BAYESIAN OPTIMIZATION AND UNCERTAINTY ANALYSIS OF COMPLEX ENVIRONMENTAL MODELS, WITH APPLICATIONS IN WATERSHED MANAGEMENT

by

Able Mashamba

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

of

Doctor of Philosophy

in

Industrial Engineering

MONTANA STATE UNIVERSITY
Bozeman, Montana

November 2010

APPROVAL

of a dissertation submitted by

Able Mashamba

This dissertation has been read by each member of the dissertation 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. Lucy Marshall and Dr. Edward Mooney

Approved for the Department of Mechanical and Industrial Engineering

Dr. Chris Jenkins

Approved for the Division of Graduate Education

Dr. Carl A. Fox

STATEMENT OF PERMISSION TO USE

In presenting this dissertation in partial fulfillment of the requirements for a doctoral degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library. I further agree that copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for extensive copying or reproduction of this dissertation should be referred to ProQuest Information and Learning, 300 North Zeeb Road, Ann Arbor, Michigan 48106, to whom I have granted "the exclusive right to reproduce and distribute my dissertation in and from microform along with the non-exclusive right to reproduce and distribute my abstract in any format in whole or in part."

Able Mashamba

November 2010

DEDICATION

I dedicate this work and effort to Taropafadzwa Mashamba

TABLE OF CONTENTS

TABLE OF CONTENTS - CONTINUED

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

LIST OF FIGURES - CONTINUED

Figure Page

GLOSSARY

AM – Adaptive Metropolis (algorithm)

AWBM – Australian water balance model (case study model)

BMP – Best management practices

CI – Confidence interval

CN2 – Curve number of moist soils

DHSVM – Distributed hydrology soil-vegetation model

DRAM – Delayed Rejection Adaptive Metropolis (algorithm)

EXT – Extended (region)

FCOS – Cosine model (synthetic model)

FLO – Storage model (synthetic model)

FX2 – Quadratic model (synthetic model)

GLUE – Generalized likelihood uncertainty estimation

HPD – High posterior density (region)

HRU – Hydrologic response unit

LH-OAT – Latin hypercube – One factor at a time (algorithm)

MCMC – Markov chain Monte Carlo

MLR – Multiple linear regression

MSU-ES – Montana State University – Extension Services

NRIS – Natural Resources Information System

NRCS – Natural Resources Conservation Services

NSE – Nash Sutcliffe efficiency

ParaSol – Parameter solutions (algorithm)

GLOSSARY – CONTINUED

RBF – Radial basis function

RLF – Random local fitting

RSM – Response surface methodology

SCEUA – Shuffled complex evolution uncertainty analysis

SCS – Soil Conservation Services

SLHD – Symmetric Latin hypercube design

SSR – Sum of squared residual

STATSGO – State Soils Geographic

SUNGLASSES – Sources of Uncertainty Global Assessment using Split SampleES (algorithm)

SWAT – Soil and Water Assessment Tool

TCEF – Tenderfoot Creek Experimental Forest (case study watershed)

USDA – United States Department of Agriculture

USGS – United States Geological Survey

USLE – Universal soil loss equation

USLE_C – Universal soil loss equation due to vegetation cover

USLE_P – Universal soil loss equation due to support practices

ABSTRACT

This dissertation presents results of research in the development, testing and application of an automated calibration and uncertainty analysis framework for distributed environmental models based on Bayesian Markov chain Monte Carlo (MCMC) sampling and response surface methodology (RSM) surrogate models that use a novel random local fitting algorithm. Typical automated search methods for optimization and uncertainty assessment such as evolutionary and Nelder-Mead Simplex algorithms are inefficient and/or infeasible when applied to distributed environmental models, as exemplified by the watershed management scenario analysis case study presented as part of this dissertation. This is because the larger numbers of non-linearly interacting parameters and the more complex structures of distributed environmental models make automated calibration and uncertainty analysis more computationally demanding compared to traditional basin-averaged models. To improve efficiency and feasibility of automated calibration and uncertainty assessment of distributed models, recent research has been focusing on using the response surface methodology (RSM) to approximate objective functions such as sum of squared residuals and Bayesian inference likelihoods. This dissertation presents (i) results on a novel study of factors that affect the performance of RSM approximation during Bayesian calibration and uncertainty analysis, (ii) a new 'random local fitting' (RLF) algorithm that improves RSM approximation for large sampling domains and (iii) application of a developed automated uncertainty analysis framework that uses MCMC sampling and a spline-based radial basis approximation function enhanced by the RLF algorithm to a fully-distributed hydrologic model case study. Using the MCMC sampling and response surface approximation framework for automated parameter and predictive uncertainty assessment of a distributed environmental model is novel. While extended testing of the developed MCMC uncertainty analysis paradigm is necessary, the results presented show that the new framework is robust and efficient for the case studied and similar distributed environmental models. As distributed environmental models continue to find use in climate change studies, flood forecasting, water resource management and land use studies, results of this study will have increasing importance to automated model assessment. Potential future research from this dissertation is the investigation of how model parameter sensitivities and inter-dependencies affect the performance of response surface approximation.

CHAPTER 1

INTRODUCTION

Background

In watershed hydrology, the use of simulation models to study the origins, storage and movement of water, dissolved pollutants and sediment in a watershed is wide-spread (Yang, Herath et al. 2000; Beven and Feyen 2002; Dunn, Freer et al. 2008; Soulsby, Neal et al. 2008). Knowledge gained from hydrologic modeling and simulation studies guides a variety of engineering, scientific and legislative applications, including the design of civil works and human settlements, the formulation of water management and land use policies and the prediction of critical phenomena in a watershed. Such phenomena could include water scarcity, flooding, erosion, river morphology, surface and ground water pollution, threats to aquatic and riparian ecologies or impacts of climatic changes to a watershed (Black 1996).

When computing power was limited, watershed modeling mainly involved studying or predicting the average behavior, response, or variables of a watershed (Beven 2001). These watershed-averaged (or lumped) models are often adequate for studying water balance, long-term behavior, and mean outlet discharge in the watershed (Sivakumar 2008). However, lumped models do not represent the spatial distribution of internal behavior, inputs, or responses of a watershed. For large drainage basins, the spatial heterogeneity of precipitation, weather, channel networks, soils, land use, and land cover is significant. Models that represent spatial heterogeneity in watershed inputs, interactions, and responses tend to have better explanatory power and also have the

ability to identify and isolate a locale with critical watershed responses (Moreda, Koren et al. 2006; Hansen, Refsgaard et al. 2007). Such models are called distributed hydrologic models. They may also be referred to as complex models due to their high structural detail, their representation of system non-linearity, and their consequently high demand for computational power (Hansen, Refsgaard et al. 2007).

Figure 1 shows a summary of the hydrologic processes that are often simulated in land-phase hydrologic studies of watersheds. The square grids in Figure 1 represent the spatial resolution or scale that a distributed modeling paradigm might use to capture homogeneity and heterogeneity. Within each cell, parameters, hydrologic factors and responses are homogeneous, while different grids can freely have different levels of the parameters, factors and responses under study.

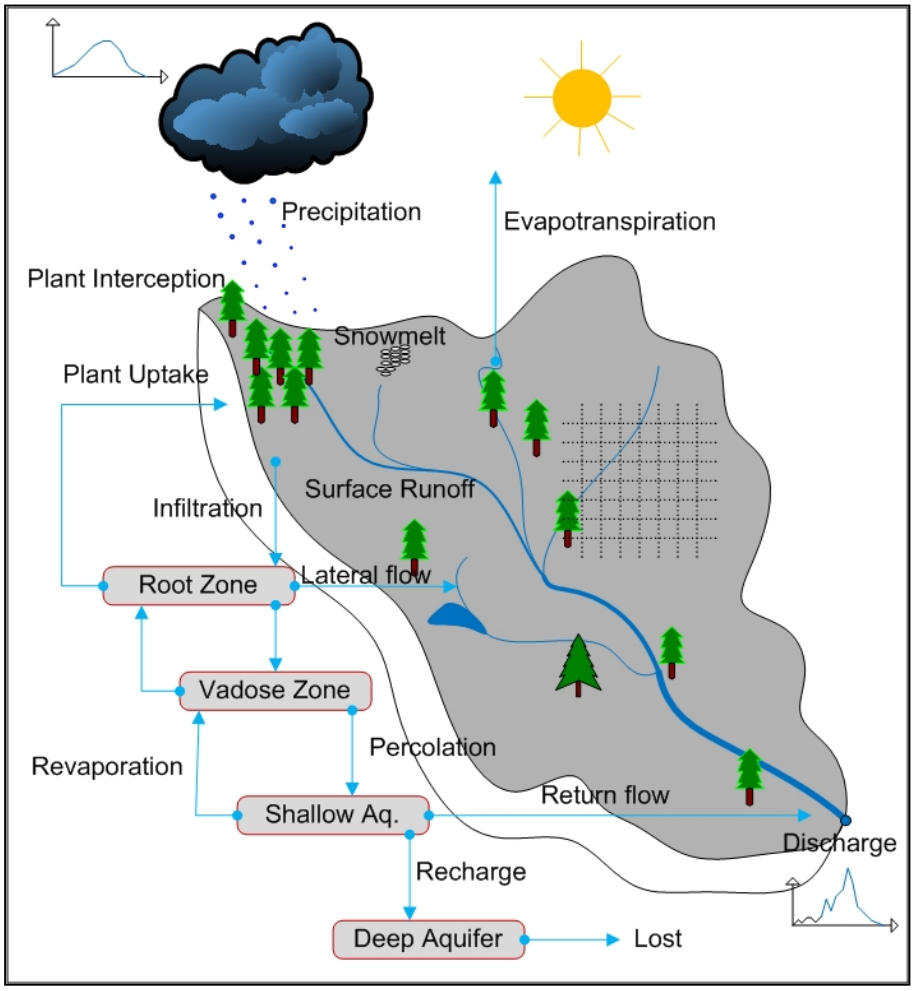

Figure 1: Distributed hydrologic processes

Distributed models need not represent continuous spatial variability as found in real systems. Instead, they often employ various schemes to sub-divide or delineate a watershed into regular or regular sub-watersheds each containing homogenous hydrologic factors under study. The entire watershed can thus be described by the aggregation of the individual lumped sub-models.

Models that have a limit on the resolution to which heterogeneity can be represented, such as the Soil Water Assessment Tool (SWAT) (Arnold, Srinivasan et al. 1998) and HYDROGEIOS (Efstratiadis, Nalbantis et al. 2007), are better known as semi

or partially distributed models (Hansen, Refsgaard et al. 2007). Models with no limit to the spatial resolution (scale) that they can represent, such as the Distributed Hydrology Soil-Vegetation Model, DHSVM, (Wigmosta, Vail et al. 1994; Wigmosta, Nijssen et al. 2002) and MikeSHE (Danish Hydraulics Institute), are considered fully distributed. The higher the spatial resolution the greater the computational demand a distributed model will have in simulation, calibration and analysis of sensitivity and uncertainty (Montanari and Brath 2004; Muleta and Nicklow 2005). This is because the number of parameters or structural constants of a model (and hence its computational complexity) increases with resolution. Parameters of distributed models can theoretically be measured from the sub-basin being modeled (if they correspond to measurable factors), researched from literature or obtained using automated search algorithms such as the Nelder-Mead, evolutionary, genetic, and Bayesian Markov chain Monte Carlo methods (Eckhardt, Fohrer et al. 2005). This process of manually or automatically adjusting parameter values so that the output of a model best matches a set of observed data is called optimization or calibration. Parameter sensitivity analysis is the study of how changes to model parameters affect model output. Sensitivity analysis (SA) is and should be conducted before calibration and uncertainty analysis so as to exclude insensitive parameters from further study and hence reduce the computational costs of calibration and uncertainty analyses. SA is also useful in assessing the robustness and stability of the model to changes in inputs and in ordering the strengths and relevance of inputs. Parameter uncertainty analysis involves determining the likely probability distributions or confidence intervals of parameter values and hence the likely distributions or confidence intervals of the predictions of a model. Model predictive uncertainty information is

critical to risk assessment, improved objectivity, creditability and hence acceptability of simulation output (Montanari 2007; Sivakumar 2008).

