate in the calibration period compared to scenarios with all sources of uncertainty included. In addition, Scenarios V and VI, without applying watershed behavior constraints, have a lower behavior rate than Scenario VII and VIII where intra-watershed responses constraints are applied. In the evaluation period, the behavior rate is typically low because the first year of evaluation (2001) had a severe drought event where simulated outputs have difficulty matching the patterns of observed data.

5.4.3.3. BMA Weights

Two different approaches are applied in the SWAT model to perform the surface runoff calculation. Therefore, the BMA has been implemented to investigate predictive uncertainty. Table 5.4 (BMAI represents Scenario V and VI) and 5.5 (BMAII represents Scenario VII and

VIII) show the BMA weights of scenarios which include all source of uncertainty simultaneously. From Table 5.4 and 5.5, not all total nitrate stations are assigned dominant weights, compared to the results from Chapter 4 where all total nitrate stations have dominant weights for SCSII associated scenarios. This shows that the inclusion of all sources of uncertainty has not only changed the pattern calibration processes but also altered the priority of selection from different components in structure (e.g. SCSI and SCSII).

Table 5.3 Behavior Rate of Parameter Sets of All Case Studies for Calibration and Evaluation Periods

Behavior Rate (%)	SI	SII	SIII	SIV	SV	SVI	SVII	SVIII	BMAI	BMAII
Calibration	67.19	69.06	75.18	72.41	43.27	21.34	68.47	47.80	37.65	69.68
Evaluation	0.30	0.00	3.01	8.31	4.82	0.08	0.11	0.47	0.35	0.18

Behavior rate in Table 5.1 is applying the "Satisfactory" category in GPR

Table 5.4 BMA Weights for BMAI

BMAI	ω @ st.35	ω @ st.32	ω @ st.27	ω @ st.22	ω @ st.20
Scenario V	0.7286	0.1616	0.1646	0.1344	0.9990
Scenario VI	0.2714	0.8384	0.8354	0.8656	0.0010

ω: BMA weight (BMA weights in each gauge station is summed to one)
st.: gauge station number on Figure 5.1

Table 5.5 BMA Weights for BMAII

BMAII	ω @ st.35	ω @ st.32	ω @ st.27	ω @ st.22	ω @ st.20
Scenario VII	0.3536	0.3158	0.2602	0.5138	0.4879
Scenario VIII	0.6464	0.6842	0.7398	0.4862	0.5121

ω: BMA weight (BMA weights in each gauge station is summed to one)
st.: gauge station number on Figure 5.1

5.4.3.4. Results of statistics and BMA applications

Table 5.6, 5.7, 5.8, 5.9 show the error statistics of Nash-Sutcliffe efficiency coefficient (NSE) and percent bias (PBIAS) (definition of NSE and PBIAS can be found in [*Moriasi et al.*, 2007]) corresponding to the best solution of the objective function. In Table 5.6 and 5.7, intra-watershed responses constraints are not considered where statistics are not showing noticeable differences in calibration period for all case scenarios. However, the results for total nitrate stations show

moderate degeneration for both NSE and PBIAS in the evaluation period. In Table 5.8 and 5.9, the best solutions are selected only when both intra-watershed responses constraints are satisfied. Statistics under the additional requirements significantly declined in scenarios with only one source of uncertainty included (except for parameter uncertainty). Results of best objective function values and the corresponding watershed behavior outputs (denitrification and the ratio of N-nitrate losses yield from subsurface and surface flow) are shown in Table 5.10 and 5.11. The best solutions based on the objective function values for the first four scenarios are comparable with scenarios with all sources of uncertainty included. However, at least one watershed behavior constraint is violated. By considering both criteria (objective function value and watershed behavior constraint), the best objective function values decline notably. The same situation also happens during the evaluation period; however, this is not shown in this chapter.

Table 5.6 Results of Statistics for the Calibration Period*

Calibration (1997 - 2000)	Stats	st.35	st.32	st.27	st.22	st.20
Scenario I	PBIAS (%)	11.94	-6.27	7.69	0.22	21.08
	NSE	0.92	0.54	0.78	0.62	0.93
Scenario II	PBIAS (%)	-0.57	19.35	-3.02	10.66	-2.47
	NSE	0.87	0.85	0.59	0.83	0.76
Scenario III	PBIAS (%)	3.39	2.09	13.75	-5.37	21.43
	NSE	0.91	0.85	0.93	0.76	0.95
Scenario IV	PBIAS (%)	-10.80	13.76	-8.81	2.70	-6.29
	NSE	0.85	0.92	0.70	0.92	0.89
Scenario V	PBIAS (%)	9.34	-4.31	-22.92	-14.13	-32.20
	NSE	0.84	0.85	0.56	0.64	0.42
Scenario VI	PBIAS (%)	9.39	32.30	21.01	8.39	-5.09
	NSE	0.80	0.78	0.71	0.94	0.91
Scenario VII	PBIAS (%)	3.64	11.14	-18.27	2.81	-8.02
	NSE	0.88	0.95	0.67	0.87	0.79
Scenario VIII	PBIAS (%)	0.25	0.16	-29.42	-0.96	-21.61
	NSE	0.78	0.79	0.50	0.69	0.62
BMAI	PBIAS (%)	16.19	-11.67	6.14	15.35	-6.77
	NSE	0.92	0.87	0.94	0.77	0.85
BMAII	PBIAS (%)	6.81	-18.13	-0.86	-23.70	6.73
	NSE	0.93	0.73	0.77	0.64	0.90

*: Actual watershed behavior is not considered in this table.
st.: gauge station number on Figure 5.1

Table 5.7 Results of Statistics for the Evaluation Period*

Evaluation (2001 - 2003)	Stats	st.35	st.32	st.27	st.22	st.20
Scenario I	PBIAS (%)	14.18	-23.09	-1.60	-32.02	20.54
	NSE	0.13	0.30	1.00	-0.35	0.58
Scenario II	PBIAS (%)	15.69	42.39	8.12	30.22	18.97
	NSE	0.76	0.15	-0.09	0.10	0.09
Scenario III	PBIAS (%)	4.26	-23.73	-5.76	-55.94	5.91
	NSE	0.92	0.26	0.96	-3.14	0.97
Scenario IV	PBIAS (%)	0.04	46.93	29.39	34.76	26.88
	NSE	0.84	0.20	0.32	0.26	0.35
Scenario V	PBIAS (%)	9.34	-4.31	-22.92	-14.13	-32.20
	NSE	0.84	0.85	0.56	0.64	0.42
Scenario VI	PBIAS (%)	9.39	32.30	21.01	8.39	-5.09
	NSE	0.80	0.78	0.71	0.94	0.91
Scenario VII	PBIAS (%)	3.64	11.14	-18.27	2.81	-8.02
	NSE	0.88	0.95	0.67	0.87	0.79
Scenario VIII	PBIAS (%)	0.25	0.16	-29.42	-0.96	-21.61
	NSE	0.78	0.79	0.50	0.69	0.62
BMAI	PBIAS (%)	16.19	-11.67	6.14	15.35	-6.77
	NSE	0.92	0.87	0.94	0.77	0.85
BMAII	PBIAS (%)	6.81	-18.13	-0.86	-23.70	6.73
	NSE	0.93	0.73	0.77	0.64	0.90

