A CONCEPTUAL PRECIPITATION-RUNOFF MODELING SUITE: MODEL SELECTION, CALIBRATION AND PREDICTIVE UNCERTAINTY ASSESSMENT

by

Tyler Jon Smith

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Civil Engineering

MONTANA STATE UNIVERSITY
Bozeman, Montana

November 2008

APPROVAL

of a thesis submitted by

Tyler Jon Smith

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citation, bibliographic style, and consistency, and is ready for submission to the Division of Graduate Education.

Dr. Joel E. Cahoon

Dr. Lucy A Marshall

Approved for the Department of Civil Engineering

Dr. Brett W. Gunnink

Approved for the Division of Graduate Education

Dr. Carl A. Fox

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library.

If I have indicated my intention to copyright this thesis by including a copyright notice page, copying is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for permission for extended quotation from or reproduction of this thesis in whole or in parts may be granted only by the copyright holder.

Tyler Jon Smith

November 2008

ACKNOWLEDGEMENTS

I would like to offer my appreciation to all of the members of my graduate committee, Dr. Lucy Marshall, Dr. Joel Cahoon, Dr. Brian McGlynn and Dr. Otto Stein for their endless assistance to me in this endeavor. I wish to especially thank Dr. Lucy Marshall for her continuous support and encouragement toward the completion of this research project. Additionally, I would like to thank Dr. Joel Cahoon for making the process of performing research in the Department of Land Resources and Environmental Sciences, while pursuing a degree from the Department of Civil Engineering so uncomplicated. This research was made possible thanks to partial funding by a USGS 104b grant administered by the Montana Water Center. Logistical support was provided by Dr. Ward McCaughey and the US Forest Service Rocky Mountain Research Station. Further, I would like to express my appreciation to the members of the Watershed Hydrology Lab at MSU, and especially Kelsey Jencso, for providing much of the data used in this study. Finally, I would like to thank my adoring wife, Kelly, who happily suffered through countless conversations about the details of conceptual hydrologic modeling, model calibration algorithms, MATLAB® programming, etc.

TABLE OF CONTENTS

1. INTRODUCTION 1

Scope and Context 1
Research Motivation 3
Outline of Thesis 4

2. LITERATURE REVIEW 6

Top-Down Model Development 6
Model Calibration Methods 9
Uncertainty Methods 12
Model Assessment 15

3. BAYESIAN METHODS IN HYDROLOGIC MODELING: A STUDY OF RECENT ADVANCEMENTS IN MARKOV CHAIN MONTE CARLO TECHNIQUES 20

Introduction 21
Bayesian Methods in Hydrology 23
Advancements in Markov Chain Monte Carlo Techniques 25
Adaptive Metropolis Algorithm 26
Delayed Rejection Adaptive Metropolis Algorithm 27
Differential Evolution Markov Chain Algorithm 30
Case Studies 32
Case Study 1: Synthetic Bimodal Gaussian Mixture 33
Case Study 2: Tenderfoot Creek Experimental Forest, MT 36
Site Description 36
Precipitation-Runoff Model Description 37
Input Data 38
Algorithm Evaluations 39
Conclusions 43
Acknowledgements 45

4. DEVELOPMENT OF A CONCEPTUAL PRECIPITATION-RUNOFF MODELING SUITE 46

Introduction 46
Site Description 47
Tenderfoot Creek Experimental Forest 47
Model Data 49
Conceptual Modeling Suite Development 52

TABLE OF CONTENTS – CONTINUED

Soil Moisture Accounting and Runoff Routing Base Structures 53
Bucket Model 54
Probability Distributed Model 56
TOPMODEL 59
Soil Moisture Accounting Data Requirements 62
Snowmelt Accounting 63
Temperature Index Approach 64
Radiation Index Approach 64
Temperature and Radiation Index Approach 65
Precipitation Semi-Distribution 66
Semi-Distribution by Aspect 67
Semi-Distribution by Elevation 69
Semi-Distribution by Aspect and Elevation 71
Conclusion 72

5. PREDICTIVE PERFORMANCE AND UNCERTAINTY ASSESSMENT OF A CONCEPTUAL MODELING FRAMEWORK APPLIED TO STRINGER CREEK WATERSHED 74

Introduction 74
A Review of the Bayesian Methodology Applied in this Study 76
Assessment Protocol for Modeling Suite Applied to the Stringer Creek Watershed 77
Performance-Based Assessment 77
Model Calibration 78
Results and Discussion of Model Performance Under Calibration 79
Optimum Model Structure Under Calibration 86
Assessment of Model Performance Using Independent Data Source 89
Comparison of Likelihood Function Assumptions 96
Uncertainty-Based Assessment 98
Uncertainty Assessment of Individual Model Structures 100
Realism-Based Assessment 106
Conclusion 113

6. SIMULATION AND PREDICTION LAB FOR ANALYSIS OF SNOWMELT HYDROLOGY 117

Introduction 117
Model Selection 119
Model Calibration 121
Calibration Data 123

TABLE OF CONTENTS – CONTINUED

Watershed Data 124
Model Parameter Values 124
Algorithm Parameter Values 125
Available Plots 125
Available Data Outputs 126
Parameter Set Evaluation 127
Uncertainty Analysis 128
Conclusion 131

7. CONCLUSIONS 133

Summary of Research 133
Comparison of Recent Advancements in Markov Chain Monte Carlo Sampling 134
Formation of a Hydrologic Modeling Suite 136
Evaluation of Alternate Hydrologic Model Structures 136
Development of a Graphical User Interface 138

REFERENCES CITED 140

APPENDICES 155

APPENDIX A: 95% Prediction Intervals 156
APPENDIX B: SPLASH – Sample Plots 172

LIST OF TABLES

Table Page

3.1 The 2.5 and 97.5 Percentile Sampled Parameter Values for the AM, DRAM, and DE-MC Algorithms and Values Corresponding to the Maximum log-Likelihood 42

4.1 Onion Park SNOTEL Meteorological Sensors of Interest 52

4.2 Effective Parameters of the Bucket Model Soil Moisture Accounting and Runoff Routing Component 55

4.3 Effective Parameters of the Probability Distributed Model Soil Moisture Accounting and Runoff Routing Component 57

4.4 Effective Parameters of the TOPMODEL Soil Moisture Accounting and Runoff Routing Component 60

5.1 Model Performance-Based Assessment Criteria for the Calibration Period ... 80

5.2 Model Performance-Based Assessment Criteria for the Validation Period 91

5.3 Comparison Likelihood Functions for the Calibration of the Probability Distributed Model Component 97

5.4 Model Uncertainty-Based Assessment Criteria 104

5.5 Nash-Sutcliffe Efficiencies for the Simulation of Snow Water Equivalents on the Calibration and Validation Periods 107

6.1 Sample Calibration Data Format 123

6.2 Description of Algorithm Parameters 125

6.3 Description of Plots Available Within the Model Calibration Tab of SPLASH 126

6.4 Description of Data Outputs Available Within the Model Calibration Tab of SPLASH 126

6.5 Description of Plots and Data Outputs Available Within the Parameter Set Evaluation Tab of SPLASH 128

LIST OF TABLES – CONTINUED

Table Page

6.6 Description of Plots and Data Outputs Available Within the Uncertainty Analysis Tab of SPLASH 131

LIST OF FIGURES

Figure Page

3.1 Hypothetical parameter surface illustrating the ability of the AM and DRAM algorithms to explore the parameter surface 29

3.2 Comparison of parameter estimation between the AM, DRAM and DE-MC algorithms for a synthetic case study 35

3.3 The Probability Distributed Model with snowmelt as an additional precipitation input 37

3.4 Comparison of AM, DRAM and DE-MC parameter searches for a hydrologic case study 41

4.1 Tenderfoot Creek Experimental Forest; (a) location and (b) instrumentation 48

4.2 (a) Onion Park SNOTEL site and (b) Stringer Creek flume 50

4.3 A schematic representation of the bucket model soil moisture accounting and runoff routing component 54

4.4 A schematic representation of the Probability Distributed Model soil moisture accounting and runoff routing component 57

4.5 A schematic representation of the TOPMODEL soil moisture accounting and runoff routing component 59

4.6 Semi-distribution of precipitation by aspect 68

4.7 Semi-distribution of precipitation by elevation 70

4.8 Semi-distribution of precipitation by aspect and elevation 72

5.1 Calibration results showing maximum, mean and minimum log-likelihood value for each model dimension 83

5.2 Calibration results grouped by the soil moisture accounting and runoff routing component 84

5.3 Calibration results grouped by the snowmelt accounting component 85

LIST OF FIGURES – CONTINUED

Figure Page

5.4 Calibration results grouped by the precipitation semi-distribution component....86

5.5 Parameter traces from the calibration of the PDMsdE_TRn model structure....87

5.6 Hydrograph simulation for the PDMsdE_TRn model structure and the calibration data set....88

5.7 Hydrograph simulation for the PDMsdE_TRn model structure and the validation data set....90

5.8 Validation results grouped by maximum, mean and minimum log-likelihood value for each model dimension....92

5.9 Validation results grouped by the soil moisture accounting and runoff routing component....93

5.10 Validation results grouped by the snowmelt accounting component....94

5.11 Validation results grouped by the precipitation semi-distribution component....95

5.12 Likelihood function comparison for the best Probability Distributed Model based structure at each dimension....97

5.13 Range of predictions on the calibration data set from the collection of model structures investigated....99

5.14 Range of predictions on the validation data set from the collection of model structures investigated....100

5.15 Parameter posterior histograms from the calibration of the PDMsdE_TRn model structure....101

5.16 Comparison of sharpness and likelihood versus model dimension....103

5.17 Range of predictions on the 95% prediction intervals for the collection of model structures investigated....105

LIST OF FIGURES – CONTINUED

Figure Page

5.18 Snow water equivalents simulation results grouped by the soil moisture accounting and runoff routing component for each model dimension over the calibration period 108

5.19 Snow water equivalents simulation for the PDMsdE_TRn model structure and the calibration data set....109

5.20 Snow water equivalents simulation for the PDMsdE_TRn model structure and the validation data set....110

5.21 Snow water equivalents ensemble simulations for each model structure and the calibration data set....111

5.22 Snow water equivalents ensemble simulations for each mode structure and the validation data set....112

6.1 SPLASH modeling software....118

6.2 Suite of models available in SPLASH....120

6.3 Model calibration tab within SPLASH....122

6.4 Evaluate parameter set tab within SPLASH....127

6.5 Uncertainty analysis tab within SPLASH....130

ABSTRACT

In Montana and much of the Rocky Mountain West, the single most important parameter in forecasting the controls on regional water resources is snowpack. Despite the heightened importance of snowpack, few studies have considered the representation of uncertainty in coupled snowmelt/hydrologic conceptual models. Uncertainty estimation provides a direct interpretation of the risk associated with predictive modeling results. Bayesian inference, through the application of Markov chain Monte Carlo methods, provides a statistical means of approximating uncertainty associated with both the parameters and the model structure. This thesis addresses the complexities of predictive modeling in hydrology through the development, implementation and analysis of a suite of conceptual hydrologic models under a Bayesian inference framework. The research is presented in four main sections. First, a comparative assessment of three recently developed Markov chain Monte Carlo algorithms, based on their performance across two case studies, is performed. This study has revealed that the extreme complexity of the parameter space associated with simple, conceptual models is best explored by the Delayed Rejection Adaptive Metropolis algorithm. Second, a complete description of the models and study site incorporated in the study are presented, building on established theories of model development. Third, an investigation of the value of each model structure, considering predictive performance, uncertainty and physical realism is presented. This section builds on results of the first section, through the application of the Delayed Rejection Adaptive Metropolis algorithm for model calibration and uncertainty quantification under Bayesian principles. Finally, a discussion of the Simulation and Prediction Lab for Analysis of Snowmelt Hydrology, developed to incorporate the tasks of model selection, calibration and uncertainty analysis into a simple graphical user interface is explained. The application of a complete modeling framework from model selection to calibration and assessment presented in this thesis represents a holistic approach to the development of improved understanding of snow-dominated watersheds through prediction by coupled snowmelt/hydrologic modeling strategies.

INTRODUCTION

Scope and Context

In Montana and much of the Rocky Mountain West, the single most important parameter in forecasting the controls on regional water resources is snowpack [*Williams et al.*, 1999]. In the mountainous areas of Montana, in particular, the runoff of snowmelt is the key driving force in downstream supplies of water. Water is deposited at much higher volumes in the form of snow in these mountainous areas, which results in greater accumulation and storage and diminished evaporation [*Hauer et al.*, 2007].

Given the importance of upstream precipitation and streamflow generation on water availability for downstream communities, predictive hydrologic models are an important tool for water supply forecasting. Much work has been done in the past to help characterize and identify the controls on snowpack dynamics in hydrologic models, from how a snowpack undergoes ripening to the eventual release and hydrologic pathways of the water associated with the snowpack itself. However, the difficulty associated with relating first-order controls to watershed characteristics and in understanding the balance between model complexity, parameterization and uncertainty is significant.

The main problem faced in hydrologic modeling is in finding an appropriate mathematical representation of the natural system at a level of complexity suitable for the modeling exercise. While it is a widely accepted belief that no single model is suitable for all potential hydrologic conditions, it must also be acknowledged that in many cases the criteria for determining a suitable model given a set of watershed characteristics are still

largely undefined. This type of ambiguity is directly attributable to difficulties in quantifying the primary watershed drivers; those processes that can be singled out as most significant to successful prediction across a range of possible conditions. Linked to this is a fundamental approach common in hydrology of developing a single model and contending its appropriateness, rather than approaching the problem from multiple avenues and identifying underlying patterns associated with the efficiencies and deficiencies of multiple model structures [*Smith et al.*, 2004].

Model development approaches cover a broad spectrum from physically-based to empirical models. As calibration techniques have become more sophisticated (and trusted) there has been a shift from a physically-based approach to a conceptual approach to model development. Conceptual models are built in a downward fashion, attempting to include as many of the known processes of the system as possible, while lumping all others into model parameters. Such an approach has been strongly advocated based on the uncertainty associated with appropriate process representation in physically-based models and the reliability of model calibration techniques [*Beven*, 1989; *Klemes*, 1983].

All conceptual models are confronted with the necessity of calibrating immeasurable, effective parameters. There has been an evolution in hydrology from manual calibration methods to fully computerized approaches. Important to automatic calibration is some form of objective mathematical function for the evaluation of goodness of fit. While all calibration methods seek to find the best set of parameters (given the model and data), there has been a recent emergence of Bayesian inference in an effort to quantify the uncertainty associated with the optimized parameter set.

Bayesian inference is strong in its probabilistic characterization of parameter posterior distributions and in its ability to accommodate multiple model comparisons.

Research Motivation

Development and implementation of a complete hydrologic modeling framework is necessary in the face of uncertain data and uncertain representation of the physical system. However, few studies in the past have addressed the complexities of hydrologic modeling through a holistic approach, in the sense of considering modeling calibration and uncertainty analysis across a suite of conceptual modular structures developed under a downward approach.

In the context of setting the foundation for a holistic strategy to predictive hydrologic modeling, the research presented in this thesis focuses on the application of a Bayesian approach to the problem of model selection and assessment particular to a coupled snowmelt/hydrologic system. The specific objectives of the study are:

- Comparison of three recently developed Markov chain Monte Carlo (MCMC) algorithms for specifying hydrologic models using Bayesian inference. Evaluation of each algorithm will be focused on their effectiveness and efficiency when applied to both a synthetic case study and a hydrologic case study.
- Specification of a suite of alternative precipitation-runoff models. The collection of models will be conceptual in their representation of the physical system, targeting description of the primary influences on overall watershed response (streamflow).

- Assessment of the suite of alternate model structures under a Bayesian inferential framework. Model calibration and uncertainty assessment will be carried out via MCMC simulation and model evaluation will be based on the criteria of performance, uncertainty and realism; with effort targeted to assessing different quantifications of primary watershed drivers.
- Development of a graphical user interface that can serve as an amalgamation of the broader research objectives; that is, creation of a framework encompassing model selection, model calibration and uncertainty analysis under a Bayesian context.

Outline of Thesis

Chapter 2 provides the necessary background on the theory serving as the basis for the research offered in this thesis. The key principles relating to model development strategies are discussed, including a review of the uses of multiple types of data in model specification. The evolution of model calibration and its relationship with uncertainty analysis methodologies are summarized. Finally, an appraisal of current model assessment approaches and typical implementations are offered.

Chapter 3 offers a comparative study of three newly designed MCMC algorithms implemented under a Bayesian framework to parameter estimation and uncertainty analysis. The efficacy of each is considered through their application to both a synthetic case study and a hydrologic case study. The results of this study then serving as the

motivating factor in the selection of the algorithm to be used for calibration of the models discussed in Chapter 4.

In Chapter 4 the suite of conceptual hydrologic models considered in this research is introduced, providing specific details on each model component and its associated parameters. The motivation for the modeling suite's construction is considered in further detail. Additionally, the study location and its relevant characteristics are presented.

Chapter 5 presents an evaluation of the modeling suite developed in Chapter 4, from model calibration to model assessment. The Delayed Rejection Adaptive Metropolis (DRAM) algorithm is selected for parameter estimation and uncertainty analysis under the Bayesian framework. Model assessment is considered on three fronts: predictive performance, uncertainty analysis and realism.

Chapter 6 introduces the graphical user interface designed to incorporate each of the salient tasks associated with predictive hydrologic modeling. The simplification associated with the implementation of a complete modeling framework is achieved through an easy to use software package.

Chapter 7 summarizes the conclusions related to the research presented in each of the chapters. The references cited in the body of the thesis follows, along with the appendices containing auxiliary findings relevant to the research.

LITERATURE REVIEW

This chapter introduces and reviews the progression of ideas in the domain of predictive modeling in hydrology. In line with the research focus of this thesis, the topics reviewed will cover model development techniques, especially focusing on conceptual approaches that blend black-box and physically-based formulations; the treatment of calibration and uncertainty analysis for effective model parameters, including a discussion of the evolution of approaches from past to present; and the assessment of model suitability, considering performance, uncertainty and realism.

Top-Down Model Development

Model development can be classified as following one of two routes in hydrology, upward or downward [*Sivapalan et al.*, 2003]. Where the upward approach attempts to scale theoretical knowledge of a small-scale process up to a higher scale (e.g., the hydrological scale), the downward approach attempts to identify a large-scale process and then identify the possible steps that could have been taken to reach the result of the large-scale process [*Klemes*, 1983].

Extensive effort has been put forth to develop physically-based models, using upward model development, resulting in a variety of hydrologic models [see *Beven*, 1989, for example]. The goal of these models was to work from the finest details of the system upward, incorporating all the processes that are involved in the resulting watershed dynamics (i.e., streamflow). Physically-based models were seen as an improvement upon the previous model structures of the time, in that the parameters were

thought to be of a physical interpretation that would reduce the burdens of data required for parameter calibration/validation [e.g., *Abbott et al.*, 1986; *Bathurst*, 1986].

While, in concept, this is a very appealing type of model, it has been found to be burdened with complications, noting that “physical” models may still be a very simplified representation of reality [*Beven*, 1989] and are vulnerable to the issue of scale transferability [*Vache and McDonnell*, 2006]. Only the scientifically understood and accepted hydrologic processes are generally included in physically-based models and the understanding of these dominant processes is rooted in small-scale studies which have been shown to be unreliable when scaled up to watershed implementations [*Beven*, 2000]. It may be argued that confidence in the physical nature of the model is overwhelmingly greater than the confidence in a parameter calibration scheme required for conceptual, downward developed models [*Bathurst*, 1986]. *Beven* [1989, p. 158], however, has argued that the computing power required to implement trustworthy calibration techniques is sufficient, while physically-based models are still never proven to be of particular utility in “real-world” applications and believed that they are "considerably in error, despite or even because of the so-called 'physical basis' of the model".

Recently, a hybrid of the top-down and bottom-up approaches has gained more popularity, incorporating well-understood physical processes, while lumping all other processes into “tuning” parameters. This is a concept that was argued for 25 years ago by *Klemes* [1983, p. 8] who stated that

> … the most promising route to significant new discoveries in hydrology is to combine the upward and downward search based on the existing facts and knowledge as well as on imagination and intuition, to form testable hypotheses—i.e. to apply the time-honored scientific method.

Littlewood et al. [2003] renewed the call for such an approach that would allow scientific knowledge to be combined with the information provided by the data to develop the most parsimonious and realistic model structure possible.

Since then, many studies have shown that comparatively very simple models were able to equal or outperform their complex counterparts in representing the observed output of a watershed [e.g., *Beven*, 1989; *Hooper et al.*, 1988; *Hornberger et al.*, 1985; *Jakeman and Hornberger*, 1993; *Loague and Freeze*, 1985]. Indeed, as computing ability has increased, the implementation of more advanced calibration techniques have enhanced the utility of conceptual models, while physically-based models have continued to languish in the unresolved problems of scale transferability [*Beven*, 2000].

Downward modeling lends itself to a hierarchical scheme that allows the first-order controls to be extracted by varying the model complexity and learning from the data [*Sivapalan et al.*, 2003]. Foremost on the list of desirable benefits of top-down techniques is its ability to reduce overall model complexity without affecting a drastic increase on predictive uncertainty [*Chirico et al.*, 2003].

Atkinson et al. [2003] utilized the downward approach to develop a suite of models of varying complexity to better understand the key processes involved in accurately predicting streamflow at the watershed outlet, on an hourly basis. The use of such an approach made it possible to better define the complexity required to match the external response of the watershed, as well as the value of climate and watershed data and spatial-temporal interactions [*Jothityangkoon et al.*, 2001].

Top-down modeling is also very valuable in not only determining the controls on the external dynamics of the watershed, but was successfully implemented to establish the controls on internal watershed processes through the use of "soft" data [*Vache and McDonnell*, 2006]. This is possible because downward modeling places a high value on the information provided by observed data, with modeled processes being inferred from data rather than preconceived notions of the watershed dynamics [*Sivapalan et al.*, 2003].

However, to date, downward modeling has not yet reached its full potential in hydrology [*Sivapalan et al.*, 2003], just as it took time in other sciences before the "big picture" concepts were broken down into their most basic components, it will take time for the evolution of hydrologic understanding to occur. In the mean time, an approach that incorporates as much of the known processes should be employed to maintain the balance of realism in predictive modeling.

Model Calibration Methods

The implementation of methods such as the top-down modeling philosophy often give rise to model structures that contain parameters that are immeasurable directly [*Sivapalan et al.*, 2003]. Instead, these "effective" parameters must be estimated through indirect means, often through the use of some form of calibration. In hydrology, the goal of model calibration has been focused on finding values for the immeasurable model parameters that result in the best fit to the observed data, most often streamflow at the watershed outlet.

The task of hydrologic model calibration in its earliest form was performed manually, with the modeler tuning effective parameters in an attempt to provide the best fit of the simulated streamflow to the measured streamflow, often using only a visual closeness-of-fit criteria [*Madsen et al.*, 2002]. While the implementation of manual calibration methods have been successful in the past, the cost associated with manual techniques is high due to the amount of time and training required [*Boyle et al.*, 2000]. However, perhaps the biggest shortcoming of manual calibration is its lack of consistency and reproducibility caused by the inherent subjective decisions required in its execution [*Janssen and Heuberger*, 1995].

Over the past several decades, much effort has been made to develop calibration approaches that transcend manual methods and harness the availability of ever-advancing computing power for use in automatic calibration [e.g., *Beven and Binley*, 1992; *Duan et al.*, 1992; *Kuczera and Parent*, 1998; *Sorooshian and Dracup*, 1980]. Automatic calibration encompasses a wide range of different calibration methodologies, from global optimization [e.g., *Duan et al.*, 1994; *Gan and Biftu*, 1996; *Yapo et al.*, 1996] to Monte Carlo [e.g., *Beven and Binley*, 1992; *Freer et al.*, 1996; *Uhlenbrook et al.*, 1999], with additional consideration having been given to multi-objective methods [e.g., *Vrugt et al.*, 2003b; *Yapo et al.*, 1998]. The goal of each of the aforementioned automatic calibration approaches is to reduce the subjectivity common in manual calibration by implementing mathematical functions to evaluate goodness-of-fit [*Vrugt et al.*, 2003a], while at the same time taking advantage of available computing resources.

Global optimization techniques used as a means of automatic calibration of conceptual rainfall-runoff models gained significant popularity due to the desire to break away from the tedium of manual calibration. The obvious benefit of such methods was to obtain the parameter values associated with the global optimum fit of predicted and measured data. However, *Duan et al.* [1992] discovered five characteristics of the response surface that were complicating the problem of optimization in conceptual hydrologic models, namely, regions of attraction, minor local optima, roughness, sensitivity and shape. In a response to the inability of the optimization techniques to handle the complicating factors associated with the response surface, the Shuffled Complex Evolution (SCE-UA) method [*Duan et al.*, 1992] was introduced and has been successfully implemented in several hydrologic studies [e.g., *Gan and Biftu*, 1996; *Hogue et al.*, 2000; *Kuczera*, 1997; *Luce and Cundy*, 1994]. Despite the successes of global optimization techniques, poor parameter identifiability can lead to uncertainty in the modeled results, seriously limiting the usefulness of such methods [*Vrugt et al.*, 2003b]. Due to a lack of uncertainty quantification associated with manual calibration methods [*Janssen and Heuberger*, 1995], as well as with global optimization techniques [*Vrugt et al.*, 2003b], new calibration techniques have been devised with a desire to optimize the simulated result and quantify the uncertainty associated with the parameter values.

Monte Carlo techniques are simple and commonly used approaches to calibration that require the generation of a number of parameter and modeled response realizations [*Binley et al.*, 1991]. A great advantage of Monte Carlo methods is the simplicity of their implementation; the method is useful in that nonlinear parameter interactions are

maintained, but can be computationally demanding in the number of realizations required to adequately sample the parameter surface [*Beven*, 1989].

Multi-objective calibration approaches, aimed at considering a multitude of aspects associated with system behavior, are founded on the principle that no single comparison function can properly account for all the important characteristics of the data [*Vrugt et al.*, 2003a]. Multi-objective calibration can be performed in either a concurrent evaluation procedure [e.g., *Vrugt et al.*, 2003a; *Yapo et al.*, 1998] or as a single-objective procedure done repeatedly, weighting the results of multiple functions considered [e.g., *Madsen*, 2000]. In either case, the goal of considering several different objective functions is to gain a better understanding and interpretation of the fit of the modeled result to the observed system.

Uncertainty Methods

Approaches to uncertainty estimation in hydrologic modeling are closely related to model calibration, but have the added ability to quantify the uncertainty associated with parameter estimates; such approaches range from Monte Carlo to Markov chain Monte Carlo [e.g., *Bates and Campbell*, 2001; *Kuczera and Parent*, 1998] to combination approaches [e.g., *ter Braak*, 2006; *Vrugt et al.*, 2003b].

Of the Monte Carlo approaches, the Generalized Likelihood Uncertainty Estimation (GLUE) technique [*Beven and Binley*, 1992] has become a widely-available and widely-used implementation. GLUE is rooted in the concept that many model realizations may produce equally acceptable results, an occurrence known as equifinality

[e.g., *Beven and Binley*, 1992; *Beven*, 1993; *Freer et al.*, 1996]. As a Monte Carlo method, GLUE involves the random sampling of parameter values, whose response is characterized by the calculation of a generalized likelihood function (often referred to as an objective function); additionally, the GLUE methodology requires that realistic parameter ranges are defined at the onset of the procedure [*Freer et al.*, 1996]. In compliance with the equifinality hypothesis, a threshold is set in the methodology, above which realizations are considered to be "behavioral" responses and below which are considered to be "non-behavioral" responses [*Beven and Binley*, 1992]. The behavioral responses are then considered in developing the uncertainty associated with the parameter values of a particular model structure [*Beven*, 1993]. Many studies have successfully implemented the GLUE framework in their calibration of parameters and analysis of uncertainty [e.g., *Aronica et al.*, 1998; *Beven and Binley*, 1992; *Beven and Freer*, 2001; *Borga et al.*, 2006; *Ewen et al.*, 2006; *Freer et al.*, 1996; *Jacquin and Shamseldin*, 2007].

Markov chain Monte Carlo (MCMC) methods have gained attention as a means of automatic calibration and uncertainty analysis due to their general applicability to complex models [*Gilks et al.*, 1996] , such as conceptual hydrologic models. MCMC sampling seek to sample from the posterior distribution of the parameters using a random walk approach [*Kuczera and Parent*, 1998].

As with Monte Carlo methods, objective mathematical measures are used to evaluate the fit between observed and predicted data; however, in the case of MCMC methods these measures are formal likelihood functions unlike the informal likelihoods used within a Monte Carlo approach [*Mantovan and Todini*, 2006]. When utilized within

a Bayesian inference framework, MCMC methods have the added characteristic of describing the uncertainty associated with the parameter calibration [*Marshall et al.*, 2004]. Under a Bayesian frame, uncertainty assessment is based on the probabilistic principles associated with the posterior distributions of model parameters, sampled by the selected MCMC algorithm and calculated via a formal likelihood measure [*Samanta et al.*, 2007]. A number of hydrologic studies have looked at the utility of MCMC methods (within a Bayesian inference framework) as a calibration/uncertainty analysis technique [e.g., *Bates and Campbell*, 2001; *Gallagher and Doherty*, 2007; *Kavetski et al.*, 2006; *Kuczera and Parent*, 1998; *Marshall et al.*, 2004; *Qian et al.*, 2003; *Renard et al.*, 2006; *Smith and Marshall*, 2008].

In an effort to secure the advantages of multiple automatic calibration methods, while minimizing the disadvantages common to each, a new class of combination calibration techniques has emerged, based on combining the principles of two or more other calibration methods. The use of these methods has become more and more common in hydrologic studies [e.g., *Feyen et al.*, 2007; *Smith and Marshall*, 2008; *Vrugt et al.*, 2003b], as the understanding of the model calibration problem becomes more understood. The most common of these combination approaches seeks to combine the usefulness of global optimization methods in finding the global optimum with Markov chain Monte Carlo approaches to uncertainty evaluation; such as the Shuffled Complex Evolution Metropolis (SCEM-UA) algorithm of *Vrugt et al.* [2003b] and the Differential Evolution Markov Chain (DE-MC) algorithm of *ter Braak* [2006]. While these hybrid approaches can be useful in hydrologic modeling, many of the methods making use of Bayesian

principles (like SCEM-UA) have not been shown to satisfy the basic principles of MCMC approaches such as ergodicity.

Model Assessment

In general terms, model assessment can refer to two primary processes, (1) the ability of a model to represent a variable of the physical system (e.g., stream discharge) and (2) the differentiation between two or more models on the basis of their simulated results. The task of assessing a model's quality of representation of a component of the physical system is a problem that has been oft-addressed in the past, with assessment strategies ranging from statistical approaches that summarize observed data and model fit to others that attempt to diagnose different parts of the model structure. The further obligation to distinguish a “best” model from a set of models has long been studied in statistical modeling with recent applications in a hydrologic context.

Validation testing of hydrologic models is a common practice in assessing the predictive ability of the model through use of a data set independent to that used for model calibration. Over two decades ago, *Klemes* [1986] introduced a validation framework aimed at standardizing the verification of a model's performance. This validation framework was established in a hierarchical fashion to reflect differing scenarios of interest for which hydrologic models could be applied. Today, the principles of Klemes' validation scheme are often evoked during assessment of precipitation-runoff models, perhaps most commonly through use of the split-sample test approach to validation [e.g., *Donnelly-Makowecki and Moore*, 1999; *Refsgaard and Knudsen*, 1996].

Such an approach requires the data set to be split into two samples, a calibration data set and a validation data set, to allow for independent testing of the model performance.

In the application of model validation methods, the use of mathematical functions to evaluate the performance of the models is fundamental in the hydrological sciences [*Krause et al.*, 2005]. In a model comparison sense, objective functions can be evaluated to determine which of a set of competing model structures is better suited to predicting a given setting (topography, climate, etc.). However, *Fenicia et al.* [2007] point out that because no single efficiency criteria can adequately consider all the components related to a model's output, several different measures often need to be utilized and weighted to consider multiple aspects of the modeled response. Several studies have utilized a multi-objective validation approach successfully in hydrology [e.g., *Fenicia et al.*, 2008; *Guntner et al.*, 1999; *Wagener and Kollat*, 2007].

Along with considering multiple objective measures to characterize the model fit, there are a variety of data sources that can be utilized in watershed model evaluation. Traditionally in hydrologic modeling, the focus has been to calibrate and validate the model based on the most common variable of interest, streamflow [*Seibert and McDonnell*, 2002]. As streamflow is the external integration of internal watershed processes, it represents the most important characteristic in terms of water resources management and is of heightened importance in mountain watersheds because of its implications on water supply for downstream communities [*Marks et al.*, 1999].

Ambroise et al. [1995, p. 1467] reasoned that a superior implementation "…should include some test of the internal consistency of the distributed results…". The

benefits of such an implementation are well documented [e.g., *Hooper et al.*, 1988; *Son and Sivapalan*, 2007], where it has been shown that auxiliary data sources have the ability to reduce the predictive uncertainty of modeling results.

Besides the potential for reducing predictive uncertainty, a multiresponse data approach has the added benefit of improving modeling results (parameter identifiability), while (at least partially) preserving parsimonious model structure [*Kuczera and Mroczkowski*, 1998]. The advantages of multiresponse data are seen as especially central in downward model development and the establishment of appropriate structural complexity required to describe the key traits of the system [*Son and Sivapalan*, 2007].

Importantly, though, has been the recognition that not all secondary data is useful. *Kuczera and Mroczkowski* [1998] found that groundwater level data did little to reduce parameter uncertainty, while stream salinity substantially reduced parameter uncertainty in their application of a hydro-salinity model. The challenge is determining what type of data is appropriate for the watershed of interest; this requires insight into the dominant hydrologic forces guiding the system's response and its relationship with the structural development of the model [*Kuczera and Mroczkowski*, 1998]. The multicriteria method is more likely to provide a more structurally appropriate model, better able to predict in conditions other than those for which the model was calibrated [*Mroczkowski et al.*, 1997; *Seibert*, 2000].

Although a major limitation in multiresponse model evaluation is the necessary availability of multiple sources of data (for example, streamflow, groundwater levels, soil moisture or evapotranspiration), some attempt can be made to escape the demand for data

in a multicriteria scheme by utilizing "soft" data sources, fuzzy measures that do not have a specific measured value but represent expert knowledge of the modeled system [*Seibert and McDonnell*, 2002]. The use of soft data in model assessment has been shown to be a useful tool in analyzing model structure and parameter reasonability [*Vache and McDonnell*, 2006].

The importance of validating a given model calibration is of utmost importance, and the use of several sources of data can prove or disprove the reasonability of the modeling realization across multiple levels [*Uhlenbrook and Leibundgut*, 2002]. The use of such an approach is crucial and has the ability of improving model structure, performance and uncertainty across the board [*Bergstrom et al.*, 2002; *Uhlenbrook and Sieber*, 2005].

While the assessment of a given model formulation to an aspect of the physical system is important in understanding the capabilities (and inabilities) of a particular structure, the ability to distinguish the preferred model from a collection of models is also of importance in model assessment. Model selection strategies often seek to balance performance and complexity in selecting the most parsimonious model capable of describing the data [*Atkinson*, 1980].

A variety of measures have been developed for model selection from Akaike's information criterion [AIC, *Akaike*, 1974] to the Bayes' information criterion [BIC, *Schwarz*, 1978] to Kashyap's information criterion [KIC, *Kashyap*, 1982]. Such measures have been considered in past hydrologic studies [e.g., *Neuman*, 2003; *Ye et al.*, 2004].

In understanding the need for the establishment of a generalized methodology to model evaluation, *Wagener* [2003] proposed that a multifaceted approach that would consider performance, uncertainty and realism in the assessment of a model's usefulness. While current approaches are available to consider each of these evaluation categories, few studies have implemented assessment procedures that take into account each.

BAYESIAN METHODS IN HYDROLOGIC MODELING: A STUDY OF RECENT ADVANCEMENTS IN MARKOV CHAIN MONTE CARLO TECHNIQUES

Bayesian methods, and particularly Markov chain Monte Carlo (MCMC) techniques, are extremely useful in uncertainty assessment and parameter estimation of hydrologic models. However, MCMC algorithms can be difficult to implement successfully due to the sensitivity of an algorithm to model initialization and complexity of the parameter space. Many hydrologic studies, even relatively simple conceptualizations, are hindered by complex parameter interactions where typical uncertainty methods are harder to apply.

In this chapter, comparisons between three recently introduced MCMC approaches are presented. The algorithms compared are the Adaptive Metropolis (AM), the Delayed Rejection Adaptive Metropolis (DRAM) and the Differential Evolution Markov Chain (DE-MC). Assessments of the algorithms are made via two case studies; (1) a synthetic Gaussian mixture with five parameters and two modes and (2) a real-world hydrologic modeling scenario where each algorithm will serve as the uncertainty and parameter estimation framework for a conceptual precipitation-runoff model. The research offered in this chapter has been published in the journal Water Resources Research [*Smith and Marshall*, 2008]. *Reproduced by permission of American Geophysical Union.*

Introduction

With a growing desire to better quantify watershed processes and responses, many modeling studies have been undertaken ranging from attempts at developing models based completely on physical process understanding to simple black box methods [e.g., *Atkinson et al.*, 2003; *Beven*, 1989]. Recently, there has been a movement toward a combination approach, including as much of the known physics of the system as possible, while maintaining model structure parsimony [*Littlewood et al.*, 2003]. A driving force in this has been the advent of more sophisticated algorithms capable of automatic parameter estimation and uncertainty quantification, given such model parameters are "effective", non-measureable values.

Model calibration techniques have changed with the availability of ever-faster computing systems, from simple trial-and-error methods to fully computerized algorithms designed to completely investigate the parameter space [*Vrugt et al.*, 2003a]. Automatic calibration techniques are varied in how they attempt to implement objective mathematical procedures and search the parameter space to optimize the model simulation. Commonly used calibration methods range from stochastic global optimization techniques [e.g., *Duan et al.*, 1992; *Sorooshian and Dracup*, 1980; *Thyer et al.*, 1999] to Monte Carlo methods [e.g., *Beven and Binley*, 1992; *Freer et al.*, 1996; *Uhlenbrook et al.*, 1999] to Markov chain Monte Carlo routines [e.g., *Bates and Campbell*, 2001; *Campbell et al.*, 1999; *Kuczera and Parent*, 1998]. In a push to characterize the predictive uncertainty associated with estimated parameter values, Monte Carlo-based approaches have moved to the forefront of automatic calibration routines

[*Feyen et al.*, 2007]. The most frequently implemented variants of Monte Carlo methods include uniform random sampling (URS) often implemented in the popular generalized likelihood uncertainty estimation (GLUE) approach and Markov chain Monte Carlo (MCMC) schemes [e.g., *Bates and Campbell*, 2001; *Marshall et al.*, 2004].

While these methods have been successfully implemented in hydrologic studies, in many cases they also suffer from a variety of problems. In general, all Monte Carlo-based techniques suffer from inefficiency in the exploration of the parameter space [*Bates and Campbell*, 2001]. This is especially true for highly parameterized models, where parameter interactions can be very involved and not adequately explored by the algorithm without an extremely large number of samples [*Kuczera and Parent*, 1998]. Markov chain Monte Carlo-based approaches are more adept at exploring the parameter space in an "intelligent" manner. However, such approaches often suffer greatly from initialization problems associated with the variance of the proposal being either too large or too small, preventing the algorithm from efficiently reaching convergence [*Haario et al.*, 2006]. The additional problem of convergence to the posterior distribution is of significant concern for hydrologic models, as their non-linear nature often leads to a complex parameter response surface with many local optima [*Duan et al.*, 1992].

The present chapter provides a comparison of three recently developed Markov chain Monte Carlo algorithms intended to overcome some of the inefficiencies common to other well established MCMC methods. This chapter is organized into the following sections: a brief description of Bayesian methods in hydrology and introduction to the MCMC algorithms; presentation of the two case studies with their results; and discussion

of the conclusions drawn from this research and applicable to a variety of modeling problems.

Bayesian Methods in Hydrology

Bayesian inference provides a framework for explicitly accounting for modeling uncertainties. At the core of Bayesian inference is the use of formal likelihood functions to analyze parameter uncertainty. Bayesian methods are of particular popularity in environmental science, as they allow for the incorporation of expert knowledge and have the ability to learn from additional data as it becomes available [*Mantovan and Todini*, 2006], an especially desirable property in hydrologic forecasting.

The essential characteristic of Bayesian inference is the description of parameter and model uncertainty as probability distributions, namely posterior distributions. The posterior distribution is calculated based on prior knowledge of the parameters and a likelihood function for the data, given the parameters. The posterior distribution may be determined through the implementation of Bayes' Theorem:

$$P(\theta|Q) = \frac{p(\theta)P(Q|\theta)}{P(Q)} \tag{3.1}$$

where θ are the model parameters, Q are the data, $p(\theta)$ is the prior distribution, $P(Q|\theta)$ is the likelihood function and $P(Q)$ is a proportionality constant. The shape of the posterior distribution is dependent upon the information included in the prior distribution and the information included in the measured data. If the data agrees well with the included prior, the posterior will resemble the prior; however, if the data disagrees with the prior (and there is sufficient data available) the shape of the posterior will be

dissimilar to the prior. Bayesian methods are often of particular use in situations where a portion of the data may be insufficient and expert knowledge can be used to supplement the measured observations.

Given the analytically intractable nature of many hydrologic models, implementation of Bayesian methods are usually aided by Markov chain Monte Carlo (MCMC) techniques. MCMC is based on the favorable mathematical properties of Markov chains, as they relate to Monte Carlo sampling and distribution estimation [*Tierney*, 1994]. Markov chains can be used to generate samples of the posterior distribution of the model parameters, using a random walk approach [*Kuczera and Parent*, 1998]. Many MCMC algorithms have been developed with the aim of constructing useful and statistically relevant Markov chains [see, for example, *Brooks*, 1998]. Formal likelihood functions are used to analyze parameter uncertainty, following parameter estimation by the selected MCMC algorithm [*Bates and Campbell*, 2001].

While there have been a variety of MCMC algorithms developed, the Metropolis algorithm [*Metropolis et al.*, 1953] will be detailed here due to its frequent application to a range of problems, especially those related to hydrologic modeling. The general algorithmic procedure [*Kuczera and Parent*, 1998] is as follows (using previously defined notation):

1. For the i^{th} iteration, randomly draw a sample from a symmetric multivariate jump probability distribution $\theta^* \leftarrow J(\theta|\theta_{i-1})$ where θ_{i-1} is the parameter set sampled at the previous iteration. A common selection for the multivariate jump probability distribution is the multivariate Gaussian distribution.

2. Evaluate the likelihood function, $P(Q|\theta^*)$.
3. Accept the proposed state with probability equal to the ratio:

$$\alpha = \frac{P(\theta^*|Q)}{P(\theta_{i-1}|Q)} \tag{3.2}$$

   If the algorithm samples illogical values for θ, then $\alpha = 0$ and the algorithm remains at the current position.

4. Increment i. If i is less than N, go to step 1.

As with Monte Carlo techniques, Markov chain Monte Carlo methods are not free of deficiencies. Often times, MCMC algorithms struggle to reach convergence due to problems with initialization and selection of an appropriate jump distribution [*Haario et al.*, 2006]. Because of these issues, pre-calibration and tuning runs are often required to identify points of high posterior parameter density [*Haario et al.*, 2005].

Advancements in Markov chain Monte Carlo Techniques

Recent efforts have yielded several new methods designed to improve the efficiency and effectiveness of MCMC algorithms [e.g., *Kou et al.*, 2006; *Laskey and Myers*, 2003; *Tierney and Mira*, 1999]. Three such advancements will be compared here by way of two case studies, the Adaptive Metropolis (AM) algorithm [*Haario et al.*, 2001], the Delayed Rejection Adaptive Metropolis (DRAM) algorithm [*Haario et al.*, 2006] and the Differential Evolution Markov Chain (DE-MC) algorithm [*ter Braak*, 2006].

Adaptive Metropolis Algorithm

The Adaptive Metropolis (AM) algorithm is a modification to the standard random walk, Metropolis-based Markov chain Monte Carlo methods, conceived by *Haario et al.* [2001]. The key attribute of the AM algorithm is its continuous adaptation toward the target distribution via its calculation of the covariance of the proposal distribution using all previous states. Utilizing this attribute, the proposal distribution is updated using the information gained from the posterior distribution thus far. At step i, *Haario et al.* [2001] consider a multivariate normal proposal with mean given by the current value and covariance matrix C_i. The covariance C_i has a fixed value C_0 for the first i_0 iterations and is updated subsequently:

$$C_i = \begin{cases} C_0 & i \leq i_0 \\ s_d cov(\theta_0, \cdots, \theta_{i-1}) + s_d \varepsilon I_d & i > i_0 \end{cases} \tag{3.3}$$

where ε is a small parameter chosen to ensure C_i does not become singular, I_d denotes the *d*-dimensional identity matrix and s_d is a scaling parameter depending on the dimensionality, d, of θ, the parameter set, to ensure reasonable acceptance rates of the proposed states. As a basic guideline, *Haario et al.* [2001] suggest choosing s_d for a model of a given dimension as $2.4^2/d$.

An initial, arbitrary covariance, C_0, must be defined for the proposal covariance to be calculated. To this end, the initial covariance is set as the covariance of the parameters under their prior distributions. The steps involved in implementing the AM algorithm then become the following [see *Marshall et al.*, 2004]:

1. Initialize $i = 0$.
2. (a) Select C_i for the current iteration number i using equation (3.3).

(b) Generate a proposed value θ^* for θ where $\theta^* \sim N(\theta^i, C_i)$.

(c) Accept the proposed state with probability equal to α using equation (3.2).

3. Repeat step 2.

Block updating (sampling all parameters concurrently) is utilized in the AM algorithm, enhancing the computational efficiency and reducing run time. While the AM algorithm has many beneficial traits, it can potentially experience difficulties with initialization (sampling an appropriate starting parameter set from a place of high posterior density) and in exploring the parameter space if the parameters are considerably non-Gaussian, given the proposal distribution is a multivariate Gaussian. Although the Adaptive Metropolis algorithm is not a true Markov chain due to the adaptive component, results establishing the validity and the ergodic properties of the approach have been proven [*Haario et al.*, 2001].

Delayed Rejection Adaptive Metropolis Algorithm

The Delayed Rejection Adaptive Metropolis (DRAM) algorithm is based on the combination of the Adaptive Metropolis algorithm and the Delayed Rejection (DR) algorithm, introduced by *Tierney and Mira* [1999] and applied to a basic Metropolis-Hastings approach. A detailed explanation of the steps involved in combining the DR and AM algorithms and implementing the DRAM algorithm is provided by *Haario et al.* [2006].

The DR algorithm is based on the concept that the performance of MCMC methods is improved by decreasing the probability that the algorithm remains at the current state, as proven by *Peskun* [1973] for finite state spaces and *Tierney* [1998] for

general state spaces. In general terms, the DR algorithm allows for a more expansive search than traditional MCMC algorithms, as the parameter space explored by the algorithm can initially be set to attempt to sample regions of the response surface that are far from the current location by using a larger variance of the proposal. The variance of the proposal is then reduced if the parameter walk does not move; thereby searching a region of the parameter space closer to the current position. By creating multiple proposal stages the probability of remaining at the current state is reduced.

A stage one proposal is generated and accepted or rejected in the same manner as in the Metropolis algorithm, however, if the stage one proposal is rejected, a stage two proposal is drawn from a region closer to the current position. If the stage two proposal is accepted, it becomes the current position. If the stage two proposal is rejected the position held prior to the failed stage one and two proposals is retained. While only two proposal stages are outlined here, this structure allows for as many proposal stages as desired.

In mathematical terms, suppose the current position is x and a proposal to move to a new candidate, y, is generated from $q_1(x, \cdot)$, a multivariate proposal distribution, with the acceptance probability

$$\alpha_1(x, y) = min\left\{1, \frac{\pi(y)}{\pi(x)} \frac{q_1(y, x)}{q_1(x, y)}\right\} \tag{3.4}$$

where $\pi(\cdot)$ represents the likelihood function evaluated at the given proposal (x, y, z). Upon rejection, instead of remaining at the current state $\theta_{i+1} = \theta_i$, a second stage proposal, z, is produced. This second stage proposal depends on the current state and the suggested and failed stage one proposal, $q_2(x, y, \cdot)$; the second stage is accepted with the probability

$$\alpha_2(x,z) = min\left\{1, \frac{\pi(z)}{\pi(x)}\frac{q_1(z,y)}{q_1(x,y)}\frac{q_2(z,y,x)}{q_2(x,y,z)}\frac{[1-\alpha_1(z,y)]}{[1-\alpha_1(x,y)]}\right\} \quad (3.5)$$

If the stage two candidate is rejected, either a stage three proposal is invoked or the algorithm remains at the current position.