While classical statistical methods are often used to identify and quantify parameter and predictive uncertainty, the large number of runs required to do so can be prohibitive for computationally expensive models such as distributed models (Bates and Campbell 2001). In fact, early work on calibrating a semi-distributed SWAT model (see Chapter 2) for a case study Lower Yellowstone watershed in Eastern Montana using the native evolutionary algorithm, the Shuffled Complex Evolution – Uncertainty Analysis, SCEUA, (Duan, Gupta et al. 1993) highlighted the computational difficulties associated with conditioning distributed models. As Chapter 2 shows, automatic calibration of the SWAT model would be unfeasibly long for urgent studies or for comprehensive uncertainty analysis schemes. Second, the native routines of the SWAT model, like other classical statistics frameworks, used "good enough'" parameter sets in constructing uncertain distributions and confidence intervals. Within this paradigm, good enough parameter sets are those that give the smallest sum of squared differences (residuals) between simulated output and conditioning observed data. However, the often debated issue is that the selection of "good-enough" parameter values depends on arbitrary thresholds around the optimum parameter set (Selker 2002; Beven, Smith et al. 2006; Mantovan and Todini 2006; Beven, Smith et al. 2007; Mantovan, Todini et al. 2007). There is no rigorous statistical argument or justifiable assumptions that can be provided to support these cut-off points on parameter values based on model residuals (Mantovan and Todini 2006; Mantovan, Todini et al. 2007).

Bayesian Markov chain Monte Carlo (MCMC) methods, which use Bayes' conditional probability rule, offer a statistically coherent alternative to model calibration and parameter uncertainty analysis (Kaheil, Gill et al. 2006). The MCMC paradigm to uncertainty analysis is based on the statistical reasoning that, (i) if a joint-prior parameter probability distribution is reasonably estimated and (ii) a parameter likelihood function is chosen based on the assumptions about the probability distribution of the residual of a model, then (iii) an ergodic Markov chain set on Bayes' rule will always converge to the target joint-posterior probability distribution of the parameters, as the chain gets long and/or as observed data increases (Gelman, Carlin et al. 2004).

While using Bayesian MCMC methods is appealing and has been successfully used because of existing rigorous arguments that support the statistical coherency of the resulting uncertainty analysis, MCMC methods do not solve the problem of long calibration times for distributed models. Fortunately, by considering the simulation of a distributed model for a given period as a being equivalent to evaluating a computationally expensive black-box function, modelers can use response surface methodology (RSM) to replace lengthy simulations during the calibration of distributed models (Gutmann 2001(a); Regis and Shoemaker 2004; Regis and Shoemaker 2005; Regis and Shoemaker 2007; Bliznyuk, Ruppert et al. 2008). The response surface methodology is a statistical tool that uses quantitative data from appropriate experiments to establish or explore any existing relationships between several explanatory variables and one or more response variables.

Figure 2 is a schematic of how computationally expensive distributed hydrologic models can be replaced by emulators during the simulation run or by RSM models during

calibration. An emulator is any system that mimics the responses of another system, preferably at a lower cost or avoiding some other difficulty. Unlike RSM fitted functions, emulators need not be developed from designed experiments. Adaptive sampling and neural network algorithms are often used in emulators to mimic a set of observation or characteristics. The surrogate models are generally statistical or mathematical functions whose parameters (β) are obtained from fitting inputs to responses of suitable sampled model parameters (Θ) of the computationally expensive distributed hydrologic models, as illustrated in Figure 3. The RSM or emulator functions are orders of magnitude faster to evaluate than the actual models and hence can be used to speed up the modeling and/or calibration and uncertainty assessment.

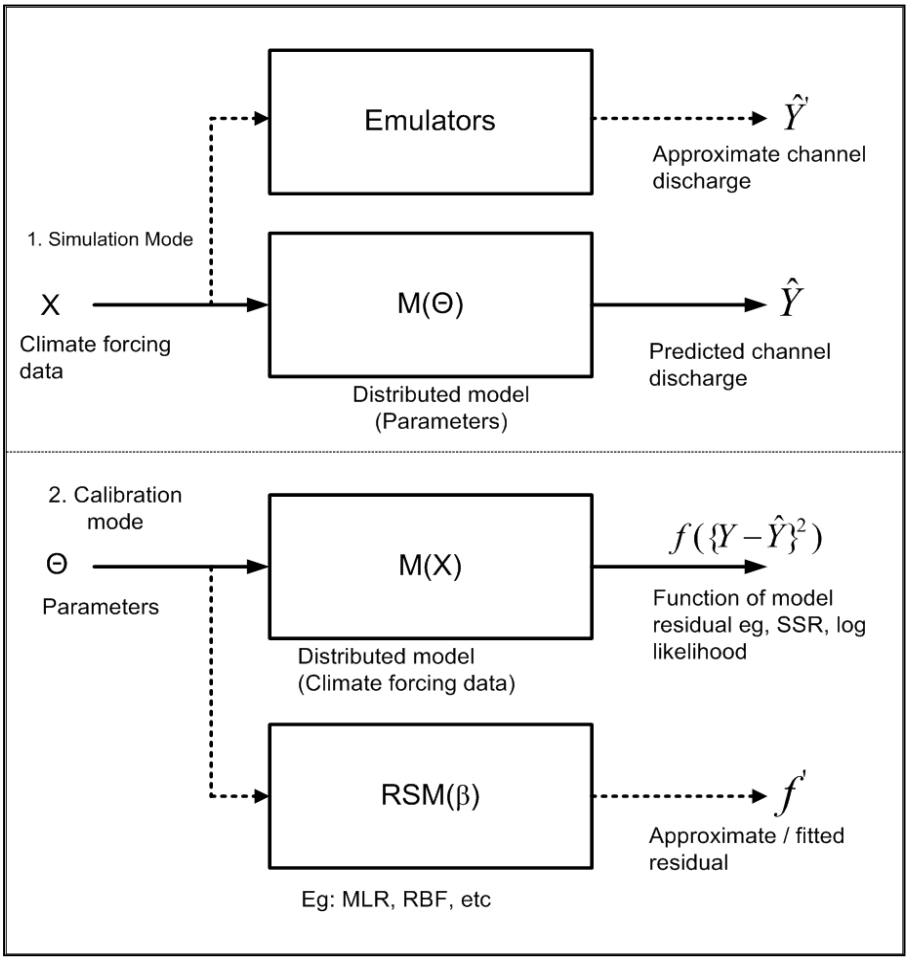

Figure 2: Schematic of surrogate models

Two common RSM approximation functions are radial basis and multiple linear regression functions. These two RSM functions are particularly popular in environmental applications for their conceptual simplicity and fitting properties assuring a small fitting region that covers the neighborhood of a global optimum point. Figure 3 is a one-dimensional (parameter) illustration of how the radial basis and a second order multiple linear RSM functions may be sampled and fitted to the response or sum of squared residual of the exact hydrologic model. In Figure 3, the big dots represent uniformly sampled points (parameter values) used for running the model and hence at which to

evaluate the actual expensive model sum of squared residuals. Radial basis function (RBF) fitting passes through all the previously evaluated design points, while a second order multiple linear regression (MLR) function fits the design points to a single optimum (uni-modal) curvature. Unlike in the hypothetical Figure 3, within the neighborhood of most real global optima, the response surface tends to be uni-modal and hence is often approximated by MLR curves.

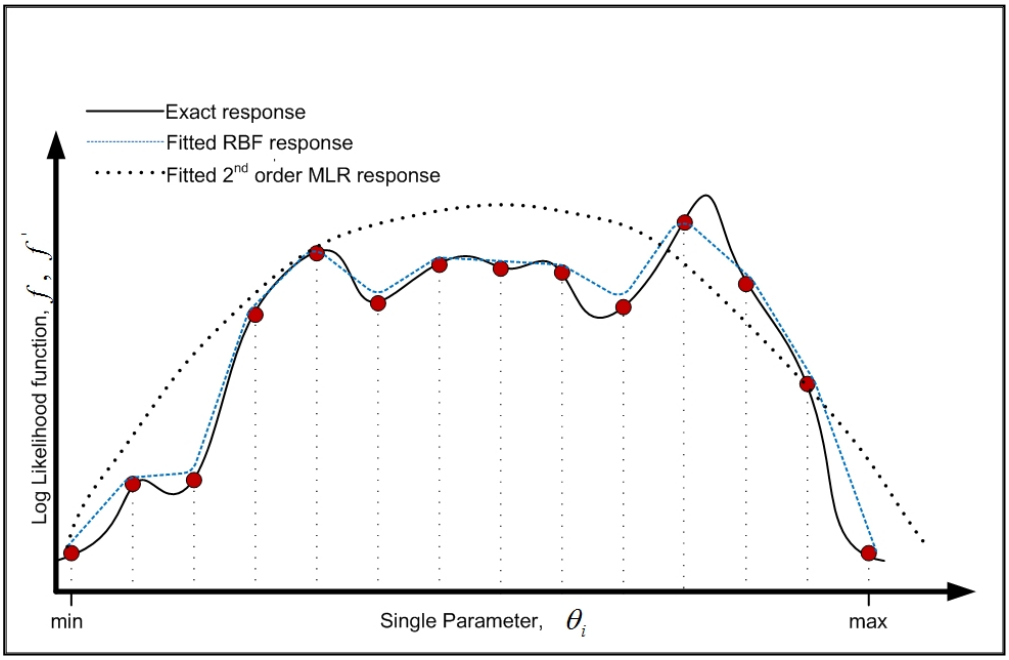

Figure 3: Exact versus fitted RSM function response

Use of RSM for approximating computationally expensive black-box functions has been studied (Gutmann 2001(a); Regis and Shoemaker 2004; Regis and Shoemaker 2005; Regis and Shoemaker 2007; Bliznyuk, Ruppert et al. 2008). However, there is limited research in the use of RSM methods during the calibration and uncertainty assessment of distributed hydrologic models. Further, there has not been any published research on the use of RSM methods specifically for the Bayesian calibration and

uncertainty analysis of distributed hydrologic models. The research in this dissertation is in part an attempt to fill that gap in application. Related to investigating the use of MCMC methods on RSM fitting functions for hydrologic model optimization and uncertainty is the study of factors that affect the performance of such an approach. Before the research presented in this dissertation, there had not been any such performance factors study. As a result of the MCMC-RSM performance factors study, a new routine for response surface approximation was also developed to improve fitting performance. Ultimately, the main contribution of research in this dissertation is to develop and understand methods for automated Bayesian MCMC optimization and uncertainty analysis for use with complex distributed hydrologic modeling.