*: Actual watershed behavior is not considered in this table.
st.: gauge station number on Figure 5.1

Table 5.8 Results of Statistics for the Calibration Period**

Calibration (1997 - 2000)	Stats	st.35	st.32	st.27	st.22	st.20
Scenario I	PBIAS (%)	22.31	-29.49	-25.35	-12.40	-17.13
	NSE	0.55	0.23	0.51	0.54	0.69
Scenario II	PBIAS (%)	-73.89	42.39	8.12	30.22	18.97
	NSE	-0.51	0.15	-0.09	0.10	0.09
Scenario III	PBIAS (%)	13.18	19.13	7.11	-22.74	-19.13
	NSE	0.64	0.41	0.62	0.43	0.63
Scenario IV	PBIAS (%)	25.61	-27.87	-97.86	-49.35	-81.87
	NSE	0.78	0.16	-5.01	-1.10	-2.76
Scenario V	PBIAS (%)	-0.98	22.46	-0.41	10.04	-3.42
	NSE	0.79	0.71	0.68	0.77	0.71
Scenario VI	PBIAS (%)	2.42	34.42	-0.09	18.99	3.70
	NSE	0.72	0.74	0.65	0.76	0.64
Scenario VII	PBIAS (%)	4.65	5.34	-31.96	1.33	-10.51
	NSE	0.81	0.67	0.54	0.64	0.69
Scenario VIII	PBIAS (%)	-1.99	31.90	-8.45	29.51	19.97
	NSE	0.75	0.76	0.44	0.75	0.77
BMAI	PBIAS (%)	-0.04	2.49	17.42	0.05	22.56
	NSE	0.96	0.91	0.60	1.00	0.50
BMAII	PBIAS (%)	0.37	10.39	22.12	-20.25	18.83
	NSE	0.94	0.86	0.35	0.46	0.65

**: Actual watershed behavior constraints are implemented in this table.
st.: gauge station number on Figure 5.1

Table 5.9 Results of Statistics for the Evaluation Period **

Evaluation (2001 - 2003)	Stats	st.35	st.32	st.27	st.22	st.20
Scenario I	PBIAS (%)	35.86	-172.20	-288.30	-210.00	-273.40
	NSE	0.68	-7.06	-32.44	-15.16	-24.43
Scenario II	PBIAS (%)	26.89	-20.60	-118.10	-62.67	-109.50
	NSE	0.72	0.01	-9.65	-3.30	-7.56
Scenario III	PBIAS (%)	20.16	-44.66	-132.00	-64.44	-99.05
	NSE	0.80	-0.18	-7.76	-1.89	-4.09
Scenario IV	PBIAS (%)	25.61	-27.87	-97.86	-49.35	-81.87
	NSE	0.78	0.16	-5.01	-1.10	-2.76
Scenario V	PBIAS (%)	9.34	-4.31	-22.92	-14.13	-32.20
	NSE	0.84	0.85	0.56	0.64	0.42
Scenario VI	PBIAS (%)	9.39	32.30	21.01	8.39	-5.09
	NSE	0.80	0.78	0.71	0.94	0.91
Scenario VII	PBIAS (%)	3.64	11.14	-18.27	2.81	-8.02
	NSE	0.88	0.95	0.67	0.87	0.79
Scenario VIII	PBIAS (%)	0.25	0.16	-29.42	-0.96	-21.61
	NSE	0.78	0.79	0.50	0.69	0.62
BMAI	PBIAS (%)	16.19	-11.67	6.14	15.35	-6.77
	NSE	0.92	0.87	0.94	0.77	0.85
BMAII	PBIAS (%)	6.81	-18.13	-0.86	-23.70	6.73
	NSE	0.93	0.73	0.77	0.64	0.90

**: Actual watershed behavior constraints are implemented in this table.
st.: gauge station number on Figure 5.1

Table 5.10 Results of Best Objective Function Values and the Corresponding Watershed Behavior Outputs for the Calibration Period*

Scenarios	Best Objective Function value	Denitrification (kg/ha)	SSQ Ratio
Scenario I	342.1	121.4	0.98
Scenario II	348.6	214.2	0.97
Scenario III	343.1	243.3	0.96
Scenario IV	342.4	211.4	0.98
Scenario V	340.1	36.1	0.96
Scenario VI	345.2	14.5	0.63
Scenario VII	343.0	36.1	0.92
Scenario VIII	344.7	49.8	0.79

*: Actual watershed behavior is not considered in this table.

Table 5.11 Results of Best Objective Function Values and the Corresponding Watershed Behavior Outputs for the Calibration Period**

Scenarios	Best Objective Function value	Denitrification (kg/ha)	SSQ Ratio
Scenario I	399.6	16.9	0.94
Scenario II	379.4	30.4	0.99
Scenario III	373.1	7.1	0.98
Scenario IV	399.6	49.7	0.98
Scenario V	340.1	36.1	0.96
Scenario VI	345.2	14.5	0.63
Scenario VII	343.0	36.1	0.92
Scenario VIII	344.7	49.8	0.79

**: Actual watershed behavior constraints are implemented in this table.

5.4.3.5. Uncertainty analysis

Table 5.12 summarizes the information for predictive uncertainty for all scenarios. Inclusion rate is computed by counting data points of observed data located within 95% of the predictive uncertainty intervals. The spread is the average width of the corresponding uncertainty band.

From the first four scenarios in Table 5.12, Scenario II shows wider spread of uncertainty band compare to other three scenarios. However, it only stands when input uncertainty is bound with and parameter uncertainty which cannot be entirely separated for all scenarios.