The relationship between the AM and DRAM algorithms may be illustrated by way of a simple diagram. Figure 3.1 shows a hypothetical, two-dimensional parameter space, with the model's log-likelihood function on the vertical axis; we initialize the algorithm at the parameter set indicated by point *X*. The ability of the DRAM algorithm to search a larger portion of the parameter space in its stage one proposal (indicated by

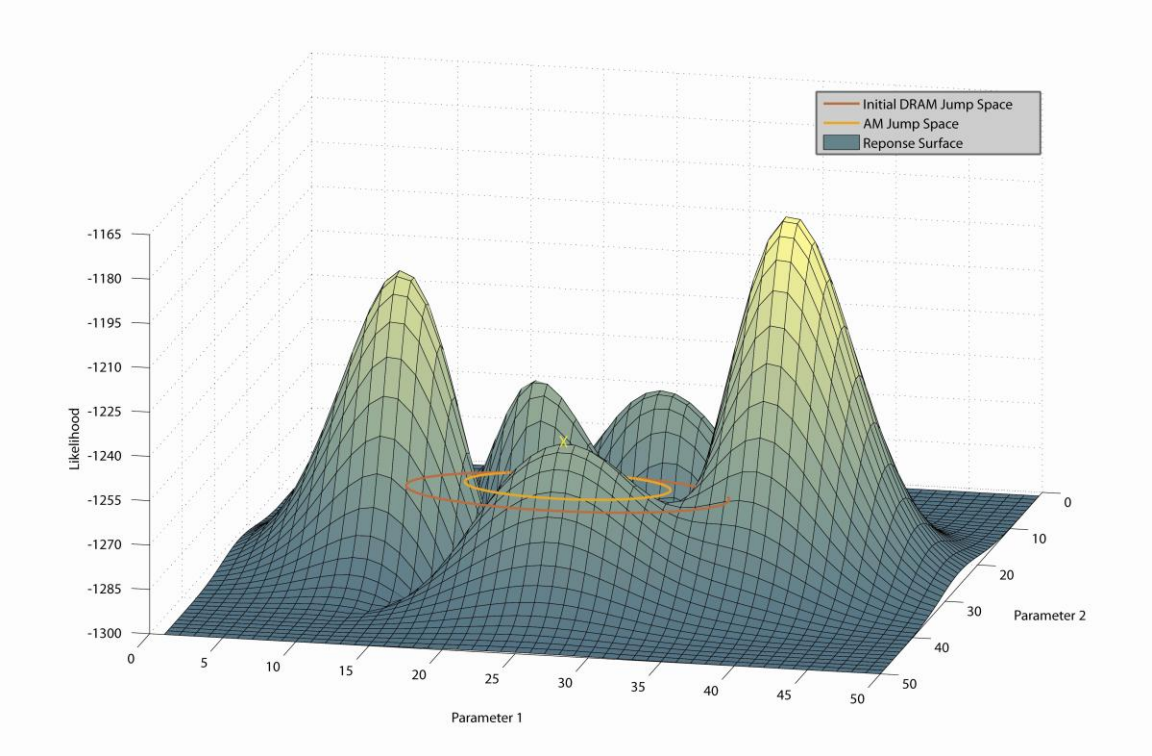


Figure 3.1: Hypothetical parameter surface illustrating the ability of the AM and DRAM algorithms to explore the parameter surface. Rings represent distance from the current location an algorithm can explore.

the red ellipse) than the AM (indicated by the orange ellipse) is illustrated. If the stage one proposal of the DRAM fails, its search of the parameter space is reduced by a user-

set factor. Although the AM is a sub-algorithm of the DRAM, the computational efficiency of DRAM is reduced by having to perform multiple executions of the model for a single algorithm iteration. In this manner, the AM algorithm can be seen as maintaining the benefits of a computational efficiency advantage over the DRAM algorithm.

The DRAM algorithm benefits from its ability to make use of two concepts that, alone, have made improvements over the standard MCMC approaches. Namely, DRAM has the ability to adapt continuously toward the target distribution, constantly update its calculation of the covariance of the proposal distribution using all previous states and more efficiently search the parameter space by reducing the probability that the algorithm will remain at the current state. *Haario et al.* [2006] point out that from the DR point of view the first stage is aided by the adaptiveness of the AM component, while from the AM point of view the DR component offsets the initialization difficulties often associated with the AM component by searching for sets far away from the current position in stage one proposals and sets close to the current position in stage two proposals. The AM component can be seen as responsible for "global" adaptation, while the DR component is responsible for "local" adaptation within the DRAM algorithm.

Differential Evolution Markov Chain Algorithm

The Differential Evolution Markov Chain algorithm is formed by combining the Differential Evolution algorithm of *Storn and Price* [1997] designed for global optimization in real parameter spaces with MCMC simulation, utilizing standard

Metropolis principles. The result is a population MCMC algorithm, where multiple chains are run in parallel and allowed to interact with each other.

This combination intends to overcome the difficulties common to MCMC methods of choosing an appropriate scale and orientation for the proposal distribution, while also addressing the issue of computational efficiency related to time to reach convergence [*ter Braak*, 2006]. There are only two parameters to be defined by the user, a scaling factor, γ, and the number of parallel chains, N.

ter Braak [2006] provides the basic algorithmic setup for DE-MC. N chains are run in parallel and the jumps for a current chain are derived from the remaining $N-1$ chains. The DE-MC algorithm generates proposals based on two randomly selected chains that are subtracted from each other, multiplied by a scaling factor and added to the current chain

$$\theta_p = \theta_i + \gamma(\theta_{R1} - \theta_{R2}) + \varepsilon \tag{3.6}$$

where θ_p is the proposed parameter set, θ_i is the current parameter set, θ_{R1} and θ_{R2} are randomly selected parameter sets from the population excluding θ_i and γ is the scaling factor. The final term, ε, is intended to introduce a probabilistic acceptance rule to the DE algorithm, with $\varepsilon \sim N(0, b)^d$ and b being small. The proposals are then accepted or rejected based on the Metropolis ratio, defined in equation (3.2) previously.

While the DE-MC algorithm is fairly simple to implement, it is also quite useful in practice. The algorithm's multiple chains maintain asymptotic independence that allows convergence testing using the *Gelman and Rubin* [1992] *R*-statistic to be performed from one DE-MC run.

Case Studies

In this section, we present two case studies to illustrate the differences in performance between the AM, DRAM and DE-MC algorithms. First, a synthetic example is constructed based on a mixture of two multivariate normal distributions, generating a bimodal response surface. The second example illustrates the capabilities of the algorithms applied to a hydrologic setting, featuring a simple conceptual precipitation-runoff model with data from the Tenderfoot Creek Experimental Forest, MT. In each example, the three algorithms will be compared in their ability to explore the parameter space and converge to the posterior, with additional consideration given to computational efficiency.

The potential benefits of these newly introduced MCMC approaches for hydrologic studies are of great interest. The ability to more accurately and efficiently search the parameter space provides potentially enduring gains to the field of conceptual precipitation-runoff modeling. Multimodality is of great concern in hydrologic modeling, and the ability of each algorithm to handle such complexities should be considered. The AM and DRAM algorithms have the same adaptive component added to the traditional Metropolis formulation, with the DRAM having the further component of delayed rejection. The DE-MC algorithm is a combination of a genetic algorithm with a traditional Metropolis approach. Each algorithm has an associated efficiency due the specific components included in its particular formulation; the tradeoff between the algorithm's computational effort and its efficiency in characterizing the true posterior distribution will be considered for each of the three algorithms.

Case Study 1: Synthetic Bimodal Gaussian Mixture

In order to provide a controlled setting in which to compare the algorithms, a multivariate Gaussian model was defined to emulate a parameter surface with multiple modes due to the commonality of multimodal response surfaces in hydrologic modeling. The likelihood of this mixture of normals is proportional to

$$L = 0.3 \cdot \exp\left(-(\theta - \mu_1) \cdot \Sigma_1^{-1} \cdot (\theta - \mu_1)^T\right) + \cdots \\ 0.7 \cdot \exp\left(-(\theta - \mu_2) \cdot \Sigma_2^{-1} \cdot (\theta - \mu_2)^T\right) \tag{3.7}$$

where μ represents the mean of the given mode, Σ is the covariance of the given mode and θ is the parameter set. For this study, we set the model dimension at five variables.

The AM, DRAM and DE-MC algorithms were applied to the defined model. Each algorithm was run at two fixed numbers of iterations, 10,000 and 50,000, to understand the parameter space exploration and computational efficiency for each algorithm under fixed iteration budgets of different sizes. The scaling factors for the DRAM algorithm were set up such that the stage one proposal was approximately 1.35 times larger than the AM algorithm proposal and the stage two proposal was one-hundredth as wide as the stage one proposal. The DE-MC was initialized with γ set to the suggested value of $2.38/(2d)^{0.5}$ recommended by *ter Braak* [2006], where d is the number of model parameters and N was set to 10 (twice the number of parameters in the study). Although it is suggested that the number of parallel chains required for multimodal response surfaces is at least ten times the model dimension when using the DE-MC algorithm, this requirement was found to be prohibitive for even relatively modest numbers of iterations (50,000 iterations would require 2.5 million evaluations for 50 chains).

Figure 3.2 shows the sampled parameter values for each algorithm run for 50,000 iterations, as well as providing the true distribution of the Gaussian mixture. The results for each algorithm were similar for each of the iteration levels (10,000 and 50,000 iterations) with the AM algorithm only identifying one mode, while the DRAM and DE-MC algorithms sampled from each of the modes. It was found for all cases that even 50,000 iterations were insufficient to completely characterize the response surface. Each algorithm maintained acceptance rates within the recommended range of approximately 20-45% suggested to obtain optimal algorithm efficiency [e.g., *Gelman et al.*, 2004]; moreover, all algorithms maintained similar acceptance rates across the two levels of iterations examined.

From this simple case study, it is clear that both the DE-MC and DRAM algorithms more readily traverse the parameter space in complex, multimodal situations than the AM algorithm (Figure 3.2). Not surprisingly, the DRAM algorithm proved to be an improvement upon the Adaptive Metropolis algorithm because of its increased probability of moving from the current state, brought about by the DR component of the algorithm. While the DE-MC algorithm also proved to be more efficient than the AM algorithm (and perhaps the DRAM algorithm as well), it does benefit from the ability of having multiple parallel chains learning from each other. In this case, ten parallel chains were used, providing the DE-MC algorithm with ten times the number of sample draws.

Investigating this further, both the AM and DRAM algorithms were set up to perform the same number of total sample draws as the DE-MC performed for the 10,000 iteration level (100,000 for this study). Under this setup, the AM was able to characterize

both modes and the DRAM's characterization of the entire distribution was more complete than at 10,000 and 50,000 iterations.

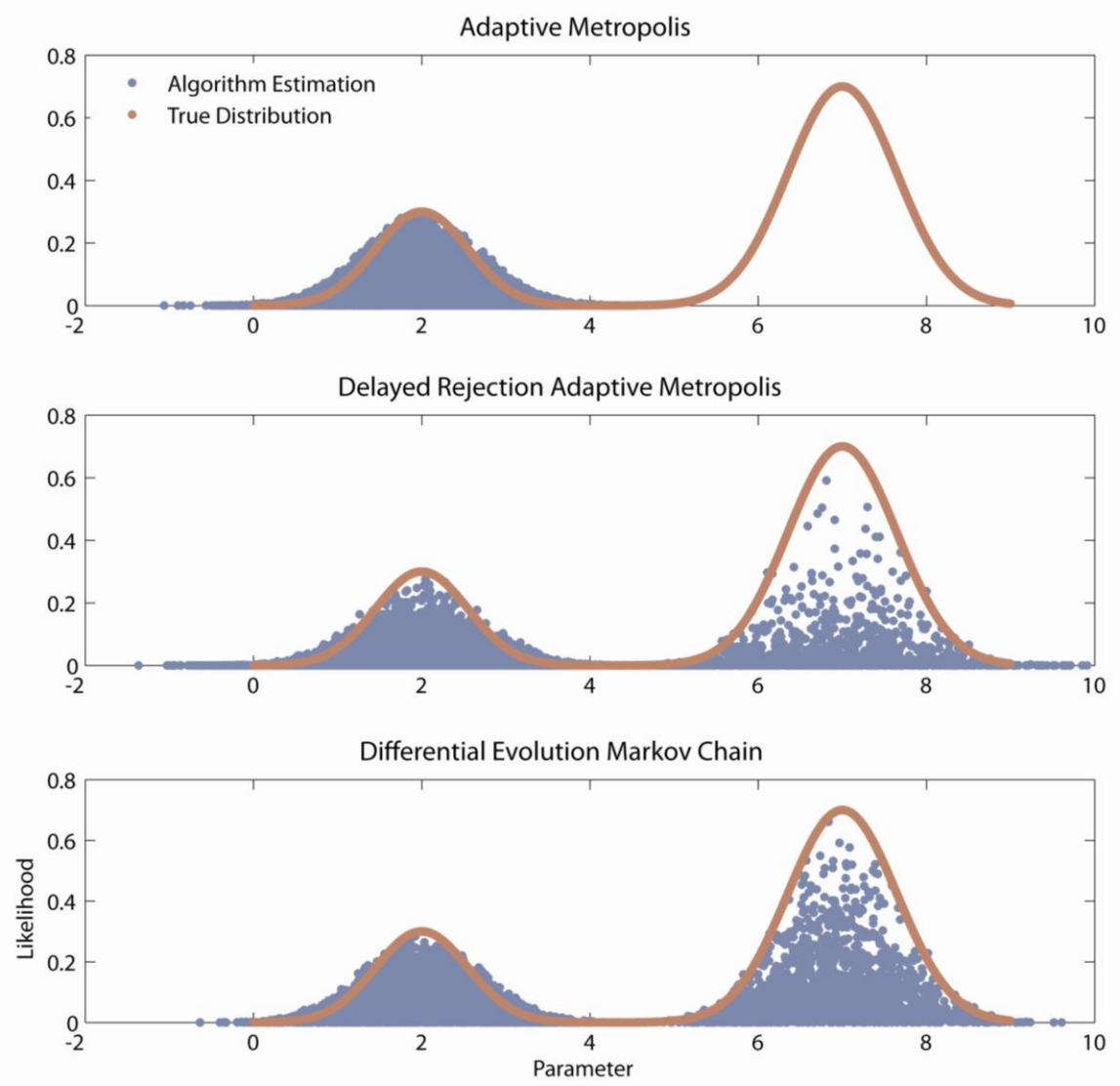


Figure 3.2: Comparison of parameter estimation between the AM, DRAM and DE-MC algorithms for 50,000 iterations and one parameter for a five-dimensional Gaussian mixture. Values shown in blue were sampled in 5-D space and were projected into 1-D space. The exploration of the parameter surface is weakest in the AM algorithm, with only one mode sampled, while the DRAM and DE-MC algorithms manage to sample each mode.

This simple mixture of multivariate normal distributions provides a relatively objective testing ground for the three algorithms of concern, where the end result is

known a priori. The results of this case study reveal the ability of the DRAM and the DE-MC algorithms to explore the parameter space in a wider reaching, more efficient manner than the AM algorithm. While the Differential Evolution Markov Chain algorithm had a potentially superior ability to characterize the multimodal distribution compared to the DRAM, it benefits from multiple parallel chain interaction unlike either the AM or DRAM algorithms and hence suffers from increased computational cost. The increased efficiency of the DRAM and DE-MC algorithm searches, however, holds particular importance to hydrologic modeling because of the tremendous complexity of the system being modeled.

Case Study 2: Tenderfoot Creek Experimental Forest, MT

This second case study applies the AM, DRAM and DE-MC algorithms to a hydrologic model for an experimental watershed. The following sub-sections discuss the site, the conceptual precipitation-runoff model, the input data and the outcomes associated with each algorithm.

Site Description: The Tenderfoot Creek Experimental Forest (TCEF) is located at the headwaters of Tenderfoot Creek in the Little Belt Mountains of the Lewis and Clark National Forest in Montana, USA. TCEF was established in 1961 and is representative of the vast expanses of lodgepole pine (*Pinus contorta*) found east of the continental divide. The watershed encompasses an area of nearly 3,700 hectares, in all.

TCEF consists of seven distinct watersheds with differing vegetation, topographic characteristics and silvicultural treatments. The watershed studied in this project was the

555 hectare Stringer Creek Watershed. For a more detailed description of TCEF, see *Jencso et al.* [2008, In Review] and Chapter 4 in this thesis.

Precipitation-Runoff Model Description: This study implemented a model based on the Probability Distributed Model (PDM, Figure 3.3), developed by *Moore* [1985; 2007]. The PDM is a conceptual rainfall-runoff model that seeks to balance model structural parsimony with watershed physical complexities. As a conceptual model, the PDM is concerned only in "the frequency of occurrence of hydrological variables of certain magnitudes over the basin without regard to the location of a particular occurrence within the basin" [*Moore*, 1985, p. 274].

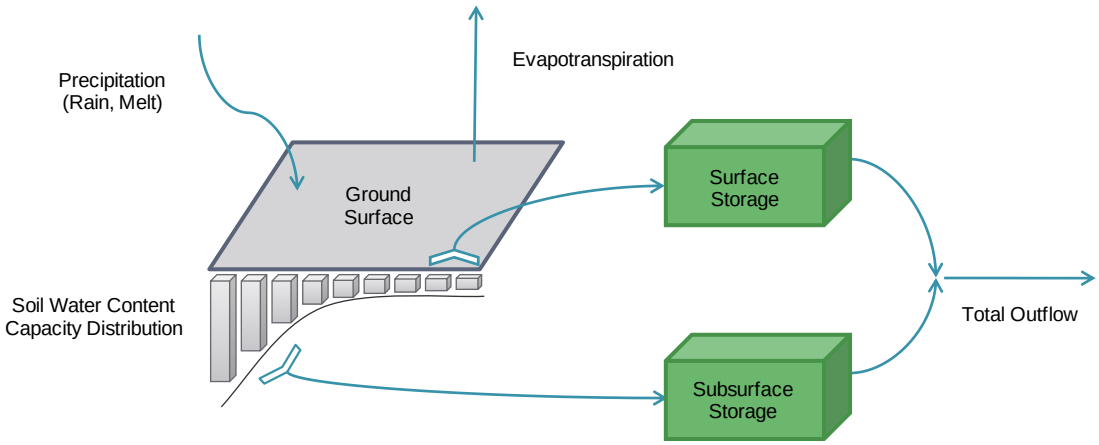


Figure 3.3: The Probability Distributed Model with snowmelt as an additional precipitation input.

Soil absorption capacity controls the runoff produced by the model, based on the spatial variability of soil capacities across the watershed. Water in excess of the soil capacity is routed to the surface storage component, while infiltrated water eventually enters the subsurface storage component; the combination of the outflows of the two storage components comprises the total outflow of the watershed.

The model was supplemented with a snowmelt routine to account for the dominant form of precipitation common to the study area. In this case, a relatively simple combined temperature- and radiation-index approach was used. Previous studies [e.g., *Brubaker et al.*, 1996; *Kustas et al.*, 1994] have shown that incorporating radiation into the temperature-index approach can greatly improve modeled results, while also improving the physical reality of the model.

The model is then comprised of nine effective parameters: maximum soil storage capacity (*CMAX*, [mm]), spatial variability within the watershed (*B*, [-]), rate of drainage into subsurface storage (*KB*, [12-hr^{-1}]), fraction of subsurface storage released to outflow (*TRES1*, [-]), fraction of surface storage released to outflow (*TRES2*, [-]), soil storage threshold for subsurface inflow (*CF*, [mm]), threshold temperature for snowmelt (*TF*, [^{o}C]), degree-day factor for snowmelt (*DDF*, [mm/^{o}C]) and net radiation factor for snowmelt (*NRF*, [-]).

Input Data: The required inputs to the model include precipitation (rain, snow), evapotranspiration, temperature and net radiation. Data required for model operation was obtained from the Onion Park SNOTEL site located within TCEF, featuring sensors recording snowpack depth, snow water equivalents, accumulated precipitation (snow and rain), temperature, radiation and others. Based on a desire to better characterize the diurnal dynamics of the system, a 12-hour time step was selected and data was retrieved from the NRCS National Weather and Climate Center, managing data collection for all SNOTEL sites. All stream flow data was obtained from the network of stream gauges

located within TCEF managed by the Rocky Mountain Research Station of the U.S. Forest Service.

Algorithm Evaluations: For this case study the model and watershed data were kept the same for all analyses and each of the three MCMC algorithms were implemented (AM, DRAM, DE-MC). Each algorithm featured differing internal parameters (scaling factors, number of chains, etc.) which were selected to maintain appropriate acceptance rates. The stage one jump space of the DRAM algorithm was approximately 1.67 times as wide as the jump space of the AM algorithm, with the stage two jump space of the DRAM algorithm being one-hundredth as wide as its stage one proposal. The DE-MC algorithm was initialized with its jumping factor (γ) set to the suggested value of $2.38/(2d)^{0.5}$. As with the synthetic case study, the number of parallel chains was set to twice the model dimension due to computational restraints. All algorithms maintained appropriate acceptance rates [see *Gelman et al.*, 2004] in all cases. The algorithms were compared in their ability to converge to the posterior density and their effectiveness in searching the posterior parameter space.

To diagnose the necessary number of iterations to be performed for the algorithm to obtain convergence, multiple runs of the algorithms were conducted and the between- and within-chain variance was estimated, using the method described by *Gelman and Rubin* [1992]. Two hundred thousand iterations (for each parallel run) were chosen for this study to ensure that the estimated potential scale reduction (*R*-statistic) for each parameter was less than 1.2.

Two situations were considered in this case study, the performance of each algorithm when started from points of known, high posterior density and the performance of each when started from points of known, low posterior density, such as a localized optimum. In each case, the points of known posterior density were determined from tuning runs and a series of multiple parallel runs were performed using this pilot information.

Starting at a point of low posterior density, Figure 3.4 shows the trace of the log-likelihood value for each algorithm. At a large number of iterations (approximately 195,000) the DRAM samples a region of higher posterior density, while the AM and DE-MC algorithms falsely converge to a local optimum for all sampled values. Though the AM algorithm is more likely than the DRAM to converge to local optima (due to the single stage proposal), the convergence of the DE-MC algorithm to a local optimum is somewhat surprising given its beneficial interaction between multiple chains. Based on the convergence criteria of *Gelman and Rubin* [1992], the results for both the DE-MC and AM algorithms indicate convergence for all ten model parameters. The DRAM algorithm, on the other hand, has not converged based on the criteria, which can be clearly seen in Figure 3.4 where the likelihood trace experiences a large increase near 195,000 iterations. Because the DRAM algorithm locates a region of higher posterior density, the within-chain and between-chain variability is no longer such that convergence is confirmed.

When considering the other scenario of interest, starting from a known, high posterior density, the convergence criteria suggests that all three of the studied algorithms

have converged (based on ten parallel runs for AM and DRAM, 20 for the DE-MC). While each algorithm converges to the same maximum log-likelihood value, the DRAM algorithm appears to better sample from the tails of the parameter distributions.

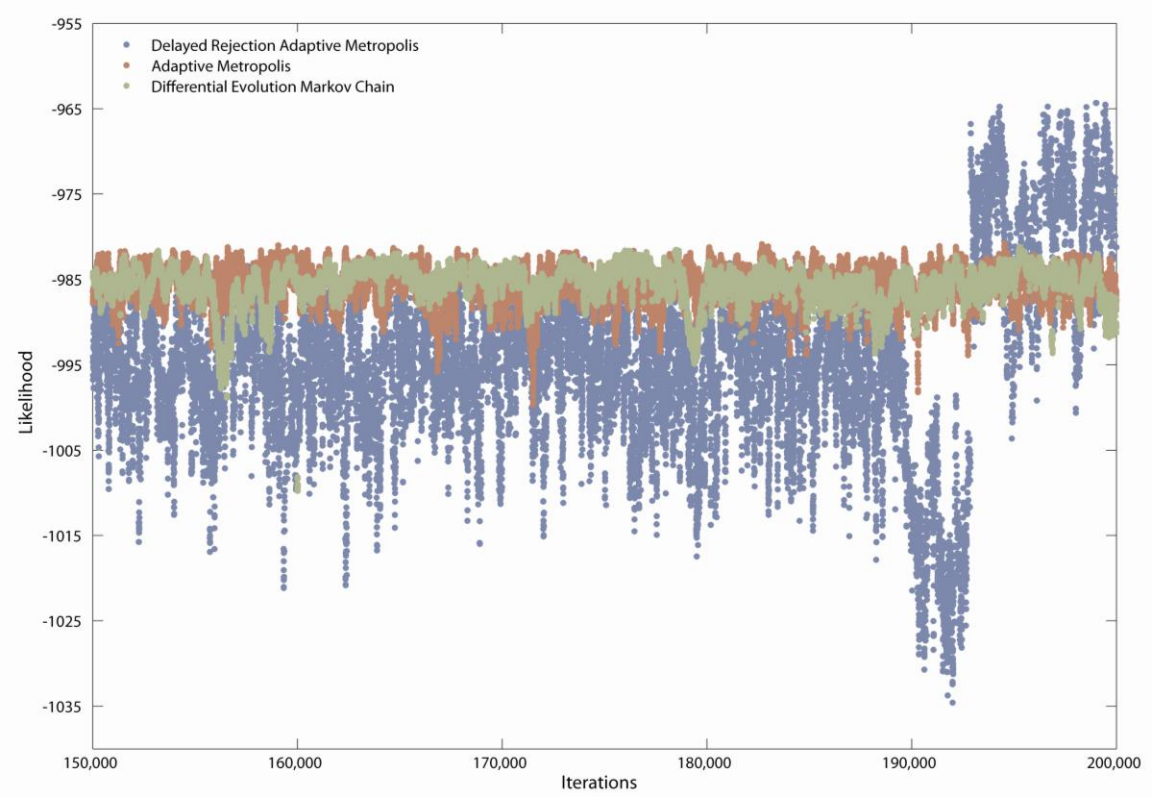


Figure 3.4: Comparison of AM, DRAM and DE-MC parameter searches, started from a point of known low posterior density. The AM and DE-MC falsely converge to a local optimum, while the DRAM locates a region of higher posterior density.

Table 3.1 shows the 2.5 and 97.5 percentiles for each parameter and the value of each parameter resulting in the maximum log-likelihood, for each algorithm. The DRAM algorithm displays a superior ability to sample a wider range of values for each parameter, in contrast to the Adaptive Metropolis algorithm and even the DE-MC algorithm. However, the improved sampling coverage comes at a cost to the computational effort.

In considering the efficiency of the algorithms evaluated in this study, the number of model runs performed for each is of significant concern. In all cases, the Adaptive Metropolis algorithm performed the same number of model evaluations as algorithm

Table 3.1: The 2.5 and 97.5 Percentile Sampled Parameter Values for the AM, DRAM, and DE-MC Algorithms and Values Corresponding to the Maximum log-Likelihood[a]

Parameter	AM			DRAM			DE-MC		
	2.5%	*Max log-Likelihood Value*	*97.5%*	*2.5%*	*Max log-Likelihood Value*	*97.5%*	*2.5%*	*Max log-Likelihood Value*	*97.5%*
CMAX	187.30	188.00	189.10	186.90	188.00	189.90	187.30	188.00	188.70
B	0.0021	0.0100	0.0120	0.0004	0.0099	0.0153	0.0033	0.0100	0.0122
KB	0.0334	0.0336	0.0342	0.0326	0.0337	0.0344	0.0334	0.0336	0.0342
TRES1	0.0259	0.0263	0.0276	0.0252	0.0267	0.0282	0.0259	0.0264	0.0275
TRES2	0.2170	0.2282	0.2427	0.2045	0.2351	0.2490	0.2153	0.2312	0.2423
CF	175.50	176.00	177.70	175.30	176.00	178.40	175.50	176.00	177.30
TF	2.1170	2.1280	2.1300	2.1170	2.1220	2.1300	2.1170	2.1290	2.1300
DDF	0.2718	0.2823	0.2984	0.2708	0.2833	0.3090	0.2710	0.2852	0.2996
NRF	10.290	10.420	10.510	10.210	10.410	10.530	10.280	10.390	10.510
VARP	0.1171	0.1228	0.1297	0.1139	0.1252	0.1364	0.1168	0.1241	0.1295

[a]Parameter values start from a point of known high posterior density. For each parameter, the DRAM samples a wider range of values, indicating its ability to better characterize the tails of the distribution.

iterations, 200,000 evaluations for each run of the algorithm. The Differential Evolution Markov Chain algorithm, due to its requirement of multiple parallel chains, performed 4 million model evaluations, 20 times (the value of the algorithm parameter N) as many evaluations as algorithm iterations. The DRAM algorithm experiences a variable number of model evaluations for any given run of the algorithm, based on the multiple proposal stages of the delayed rejection step. In the case of known, high posterior density starting points, the DRAM algorithm performed approximately 360,000 model evaluations for each algorithm chain. For the case of known, low posterior density starting values, approximately 350,000 model evaluations were performed for each run of the DRAM algorithm. The delayed rejection step of the DRAM algorithm is both responsible for the

increased computational cost of the algorithm and the enhanced efficiency of characterizing the true posterior distribution.

Conclusions

The implementation and use of Markov chain Monte Carlo algorithms in hydrologic modeling is becoming more and more common, especially as computing power becomes less of a limiting factor. While the use of traditional MCMC algorithms is not an uncommon event in hydrology [e.g., *Bates and Campbell*, 2001; *Kuczera and Parent*, 1998; *Qian et al.*, 2003; *Renard et al.*, 2006], few studies have taken advantage of the recent advances in such algorithms.

Our case studies show that the Delayed Rejection Adaptive Metropolis algorithm has the ability to improve modeling results where the posterior density is low and the parameter surface is extremely complex, with many local optima. The DRAM algorithm outperformed the Adaptive Metropolis algorithm in its ability to explore the parameter surface with a limited number of iterations in a synthetic example featuring a bimodal parameter distribution. The DE-MC algorithm proved to be best at defining the parameter surface in this case, however, benefited largely due to its ten-fold advantage in sample evaluations.

A second case study illustrated the utility of the DRAM algorithm when applied to real data and a hydrologic model. When the modeling began from a point of low posterior density, the DRAM algorithm proved to be more able to traverse the complex parameter surface than either the Differential Evolution Markov Chain or the Adaptive

Metropolis algorithms. While the AM and DE-MC falsely converged to a local optimum, the DRAM was able to discover a region of higher posterior density, providing the best set of parameter values based on the calibration data from Tenderfoot Creek Experimental Forest, MT. In addition to finding a better log-likelihood value, the DRAM algorithm more efficiently and extensively searched the tails of the parameter distributions.

This case study has illustrated the potential consequences associated with hydrologic modeling when starting from a point of low posterior density. While the DE-MC and AM algorithms became trapped in a local optimum, the DRAM algorithm, benefiting from a more robust sampling of the tails of the parameter distributions, avoided convergence to areas of local optima.

As with the first case study, the benefits of increasing the probability of moving from the current state can be clearly seen in Figure 3.4, with the DRAM locating a point of higher posterior density. Again, this is the expected result, especially for complex situations where points of high posterior density are unknown at the onset of the study. While the AM algorithm is more likely to falsely converge to a local optimum (refer to Figure 3.1), the delayed rejection component of the DRAM algorithm reduces the probability of such an outcome.

The DE-MC algorithm potentially faces tremendous computational demand for multimodal, high-dimensional problems that require large numbers of iterations. The guideline [suggested by *ter Braak*, 2006] of using at least ten times the number of chains as model dimensions severely limits the collection of problems that can be solved in a

reasonable manner using DE-MC. Although the DRAM algorithm has greater computational constraints (implementation of the algorithm code and its computational execution) than the AM algorithm due to the inclusion of the delayed rejection step, the utilization of an algorithm that can greatly reduce the risk of false convergence to local optima is fundamental to the application of hydrologic models for predictive purposes and should not be underestimated.

The success of the DRAM algorithm is ingrained in the concept that the benefits of the adaptive nature of the AM algorithm are complemented by the benefits of the delayed rejection step of the DR algorithm through their combination, with each working to alleviate the deficiencies of the other while maintaining the desirable statistical properties fundamental to MCMC.

Acknowledgements

This work was made possible due to partial funding by the Montana Water Center. We thank the U.S. Forest Service Rocky Mountain Research Station and the NRCS National Weather and Climate Center for providing the data necessary to perform this study.

DEVELOPMENT OF A CONCEPTUAL PRECIPITATION-RUNOFF MODELING SUITE

Building on the concepts presented in Chapter 2, this chapter describes the collection of precipitation-runoff models to be considered in this study, while also introducing the selected area within which the models will be applied. A framework for model development is instituted that allows for multiple site conceptualizations to be considered through 30 independent alternate modular structures.

Introduction

Hydrologic modeling has taken on many conceptual forms in past studies, covering the entire spectrum from physically-based to conceptual to black-box formulations. However, the bulk of studies focusing on the ability of a model to represent a physical system fail to consider a range of competing alternate, hypotheses aimed at addressing the modeling question [*Neuman*, 2003]. While the task of estimating uncertainty in the model and its parameter estimates has become increasingly studied [e.g., *Bates and Campbell*, 2001; *Beven and Binley*, 1992], *Neuman* [2003, p. 292] points out that "[t]he bias and uncertainty that result from reliance on an inadequate conceptual-mathematical model are typically much larger than those introduced through an inadequate choice of model parameter values."

While much work has been focused on the development of models to describe the physical system, much is still unknown about the primary drivers of watershed processes. In this chapter, a modeling framework is introduced consisting of a collection of flexible,

modular model structures and variable complexities (number of model dimensions) focusing on previously established conceptual precipitation-runoff models. The motivation of this design is to perform a comparative analysis concentrating on characterizing the relationship between structural complexity and predictive performance across the extent of the modeling framework.

Site Description

Recognizing the need to select a site that offers broad applicability to the modeling approach laid out in the research and that is also representative of the conditions typical to mountain watersheds in the Rocky Mountain West, a location in Central Montana has been selected. The selection of Tenderfoot Creek Experimental Forest for this study was motivated by a climatic pattern typical of regional mountain locations, the availability of historical long-term records, the existence of newly installed instrumentation providing the availability of detailed data and the presence of multiple watersheds with varying levels of alteration.

Tenderfoot Creek Experimental Forest

Positioned at the headwaters of Tenderfoot Creek in the Little Belt Mountains of the Lewis and Clark National Forest, Tenderfoot Creek Experimental Forest (TCEF, $46^{o}55'$ N, $110^{o}52'$ W) is approximately 115 kilometers southeast of Great Falls, Montana (Figure 4.1a). Committed to the research of subalpine forests influenced by considerable expanses of lodgepole pine (*Pinus contorta*) typically found on the east slope of the

northern Rocky Mountains, TCEF was established by the United States Department of Agriculture (USDA) in 1961.

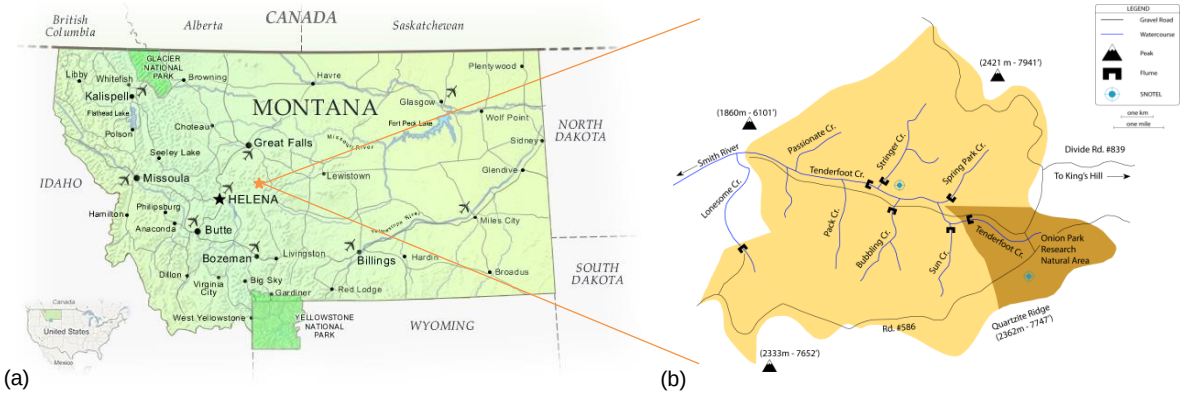


Figure 4.1: Tenderfoot Creek Experimental Forest; (a) location and (b) instrumentation. Adapted from Jencso et al. (2008, In Review) and <http://www.fs.fed.us/rm/ecology/demo/tenderfoot/>.

Tenderfoot Creek Experimental Forest, having a total area of approximately 3,700 hectares, is comprised of seven distinct watersheds. Each of TCEF's watersheds is comprised of unique characteristics (including different vegetation types and topographic traits) with areas ranging from approximately 300 to 2,270 hectares and elevations varying from approximately 1,800 to 2,400 meters. Among the dominant vegetation types, lodgepole pine and mixed lodgepole pine with Engelmann spruce and subalpine fir stands comprise about 95% of the total area of TCEF, wet meadows cover less than 0.5% and drier grass and scree slopes make up another 1.5% [*Woods et al.*, 2006].

Average annual precipitation across the entirety of TCEF is approximately 850 millimeters per year, largely in the form of snow. The precipitation pattern generally sees maximum monthly precipitation in December or January with a typical dry period during the months of July to October and a wet period from November to May. Peak streamflows commonly occur in late May or early June in a response to spring

precipitation and melting snowpacks, while low streamflows are persistent from late summer through winter.

The watersheds within TCEF are highly instrumented (see Figure 4.1b), featuring SNOTEL (SNOwpack TELemetry) sites at Stringer Creek (1,996 m) and Onion Park (2,258 m), managed by the Natural Resources Conservation Service (NRCS), with enhanced sensors recording precipitation, air temperature, snow water equivalents, snow depth, relative humidity, solar radiation, wind speed and direction, among others. Additionally, there are eight stream gauging stations (seven flumes and one weir), managed by the U.S. Forest Service's Rocky Mountain Research Station (RMRS), measuring continuous stream discharge, using stream level recorders, with historic records dating to 1996. Several new devices have been installed recently at TCEF, including twelve transects of nested wells and piezometers (140 total) across the stream channel and two eddy flux covariance towers featuring full energy budget instrumentation [*Jencso et al.*, 2008, In Review].

Model Data

The area of focus for this study is the Stringer Creek watershed (555 hectares). Specific data for the Stringer Creek watershed that were assembled for this study include meteorological data from the Onion Park SNOTEL (Figure 4.2a), streamflow data from the gauging station at the outlet of Stringer Creek (Figure 4.2b) and topographic data in the form of a digital elevation model (10 m grid). Time series data were selected for the period of study spanning from March 2002 to September 2007. The Onion Park SNOTEL

station was selected over the Stringer Creek SNOTEL station due to the similarity of its elevation to the mean elevation of the Stringer Creek watershed.

Based on the information and data available at the Onion Park SNOTEL site, data pre-processing was required. The National Water and Climate Center (NWCC) of the NRCS is charged with data collection and quality control for all SNOTEL stations across the U.S., receiving data from the Onion Park SNOTEL location on 3-hour time intervals; however, the only data that is subject to quality control and assurance is that occurring at the midnight timestamp. This data processing procedure poses a problem in the selection

Figure 4.2: (a) Onion Park SNOTEL site and (b) Stringer Creek flume.

of an appropriate time interval for modeling activities due to the belief that the system response (change in streamflow due to system gains/losses) occurs on a sub-daily time scale.

It was found that there were significant inconsistencies and time series gaps in the 3-hour data samples (many instruments had no data for times other than midnight for

months of a given water year). In an effort to balance the available quality data (24-hour interval) with system response characteristics, a 12-hour time interval was selected to reduce the amount of error introduced to the data via processing.

Upon establishing a 12-hour modeling time step, the data were collected for the sensors of interest (see Table 4.1). Problems were perceived to be inherent to the two different types of precipitation gauges for which data was available. The difficulties in obtaining "clean" precipitation data were largely due to inconsistent 3-hourly data in the accumulation gauge (e.g., loss of accumulated precipitation over a single time step) and the inability of the tipping bucket rain gauge to properly measure frozen precipitation [*Groisman et al.*, 1999].

A temperature-based decision algorithm was then implemented, such that under sub-freezing temperatures the accumulation gauge would be used for precipitation, while the mean tipping bucket response would be used otherwise. It was expected that this would minimize potential gauge biases, such as undercatch due to snow drift. The use of a 12-hour time step largely solved the issue of "negative precipitation" occurring in the accumulation gauge and through the application of the temperature-based decision algorithm frozen precipitation was measured from the wind-shielded accumulation gauge.

The SNOTEL snow pillow was to be used as a check on precipitation gauge data and the model output for model and data consistency. Because evapotranspiration is a variable of interest and is not measured directly by any of the available sensors over the time period used in the model, it was calculated using the available meteorological data (including maximum daily air temperature, minimum daily air temperature, 12-hour

mean air temperature, relative humidity and wind speed) following the guidelines of the FAO (Food and Agriculture Organization of the United Nations) Irrigation and Drainage Paper No. 56 [*Allen et al.*, 1998]. Although evapotranspiration data can be obtained from the eddy flux covariance towers, it was not utilized in this study due to a desire to generate and test modeling hypotheses based solely on widely available data. Net radiation, another unmeasured variable of interest, was obtained as an intermediate calculation in the estimation of potential evaporation. Streamflow data, collected every 15 minutes from the Stringer Creek flume, was aggregated into a 12-hour mean value.

Table 4.1: Onion Park SNOTEL Meteorological Sensors of Interest

Sensor	**Description**
PREC-1	a 10” orifice accumulated precipitation gauge with windshield; [in]
PREC-2,-3,-4	a series of three (open, partial, closed canopy) incremental precipitation tipping buckets; [in]
SWE	a pressure transducer snow pillow, works in conjunction with snow depth sensor; [in]
TOBS	air temperature observed at sample time; [^{o}C]
TMIN	minimum air temperature observed for 24 hour period; [^{o}C]
TMAX	maximum air temperature observed for 24 hour period; [^{o}C]
WIND	wind speed observed at sample time; [mph]
RH	relative humidity observed at sample time; [%]
SOLRAD	solar radiation observed at sample time; [W/m^{2}]

Conceptual Modeling Suite Development

A suite of conceptual model structures was developed for this study. Model structures were selected ranging in structural representation of the physical system and in structural complexity (i.e., the number of model dimensions), acknowledging the

complexity of hydrologic systems and the ambiguity associated with the understanding of the driving processes. In consideration of the Stringer Creek watershed, several main processes could clearly be seen as dominant, necessitating their inclusion in the conceptualization of the overall model structure. Among these processes were snowmelt generation, soil moisture accounting and runoff routing. Additional consideration has been devoted to areal distribution of precipitation inputs to the watershed itself and whether or not this merits direct conceptualization in the model structure.

Models were constructed within a modular framework that consisted of a base soil water accounting structure, a snowmelt accounting procedure (that described input to the watershed) and a semi-distribution component. Specified under such a framework, thirty unique modular structures were developed comprising all possible combinations of soil water accounting, snowmelt accounting method, water routing and semi-distribution technique (describing variability in input to the watershed system). The details associated with each module will be discussed in the following sections.

Soil Moisture Accounting and Runoff Routing Base Structures

Representing the water accounting processes of the physical system, three competing model base structures for soil moisture accounting and runoff routing were selected. The three approaches chosen for this study were a simple bucket model [BUC, e.g., *Manabe*, 1969], the Probability Distributed Model [PDM, *Moore*, 1985; 2007] and the TOPMODEL [TOP, *Beven*, 1997; *Beven and Kirkby*, 1979]. Given that each base structure is a conceptualization of reality, developed under a top-down modeling

approach, the models attempt to characterize the dominant processes while lumping all others into several effective parameters.

The value of the three components selected for soil moisture accounting and runoff routing in this study is in the unique similarities and differences among each in their first-order representations of the Stringer Creek watershed. The similarity of the bucket and Probability Distributed models allows for a better understanding of the worth associated with the soil moisture accounting procedure used in each (i.e., whether the variable soil water content storage of the PDM provides a benefit). The dissimilarity of TOPMODEL to the bucket model or the PDM allows for the opportunity to better understand the merit of the variable contributing area approach implemented in TOPMODEL. The properties of each base module will be discussed in detail in the following subsections.

Bucket Model: The bucket model is a simple, conceptual, five parameter model (Figure 4.3). Water enters the model as precipitation (either as rain or snowmelt) and

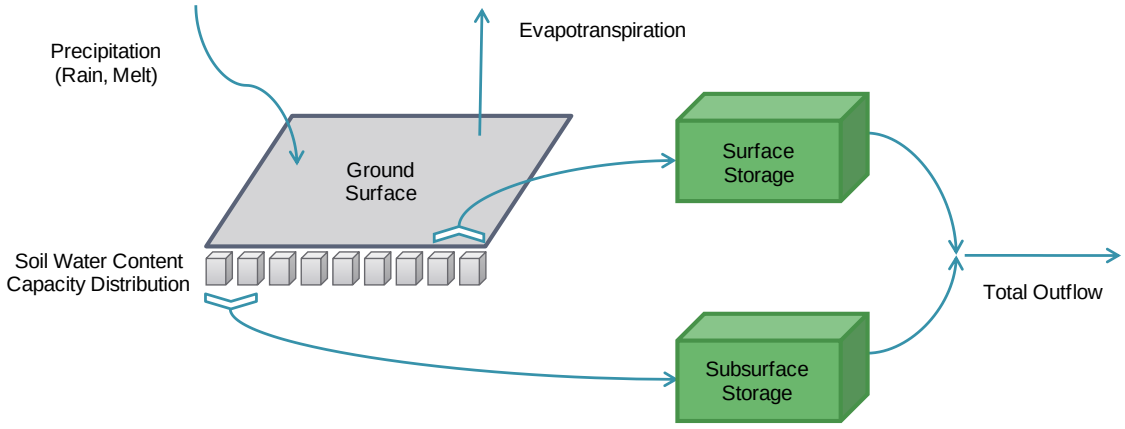


Figure 4.3: A schematic representation of the bucket model soil moisture accounting and runoff routing component.

leaves as either evapotranspiration losses or as streamflow at the outlet. Precipitation that lands on the ground surface is routed to the surface and subsurface storages based on the amount of water occupying the soil water content storages (distributed uniformly); water leaks out of the bottom of the soil water content storages and into the subsurface storage, while water that exceeds the capacity of the soil water content storages becomes overland flow that eventually enters the surface storage. The combined outflow from each of the storages becomes the total outflow at the watershed outlet. The five effective parameters associated with the bucket model component are given in Table 4.2.

Table 4.2: Effective Parameters of the Bucket Model Soil Moisture Accounting and Runoff Routing Component

Parameter	**Description**
CMAX	the maximum soil water content capacity; [mm]
KB	the rate of drainage into subsurface storage; [12-hr^{-1}]
TRES1	the fraction of subsurface storage released to outflow; [-]
TRES2	the fraction of surface storage released to outflow; [-]
CF	the soil storage threshold for subsurface inflow; [mm]

Mathematically, the bucket model partitions the contribution of precipitation to the watershed into a surface storage given by

$$u1_i = \max\begin{cases} 0 \\ P_i - CMAX + c_{i-1} \end{cases} \tag{4.1}$$

where the "effective" precipitation (i.e., precipitation that exceeds the soil water content capacity), $u1$, is determined based on the maximum soil water content capacity ($CMAX$), the soil water content depth at the previous time step (c_{i-1}) and the precipitation at the

current time step (P_i). Water not partitioned to the surface storage is routed to the subsurface storage, calculated as

$$Pnet_i = P_i - u1_i \tag{4.2}$$

where $Pnet_i$ is the water that infiltrates into the soil. Water drainage through the soil is governed by

$$\gamma_i = KB \cdot c_i \tag{4.3}$$

where γ_i represents the amount of water that enters the subsurface storage provided c_i is greater than CF, the soil storage threshold for subsurface inflow. Water in each of the storages is released to outflow at rates governed by the model parameters $TRES1$ (for subsurface storage) and $TRES2$ (for surface storage).

The bucket model conceptualizes the watershed as a uniform aggregation of the variability of the soil water content capacity; that is, the ability of the soil to hold infiltrated water is modeled as identical at every point within the watershed. Although the uniformity generalization is a great simplification to the physical system, it is in line with a top-down philosophy of model building where the simplest of model representations must be considered.

Probability Distributed Model: As a conceptual, six parameter model, the Probability Distributed Model (Figure 4.4) allows for explicit representation of the variability in the watershed storage capacity, unlike the bucket model. As with the bucket model, water enters the model as precipitation (as either rain or snowmelt) and leaves as either evapotranspiration losses or as streamflow at the outlet. Precipitation that lands on the ground surface is routed to the surface and subsurface storages based on the amount

of water occupying the soil water content storages (following a Pareto distribution); water leaks out the bottom of the soil water content storages and into the subsurface storage, while the water that exceeds the capacity of the soil water content storages becomes surface flow that eventually enters the surface storage. The combined outflow from each

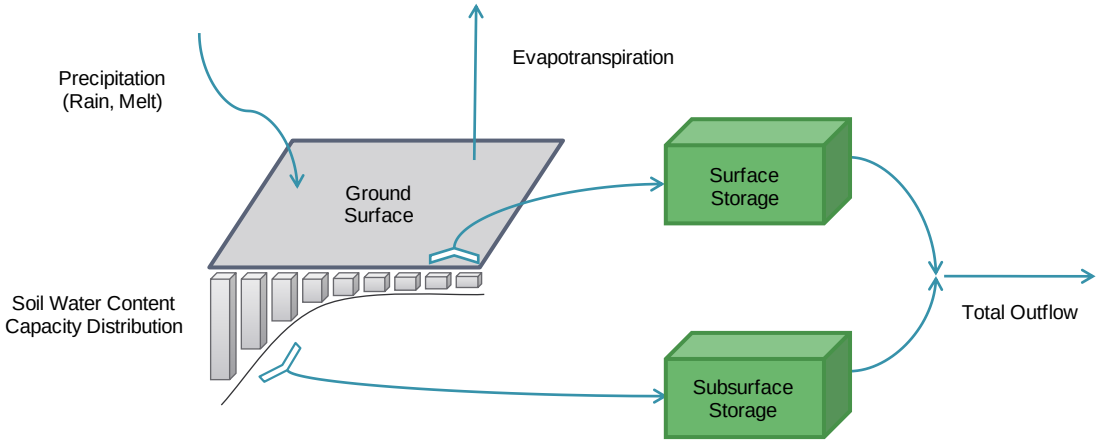


Figure 4.4: A schematic representation of the Probability Distributed Model soil moisture accounting and runoff routing component.

of the hypothetical storages becomes the total outflow at the watershed outlet. The six effective parameters associated with the Probability Distributed Model are given in Table 4.3.