Problem Statement and Scope of Research

The focus of this research is to respond to challenges arising from the use of distributed or semi-distributed hydrologic models. Distributed hydrologic models can be considered a class of complex environmental models and are finding increasing use in studying the impacts of climate change, water pollution, land use changes, human settlements and watershed management often for large watersheds. Their ability to represent the heterogeneity of physical watershed processes as they are found in geographic space gives them great explanatory power (Moreda, Koren et al. 2006; Hansen, Refsgaard et al. 2007). Unfortunately their great level of detail presents two major challenges. First, they require a much larger input dataset as compared to other environmental models (Liu and Gupta 2007). Such data may not be readily available, which makes model calibration, validation and uncertainty estimation difficult. Second,

distributed models inherently require more parameters than lumped models (Balakrishnan, Roy et al. 2003). This makes them harder and time consuming to calibrate due to the higher dimensional spaces automated routines must search through during analysis. Also, more parameters mean greater parameter uncertainty. Parameter uncertainty contributes to the predictive uncertainty, together with uncertainties in input, output and structure of the model (Kuczera, Kavetski et al. 2006). These two challenges make model calibration, uncertainty and sensitivity analysis very useful for complex environmental models.

While finding the most suitable set of parameters (calibration) is important, environmental modelers often discover that there are usually many combinations of parameter sets that give near-optimal model responses (Beven and Binley 1992; Beven and Freer (2001a); Eckhardt, Breuer et al. 2003; Vogel and Sankarasubramanian 2003; Pappenberger, Beven et al. 2005) . Such behaviour is what Beven and Freer (2001a) termed parameter equifinality and is also known in statistics as the problem of parameter identifiability. Equifinality, a term coined by Bertalanffy (1968), is the principle that in open systems a given end state can be reached by many potential means. Therefore, instead of focusing on calibration, the uncertainty paradigm offers the view that there exists a probability distribution from which acceptable values of parameters can be sampled (Eckhardt, Breuer et al. 2003). That is, parameter values sampled from a specific joint distribution would define a model that makes statistically acceptable predictions. Such a distribution, if it exists, would describe parameter uncertainty (Moradkhani, Hsu et al. 2005; Helton, Johnson et al. 2006). Based on this uncertainty assessment paradigm, calibration is secondary (Vogel and Sankarasubramanian 2003; Pappenberger, Beven et

al. 2005). This research considers both approaches by focusing on parameter optimization and on uncertainty assessment for complex distributed hydrologic models.

To achieve both model optimization and uncertainty assessment of distributed hydrological models, there are strong arguments for using a framework that involves Bayesian MCMC and RSM approaches. First, while other non-Bayesian optimization algorithms are more effective than MCMC, the near-optimal parameter sets obtained from non-Bayesian search methods cannot be used, in a statistically coherent manner, to develop parameter and predictive uncertainty. This is largely because there is no statistically rigorous justification for selecting which parameter sets to include for uncertainty estimation. Bayesian MCMC optimization routines, however, provide output that is statistically valid for use in uncertainty assessment, hence achieving both optimization and uncertainty assessment with the same runs. Second, using Bayesian or non-Bayesian search routines for calibrating and uncertainty assessment of distributed hydrologic models is computationally expensive, thus often requiring infeasible time except when using surrogate models. Lengthy times required for direct model calibration thus supports the use of response surface approximation approaches.

The main concern with Bayesian MCMC optimization and uncertainty assessment using RSM approximation for distributed hydrologic models is that this has not been tried before. The effects parameter insensitivity and interdependence on RSM approximation of, which are common to distributed hydrologic models, are currently poorly understood

Three synthetic and two case study models were used as proof of concept studies to investigate the use Bayesian methods in conjunction with response surface methodology. The developed “Bayesian MCMC on RSM” framework for the

optimization and uncertainty assessment of distributed models was implemented to calibrate a case study Distributed Hydrology Soil-Vegetation Model (DHSVM). The rationale for choosing DHSVM is that (i) the model is fully distributed and significantly computationally expensive, which is ideal for testing the developed framework and (ii) the model has gained wide acceptance and use in key hydrologic research such as climate change scenario analysis (Wiley and Palmer 2008), forest management (Wissmar, Raymond et al. 2004), land use and land cover studies (Thanapakpawina, Richeyb et al. 2007). DSHVM has its roots in the hydrology-vegetation model (Wigmosta et al 1994) that originally modeled canopy interception, evapotranspiration and snow accumulation and melt, as well as runoff generation via the saturation excess mechanisms (Wigmosta, Nijssen et al. 2002). Model inputs include boundary, meteorology, topography, soil, vegetation, stream and road network and initial state geographic information science datasets. Model outputs include main channel flow, water balance and sediment yields. What makes the DHSVM model attractive to modelers is its relative consistency of coding and documentation, clear inputs and outputs formats, flexibility of inputs and its run-time stability (Wigmosta, Vail et al. 1994; Whitaker, Alila et al. 2003; Wigmosta, Nijssen et al. 2002). However, as with other fully distributed modeling paradigms, there has been little to no published routines for automatic calibration and parameter uncertainty analysis for the DHSVM model (Yao and Yang 2009).

Research Objectives and Dissertation Outline

To fully address the problem statement, the four main objectives of this research are to:

i. Assess the feasibility of currently available non-Bayesian calibration frameworks for use with optimization and uncertainty analysis of distributed hydrologic models.
ii. Study the factors that affect the performance of Bayesian Markov chain Monte Carlo optimization and uncertainty analysis of model parameters using response surface approximation (surrogate functions) with applications to hydrologic modeling.
iii. Develop an efficient Bayesian Markov Chain Monte Carlo optimization and uncertainty analysis framework that uses response surface approximation for suitability to distributed hydrologic modeling.
iv. Evaluate the potential of the developed Bayesian MCMC – RSM calibration and uncertainty analysis framework when applied to a test case, in which the Distributed Hydrology Soil-Vegetation Model is used.

The first objective is covered in Chapter 2 (Manuscript 1), titled "A new approach to watershed management practices assessment using the Soil and Water Assessment Tool, SWAT". The second objective is addressed in Chapter 3 (Manuscript 2), titled "A case study examining factors affecting the performance of response surface modeling during Bayesian optimization and uncertainty analysis of hydrologic models". Chapter 4 (Manuscript 3), titled "Bayesian constrained optimization and uncertainty analysis using radial basis random local fitting", covers the third objective. The fourth objective is addressed in Chapter 5 (Manuscript 4), titled "Bayesian calibration and uncertainty analysis of the Distributed Hydrology Soil-Vegetation Model using radial basis functions".

Contribution of Authors and Co-Authors

Manuscript in Chapter 2:
A New Approach to Watershed Management Practices Assessment Using the Soil and Water Assessment Tool, SWAT

Author: Able Mashamba
Contributions: Data collection and collation, model development and assessment, results and manuscript write-up.

Co-author: Lucy Marshall
Contributions: Study proposal and supervision.

Manuscript Information

Able Mashamba and Lucy Marshall

Journal of Soil and Water Conservation

Status of manuscript (check one):

__✓__Prepared for submission to a peer-reviewed journal

___Officially submitted to a peer-reviewed journal

___Accepted by a peer-reviewed journal

___Published in a peer-reviewed journal

Publisher: Soil and Water Conservation Society

CHAPTER 2

A NEW APPROACH TO WATERSHED MANAGEMENT PRACTICES ASSESSMENT USING THE SOIL AND WATER ASSESSMENT TOOL, SWAT

Abstract: We present here a study of watershed management practices using a new systematic approach to scenario modeling with the Soil and Water Assessment Tool, SWAT. Typical scenario analysis involves analyzing model output from a set of parameter values selected to simulate various best management practices (BMPs). This single selection tends to be highly subjective and does not show the sensitivity of the model to the BMP being studied. The scenario analysis approach proposed here provides a less subjective framework for the selection and assessment of a range of parameter values (model scenarios). These are chosen relative to some calibrated and validated baseline scenario to study the various impacts of each selected target BMP.

In this research, we applied SWAT to a case study watershed that included a 45,580 acre irrigation district called Buffalo Rapids. Nine conservation practices were studied using 107 model scenarios that provided response trends and hence showed the criticality and sensitivity of practices and case study SWAT model respectively. Ten parameter value adjustments were used to test model response for most of the BMPs that were studied. Where possible, these parameter value adjustments were within ±25% of the baseline at 5% intervals. Where ±25% of the baseline gave values outside the feasible range of parameters, 10 other values within the feasible space and close to the baseline were selected. Conservation practices studied in this research for which it was not feasible to test as many model adjustments were irrigation delivery, conservation tillage and crop rotation. The main drawback with the new scenario analysis approach is that

using more parameter value adjustments to test one conservation practices requires more model runs.

Keywords: best management practices, scenario analysis, distributed modeling, Soil Water Assessment Tool (SWAT), calibration, parameter sensitivity.

Introduction

Studying the impact of watershed management practices helps in the selection and implementation of effective conservation policies that may minimize water pollution concerns or improve water use and thus promote sustainable watershed productivity. Hydrologic models are increasingly being used to quantitatively study the effects of watershed conservation practices on water quality and quantity (Arnold and Fohrer 2005; Jayakrishnan, Srinivasan et al. 2005). Among the models capable of basin-wide or distributed hydrologic process assessment, the Soil Water Assessment Tool (SWAT) is often cited as the most capable in representing agricultural management practices for water quality studies (Arabi, Frankenberger et al. 2007; Gassman, Reyes et al. 2007). However, the absence of widely acceptable representations of conservation practices in the model and of a systematic scenario analysis procedure continues to undermine the objectivity of any management practice assessment (Arabi, Frankenberger et al. 2007).

Currently, the main approach to watershed management scenario analysis using SWAT involves building, calibrating and validating a baseline model from which parameter values and other data inputs are then adjusted to represent different conversation practice scenarios (Jayakrishnan, Srinivasan et al. 2005; Arabi, Frankenberger et al. 2007; Gassman, Reyes et al. 2007). The baseline model is often set to represent the current case study situation (Arabi, Frankenberger et al. 2007). To model

the different scenarios, Arabi et al (2007) for example first identified target conservation practices for study and then proposed parameter value changes to represent these scenarios. However, the amounts by which parameter values were adjusted to represent the different scenarios remain subjective. Jayakrishnan et al (2005) applies SWAT for scenario analysis, in a limited way, to assess the effects of municipal wastewater and dairy manure pollution on the total maximum daily load in surface water.

A new approach to studying different watershed management scenarios presented in this chapter builds on the general approach by Arabi et al (2007). Firstly this involves building, calibrating and validating a baseline SWAT model that represents an existing case study situation. Next, we identify conservation practices to be studied and the SWAT hydrologic response units, HRUs, in which these practices are to be applied. An HRU is a contiguous area land within which hydrological properties and responses are considered homogenous. HRUs are specified as the smallest homogeneous units of land within a large heterogeneous watershed. As with Arabi et al (2007), we also identify the SWAT parameters and inputs representing the target conservation practices to be studied. However, instead of subjectively determining some parameter and input values that represent specific scenarios, the approach advocated here studies model responses for a range of conservation practice-related parameter values close to the baseline values. For baseline parameters having near limit values, arbitrary values close to the baseline value are chosen for testing; otherwise percent adjustments are used for the range of potential parameter values. As an illustration, say we have targeted a conservation practice that is likely to impact the universal soil loss equation practice factor parameter, USLE_P, which is bound by the physical constraints of [0, 1]. If the USLE_P of the baseline model

is 0.95, we simulate for 10 other arbitrary values of USLE_P close to 0.95 to assess how the model responds to minor changes of parameters from the baseline, as would be brought about by reasonably conservation practice implementation. A positive 25% parameter adjustment would be beyond the theoretical limit of 1.0. Whenever possible, we use a parameter adjustment range of ± 25% around the baseline values with each change being 5% apart thus providing 10 values to be tested for each parameter. This is in keeping with the range of parameter adjustments used by the native SCEUA algorithm in SWAT. For conservation practices that affect many parameters and inputs simultaneously, the adjustments on the different parameters would be made to represent a worsening or improving scenario throughout for each of the 10 positions making the test range. Note that different parameter and or input range schemes could also be used.