Table 5.12 Inclusion Rate of Observed Streamflow/NOX Data Included within 95% Confidence Interval and the Corresponding Spread for the Calibration Period (1997-2000)

Calibration		st. 35	st. 32	st. 27	st. 22	st. 20
Scenario I	IR (%)	33.33	18.52	33.33	14.81	33.33
	Spread	0.50	1.00	1.00	0.66	0.68
Scenario II	IR (%)	37.50	37.04	48.15	29.63	44.44
	Spread	0.52	0.99	0.94	1.12	0.66
Scenario III	IR (%)	29.17	37.04	48.15	29.63	48.15
	Spread	0.48	0.82	0.82	0.86	0.80
Scenario IV	IR (%)	25.00	48.15	44.44	40.74	55.56
	Spread	0.41	0.77	0.73	0.80	0.70
Scenario V	IR (%)	45.83	48.15	44.44	51.85	59.26
	Spread	0.67	1.78	1.59	2.35	1.33
Scenario VI	IR (%)	25.00	51.85	48.15	44.44	51.85
	Spread	0.47	41.19	49.76	43.38	22.79
Scenario VII	IR (%)	20.83	48.15	51.85	66.67	48.15
	Spread	0.53	1.24	1.22	1.25	1.18
Scenario VIII	IR (%)	33.33	44.44	44.44	51.85	44.44
	Spread	0.54	6.75	12.02	1.54	1.55
BMAI	IR (%)	20.83	25.93	29.63	22.22	40.74
	Spread	0.42	41.74	45.46	40.90	1.55
BMAII	IR (%)	20.83	29.63	37.04	18.52	18.52
	Spread	0.43	5.01	9.33	1.24	1.20

**: Actual watershed behavior constraints are implemented in this table.
st.: gauge station number on Figure 5.1

The average width of the uncertainty band increases dramatically when all sources of uncertainty are included simultaneously (Scenario V, VI, VII and VIII). For Scenario V and VI, the spread of the uncertainty band is multiple times wider than the spread for Scenario I, II, III and IV. In addition, the spread in Scenario VI is wider than that produced by Scenario V. The choice of functions in calculating surface runoff impacts predictive uncertainty. Predictive uncertainty decreased when watershed behavior constraints were used in Scenario VII and VIII. The incorporation of extra information in terms of constraints from the field not only improves the quality of calibrated solutions but also reduces predictive uncertainty.

From the results of predictive uncertainty, the spread of BMAI and BMAII is primarily affected by the BMA weights. However, the inclusion rate deceased compared to the corresponding scenarios (e.g. BMAI is the aggregation of Scenario V and VI). BMA applications show the lowest inclusion rate of all cases studies at both streamflow and total nitrate stations.

5.4.3.6. Additional information of intra-watershed responses constraints

As mentioned previously, the adoption of actual watershed behavior has direct influence on the quality of statistical results, behavior rate and predictive uncertainty. From Table 5.10 and 5.11, the best statistical results from Scenario V and VI (incorporation of all sources of uncertainty without considering actual watershed behavior) satisfy additional constraints added in Scenario VII and VIII automatically for this case study. In Table 5.13, Scenario V yields results which satisfy actual watershed behavior constraints even more than Scenario VII. As illustrated in Figure 5.4, denitrification converged to the value of 37 (kg/ha) in Scenario V but it does not show any improvement in 10,000 model evaluations and converged to 115 (kg/ha). Specifically, the CDFs of denitrification are shown in Figure 5.5(A) and 5.5(B) (take denitrification as example). None of the first four scenarios tend to search for solutions that satisfy actual watershed behavior constraints. On the other hand, scenarios with all sources of uncertainty have more reasonable range.

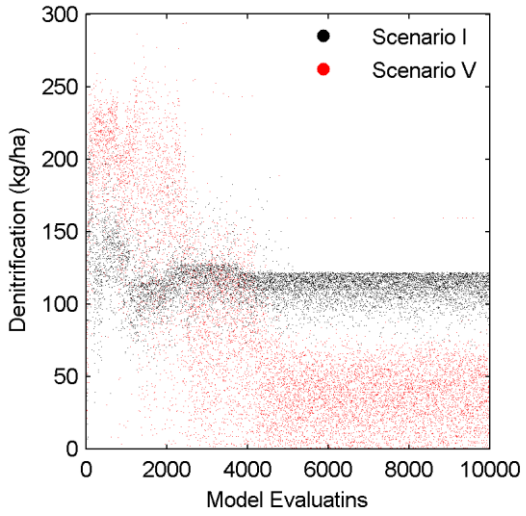

Figure 5.4 Convergence process of denitrification in Scenario I and V

Table 5.13 Results that Satisfy Both Actual Watershed Behavior Constraints (%)

Scenarios	% of Evaluations
Scenario I	0.18
Scenario II	0.28
Scenario III	0.24
Scenario IV	0.50
Scenario V	47.81
Scenario VI	13.48
Scenario VII	27.81
Scenario VIII	46.68

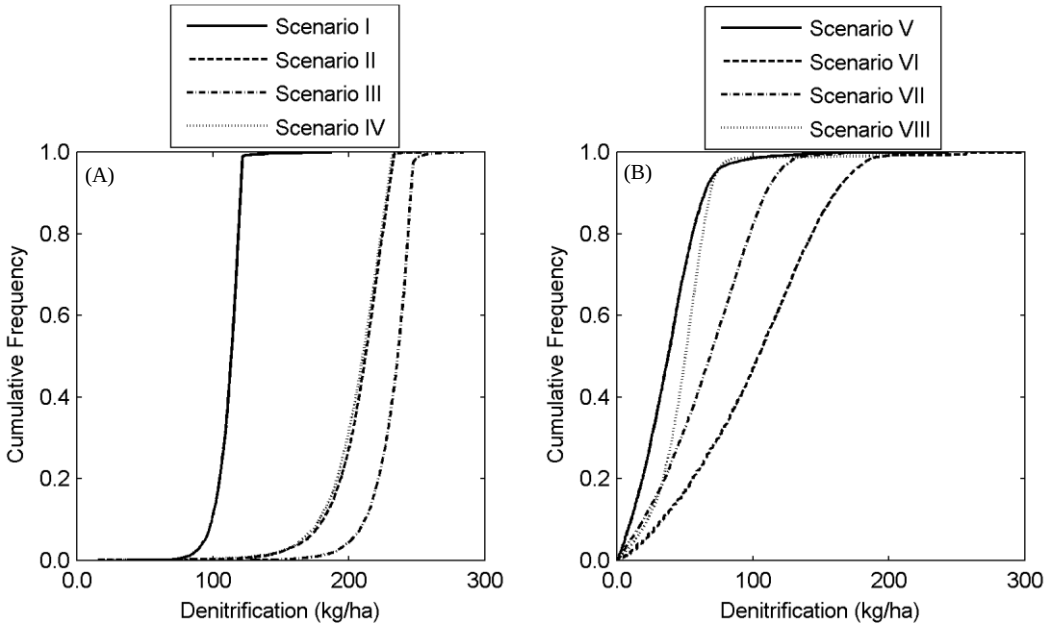

Figure 5.5 Cumulative distribution functions of the denitrification constraint applied in all case scenarios: (A) Scenario I~IV with only one source of uncertainty included (except for parameter uncertainty), (B) Scenario V~VIII with all sources of uncertainty included

5.4.3.7. Latent variables of input Uncertainty

In Figure 5.6(A), 5.6(B), 5.6(C) and 5.6(D), values for the latent variable nu range from 1.00 to 1.05 and the latent variable sigma is generally less than $5x10^{-4}$. The average values of latent variables are higher than 1 which means the original precipitation data needed to be adjusted to higher values. The results of latent variables show that the observed data of precipitation is possibly being underestimated.