Table 4.3: Effective Parameters of the Probability Distributed Model Soil Moisture Accounting and Runoff Routing Component

Parameter	**Description**
CMAX	the maximum soil water content capacity; [mm]
B	the spatial variability within the watershed; [-]
KB	the rate of drainage into subsurface storage; [12-hr^{-1}]
TRES1	the fraction of subsurface storage released to outflow; [-]
TRES2	the fraction of surface storage released to outflow; [-]
CF	the soil storage threshold for subsurface inflow; [mm]

As with the bucket model, the Probability Distributed Model calculates the partitioning of water to the surface and subsurface storages. Water is routed to the surface storage as given in equation (4.1), above, and as

$$u2_i = \max\begin{cases} 0 \\ Pnet_i - (s_i - s_{i-1}) \end{cases} \tag{4.4}$$

where s_i and s_{i-1} are the soil water content storage volumes at the current and previous time steps, respectively. Water is then routed to the surface storage as the sum of $u1$ and $u2$.

The remaining precipitation is leaked into the soil following equation (4.2). The difference between the bucket and Probability Distributed models is the inclusion of a model parameter (B) governing the spatial variability of soil water content capacity within the watershed.

To accommodate for the variable soil capacity, the storage depths (c_i) and storage volumes must be calculated at each time step as

$$s_i = \max\begin{cases} 0 \\ s_{max} \cdot \left[1 - \left(1 - \frac{c_i}{CMAX}\right)^{B+1}\right] \end{cases} \tag{4.5}$$

where s_i is the storage volume at the current time step, s_{max} is the maximum soil water content volume and other variables are as previously defined. The water released from the soil into the subsurface storage is then computed from

$$\gamma_i = KB \cdot s_i \tag{4.6}$$

The condition that c_i must exceed CF before inflow to the subsurface storage occurs applies, as before. Once more, water in each of the storages is released to outflow at rates

governed by the model parameters $TRES1$ (for subsurface storage) and $TRES2$ (for surface storage).

The PDM (which is the basis for the bucket model used in this study) incorporates spatial variability into the soil water content capacity component of the structure through use of a non-uniform, Pareto distribution. The shape of this distribution is regulated by the effective parameter B; when this parameter is equal to zero the bucket model and the Probability Distributed Model are identical.

TOPMODEL: The topography based hydrologic model, TOPMODEL, was developed originally not to be utilized in a traditional fashion, but rather as a loose framework of principles to be applied when seen fit. The formulation of TOPMODEL used for this study is expressed as a five parameter model (see Figure 4.5). Water enters

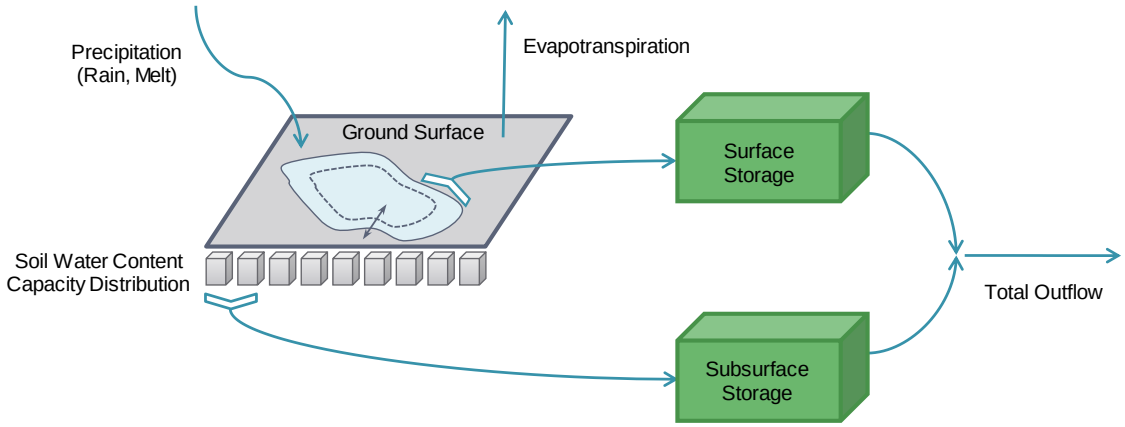


Figure 4.5: A schematic representation of the TOPMODEL soil moisture accounting and runoff routing component.

the model as precipitation (either as rain or snowmelt) and leaves as either evapotranspiration losses or as streamflow at the outlet. The primary control on partitioning surface and subsurface water is the topographic index [*Kirkby*, 1975]. The

topographic index drives the mathematical machinery of TOPMODEL, describing the tendency of a given cell within the watershed to become fully saturated and begin to produce saturation-excess overland flow; computed as

$$TI = \ln\left(\frac{\alpha}{\tan\beta}\right) \tag{4.7}$$

where TI is the topographic index for a given cell, α is the upslope contributing area per unit contour length and $\tan\beta$ is the local slope.

In this way the soil water content capacity of the watershed is spatially distributed via calculated values from each gridded cell of the watershed, rather than applying a statistical distribution to characterize this model component (as is done in the bucket and Probability Distributed models). By determining the potential wetness associated with a point within the catchment, the TOPMODEL base structure specifically accounts for spatial variability.

Table 4.4: Effective Parameters of the TOPMODEL Soil Moisture Accounting and Runoff Routing Component

Parameter	**Description**
M	the exponential storage parameter; [mm]
TMAX	the maximum watershed value of transmissivity; [$mm^2 \cdot 12\text{-}hr^{-1}$]
TDELAY	the time delay per unit deficit in the unsaturated zone; [12-hr]
SBAR0	the initial average watershed deficit; [mm]
SRMAX	the available water capacity of the root zone; [mm]

Runoff generated by way of saturation-excess overland flow is routed to a surface storage component, while infiltrated water slowly leaks into a subsurface storage component. The sum of the outflow from each of the conceptual storages is the total

outflow from the watershed. The five effective parameters associated with the TOPMODEL are given in Table 4.4.

Computationally, TOPMODEL considers the representation of multiple storages in its calculation procedure, representing the local storage deficit as

$$SD_{i,j} = \bar{S}_i + M \cdot \left(\overline{TI} - TI_j\right) \tag{4.8}$$

where i indexes the time step, j indexes the topographic index class, $\bar{S}_i$ is the catchment mean saturation deficit, M is the exponential storage parameter, $\overline{TI}$ is the catchment mean topographic index and TI_j is the topographic index for the j^{th} topographic class. The root zone deficit is described by

$$SRZ_{i,j} = SRZ_{i-1,j} - P_i \tag{4.9}$$

where $SRZ_{i-1,j}$ is the root zone deficit at the previous time step and P_i is the precipitation at the current time step. Finally, the saturation excess is calculated from

$$SatEx_{i,j} = SUZ_{i,j} - SD_{i,j} \tag{4.10}$$

where $SUZ_{i,j}$ is the unsaturated zone storage.

The partitioning of water between the surface and subsurface storages in TOPMODEL is somewhat more involved than in the bucket or Probability Distributed models. Drainage from the unsaturated zone into the saturated zone is calculated as

$$QUZ_{i,j} = \frac{SUZ_{i,j}}{\left(SD_{i,j} \cdot TDELAY\right)} \cdot ACF \tag{4.11}$$

where $TDELAY$ is the time delay per unit deficit in the unsaturated zone and ACF is the cumulative fractional area for each topographic index class. While drainage from the saturated zone is regulated by

$$Qsub_i = \exp(TMAX - \overline{TI}) \cdot \exp\left(-\frac{\bar{S}_i}{M}\right) \tag{4.12}$$

where $TMAX$ is the maximum watershed value of transmissivity and $\bar{S}_i$ is the mean catchment deficit. The release of water from the surface storage is dependent on an assessment of the wetness state of the current and previous topographic index classes expressed via

$$Qof_i = \begin{cases} Qof_i + TI_j^* \cdot \dfrac{SatEx_{i,j-1} + SatEx_{i,j}}{2} & \text{if } SatEx_{i,j} > 0 \\ Qof_i + ACF \cdot \dfrac{SatEx_{i,j-1}}{2} & \text{if } SatEx_{i,j-1} > 0 \end{cases} \tag{4.13}$$

where TI_j^* is the portion of the watershed contained by the current topographic index class and $SatEx$ is the saturation excess for the current time step and the j^{th} and $j-1^{th}$ topographic index classes. The total simulated outflow from the system is simply the sum of $Qsub_i$ and Qof_i.

In its expression of soil moisture accounting, TOPMODEL represents a vastly different conceptualization based on the wetness of any given topographic index class within the watershed. The direct estimation of the tendency of a given cell to produce saturation-excess overland flow by the calculation of topographic indices provides a clear link to the physical system.

Soil Moisture Accounting Data Requirements: The data requirements for each of the three base structures are quite similar. Each base structure requires the input of precipitation (as rainfall or snowmelt) and evapotranspiration data to produce simulated

stream discharge. The TOPMODEL has the additional data requirement of a digital elevation model of the study site, necessary for the calculation of the topographic index.

Snowmelt Accounting

Given the snowfall-dominated precipitation pattern of the Stringer Creek watershed, the model base structures must be augmented with a snowmelt accounting procedure to appropriately characterize inputs to the physical system. Continuing the use of downward model development strategies, three related, conceptual snowmelt accounting techniques were selected for use within the modeling suite; the temperature index, the radiation index and the temperature and radiation index approaches.

Although significant research has gone into the physical understanding of snowpack metamorphosis and the development of full energy balance methods to simulate snowmelt [e.g., *Anderson*, 1968; *Barry et al.*, 1990; *Bloschl et al.*, 1991; *Marks and Dozier*, 1992], the constraints posed by such methods are extensive [e.g., *Rango and Martinec*, 1995]. The use of conceptual, index-based approaches to snowmelt accounting is well established in the literature [e.g., *Bloschl and Kirnbauer*, 1991; *Braithwaite*, 1995; *Daly et al.*, 2000; *Hock*, 1999], with *Ohmura* [2001] providing a physical basis for such approaches. Conceptual, temperature index approaches have been found to often perform comparably to full energy balance methods on a catchment scale [*WMO*, 1986]. Discussion of each snowmelt accounting method will be covered in the following subsections.

Temperature Index Approach: The temperature index approach to snowmelt accounting is a widely-used conceptual model [see *Hock,* 2003 for a review] that is based on the empirical relationship between air temperature and snowmelt

$$\text{Melt}_i = DDF \cdot \max\begin{Bmatrix} 0 \\ Tbar_i - TF \end{Bmatrix} \tag{4.14}$$

where $Tbar$ is the observed temperature averaged over a 36-hour period (3 time steps) and the effective model parameters are the degree-day factor (DDF, [mm·^{o}C^{-1}]) and the threshold temperature for snowmelt (TF, [^{o}C]).

Requiring only air temperature data to implement this method, the temperature index approach is an accessible tool that has been proven to be effective in modeling snowmelt, seeing use in many common operational precipitation-runoff models [see *Lindstrom et al.,* 1997; *Martinec and Rango,* 1986] and even some physically-based hydrologic models [*Boggild et al.,* 1999].

The foundation of the temperature index approach to snowmelt accounting is strongly based on the physical processes driving snowpack evolution [*Ohmura,* 2001]. *Braithwaite and Oelsen* [1990] found that much of the temperature index approach's success originates from the high correlation of air temperature to several components of the energy balance.

Radiation Index Approach: The radiation index approach to snowmelt accounting is functionally similar to the temperature index approach, substituting the empirical relationship between air temperature and snowmelt with the relationship between net radiation and snowmelt

$$\text{Melt}_i = NRF \cdot \max\begin{Bmatrix} 0 \\ Rn_i \end{Bmatrix} \tag{4.15}$$

where Rn is the observed net radiation and the only effective model parameter is NRF, the net radiation factor.

Such a conceptualization takes aim at considering another variable important to the overall physics of the snowmelt process. The representation of this relationship is supported by the research of *Ohmura* [2001] who found that net radiation is most often the primary source of heat during the melt period. Despite its simplicity, the universal applicability of this approach is limited by the availability of net radiation data; however, many newer meteorological stations now collect radiation data making this condition less restrictive than it was in the recent past. The utilization of net radiation in snowmelt accounting is not uncommon [e.g., *Cazorzi and Fontana*, 1996], however the application of net radiation as the sole descriptor of snowmelt is much less common.

Temperature and Radiation Index Approach: The temperature and radiation index snowmelt accounting method is based on a conceptualization that considers both air temperature and net radiation to be of significant importance to the modeling of snowmelt dynamics. This approach is a simple combination of the temperature index and the radiation index methods discussed previously

$$\text{Melt}_i = DDF \cdot \max\begin{Bmatrix} 0 \\ Tbar_i - TF \end{Bmatrix} + NRF \cdot \max\begin{Bmatrix} 0 \\ Rn_i \end{Bmatrix} \tag{4.16}$$

Much research has shown that snowmelt estimates are improved by combining temperature and radiation approaches [e.g., *Brubaker et al.*, 1996; *Kustas et al.*, 1994],

motivating the incorporation of such a method in the suite of snowmelt accounting approaches for the current study. Due to the presence of radiation fluxes in the energy balance equations controlling snowmelt processes, the incorporation of net radiation also serves to add a further level of physical realism to the model [*Ohmura*, 2001].

At this point, the next logical step forward in snowmelt accounting is to move down one of two paths: (1) move to full energy balance techniques for snowmelt prediction or (2) implement a spatial distribution methodology on top of the conceptual snowmelt accounting procedures outlined above. However, the desire to better characterize primary drivers of the watershed (i.e., the processes dominating watershed response) relies on maintaining parsimonious model structures. To that end, the upcoming subsections will detail simple methods of watershed semi-distribution aimed at improving the realism of the precipitation processes in a way that upholds the principle of process parsimony.

Precipitation Semi-Distribution

The implementation of a spatial distribution scheme associated with snowmelt modeling has the ability to improve the physical representation of system processes [*Garen and Marks,* 2005], as well as the predictive result of the model simulation. In an effort to account for the spatial variability (potentially) present in the precipitation inputs to the system, three different conceptual methods have been assimilated into the modeling suite with a fourth option of assuming the watershed to be uniform with respect to precipitation (i.e., applying no precipitation semi-distribution). Each semi-distribution method relies on the availability of a digital elevation model for the study watershed.

For each method, the semi-distributed results are re-aggregated using a weighted average approach after the snowmelt accounting procedures are called, but before the precipitation is called by the model soil moisture accounting and runoff routing component; this results in a single simulated streamflow at the watershed outlet. The procedure for each semi-distribution method is outlined in the following subsections.

Semi-Distribution by Aspect: Watershed aspect is an important control on snowmelt rates due to its relationship with to the energy balance of a snowpack, manifested as solar radiation [*Cazorzi and Fontana*, 1996]. Given the desire to better characterize the physical processes driving the hydrologic system and the correlation between aspect and radiation, disaggregation of a watershed into hydrologic response units (HRUs) based on aspect is not unfounded [e.g., *Braun et al.*, 1994; *Charbonneau et al.*, 1981].

Semi-distribution by aspect (Figure 4.6) considers only the calculation of snowmelt input to the system by breaking up the watershed into three aspect classes; North-facing (315˚-45˚), South-facing (135˚-225˚) and East-West-facing (45˚-135˚ & 225˚-315˚). Each aspect class has been assigned an *aspect factor* (-1 for North-facing, 1 for South-facing and 0 for East-West-facing). The formulation of aspect discretization used for this study is based on the assumption that South-facing aspects produce snowmelt first, followed by East-West-facing and North-facing, an assumption supported by *Cazorzi & Fontana* [1996]. The aspect factor is then used as a multiplier in the calculation of snowmelt that assesses the aspect-specific snowmelt (for the North and

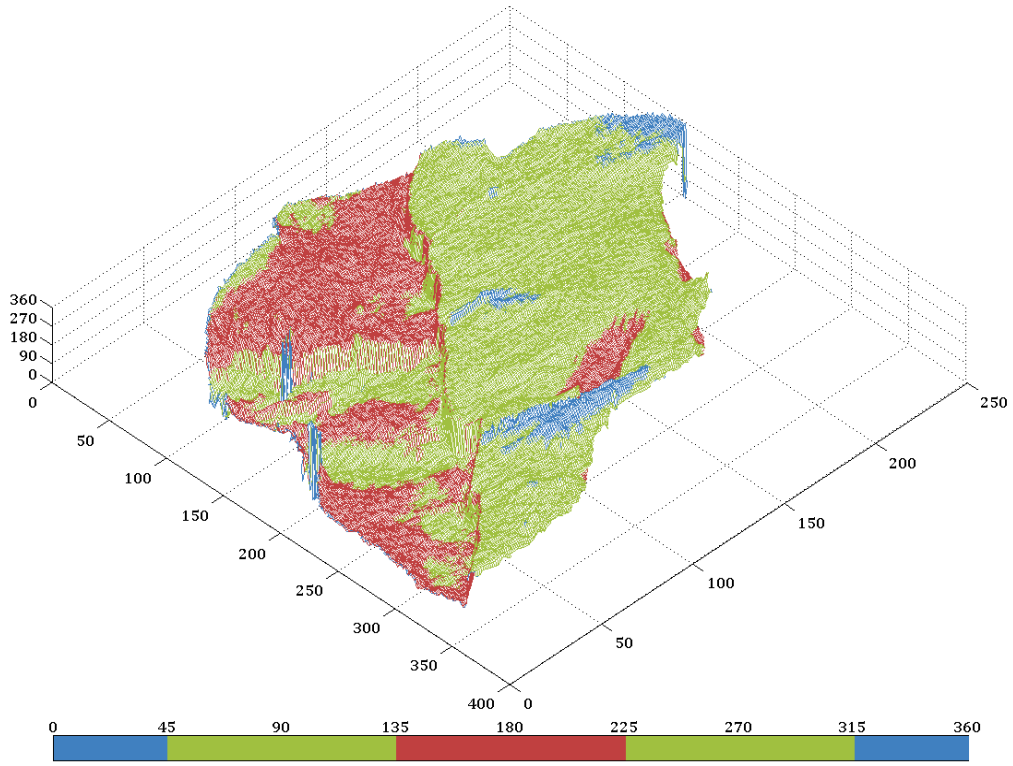


Figure 4.6: Semi-distribution of precipitation by aspect.

South aspect classes) produced above the base snowmelt level (assumed to be that of the East-West aspect class). The overall snowmelt for the watershed is calculated as a weighted mean response at the end of the snowmelt accounting routine.

$$\text{Melt}_{ij} = \left(NRF_0 + NRF_1 \cdot ASPF_j\right) \cdot \max\begin{Bmatrix} 0 \\ Rn_i \end{Bmatrix} \tag{4.17}$$

$$\text{Melt}_{ij} = DDF \cdot \max\begin{Bmatrix} 0 \\ Tbar_i - TF \end{Bmatrix} + \left(NRF_0 + NRF_1 \cdot ASPF_j\right) \cdot \max\begin{Bmatrix} 0 \\ Rn_i \end{Bmatrix} \tag{4.18}$$

where i is the index on time, j is the index on aspect class, $ASPF$ is the aspect factor, NRF_0 (model parameter) is the base net radiation factor (equivalent to NRF for non-aspect semi-distribution) and NRF_1 (model parameter) is the aspect dependent net

radiation factor; (4.17) is the equation used for radiation index snowmelt accounting and (4.18) is the equation used for temperature and radiation index snowmelt accounting.

The weight of each aspect class is determined by the ratio of the number of cells in a given class to the total number of cells. Application of the aspect semi-distribution method excludes the use of the temperature index snowmelt accounting approach because the aspect factor is radiation dependent in this formulation and the temperature index approach (4.14) is independent of radiation, the driving force behind the motivation for this approach.

Semi-Distribution by Elevation: A known relationship exists between precipitation (and temperature) and elevation [see *Daly et al.*, 1994 for a comprehensive review]. Invoking this (in a similar fashion to the previously described approach to dividing a watershed according to aspect), a watershed can be simply divided into HRUs based on evenly spaced elevation bands to help improve the link between process understanding and model conceptualization. Such a practice has gained growing support in snow-dominated hydrologic systems [e.g., *Fontaine et al.*, 2002; *Hartman et al.*, 1999].

In this application, semi-distribution by elevation considers the calculation of all precipitation (and temperature) input to the system by breaking up the watershed into five uniform width elevation classes (Figure 4.7). The precipitation input is calculated based on the use of lapse rates applied to the data, utilizing a reference point (Onion Park SNOTEL) located within the elevation range of the watershed itself. Because snowmelt

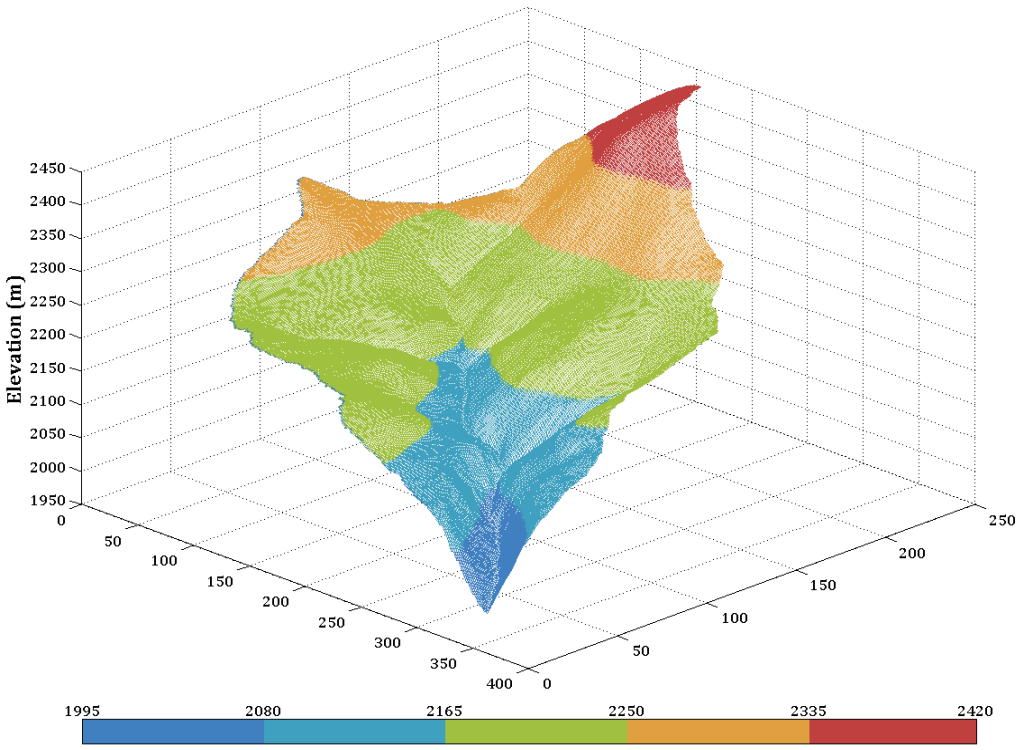


Figure 4.7: Semi-distribution of precipitation by elevation.

accounting relies on air temperature, both precipitation (and snow water equivalents) and temperature are computed for each elevation class

$$\text{Precip}_k = \text{Precip} \cdot \frac{1 + PLASPE \cdot (\text{ElClass}_k - \text{ElClass}_{REF})}{1 - PLAPSE \cdot (\text{ElClass}_k - \text{ElClass}_{REF})} \tag{4.19}$$

$$\text{SWE}_k = \text{SWE} \cdot \frac{1 + PLAPSE \cdot (\text{ElClass}_k - \text{ElClass}_{REF})}{1 - PLAPSE \cdot (\text{ElClass}_k - \text{ElClass}_{REF})} \tag{4.20}$$

$$\text{Temp}_k = \text{Temp} - TLAPSE \cdot (\text{ElClass}_k - \text{ElClass}_{REF}) \tag{4.21}$$

where k indexes the five evenly-spaced elevation classes, $PLAPSE$ (model parameter, [m^{-1}]) is the lapse rate for precipitation (and snow water equivalents), $TLAPSE$ (model parameter, [$^oC \cdot m^{-1}$]) is the lapse rate for air temperature, ElClass_k is the elevation of the k^{th} elevation class, ElClass_{REF} is the elevation of the reference point, Precip is the

observed precipitation, SWE is the observed snow water equivalents and Temp is the observed air temperature.

The total precipitation input to the soil moisture accounting and runoff routing component is then a mean weighted input of the five individual classes. The weight of each elevation class is determined by the ratio of the number of cells in a given class to the total number of cells in the watershed.

Semi-Distribution by Aspect and Elevation: The final approach to watershed semi-distribution included in the modeling framework is based on a combination of the aspect semi-distribution and the elevation semi-distribution approaches. Such an approach has been used previously to enhance the physical reality of the snowmelt accounting model component [e.g., *Braun et al.*, 1994]. Calculation of semi-distributed input data proceeds with the discretization of the watershed into three aspect classes for each of the five elevation classes (for a total of 15 classes, Figure 4.8), as described in the previous subsections.

Equations (4.19), (4.20) and (4.21) are applied to the input data, which is then passed to the snowmelt accounting routine. As with the aspect semi-distribution approach, the temperature index method for snowmelt accounting is precluded by use of this semi-distribution scheme. This exclusion is a result of the assumption in the aspect semi-distribution conceptualization that snowmelt varies with aspect as a result of differing net radiation influence on different aspects; because the temperature index snowmelt accounting approach is independent of net radiation it must also be independent of aspect, based on this underlying hypothesis. Again, snowmelt is

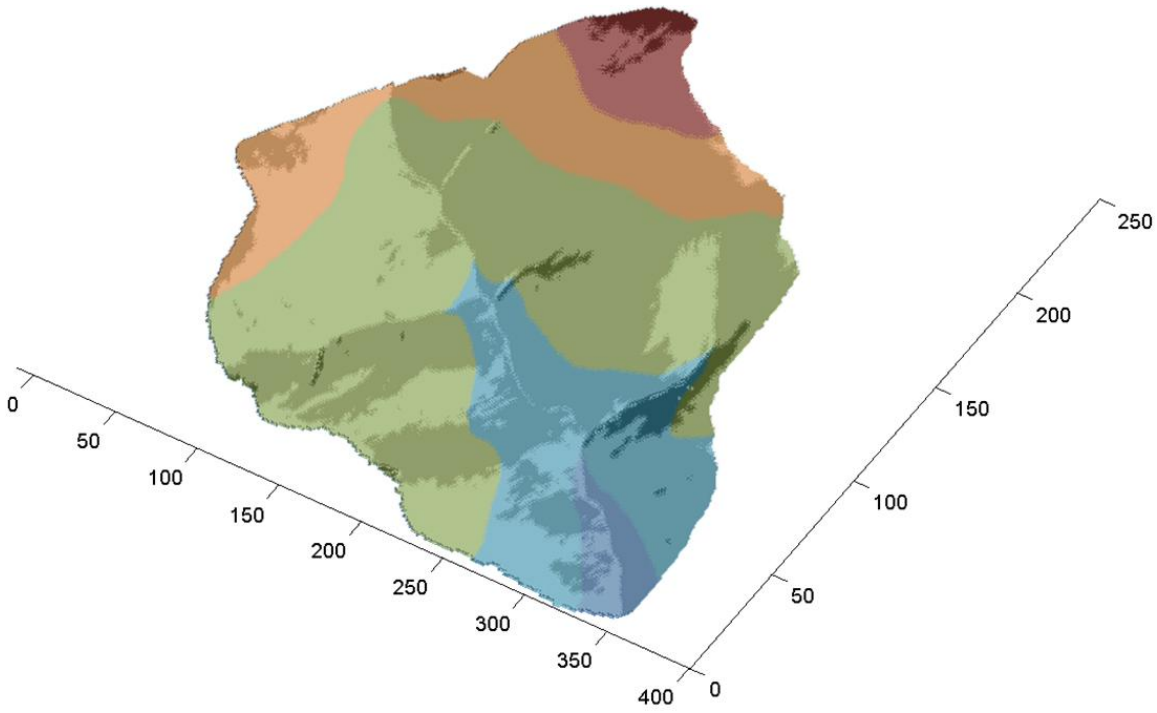


Figure 4.8: Semi-distribution of precipitation by aspect and elevation.

calculated using equations (4.17) or (4.18) depending on whether the radiation index or temperature and radiation index approach to snowmelt accounting is implemented.

Conclusion

The development of a framework for model selection, covering a range of possible representations of the physical system, is of vital importance to the principal task of improving understanding of primary watershed processes. The framework assembled and presented in this chapter makes use of well-established theories on model formation in application to the soil moisture accounting and runoff routing, snowmelt accounting and precipitation semi-distribution components. The modular format makes it possible to account for 30 independent model structures.

The following chapter will detail the results of model calibration and uncertainty analysis through an assessment of each of the thirty modular structures described previously. The material presented in Chapter 3 will serve as the basis for the selection of the most appropriate algorithm to use for assessment of the entire modeling framework, to be discussed in Chapter 5.

PREDICTIVE PERFORMANCE AND UNCERTAINTY ASSESSMENT OF A CONCEPTUAL MODELING FRAMEWORK APPLIED TO STRINGER CREEK WATERSHED

In the two previous chapters, a modular modeling framework and a review of three recently introduced Markov chain Monte Carlo algorithms necessary to the estimation of effective parameters and analysis of uncertainty were presented. The current chapter builds on these through the application of the Delayed Rejection Adaptive Metropolis algorithm to each model structure in the modeling framework. A detailed assessment of the performance of each modular structure applied to the study site is presented.

Introduction

The interrelation of model calibration, uncertainty analysis and assessment in hydrologic modeling necessitates careful consideration of a holistic approach to the precipitation-runoff problem. Many different methodologies exist and have been applied to the calibration of hydrologic models to observed data, such as manual approaches including visual goodness of fit tests [e.g., *Madsen et al.*, 2002] or fully automatic techniques including global optimization [e.g., *Duan et al.*, 1992], Monte Carlo simulation [e.g., *Beven and Binley*, 1992] or Markov chain Monte Carlo (MCMC) simulation [e.g., *Kuczera and Parent*, 1998]. Unfortunately, many methods lack the ability to objectively quantify the uncertainty associated with the calibrated parameter estimates [*Samanta et al.*, 2007].

The implementation of a Bayesian statistical approach offers the ability to perform both parameter estimation and uncertainty analysis, while also serving as an attractive framework within which multiple model structures can be compared [*Marshall et al.*, 2005]. Despite the obvious benefits of such an approach, Bayesian inference has seen limited use in hydrology until recently due to the complications associated with calculating the posterior distribution for highly complex, non-linear models. In order to calculate the analytically intractable posterior distribution, Markov chain Monte Carlo simulation must be implemented. As discussed in Chapter 3 of this thesis, there have been several recent advancements in MCMC techniques that have improved the efficiency and effectiveness of such methods in hydrology.

Building on the material presented in Chapter 3 and Chapter 4, this chapter offers an implementation of a joint traditional assessment framework with a Bayesian inferential approach to model calibration, uncertainty estimation and structural assessment. The present chapter is divided into the following sections: a brief review of the MCMC simulation applied in the study including an overview of the Delayed Rejection Adaptive Metropolis algorithm; a presentation of model calibration results for the modeling suite with a focus on the "best" conceptual model structure; a comparative assessment of each model structure on the basis of performance, uncertainty and realism; and a summary of the conclusions drawn from this investigation.

A Review of the Bayesian Methodology Applied in this Study

In hydrology, Bayesian methods usually rely on Markov chain Monte Carlo simulation for the characterization of a probability distribution for the model parameters and outputs, known as the posterior distribution. This reliance is due to the highly complex nature of many hydrologic models where the posterior distribution is analytically intractable. The distribution of model parameters is explored by MCMC algorithms via generation of a random process whose stationary distribution is the posterior distribution of the parameters [*Marshall et al.*, 2004]. While many MCMC algorithms have been developed in an effort to describe the posterior distribution, the Delayed Rejection Adaptive Metropolis [DRAM, *Haario et al.*, 2006] algorithm has been selected for calibration and uncertainty estimation for the conceptual modeling suite presented in Chapter 4.

Recalling the study performed in Chapter 3, the DRAM algorithm was established as an efficient means of characterizing complex parameter spaces in a comparison of recent improvements to MCMC simulation techniques. Reflecting the benefits of the adaptive calculation of the covariance and the delayed rejection of proposed parameter sets, the DRAM algorithm was shown to be less likely to converge to local optimum and more likely to better sample from the tails of the posterior distribution. Based on these results, the Delayed Rejection Adaptive Metropolis algorithm was selected as the MCMC method most suitable for assessment of the modeling suite.

Assessment Protocol for Modeling Suite Applied to the Stringer Creek Watershed

The assessment of model suitability can take many forms. While studies often focus primarily on how well a model's prediction matches the observed data, a more thorough approach considering at least performance, uncertainty, and realism is advocated by *Wagener* [2003]. By implementing this type of multi-faceted approach to evaluate models, a more comprehensive understanding of reasons behind model successes or failures can be extracted. To that end, each of the models contained in the modeling suite will be analyzed in the following subsections with respect to each aspect of the outlined assessment strategy.

Performance-Based Assessment

Evaluation of model performance is traditionally based on the results of model predictions relative to the observed data. In this study, data have been split into two sets, with one set being used to calibrate model parameters and the other being used for an "independent" assessment of model performance (often referred to as validation). The calibration results presented here serve primarily as a comparison point for considering the utility of one model across an array of multiple model structures (or a set of model parameters across multiple possible parameter sets). Whereas the validation results serve as a frame of reference for the ability of a model to perform in conditions separate to those under which it was calibrated, as well as allowing for the comparison across multiple structures. In addition to a performance assessment via these more conventional

techniques, examination of the relationship between model complexity (i.e., model dimension) and predictive performance has been undertaken.

Model Calibration: Model calibration was performed on each of the thirty modular structures of the modeling suite using the calibration data set collected from March 2002 through September 2005. An a priori decision must be made on the appropriate likelihood function to use in the DRAM algorithm. For this study, all models were assessed using a likelihood that assumes the errors are normally, independently and identically distributed

$$P(Q|\theta) = (2\pi\sigma^2)^{-n/2} \prod \exp\left\{-\frac{[Q_{obs} - Q_p]^2}{2\sigma^2}\right\} \quad (5.1)$$

where n is the number of data points, σ^2 is the variance, Q_{obs} is the observed streamflow and Q_p is the simulated streamflow for a given parameter set. The use of the likelihood function defined in (5.1) has been commonly used in the literature [for example, *Bates and Campbell*, 2001; *Samanta et al.*, 2007; *Thiemann et al.*, 2001; *Vrugt et al.*, 2003b] and is seen as justifiable in this case due to the few peak streamflows in the data, reducing the chance of non-constant variance of the errors.

Convergence of a model's parameters to a stationary probability distribution is of utmost importance in proper calibration of a model to the calibration data set. Criteria establishing convergence are common and varied in their approach [e.g., *Brooks and Gelman*, 1998; *Brooks and Roberts*, 1998]. For the purposes of this study, the approach proposed by *Gelman and Rubin* [1992] was selected. *Gelman and Rubin's* [1992] method seeks to compare the between and within chain (where a chain is the trace of the

parameter value versus the iteration) variance for each parameter through the calculation of the R-statistic. However, due to the sizeable number of models and substantial run-times associated with each, convergence was only directly tested for a representative subset of models within the suite. The necessary number of parameter samples required for convergence of these models was used to guide inference in other models. Those not directly checked with the R-statistic underwent visual assessment of convergence.

Results and Discussion of Model Performance Under Calibration: The resulting maximum log-likelihood for each of the thirty modular structures implemented using the likelihood function of (5.1) are given in Table 5.1 (column 3), where BUC refers to the bucket model soil moisture accounting and runoff routing component, PDM is the Probability Distributed Model soil moisture accounting and runoff routing component, TOP is the TOPMODEL soil moisture accounting and runoff routing component, sdA is aspect semi-distribution, sdE is elevation semi-distribution, sdAE is aspect and elevation semi-distribution, Rn is radiation index snowmelt accounting, T is temperature index snowmelt accounting and TRn is temperature and radiation index snowmelt accounting.

The best model structure based on the maximum calibrated log-likelihood value and grouped by soil moisture accounting and runoff routing components of Table 5.1 is given in bold. A direct ratio of likelihoods represents how much more likely the first model is to the second. For example, if we consider the BUCsdAE_Rn and BUCsdAE_TRn structures (log-likelihood values of -1279.50 and -1294.06, respectively) from Table 5.1, the ratio becomes $\exp[(-1279.50) - (-1294.06)]$. This indicates that the BUCsdAE_Rn structure is over 2 million times more likely than the BUCsdAE_TRn

structure. It is clear from the data presented in Table 5.1 that there is a significant affect imposed by the different model structural components of soil moisture accounting and runoff routing, snowmelt accounting and precipitation semi-distribution. Considered as a whole, the structures using the PDM component are generally better performing than

Table 5.1: Model Performance-Based Assessment Criteria for the Calibration Period[a]

Model Name	Model Dimension	Maximum log-Likelihood	Nash-Sutcliffe Efficiency	AIC	BIC
BUC_Rn	7	-1252.53	0.89943	2519.1	2529.0
BUC_T	7	-2224.00	0.78747	4462.0	4471.9
BUC_TRn	8	-1387.56	0.88841	2791.1	2802.4
BUCsdA_Rn	8	-1232.32	0.90100	2480.6	2492.0
BUCsdA_TRn	**9**	**-958.51**	**0.91981**	**1935.0**	**1947.8**
BUCsdE_Rn	9	-1247.12	0.89984	2512.2	2525.0
BUCsdE_T	9	-1893.06	0.83535	3804.1	3816.9
BUCsdE_TRn	10	-1160.85	0.90630	2341.7	2355.8
BUCsdAE_Rn	10	-1279.50	0.89731	2579.0	2593.1
BUCsdAE_TRn	11	-1294.06	0.89615	2610.1	2625.7
PDM_Rn	8	-1177.25	0.90511	2370.5	2381.8
PDM_T	8	-1747.60	0.85274	3511.2	3522.5
PDM_TRn	9	-960.94	0.91966	1939.9	1952.6
PDMsdA_Rn	9	-1152.01	0.90692	2322.0	2334.8
PDMsdA_TRn	10	-929.02	0.92160	1878.0	1892.2
PDMsdE_Rn	10	-1040.86	0.91455	2101.7	2115.9
PDMsdE_T	10	-1673.04	0.86096	3366.1	3380.2
PDMsdE_TRn	**11**	**-811.51**	**0.92839**	**1645.0**	**1660.6**
PDMsdAE_Rn	11	-1208.32	0.90279	2438.6	2454.2
PDMsdAE_TRn	12	-1182.83	0.90468	2389.7	2406.6
TOP_Rn	7	-2140.88	0.80065	4295.8	4305.7
TOP_T	7	-2267.48	0.78024	4549.0	4558.9
TOP_TRn	8	-1971.19	0.82507	3958.4	3969.7
TOPsdA_Rn	8	-1875.25	0.83753	3766.5	3777.8
TOPsdA_TRn	**9**	**-1112.65**	**0.90970**	**2243.3**	**2256.0**
TOPsdE_Rn	9	-2089.37	0.80840	4196.7	4209.5
TOPsdE_T	9	-2705.56	0.69208	5429.1	5441.8
TOPsdE_TRn	10	-2063.71	0.81216	4147.4	4161.6
TOPsdAE_Rn	10	-2089.60	0.80838	4199.2	4213.3
TOPsdAE_TRn	11	-1759.78	0.85136	3541.6	3557.1

[a]The values highlighted in bold indicate the best model structure for each soil moisture accounting and runoff routing component.

either the BUC or TOP components. Further detail on component-wise comparisons will be given in the following paragraphs. Similarly, the models were evaluated via the Nash-Sutcliffe Efficiency [NSE, *Nash and Sutcliffe*, 1970] measure (Table 5.1, column 4) based on the minimization of the residual sum of squares given by

$$NSE = 1 - \frac{\sum\left(SWE_i - SWE_{p_i}\right)^2}{\sum(SWE_i - \overline{SWE})^2} \tag{5.2}$$

where i indexes the data, SWE is the observed snow water equivalents, SWE_p is the predicted snow water equivalents and $\overline{SWE}$ is the mean of the observed snow water equivalents. The maximum value of the NSE is 1, occurring when the observed and predicted data are identical. As with the maximum log-likelihood, the NSE metric indicated the same models to be superior performers. This consistency is readily explained by the similarity of the likelihood function to the Nash-Sutcliffe Efficiency.

Although direct selection of the best model structure based on log-likelihood or NSE gives a broad assessment of different model predictive performances in the calibration period, the application of an objective approach to model selection that considers performance and complexity is often preferred. More complex models often provide an improvement in the calibration period due to a greater flexibility in structure afforded by additional parameters, yet more complex models often suffer from a decrease in identifiability of model parameters, leading to increased uncertainty [*Tripp and Niemann*, 2008]. Introducing criteria that penalize models for complexity, Table 5.1 (columns 5, 6) summarizes the model selection process based on Akaike's information criterion [AIC, *Akaike*, 1974] and Bayes' information criterion [BIC, *Schwarz*, 1978],

where the minimum values are preferred. Equations (5.3) and (5.4) represent the mathematical form of the AIC and BIC measures, respectively

$$AIC = -2 \cdot \ln[P(Q|\theta)] + 2k \tag{5.3}$$

$$BIC = -2 \cdot \ln\left[\left(P(Q|\theta)\right)\right] + k \cdot \ln(n) \tag{5.4}$$

where $\ln[P(Q|\theta)]$ is the maximum log-likelihood, k is the number of model parameters and n is the number of observations in the data set.

In this study, the AIC, BIC, NSE and maximum log-likelihood value give the same model ranking based on predictive performance. This result is based on the relatively substantial separation in likelihoods across the model structures. Because of this, the consideration that the AIC and BIC measures give to model complexity are largely overshadowed by its consideration of model performance.

The range of responses across model dimension for each of the thirty model structures was quite extensive. The maximum, mean and minimum results from the calibration stage, grouped by model dimension, are presented in Figure 5.1. The general upward trend of the mean response suggests that the greater the number of model parameters the more able the overall structure is in its representation of the observed data. While this is not true always, it is true on average for the data and models tested in this study.

Taking a closer look at the structural components of each of the models, Figure 5.2, Figure 5.3 and Figure 5.4 show the maximum calibrated log-likelihood versus model dimension, with each figure representing a different model group. Figure 5.2 shows the maximum log-likelihood versus model dimension for the models grouped according to

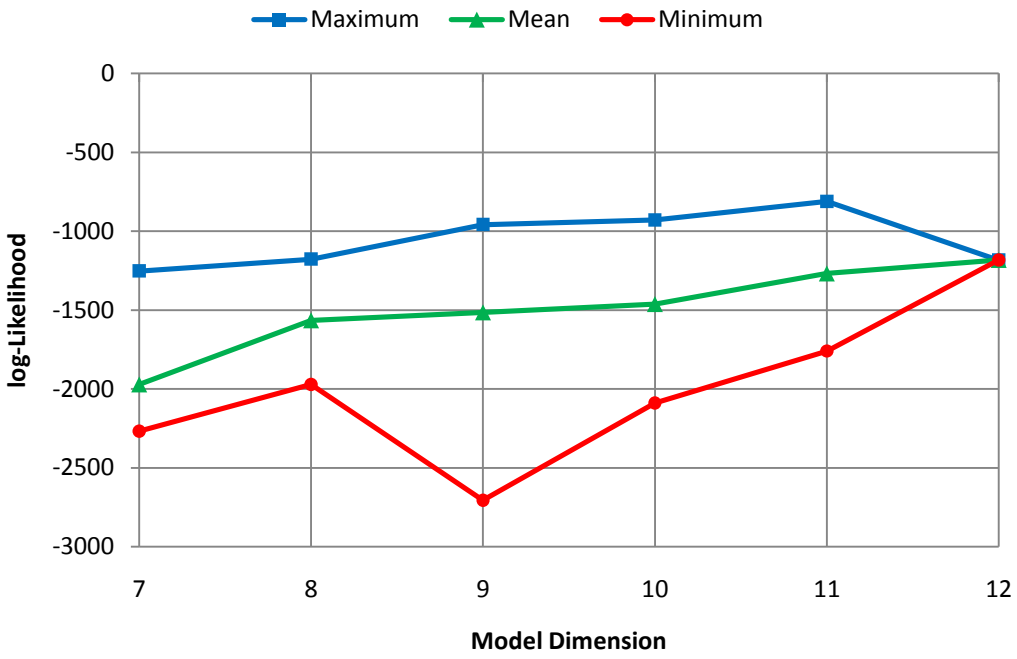


Figure 5.1: Calibration results, showing maximum, mean and minimum log-likelihood value for each model dimension.

their soil moisture accounting and runoff routing structure. Figure 5.3 and Figure 5.4 show the maximum log-likelihood results for models grouped by their snowmelt accounting and precipitation semi-distribution, respectively.

When clustered by the soil moisture accounting and runoff routing component (Figure 5.2), the TOPMODEL had the worst performance of the three components tested at every model dimension. The Probability Distributed Model was the best performer at each of its model dimensions. Note that at the nine parameter model dimension, the parameter B in the PDM model has a calibrated value close to zero, giving a performance approximately equivalent to the bucket model. The results of this classification suggest that the structural composition of the bucket and Probability Distributed models are better able to simulate the observed data than the TOPMODEL. The description of the

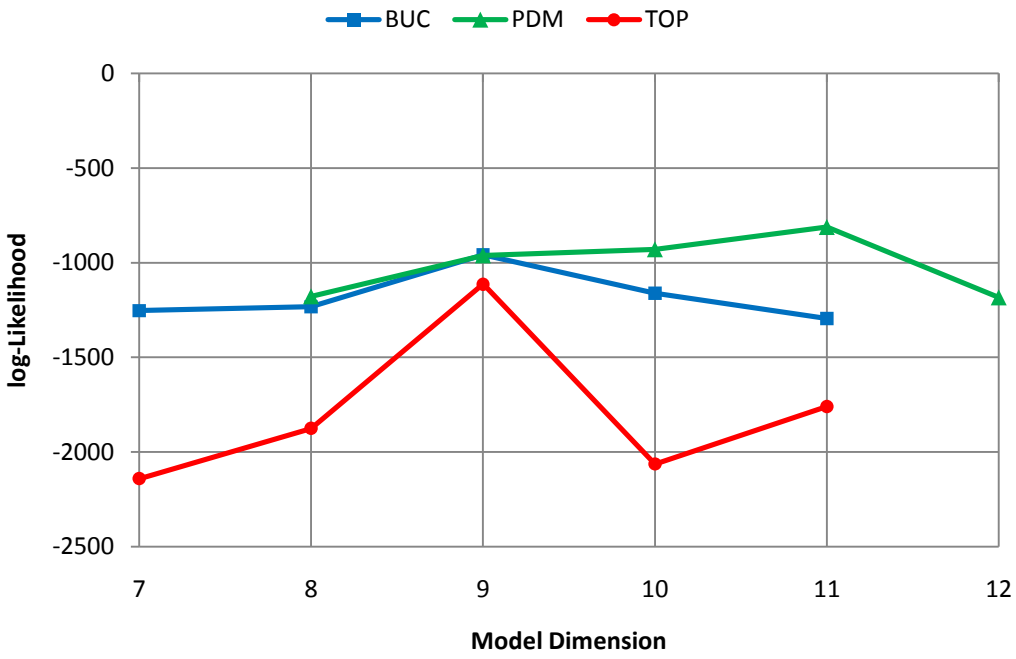


Figure 5.2: Calibration results grouped by the soil moisture accounting and runoff routing component. The PDM component results in the best performance at each dimension.

watershed using topographic indexes within the TOPMODEL soil moisture accounting and runoff routing component creates an additional level of reality and restraint over the PDM and BUC components.

Figure 5.3 presents a similar comparison, collecting the models based on the snowmelt accounting method used. In all cases the temperature index model showed a poorer performance than either the radiation index or temperature and radiation index approaches regardless of model dimension. This figure illustrates the improved model performance by the use of net radiation data in explaining the melting of a snowpack, reflecting the importance of net radiation in the energy balance as suggested by *Ohmura* [2001]. In most instances, the combination of temperature and radiation data proved to be

the most able to characterize the snowmelt process, although the radiation index approach resulted in the best eight parameter model.