Testing model response for a range of parameters or input values eliminates the need to subjectively match one set of values to a specific conservation practice. Instead, using ranges of practice-related parameter values near the baseline reveals the response and sensitivity of a model to different levels of conservation practice implementation. This extra information could help with estimating the costs versus benefits of different levels of conservation practice implementation.

This paper details the processes that analysts could follow from model building to scenario analysis using SWAT. To do this, summaries on methods and materials, sensitivity analysis, calibration, validation, uncertainty and scenario analysis as well as on results and discussions are included.

Methods and Materials

The Soil and Water Assessment Tool

The Soil and Water Assessment Tool, (Arnold, Srinivasan et al. 1998) is a conceptual, continuous time, semi distributed model developed in the early 1990s to assist water resource managers in assessing the impact of management and climate on water supplies and non-point source pollution in watersheds and large river basins. SWAT is the continuation of over 30 years of FORTRAN 90 model development within the US Department of Agriculture - Agricultural Research Service and was developed to 'scale up' past field-scale models to large river basins (Gassman, Reyes et al. 2007).

The model has also been adopted by the U.S. Environmental Protection Agency (USEPA)'s Better Assessment Science Integrating Point and Nonpoint Sources (BASINS) software package (Gassman, Reyes et al. 2007). There are currently over 250 peer-reviewed published articles that report SWAT applications, reviews of SWAT components, or other research that includes SWAT (Gassman, Reyes et al. 2007). The application of SWAT include streamflow calibration and related hydrologic analyses, climate change impacts on hydrology, pollutant load assessments, comparisons with other models, and sensitivity analyses and calibration techniques (Gassman, Reyes et al. 2007).

The main watershed processes that can be simulated in SWAT include climate, automatic weather generation, hydrological processes, plant growth, land use management, channel processes, impoundment processes and the routing of sediments, nutrients, pesticides, bacteria and heavy metals. Further details on the technical theories and assumptions underpinning SWAT are available from the software user manual, Soil and Water Assessment Tool Theoretical Documentation, Version 2005.

Case Study

The case study baseline model was built using weather, elevation, land use, soils, and vegetation data for a 4,623,100 acre lower Yellowstone River watershed in eastern Montana shown in Figure 4. Within the watershed lies a 25,580 acre irrigation district called Buffalo Rapids whose agricultural and irrigation data were used in baseline input management files settings. The Buffalo Rapids irrigation district is located in the Custer, Prairie, and Dawson counties of Montana and parallels 64 miles of the lower Yellowstone River.

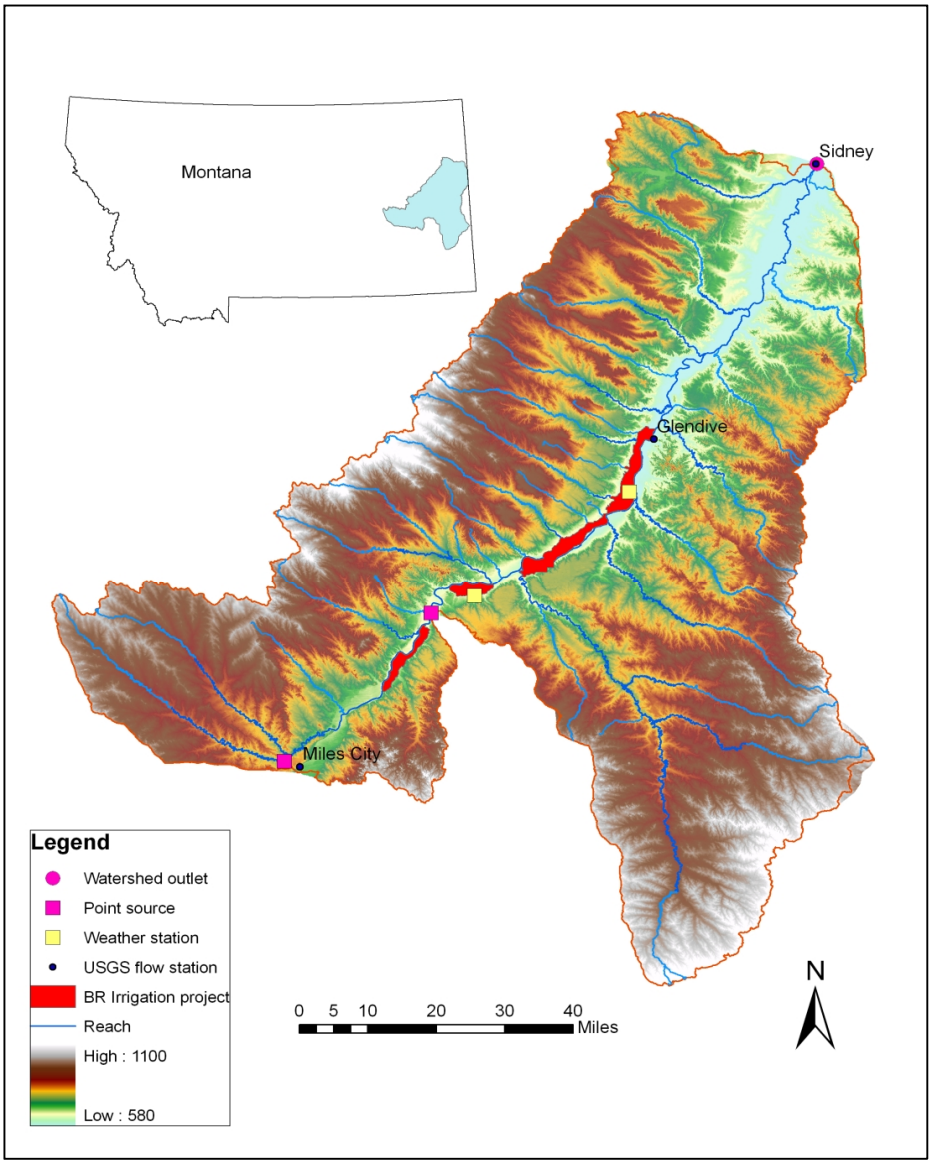

Figure 4: Lower Yellowstone Basin and the Buffalo Rapids Irrigation District

Reliance on irrigation delivery for crop production in the watershed provides an opportunity to model the effects of irrigation and agricultural conservation practices for improved water management. The Buffalo Rapids district includes 155 full-time crop and/or livestock producers with an average of 164 acres per full time farm. Issues of assessment of drinking water, pollution and prevention, and watershed management and conservation are integral to the future agricultural sustainability and economics of the

district and watershed. Three main potential causes of water-resource degradation in the watershed include; (i) irrigation-related nutrient (nitrogen and phosphorous) and sediment loading to the Yellowstone River, (ii) agriculturally-related nutrient and pesticide loading to shallow alluvial groundwater, a primary domestic water supply in the watershed, and (iii) inefficient use of water pumped from the Yellowstone River, which critically reducing in-stream flows during low-flow periods of the year. Thus results of scenario analysis from this research could help farmers and stakeholders in identifying conservation practices and watershed management policies that could help improve water quality and quantity in the watershed and irrigation district.

The large size of the case study watershed meant that the spatial variability of hydrological processes would be significant; thus making the use of spatially distributed modeling preferable. Ideally, a fully distributed model could represent the continuous spatial variability of topography, weather, soil, land use or land cover found in real catchments. However, limiting computing resources, the perceived limited gains in knowledge for doing so (Moreda, Koren et al. 2006), and the scarcity of measured data at higher spatial resolution limits the widespread application of fully distributed models (V´azquez, Feyen et al. 2002). In this paper, the semi distributed SWAT was chosen as a compromise between a fully distributed and a point modeling approach. SWAT also has had widespread published use (Arnold and Fohrer 2005), and has a preferred ability to better represent the most watershed management scenarios (Gassman, Reyes et al. 2007).

Input data for SWAT was prepared using ArcGIS 9.1 Geographic Information System (GIS) software, from the Environmental Systems Research Institute, interfaced with ArcSWAT 1.06 from the University of Texas A&M. (Winchell, Srinivasan et al.

2007). The digital elevation model, channels network shape files and river discharge data were obtained from United States Geological Survey (USGS) online web database. Land use and land cover data came from Natural Resources Information System (NRIS)-United States Department of Agriculture (USDA) online servers. The State Soils Geographic (STATSGO) soils data, from the USDA Soils Conservation Services, came bundled with ArcSWAT 1.06. Water quality data was obtained from USGS National Water Information System and from Montana State University Extension Services, MSU-ES. The United States Department of the Interior-Bureau of Reclamation-Agricultural Meteorology (Agrimet) provided daily weather data from two stations while data on watershed management practices were provided by MSU-ES, Buffalo Rapids project staff, MSU Eastern Agricultural Research Center and USDA Agricultural Research Services, Sidney-Montana.

The model application proceeded by initial sensitivity analysis, automated calibration, split-sample model validation, and model uncertainty analysis via the ParaSol algorithm (van Griensven, Meixner et al. 2006) . Following this, our newly developed approaches to scenario analysis were implemented and studied. The main new scenario analysis approach involves investigating model responses to a set of systematically selected conservation-related parameter values that are close to the calibrated baseline settings. This allows for the study of how management scenarios represented by a range of parameter values would affect model output without the need for subjectively assigning specific arbitrary parameter values to represent specific management scenarios. Model responses for a range of parameter values provide extra information on the sensitivity or rate of change of the output to changes in a specific parameter, hence to

implementation levels of the related conservation practices. Other management practice scenarios that are not represented by parameter value ranges were also studied.

Model Sensitivity Analysis

Sensitivity analysis (SA) involves testing how model output is affected by changes to parameters and or input data (Sieber and Uhlenbrook 2005). Sensitivity analysis is critical because there is often uncertainty about the true values of inputs that should be used for a particular model. Thus by investigating the stability and robustness of a model against random inputs, a picture emerges of how reliable the predictions of a model might be against possible future changes and about which parameters or inputs are critical to the model (Zheng and Keller 2006). Our implementation of the SWAT model commenced with a systematic sensitivity analysis to aid in model calibration and to identify important model parameters.

The automatic sensitivity analysis involved using Latin Hypercube-One factor At a Time (LH-OAT) to sample for up to 41 model parameters within specified boundaries as recommended in the ArcSWAT 1.06 software framework. Sensitivity is expressed by a dimensionless index *I*, which is calculated as the ratio between the relative change of sum of squared residual (SSR) and the relative change of a parameter of a model, $\{(SSR_1 - SSR_0)/SSR_0]\}/\{(\theta_1 - \theta_0)/\theta_0\}$. The subscripts 0 and 1 represent the SSR and the parameter (θ) before and after sampling and adjustment respectively. The ranges $0.00 \le |I| < 0.05,\ 0.05 \le |I| < 0.20,\ 0.20 \le |I| < 1.00,$ and $|I| \ge 1.00$ denote small to negligible, medium, high and very high sensitivity respectively (Lenhart, Eckhardt et al. 2002). The SSR is calculated as $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$; where y_i and $\hat{y}_i$ are the i^{th} observed and

simulated responses (e.g. channel discharge) respectively and the value *n* is the total number of observations used in the sensitivity analysis. Input and calibration data were used from the time period March 11, 1999 to December 31, 2008. The model output used for assessing relative sensitivity were main channel discharge and concentrations of sediment, organic nitrogen, nitrates, total nitrogen and total phosphorous.

For the case study watershed, of the 41 parameters tested, 29 showed an identifiable trend in their scatter plots and had higher relative sensitivities (overall sensitivity ranking) than the remaining parameters as shown in Table 1. Hence, these were chosen for calibration and further study.