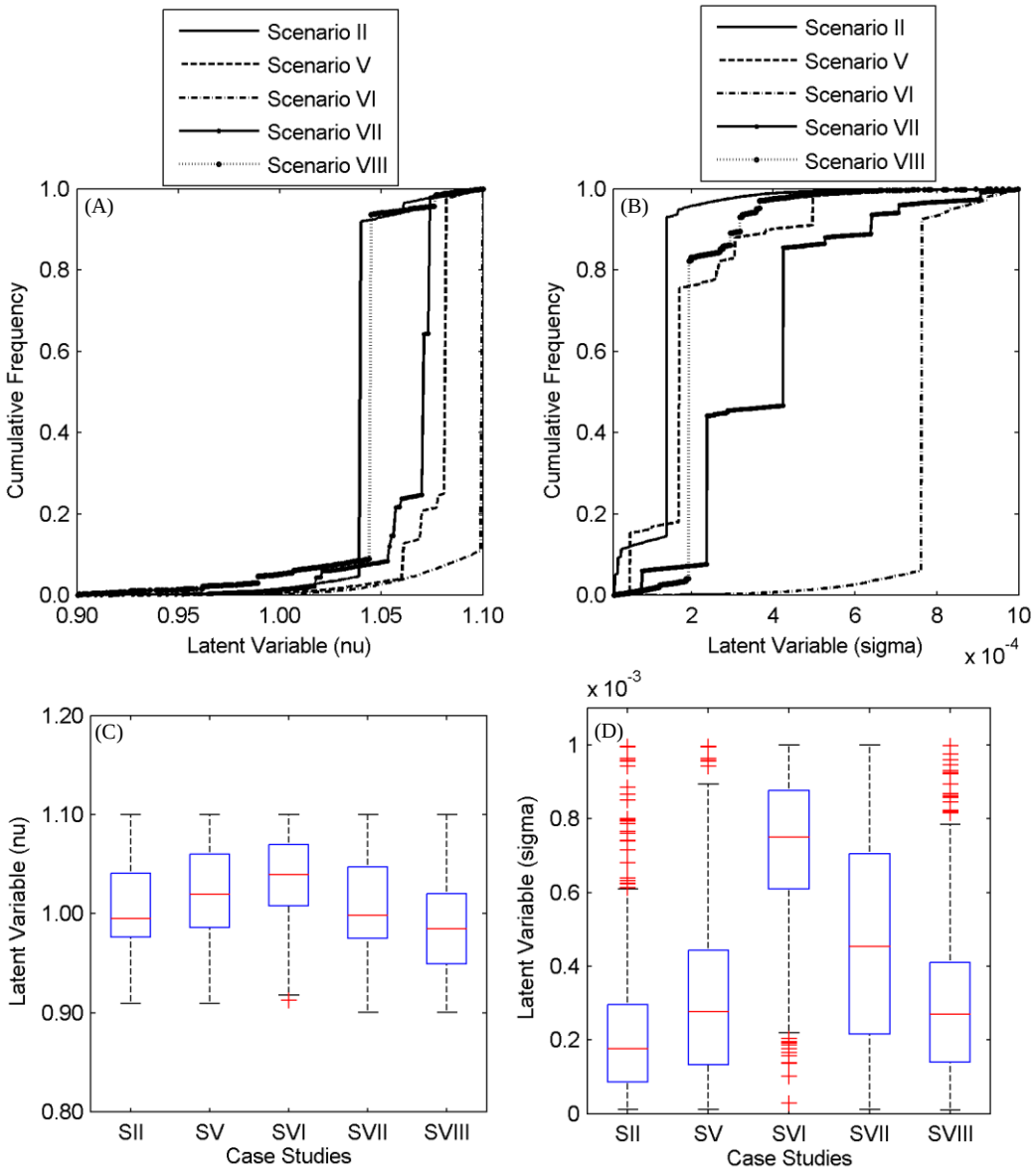

Figure 5.6 Cumulative distribution functions and marginal posterior distributions of latent variables in Scenario II, V, VI, VII and VIII: (A) CDF of latent variable nu, (B) CDF of latent variable sigma, (C) marginal posterior distributions of latent variable nu, (D) marginal posterior distributions of latent variable sigma

[In Figure 5.6(c) and 5.6(D), the central mark represents the median. The edges of each box are the 25th and 75th percentiles, and the whiskers are the most extreme data points which are not considered outliers]

5.4.3.8. Correction factors of measurement uncertainty

The average values of correction factors for streamflow and total nitrate observations are shown in Table 5.14. Correction factors ranged from 0.90 to 0.93 which means more than 90% of simulated outputs are located within the designated error range (10% for streamflow and 30% for total nitrate) [*Harmel et al.*, 2006]. The designated error adopted from literature [*Harmel et al.*, 2006] (in this study the typical scenario average is the adopted criteria) could be underestimated. Therefore, more observed data is required in the future.

Table 5.14 Correction Factors of All Gauge Stations

Scenarios	st.@35	st.@32	st.@27	st.@22	st.@20
Scenario IV	0.9489	0.9052	0.9227	0.9108	0.9342
Scenario V	0.9337	0.9192	0.9341	0.9272	0.9226
Scenario VI	0.9354	0.9513	0.9450	0.9452	0.9267
Scenario VII	0.9427	0.9328	0.9183	0.9278	0.9364
Scenario VIII	0.9090	0.9389	0.9464	0.9201	0.9421

st.: gauge station number on Figure 5.1

5.5. Discussion and Conclusion

As shown in Figure 5.3, the relative importance of each uncertainty source cannot be identified through the overall performance in finding better objective function values. A framework to incorporate input, parameter, structural and measurement uncertainty is proposed to quantify the influence from different sources towards predictive uncertainty.

The first four scenarios (Scenario I, II, III and IV) illustrate the comparative significance among each individual source of uncertainty and the input uncertainty shows largest influence on the construction of the uncertainty band and the inclusion rate of observed data points. The four scenarios V, VI, VII and VIII indicate three major findings. First, the inclusion of four uncertainty sources caused the predictive uncertainty as measured by average width of the uncertainty band to increase. Second, the application of intra-watershed responses constraints enhances the quality of solutions by yielding a higher behavior rate. Third, for this case study

intra-watershed responses were fulfilled by aggregating all uncertainty sources without implementing watershed process constraints.