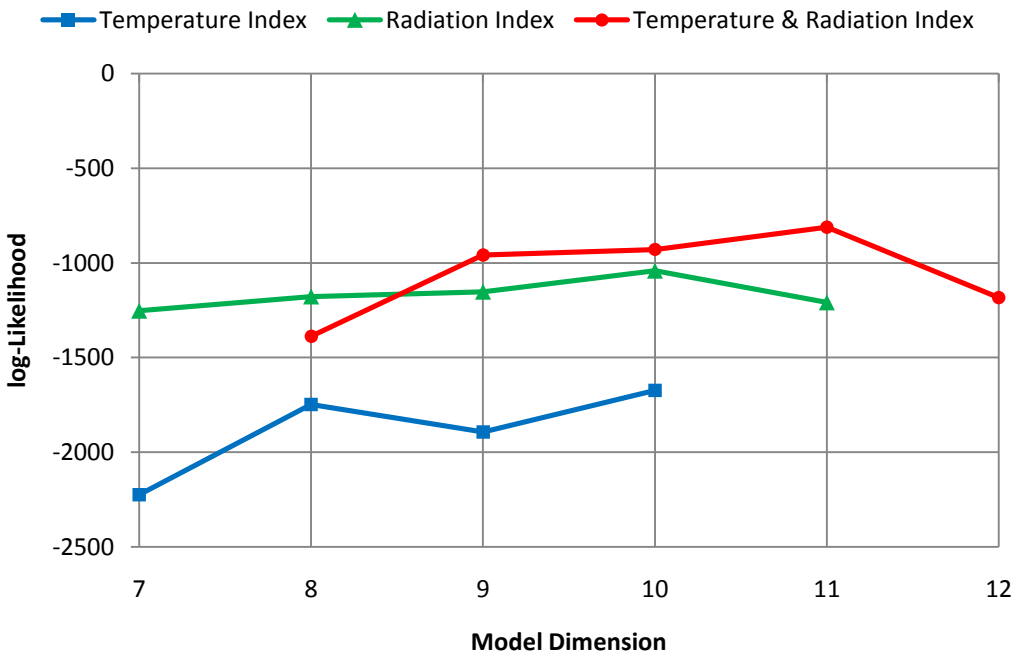


Figure 5.3: Calibration results grouped by the snowmelt accounting component. Snowmelt components utilizing net radiation data far outperform the simple temperature index approach across all model dimensions.

The final grouping, based on the semi-distribution component, is given in Figure 5.4. Due to a smaller overlap among the semi-distribution methods, this figure gives a less clear cut pattern but still offers relevant insight into the behavior of the calibration based on the accounting of precipitation (and snowmelt). The calibration performance associated with each semi-distribution component improves with increasing dimension, suggesting that additional model parameters are relevant to the overall ability of a model structure to simulate observed conditions. Additionally, it can be seen that the semi-distribution scheme with fewer distribution classes provides superior performance to its more complex counterpart at each overlapping model dimension. The combined effect of

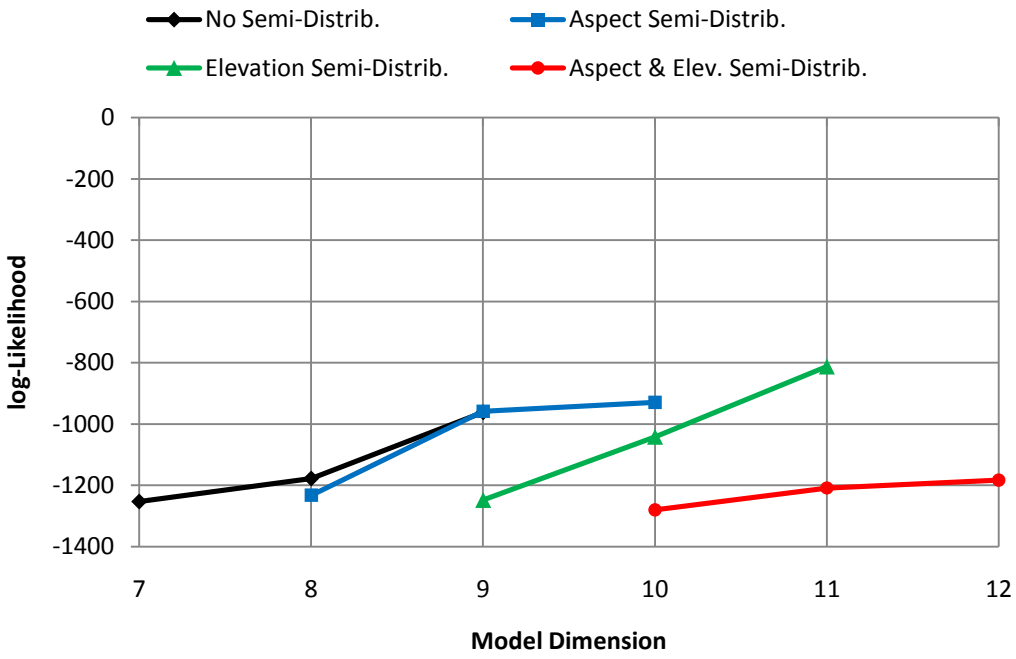


Figure 5.4: Calibration results grouped by the precipitation semi-distribution component. This comparison indicates the precipitation semi-distribution approaches provide little to the model aside from additional flexibility.

these observations strongly support the premise that the semi-distribution component is more important in its ability to add more parameters (and hence allow these parameters to be tuned to the observed data) than in its ability to provide an enhanced process-based explanation of the dominant processes on the streamflow at the watershed's outlet.

Optimum Model Structure Under Calibration: The model structure that resulted in the highest calibrated log-likelihood value, defined by (5.1), was the Probability Distributed Model, incorporating semi-distributed by elevation and accounting for snowmelt with the temperature and radiation index approach (PDMsdE_TRn). This result is given the calibration data and the assumptions of the likelihood function. We now examine the performance of this model in greater detail.

In implementing the DRAM algorithm for this model, samples from the model parameter distributions are generated. Figure 5.5 shows the sampled parameter value

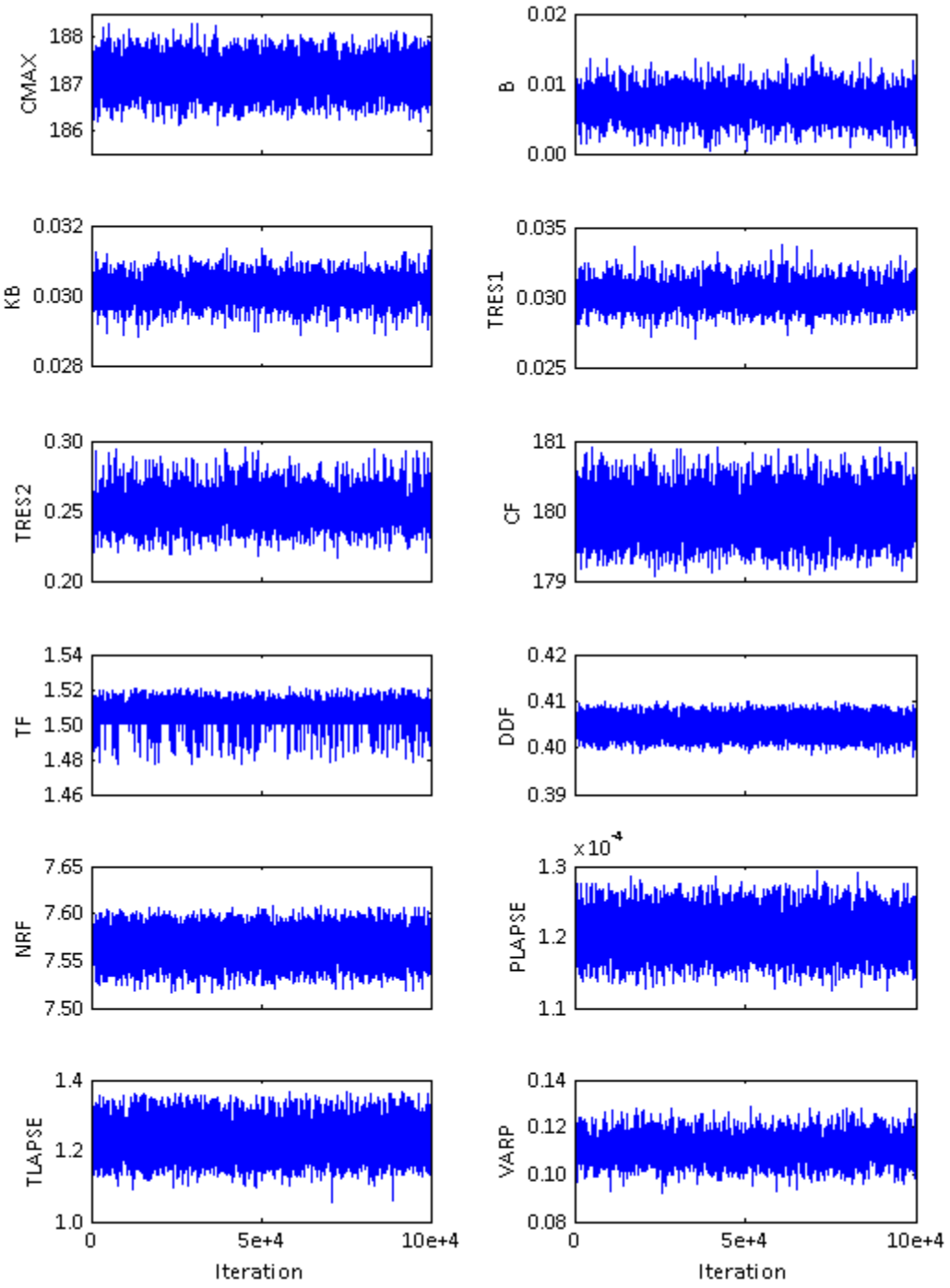


Figure 5.5: Parameter traces from the calibration of the PDMsdE_TRn model structure. Each parameter for this structure was shown to have converged based on visual inspection. Units are as previously defined.

versus iteration for each model parameter. The parameter trace data from multiple parallel runs of the calibration algorithm are used in the assessment of convergence. In this case, convergence was assessed visually based on the evidence provided by the parameter traces. Additionally, the sampled parameters show the range of values for each parameter in the model.

Applying the "best" set of calibrated parameter values to the model structure allows for a direct comparison of observed and predicted variables simulated by the model. Plotted in Figure 5.6 are the observed and predicted streamflows for the PDMsdE_TRn structure and the calibration time period. Visually, the model simulation

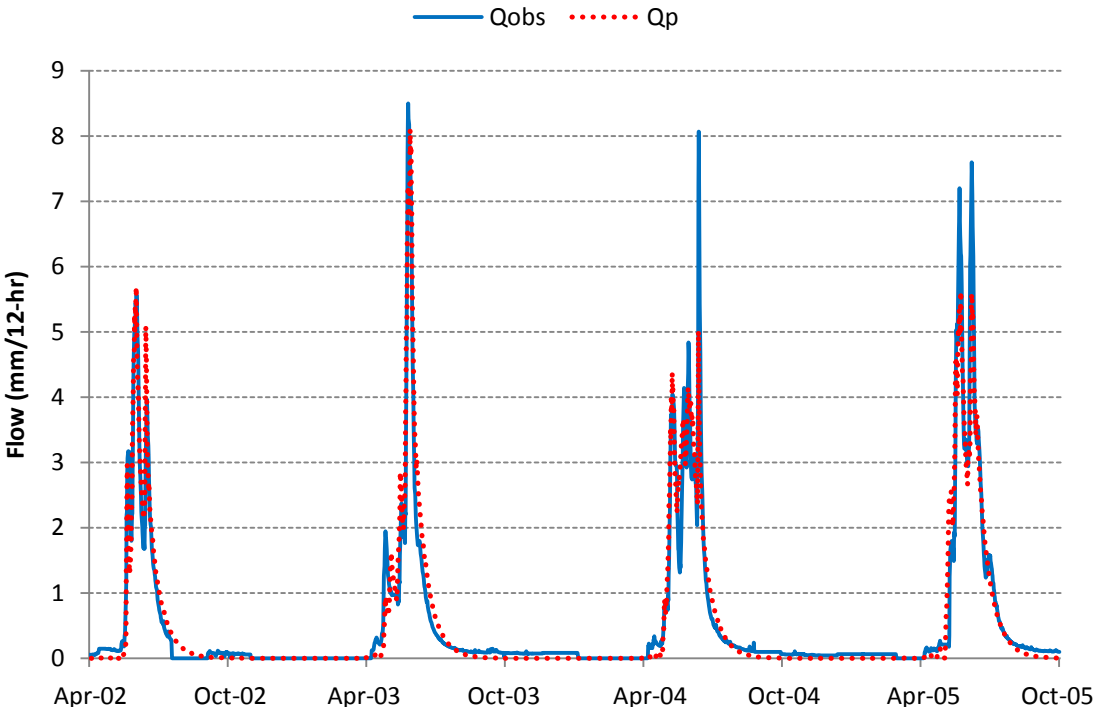


Figure 5.6: Hydrograph simulation for the PDMsdE_TRn model structure and the calibration data set.

generally fits the observed streamflow hydrograph. The shape and timing of the critical features of each hydrograph from the four year calibration period (peak runoff, peak

magnitude, rising and falling limbs) are generally reproduced, even given the variability in wetness from year to year. The greatest deficiency is the slight underestimation of peak flow magnitudes in the last two years of the calibration period corresponding to a regular underestimation of peak snow water equivalents, restricting the amount of water available to streamflow.

Assessment of Model Performance Using Independent Data Source: The assessment of model performance in a time period not used for model specification (also known as the validation phase) is of great importance in assessing a model's ability to simulate streamflow in conditions separate to those it was calibrated under, in a forecasting mode. The assessment of each of the model structures was performed for the validation period, considering data from October 2006 through September 2007 and the optimum parameter values found during the calibration period.

The simulation of the optimal calibrated model for the validation period is given in Figure 5.7. The model simulation for the first validation year is quite strong, matching closely the timing of streamflow response, though slightly under predicting the peak flow. For the second validation year, the fit is less accurate in the timing of the rising and falling limbs of the hydrograph, but does approximately match the peak flows. It should be noted that the second validation year features a much more complex hydrograph (multiple peaks) than any of the four calibration years or the first validation year. This complexity was brought about by a warming phase resulting in the release of melt waters followed by a cooling phase and further accumulation of snowpack; this pattern resulted in multiple melt periods leading to multiple peaks on the hydrograph. The model

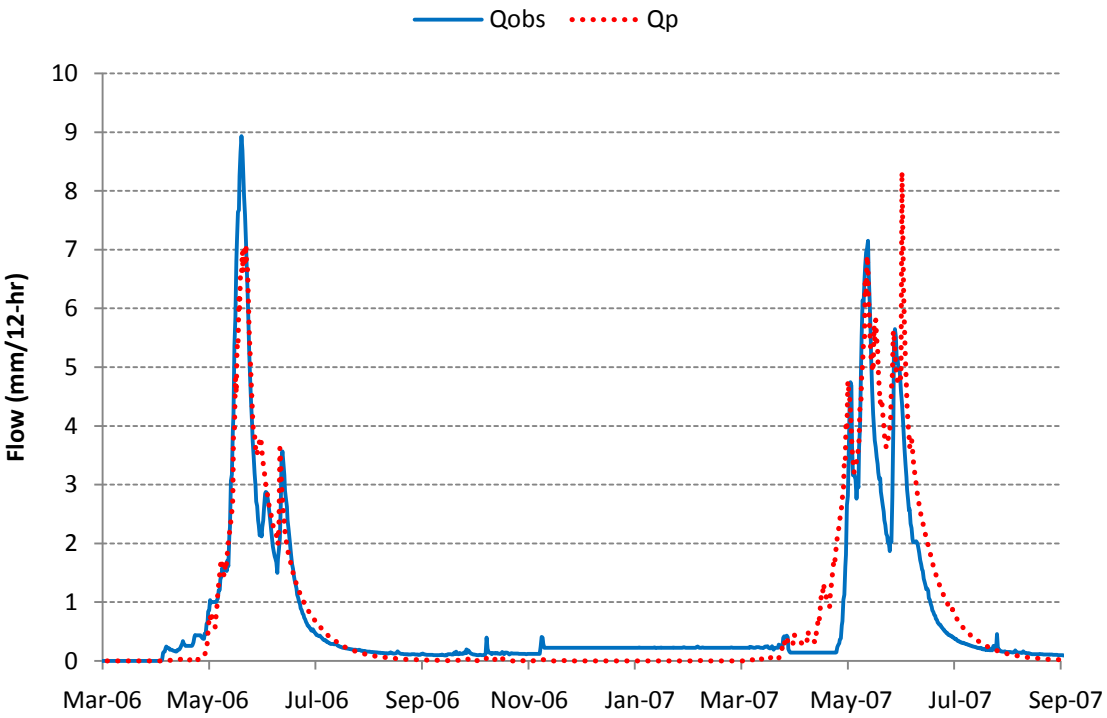


Figure 5.7: Hydrograph simulation for the PDMsdE_TRn model structure and the validation data set.

struggles to accurately fit the complexity of this shape due to a biasness of the parameters toward single peak hydrographs and the assumption of time constant snowmelt parameters that are incapable of adapting to increasingly complex snowpack ripening circumstances.

The log-likelihood value associated with simulation of the validation period for each model structure is provided in Table 5.2. Differences in precipitation events (both magnitude and timing) between the calibration and validation periods can be seen as large contributors to the reshuffling of the rank for the performance of each model structure given in Table 5.2. This is especially evident in the bucket and Probability Distributed models. Despite the climatic differences between the calibration and validation periods used in this study, the best performing TOPMODEL-based structure in

the calibration stage was also the best performing in the validation phase. The consistent performance of the TOPsdA_TRn structure suggests that this formulation may be more robust to climatic variability than other structures tested.

Table 5.2: Model Performance-Based Assessment Criteria for the Validation Period[a]

Model Name	Model Dimension	Maximum log-Likelihood	NSE
BUC_Rn	7	-2337.46	0.69163
BUC_T	7	-1208.53	0.80803
BUC_TRn	8	-1035.42	0.85699
BUCsdA_Rn	8	-1847.31	0.75993
BUCsdA_TRn	9	-1249.38	0.84826
BUCsdE_Rn	9	-2373.32	0.68392
BUCsdE_T	9	-1107.10	0.83306
BUCsdE_TRn	**10**	**-736.02**	**0.90035**
BUCsdAE_Rn	10	-1114.11	0.84929
BUCsdAE_TRn	11	-922.10	0.87472
PDM_Rn	8	-1370.44	0.823652
PDM_T	8	-1370.47	0.784985
PDM_TRn	9	-1085.22	0.865853
PDMsdA_Rn	9	-1816.28	0.770846
PDMsdA_TRn	10	-1008.16	0.87586
PDMsdE_Rn	10	-2464.26	0.708855
PDMsdE_T	10	-941.70	0.867032
PDMsdE_TRn	11	-1277.55	0.854237
PDMsdAE_Rn	11	-964.88	0.871012
PDMsdAE_TRn	**12**	**-773.20**	**0.895679**
TOP_Rn	7	-2402.56	0.498395
TOP_T	7	-1199.50	0.810923
TOP_TRn	8	-1061.04	0.843417
TOPsdA_Rn	8	-2332.16	0.575963
TOPsdA_TRn	**9**	**-778.41**	**0.896242**
TOPsdE_Rn	9	-2339.91	0.526481
TOPsdE_T	9	-1590.32	0.676818
TOPsdE_TRn	10	-1560.74	0.722812
TOPsdAE_Rn	10	-2317.75	0.528169
TOPsdAE_TRn	11	-1656.70	0.727669

[a]The values highlighted in bold indicate the best model structure for each soil moisture accounting and runoff routing component.

As with the calibration period, Figure 5.8 reveals that the general response to increasing the model dimension is an increase in the predictive performance. However the trends present in the calibration phase when the model structures were grouped based on soil moisture accounting and runoff routing, snowmelt accounting and precipitation semi-distribution are less clear for the validation phase (see Figure 5.9, Figure 5.10 and Figure 5.11). These results further highlight the differences in data contained in the calibration and validation data sets.

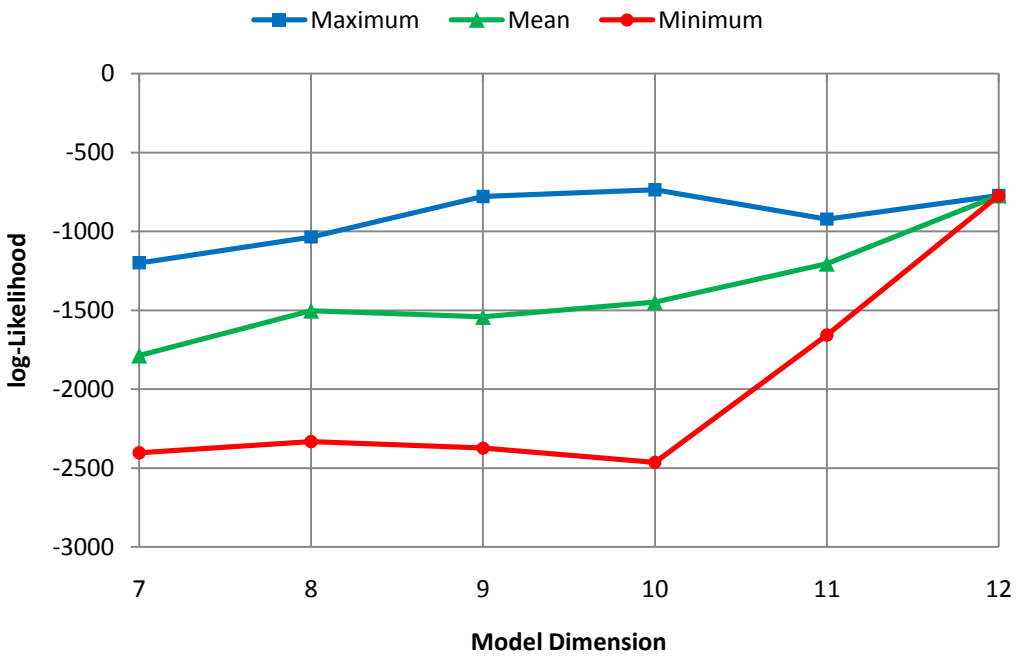


Figure 5.8: Validation results grouped by maximum, mean and minimum log-likelihood value for each model dimension. On average, there is an increase in predictive performance with an increase in complexity, similar to the calibration period.

While the TOPMODEL-based structures performed the poorest across each model dimension in the calibration phase and the Probability Distributed Model structures performed the best, in the validation phase this relationship no longer holds. Figure 5.9 shows the significant crossover in performance with increasing model

dimension, where no single base structure component establishes an advantage in predictive performance. It can be seen, however, that the BUC and PDM components tend to result in improved performance with increasing dimension, while the TOP component experiences a clear maximum performance for the median dimension level (nine model parameters).

Focusing on the grouping of models based on the snowmelt accounting component, the relationship between performance and dimension from the calibration

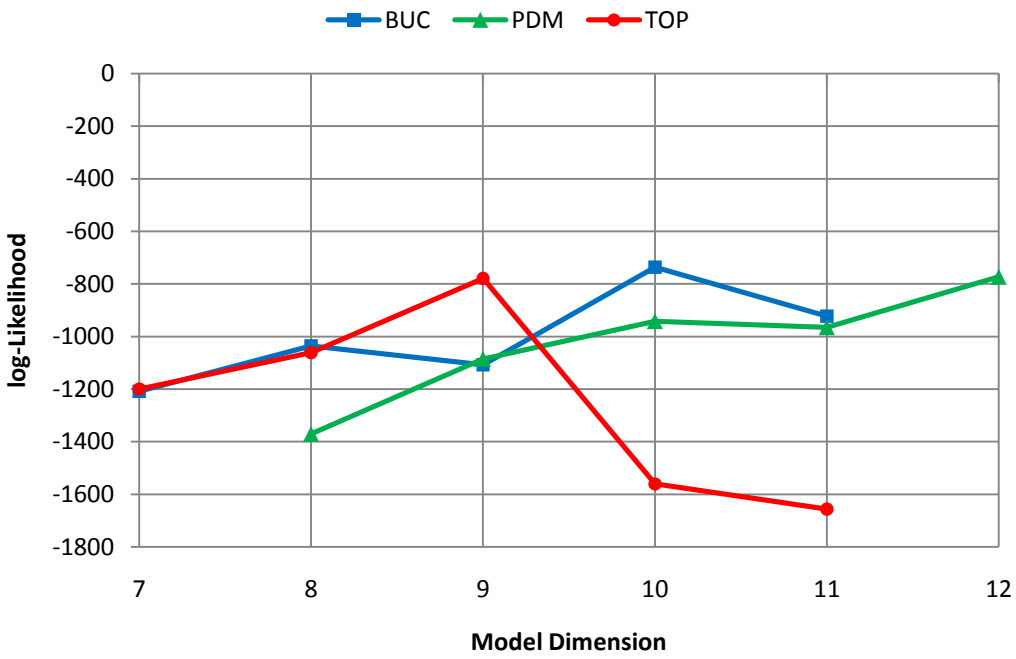


Figure 5.9: Validation results grouped by the soil moisture accounting and runoff routing component.

phase is partially maintained. Figure 5.10 illustrates the continuing dominance of the temperature and radiation index approach in the validation phase, however, the temperature index approach has replaced the radiation index approach as the next best description of the snowmelt process for the validation time period. The decline in performance for the radiation index approach suggests that the unique information

provided by the combination of the temperature and net radiation data may be more robust than either the net radiation or temperature data alone.

Similar to the change occurring between calibration and validation periods for the structures when grouped by base structure component, when the structures were grouped by the precipitation semi-distribution component the consistent relationship seen in the

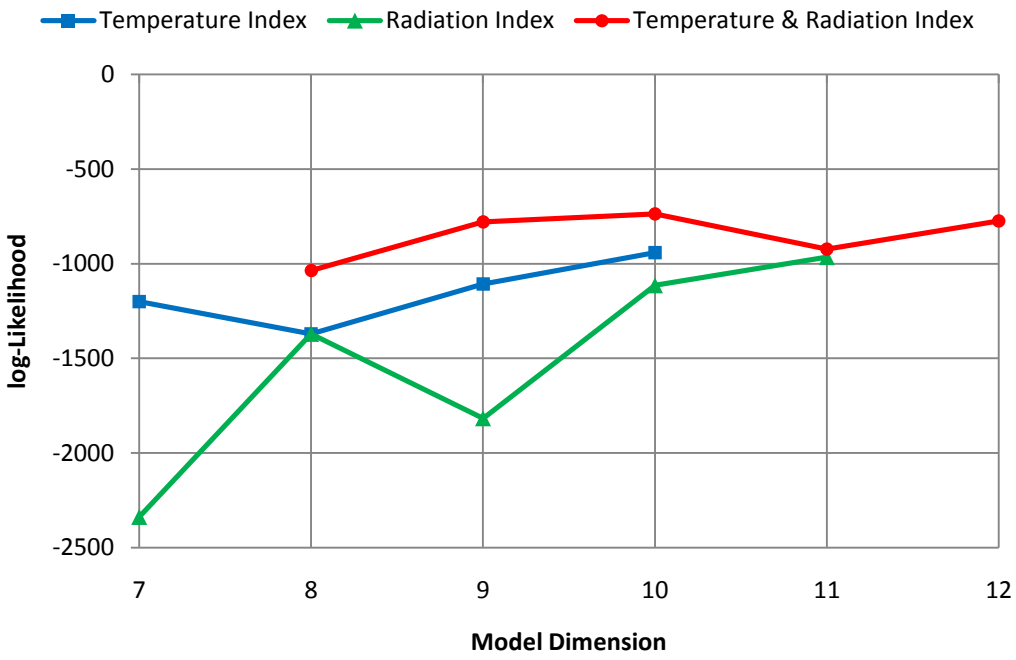


Figure 5.10: Validation results grouped by the snowmelt accounting component.

calibration period became inconsistent for the validation period. Figure 5.11 reveals no clear pattern providing explanation for the relevance of one semi-distribution technique over another.

The differences in values of the likelihood function and patterns associated with the different structural components of each model between the calibration and validation phases further complicate the issue of model selection. Of course, in an operational setting all the available data would be used to calibrate the model. In this study, the

overall goal was not to identify the best model structure for analyzing the Stringer Creek watershed, but rather to identify the components that are best able to reflect the true processes they model. Ultimately, in the calibration phase the PDM component proved to perform the best, suggesting that this structure could best represent the soil water

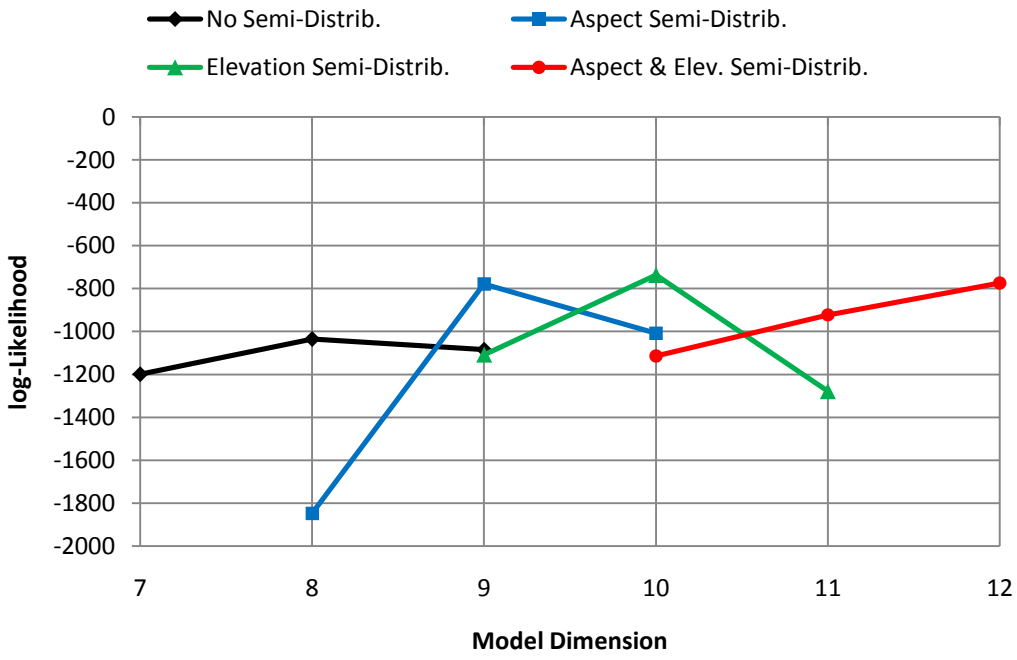


Figure 5.11: Validation results grouped by the precipitation semi-distribution component. The relationship of increased performance with increasing dimension seen in the calibration period breaks down for the validation period.

variability of the Stringer Creek watershed. However, the change in model ranking during the validation phase may suggest that the PDM component is not necessarily the most appropriate, and in fact may have performed the best in the calibration phase due to a greater data fitting flexibility. The variability in the validation data does not discredit this conclusion, but does diminish the persuasiveness of the validation phase assessment. The disconnect between calibration and validation phase assessment results points to the

conclusion that the models are best used under "typical" conditions, with more extreme conditions causing a breakdown in model performance.

Comparison of Likelihood Function Assumptions: In order to better investigate the assumptions on the error model brought about by the application of a likelihood function, the ten model structures associated with the Probability Distributed Model component were further examined using a second likelihood function assuming heteroscedastic and uncorrelated errors

$$P(Q|\theta) = (2\pi\sigma^2)^{-n/2} \prod \left[\exp\left\{ -\frac{\left\{\log\left[\frac{Q_{obs}+0.3}{Q_p+0.3}\right]\right\}^2}{2\sigma^2} \right\} \cdot (Q_{obs}+0.3)^{-1} \right] \quad (5.5)$$

where all terms are as defined previously. By considering multiple likelihood functions it becomes possible to better understand the implications of such selections.

Table 5.3 compares the maximum calibrated value for each of the two likelihood functions considered for the Probability Distributed Model soil moisture accounting and runoff routing structures. Applying two different likelihood functions to the same model and data permits the two to be compared directly. In this case, the likelihood function assuming heteroscedastic Gaussian errors (Eqn 5.5) substantially outperforms the likelihood function assuming homoscedastic errors (Eqn 5.1). This result strongly suggests that the errors of the calibration data set are better characterized by the second likelihood (i.e., heteroscedastic and uncorrelated rather than normal, independent and identical). Because of the difference in how the errors are described between the two

Table 5.3: Comparison of Likelihood Functions for Calibration of the Probability Distributed Model Component[a]

Model Name	Model Dimension	log-Likelihood (Eqn 5.1)	log-Likelihood (Eqn 5.5)
PDM_Rn	8	-1177.3	2442.2
PDM_T	8	-1747.6	1824.3
PDM_TRn	9	**-960.9**	2543.9
PDMsdA_Rn	9	-1152.0	**2581.8**
PDMsdA_TRn	10	**-929.0**	**2597.9**
PDMsdE_Rn	10	-1040.9	**2561.4**
PDMsdE_T	10	-1673.0	2147.4
PDMsdE_TRn	11	**-811.5**	2556.0
PDMsdAE_Rn	11	-1208.3	2344.4
PDMsdAE_TRn	12	-1182.8	2428.3

[a]The values highlighted in bold represent the three best model structures, as identified by each of the likelihood functions.

likelihood functions the overall best structures are not consistent across the likelihoods, although, models that perform well based on the first likelihood function tended to perform well based on the second likelihood function.

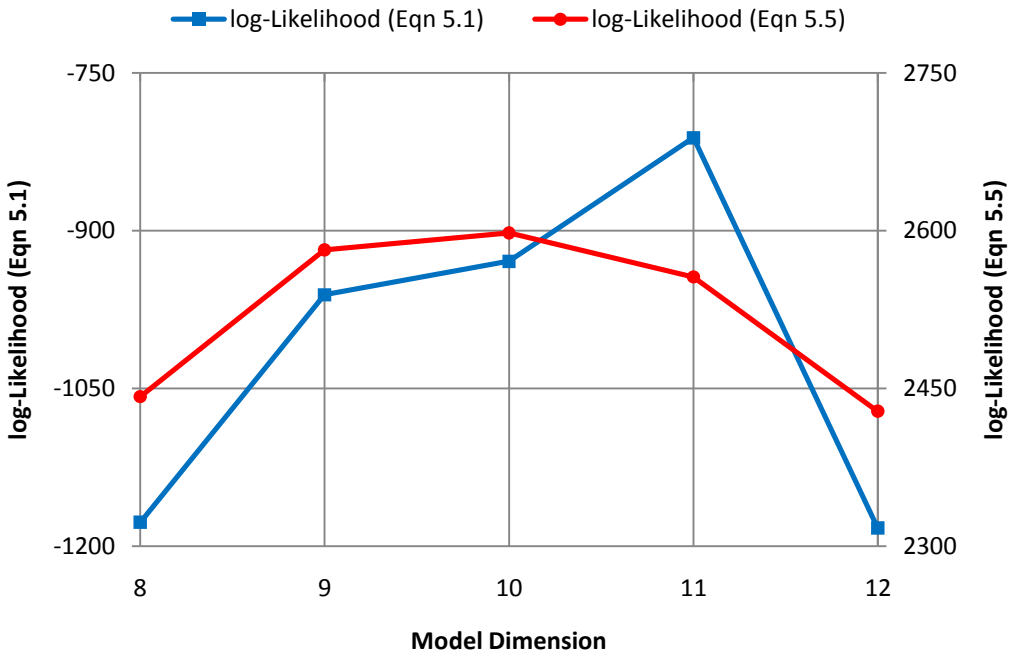


Figure 5.12: Likelihood function comparison for the best Probability Distributed Model based structure at each dimension. A similar pattern exists across the differing likelihood functions relating performance and complexity.

Figure 5.12 further illustrates this general relationship, plotting the best performing model structure at each of the five model dimensions versus the value of the likelihood function. Both likelihood functions describe some amount of diminishing returns associated with adding more parameters to the model, the primary difference being the exact location of the balance point (11 for Gaussian errors and 10 for heteroscedastic Gaussian errors). Figure 5.12 highlights the trade-off between performance and complexity typified by top-down model construction. Additional parameters do not necessarily provide a better description of the data, however, the structural components themselves become important to the model prediction.

Uncertainty-Based Assessment

The uncertainty associated with any single model structure is of substantial importance to the understanding of the risk associated with its predictions [*Chiang et al.*, 2007]. An appropriate assessment of uncertainty has the ability to temper or intensify the weight of the performance-based assessment strategies discussed previously, by introducing further criteria for selecting the optimal structure. In this section, the structural uncertainty across the thirty models will be assessed, as well as the uncertainty in the individual model parameters and structures.

The structural uncertainty present in the modeling suite is brought about by alternative representations of the key watershed processes of soil moisture accounting and runoff routing, snowmelt accounting and precipitation semi-distribution. By pooling the model simulations of each model structure at each time step, the overall structural uncertainty of the modeling suite is given. This is shown in Figure 5.13. The variability in

the predicted response of streamflow across all of the modular structures was examined, where the range of predictions is plotted for each time step. In general, the predicted

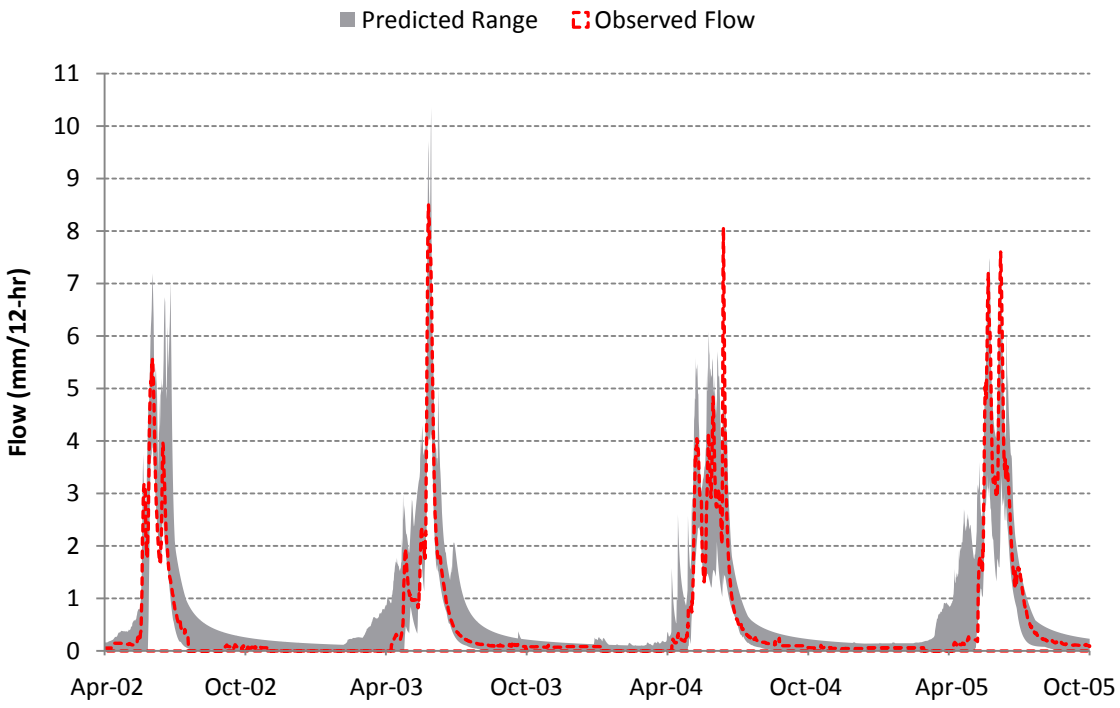


Figure 5.13: Range of predictions on the calibration data set from the collection of model structures investigated.

range generated from the modeling suite captures the timing of the observed flow, but is more flashy in its simulation (i.e., gives a rapid response of streamflow to precipitation events). The width of the predicted range is much larger during the rise and fall of the hydrograph, signaling that this is the location where the true differences between model structures lie. The predicted range on the calibration data encapsulated approximately 73% of the observed data record.

A similar representation was developed for the modeled simulations on the validation data period. Again, Figure 5.14 displays the predicted range of simulations for

each time step. Much as with the predicted ranges for the calibration period, the predicted ranges for the validation period illustrate the flashiness in the response of the simulated

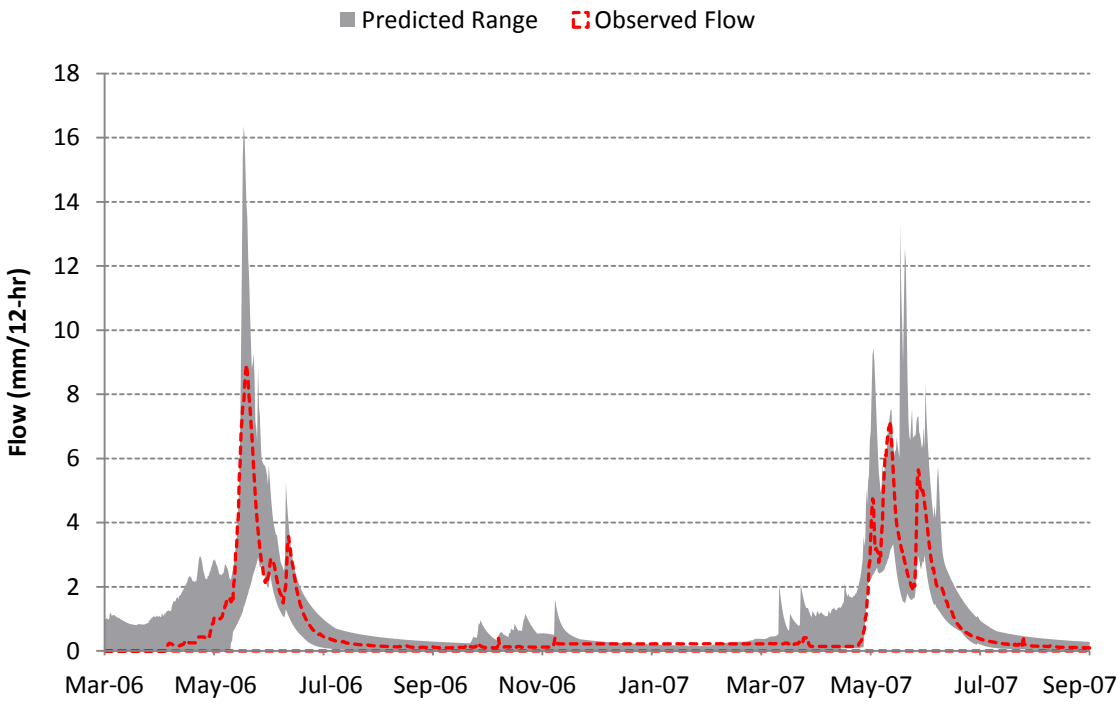


Figure 5.14: Range of predictions on the validation data set from the collection of model structures investigated.

streamflow to (often minor) precipitation events. The predicted range on the validation period resulted in a greater percent of capture of the observed flow data (at approximately 91%) than was found with the predicted range for the calibration period signifying a general increase in uncertainty about the structures during the validation period.

Uncertainty Assessment of Individual Model Structures: Although Figure 5.13 and Figure 5.14 represent the uncertainty among the thirty model structures considered, the uncertainty associated with each structure in particular may be accounted for, as well. In a Bayesian approach, the parameter uncertainty is expressed by the posterior

distributions of the model parameters. The information contained in the parameter traces of a converged model (see Figure 5.5) are used to construct the posterior probability distribution for each model parameter. Figure 5.15 contains histograms of the posterior

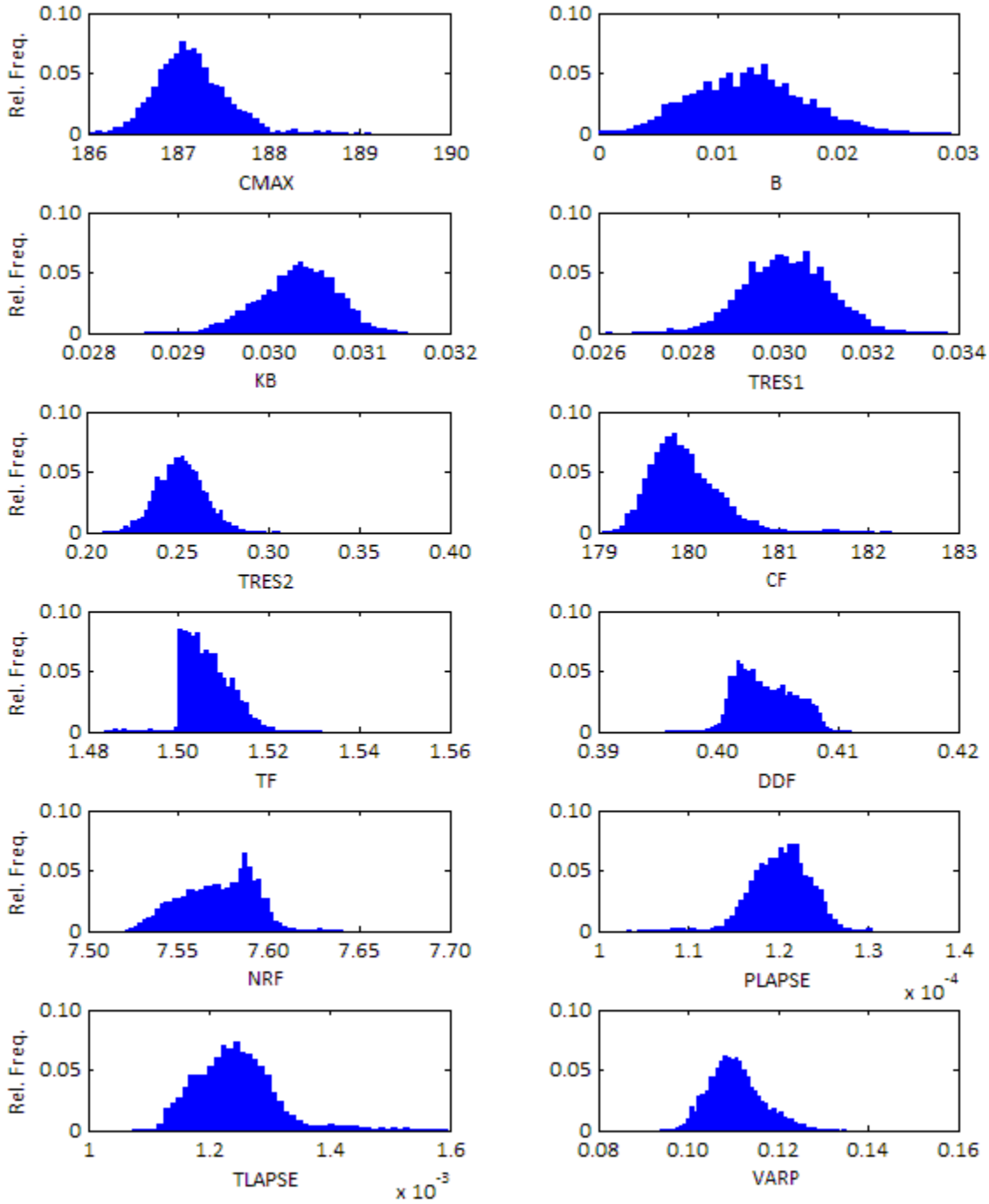


Figure 5.15: Parameter posterior histograms from the calibration of the PDMsdE_TRn model structure.

distribution samples for the PDMsdE_TRn model structure, representing the parameter uncertainty given the calibration data. The histograms of Figure 5.15 represent the relative frequency of occurrence for a particular parameter value and characterize the degree of confidence in the parameter. Shape qualities important to the histogram interpretation, such as bimodality, threshold behavior or peakedness reveal parameter behaviors. The threshold behavior of the TF parameter (the threshold temperature for snowmelt) suggests this parameter has strong interactions with other parameters, while the lack of a peaked histogram for the B parameter (spatial variability of watershed soil moisture distribution) suggests that this parameter is less sensitive in the model than a parameter with a more peaked histogram as is the case with the CMAX or CF parameters (maximum soil water content capacity and soil moisture threshold for subsurface storage inflow, respectively).

Further identification of uncertainty in the parameters and model structure can be done via the calculation of prediction intervals for the simulation on the validation period. Evaluation of uncertainty in this manner makes use of two important statistics: the *reliability* and the *sharpness* of the model's prediction. The reliability describes the percent of observed data captured by the prediction interval and sharpness is the mean width of the prediction interval [*Yadav et al.*, 2007].