Table 1: SWAT model sensitivity analysis parameter ranking

	Parameter	Lower Bound	Upper Bound	Overall Sensitivity Ranking
1	Alpha_Bf, Baseflow alpha factor [days]	0	1	1
2	Surlag, Surface runoff lag time [days]	0	10	2
3	Cn2, Initial SCS curve number II value	-25	25	3
4	Timp, Snow pack temperature lag factor	0	1	4
5	Ch_K2, Channel effective hydraulic conductivity [mm/r]	0	150	5
6	Canmx, Maximum canopy storage [mm]	0	10	6
7	Ch_N, Main channel Manning's n value	0	1	7
8	Esco, Soil evaporation compensation factor	0	1	8
9	Blai, Maximum potential leaf area index	0	1	9
10	Sol_Z, Soil depth [mm]	-25	25	10
11	Smtmp, Snow melt base temperature [degC]	-25	25	11
12	Sol_Awc, Soil Available water capacity [mm H2O/mm soil]	-25	25	12
13	Slope, Average slope steepness [m/m]	-25	25	13
14	Biomix, Biological mixing efficiency	0	1	14
15	Usle_P, Universal soil loss equation support practice factor	0	1	14
16	Epco, Plant uptake compensation factor	0	1	16

Table 1 – Continued: SWAT model sensitivity analysis parameter ranking				
Parameter		Lower Bound	Upper Bound	Overall Sensitivity Ranking
17	Revapmin, Threshold water depth in the shallow aquifer for revaporation [mm]	-100	100	17
18	Sol_K, Soil saturated hydraulic conductivity [mm/hr]	-25	25	18
19	Slsubbsn, Average slope length [m]	-25	25	19
20	Rchrg_Dp, Deep aquifer percolation fraction	0	1	21
21	Gw_Delay, Groundwater delay [days]	-10	10	22
22	Nperco, Nitrogen percolation coefficient	0	1	23
23	Sol_Alb, Moist soil albedo	-25	25	24
24	Gw_Revap, Groundwater revaporation coefficient	-0.036	0.036	25
25	Usle_C, Minimum universal soil loss equation cover factor	-25	25	26
26	Phoskd, Phosphorus soil partitioning coefficient	100	200	27
27	Spcon, Linear re-entrainment parameter for channel sediment routing	0	0.01	28
28	Ch_Cov, Channel cover factor	0	1	29
29	Ch_Erod, Channel erodibility factor	0	1	30
30	Spexp, Exponetial re-entrainment parameter for channel sediment routing	1	2	31
31	Pperco, Phosphorus percolation coefficient	10	18	32
32	Gwno3, Concentration of nitrate in groundwater contribution [mg N/L]	0	10	33
33	Gwqmn, Threshold water depth in the shallow acquifer for flow [mm]	-1000	1000	33
34	Sftmp, Snowfall temperature [degC]	0	5	33
35	Smfmn, Melt factor for snow on December 21 [mm H2O/decC-day]	0	10	33
36	Smfmx, Melt factor for snow on June 21 [mm H2O/degC-day]	0	10	33
37	Sol_Labp, Initial labile phosphorus concentration [mg/kg]	-25	25	33
38	Sol_NO3, Initial nitrate concentration [mg/kg]	-25	25	33
39	Sol_Orgn, Initial organic nitrogen	-25	25	33
40	Sol_Orgp, Initial organic phosphorus concentration [mg/kg]	-25	25	33

Table 1 – Continued: SWAT model sensitivity analysis parameter ranking				
Parameter		Lower Bound	Upper Bound	Overall Sensitivity Ranking
41	Tlpas, Temperature lapse rate [degC/km]	0	50	33

Table 1 identifies the ranking of parameters that were sensitive in simulating main channel discharge. Note that channel discharge is greatly influenced by parameters that are related to changes in watershed management practices (e.g. parameters related to channel roughness or infiltration capacity). A great potential then exists to improve water quantity targets in the channels by implementing suitable conservation practices.

Sediment yield was sensitive to the linear re-entrainment factor for sediment routing, channel cover factor, channel erodibility factor, the universal soil loss equation support practice factor and the universal soil loss equation cover factor. As for discharge, the sensitive sediment yield parameters are mainly related to watershed management practices. We expect the implementation of conservation practices to control sediment creation and transport to have a positive effect on our case study.

Nitrogen and Phosphorus yields responded to their respective percolation coefficients as well as the deep aquifer percolation fraction. This implies that, for the case study, nutrients are carried to the channels mainly by infiltration into the shallow aquifer and then into the streams by return flow; and not by surface runoff. Conservation practices to reduce infiltration, such as avoiding flood irrigation, conservation tillage and vegetative cover can be predicted to positively reduce nutrient loadings into the channels in the case study watershed.

Model Calibration

Calibration involves selecting model parameter values, within the feasible parameter space, so as to minimize some measure of error, such as SSR, between model output and observation. For a model with many parameters, automated global search algorithms have to be used to find parameter values which give the closest fit between output and observation. Typically, modelers use prior knowledge about actual systems to manually adjust parameter values before or after automatic calibration. Manual adjustment of parameter values before automatic calibration could help reduce the time to find, as well as possibly improve the quality of, the optimum. Adjusting parameter values after automatic optimization might be appropriate for making the model structure more realistic, at the expense of its predictive accuracy.

For the case study watershed, sensitivity analysis showed that the model is sensitive to 29 of the 41 model parameters tested (Table 1). Insensitive parameters were then fixed to suitable mean values. For the sensitive parameters, sensitive ranges defined based on sensitivity analysis to provide the search space for model calibration. No prior knowledge of parameter values was assumed so manual calibration was omitted. The Shuffled Complex Evolution-Uncertainty Analysis, SCE-UA (Duan, Gupta et al. 1993), algorithm was used to automatically calibrate the model, first for daily discharge and then for the 4-days-per-month concentrations of sediment, organic nitrogen, nitrate, total nitrogen and total phosphorous. The year 1999 was used as a warm-up period, and calibration was for the period January 1, 2000 to October 31, 2007. After calibration, model simulations were assessed via the Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe 1970), Equation 2.1.

$$NSE = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (2.1)$$

Where y_i and $\hat{y}_i$ represent the i^{th} measured and simulated response, such as channel discharge, for an analysis period with *n* time steps. The $\bar{y}$ is the average of the observed (measured) response for the entire period of analysis. Daily discharge calibration gave a high Nash Sutcliffe efficiency of 0.951.

Model Validation

Split-sample model validation involved comparing simulated discharge to observed discharge (m^3/s) for the Yellowstone River at Sidney, Montana, for the period 1 November 2007 through 31 December 2008. Due to the scant availability of water quality data, split-sample validation of the model based on water quality predictions could not be meaningfully performed. Results for the split-sample validation process based on main channel discharge are shown in Figure 5.

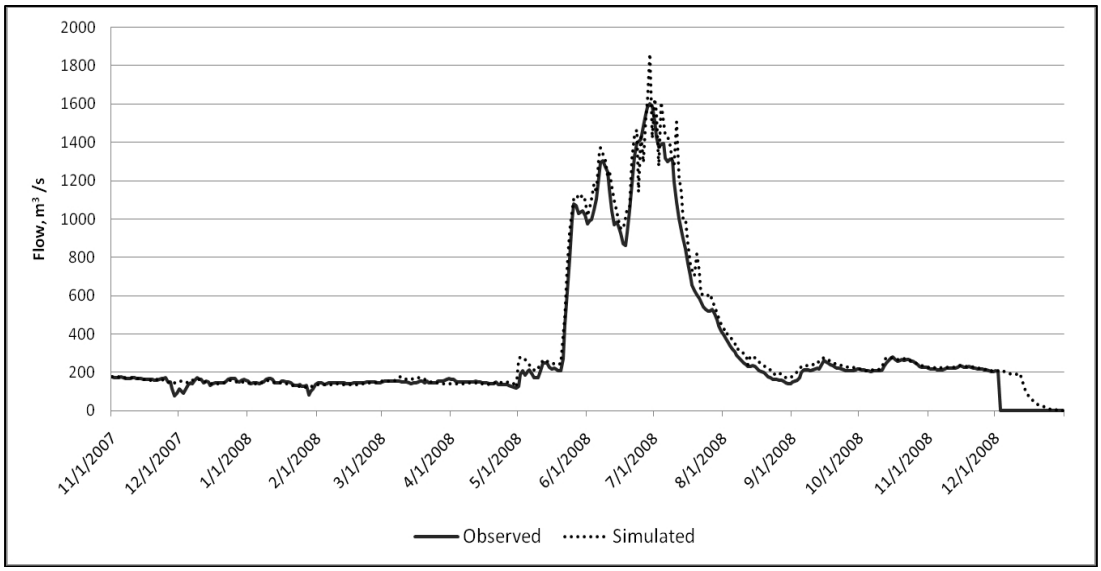

Figure 5: Observed and simulated main channel discharge for the validation period, Oct 2007-Dec 2008

The observed main channel flow of zero is due to the Yellowstone River freezing at Sidney, the outlet station, in December of 2008. Compared to the years used in calibration, the validation period, especially the 2008 season, received above normal precipitation. Even then, the model managed to have an NSE of 0.968.

Uncertainty Analysis

Quantifying the predictive uncertainty of a model is critical in assessing the risk or reliability associated with model-based decision making as well as in improving the objectivity, creditability, and hence acceptability of simulation output (Montanari 2007; Sivakumar 2008). A distinction between error and uncertainty as proposed by Huard and Mailhot (2006) is that error refers to the difference between true values (assumed known) and measurements or predictions while uncertainty is a probabilistic measure of our incomplete state of knowledge about the true values.

SWAT 2005 implements two uncertainty analysis algorithms, the Parameter Solutions (ParaSol) and the Sources of UNcertainty GLobal Assessment using Split SamplES (SUNGLASSES) algorithms (van Griensven and Meixner 2006). ParaSol is a global optimization and uncertainty assessment algorithm that relies on results from the SCEUA evolutionary algorithm. ParaSol records all the modified SCE-UA optimization parameter sets and uses either a chi-squared or Bayesian threshold to select which sets to use in building uncertainty bounds, similar to the GLUE method (van Griensven and Meixner 2006). On the other hand, SUNGLASSES is like ParaSol except that it uses an evaluation period in addition to the calibration period so as to detect other sources of uncertainty besides parameters (van Griensven and Meixner 2006). Since we are studying conservation practices which are strictly related to SWAT model parameters, we used the

ParaSol algorithm for uncertainty analysis. The scant 4-days-per-month water quality data available meant we could only meaningfully assess the uncertainty associated with the 21 sensitive parameters that affect daily discharge.

Parameter uncertainty analysis indicated that the 97.5% confidence interval for the main channel average discharge for the years 2000 to 2008 is 0.43m^3 /s. There is a positive 0.36 correlation between the observed discharge and the size of the confidence interval. This means that the greater the observed discharge being simulated the greater the predictive uncertainty. Table 2 shows the 97.5 % confidence intervals for sensitive parameters that control main channel discharge.