Predictive uncertainty was explored by applying the BMA technique. Statistical results of the BMA applications (BMAI and BMAII) are directly influenced by the BMA weights. In other words, the statistical results from the BMA are closer to particular model (e.g. higher weights are generally assigned for SCSII at total nitrate stations) since BMA tends to favor the method that performs better in matching simulation and observation values. In the context of predictive uncertainty, BMA applications also show similar characteristics where the spread of the uncertainty band is closer to the method with higher BMA weights. However, the inclusion rate decreases considerably because the way BMA aggregates model outputs may not be able to capture observation data within the newly constructed uncertainty band.

In this study, four sources of uncertainty were investigated using the framework to simultaneously propagate input, parameter, structural and measurement uncertainty for a watershed calibration problem. Several conclusions can be made by comparing the different scenarios:

1. Predictive uncertainty is reduced with the inclusion of all four sources of uncertainty. In addition, predictive uncertainty is reduced and the inclusion rate is generally increased by the application of additional constraints which reflect intra-watershed responses.
2. With the inclusion of four uncertainty sources the best statistical solutions tend to satisfy watershed behavior constraints for this case study. The denitrification and the ratio of N-nitrate yielded from subsurface and surface flow converged to the range of constraints applied in Scenario VII and VIII automatically. On the other hand, the best statistical

solutions with only one source of uncertainty included, except for parameter uncertainty, satisfy both watershed behavior constraints.

3. Comparing the results from scenarios V, VI, VII and VIII, the incorporation of additional information from intra-watershed responses constraints improves not only the quality of the calibration results by yielding more behavioral solutions, but also reduces the predictive uncertainty yielding a narrower spread in the predictive uncertainty bands.
4. In Scenario II, V, VI, VII and VIII, two latent variables (nu and sigma) converged to similar values for all gauge stations. In addition, the converged value of θ is slightly higher than 1 which means the observed precipitation data may be overestimated.
5. The involvement of measurement uncertainty enhances the predictive uncertainty along with inclusion rate. However, the average values of the correction factors for measurement uncertainty converge to the similar ranges (0.9 ~ 0.94) with small differences for streamflow discharge and total nitrate stations.
6. The priority of selection from different available functions in the SWAT model was changed by the inclusion of fours sources of uncertainty. Applications of BMA do not show significant influence on predictive uncertainty; however, the inclusion rate declined.

For a complex large-scale watershed simulation model (in this study the SWAT model is implemented), calibration work cannot be appropriately executed without considering the element of uncertainty from many different possible sources. The proposed framework is an innovative tool to investigate and explore the significance of different uncertainty sources individually and jointly. In summary, watershed calibration problems can take advantage of the proposed framework for future studies.

References

Ahmadi, M., M. Arabi, J. C. Ascough II, D. G. Fontane, and B. A. Engelc (2012), Toward improved calibration of watershed models: Multisite many objective measures of information, *Submitted to: Advances in Water Resources*.

Ajami, N. K., Q. Duan, and S. Sorooshian (2007), An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction, *Water Resources Research*, *43*(1), 1–19, doi:10.1029/2005WR004745. [online] Available from: http://www.agu.org/pubs/crossref/2007/2005WR004745.shtml (Accessed 28 June 2011)

Arnold, J. G., P. M. Allen, and G. Bernhardt (1993), A comprehensive surface-groundwater flow model, *Journal of Hydrology*, *142*, 47–69.

Arnold, J. G., P. M. Allen, M. Volk, J. R. Williams, and D. D. Bosch (2010), Assessment of Different Representations of Spatial Variability on SWAT Model Performance, *Transactions of the ASABE*, *53*(5), 1433–1443.

Arnold, J. G., R. Srinivasan, R. S. Muttiah, and J. R. Williams (1998), Large area hydrologic modeling and assessment. I. Model development, *Journal of the American Water Resources Association*, *34*(1), 73–89.

Bai, Y., T. Wagener, and P. Reed (2009), A top-down framework for watershed model evaluation and selection under uncertainty, *Environmental Modelling & Software*, *24*(8), 901–916, doi:10.1016/j.envsoft.2008.12.012. [online] Available from: http://linkinghub.elsevier.com/retrieve/pii/S1364815208002399 (Accessed 4 May 2012)

Beck, M. B. et al. (2002), Developing a concept of adaptive community learning: Case study of a rapidly urbanizing watershed, *Integrated Assessment*, *3*(4), 299–307, doi:10.1076/iaij.3.4.299.13583. [online] Available from: http://www.tandfonline.com/doi/abs/10.1076/iaij.3.4.299.13583

Box, G. E. P., and G. C. Tiao (1992), *Bayesian inference in statistical analysis*, John Wiley & Sons.

Chiang, L., I. Chaubey, M. W. Gitau, and J. G. Arnold (2010), Differentiating impacts of land use changes from pasture management in a CEAP watershed using the SWAT model, *Transactions of the ASABE*, *53*(2), 1569–1584.

Clark, M. P., A. G. Slater, D. E. Rupp, R. a. Woods, J. a. Vrugt, H. V. Gupta, T. Wagener, and L. E. Hay (2008), Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, *Water Resources Research*, *44*, 1–14, doi:10.1029/2007WR006735. [online] Available from: http://www.agu.org/pubs/crossref/2008/2007WR006735.shtml (Accessed 22 July 2011)

Douglas-Mankin, K. R., R. Srinivasan, and J. G. Arnold (2010), Soil and Water Assessment Tool (SWAT) model: current developments and applications, *Transactions of the ASABE*, *53*(5), 1423–1431.

Du, B., J. G. Arnold, A. Saleh, and D. B. Jaynes (2005), Development and application of SWAT to landscapes with tiles and potholes, *Transactions of the ASABE*, *48*(3).

Duan, Q., N. K. Ajami, X. Gao, and S. Sorooshian (2007), Multi-model ensemble hydrologic prediction using Bayesian model averaging, *Advances in Water Resources*, *30*(5), 1371–1386, doi:10.1016/j.advwatres.2006.11.014. [online] Available from: http://linkinghub.elsevier.com/retrieve/pii/S030917080600220X (Accessed 12 March 2012)

Duan, Q., S. Sorooshian, and V. Gupta (1992), Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resources Research*, *28*(4), 1015–1031.

Gallagher, M., and J. Doherty (2007), Parameter estimation and uncertainty analysis for a watershed model, *Environmental Modelling & Software*, *22*(7), 1000–1020, doi:10.1016/j.envsoft.2006.06.007. [online] Available from: http://linkinghub.elsevier.com/retrieve/pii/S1364815206001629 (Accessed 12 July 2011)

Ghebremichael, L. T., T. L. Veith, and M. C. Watzin (2010), Determination of critical source areas for phosphorus loss: Lake Champlain Basin, Vermont, *Transactions of the ASABE*, *53*(5), 1595–1604.