A 95% prediction interval was calculated for each model structure's prediction of streamflow provided the validation data set, given by

$$PI = Q_p \pm z^* \cdot \sqrt{VARP} \tag{5.6}$$

where Q_p is the predicted streamflow, z^* is the critical value of Student's t distribution with infinite degrees of freedom (1.96 for 95% prediction interval) and $VARP$ is the statistical variance (a sampled parameter in model calibration estimated via the likelihood function). Table 5.4 gives the results of the reliability and sharpness metrics for each structure. The reliability metric is primarily a check on prediction interval itself, where a

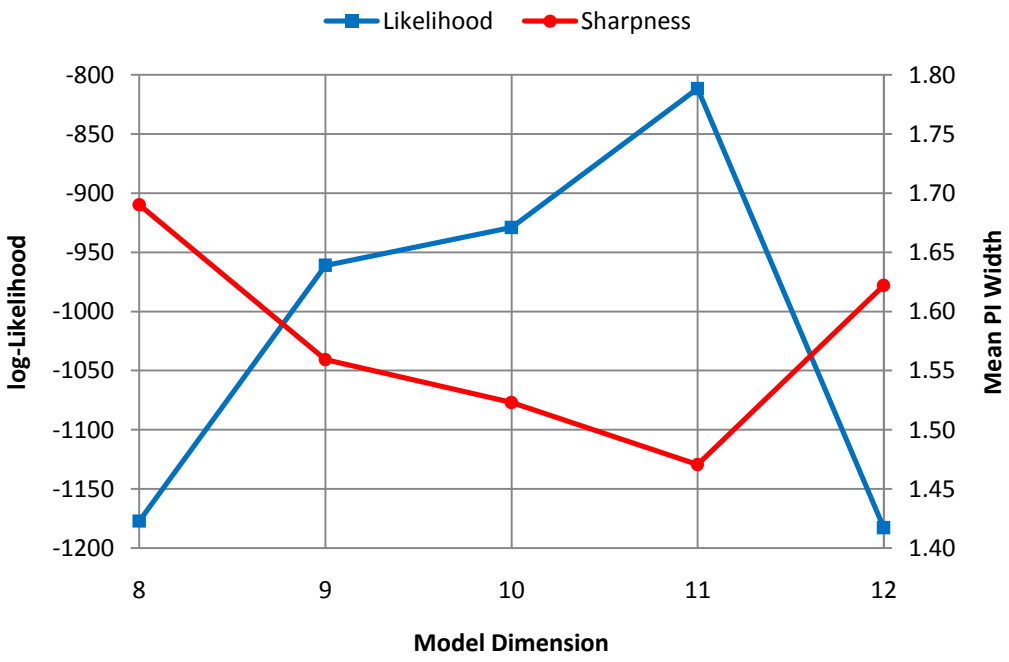


Figure 5.16: Comparison of sharpness and log-likelihood versus model dimension. There is a general inverse relationship seen between sharpness and log-likelihood, suggesting that the sharpest models are the best performing.

reliability value near 95% is desirable, values lower mean the interval is too narrow and higher mean the interval is too wide. In this case, reliability ranged from 89.6% to 97.6% with a mean of 93.9%. Sharpness ranged from approximately 1.5 to 3.1 with a mean of 2.0.

Focusing on the relationship between likelihood and sharpness, the best performing model at each dimension for the Probability Distributed Model based

Table 5.4: Model Uncertainty-Based Assessment Criteria[a]

Model Name	Model Dimension	Reliability	Sharpness
BUC_Rn	7	90.34%	1.715
BUC_T	7	94.38%	2.481
BUC_TRn	8	95.72%	1.830
BUCsdA_Rn	8	90.82%	1.720
BUCsdA_TRn	**9**	90.34%	**1.562**
BUCsdE_Rn	9	94.06%	1.719
BUCsdE_T	**9**	**95.38%**	2.203
BUCsdE_TRn	10	96.20%	1.703
BUCsdAE_Rn	10	92.96%	1.664
BUCsdAE_TRn	11	94.13%	1.684
PDM_Rn	8	94.34%	1.690
PDM_T	**8**	**95.31%**	2.164
PDM_TRn	9	92.48%	1.559
PDMsdA_Rn	9	91.93%	1.675
PDMsdA_TRn	10	92.34%	1.523
PDMsdE_Rn	10	94.00%	1.597
PDMsdE_T	10	96.27%	2.142
PDMsdE_TRn	**11**	92.89%	**1.471**
PDMsdAE_Rn	11	95.45%	1.751
PDMsdAE_TRn	12	95.38%	1.622
TOP_Rn	**7**	**95.10%**	2.447
TOP_T	7	94.06%	2.539
TOP_TRn	8	95.65%	2.265
TOPsdA_Rn	8	95.45%	2.241
TOPsdA_TRn	**9**	94.82%	**1.630**
TOPsdE_Rn	9	89.58%	2.278
TOPsdE_T	9	96.14%	3.074
TOPsdE_TRn	10	97.58%	2.383
TOPsdAE_Rn	10	90.61%	2.385
TOPsdAE_TRn	11	92.27%	2.192

[a]The values highlighted in bold indicate the most reliable (nearest to 95%) and sharpest (smallest sharpness) model structure for each soil moisture accounting and runoff routing component.

structures are plotted in Figure 5.16. From Figure 5.16 (above) it becomes apparent that there is a general inverse relationship between likelihood and sharpness. Models with a better likelihood tend to have a narrower mean prediction interval width. This result is

not surprising given that the better performing models tended to have smaller sampled variances, leading to the thinner prediction intervals. However, as mean prediction interval width decreases, the percent of observed data captured also tends to decrease (i.e., increased sharpness comes at the cost of decreased reliability).

Figure 5.17 displays the range of predictions from the pooled collection of model structures' associated 95% prediction intervals. The predicted range in this case captures

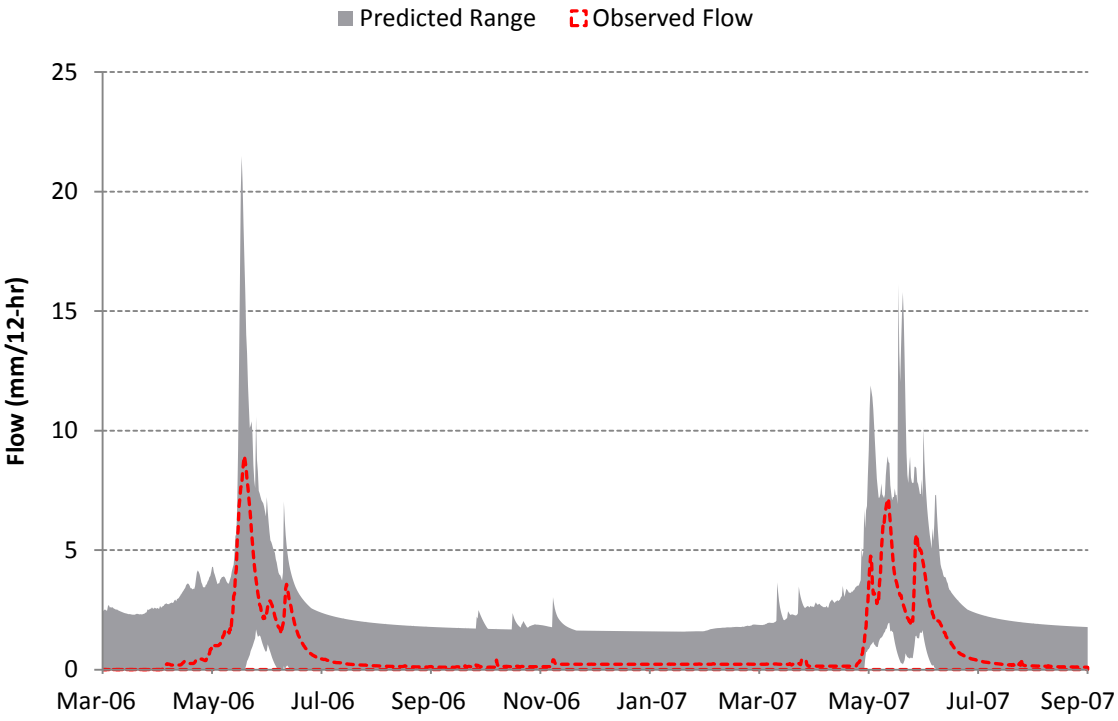


Figure 5.17: Range of predictions on the 95% prediction intervals over the validation data set from the collection of model structures investigated.

100% of the observed streamflow data, indicating that if parameter uncertainty, model predictive uncertainty and structural uncertainty are considered then all of the observed data is captured by the model simulations. The individual prediction intervals on the validation data for each model structure are provided in Appendix A and describe each

individual model structure's ability to capture the observed data considering each of the uncertainties mentioned previously.

Continuing the pursuit of identifying the model components best able to simulate observed streamflows in the Stringer Creek watershed, the uncertainty-based assessment was designed to provide another check on model suitability beyond typical performance-based methods. The result generally confirms the PDMsdE_TRn structure as the best model when considering the uncertainty propagated through the model in the calculation of a 95% prediction interval. The PDMsdE_TRn model was found to have the narrowest average prediction interval width and a correspondingly low reliability of 92.9%, in addition to having the best performance via the likelihood function. The final subsection will consider model realism as a final test of overall model appropriateness.

Realism-Based Assessment

The assessment of precipitation-runoff models based on one or more internal system processes not used in model calibration is valuable in relation to a model's overall success in a forecasting mode. Previous studies have found that precipitation and streamflow data lack the information content required to prove a structure to be a realistic description of the watershed [e.g., *Jakeman and Hornberger*, 1993; *Uhlenbrook et al.*, 1999]. In conceptual modeling, the process of linking model variables to actual watershed states is especially difficult due to the intrinsically aggregated nature of the true spatial variability; however, while difficult, studies have attempted to relate conceptual models to watershed processes [*Seibert and McDonnell*, 2002].

Accepting the limitations of the forcing data (precipitation, streamflow), snow water equivalents data was used as a first step toward checking the model reality. This data represents an independent view on the model's simulation of a vital watershed

Table 5.5: Nash-Sutcliffe Efficiencies for the Simulation of Snow Water Equivalents on the Calibration and Validation Periods[a]

Model Name	Model Dimension	NSE_{CAL}	NSE_{VAL}
BUC_Rn	7	0.925	0.927
BUC_T	7	0.840	0.924
BUC_TRn	8	0.901	0.937
BUCsdA_Rn	8	0.922	0.927
BUCsdA_TRn	9	0.929	0.937
BUCsdE_Rn	9	0.905	0.941
BUCsdE_T	9	0.833	0.927
BUCsdE_TRn	10	0.891	0.934
BUCsdAE_Rn	10	0.936	**0.944**
BUCsdAE_TRn	11	**0.937**	0.938
PDM_Rn	8	0.908	0.919
PDM_T	8	0.836	0.931
PDM_TRn	9	0.917	0.933
PDMsdA_Rn	9	0.922	0.927
PDMsdA_TRn	10	0.915	0.931
PDMsdE_Rn	10	0.919	0.932
PDMsdE_T	10	0.881	**0.950**
PDMsdE_TRn	11	0.902	0.937
PDMsdAE_Rn	11	**0.936**	0.944
PDMsdAE_TRn	12	**0.936**	0.937
TOP_Rn	7	0.697	0.730
TOP_T	7	0.528	0.680
TOP_TRn	8	0.700	0.829
TOPsdA_Rn	8	0.895	0.916
TOPsdA_TRn	9	**0.903**	**0.941**
TOPsdE_Rn	9	0.752	0.798
TOPsdE_T	9	0.638	0.718
TOPsdE_TRn	10	0.816	0.926
TOPsdAE_Rn	10	0.752	0.797
TOPsdAE_TRn	11	0.709	0.798

[a]The values highlighted in bold represent the best value associated with each of the three soil moisture accounting and runoff routing components for the calibration and validation periods.

process, specifically the development and melting of a snowpack. Although useful in a broad sense, this data on snowpack is invariably biased due to the extrapolation of single site measurements to the greater system.

Table 5.5 (above) presents the efficiency of the model structures in simulating the snow water equivalents data for the calibration and validation data sets, again using the Nash-Sutcliffe Efficiency metric. The typical response of the thirty modular structures points to a general increase in NSE values with increasing model dimension for both the calibration period (NSE_{CAL}) and the validation period (NSE_{VAL}). A comparison of the efficiency of the structures during the calibration period based on pooling the structures for the snowmelt accounting structural component is given in Figure 5.18.

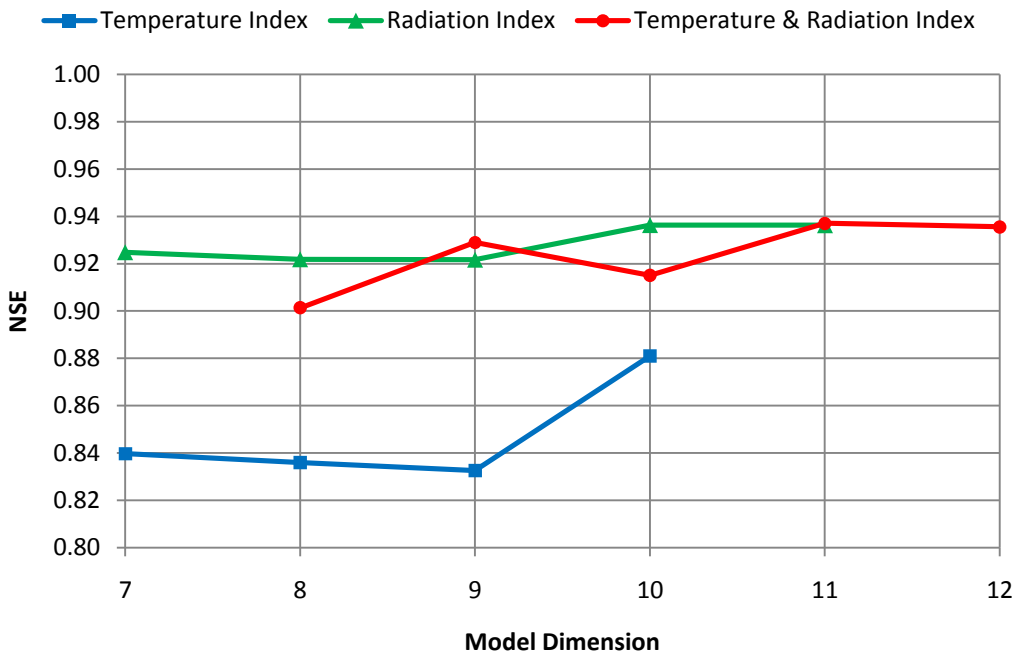


Figure 5.18: Snow water equivalents simulation results grouped by snowmelt accounting component for each model dimension over the calibration period. The temperature index approach is shown to be the least effective snowmelt accounting method over this period.

Much as was the case with the simulation of streamflow, the simulation of snow water equivalents during the calibration period strongly reflects the importance of the net radiation data in providing the most accurate prediction (Figure 5.18). Both the radiation index approach and the temperature and radiation index approach prove to provide the predictions with appreciably better NSE values as compared to the temperature index approach. During the validation period, the relationship between snowmelt accounting components and NSE break down considerably, with no discernable trends persisting.

Concentrating on the PDMsdE_TRn model structure (previously identified as having the best overall predictive performance), Figure 5.19 provides the simulation of snow water equivalents over the calibration period. As with streamflow (see Figure 5.6),

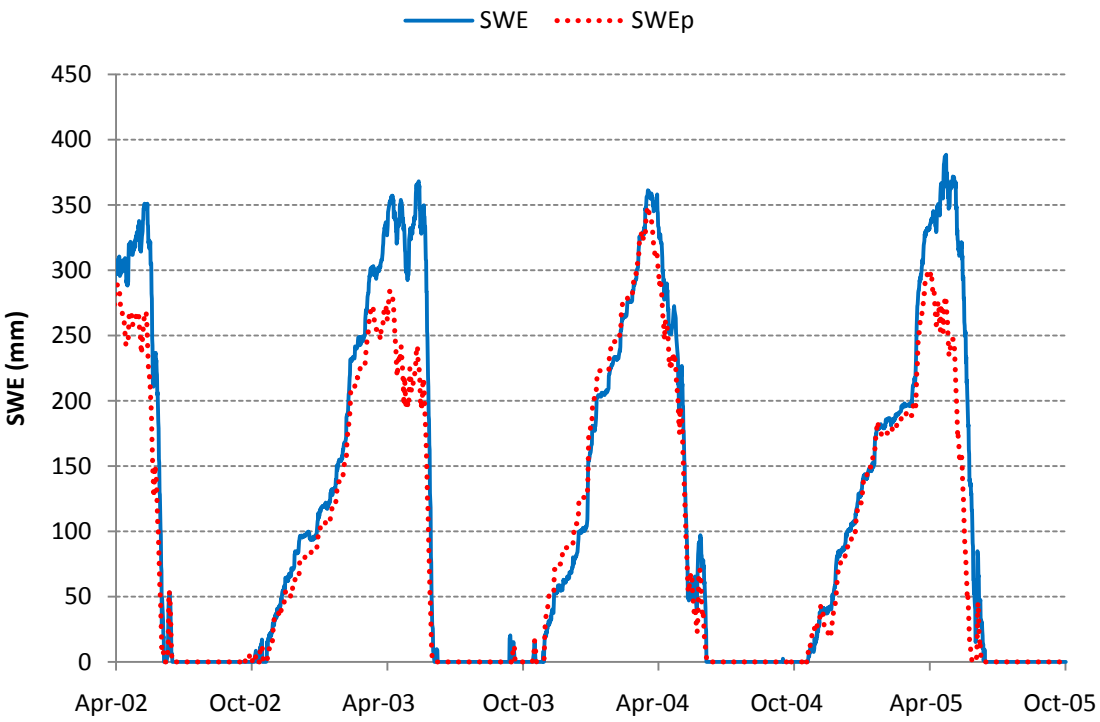


Figure 5.19: Snow water equivalents simulation for the PDMsdE_TRn model structure and the calibration data set.

the timing of the predicted snow water equivalents is rather similar to the observed data, including the development of the snowpack and the timing of melt events. The primary shortcomings are reflected in a consistent underestimation of maximum snow water equivalents for years 1, 2 and 4 of the calibration period. This underestimation in the SWE predictions suggests possible bias in the data or in the rationing of precipitation as either solid or liquid by the decision algorithm discussed in Chapter 4.

For the validation period, Figure 5.20 shows a similar pattern for the October 2005 through September 2006 water year; however, this pattern breaks down considerably in the second validation year. Over the October 2006 through September 2007 water year, the timing of events is relatively consistent, but the trend in snow water

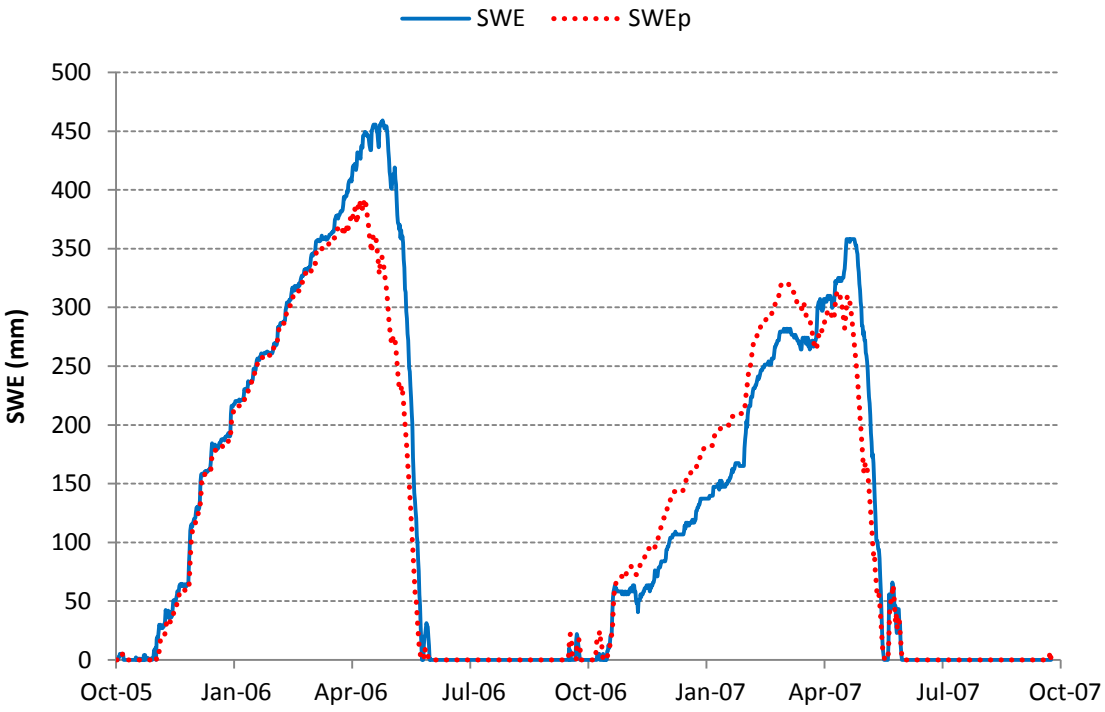


Figure 5.20: Snow water equivalents simulation for the PDMsdE_TRn model structure and the validation data set.

equivalents simulation shifts from a general under prediction of magnitudes to an over prediction during snowpack formation. Again, this change in pattern behavior reflects the difficulty in using a model calibrated to a set of climatic conditions for prediction on another. The occurrence of an early melt followed by a return to normal conditions in the second validation year has caused significant difficulties in the model's simulation of snow water equivalents.

Looking at snow water equivalents simulation as an ensemble of each of the model structures, the predicted range of simulations for the calibration period and validation period are given in Figures 5.21 and 5.22, respectively. These plots show a similar trend among all model structures that was observed for the best structure. For the

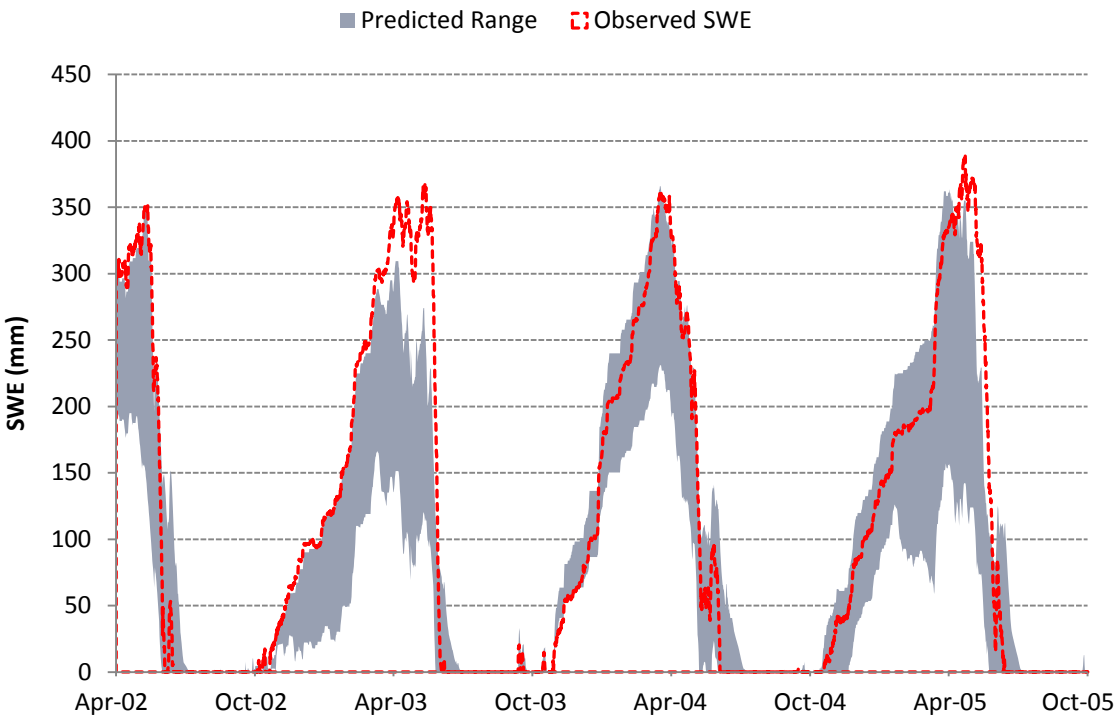


Figure 5.21: Snow water equivalents ensemble simulations for each model structure and the calibration data set.

calibration period, October 2002 to September 2003 is the worst simulated of the four years of data, while the first and third years have the best overall fit. Approximately 74% of the observed data was captured by the range of model predictions, with an average width of 68.8 mm.

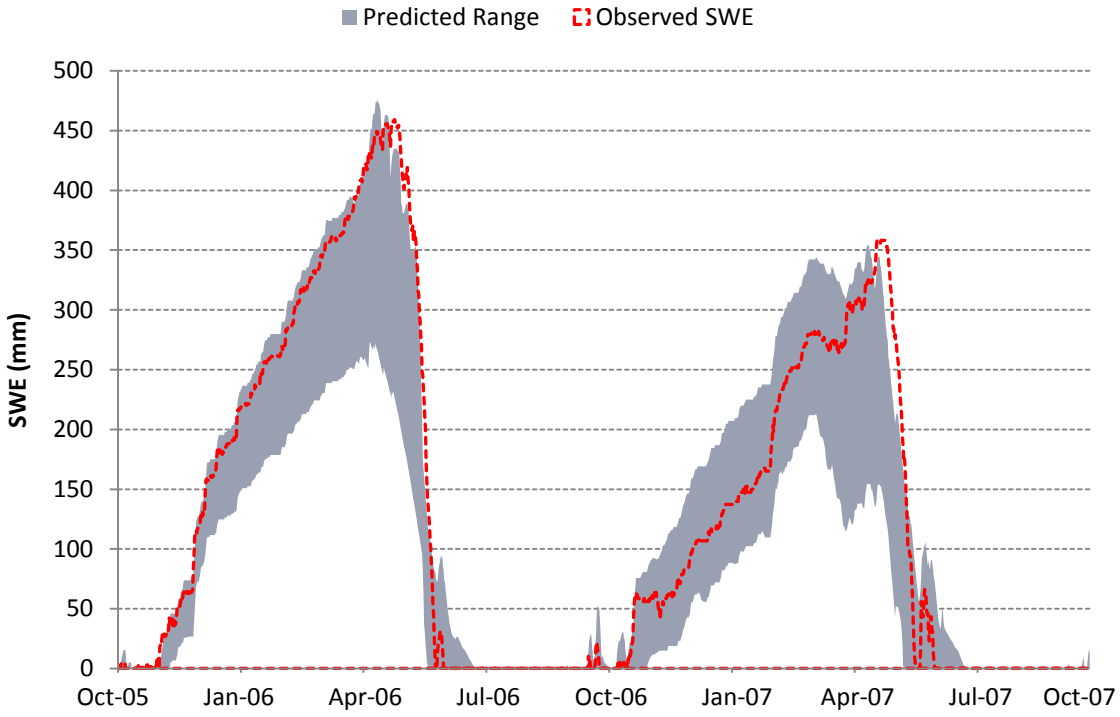


Figure 5.22: Snow water equivalents ensemble simulations for each model structure and the validation data set.

Moving to the validation period, the ensemble predictions again are strongly reminiscent of the general relationship between observed and predicted flows for the PDMsdE_TRn simulation. Year one of the validation period is generally underestimated, but the timing of major events is well matched in the range of predictions across the thirty model structures. The mean width of the predictions was 76.1 mm, which captured roughly 89% of the observed record. The increase in mean width and percent captured

reflect on the increased uncertainty of the structures during the validation period due to the change in conditions of the forcing data.

Based solely on prediction of snow water equivalents during the calibration period a different set of structures are preferred than were for prediction of streamflow. However, several patterns remain consistent between prediction of snow water equivalents and streamflow including (1) a general favoring of bucket model and Probability Distributed Model components over the TOPMODEL component; (2) a preference for snowmelt accounting components that consider net radiation; and (3) an average improvement in goodness of fit with increasing model dimension. The results of this realism-based assessment support the conclusion that the PDMsdE_TRn model provides the most favorable representation of the study location when considered in conjunction with the results of the performance and uncertainty assessments of the previous subsections.

Conclusion

The overall assessment of hydrologic models is of crucial importance when considering multiple alternative representations of the physical system. Evaluation of each of the thirty model structures contained in the modeling suite established in Chapter 4 has been conducted on the basis of performance, uncertainty and realism. The assessment of performance focused on the comparison of several criteria (AIC, BIC, NSE, log-likelihood) considering complexity and performance for both the calibration and validation periods. Quantification of uncertainty centered on an evaluation of

structural uncertainty via the range of predictions associated with all of the structures investigated and parameter uncertainty through the calculation of 95% prediction intervals, including a summary of sharpness and reliability of each. Finally, the collection of models was assessed in their ability to predict snow water equivalents, a variable internal to the simulation of streamflow. Consideration was also given to the selection of the likelihood function through the comparison of two different forms; first assuming model errors were distributed normally, independently and identically and second assuming a heteroscedastic and uncorrelated form. Further examination was given to the model structure identified as having the highest predictive performance.

The performance-based assessment yielded several important results relating to model structural components and predictive performance. In an average sense, predictive performance was found to increase with increasing model dimension. The disaggregation of the pooled models into precipitation semi-distribution groupings revealed a similar trend between complexity and performance. However, when the models were grouped by the primary structural components of soil moisture accounting and runoff routing and snowmelt accounting such relationships were not strong. Further, the TOPMODEL component was shown to be consistently inferior to the bucket model and Probability Distributed Model components, signaling a more restrictive representation of the system brought about by the topographic index. Meanwhile, the temperature index snowmelt accounting approach was inferior to those containing net radiation data, a reflection of the importance of net radiation to the energy balance associated with a snowpack. The optimal model structure was found to be the PDMsdE_TRn.

The uncertainty-based assessment focused primarily on the uncertainty of the parameters and the model predictive uncertainty (described by the statistical variance in the likelihood function) as propagated through the model structure in the calculation of 95% prediction intervals. Two statistics were calculated for every model's prediction interval, the reliability and sharpness. Based on the narrowest mean prediction interval width, the PDMsdE_TRn model was selected as the least uncertain structure for the validation period, however, the reliability for this model was less than 95% suggesting that the prediction interval was too narrow.

Finally, a check on the realism of each structure in their simulation of snow water equivalents was performed. The best overall model in simulating snow water equivalents was the BUCsdAE_TRn structure. Although different from the optimal models of the first two assessment appraisals, this selection was consistent with the trend of more complex models resulting in better predictive performance using the Nash-Sutcliffe Efficiency, as discussed by *Marshall et al.* [2005]. TOPMODEL based structures were further shown to be deficient in their simulation of snow water equivalents compared with the bucket and Probability Distributed models, providing consistency with the results of the assessment based on streamflows.

While the results of this study do not conclusively establish the best model for mountainous, snow-dominated watersheds, they do offer insight into the interplay among model structural components and predictive performance. The overriding conclusions being that (1) precipitation semi-distribution methods did little more than provide another degree of freedom to the model structures; (2) net radiation is an important component in

conceptualization of snowmelt accounting methods; and (3) the TOPMODEL component's rigidity brought on by the inclusion of topographic index classes may inhibit the overall calibration performance but may also provide an amount of physical realism that is beneficial in a forecasting mode.

SIMULATION AND PREDICTION LAB FOR UNCERTAINTY ANALYSIS OF SNOWMELT HYDROLOGY

This chapter intends to introduce a newly developed software package, the Simulation and Prediction Lab for Uncertainty Analysis of Snowmelt Hydrology (SPLASH). SPLASH serves as an integration of all of the tasks associated with predictive hydrologic modeling discussed in previous chapters, including model selection, calibration and assessment.

Introduction

The enduring characteristic of many successful applied research studies is a quality of usability beyond its research scope. In the modeling of precipitation-runoff processes in particular, there has been a trend toward computer-intensive methods that often rely strongly on the application of complex statistical algorithms [e.g., *Bates and Campbell*, 2001; *Vrugt et al.*, 2003b]. While numerous studies praise the merits of such methods [e.g., *Marshall et al.*, 2004], the audience is left with the task of researching, programming and implementing all of the interrelated components associated with such advantageous methodologies.

While the true utility of many emergent approaches to hydrologic modeling is in their application, traditional, well-understood and easily available approaches will logically be the first choice in most operational settings due to the difficulty associated with implementing the cutting-edge [*Mantovan and Todini*, 2006]. The Simulation and Prediction Lab for Uncertainty Analysis of Snowmelt Hydrology (SPLASH, Figure 6.1)

is a fully distributable, standalone graphical user interface (GUI) that aims to provide a simple, easy to use interface. It allows the user to carry out the tasks of model selection, calibration, evaluation and validation under a Bayesian framework that are described in this thesis.

SPLASH was developed using the Graphical User Interface Development Environment (GUIDE) in MATLAB®. GUIDE is an interactive GUI tool that simplifies the processes of layout and programming, where the user selects and places pre-built components such as buttons, axes or text into the layout area. A basic programmatic framework is generated based on the components placed in the layout area that initializes the GUI; however, the developer must specifically program each component to perform the desired actions (running a model, plotting data, exiting the GUI, etc.).

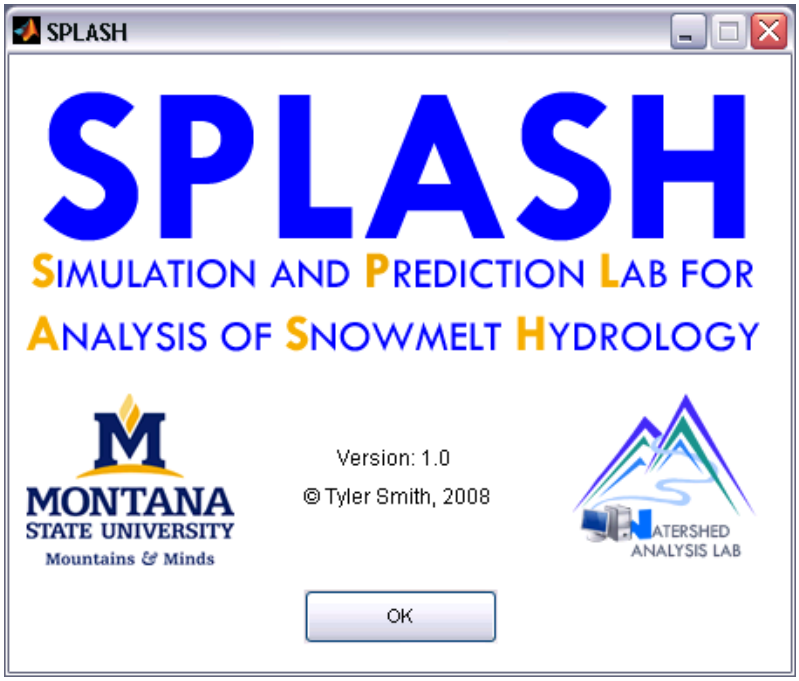


Figure 6.1: SPLASH modeling software.

SPLASH was built as a complete modeling tool, with the capability to perform the varied tasks required by conceptual modeling under uncertainty. This chapter focuses on the characteristics of the SPLASH graphical user interface and its associated benefits. The chapter is divided into the following sections, closely mimicking the actual layout of the SPLASH interface: model selection, model calibration, parameter set evaluation, uncertainty analysis and conclusions.

Model Selection

Selection of the appropriate model for prediction in a precipitation-runoff environment can be difficult, owing to the fact that multiple model structures may acceptably simulate or predict the variable of interest (i.e., streamflow). Such a recognition was made by *Beven* [1993] in defining equifinality as the ability of multiple model structures or parameter sets to result in equally good representations of the system, where the cause of such equifinality can be accredited (at least in part) to the inadequacy of measured data, physical understanding and uniqueness of place [*Beven*, 2000].

In accepting the equifinality hypothesis the modeler is forced to abandon the pursuit for the true 'optimal' solution, instead focusing on finding the best representation(s) of the system given the available data and physical understanding associated with the problem at hand. To that end, a suite of conceptual models ranging in process representation and structural complexity (i.e., number of model parameters) have been included in SPLASH (Figure 6.2). While the construction of modular modeling frameworks like the one featured in the SPLASH is not uncommon [e.g., *Wagener et al.*,

2001], the integration of such a design under a greater Bayesian framework for parameter estimation and uncertainty analysis is a novel approach.

The elemental construction of SPLASH with regard to model structure permits the user to select from a total of 30 unique model structures. The flexibility of the design allows the hydrologist to analyze the watershed to be studied and specify the model based on the intent, perceived dominant influences and available data. Designed for regions influenced heavily by snow-dominated precipitation patterns, the modeling suite offers three competing explanations for the movement of water through the watershed via the

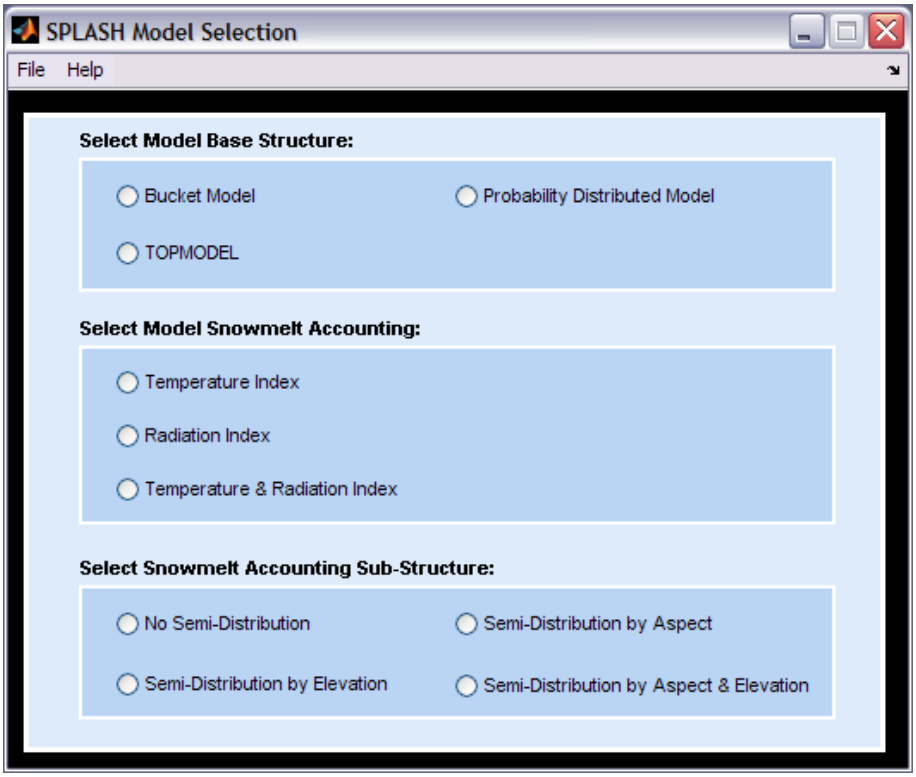


Figure 6.2: Suite of models available in SPLASH.

soil moisture accounting and runoff routing component (specifying the distribution of soil water storage over the watershed), three alternative approaches to snowmelt accounting based on data-sensible techniques and four conceptualizations of watershed semi-

distribution regarding precipitation inputs to the system. The specific details of each model component can be found in Chapter 4 of this manuscript.

Model Calibration

Once a model structure has been selected the primary SPLASH interface is launched, allowing the user to perform the desired modeling task, including model simulation (also known as 'parameter set evaluation'), model calibration or uncertainty analysis. The primary interface of SPLASH (Figure 6.3) was built in a tabbed style, with separate tabs for each of these tasks.

Due to the conceptual nature of the models included in SPLASH, some form of model calibration is required to estimate the proper values for the "effective", immeasurable parameters that each contain [*Sivapalan et al.*, 2003]. SPLASH has the ability to accommodate manual calibration by the evaluation of user-selected parameter sets (in the "Evaluate Parameter Set" tab) and automatic calibration, employing the Delayed Rejection Adaptive Metropolis algorithm [DRAM, *Haario et al.*, 2006] to perform both model calibration (parameter estimation) and uncertainty analysis (in the "Model Calibration" tab).

Recalling results presented in Chapter 3, the DRAM algorithm was designed to take advantage of two recent advancements to the Metropolis algorithm [*Metropolis et al.*, 1953], specifically the adaptive calculation of the covariance implemented in the

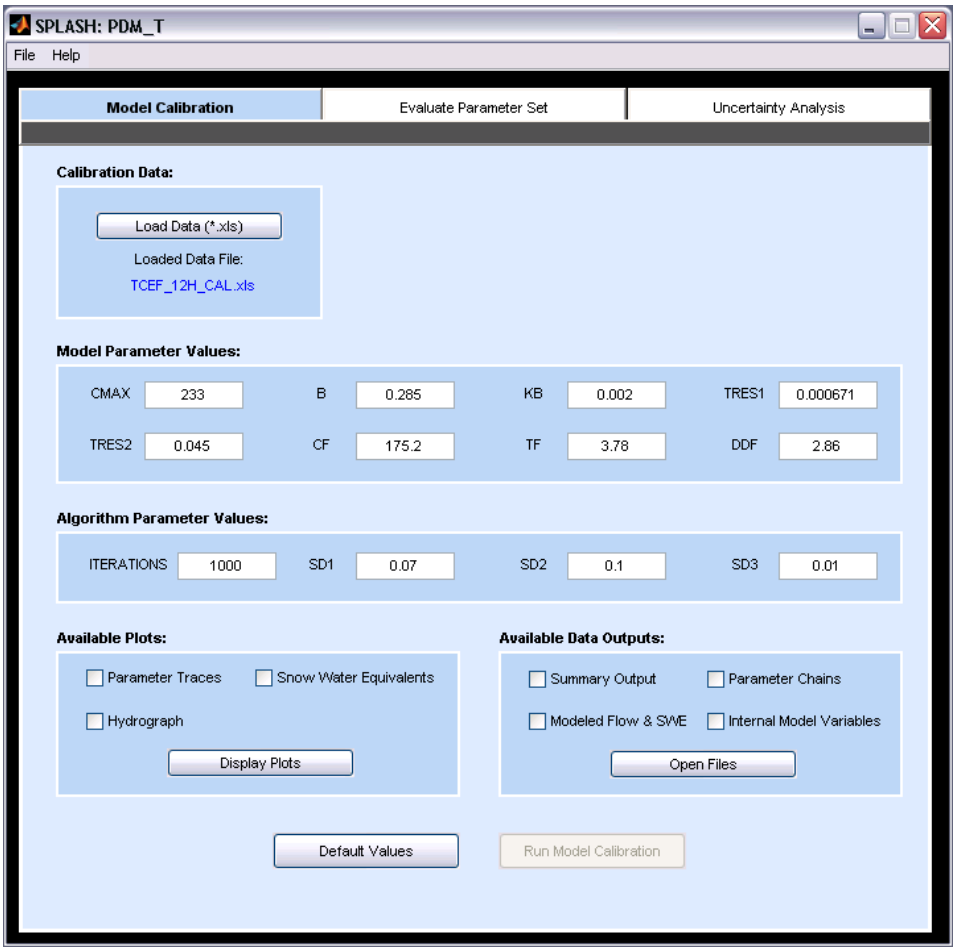


Figure 6.3: Model calibration tab within SPLASH.

Adaptive Metropolis algorithm [*Haario et al.*, 2001] and the improved sampling brought about by the delayed rejection of sampled parameter values exploited in the Delayed Rejection algorithm [*Tierney and Mira*, 1999].

Associated with model selection and calibration is the specification of an appropriate likelihood function. Currently, SPLASH utilizes a single likelihood function for all models, assuming normally, independently and identically distributed data

$$P(Q|\theta) = (2\pi\sigma^2)^{-n/2} \prod \exp\left\{-\frac{[Q_{obs} - Q_p]^2}{2\sigma^2}\right\} \tag{6.1}$$

where $P(Q|\theta)$ is the likelihood function, σ^2 is the variance (a model parameter), Q_{obs} is the observed streamflow data and Q_p is the modeled streamflow. While DRAM works as a pseudo-optimization algorithm, it does not work efficiently to find global optimum parameters over a complex parameter surface. Thus, the model calibration method used by SPLASH works much more efficiently when the initial values for each model parameter are set to reasonable values. The following subsections will provide specific details on each component of SPLASH's model calibration tab.

Calibration Data

The calibration data field is used to load watershed-specific meteorological data into the framework. SPLASH loads data saved in the format associated with Microsoft Excel® (*.xls). Table 6.1 provides the proper specific format required for data to be used in the SPLASH interface, where $Counter$ indexes the rows of data, $Precip$ contains data for both rain and snow events, SWE is observed snow water equivalents, T_{obs} is the observed air temperature, PET is the potential evapotranspiration, Q_{obs} is the observed streamflow and R_n is the net radiation; typical units for each has been included in Table 6.1. The net radiation data is not required for models using the temperature index snowmelt accounting method.

Table 6.1: Sample Calibration Data Format

Counter	**Precip** (mm/time)	**SWE** (mm/time)	**T_{obs}** (^oC)	**PET** (mm/time)	**Q_{obs}** (mm/time)	**R_n** (mm/time)
1	0.00	256.54	-1.21	0.1835	0.0567	0.2904
2	2.54	259.08	-0.26	0.2357	0.0567	0.2922
3	0.00	255.52	-8.57	0.0956	0.0564	0.1812

Watershed Data

Watershed data is that topographic data associated with the physical watershed that must be loaded for all models except those utilizing the *No Semi-Distribution* option available in the model selection window. The input format is again required to be in that of Microsoft Excel® (*.xls). The file should contain data representing cells of the watershed (a grid of x,y data). The elevation data is simply that information contained in a digital elevation model (DEM) of the watershed (in units of meters), while the aspect and topographic index data may be calculated from a DEM. Missing values in the DEM must be represented by "NaN". Both the *Topographic Index Data* and the *Elevation Data* have additional input boxes, the former allows for the selection of the number of topographic index classes to use for the model TOPMODEL (default value is ten) and the latter requires that the elevation of the meteorological data station from which the calibration data originates be defined.

Model Parameter Values

The *Model Parameter Values* panel is where the initial values for each model parameter are input. Pressing the *Default Values* button located at the bottom of the window places initial values for each model parameter for the sample calibration data set provided with SPLASH. For further information about the model parameters themselves, the user is able to refer to the *Model Selection Help* available from the Help Menu.

Algorithm Parameter Values

The *Algorithm Parameter Values* panel is the location in which values are input to control the behavior of the model calibration (DRAM) algorithm. Table 6.2 provides a brief description of each algorithm parameter and its affect on the overall algorithm. The first two scaling factors (SD1 & SD2) should be adjusted only if the algorithm is not jumping at the appropriate rate of approximately 20%-45% [*Gelman and Rubin,* 1992]. The third scaling factor (SD3) should not need to be adjusted. Pressing the *Default Values* button located at the bottom of the window enters values for each algorithm parameter; for further information regarding the algorithm parameters refer to *Smith and Marshall* [2008].

Table 6.2: Description of Algorithm Parameters

Algorithm Parameter	**Description**
ITERATIONS	the number of iterations performed by the algorithm
SD1	an initial scaling factor
SD2	a scaling factor that determines how often the algorithm jumps
SD3	a scaling factor associated with the delayed rejection step of the algorithm

Available Plots

The *Available Plots* panel displays all the available plots associated with the selected model structure (plots become available after the completion of the model calibration). Table 6.3 provides a list of all the possible plots and a brief description of each (examples of each are presented in Appendix B). The calibration is performed based on streamflows, but a visual check on snow water equivalents can help to inform the

Table 6.3: Description of Plots Available Within the Model Calibration Tab of SPLASH

Available Plot	**Description**
Parameter Traces	displays the value of the parameter versus the number of iterations
Hydrograph	displays the observed and predicted streamflows
Snow Water Equivalents	displays the observed and predicted snow water equivalents
Watershed Discretization	displays the appropriate plot(s) showing watershed partitioning (by aspect, elevation)

model selection process, in a multi-criteria sense [*Seibert and McDonnell*, 2002]. Parameter traces show a plot of the parameter values sampled at each iteration in the DRAM algorithm, and are useful in understanding whether or not the algorithm is nearing convergence to a parameter set. If the parameter trace is highly scattered (the parameter value changes greatly across the iterations) it can be assumed that the model is not near convergence.

Available Data Outputs

The *Available Data Outputs* panel displays all available data (appear upon completion of run) associated with the calibration run. Table 6.4 provides a list of all the data outputs and a brief description of each. Note that the data files are saved with a

Table 6.4: Description of Data Outputs Available Within in the Model Calibration Tab of SPLASH

Available Data Output	**Description**
Summary Output	contains summary information (jump rates, max. likelihood, optimum parameter values) (*DRAM_⋯_⋯_Outputs.txt*)
Parameter Chains	contains data used to create Parameter Traces plots (parameter value versus iteration) (*DRAM_⋯_⋯_PARS.xls*)
Modeled Flow & SWE	contains the observed and predicted values for streamflow and snow water equivalents (*DRAM_⋯_⋯_QSWE.xls*)
Internal Model Variables	contains relevant variables internal to the model (useful as a troubleshooting tool to help inform model selection) (*DRAM_⋯_⋯_MODVARS.xls*)

generic name and will be overwritten upon the next run of the selected model structure unless renamed or moved from the current directory.