Table 2: 97.5 % Confidence interval values for flow parameters

Parameter	97.5% CI min	97.5% CI max	Baseline values
Alpha_Bf, Base flow alpha factor	0.9897	1.0000	0.9990
Biomix, Biological mixing efficiency	0.0000	0.3072	0.0923
Blai, Maximum potential leaf area index	0.3242	1.0000	0.5425
Canmx, Maximum canopy storage	0.0000	3.4770	0.7828
Ch_K2, Channel effective hydraulic conductivity	42.1900	57.2930	50.2570
Ch_N, Manning's n-value for main channel	0.0499	0.0525	0.0519
CN2, Initial SCS CN II value	-25.0000	-15.7010	-23.8950
Epco, Plant uptake compensation factor	0.2083	1.0000	0.9091
Esco, Soil evaporation compensation factor	0.0000	0.2379	0.0181
Gw_Delay, Groundwater delay	-9.2371	10.0000	4.3954
Gw_Revap, Groundwater revaporation coefficient	0.0360	0.2287	0.0380
Gwqmn, Threshold shallow aq. water depth for flow	-1000.000	-226.930	-1000.000
Revapmin, Threshold shallow aquifer water depth for revaporation	-100.0000	-23.6920	-73.1750
Slope, Average slope steepness	-25.0000	5.2005	-25.0000
Smtmp, Snowmelt base temperature	-25.0000	-2.3227	-15.5070
Sol_Alb, Moist soil albedo	0.5367	25.0000	19.0100
Sol_Awc, Soil available water capacity	18.3770	25.0000	23.9500
Sol_K, Saturated hydraulic conductivity	-25.0000	-18.5570	-22.3680
Sol_Z, Soil depth	-3.8554	25.0000	18.6280
Surlag, Surface runoff lag time	0.0000	0.0003	0.0002
Timp, Snow pack temperature lag factor	0.2765	1.0000	0.8588

Due to how the SWAT model variably adjusts the parameters with space, the values in Table 2 are mostly percentages for adjusting the actual parameter in the different delineations. Biomix, Ch_K2, Ch_N and CN2 are not just sensitive but are also determined by the conservation practice in use and thus are important in scenario analysis. The low values for surface runoff lag times correspond to low surface runoff for our snowmelt-dominated watershed. Instead, water to the channels is mostly by shallow aquifer return flow, as shown by the low Gwqmn values. Figure 6 shows the predictive

uncertainty of the SWAT model for discharge of the Yellowstone River between the years 2000 and 2008.

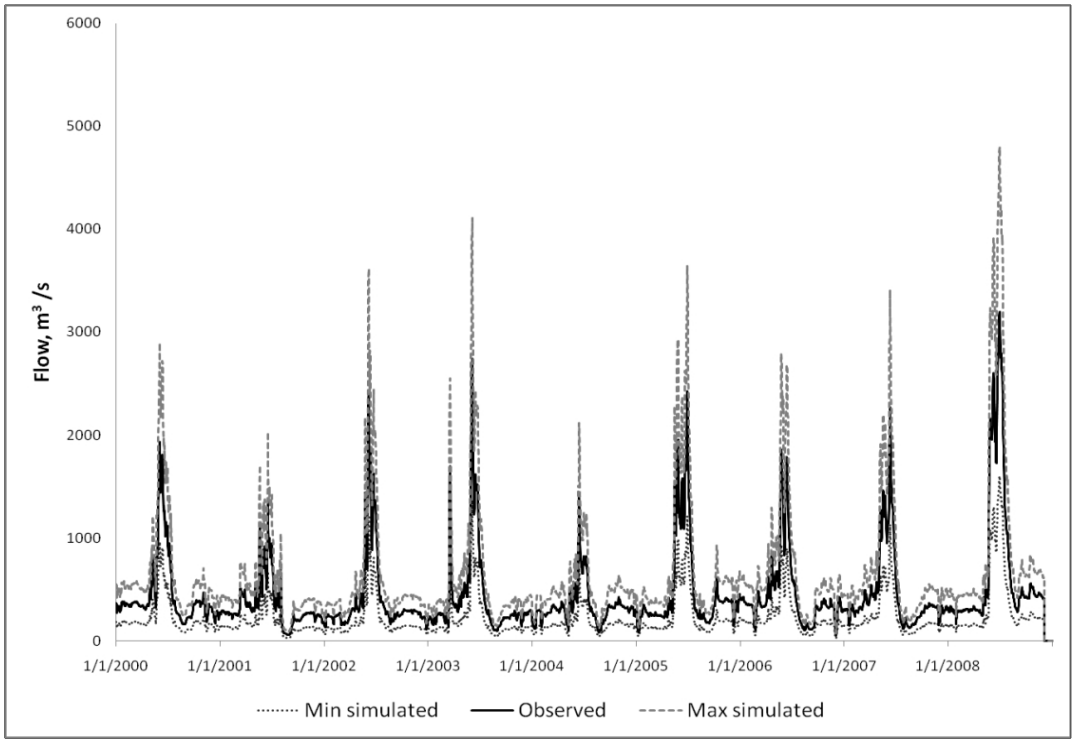

Figure 6: Predictive uncertainty of the Yellowstone River flow

The predictive uncertainty in Figure 6 is related to the parameter uncertainty of Table 2. The 97.5% CI values of the parameters in Table 2 above allow for a combination of parameters that give the maximum and minimum predicted discharge as in Figure 6.

Conservation Practices Scenario Analysis, A New Approach

Watershed conservation practices (a.k.a best management practices, BMPs) scenario analyses involved identifying sensitive parameters associated with conservation practices to be investigated, building a baseline scenario used to assess relative effects of alternatives and finally building and testing alternative scenarios. Discharge and

concentrations of sediment and nutrients in the main channel were used as the target responses of the model during scenario analysis. Crop yield was the other model output that was monitored to evaluate the desirability of a scenario.

In implementing our approach, important agricultural conservation practices were first selected for representation with the SWAT2005 model of the case study. The selection was based on the applicability of the practice to the target area, the sensitivity of the parameters representing the practices, and in consultation with the irrigation district and MSU Water Quality Extension Service. Several USDA-NRCS conservation practices (http://www.nrcs.usda.gov/tecnical/standards/nhcp.html) were not included in the analysis as they were unlikely to be used in the case study watershed and had not been identified by the Buffalo Rapids irrigation district. For instance, contour farming, strip-cropping and parallel terraces were excluded for investigation in this analysis because the case study watershed has gentle slopes (96.7% of the land has slopes of less than 7%) and the surface runoff is predominantly generated by snow-melt. That is, runoff from snow-melt is not expected to be as significantly erosive as that by rainfall, and hence there is no need for conservation practices that reduce erosion.

Table 3 shows the 9 selected USDA-NRCS conservation practices that were considered for conservation practice scenario analysis in the case study. These practices were considered relevant and possibly useful to the case study watershed based on prior knowledge of the area and on modeling data collected about the area.

Table 3: Selected BMPs for Scenario Analysis

	Selected USDA-NRCS Conservation Practices (BMPs)	Target improvements
1	Channel stabilization	Reduces channel erosion and deposition
2	Conservation crop rotation	Reduces field soil erosion and requirement of pesticide and nutrient. Improves moisture efficiency
3	Conservation tillage (no tillage)	Reduces field soil erosion and improves moisture efficiency
4	Lined waterways	Reduces erosion by channel flows
5	Field border/Strips	Reduces field soil erosion
6	Irrigation delivery system (gated pipes, wheel lines, center pivots)	Reduces infiltration of water and nutrients in loamy/sandy soil and reduces irrigation water demand
7	Nutrient management	Reduces nutrient available for infiltration and wash-away and nutrient requirement
8	Stream-bank and shoreline protection; use of vegetation and structures	Reduces channel/gully erosion
9	Critical area and range planting	Reduces soil erosion and surface runoff

Results from sensitivity analyses of 41 parameters of the case study model were used to select the parameters to be adjusted during conservation practice scenario analyses. Of the 29 sensitive parameters, those that are associated with conservation practices (Santhi, Srinivasan et al. 2006) are listed in Table 4. This list includes sensitive parameters for discharge, sediment and nutrients modeling that are determined by conservation practices being studied.

Table 4: Sensitive parameters and operation associated with conservation practices

Sensitive parameters associated with conservation practices	Management file parameters and operations associated with conservation practices
Biomix, biological mixing efficiency	FILTERW, field filter width
Ch_K2, channel hydraulic effective conductivity	CNOP, SCS Curve number II for management operations
Ch_N, manning's channel roughness	Irrigation operations
CN2, initial SCS Curve number II value	Irrigation amounts
Ch_Cov, channel cover factor	Fertilizer operations
Ch_Erod, channel erodibility factor	Fertilizer amounts
USLE_C, minimum USLE cover factor	Crop rotation
USLE_P, USLE support practice factor	Tillage operations

Having identified sensitive parameters associated with target conservation practices, the next step was to develop a baseline management scenario from which the relative effects on discharge, sediment and nutrient concentrations of alternative scenarios would be measured. The sensitive parameters in Table 4 used baseline values obtained from calibration for flow and for concentrations of sediment, organic nitrogen, nitrate, total nitrogen and total phosphorous in the main channel. From field work by MSU-ES, baseline management operations for the model included gated pipe irrigation, nitrogen and phosphorous fertilizer application and tillage operations. On the other hand, crop rotation and the use of center pivots and wheel lines for irrigation delivery were considered as alternative scenarios to be investigated together with other conservation practices.

The case study SWAT model had 70 agricultural hydrologic response units (HRUs) and 145 range land HRUs. Table 5 (Kirkpatrick, Browning et al. 2006) summarizes the cropping data used to build the baseline scenario for the agricultural

hydrologic response units. Except for irrigation, which was implemented only for Buffalo Rapids, other agricultural management scenarios were applied to all the agricultural HRUs; a total area roughly one fifth of the entire watershed.

Table 5: Baseline Agricultural Scenario Data for Case Study

Agricultural Operation	Major Buffalo Rapids Crop					
	Sugar beets	Alfalfa	Barley	Wheat	Corn	Beans
Pre-planting tillage	2 weeks before planting	2 weeks before planting	2 weeks before planting	2 weeks before planting	2 weeks before planting	2 weeks before planting
Start of growing season	4/30	5/15	4/23	5/2	5/11	5/16
N fertilizer date	Once At planting	n/a	Once at planting	Once at planting	50% at planting, the bal at 6/15	n/a
N kg/ha	235	0	55	65	90	0
P fertilizer date	4/30	5/15	4/23	5/2	5/11	5/16
P kg/ha	56	40	30	35	40	35
Secondary tillage	6/01	No till	No till	No till	6/15	No till
Harvest and kill date	9/25	8/27 of year 2	7/30	8/7	10/22	9/6
Heat Units	1925	5000	1400	1475	1605	1190
Available crop acreage data	1530.4	978.8	875.9	783.4	704.8	424.2
% crop acreage for available data	28.9	18.5	16.5	14.8	13.3	8.0
Approximate area under irrigation, km2	57.80	37.00	33.00	29.60	26.60	16.00
Area under cultivation, watershed-wide, km2	1319.74	844.82	753.48	675.85	607.35	365.32

Main sources of data for Table 5 were the annual cropping reports for 1996-2008 from USDA Montana-Sidney office for the entire Lower Yellowstone watershed, the MSU-Extension Services field data as well as Buffalo Rapids project sources. Crop

acreages, totaling 5,297.1, are annual averages for a part of the Buffalo Rapids irrigation project, where such data was available. However, the entire Buffalo Rapids project irrigates 22,719 acres (http://www.usbr.gov/projects/Projects.jsp?proj_Name=Buffalo +Rapids+Project). The total area under irrigation for the baseline model is 49421 acres, slightly more than double the size of the Buffalo Rapids irrigation project (MSU-ES). This includes irrigation acreages for individual or private irrigation schemes scattered throughout the watershed.

The total agricultural acreage for the entire case study watershed is 1,141,100. Hence most of the agriculture (1,091,679 acres) is dryland cropping. There are three main varieties of wheat cultivated in the irrigation district; spring, winter and durum wheat. For simplification, winter wheat (planting date 10/10, harvesting date 7/23 of the following year) is not being represented in the planting dates of the "wheat" crop in Table 5. This is because winter wheat is largely grown under dryland cultivation thus would not help with the investigation of irrigation scenarios.