Green, C. H., M. D. Tomer, M. D. Luzio, and J. G. Arnold (2006), Hydrologic Evaluation if the Soil and Water Assessment Tool for a Large Tile-Drianed Watershed in IOWA, *Transactions of the ASABE*, *49*(2), 413–422.

Harmel, R. D., R. Cooper, R. Slade, R. Haney, and J. Arnold (2006), Cumulative uncertainty in measured streamflow and water quality data for small watersheds, *Transactions of the ASABE*, *49*(3), 689. [online] Available from: http://asae.frymulti.com/abstract.asp?aid=20488&t=1 (Accessed 22 June 2011)

Harmel, R. D., and P. Smith (2007), Consideration of measurement uncertainty in the evaluation of goodness-of-fit in hydrologic and water quality modeling, *Journal of Hydrology*, *337*(3-4), 326–336, doi:10.1016/j.jhydrol.2007.01.043. [online] Available from: http://linkinghub.elsevier.com/retrieve/pii/S0022169407000741 (Accessed 10 June 2011)

Harmel, R. D., P. K. Smith, and K. W. Migliaccio (2010), Modifying goodness-of-fit in indicators to incorporate both measurement and model uncertainty in calibration and Evaluation, *Transactions of the ASABE*, *53*(1), 55–63.

Hassan, A. E., H. M. Bekhit, and J. B. Chapman (2009), Using Markov Chain Monte Carlo to quantify parameter uncertainty and its effect on predictions of a groundwater flow model, *Environmental Modelling & Software*, *24*(6), 749–763, doi:10.1016/j.envsoft.2008.11.002. [online] Available from: http://linkinghub.elsevier.com/retrieve/pii/S1364815208002016 (Accessed 12 July 2011)

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999), Bayesian Model Averaging : A Tutorial, *Statistical Science*, *14*(4), 382–417.

Jayakrishnan, R., R. Srinivasan, C. Santhi, and J. G. Arnold (2005), Advances in the application of the SWAT model for water resources management, *Hydrological Processes*, *19*(3), 749–762, doi:10.1002/hyp.5624. [online] Available from: http://doi.wiley.com/10.1002/hyp.5624 (Accessed 18 March 2012)

Kavetski, D., S. W. Franks, and G. Kuczera (2002), Confronting input uncertainty in environmental modeling, *Water Science and Application 6*, 49–68, doi:10.1016/j.envsoft.2005.04.010. [online] Available from: http://linkinghub.elsevier.com/retrieve/pii/S1364815205000873 (Accessed 20 June 2011)

Kavetski, D., G. Kuczera, and S. W. Franks (2006a), Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory, *Water Resources Research*, *42*(3), 1–9, doi:10.1029/2005WR004368. [online] Available from: http://www.agu.org/pubs/crossref/2006/2005WR004368.shtml (Accessed 26 July 2011)

Kavetski, D., G. Kuczera, and S. W. Franks (2006b), Bayesian analysis of input uncertainty in hydrological modeling: 2. Application, *Water Resources Research*, *42*(3), 1–10, doi:10.1029/2005WR004376. [online] Available from: http://www.agu.org/pubs/crossref/2006/2005WR004376.shtml (Accessed 26 July 2011)

Kim, N. W., A. H. Shin, and J. Lee (2010), Effects of streamflow routing schemes on water quality with SWAT, *Transactions of the ASABE*, *53*(5), 1457–1468.

Kuczera, G., and E. Parent (1998), Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm, *Journal of Hydrology*, *211*(1-4), 69–85, doi:10.1016/S0022-1694(98)00198-X. [online] Available from: http://linkinghub.elsevier.com/retrieve/pii/S002216949800198X (Accessed 24 June 2011)

Loosvelt, L., V. R. N. Pauwels, W. M. Cornelis, G. J. M. De Lannoy, and N. E. C. Verhoest (2011), Impact of soil hydraulic parameter uncertainty on soil moisture modeling, *Water Resources Research*, *47*(3), 1–16, doi:10.1029/2010WR009204. [online] Available from: http://www.agu.org/pubs/crossref/2011/2010WR009204.shtml (Accessed 12 July 2011)

Meng, H., A. M. Sexton, M. C. Maddox, A. Sood, C. W. Brown, R. R. Ferraro, and R. Murtugudde (2010), Modelling Rappahannock river basin using SWAT – Pilot for Chesapeake Bay watershed, *Applied Engineering in Agriculture*, *26*(5), 795–806.

Moriasi, D. N., J. G. Arnold, B. A. Engel, and C. G. Rossi (2009), Incorporation of a new shallow water table depth algorithm into SWAT2005, *Transactions of the ASABE*, *52*(3), 771–784.

Moriasi, D. N., J. G. Arnold, M. W. V. Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith (2007), Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, *Transactions of the ASABE*, *50*(3), 885–900.

Neitsch, S. L., J. G. Arnold, J. R. Kiniry, and W. J. R (2011), Soil & Water Assessment Tool Theoretical Documentation Version 2009, *Texas Water Resources Institute Technical Report No. 406, Texas A&M University System*.

Newman, J. E. (1997), *Our Changing Climate, In: The Natural Heritage of Indiana, Edited M. T. Jackson*.

Osidele, O. O., W. Zeng, and M. B. Beck (2006), A random search methodology for examining parametric uncertainty in water quality models, *Water Science & Technology*, *53*(1), 33, doi:10.2166/wst.2006.005. [online] Available from: http://www.iwaponline.com/wst/05301/wst053010033.htm (Accessed 28 June 2011)

Raftery, A. E., F. Balabdaoui, T. Gneiting, and M. Polakowski (2003), Using Bayesian model averaging to calibrate forecast ensembles, *Tech. Rep. 440, Dep. of Stat., Univ. of Wash., Seattle*.

Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski (2005), Using Bayesian Model Averaging to Calibrate Forecast Ensembles, *American Meteorological Society*, 1155–1174.

Refsgaard, J. C., J. P. van der Sluijs, J. Brown, and P. van der Keur (2006), A framework for dealing with uncertainty due to model structure error, *Advances in Water Resources*, *29*(11), 1586–1597, doi:10.1016/j.advwatres.2005.11.013. [online] Available from: http://linkinghub.elsevier.com/retrieve/pii/S0309170805002903 (Accessed 29 June 2011)

Sorooshian, S., and J. A. Dracup (1980), Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: Correlated and heteroscedastic error cases, *Water Resources Research*, *16*(2), 430, doi:10.1029/WR016i002p00430.

Srinivasan, R. X. Zhang, J. A. (2010), Swat Ungaged: Hydrological Budget and Crop Yield Predictions in the Upper Mississippi River Basin, *Transactions of the ASABE*, *53*(5), 1533–1546.