Parameter Set Evaluation

The parameter set evaluation tool was included in SPLASH (Figure 6.4) to provide the user the option of easily obtaining model predictions given a set of specific parameter values or for manual model calibration. A benefit of this tool is in its ability to

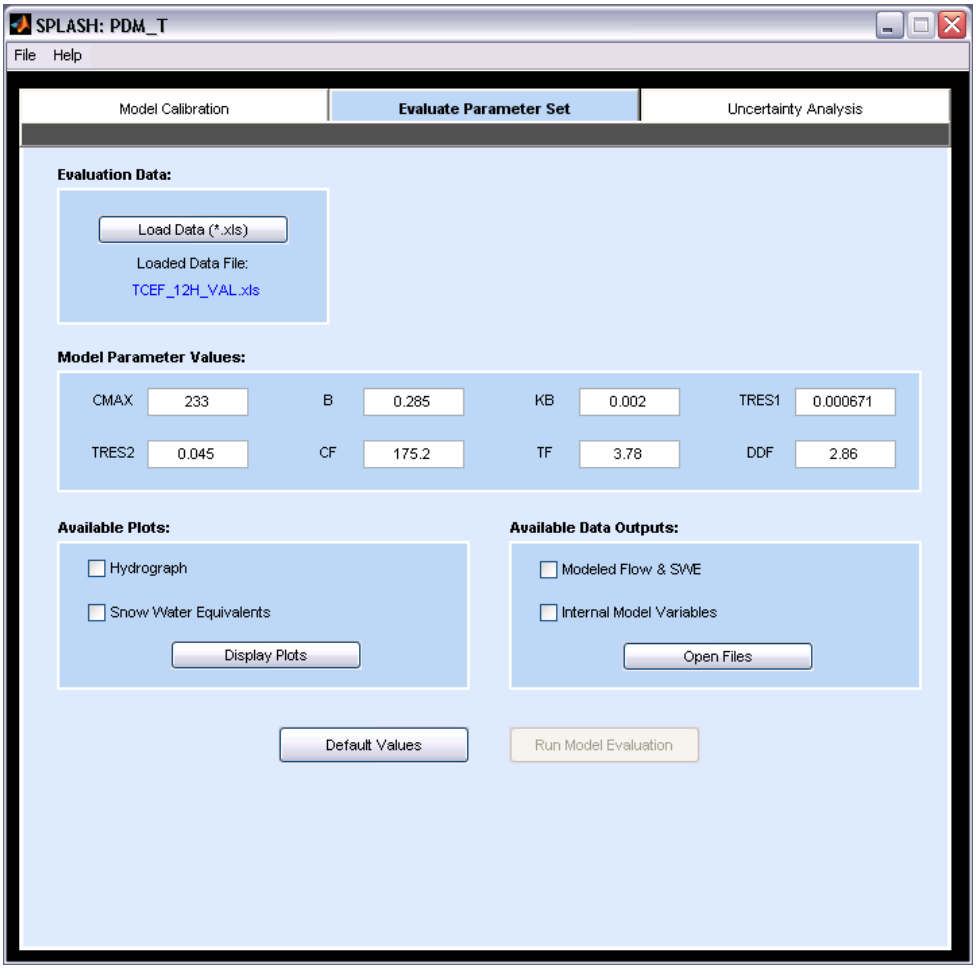


Figure 6.4: Evaluate parameter set tab within SPLASH.

analyze the influence of changing data for a given watershed study area and its calibrated parameter values.

As with the model calibration tab, the parameter set evaluation tab requires the user to load meteorological data (evaluation data) and watershed data (elevation, aspect or topographic index data; provided a semi-distribution option was chosen during model selection) and input values for the model parameters. Because the parameter set evaluation tab only executes the model and does not perform any calibration services, the available plots and data outputs are slightly different than those associated with the model calibration tab; descriptions of the plots (examples of each are presented in Appendix B) and data outputs are given in Table 6.5.

Table 6.5: Description of Plots and Data Outputs Available Within in the Parameter Set Evaluation Tab of SPLASH

Available Plot	**Description**
Hydrograph	displays the observed and predicted streamflows
Snow Water Equivalents	displays the observed and predicted snow water equivalents
Watershed Discretization	displays the appropriate plot(s) showing watershed partitioning
Available Data Output	**Description**
Modeled Flow & SWE	contains the observed and predicted values for streamflow and snow water equivalents (*DRAM_⋯_⋯_QSWE.xls*)
Internal Model Variables	contains relevant variables internal to the model; useful as a troubleshooting tool to help inform model selection (*DRAM_⋯_⋯_MODVARS.xls*)

Uncertainty Analysis

Given that the true worth of any modeled result is in an understanding of its predictions and associated risks, the final component of SPLASH (Figure 6.5) addresses the issue of uncertainty associated with the calibrated parameters and resulting modeled

response. Uncertainty analysis is an important step in the overall assessment of the appropriateness of a model given a specific watershed and its data [*Huard and Mailhot*, 2008]. Because SPLASH is cast in a Bayesian framework and implemented through the use of the Delayed Rejection Adaptive Metropolis algorithm (in the model calibration tab), it becomes possible to analyze the uncertainty associated with the calibrated parameters through the application of Bayesian principles.

In Bayesian statistics, the deterministic hydrologic model is transformed into a probabilistic form [*Samanta et al.*, 2007]; where the Markov chain Monte Carlo (MCMC) simulation employed via the DRAM algorithm samples values of the model parameters. Upon convergence of the parameter estimation (where parameter values sampled in model calibration form a stationary probability distribution), all future draws of the model parameters are assumed to be draws from their respective posterior distributions. Given sufficient time (iterations) in the model calibration phase, the posterior distributions for each model parameter can be characterized as the frequency of occurrence of a parameter value in the post-convergence samples.

SPLASH can perform uncertainty analysis under a split-sample validation testing context, where the model is validated using a data set independent to that used for model calibration. In this context, SPLASH requires the user to input the appropriate validation data (and watershed data if applicable), in addition to parameter data. The parameter data is a data output of the model calibration tab and is a matrix containing the sampled values of each parameter; the uncertainty analysis component assumes that the parameter data

that is loaded is based on a set of converged parameters. SPLASH draws samples from the parameter data to create the posterior distributions for each parameter.

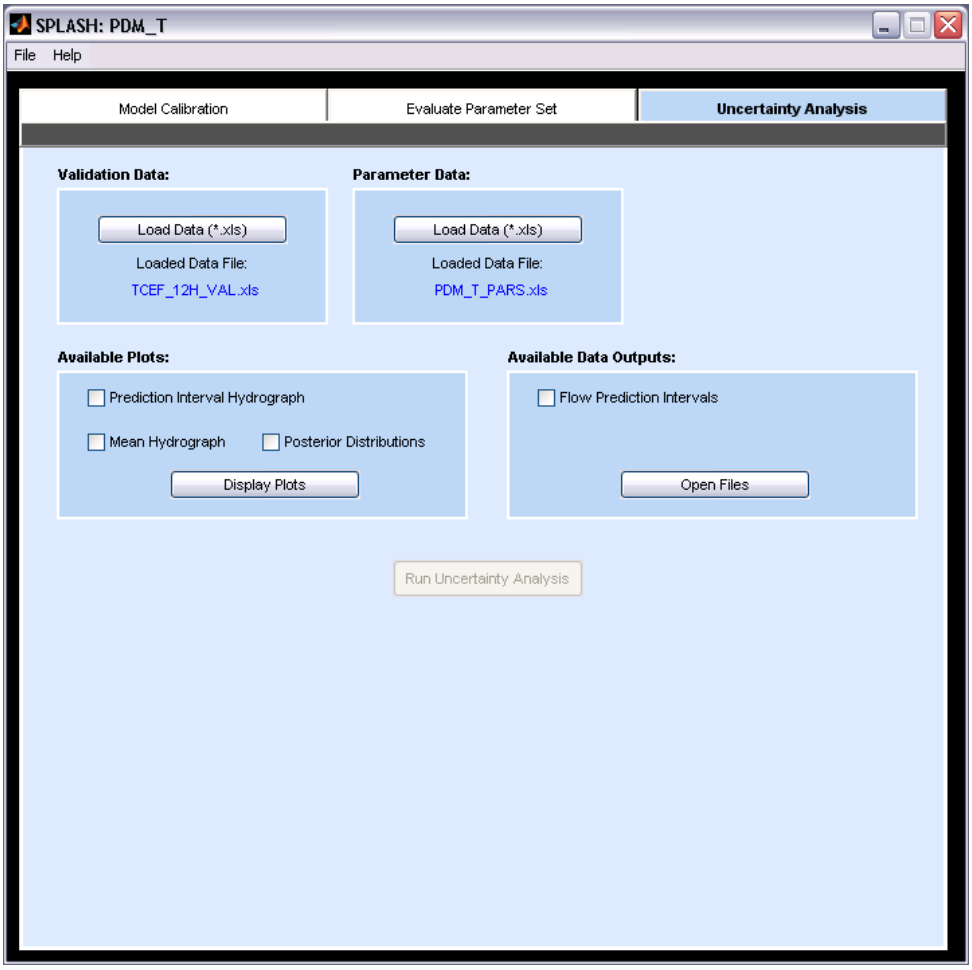


Figure 6.5: Uncertainty analysis tab within SPLASH.

Furthermore, SPLASH calculates the 95% prediction interval for the model simulation on the validation data set

$$PI = Q_p \pm z^* \cdot SE(Q_p) \tag{6.2}$$

where Q_p is the predicted streamflow, z^* is the critical value of Student's t distribution with infinite degrees of freedom (1.96 for a 95% prediction interval) and $SE(Q_p)$ is the standard error (the square root of the variance, a calibrated model parameter). The

prediction interval provides an estimate of the uncertainty associated with the model prediction (as both parameter and model structural uncertainty) when applied to an independent data set (i.e., in a forecasting mode). Finally, the mean predicted response is calculated, based on the mean of the upper and lower prediction interval bounds. Table 6.6 provides a summary of the available plots (examples of each are presented in Appendix B) and data outputs associated with the uncertainty analysis tab. The requirements for all data input are the same as those outlined in the model calibration section, including format and content.

Table 6.6: Description of Plots and Data Outputs Available Within in the Uncertainty Analysis Tab of SPLASH

Available Plot	**Description**
Prediction Interval Hydrograph	displays the 95% prediction interval for streamflow, based on the calibrated parameters and observed streamflow from the validation data set
Mean Hydrograph	displays the mean predicted streamflow response from the prediction interval hydrograph
Posterior Distributions	displays the posterior distribution associated with each model parameter
Available Data Output	**Description**
Flow Prediction Intervals	contains the lower and upper bounds of the 95% prediction interval (*DRAM_⋯_⋯_QPI.xls*)

Conclusion

When faced with a complex modeling problem, like those common in hydrology, the utilization of the most advanced technologies available to answer the modeling question is paramount. However, conventional approaches are often selected due to familiarity and ease of implementation at the expense of the solution. The Simulation and Prediction Lab for Uncertainty Analysis of Snowmelt Hydrology has been designed in an

effort to reduce the lag time associated with the operational implementation of new technologies, such as Bayesian analysis, in computational hydrology.

Constructed under a Bayesian approach to uncertainty analysis and parameter estimation, SPLASH embodies the principles of a conceptual model development strategy, allowing for up to 30 competing alternate models to be implemented. Simplification of several predictive modeling tactics, including model parameter calibration and uncertainty analysis, has been achieved through their implementation in straightforward graphical user interface.

CONCLUSION

Summary of Research

Water resource planning is of critical importance in regions where water availability is limited, water supply is seasonal or there is strong inter-annual variability of water supply. This can be especially true in regions that rely heavily on snowmelt as the primary driver of streamflows, as the majority of precipitation input to the system is stored in the snowpack and released rapidly during the short melt season. Given such importance, the capability to predict watershed streamflows based on precipitation data is a useful tool for water management. Complicating the prediction process is an imprecise knowledge of hydrologic systems and difficulties in quantifying the processes that drive watershed response to precipitation events. In snow-dominated watersheds the system can be considered inherently more complex due to the additional processes associated with the transformation of accumulated snowpack into liquid water.

Historically, the development of models for the purpose of streamflow simulation has been extensively researched. Model development strategies can be broadly classified as either upward (attempt to scale up the detailed description of small scale processes) or downward (attempt to describe the larger scale dominant processes by employing effective parameters). The downward approach to model development has become increasingly popular and advocated due to a growing belief in the practicality of model calibration techniques and skepticism in the ability of models to represent small scale processes.

Because of uncertain mathematical descriptions of the physical hydrologic system, hydrologists are faced with uncertainty in the prediction of streamflow from a watershed. The estimation of the uncertainty of a model prediction, based on both the estimates of model parameters and the model conceptualization itself, allows for the evaluation of the risk associated with the simulated result. One method for performing uncertainty analysis is Bayesian statistical inference. Bayesian inference offers a means of estimating parameter values while also characterizing probability distributions (the posterior probability distributions) for each model parameter.

The primary objectives of this thesis were the comparison of three recently developed Markov chain Monte Carlo algorithms for specifying hydrologic models centered on efficiency and effectiveness; the specification of a suite of conceptual precipitation-runoff models with emphasis on modeling snowmelt; the assessment of this modeling suite under a Bayesian inferential framework focused on the suitability of individual structural components rather than selecting an overall "best" structure via evaluation of performance, uncertainty and realism; and the development of a graphical user interface that integrates the components of model selection, calibration, evaluation and uncertainty analysis under the Bayesian approach. The key findings of each research component are presented in the following sections.

Comparison of Recent Advancements in Markov Chain Monte Carlo Sampling

Markov chain Monte Carlo (MCMC) simulation techniques enable Bayesian inference to be applied in situations where the posterior distribution of a model's

parameters is otherwise analytically intractable, such as for non-linear hydrologic models. In Chapter 3, three newly developed MCMC algorithms were compared based on efficiency and effectiveness in characterizing the posterior distribution for complex parameter surfaces. The MCMC techniques examined were the Adaptive Metropolis [AM, *Haario et al.*, 2001], the Delayed Rejection Adaptive Metropolis [DRAM, *Haario et al.*, 2006] and the Differential Evolution Markov Chain [DE-MC, *ter Braak*, 2006].

Two case studies were presented, (1) a synthetic bimodal mixture of multivariate normal distributions and (2) application to a conceptual precipitation-runoff model. The results of the first case study revealed that the DE-MC was best able to characterize the posterior distribution, but greatly benefited from its generation of multiple, interacting chains absent from the formulations of both the AM and DRAM. In the second case study, the AM and DE-MC algorithms both falsely converged to a local optimum, while the DRAM algorithm was able to find a region of higher posterior density. Additionally, the DRAM algorithm proved to better characterize the tails of the distributions by sampling a wider range of values for all parameters. The results of the second case study were surprising in that the DE-MC algorithm proved to be efficient in the first case study and was not in the second. This indicates the importance of selecting an appropriate MCMC approach dependent on the modeling exercise. The information presented in Chapter 3 has been published in Water Resources Research [*Smith and Marshall*, 2008].

Formation of a Hydrologic Modeling Suite

The specification of a model or suite of models is the basis for every hydrologic modeling study. In Chapter 4 a collection of conceptual precipitation-runoff structures are introduced. The modeling framework was designed in a modular fashion with components for soil moisture accounting and runoff routing, snowmelt accounting and precipitation semi-distribution. The goal of this modeling suite was to investigate the predictive performance of alternate conceptualizations of snowmelt and runoff using a downward approach to model formation.

The individual components allowed for alternate descriptions of each of the processes seen to be important watershed drivers for the study location at Stringer Creek in Central Montana. In all, thirty unique model structures were assembled and served as the foundation for the development of a graphical user interface in Chapter 6.

Evaluation of Alternate Hydrologic Model Structures

The problem of watershed model evaluation has been widely discussed in the literature, with methods commonly focusing on the fit of predicted to observed hydrologic variables, typically streamflow during a validation period [e.g., *Donnelly-Makowecki and Moore*, 1999; *Klemes*, 1986; *Refsgaard and Knudsen*, 1996]. More recently, there has been a movement to analyze the ability of a model to reproduce internal watershed variables [e.g., *Mroczkowski et al.*, 1997; *Seibert and McDonnell*, 2002; *Vache and McDonnell*, 2006]. These methodologies, however, do not directly take into account the inherent uncertainties present in the data, parameters or model structure.

Approaches to characterize this uncertainty as a constraint on the risk associated with simulated predictions have gained increasing attention as model structures requiring parameter calibration have expanded in popularity [e.g., *Beven and Binley*, 1992; *Kuczera and Parent*, 1998].

This thesis focused on the additional task of comparing multiple competing structural representations of the physical system. This type of assessment is well-documented in the literature [e.g., *Chiew et al.*, 1993; *Franchini and Pacciani*, 1991; *Ye et al.*, 1997]; however, the majority of such studies have in the past sought to indentify a single model as superior to the others in the comparison. In this study, the goal of the comparison of multiple structures was to identify not a single structure, but rather the individual building blocks of the structures themselves.

Assessment of each of the thirty model structures, described previously in Chapter 4, was performed with consideration given to the three factors of model performance, uncertainty and realism. In general, it was found that model predictions for both the calibration and the validation data sets were improved with increasing model dimension. However, a closer look at each of the three primary model components (soil moisture accounting and runoff routing, snowmelt accounting and precipitation semi-distribution) revealed that the precipitation semi-distribution component did little to provide realism to the model (as assessed via snowpack accumulation and melt data), rather simply providing an additional parameter to aid in model fitting. The benefit of net radiation data in the calculation of snowmelt was shown both in the prediction of streamflow and of snow water equivalents, with the temperature index approach consistently under-

performing relative to the radiation index and temperature and radiation index methods. Finally, analysis of the soil moisture accounting and runoff routing component revealed a much greater performance in the models containing either the bucket model (describing an effective or spatially homogenous measure of soil water depth) or Probability Distributed Model (describing a spatially heterogeneous measure of soil water depth). The TOPMODEL component (describing potential soil water depth using the topographic index) consistently resulted in poorer simulation of streamflow and of snow water equivalents; however, the optimum TOPMODEL-based structure selected for the calibration period and the validation period for the simulation of streamflow and of snow water equivalents was consistent. The selection of the same model for each of these four cases was unique to this occurrence and suggests robustness in the description of the watershed provided by the TOPMODEL, semi-distributed by aspect and accounting for snowmelt with the temperature and radiation index approach.

Development of a Graphical User Interface

In Chapter 6 a graphical user interface, the Simulation and Prediction Lab for Analysis of Snowmelt Hydrology (SPLASH) was introduced. The development of this software package focused on the integration of multiple tasks necessary for the complete application of a hydrologic model under a Bayesian framework including (1) model selection; (2) model calibration; (3) direct parameter set evaluation; and (4) uncertainty analysis.

SPLASH currently allows for the selection of 30 unique model structures based on a modular framework construction requiring a soil moisture accounting and runoff routing component, a snowmelt accounting component and a precipitation semi-distribution component, as described in Chapter 4. Model calibration and uncertainty analysis are performed via the Delayed Rejection Adaptive Metropolis algorithm, selected as the most efficient MCMC approach in the comparison performed in Chapter 3. The ultimate aspiration being that SPLASH provides an easy-to-use tool for the implementation of progressive and advanced techniques in an operational setting.

REFERENCES CITED

Abbott, M. B., J. C. Bathurst, J. A. Cunge, P. E. O'Connell, and J. Rasmussen (1986), An introduction to the European Hydrological System - Systeme Hydrologique Europeen, "SHE", 1: History and philosophy of a physically-based, distributed modelling system, *Journal of Hydrology*, *87*(1-2), 45-59.

Akaike, H. (1974), A new look at statistical model identification, *IEEE Transactions on Automatic Control*, *19*(6), 716-723.

Allen, R. G., L. S. Pereira, D. Raes, and M. Smith (1998), Crop evapotranspiration - Guidelines for computing crop water requirements, Food and Agriculture Organization of the United Nations, Rome, Italy.

Ambroise, B., J. L. Perrin, and D. Reutenauer (1995), Multicriterion validation of a semidistributed conceptual model of the water cycle in the Fecht Catchment (Vosges Massif, France), *Water Resources Research*, *31*(6), 1467-1481.

Anderson, E. A. (1968), Development and Testing of Snow Pack Energy Balance Equations, *Water Resources Research*, *4*(1), 19-37.

Aronica, G., B. Hankin, and K. Beven (1998), Uncertainty and equifinality in calibrating distributed roughness coefficients in a flood propagation model with limited data, *Advances in Water Resources*, *22*(4), 349-365.

Atkinson, A. C. (1980), A Note on the Generalized Information Criterion for Choice of a Model, *Biometrika*, *67*(2), 413-418.

Atkinson, S. E., M. Sivapalan, R. A. Woods, and N. R. Viney (2003), Dominant physical controls on hourly flow predictions and the role of spatial variability: Mahurange catchment, New Zealand, *Advances in Water Resources*, *26*, 219-235.

Barry, R., M. Prevost, J. Stein, and A. P. Plamondon (1990), Application of a Snow Cover Energy and Mass Balance Model in a Balsam Fir Forest, *Water Resources Research*, *26*(5), 1079-1092.

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 Resources Research*, *37*(4), 937-947.

Bathurst, J. C. (1986), Physically-based distributed modelling of an upland catchment using the Systeme Hydrologique Europeen, *Journal of Hydrology*, *87*(1-2), 79-102.

Bergstrom, S., G. Lindstrom, and A. Pettersson (2002), Multi-variable parameter estimation to increase confidence in hydrological modelling, *Hydrological Processes*, *16*(2), 413-421.

Beven, K. (1989), Changing ideas in hydrology - The case of physically-based models, *Journal of Hydrology*, *105*(1-2), 157-172.

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

Beven, K. (1993), Prophecy, reality and uncertainty in distributed hydrological modelling, *Advances in Water Resources*, *16*(1), 41-51.

Beven, K. (1997), TOPMODEL: A critique, *Hydrological Processes*, *11*(9), 1069-1085.

Beven, K. (2000), Uniqueness of place and process representations in hydrological modelling, *Hydrology and Earth System Sciences*, *4*(2), 203-213.

Beven, K., and J. Freer (2001), Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, *Journal of Hydrology*, *249*(1-4), 11-29.

Beven, K. J., and M. J. Kirkby (1979), A physically based, variable contributing area model of basin hydrology, *Hydrological Sciences Bulletin*, *24*(1), 43-69.

Binley, A. M., K. J. Beven, A. Calver, and L. G. Watts (1991), Changing responses in hydrology: Assessing the uncertainty in physically based model predictions, *Water Resources Research*, *27*(6), 1253-1261.

Bloschl, G., and R. Kirnbauer (1991), Point snowmelt models with different degrees of complexity - Internal processes, *Journal of Hydrology*, *129*(1-4), 127-147.

Bloschl, G., R. Kirnbauer, and D. Gutknecht (1991), Distributed Snowmelt Simulations in an Alpine Catchment 1. Model Evaluation on the Basis of Snow Cover Patterns, *Water Resources Research*, *27*(12), 3171-3179.

Boggild, C. E., C. J. Knudby, M. B. Knudsen, and W. Starzer (1999), Snowmelt and runoff modelling of an Arctic hydrological basin in west Greenland, *Hydrological Processes*, *13*(12-13), 1989-2002.

Borga, M., S. D. Esposti, and D. Norbiato (2006), Influence of errors in radar rainfall estimates on hydrological modeling prediction uncertainty, *Water Resources Research*, *42*, W08409.

Boyle, D. P., H. V. Gupta, and S. Sorooshian (2000), Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods, *Water Resources Research*, *36*(12), 3663-3674.

Braithwaite, R. J., and O. B. Olesen (1990), Response of the energy balance on the margin of the Greenland ice sheet to temperature changes, *Journal of Glaciology*, *36*(123), 217-222.

Braithwaite, R. J. (1995), Positive degree-day factors for ablation on the Greenland ice sheet studied by energy-balance modelling, *Journal of Glaciology*, *41*(137), 153-160.

Braun, L. N., E. Brun, Y. Durand, E. Martin, and P. Tourasse (1994), Simulation of discharge using different methods of meteorological data distribution, basin discretization and snow modelling, *Nordic Hydrology*, *25*, 129-144.

Brooks, S. P. (1998), Markov chain Monte Carlo method and its application, *The Statistician*, *47*(1), 69-100.

Brooks, S. P., and A. Gelman (1998), General methods for monitoring convergence of iterative simulations, *Journal of Computational and Graphical Statistics*, *7*(4), 434-455.

Brooks, S. P., and G. O. Roberts (1998), Convergence assessment techniques for Markov chain Monte Carlo, *Statistics and Computing*, *8*, 319-335.

Brubaker, K., A. Rango, and W. Kustas (1996), Incorporating radiation inputs into the snowmelt runoff model, *Hydrological Processes*, *10*(10), 1329-1343.

Campbell, E. P., D. R. Fox, and B. C. Bates (1999), A Bayesian approach to parameter estimation and pooling in nonlinear flood event models, *Water Resources Research*, *35*(1), 211-220.

Cazorzi, F., and G. D. Fontana (1996), Snowmelt modelling by combining air temperature and a distributed radiation index, *Journal of Hydrology*, *181*(1-4), 169-187.

Charbonneau, R., J. P. Lardeau, and C. Obled (1981), Problems of modelling a high mountainous drainage basin with predominant snow yields, *Hydrological Sciences Bulletin*, *26*(4), 345-361.

Chiang, S., Y. Tachikawa, and K. Takara (2007), Hydrological model performance comparison through uncertainty recognition and quantification, *Hydrological Processes*, *21*(9), 1179-1195.

Chiew, F. H. S., M. J. Stewardson, and T. A. McMahon (1993), Comparison of six rainfall-runoff modelling approaches, *Journal of Hydrology*, *147*(1-4), 1-36.

Chirico, G. B., R. B. Grayson, and A. W. Western (2003), A downward approach to identifying the structure and parameters of a process-based model for a small experimental catchment, *Hydrological Processes*, *17*(11), 2239-2258.

Daly, C., R. P. Neilson, and D. L. Phillips (1994), A statistical-topographic model for mapping climatological precipitation over mountainous terrain, *Journal of Applied Meteorology*, *33*(2), 140-158.

Daly, S. F., R. Davis, E. Ochs, and T. Pangburn (2000), An approach to spatially distributed snow modelling of the Sacramento and San Joaquin basins, California, *Hydrological Processes*, *14*(18), 3257-3271.

Donnelly-Makowecki, L. M., and R. D. Moore (1999), Hierarchical testing of three rainfall-runoff models in small forested catchments, *Journal of Hydrology*, *219*(3-4), 136-152.

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.

Duan, Q., S. Sorooshian, and V. K. Gupta (1994), Optimal use of the SCE-UA global optimization method for calibrating watershed models, *Journal of Hydrology*, *158*(3-4), 265-284.

Ewen, J., G. Donnell, A. Burton, and E. Connell (2006), Errors and uncertainty in physically-based rainfall-runoff modelling of catchment change effects, *Journal of Hydrology*, *330*(3-4), 641-650.

Fenicia, F., H. H. Savenije, P. Matgen, and L. Pfister (2007), A comparison of alternative multiobjective calibration strategies for hydrological modeling, *Water Resources Research*, *43*, W03434.

Fenicia, F., H. H. Savenije, P. Matgen, and L. Pfister (2008), Understanding catchment behavior through stepwise model concept improvement, *Water Resources Research*, *44*, W01402.

Feyen, L., J. A. Vrugt, B. O. Nuallain, J. van der Knijff, and A. De Roo (2007), Parameter optimisation and uncertainty assessment for large-scale streamflow simulation with the LISFLOOD model, *Journal of Hydrology*, *332*(3-4), 276-289.

Fontaine, T. A., T. S. Cruickshank, J. G. Arnold, and R. H. Hotchkiss (2002), Development of a snowfall-snowmelt routine for mountainous terrain for the soil water assessment tool (SWAT), *Journal of Hydrology*, *262*(1-4), 209-223.

Franchini, M., and M. Pacciani (1991), Comparative analysis of several conceptual rainfall-runoff models, *Journal of Hydrology*, *122*(1-4), 161-219.

Freer, J., K. Beven, and B. Ambroise (1996), Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach, *Water Resources Research*, *32*(7), 2161-2173.

Gallagher, M., and J. Doherty (2007), Parameter estimation and uncertainty analysis for a watershed model, *Environmental Modelling and Software*, *22*(7), 1000-1020.

Gan, T. Y., and G. F. Biftu (1996), Automatic calibration of conceptual rainfall-runoff models: Optimization algorithms, catchment conditions, and model structure, *Water Resources Research*, *32*(12), 3513-3524.

Garen, D. C., and D. Marks (2005), Spatially distributed energy balance snowmelt modelling in a mountainous river basin: estimation of meteorological inputs and verification of model results, *Journal of Hydrology*, *315*(1-4), 126-153.

Gelman, A., and D. B. Rubin (1992), Inference from iterative simulation using multiple sequences, *Statistical Science*, *7*(4), 457-472.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2004), *Bayesian Data Analysis*, 2nd ed., Chapman & Hall/CRC, New York.

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter (Eds.) (1996), *Markov Chain Monte Carlo in Practice*, Chapman & Hall, London.

Groisman, P. Y., E. L. Peck, and R. G. Quayle (1999), Intercomparison of Recording and Standard Nonrecording U.S. Gauges, *Journal of Atmospheric and Oceanic Technology*, *16*(5), 602-609.

Guntner, A., S. Uhlenbrook, J. Seibert, and C. Leibundgut (1999), Multi-criterial validation of TOPMODEL in a mountainous catchment, *Hydrological Processes*, *13*, 1603-1620.

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

Haario, H., E. Saksman, and J. Tamminen (2005), Componentwise adaptation for high dimensional MCMC, *Computational Statistics*, *20*, 265-273.

Haario, H., M. Laine, A. Mira, and E. Saksman (2006), DRAM: Efficient adaptive MCMC, *Statistics and Computing*, *16*(4), 339-354.

Hartman, M. D., J. S. Baron, R. B. Lammers, D. W. Cline, L. E. Band, G. E. Liston, and C. Tague (1999), Simulations of snow distribution and hydrology in a mountain basin, *Water Resources Research*, *35*(5), 1587-1603.

Hauer, F. R., J. A. Stanford, and M. S. Lorang (2007), Pattern and process in northern rocky mountain headwaters: ecological linkages in the headwaters of the crown of the continent, *Journal of the American Water Resources Association*, *43*(1), 104-117.

Hock, R. (1999), A distributed temperature-index ice- and snowmelt model including potential direct solar radiation, *Journal of Glaciology*, *45*(149), 101-111.

Hock, R. (2003), Temperature index melt modelling in mountain areas, *Journal of Hydrology*, *282*(1-4), 104-115.

Hogue, T. S., S. Sorooshian, H. Gupta, A. Holz, and D. Braatz (2000), A multistep automatic calibration scheme for river forecasting models, *Journal of Hydrometeorology*, *1*(6), 524-542.

Hooper, R. P., A. Stone, N. Christophersen, E. de Grosbois, and H. M. Seip (1988), Assessing the Birkenes model of stream acidification using a mulitsignal calibration methodology, *Water Resources Research*, *24*, 1308-1316.

Hornberger, G. M., K. J. Beven, B. J. Cosby, and D. E. Sappington (1985), Shenandoah watershed study: Calibration of a topography-based, variable contributing area hydrological model to a small forested catchment, *Water Resources Research*, *21*, 1841-1859.

Huard, D., and A. Mailhot (2008), Calibration of hydrological model GR2M using Bayesian uncertainty analysis, *Water Resources Research*, *44*, W02424.

Jacquin, A. P., and A. Y. Shamseldin (2007), Development of a possibilistic method for the evaluation of predictive uncertainty in rainfall-runoff modeling, *Water Resources Research*, *43*, W04425.

Jakeman, A. J., and G. M. Hornberger (1993), How much complexity is warranted in a rainfall-runoff model?, *Water Resources Research*, *29*(8), 2637-2649.

Janssen, P. H. M., and P. S. C. Heuberger (1995), Calibration of process-oriented models, *Ecological Modelling*, *83*(1-2), 55-66.

Jencso, K. G., B. L. McGlynn, M. N. Gooseff, S. M. Wondzell, and K. E. Bencala (2008, In Review), Hydrologic connectivity between landscapes and streams: Transferring reach and plot scale understanding to the catchment scale, *Water Resources Research*.

Jothityangkoon, C., M. Sivapalan, and D. L. Farmer (2001), Process controls of water balance variability in a large semi-arid catchment: downward approach to hydrological model development, *Journal of Hydrology*, *254*(1-4), 174-198.

Kashyap, R. L. (1982), Optimal Choice of AR and MA Parts in Autoregressive Moving Average, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *4*(2), 99-104.

Kavetski, D., G. Kuczera, and S. W. Franks (2006), Bayesian analysis of input uncertainty in hydrological modeling: Theory, *Water Resources Research*, *42*, W03407.

Kirkby, M. J. (1975), Hydrograph modelling strategies, in *Process in Physical and Human Geography*, edited by R. Peel, et al., pp. 69-90, Heinemann, London.

Klemes, V. (1983), Conceptualization and scale in hydrology, *Journal of Hydrology*, *65*, 1-23.

Klemes, V. (1986), Operational testing of hydrological simulation models, *Hydrological Sciences Journal*, *31*(1), 13-24.

Kou, S. C., Q. Zhou, and W. H. Wong (2006), Equi-energy sampler with applications in statistical inference and statistical mechanics, *Annals of Statistics*, *34*(4), 1581-1619.

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

Kuczera, G. (1997), Efficient subspace probabilistic parameter optimization for catchment models, *Water Resources Research*, *33*(1), 177-185.

Kuczera, G., and M. Mroczkowski (1998), Assessment of hydrologic parameter uncertainty and the worth of multiresponse data, *Water Resources Research*, *34*(6), 1481-1489.

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.

Kustas, W. P., A. Rango, and R. Uijlenhoet (1994), A simple energy budget algorithm for the snowmelt runoff model, *Water Resources Research*, *30*(5), 1515-1527.

Laskey, K. B., and J. W. Myers (2003), Population Markov Chain Monte Carlo, *Machine Learning*, *50*(1-2), 175-196.

Lindstrom, G., B. Johansson, M. Persson, M. Gardelin, and S. Bergstrom (1997), Development and test of the distributed HBV-96 hydrological model, *Journal of Hydrology*, *201*(1-4), 272-288.

Littlewood, I. G., B. F. W. Croke, A. J. Jakeman, and M. Sivapalan (2003), The role of 'top-down' modelling for Prediction in Ungauged Basins (PUB), *Hydrological Processes*, *17*(8), 1673-1679.

Loague, K. M., and R. A. Freeze (1985), A comparison of rainfall-runoff modelling techniques on small upland catchments, *Water Resources Research*, *21*, 229-248.

Luce, C. H., and T. W. Cundy (1994), Parameter identification for a runoff model for forest roads, *Water Resources Research*, *30*(4), 1057-1069.

Madsen, H. (2000), Automatic calibration of a conceptual rainfall-runoff model using multiple objectives, *Journal of Hydrology*, *235*(3-4), 276-288.

Madsen, H., G. Wilson, and H. C. Ammentorp (2002), Comparison of different automated strategies for calibration of rainfall-runoff models, *Journal of Hydrology*, *261*(1-4), 48-59.

Manabe, S. (1969), Climate and the Ocean Circulation: 1. The Atmospheric Circulation and the Hydrology of the Earth's Surface, *Monthly Weather Review*, *97*(11), 739-774.

Mantovan, P., and E. Todini (2006), Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology, *Journal of Hydrology*, *330*(1-2), 368-381.

Marks, D., and J. Dozier (1992), Climate and Energy Exchange at the Snow Surface in the Alpine Region of the Sierra Nevada 2. Snow Cover Energy Balance, *Water Resources Research*, *28*(11), 3043-3054.

Marks, D., J. Domingo, D. Susong, T. Link, and D. Garen (1999), A spatially distributed energy balance snowmelt model for application in mountain basins, *Hydrological Processes*, *13*(12-13), 1935-1959.

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*, W02501.

Marshall, L., D. Nott, and A. Sharma (2005), Hydrological model selection: A Bayesian alternative, *Water Resources Research*, *41*, W10422.

Martinec, J., and A. Rango (1986), Parameter values for snowmelt runoff modelling, *Journal of Hydrology*, *84*(3-4), 197-219.

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

Moore, R. J. (1985), The probability-distributed principle and runoff production at point and basin scales, *Hydrological Sciences*, *30*(2), 273-297.

Moore, R. J. (2007), The PDM rainfall-runoff model, *Hydrology and Earth System Sciences*, *11*(1), 483-499.

Mroczkowski, M., G. P. Raper, and G. Kuczera (1997), The quest for more powerful validation of conceptual catchment models, *Water Resources Research*, *33*(10), 2325-2335.

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

Neuman, S. P. (2003), Maximum likelihood Bayesian averaging of uncertain model predictions, *Stochastic Environmental Research and Risk Assessment*, *17*, 291-305.

Ohmura, A. (2001), Physical basis for the temperature-based melt-index method, *Journal of Applied Meteorology*, *40*(4), 753-763.

Peskun, P. H. (1973), Optimum Monte-Carlo sampling using Markov chains, *Biometrika*, *60*, 607-612.

Qian, S. S., C. A. Stow, and M. E. Borsuk (2003), On Monte Carlo methods for Bayesian inference, *Ecological Modelling*, *159*(2-3), 269-277.

Rango, A., and J. Martinec (1995), Revisiting the degree-day method for snowmelt computations, *Journal of the American Water Resources Association*, *31*(4), 657-669.

Refsgaard, J. C., and J. Knudsen (1996), Operational validation and intercomparison of different types of hydrological models, *Water Resources Research*, *32*(7), 2189-2202.

Renard, B., V. Garreta, and M. Lang (2006), An application of Bayesian analysis and Markov chain Monte Carlo methods to the estimation of a regional trend in annual maxima, *Water Resources Research*, *42*, W12422.

Samanta, S., D. S. Mackay, M. K. Clayton, E. L. Kruger, and B. E. Ewers (2007), Bayesian analysis for uncertainty estimation of a canopy transpiration model, *Water Resources Research*, *43*, W04424.

Schwarz, G. (1978), Estimating the Dimension of a Model, *The Annals of Statistics*, *6*(2), 461-464.

Seibert, J. (2000), Multi-criteria calibration of a conceptual runoff model using a genetic algorithm, *Hydrology and Earth System Sciences*, *4*(2), 215-224.

Seibert, J., and J. J. McDonnell (2002), On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration, *Water Resources Research*, *38*(11), 1241-1255.

Sivapalan, M., G. Bloschl, L. Zhang, and R. Vertessy (2003), Downward approach to hydrological prediction, *Hydrological Processes*, *17*(11), 2101-2111.

Smith, M. B., D.-J. Seo, V. I. Koren, S. M. Reed, Z. Zhang, Q. Duan, F. Moreda, and S. Cong (2004), The distributed model intercomparison project (DMIP): motivation and experiment design, *Journal of Hydrology*, *298*(1-4), 4-26.

Smith, T. J., and L. A. Marshall (2008), Bayesian methods in hydrology: A study of recent advancements in Markov chain Monte Carlo techniques, *Water Resources Research*, *44*, W00B05.

Son, K., and M. Sivapalan (2007), Improving model structure and reducing parameter uncertainty in conceptual water balance models through the use of auxiliary data, *Water Resources Research*, *43*, W01415.

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), 430-442.

Storn, R., and K. Price (1997), Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces, *Journal of Global Optimization*, *11*(4), 341-359.

ter Braak, C. J. F. (2006), A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces, *Statistics and Computing*, *16*(3), 239-249.

Thiemann, M., M. Trosset, H. V. Gupta, and S. Sorooshian (2001), Bayesian recursive parameter estimation for hydrologic models, *Water Resources Research*, *37*(10), 2521-2535.

Thyer, M., G. Kuczera, and B. C. Bates (1999), Probabilistic optimization for conceptual rainfall-runoff models: A comparison of the shuffled complex evolution and simulated annealing algorithms, *Water Resources Research*, *35*(3), 767-773.

Tierney, L. (1994), Markov chains for exploring posterior distributions, *The Annals of Statistics*, *22*(4), 1701-1728.

Tierney, L. (1998), A note on Metropolis-Hastings kernels for general state spaces, *The Annals of Applied Probability*, *8*, 1-9.

Tierney, L., and A. Mira (1999), Some adaptive Monte Carlo methods for Bayesian inference, *Statistics in Medicine*, *18*, 2507-2515.

Tripp, D. R., and J. D. Niemann (2008), Evaluating the parameter identifiability and structural validity of a probability-distributed model for soil moisture, *Journal of Hydrology*, *353*(1-2), 93-108.

Uhlenbrook, S., J. Seibert, C. Leibundgut, and A. Rodhe (1999), Prediction uncertainty of conceptual rainfall-runoff models caused by problems in identifying model parameters and structure, *Hydrological Sciences*, *44*(5), 779-797.

Uhlenbrook, S., and C. Leibundgut (2002), Process-oriented catchment modelling and multiple-response validation, *Hydrological Processes*, *16*, 423-440.

Uhlenbrook, S., and A. Sieber (2005), On the value of experimental data to reduce the prediction uncertainty of a process-oriented catchment model, *Environmental Modelling and Software*, *20*(1), 19-32.

Vache, K. B., and J. J. McDonnell (2006), A process-based rejectionist framework for evaluating catchment runoff model structure, *Water Resources Research*, *42*, W02409.

Vrugt, J. A., H. V. Gupta, L. A. Bastidas, and W. Bouten (2003a), Effective and efficient algorithm for multiobjective optimization of hydrologic models, *Water Resources Research*, *39*(8), 1214-1233.

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

Wagener, T., D. P. Boyle, M. J. Lees, H. S. Wheater, H. V. Gupta, and S. Sorooshian (2001), A framework for development and application of hydrological models, *Hydrology and Earth System Sciences*, *5*(1), 13-26.

Wagener, T. (2003), Evaluation of catchment models, *Hydrological Processes*, *17*, 3375-3378.

Wagener, T., and J. Kollat (2007), Numerical and visual evaluation of hydrological and environmental models using the Monte Carlo analysis toolbox, *Environmental Modelling and Software*, *22*(7), 1021-1033.

Williams, M. W., D. Cline, M. Hartman, and T. Bardsley (1999), Data for snowmelt model development, calibration, and verification at an alpine site, Colorado Front Range, *Water Resources Research*, *35*(10), 3205-3209.

WMO (1986), *Intercomparison of models for snowmelt runoff*, Geneva.

Woods, S. W., R. Ahl, J. Sappington, and W. McCaughey (2006), Snow accumulation in thinned lodgepole pine stands, Montana, USA, *Forest Ecology and Management*, *235*, 202-211.

Yadav, M., T. Wagener, and H. Gupta (2007), Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins, *Advances in Water Resources*, *30*(8), 1756-1774.

Yapo, P. O., H. V. Gupta, and S. Sorooshian (1996), Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data, *Journal of Hydrology*, *181*(1-4), 23-48.

Yapo, P. O., H. V. Gupta, and S. Sorooshian (1998), Multi-objective global optimization for hydrologic models, *Journal of Hydrology*, *204*(1-4), 83-97.

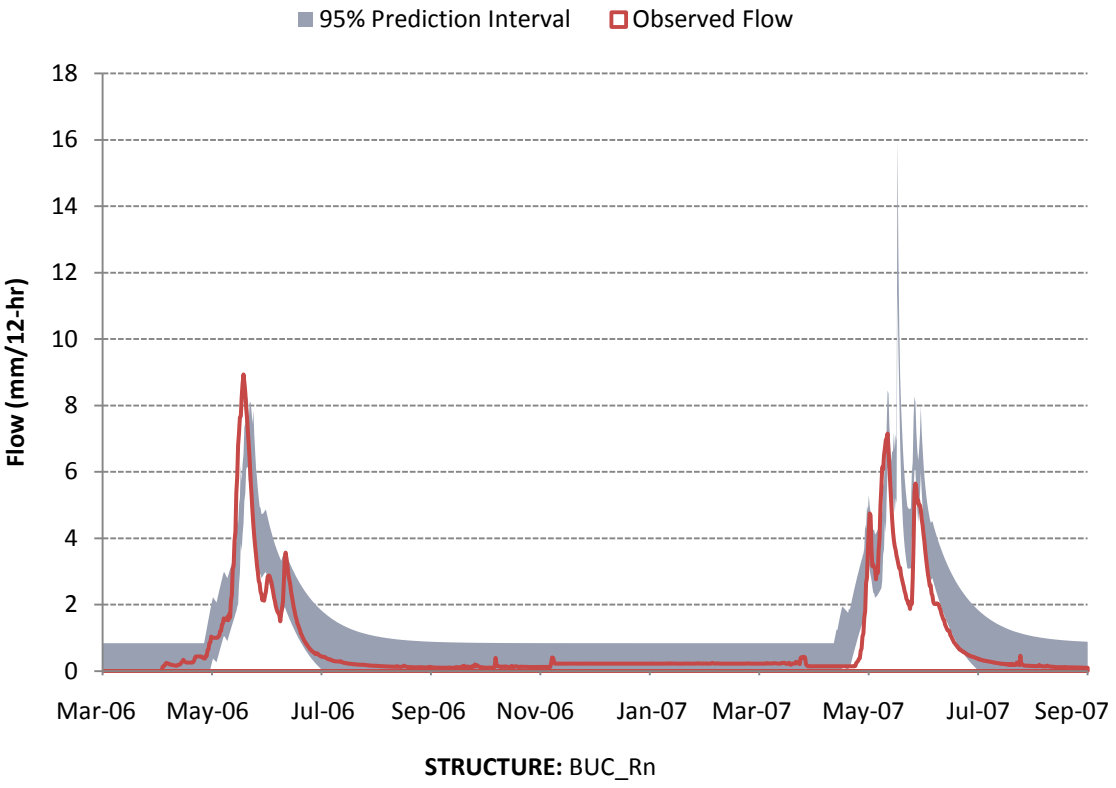
Ye, M., B. C. Bates, N. R. Viney, and M. Sivapalan (1997), Performance of conceptual rainfall-runoff models in low-yielding ephemeral catchments, *Water Resources Research*, *33*(1), 153-166.

Ye, M., S. P. Neuman, and P. D. Meyer (2004), Maximum likelihood Bayesian averaging of spatial variability models in unsaturated fractured tuff, *Water Resources Research*, *40*, W05113.