Nitrogen and phosphorus fertilizer application rates and secondary tillage schedules for the different crops were as recommended by MSU-Extension Services' staff (Jim Bauder, pers. comm.). Each crop's heat units were calculated from standard formula using data on planting to harvesting periods, maximum and minimum daily temperatures, and the crops' optimal growth and base temperatures. Crop allocation to the 70 agricultural hydrologic response units (HRUs) in the watershed was randomized, while matching percent crop acreage to percent area of allocated agricultural HRUs.

Once a baseline scenario was built using the best available data for the approximation of the current situation in the watershed, the final step was to select, build

and run alternative management scenarios. The four main approaches to scenario analysis were (i) use of crop rotation, (ii) changes to irrigation delivery systems (and hence irrigation amounts), (iii) use of zero tillage and (iv) the systematic adjustments of sensitive parameters associated with target conservation practices.

Three irrigation delivery systems that were investigated are the gated pipes, wheel lines and center pivots. These are the main systems currently in use in the district. Gated pipe irrigation delivery, the most prevalent system in the case study watershed was used as the baseline setting. Table 6 shows the average irrigation management scenarios and schedules that were investigated.

Table 6: Irrigation scenarios for study (Kirkpatrick, Browning et al, 2006)

	Gated pipes (baseline)	Wheel lines	Center Pivots
Start of irrigation season	5/01	5/01	5/01
Frequency	2 - 3 weeks	1wk 3days - 2 weeks	5 days
Amount per application (mm)	152	88	35
Total Amount (mm)	760	528	460
End of irrigation season	8/25	8/25	8/25

In consultation with MSU-ES experts, the irrigation amounts were assumed to be largely determined by the irrigation system in use and the irrigation schedules used. The types of crops do not significantly determine the amount of water that will be delivered to the fields on a particular irrigation date. Irrigation begins early May and water delivery to the fields is staggered according to sections and farms to reduce demand on any specific day. For the period of simulation, the observed minimum daily average main channel flow was 28.60 m^3/s, in late August of 2001. To prevent zero or negative flow, baseline

irrigation schedules for the model were devised such that the main channel flow does not fall below this minimum average discharge while ensuring that the channel provides the full daily irrigation demand. To study the relative effects of crop rotation, the baseline scenario did not include crop rotation. Sugar beets were rotated with barley and wheat while beans were rotated with corn in the alternative scenario. Alfalfa is typically rotated with pasture bi-annually; this rotation was not investigated to limit modeling complexity. Another pertinent scenario that was investigated was the effect of zero tillage (conservation tillage) on main channel discharge.

Finally, quantitative alternative scenarios were built from the baseline scenario by using the new approach proposed in this paper. This involved making and testing "one factor at a time" adjustments to the sensitive conservation-related parameters from their baseline values. Ten adjustments were made and tested for each of the numerical parameters in the agricultural HRUs of the case study. The adjustments were for a range of ±25 percent of the baseline values at 5% steps, where percent changes would not fall out of the feasible parameter limits. This was to ensure parameters remained within reasonable limits while we investigate how the degree of intervention might impact the results. However, the baseline values for Biomix, FILTERW, USLE_P, Ch_Cov and Ch_Erod were close to feasible limits, hence the ten alternative scenario values were selected based on recommendations of typical values from SWAT 2005 manual and also as guided by sensitivity analysis results. The results and discussions of these scenarios are given in the results and discussions below.

Results and Discussions

Baseline Scenario

Figure 7 and 8 show the yearly and monthly trends for discharge, concentrations of sediment and yields for total nitrogen and total phosphorus for the baseline model scenario.

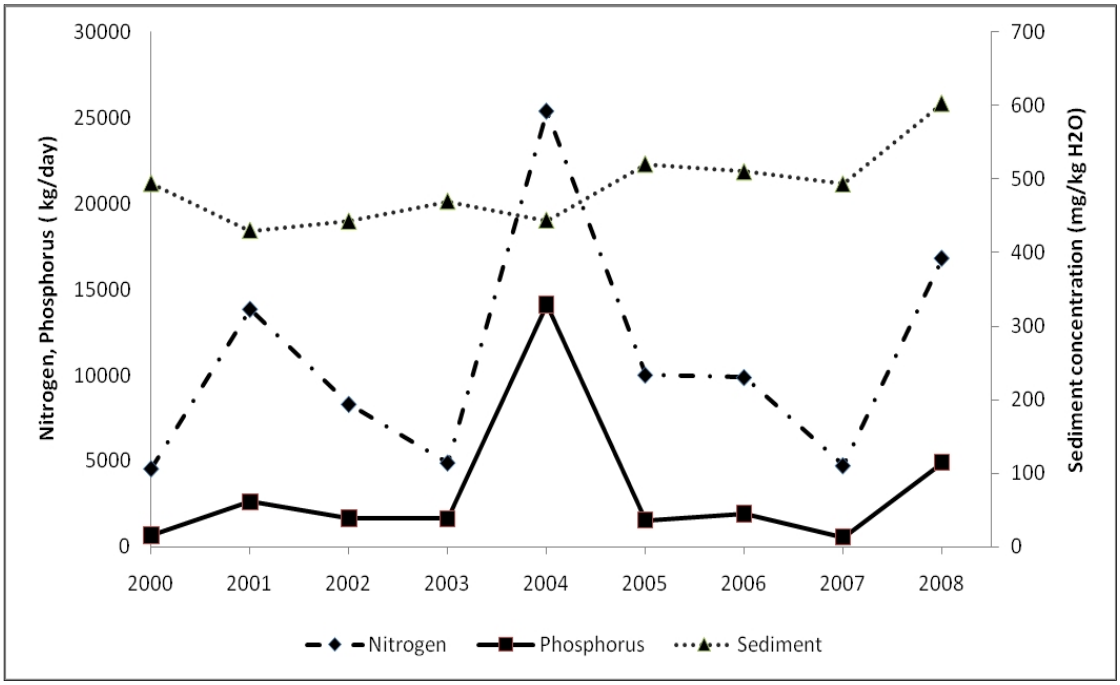

Figure 7: Yearly average nitrogen, phosphorus and sediment for baseline scenario

Figure 7 shows the average yearly trend for sediment to be increasing for the years 2000 to 2008. The close similarity between the average yearly trends of the simulated total nitrogen and phosphorus may suggest that the two nutrients are mainly from fertilizer application. The monthly average output for the baseline scenario, Figure 8, shows that simulated sediment concentration is highest in the months of May and June, which happens to be when snowmelt induced discharge peaks.

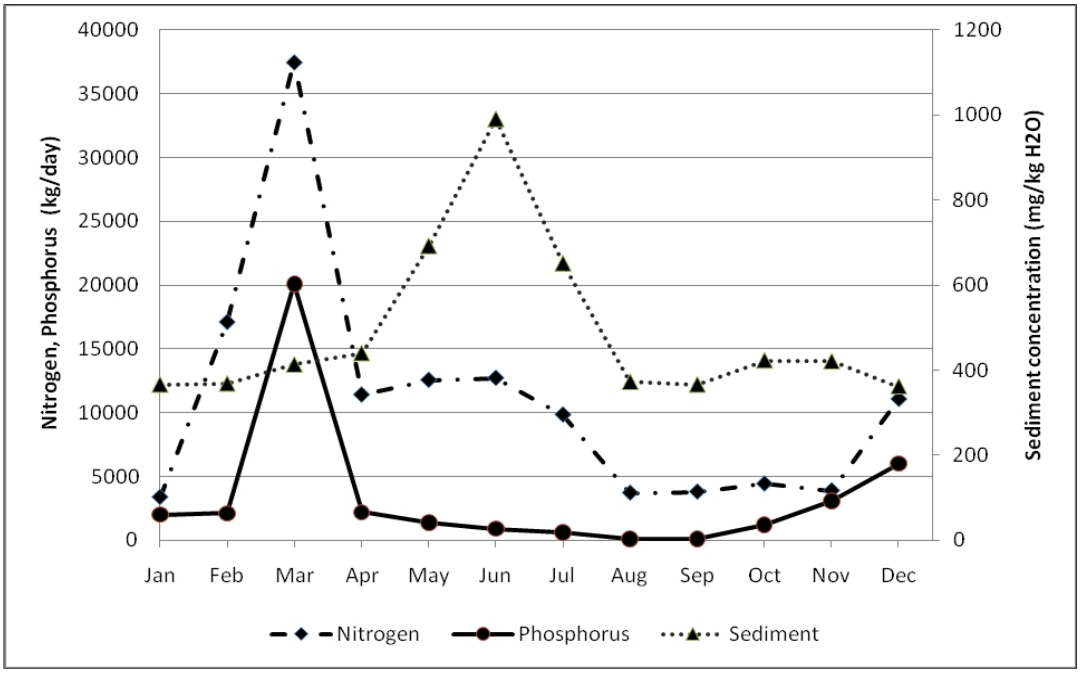

Figure 8: Monthly average nitrogen, phosphorus and sediment for the baseline scenario

Figure 8 shows the highest jump in predicted nutrients yields in March. Since this jump coincides with the beginning of the planting season, when fertilizer application is also at its highest, this could be confirming that nutrients yields in the main channel of the watershed are due to fertilizers. In Figure 8, total nitrogen remains relatively high until August mainly because of continued application during the rest of the plant growth period and also from legumes' nitrogen fixing. The rise in nutrient in winter might be due to a combination of an increase fallen foliage and decreased biological activity.

Irrigation Scenarios

Figure 9(a) shows discharges, sediment concentrations (9b), and yields of total nitrogen (9c) and total phosphorus (9d) for the gated pipes (baseline), wheel lines and center pivots irrigation scenarios. The values are for simulated daily output for the period

2000 to 2008, with 1999 omitted as the model warm-up year. The maximum outliers have been are not represented in Figure 9 for readability.

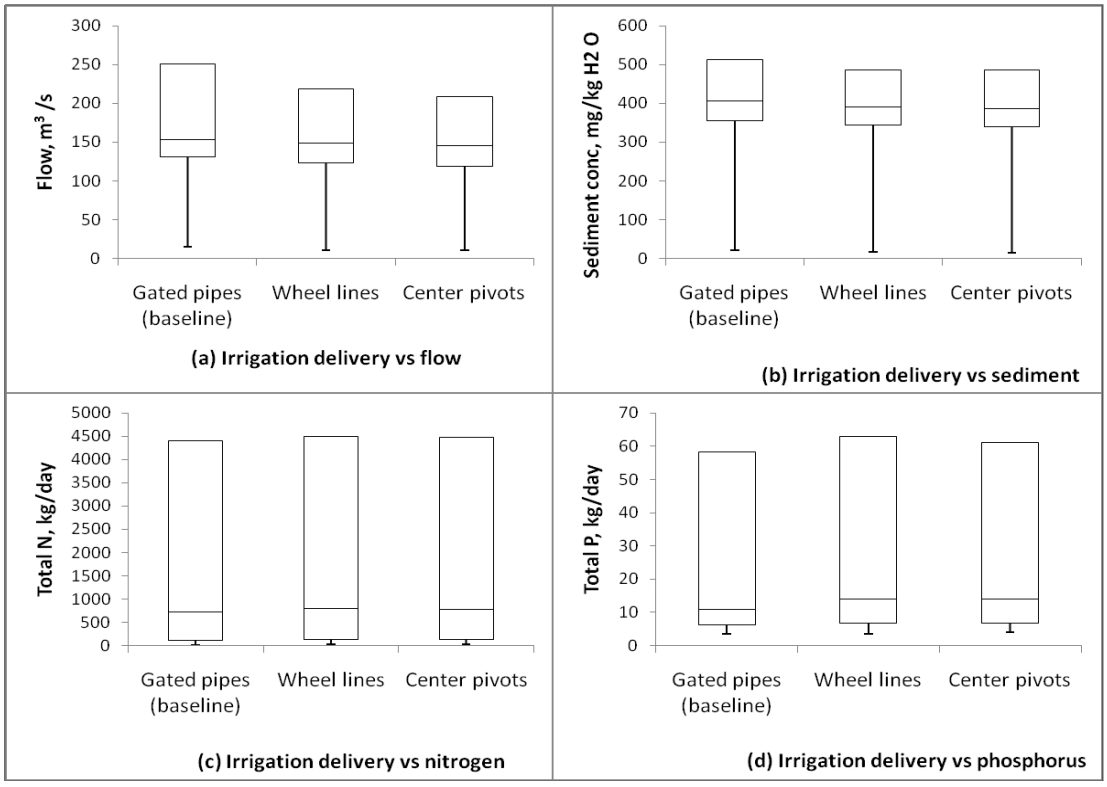

Figure 9: Irrigation delivery scenarios

From gated pipes to center pivots, the amount of irrigation water per application decreases while the frequency increases as given in Table 6. Although the total amount of water applied at the end of the irrigation season decreases with each scenario along the x-axis of Figure 9(a), counter-intuitively, the average main channel discharge also decreases. We theorize this is due to larger amounts of irrigation water per application with lower frequencies that allow for more infiltration into the shallow aquifer and hence more return flow. Lower amounts of irrigation at higher application frequencies encourage more water loss from the watershed due to evapotranspiration. However, the

decrease in discharge from wheel lines to center pivots falls within 97.5% confidence limits for wheel lines and is thus not differentiable from random error (Figure 9a).