Stedinger, J. R., R. M. Vogel, S. U. Lee, and R. Batchelder (2008), Appraisal of the generalized likelihood uncertainty estimation (GLUE) method, *Water Resources Research*, *44*, 1–17,

doi:10.1029/2008WR006822. [online] Available from: http://www.agu.org/pubs/crossref/2008/2008WR006822.shtml (Accessed 22 June 2011)

Tolson, B. A., and C. A. Shoemaker (2007), Dynamically dimensioned search algorithm for computationally efficient watershed model calibration, *Water Resources Research*, *43*(1), 1–16, doi:10.1029/2005WR004723. [online] Available from: http://www.agu.org/pubs/crossref/2007/2005WR004723.shtml (Accessed 28 June 2011)

Vrugt, J. A., C. J. F. Braak, C. G. H. Diks, B. A. Robinson, J. M. Hyman, and D. Higdon (2009a), Accelerating Markov Chain Monte Carlo Simulation by Differential Evolution with Self-Adaptive Randomized Subspace Sampling, *International Journal of Nonlinear Sciences & Numerical Simulation*, *10*(March), 271–288.

Vrugt, J. A., C. J. F. Braak, V. K. Gupta, and B. A. Robinson (2009b), Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?, *Stochastic Environmental Research and Risk Assessment*, *23*(7), 1011–1026, doi:10.1007/s00477-008-0274-y.

Wöhling, T., and J. a. Vrugt (2008), Combining multiobjective optimization and Bayesian model averaging to calibrate forecast ensembles of soil hydraulic models, *Water Resources Research*, *44*(12), 1–18, doi:10.1029/2008WR007154. [online] Available from: http://www.agu.org/pubs/crossref/2008/2008WR007154.shtml (Accessed 5 March 2012)

Yang, J., P. Reichert, K. Abbaspour, J. Xia, and H. Yang (2008), Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China, *Journal of Hydrology*, *358*(1-2), 1–23, doi:10.1016/j.jhydrol.2008.05.012. [online] Available from: http://linkinghub.elsevier.com/retrieve/pii/S0022169408002370 (Accessed 10 June 2011)

Yen, H., R. Bailey, and M. Arabi (2012a), Evaluation of improved parameter estimation and uncertainty analysis of watershed models: The role of general information of watershed processes, *Under internal review*.

Yen, H., R. M. Records, M. Arabi, and B. A. Tolson (2012b), Appraisal of sampling techniques for parameter estimation and uncertainty analysis in watershed modeling, *Under internal review*.

Chapter 6

Conclusion and Discussion

6.1. Major findings

The proposed framework provides a great tool in implementing various topics in the field of watershed calibration and watershed modeling. Results of various applications on sampling strategies and predictive uncertainty analysis in this study can be summarized into major findings as follows:

(1) The proposed framework provides an innovation structure to evaluate sampling strategies by the achievement of optimality for objective function, behavior rate based on additional behavior definition and also by exploring identifiability on the overall parameter space. Users are able to execute a meticulous assessment on not only the performance of parameter estimation techniques but also to have a better understanding of the fashion in parameter sampling on watershed calibration problems.

(2) The application of actual watershed behavior measures by additional constraints enhances the quality of watershed calibration work where well calibrated results are convincingly more realistic corresponding to real world hydrologic/nutrient processes in field.

(3) The optional function in calculating surface runoff in the SWAT model by implementing information from plant evapotranspiration has been proven to have outstanding capability

in performing total nitrate simulation over the default method (applies antecedent soil moisture condition) regardless the precipitation condition over dry/wet seasons.

(4) The BMA technique does not show impressive performance on watershed calibration by aggregating results from streamflow discharge and total nitrate processes in the SWAT model. In addition, predictive uncertainty is generally increased with incorporation of the BMA in both discharge and total nitrate.

(5) The proposed framework in evaluating uncertainty from different sources individually/jointly provides logical and informative guidance towards predictive uncertainty for watershed simulation models.

(6) Watershed calibration with the inclusion of input, parameter, structural and measurement uncertainty shows significant impact on predictive uncertainty (predictive uncertainty is considerably increased). However, it tends to converge to solutions which satisfy actual watershed behavior without being constrained by additional penalty functions. This notable finding may prompt watershed calibration work into another stage where hydrologic/nutrient processes and actual watershed behavior can both be fulfilled with uncertainty sources considered concurrently.

6.2. Limitations and shortcomings

To apply the proposed framework in evaluating sampling strategies and predictive uncertainty, a number of assumptions have to be made which could also lead to potential limitations and shortcomings as follows:

(1) The current framework is incorporated with the SWAT model where 119 model parameters (2 additional latent variables are not included) can be adjusted during

calibration process. The current adjustable parameters are sufficient enough to calibrate a well setup watershed such as the ECW with fine solutions. However, all system parameters should be included to sustain the integrity to perform of a rigorous calibration work.

(2) The default termination criteria vary for different parameter estimation techniques. Therefore, a unified termination criterion should be developed for the evaluation process of sampling strategies.

(3) For the input uncertainty, the only forcing input which is incorporated with the current framework is precipitation. However, other forcing inputs could also lead to considerable impact toward predictive uncertainty (e.g. temperature, humidity). More alternative functions should be expanded to serve more topics for the following studies.

(4) The same set of default ranges for latent variables in literature [*Ajami et al.*, 2007] is adopted for the input uncertainty explored in this study. Therefore, the influence caused by forcing input (precipitation in this study) is conditional to the default settings. The results show that input uncertainty causes the greatest impact toward predictive uncertainty (for scenarios with only one source of uncertainty included besides parameter uncertainty) may be different with the settings of latent variables being altered.

(5) The application of the input uncertainty requires two additional latent variables (nu and sigma) to be validated during calibration process. The following predictive uncertainty analysis is not investigated for the same dimension (e.g. case scenarios with input uncertainty involved are 2 dimensional more than other scenarios).

(6) Similar to the dimensional issue for the input uncertainty, the application of SCSII requires one additional model parameter (CNCOEF) to be validated during calibration

process. In this case, the following predictive uncertainty analysis is one dimensional more than other scenarios (e.g. if the SCSI application is a 28-D problem, SCSII entails 29-D).

(7) For measurement uncertainty, the recommended error estimation from literature [*Harmel et al.*, 2006] is adopted in the proposed framework. However, the source of measurement uncertainty can also be classified into various categories (e.g. sample collection, laboratory analysis [*Harmel et al.*, 2006]). Measurement uncertainty should be further examined and formulated in multiple detail features.

(8) During the application of the statistical valid likelihood function [*Ahmadi et al.*, 2012], the transformed data (residuals) are still not entirely normally distributed, independent, and homoscedastic.

(9) Real-time forecast/operation is still computational expensive for the proposed framework because each model evaluation of the SWAT model is at least one minute or more (for the large number of watershed input files) in the current computer technology and it may take weeks to reach the termination criteria during calibration process.