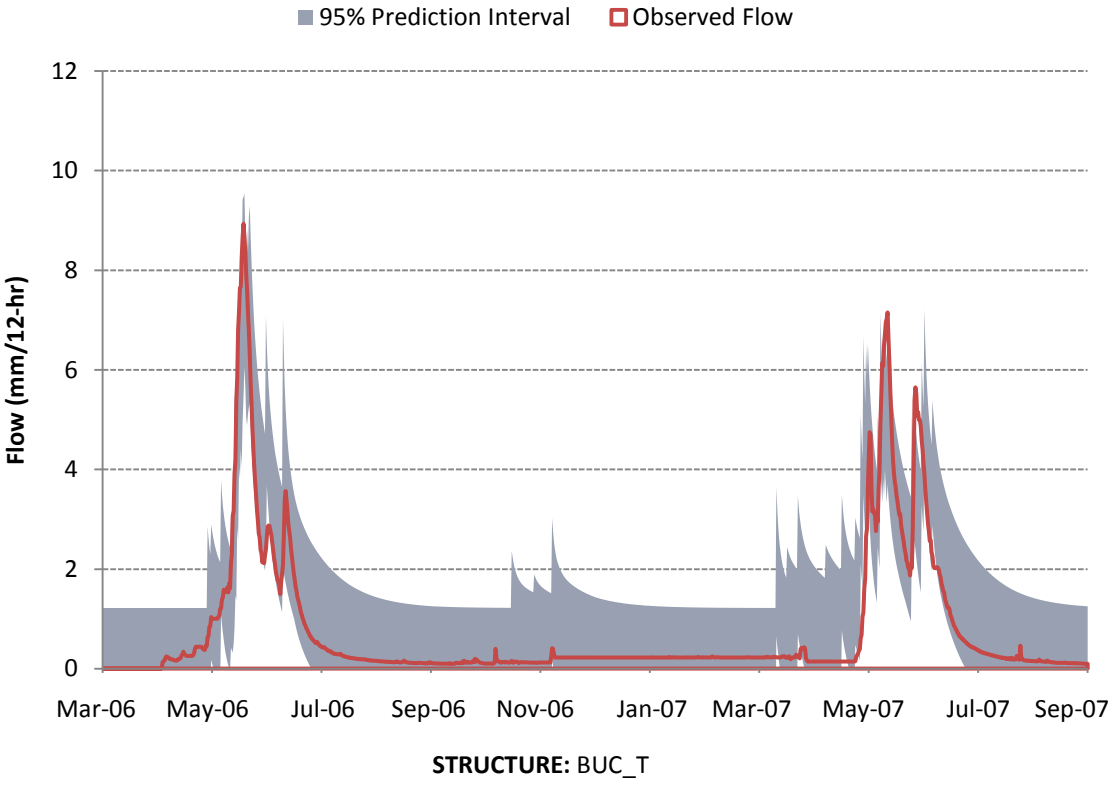
APPENDICES

APPENDIX A

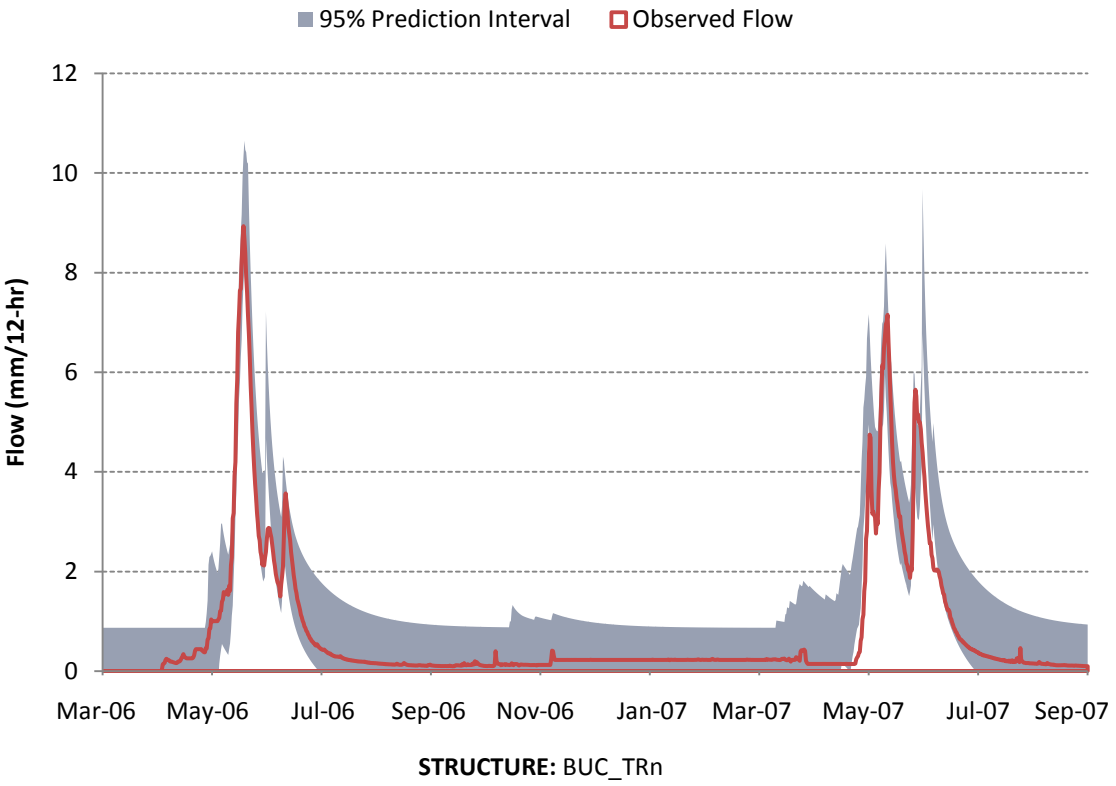
95% PREDICTION INTERVALS



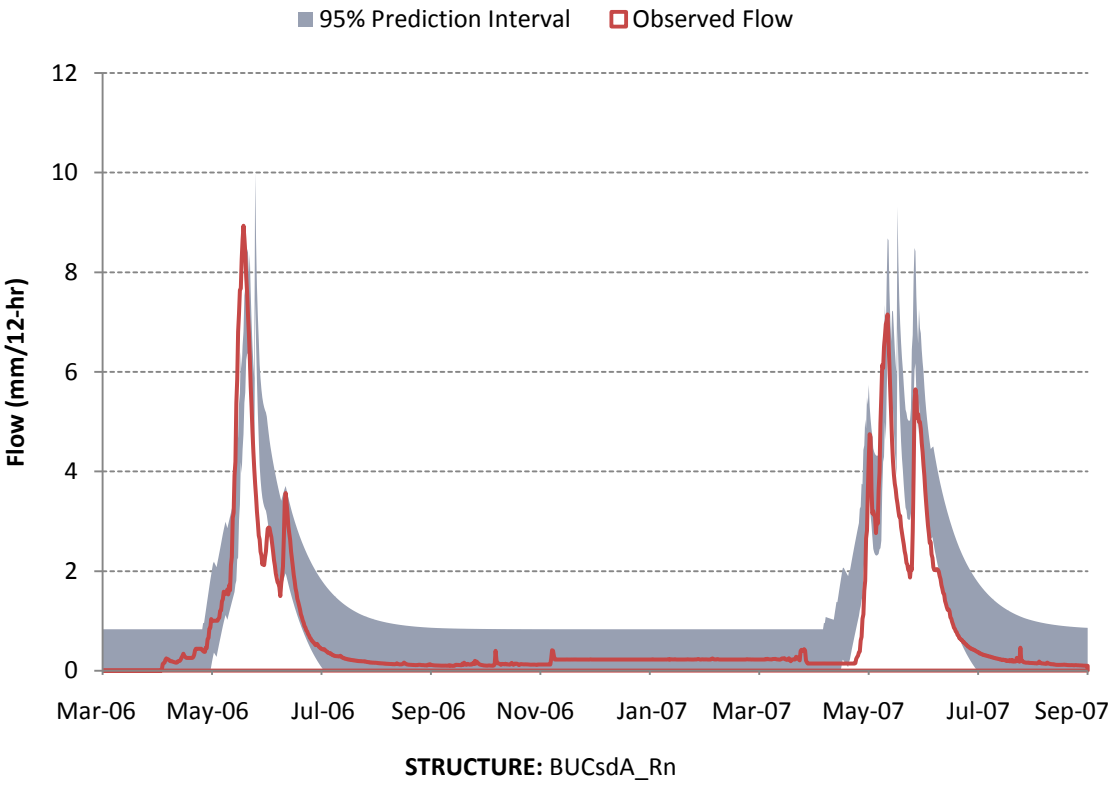
STRUCTURE: BUC_Rn



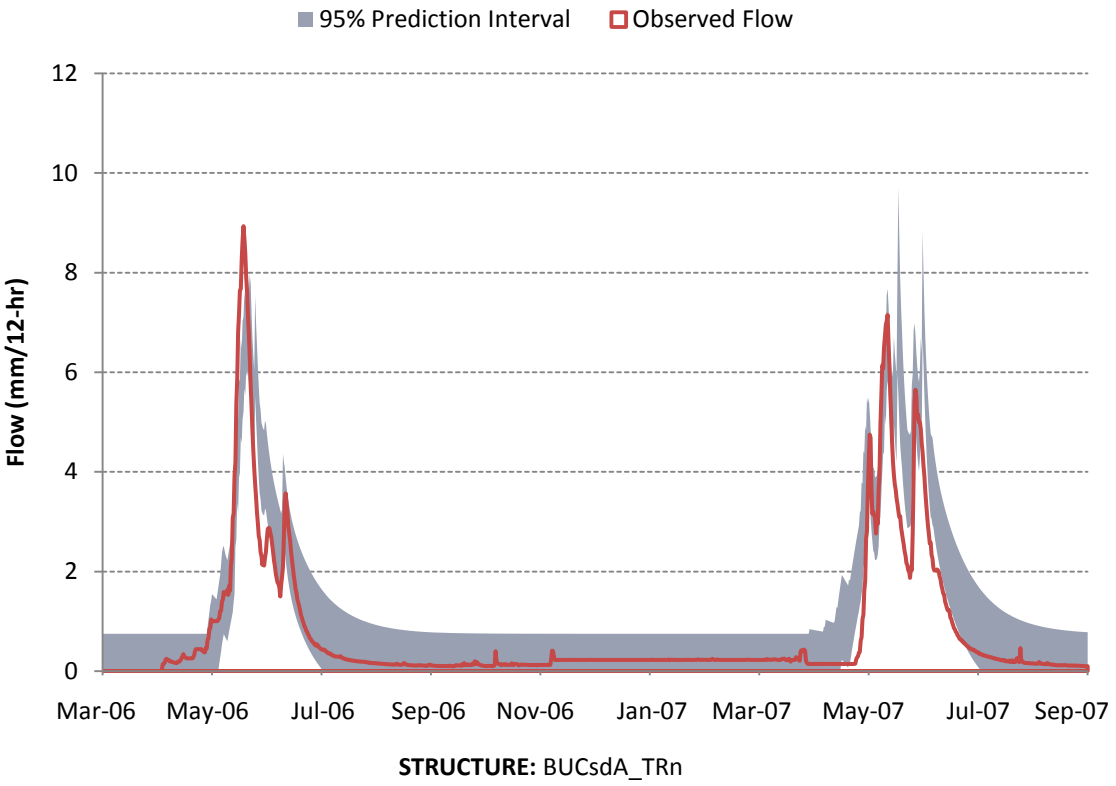
STRUCTURE: BUC_T



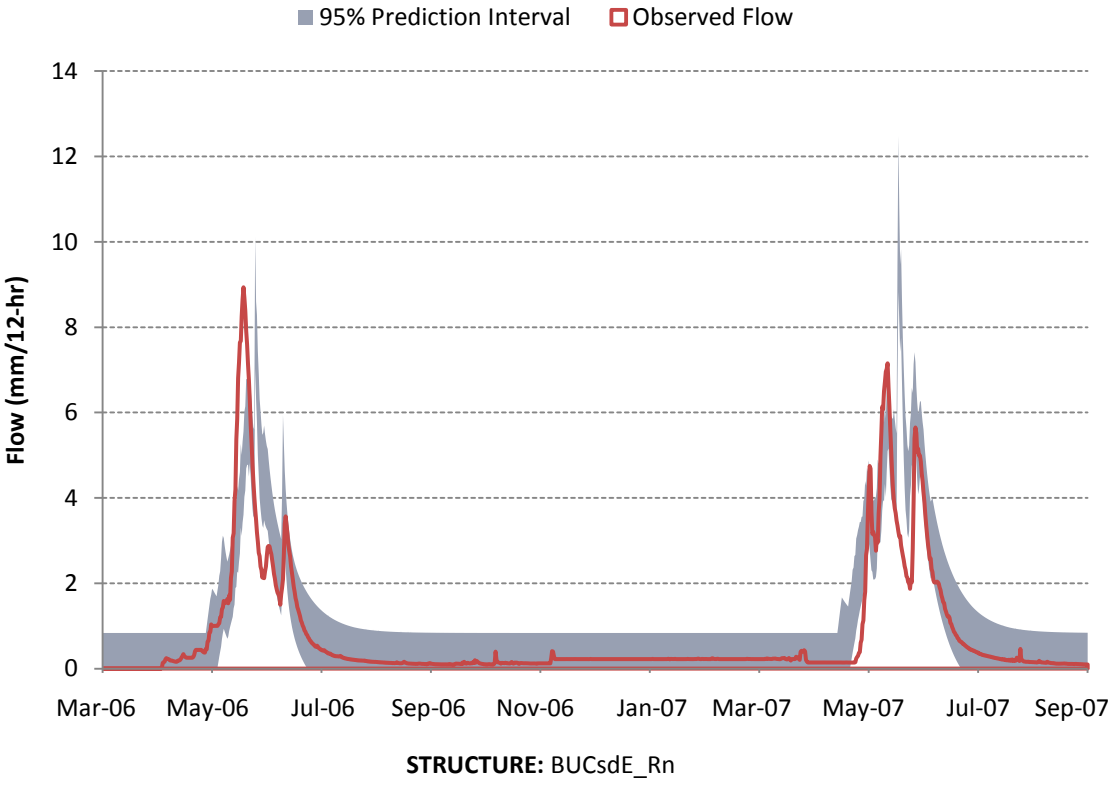
STRUCTURE: BUC_TRn



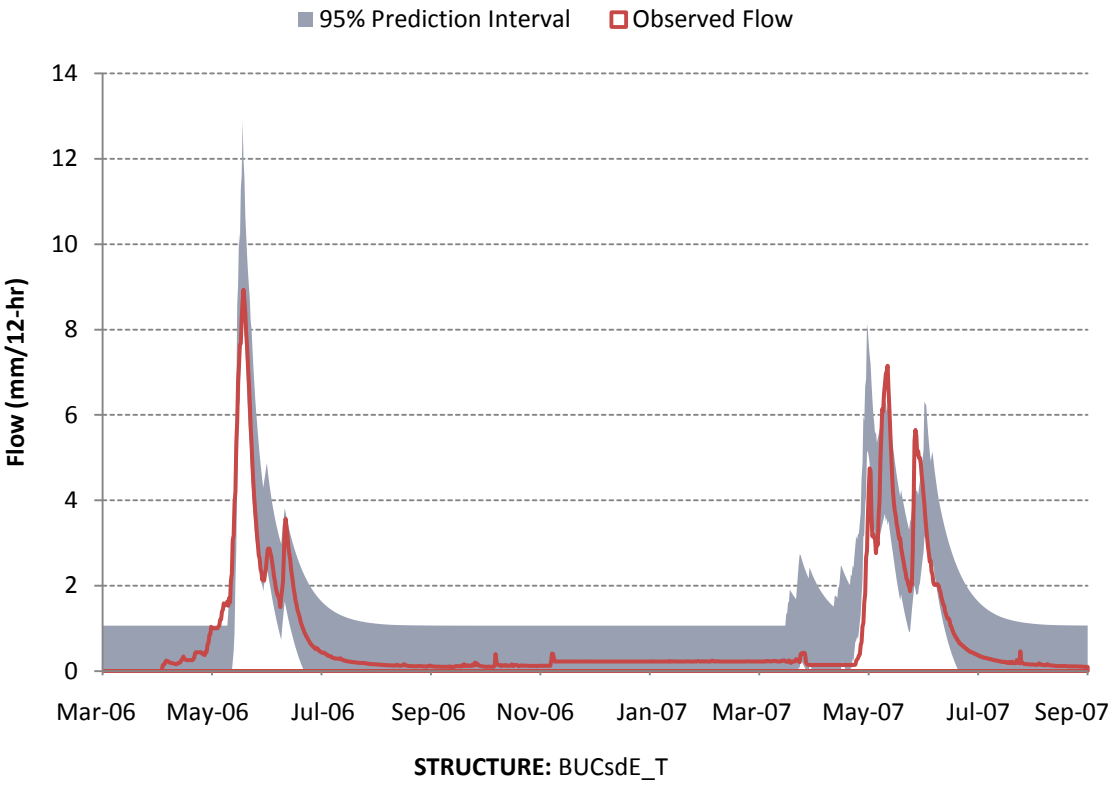
STRUCTURE: BUCsdA_Rn



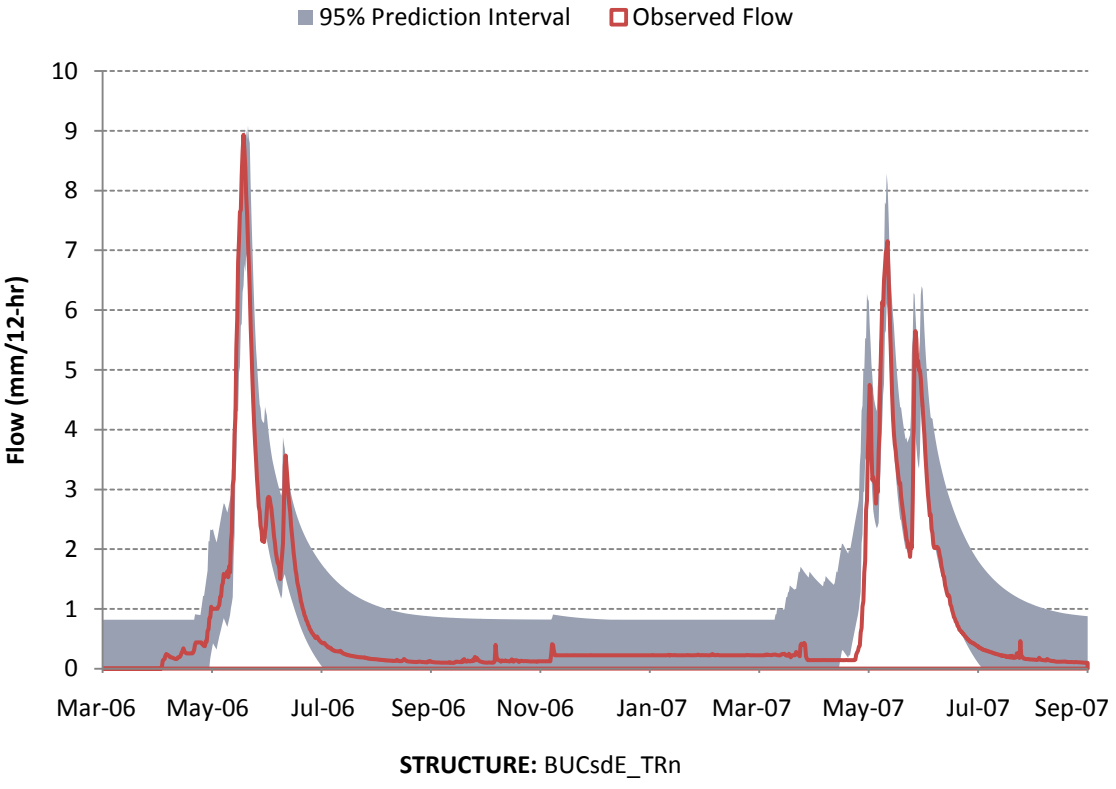
STRUCTURE: BUCsdA_TRn



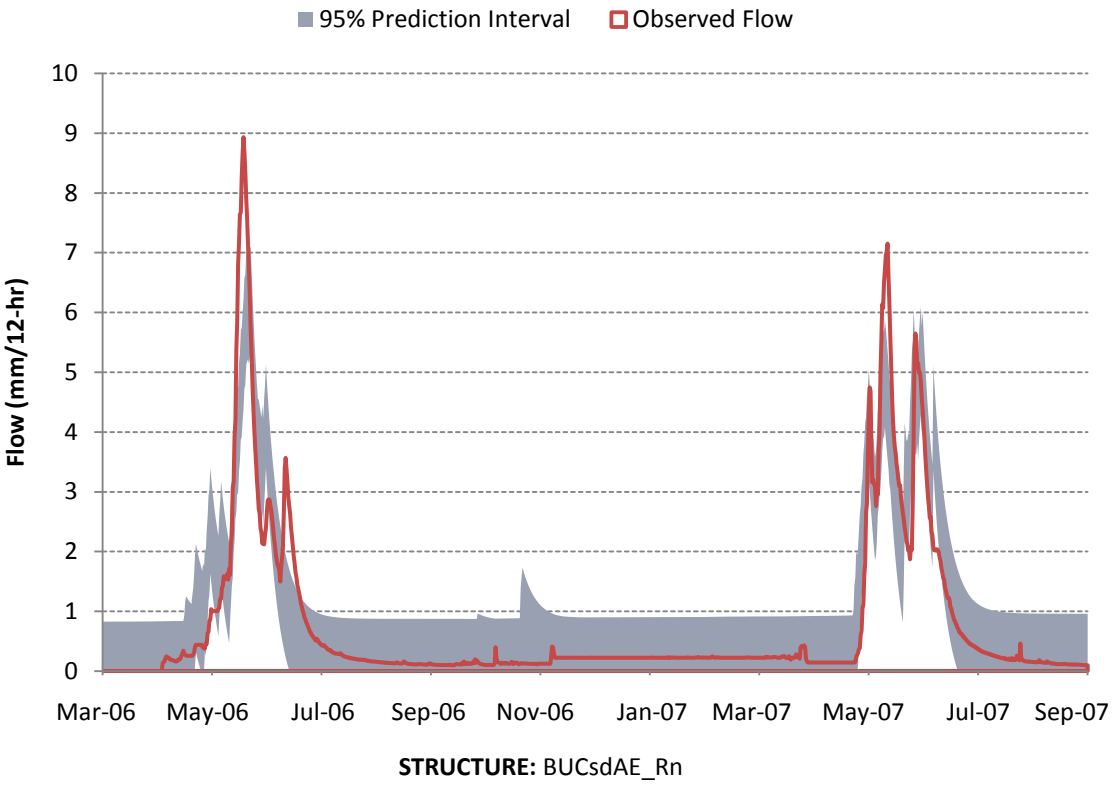
STRUCTURE: BUCsdE_Rn



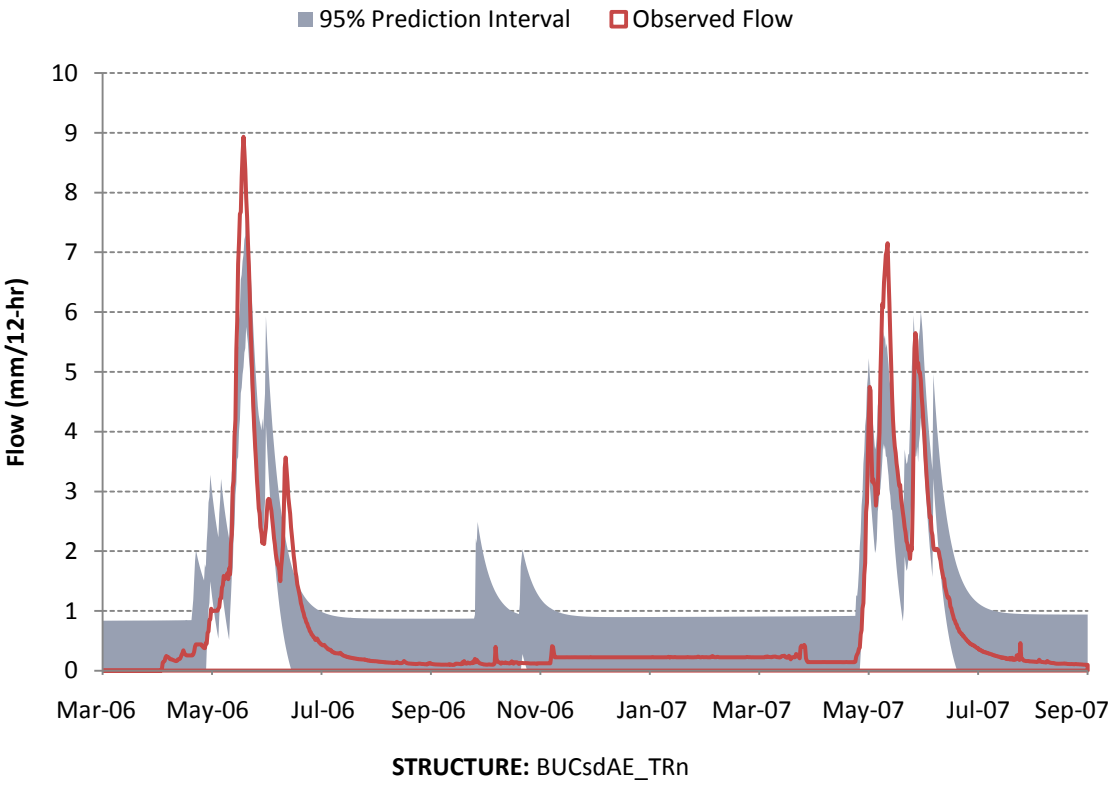
STRUCTURE: BUCsdE_T



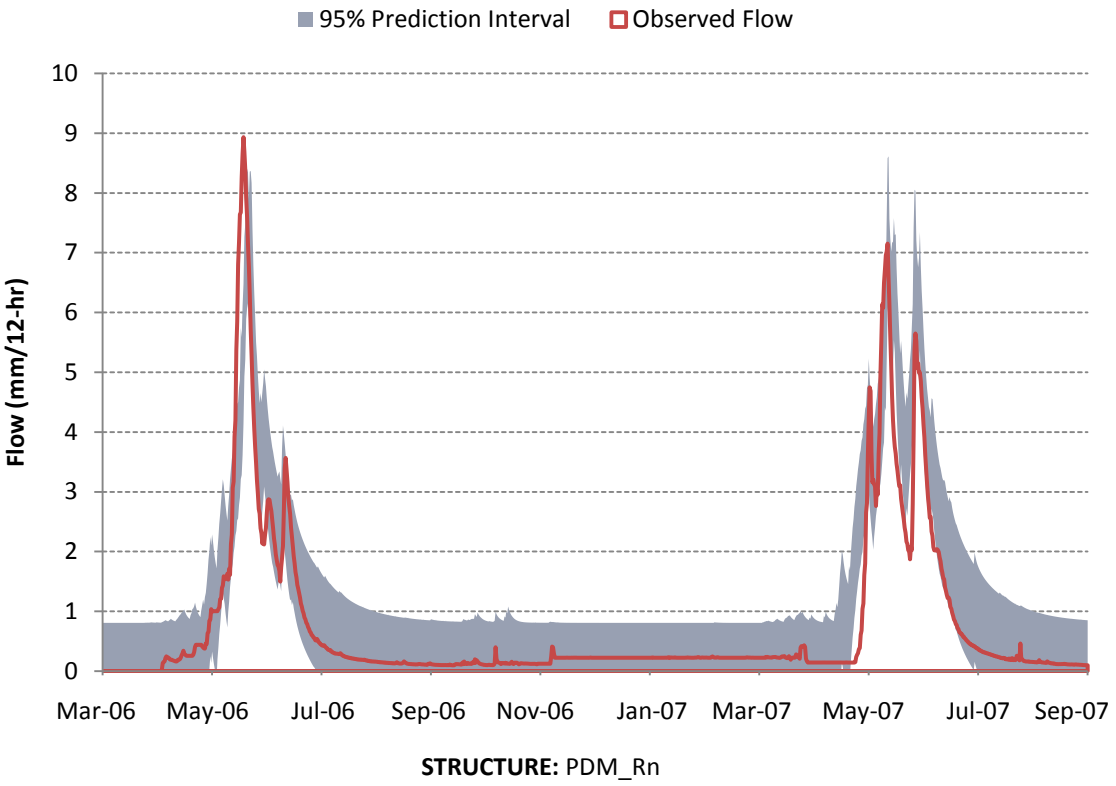
STRUCTURE: BUCsdE_TRn



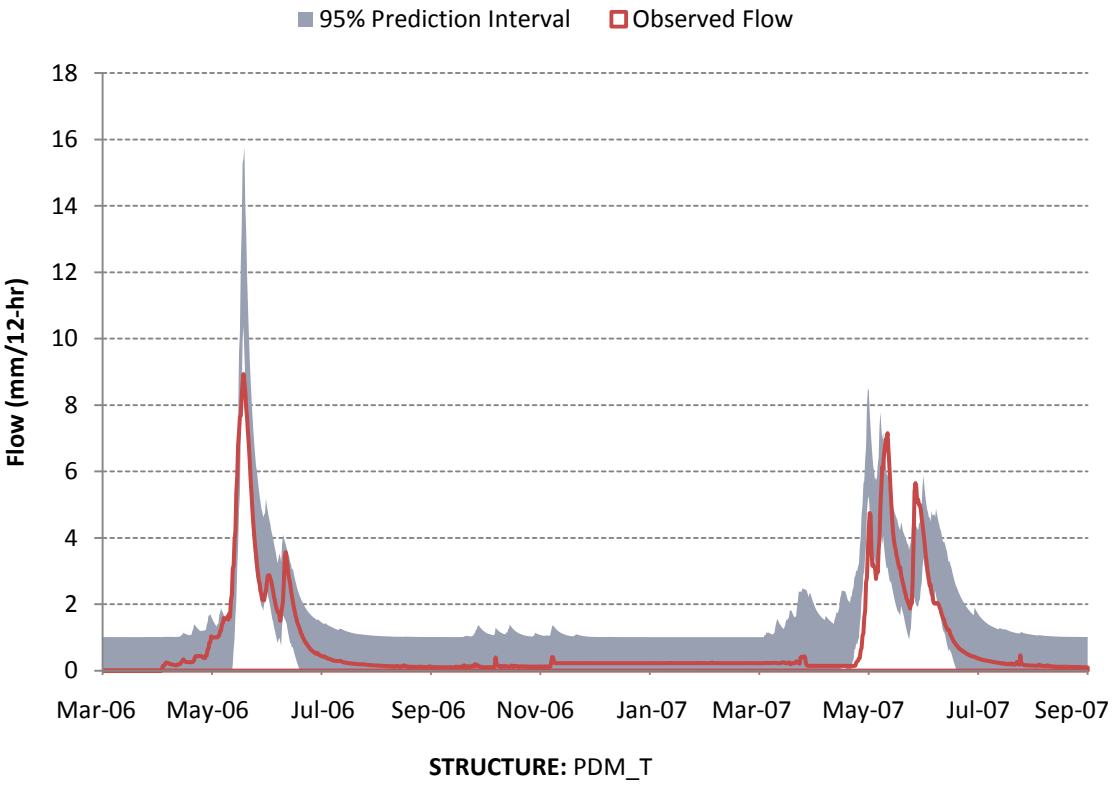
STRUCTURE: BUCsdAE_Rn



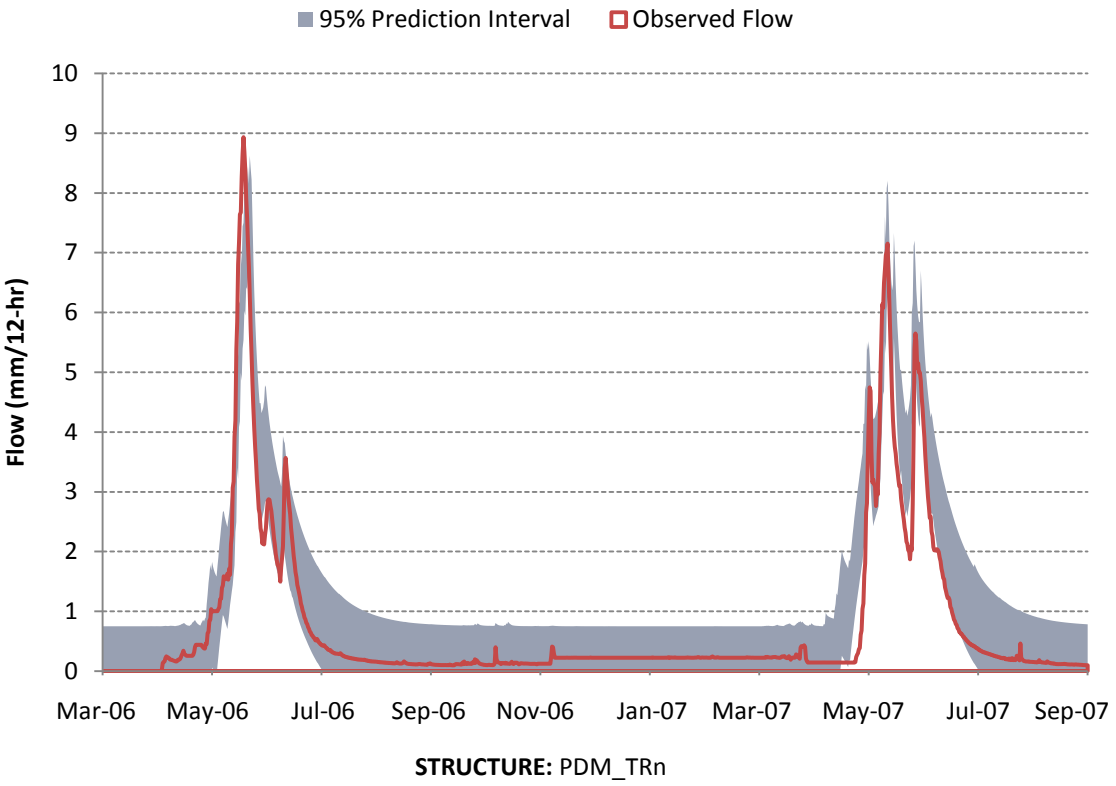
STRUCTURE: BUCsdAE_TRn



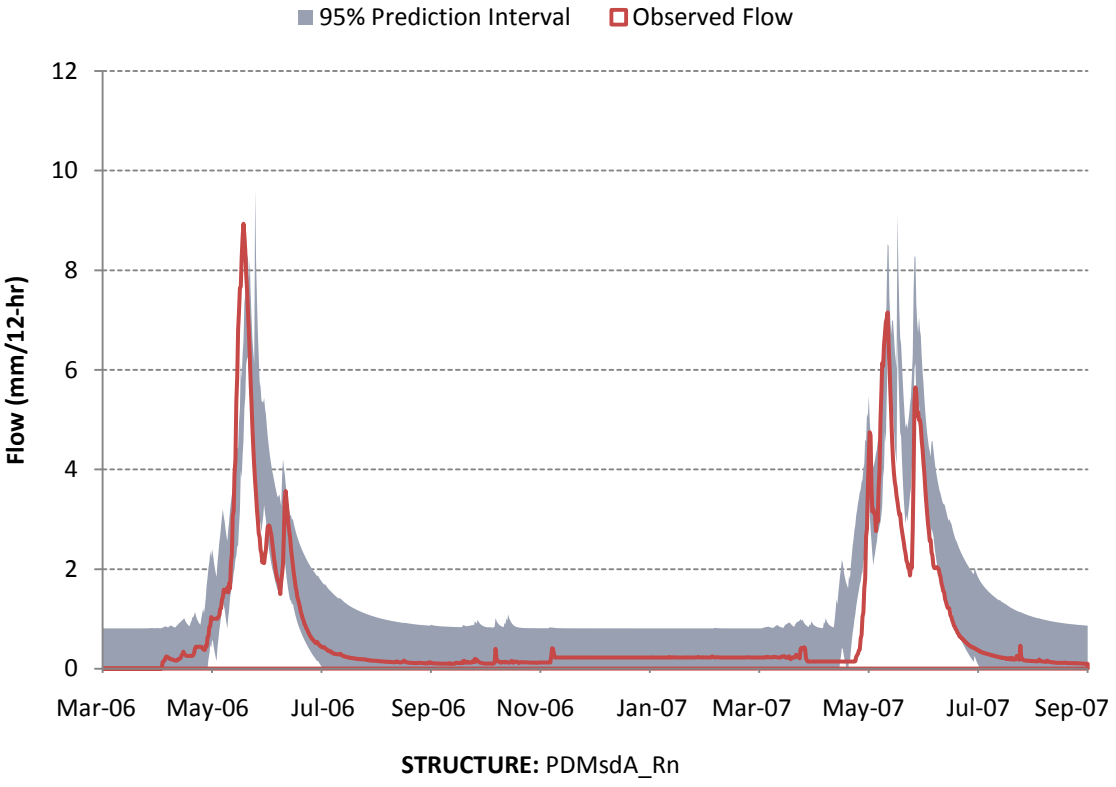
STRUCTURE: PDM_Rn



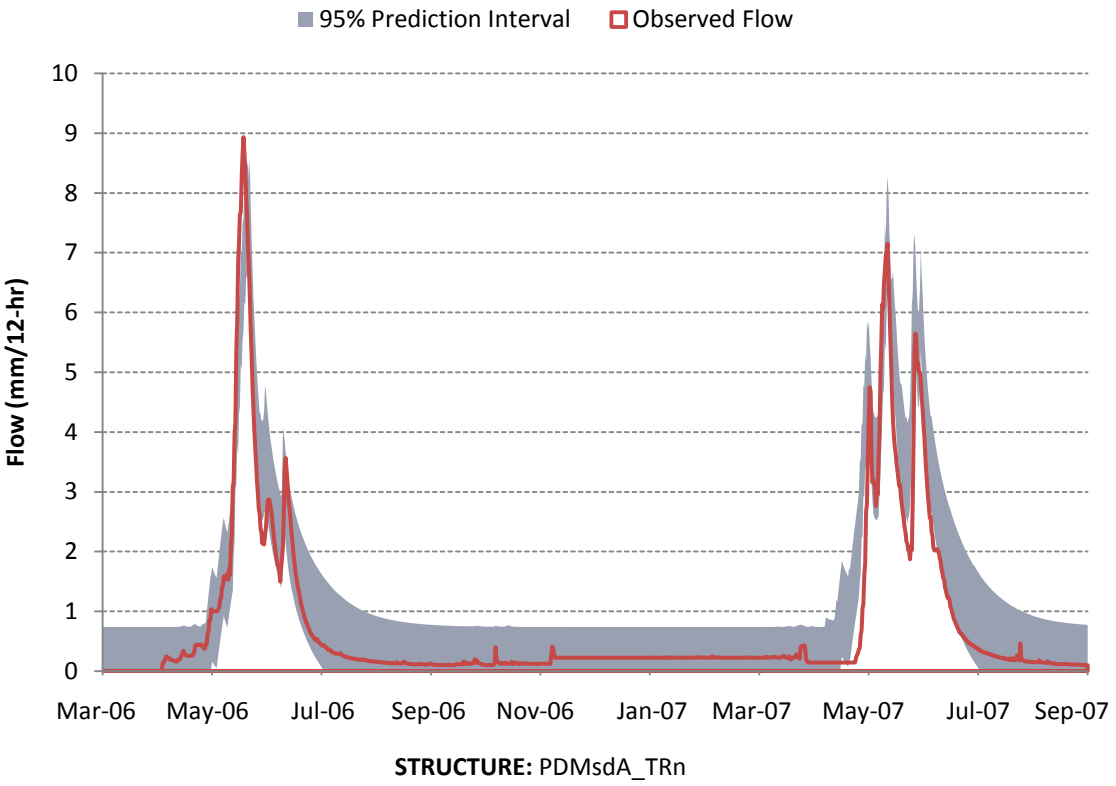
STRUCTURE: PDM_T



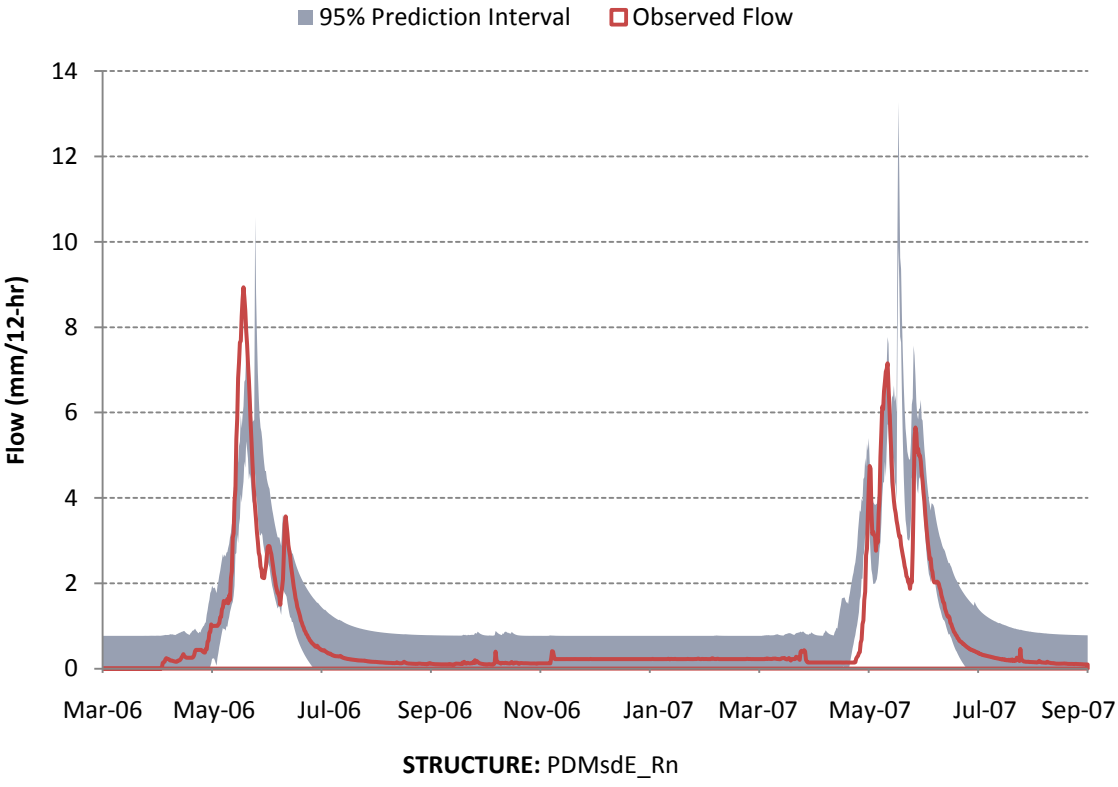
STRUCTURE: PDM_TRn



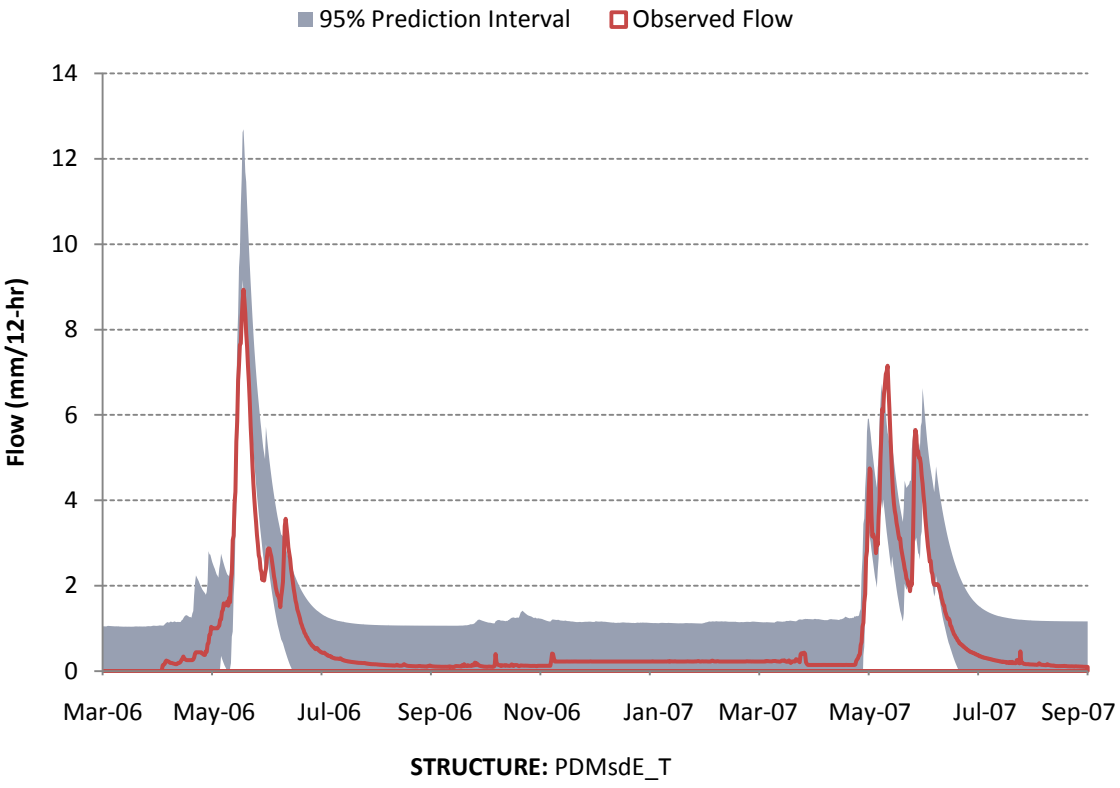
STRUCTURE: PDMsdA_Rn



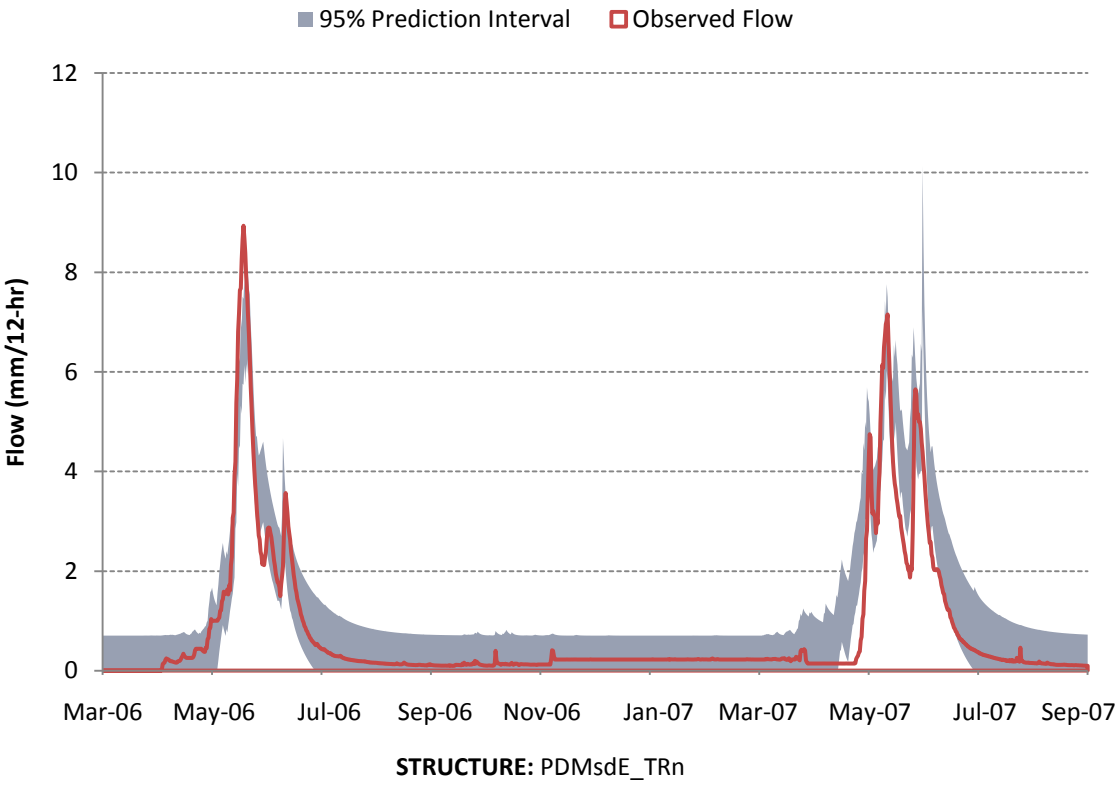
STRUCTURE: PDMsdA_TRn



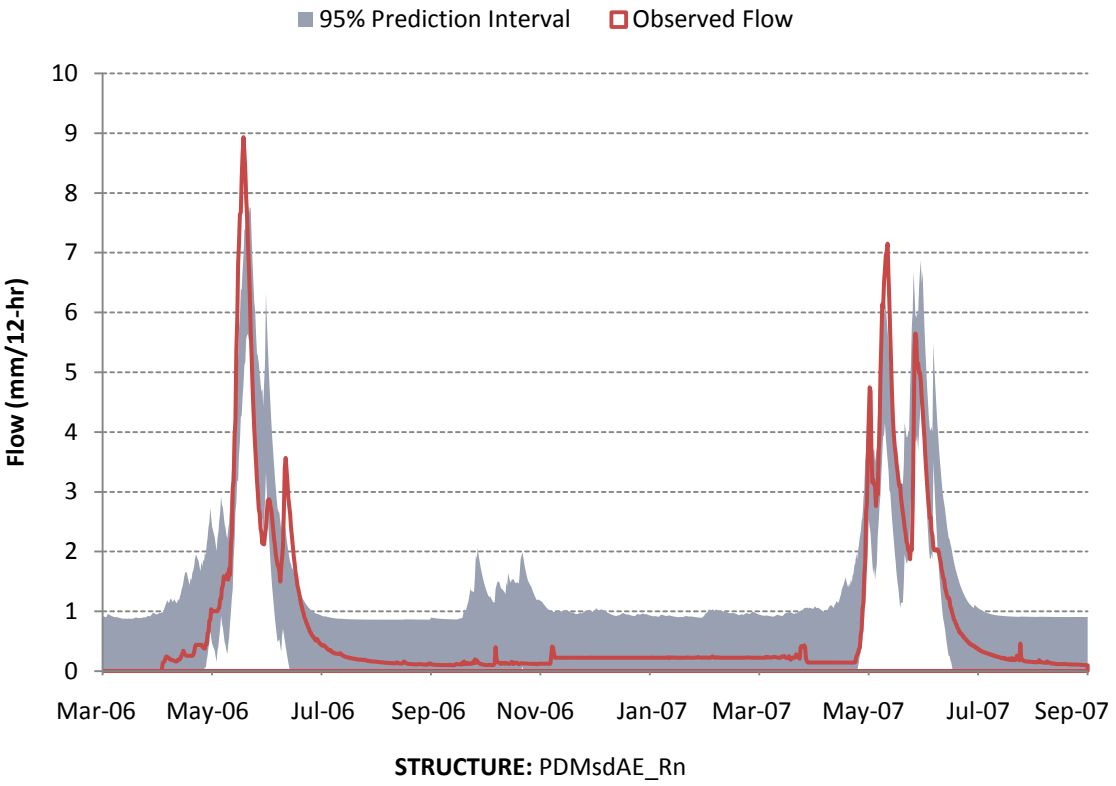
STRUCTURE: PDMsdE_Rn



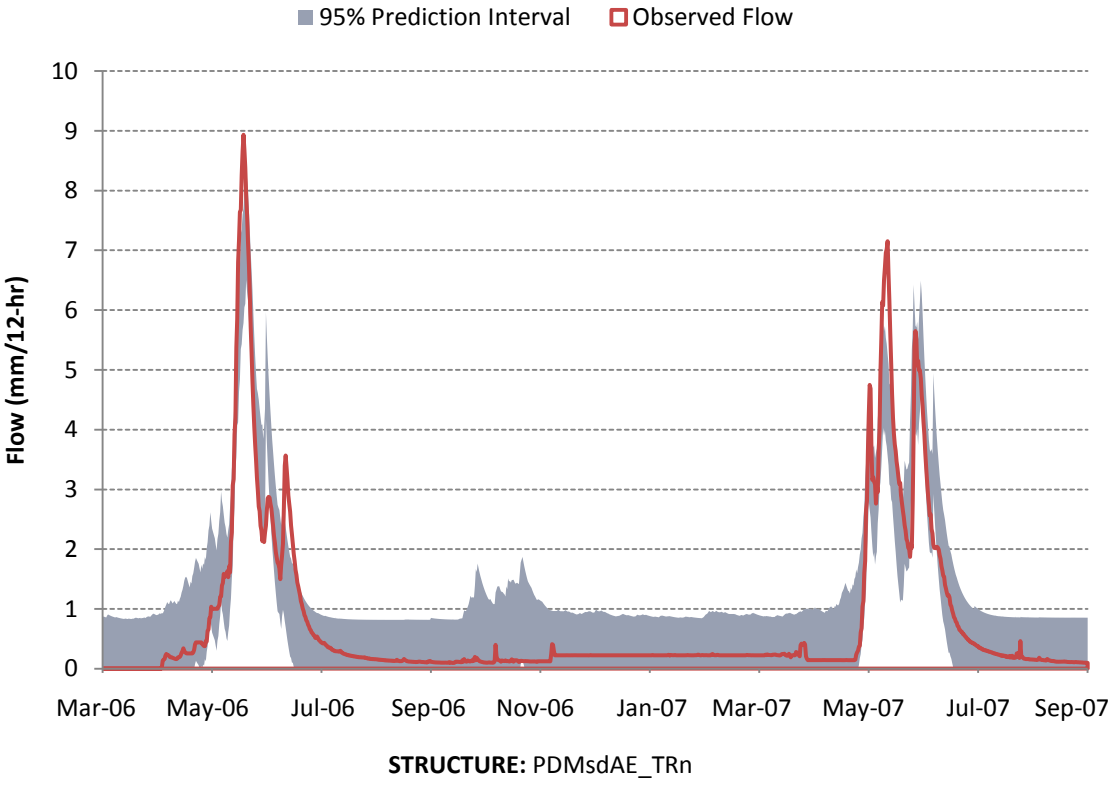
STRUCTURE: PDMsdE_T



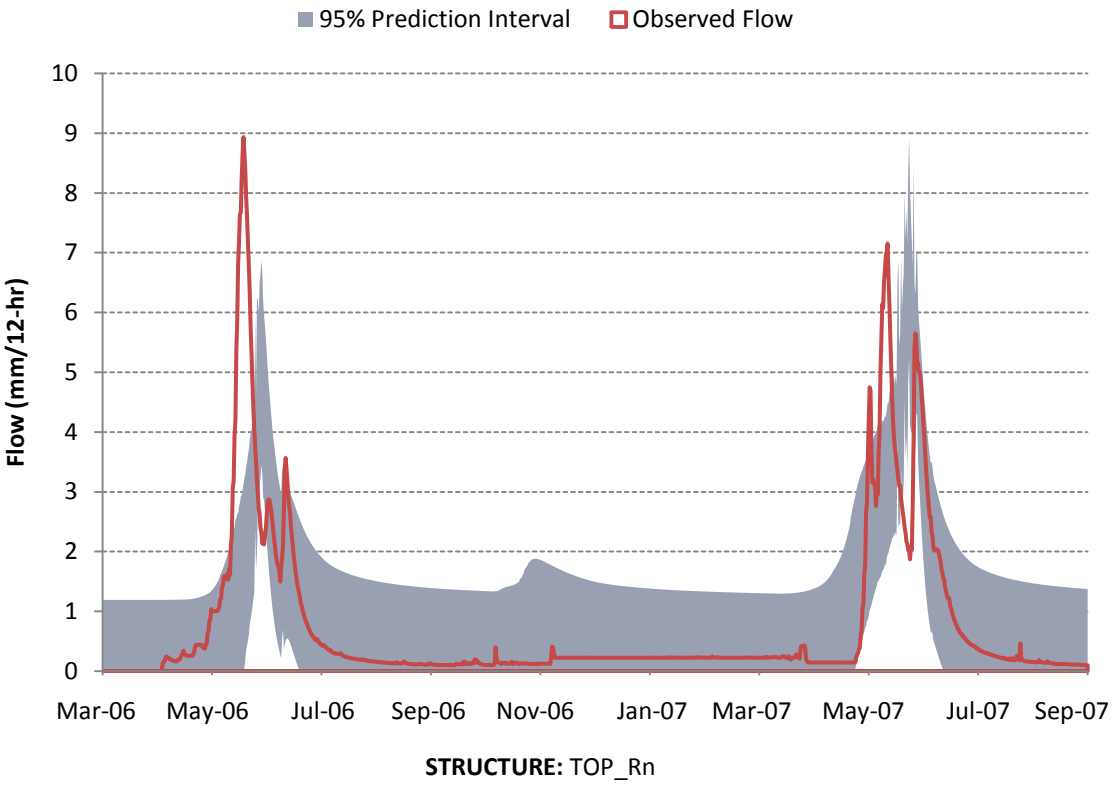
STRUCTURE: PDMsdE_TRn



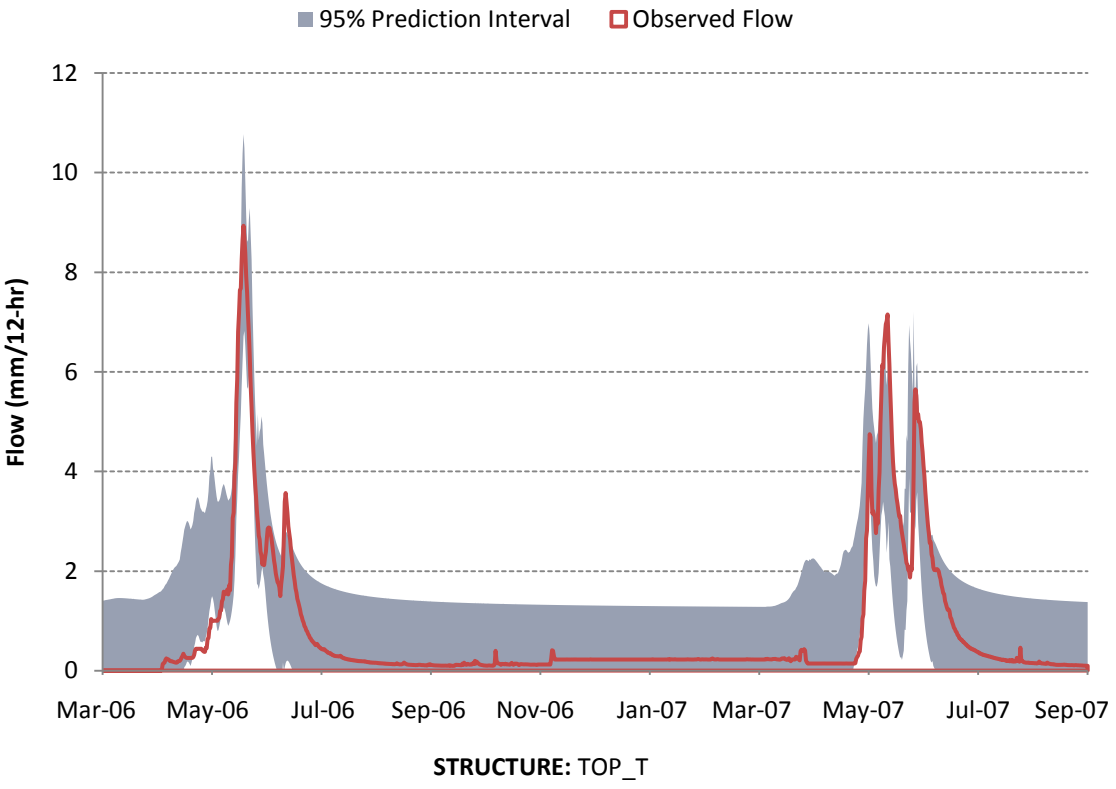
STRUCTURE: PDMsdAE_Rn