6.3. Broader impact of the proposed framework

The proposed framework is flexible to be expanded based on the current capacity in managing various purposes. For example, users can apply additional constraints for actual watershed behavior other than denitrification and ratio of N-nitrate loss from surface and subsurface flow (e.g. instead of constraining the ratio of N-nitrate loss, users can also assign constraints for actual values of N-nitrate loss). Potential broader impact of the proposed framework is described as follows:

(1) The proposed framework provides a general platform for future studies in evaluating sampling strategies in a rigorous and systematic fashion. Evaluations based on the proposed framework are having more objective comparisons and also reduce the concern of various evaluation standards.

(2) The application of behavior definition enhances the quality (solutions are further refined by implementing behavior definition) of watershed calibration. In addition, behavior definition can also be served as a supplementary tool in examining the effectiveness of different parameter estimation techniques (e.g. two methods are showing compatible results in achieving objective function values but one is having higher behavior rate over the other). Well-known sampling strategies proposed previously are not showing competitive performance after the behavior definition is applied in this study (e.g. SCE-UA and DREAM are not performing as well as DDS).

(3) The application of TSDE (to explore identifiability on the overall parameter space) allows researchers to have a better understanding on not only "how" a sampling strategy can perform (ability in achieving global optimality) but also provides the reason "why" (e.g. SCE-UA explores a much large area in the feasible domain space compares to DDS) it can (or cannot) perform well. Future applications of watershed calibration can take advantage of this before making decision in selecting the most appropriate sampling strategy.

(4) The idea of implementing the actual watershed behavior measures explains that the watershed calibration process without considering additional information from the field may generate physically meaningless final results. The proposed framework provides an

innovative approach by including additional constraints can effectively avoid the possible mistaken situation for future calibration work.

(5) By the applications of the two functions in calculating surface runoff in the SWAT model, SCSII shows superior performance in total nitrate simulation compares to SCSI. However, SCSII is still not extensively applied in relevant studies. Illustrations and comparisons of SCSI and SCSII in this study provide a scientific verification that future studies should implement SCSII as the default method in surface runoff calculation.

(6) The second part of the proposed framework presents valuable information that the input uncertainty reveals the greatest impact towards predictive uncertainty compares to other sources (based on scenarios only include only one source of uncertainty besides parameter uncertainty). Results of predictive uncertainty could be different if some of the major settings are altered (e.g. the default range of latent variables), however, future studies should have this finding considered as an imperative consideration before executing associated research work.

(7) The convergence patterns of denitrification and N-nitrate losses ratio (actual watershed behavior constraints are automatically satisfied without being constrained) in Chapter V indicate the importance to have all sources of uncertainty involved during calibration process. Future studies should not consider sources of uncertainty as a whole (e.g. consider all sources of uncertainty can be contributed to parameter uncertainty) but have to identify each specifically.

6.4. Future directions

The proposed framework provides a generalized platform for researchers to have a broad view towards watershed modeling problems in full dimension instead of only focusing on specific

divisions. Users can take advantage of the whole (or implement first and second part of the framework respectively) framework for relevant studies in the oncoming future. Some of the potential future directions are illustrated as follows:

(1) The proposed framework evaluates the effectiveness of sampling strategies in several different perspectives but the definition whether a parameter estimation technique is a global or a greedy searching based method is not considered. A scientifically rigorous standard in defining the nature of different sampling strategies (e.g. debate between the performance of SCE-UA and DDS indicates that SCE-UA is a global optimization algorithm but DDS is not [*Behrangi et al.*, 2008; *Tolson and Shoemaker*, 2008]) should be regulated in the future.

(2) In this study, the behavior definition adopted from literature does not represent the standard suitable for all other area. Further development and improvement for different site locations of a well-accepted standard for behavior definition is required.

(3) The importance of actual watershed behavior measures should be considered during the watershed calibration process in the future. In addition, more field survey is required since the observation data for most actual watershed behavior is still scarce in literature.

(4) Results from the proposed framework can be used to evaluate the comparative importance of individual source of uncertainty. Therefore, more resources can be invested in relevant topics (e.g. more investment in observation data survey is required if the forcing inputs present greatest impact toward predictive uncertainty).

(5) Evaluation between the two methods (SCSI and SCSII) in calculating surface runoff in the SWAT model substantiates the potential influence by the choice of optional functions within a complex watershed simulation model. Beyond the implemented

methods in this study, more optional functions (e.g. Green-Ampt infiltration method) should be compared and validated for further investigation.

(6) From the interactions (watershed behavior constraints are satisfied automatically during calibration) between the actual watershed behavior and the inclusion of input, parameter, structural and measurement uncertainty, the inclusion of fours sources of uncertainty simultaneously shows great impact toward not only predictive uncertainty but it can also be extended to enhance the quality of calibration without involving additional rules. Various new topics can be explored from this innovative finding.

(7) The results derived from this study may only be valid in the case of the ECW. More studies (e.g. applications on watersheds other than the Mid-west region) are required to understand the relationships among different sources of uncertainty are still consistent to the present settings.

References

Ahmadi, M., M. Arabi, J. C. Ascough II, D. G. Fontane, and B. A. Engelc (2012), Toward improved calibration of watershed models: Multisite many objective measures of information, *Submitted to: Advances in Water Resources*.

Ajami, N. K., Q. Duan, and S. Sorooshian (2007), An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction, *Water Resources Research*, *43*(1), 1-19, doi:10.1029/2005WR004745. [online] Available from: http://www.agu.org/pubs/crossref/2007/2005WR004745.shtml (Accessed 28 June 2011)

Behrangi, A., B. Khakbaz, J. a. Vrugt, Q. Duan, and S. Sorooshian (2008), Comment on "Dynamically dimensioned search algorithm for computationally efficient watershed model calibration" by Bryan A. Tolson and Christine A. Shoemaker, *Water Resources Research*, *44*(12), 1-3, doi:10.1029/2007WR006429. [online] Available from: http://www.agu.org/pubs/crossref/2008/2007WR006429.shtml (Accessed 6 August 2012)

Harmel, R. D., R. Cooper, R. Slade, R. Haney, and J. Arnold (2006), Cumulative uncertainty in measured streamflow and water quality data for small watersheds, *Transactions of the ASABE*, *49*(3), 689. [online] Available from: http://asae.frymulti.com/abstract.asp?aid=20488&t=1 (Accessed 22 June 2011)

Tolson, B. a., and C. a. Shoemaker (2008), Reply to comment on "Dynamically dimensioned search algorithm for computationally efficient watershed model calibration" by Ali Behrangi et al., *Water Resources Research*, *44*(12), 2-5, doi:10.1029/2008WR006862. [online] Available from: http://www.agu.org/pubs/crossref/2008/2008WR006862.shtml (Accessed 6 August 2012)