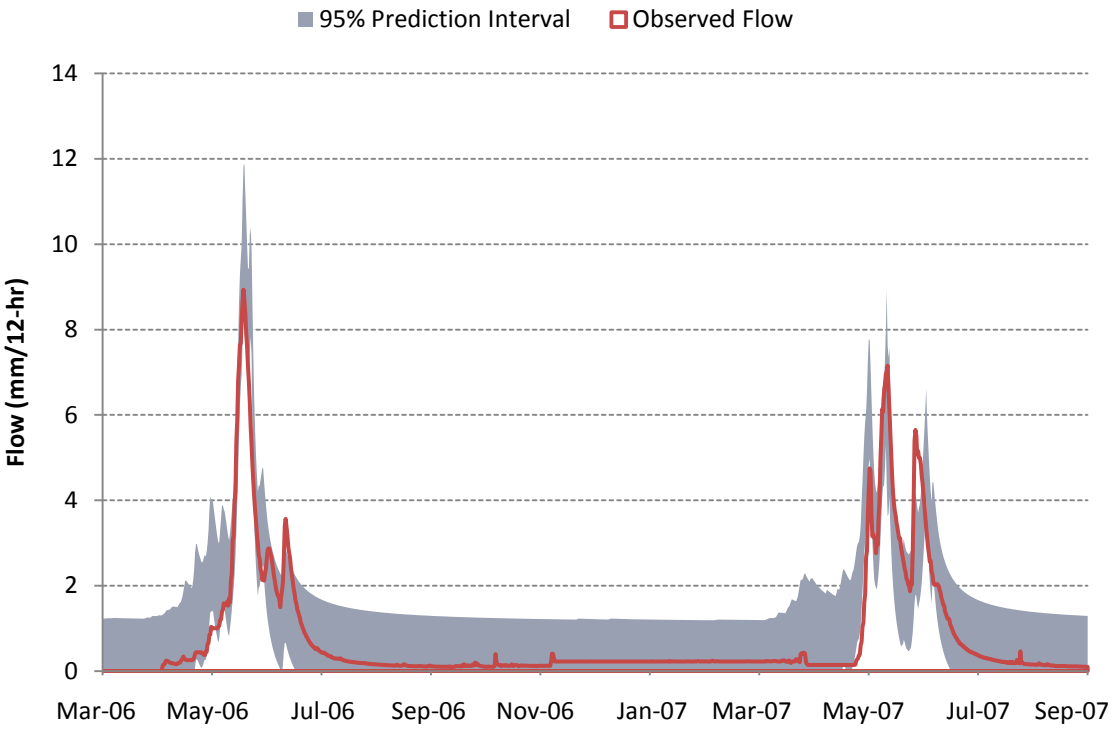
STRUCTURE: PDMsdAE_TRn



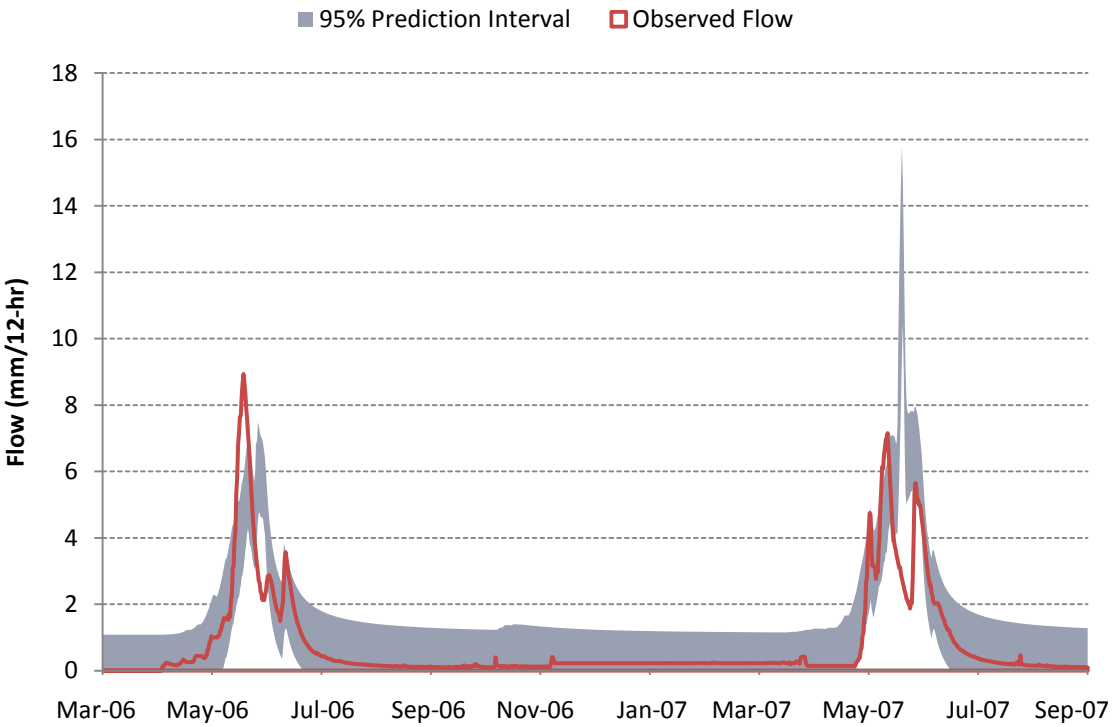
STRUCTURE: TOP_Rn



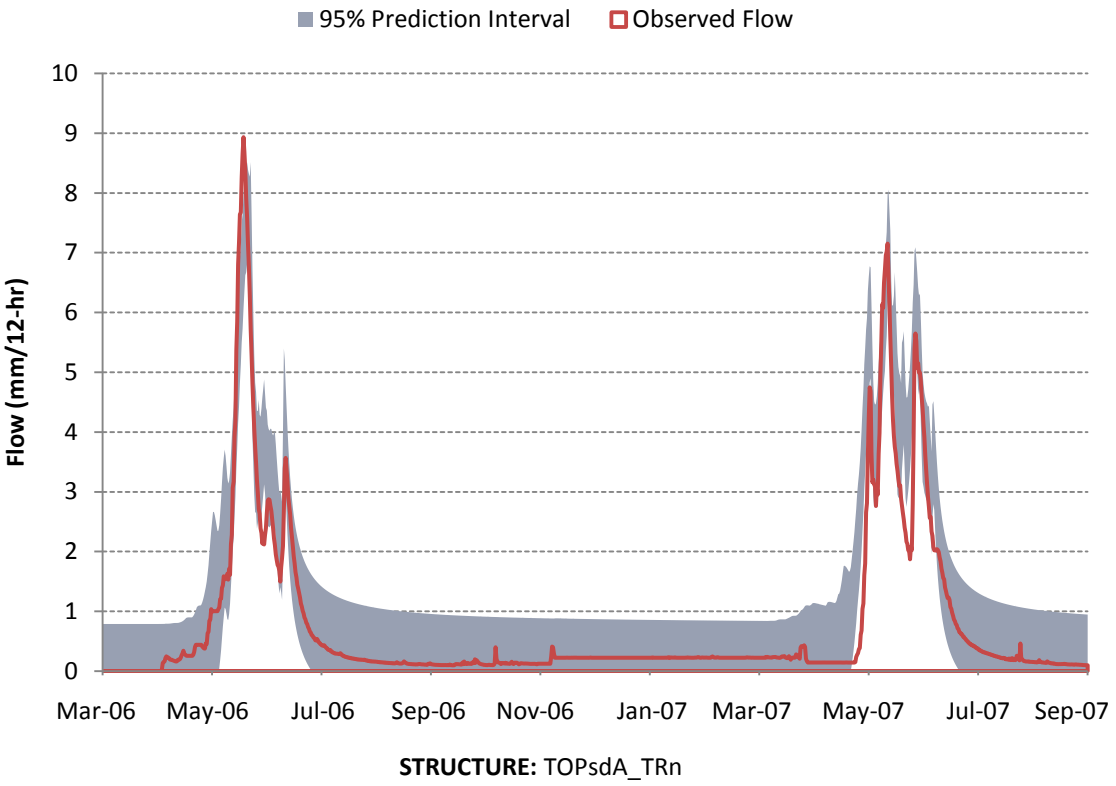
STRUCTURE: TOP_T



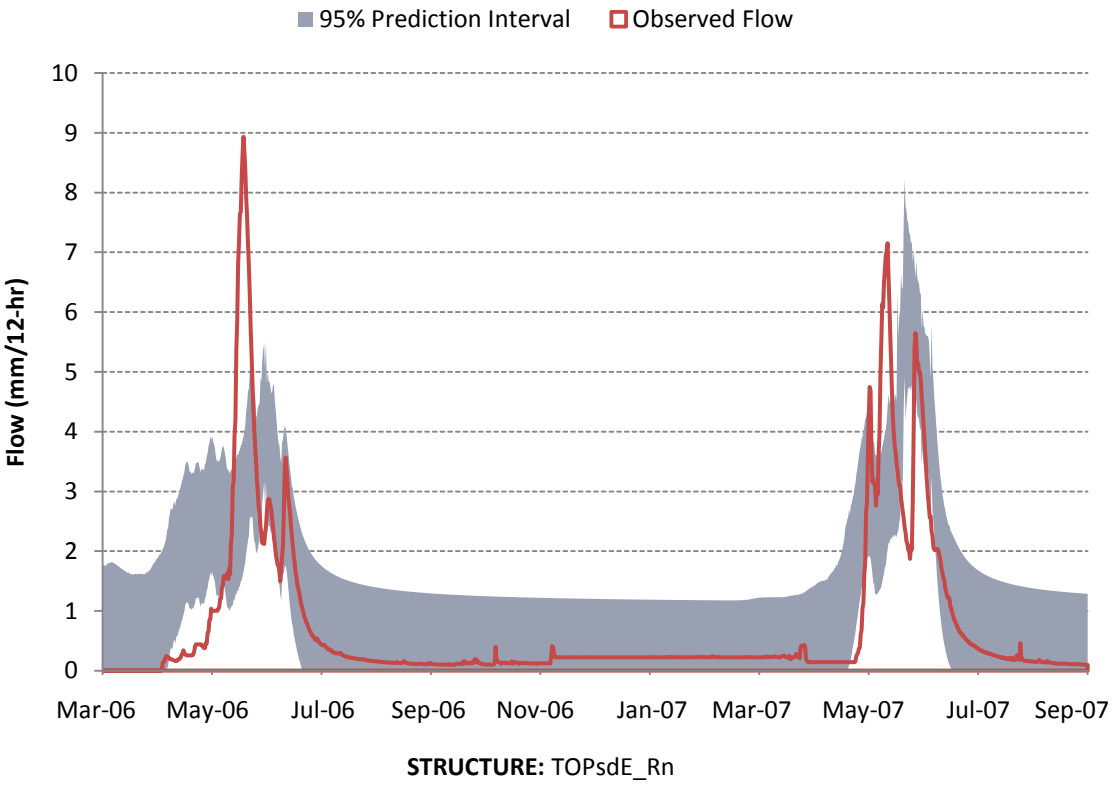
STRUCTURE: TOP_TRn



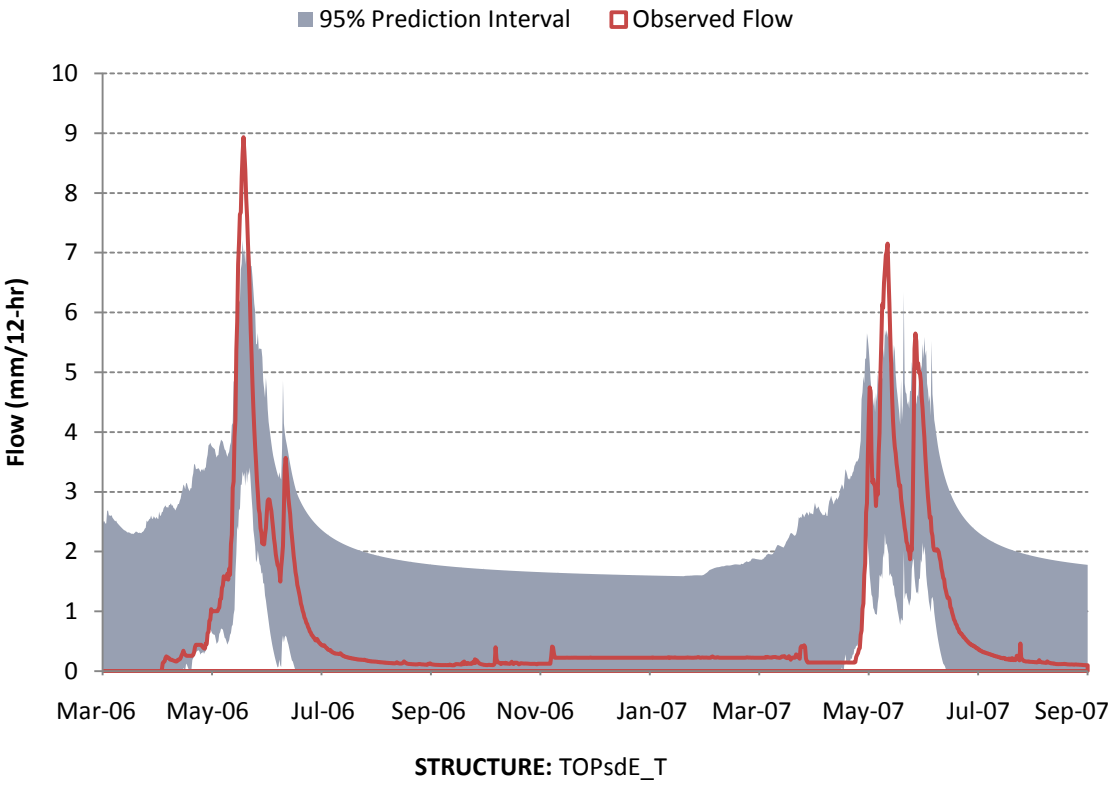
STRUCTURE: TOPsdA_Rn



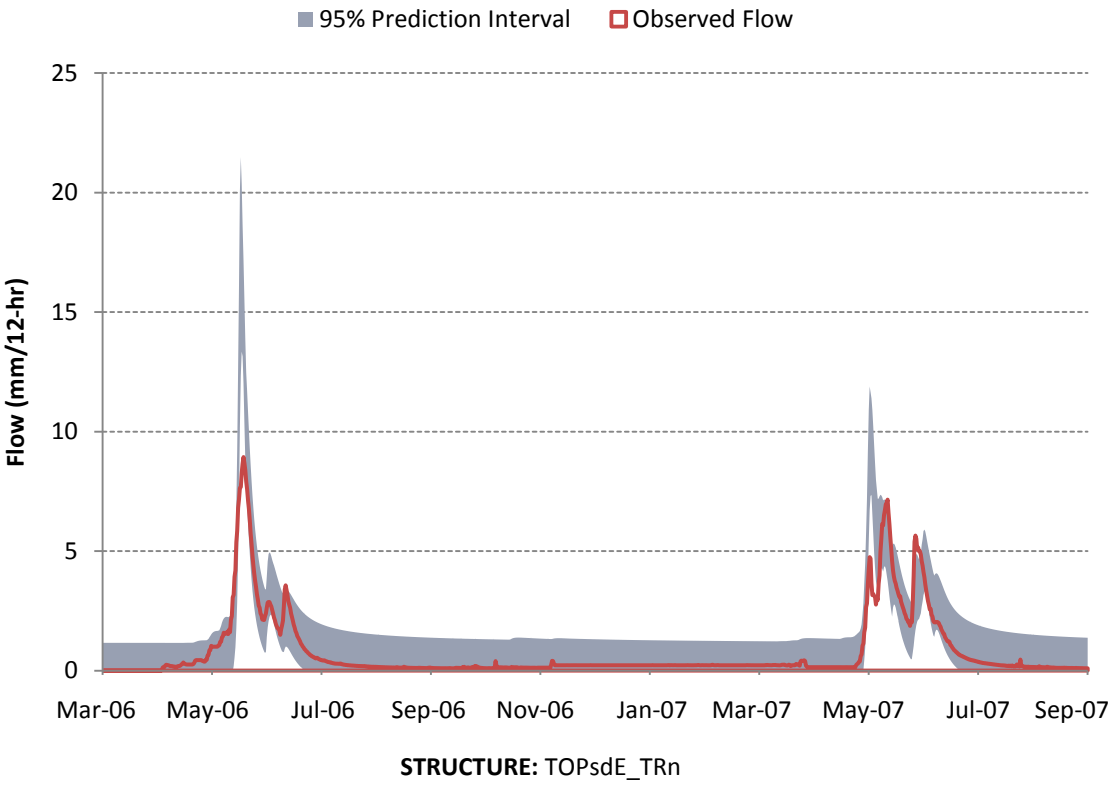
STRUCTURE: TOPsdA_TRn



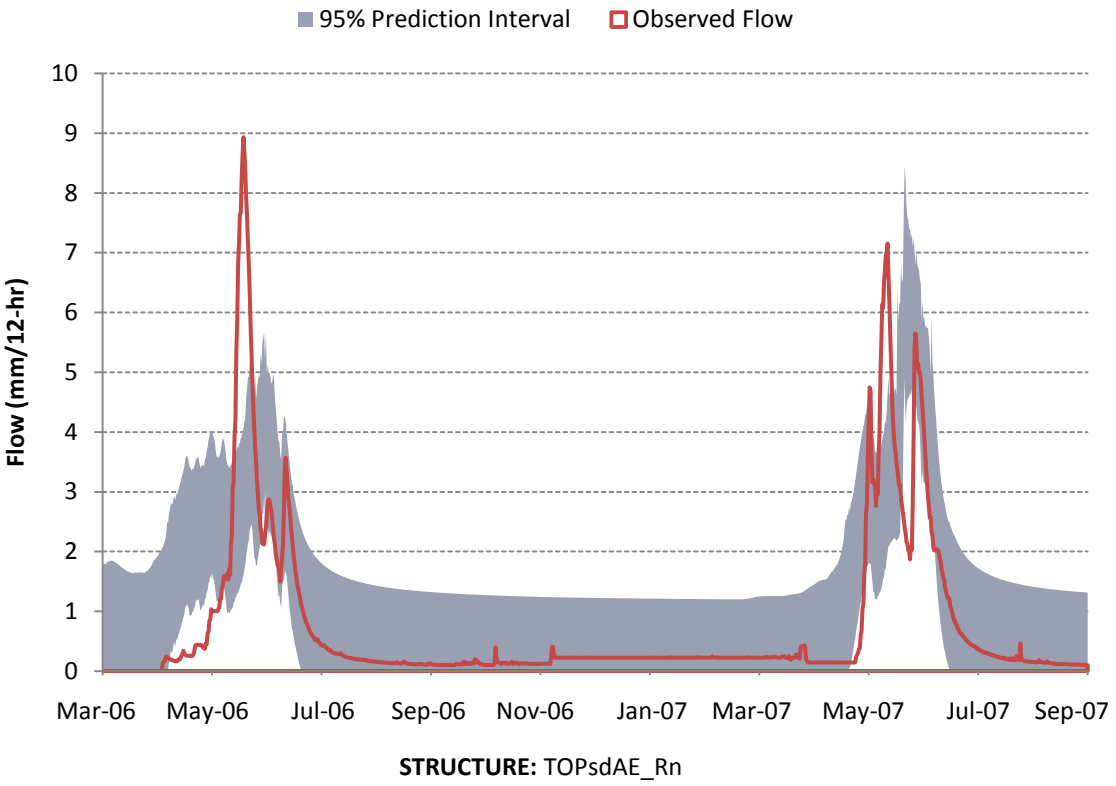
STRUCTURE: TOPsdE_Rn



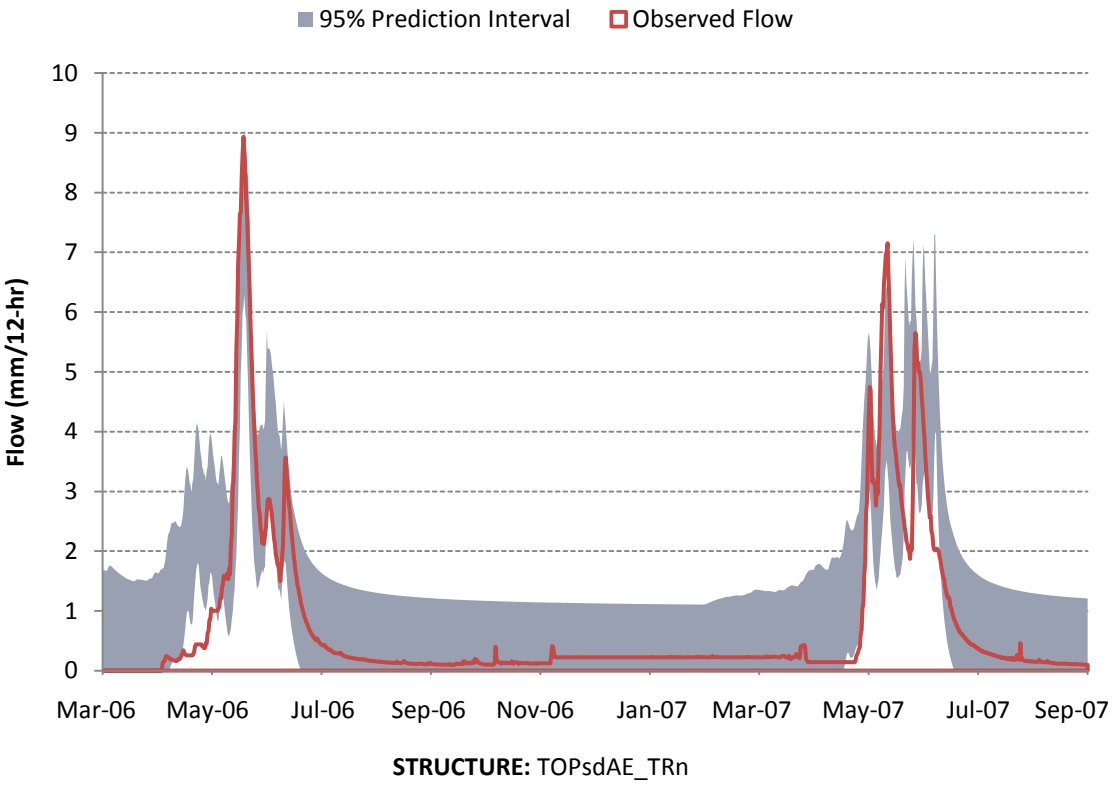
STRUCTURE: TOPsdE_T



STRUCTURE: TOPsdE_TRn



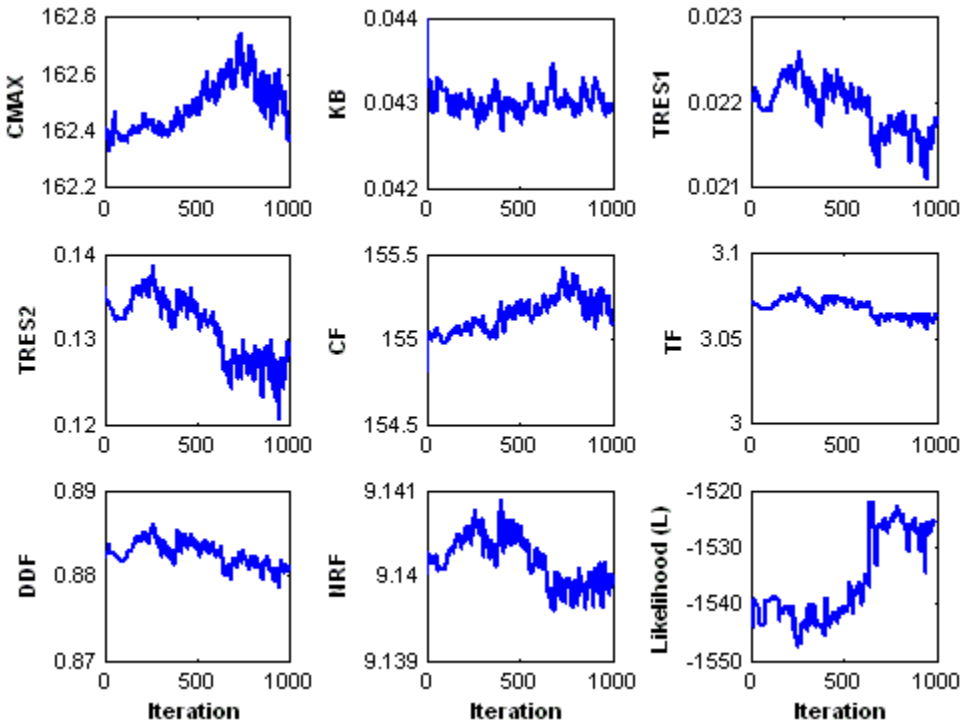
STRUCTURE: TOPsdAE_Rn



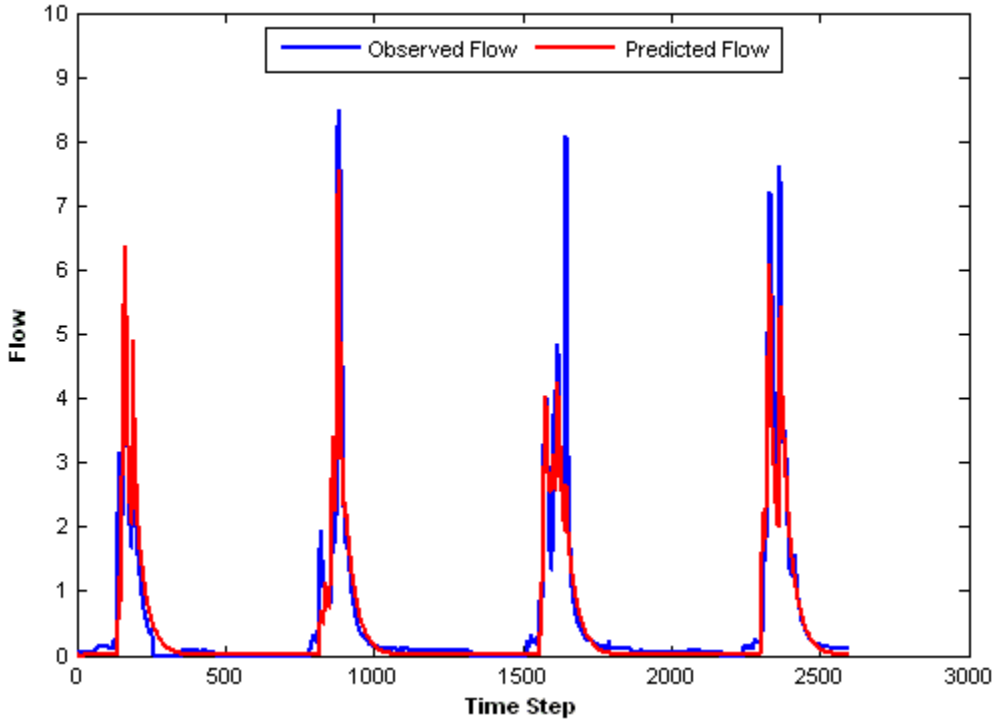
STRUCTURE: TOPsdAE_TRn

APPENDIX B

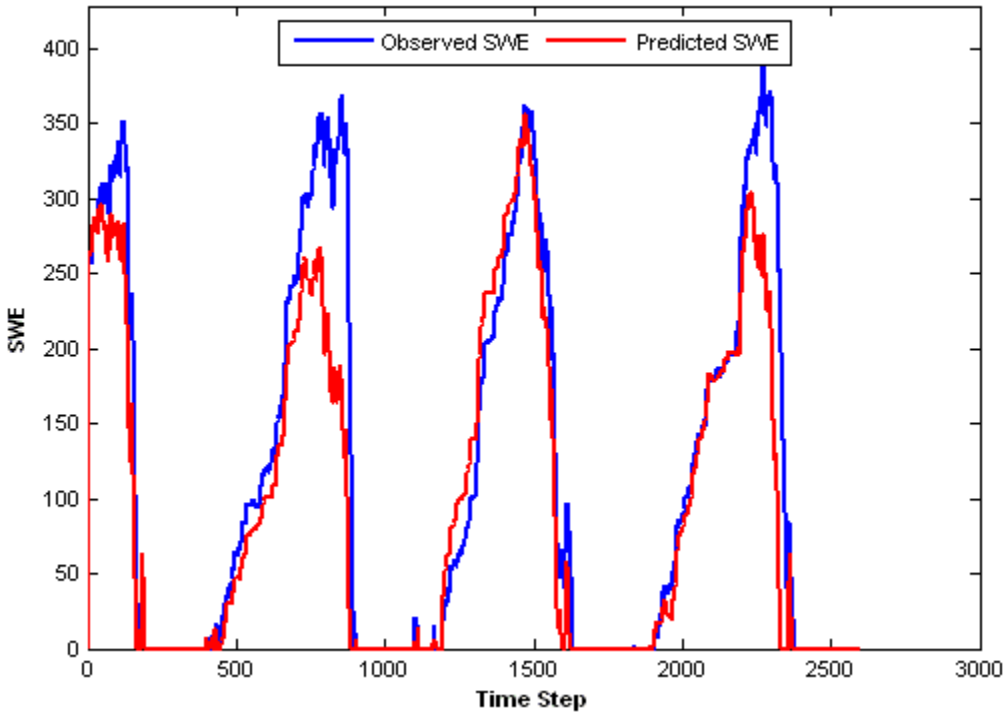
SPLASH – SAMPLE PLOTS



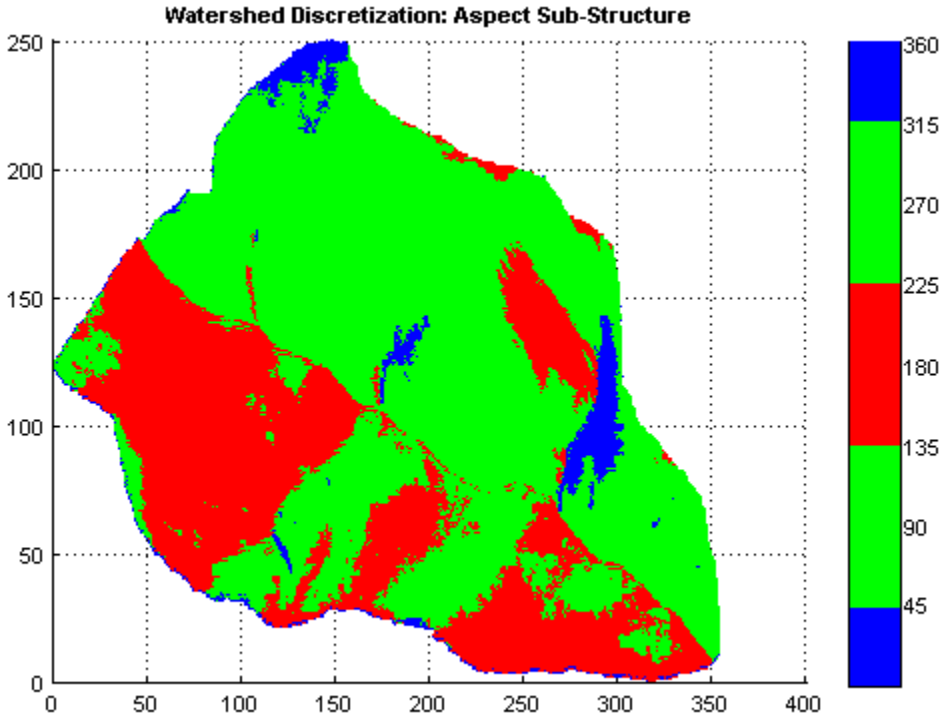
PLOT: Parameter Traces



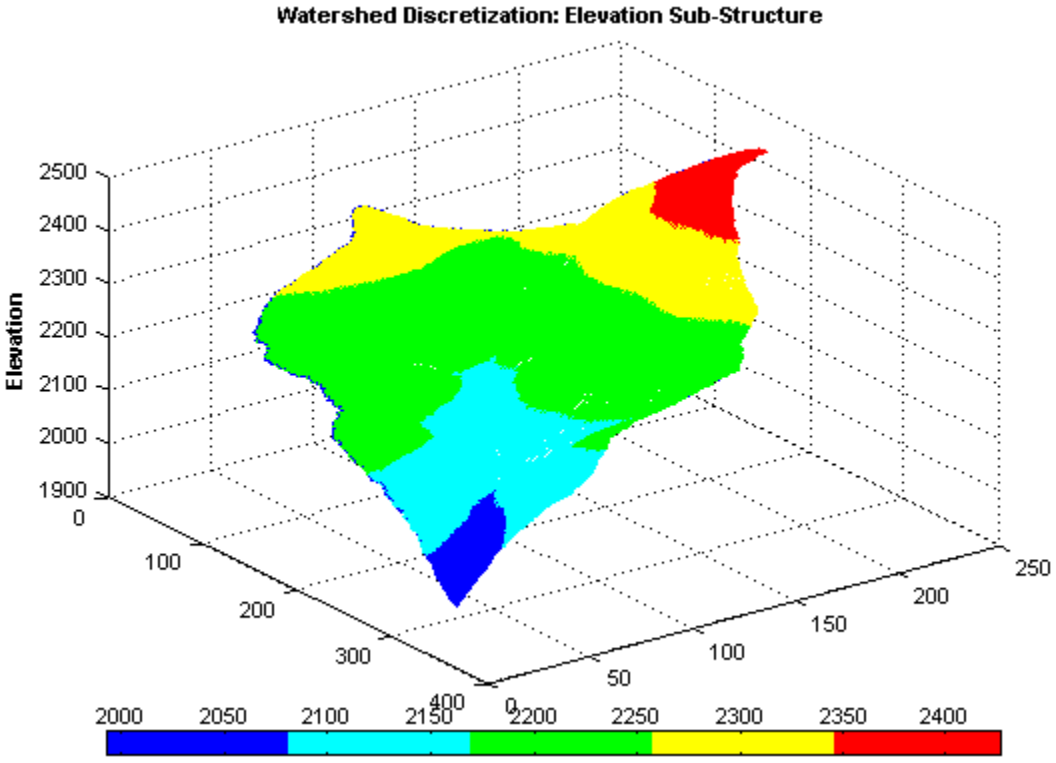
PLOT: Simulation Hydrograph



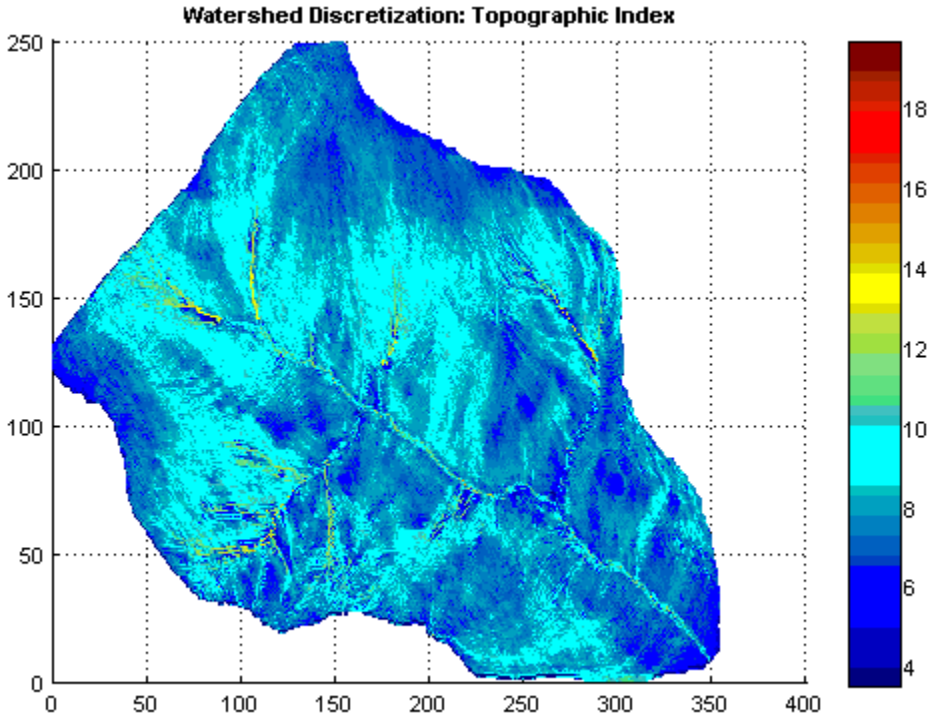
PLOT: Simulation Snow Water Equivalents



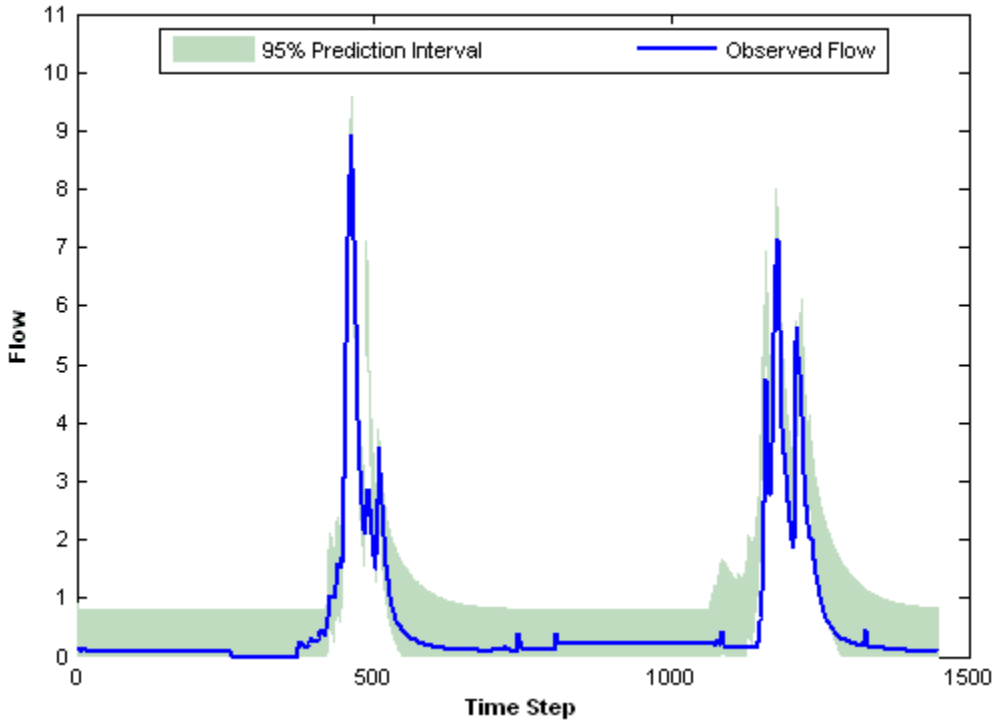
PLOT: Watershed Discretization: Aspect



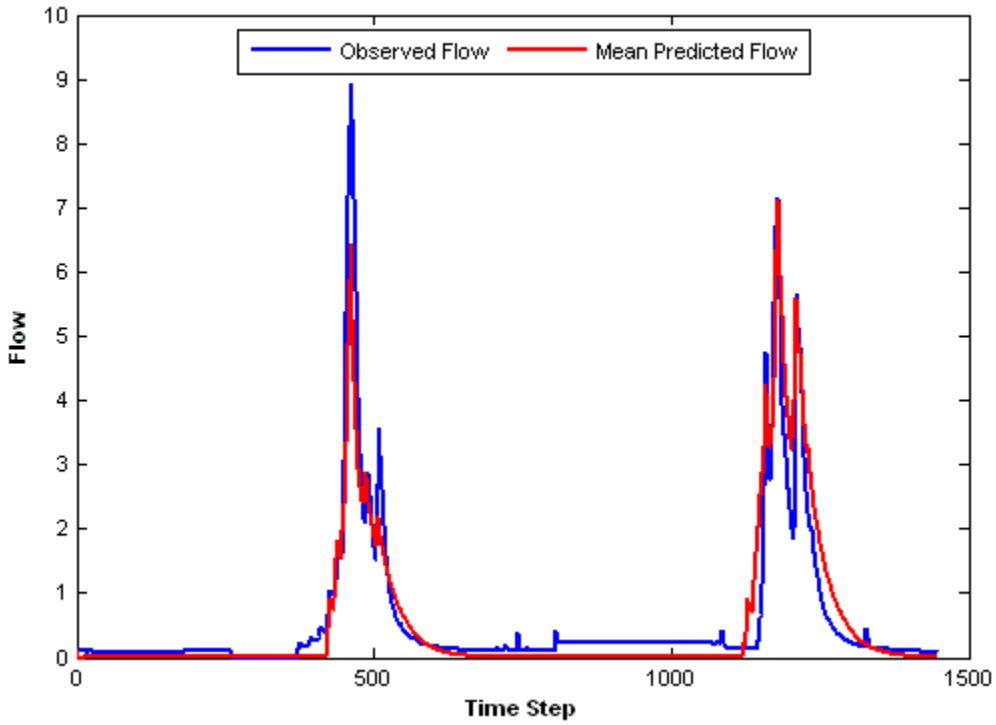
PLOT: Watershed Discretization: Elevation



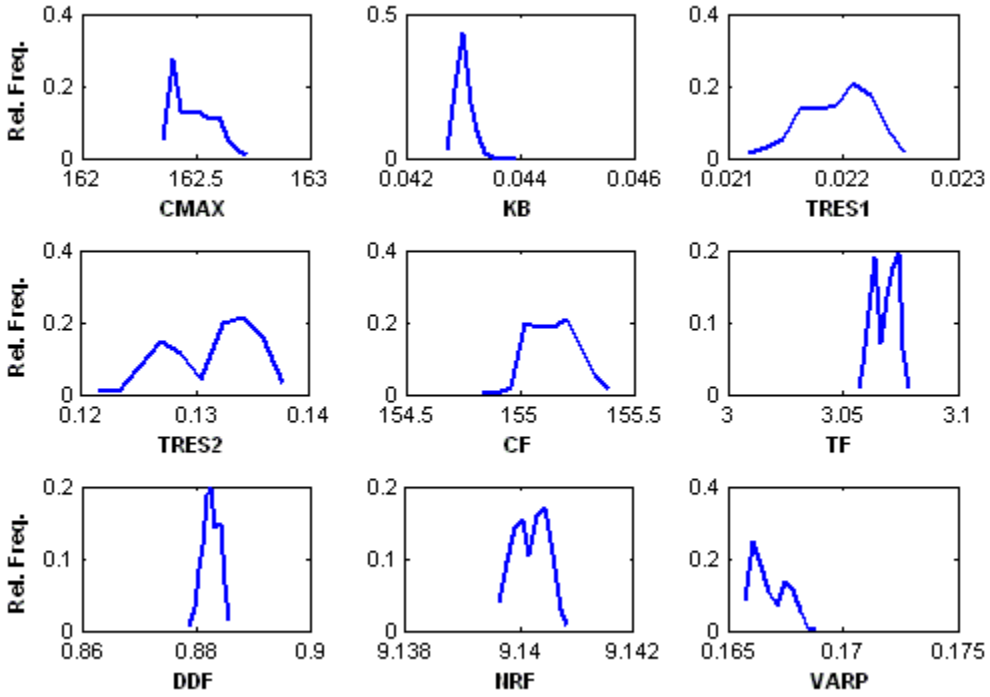
PLOT: Watershed Discretization: Topographic Index



PLOT: 95% Prediction Interval Hydrograph



PLOT: Mean Prediction Hydrograph



PLOT: Parameter Posterior Distributions