The main channel sediment concentrations of Figure 9(b) also correspond to the simulated discharges of Figure 9(a). Hence Figure 9 (a and b) shows that the baseline irrigation scenario results in higher flows and more sediment concentration in the main channel than other irrigation schemes. Figure 9(c) and 9(d) show that the baseline irrigation scenario produces the lowest yields of total nitrogen and total phosphorus in the main channel. This is unexpected since the baseline scenario also has the most irrigation amount per application and the highest total irrigation water quantity. One would expect that the higher return flows inferred to explain Figure 9(a) would also result in higher infiltration of nutrients into the main channel. To explain Figure 9(c) and 9(d) without contradicting 9(a), it is likely that the baseline irrigation water leeches the nutrients into the shallow aquifer such that surface runoff washes away less of the nutrients into the channels. Conversely, light application of irrigation water may leave more nutrients in the top soils where surface runoff from precipitation can later wash them into the channels.

Tillage and Rotation Scenarios

Compared to the baseline, which had pre-planting, and secondary tillage and no crop rotation, the no-tillage and rotation scenarios did not produce significant differences in main channel flow. However, Figure 10(a) shows that practicing conservation tillage (no till) reduces average yearly sediment concentrations in the main channel by 2%, while crop rotation has no effect as compared to the baseline.

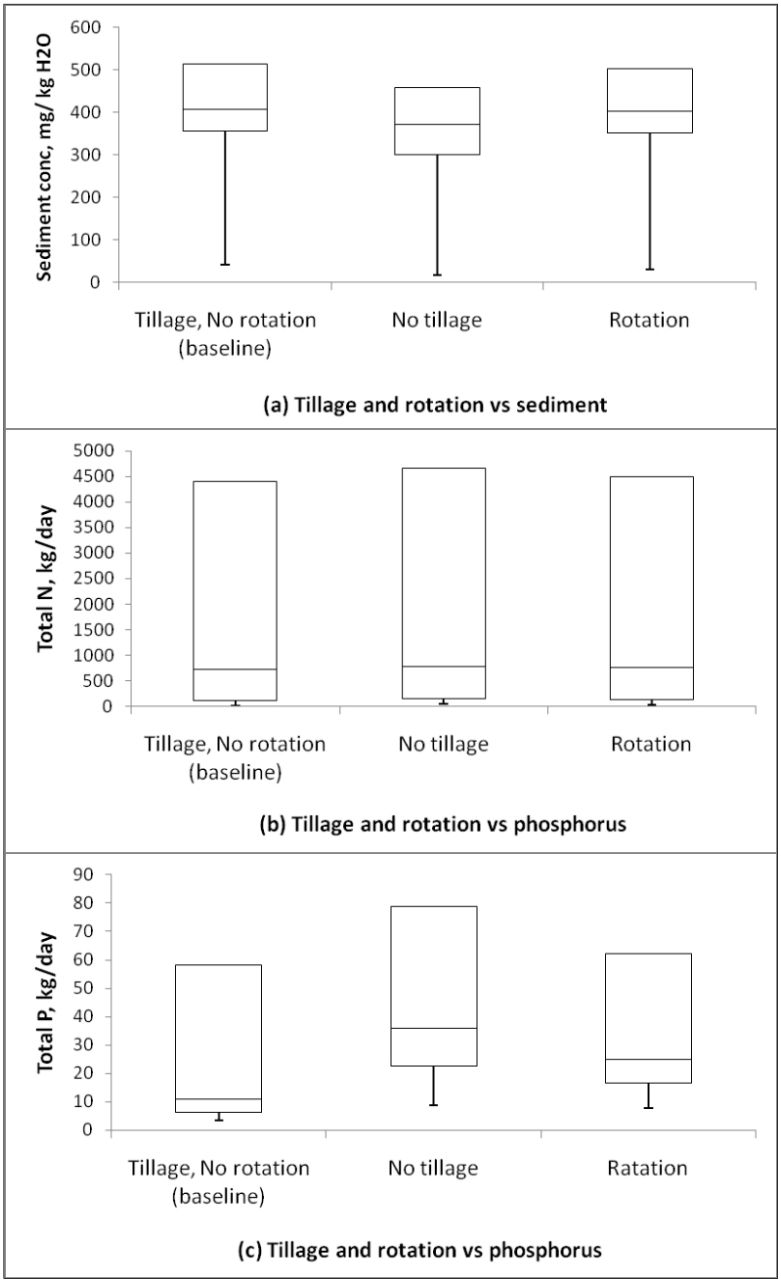

Figure 10: Tillage and crop rotation scenarios

Figure 10(b) and 10(c) show that no-tillage increases the total nitrogen and total phosphorus yields in the main channel by about 9% and 17% respectively. In this simulation, crop rotation increases total nitrogen and total phosphorus yields by about 1% and 12% respectively. The increases of nutrients yields for crop rotation are likely due to

the two-year rotation used resulting in every other year being assigned higher acreages for higher nutrient-demanding crops than for the baseline scenario. Depending on the actual rotation scheme used, lower nutrients yields can also be achieved for the same reason. While zero tillage reduces sediment in the main channel, it may not be desirable to have the higher percentage increases in nutrients loadings that result.

Biological Mixing Efficiency Scenarios

Biological mixing of soil is desirable and often increases when conservation tillage is practiced. The activities of earthworms and other soil biota improve soil hydraulic conductivity, aeration, nutrient mixing, and uptake of nutrients by plants, etc, and thus improve crop growth. Biomix is the model parameter that accounts for the extent of biological mixing for the watershed soils within the SWAT model. The Biomix efficiency parameter of the case study was calibrated at 0.0018, a very low value compared to the recommended value of 0.20. This might suggest that the watershed has very low biological mixing of soil. Ten Biomix values were selected to adequately explore the sensitivity of the model to this parameter and the results are presented and discussed below.

First, Biomix does not affect flow or sediment yields in the case study. However, Figure 11(a) and 11(b) show that greater biological mixing efficiencies significantly reduce main channel yields of total nitrogen and total phosphorus. The baseline values are shown as triangles.

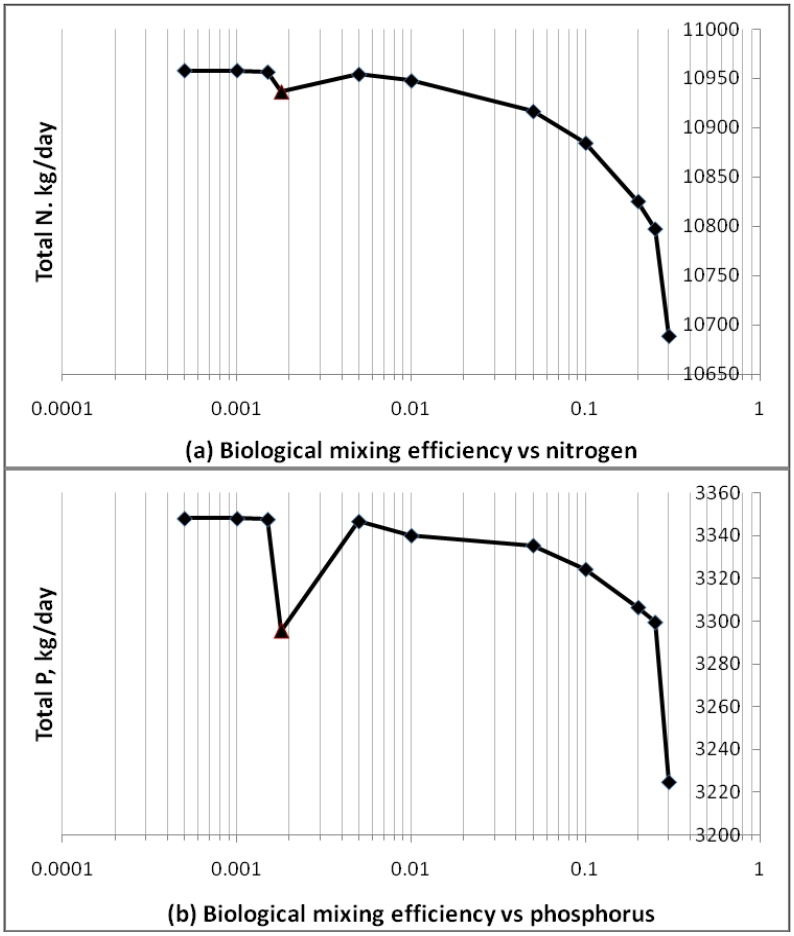

Figure 11: Biological mixing efficiency Vs nutrients scenarios

It should be noted in Figure 11 and most of the subsequent figures that baseline values are also generally the local optimum for the simulation model. For instance in Figure 11, the responses of the model are locally ideal around the calibrated Biomix of 0.0018, where both total nitrogen and total phosphorus values are relatively lower than the neighborhood. The proposed explanation for the paradoxical "off-trend local optimizations" at calibrated baseline values is that they represent the degree to which the effect of a parameter actually depends on the values of other parameters. The off-trend local optimization is a measure of the interaction effects between the parameter concerned and the rest. Adjusting a parameter from its calibrated value moves it away

from the optimum response point as well as change the strength of the interaction effects between this parameter and the rest of the calibrated parameters. Thus for the situation where this local optimum phenomenon does not manifest, this likely implies that the interaction effect of a parameter with the other parameters is weak for the specific responses, as in Figure 12(c) and 12(d) compared to Figure 12 (a) and 12(b). While this seems to suggest a multi-factor scenario analysis is desirable for the most sensitive parameters, the computational burden of the SWAT model means this is prohibitive. We intend to pursue research that will tackle the issue of improving computational efficiency for these types of applications.

SCS Curve Number Scenarios

The SCS curve number CN2, governs the amount of surface runoff increase. Generally, as the value of CN2 increases, the main channel flow increases exponentially as observed in Figure 12(a). However, the baseline, with 0% adjustment of the calibrated SCS curve numbers, again shows the local optimum mentioned earlier.

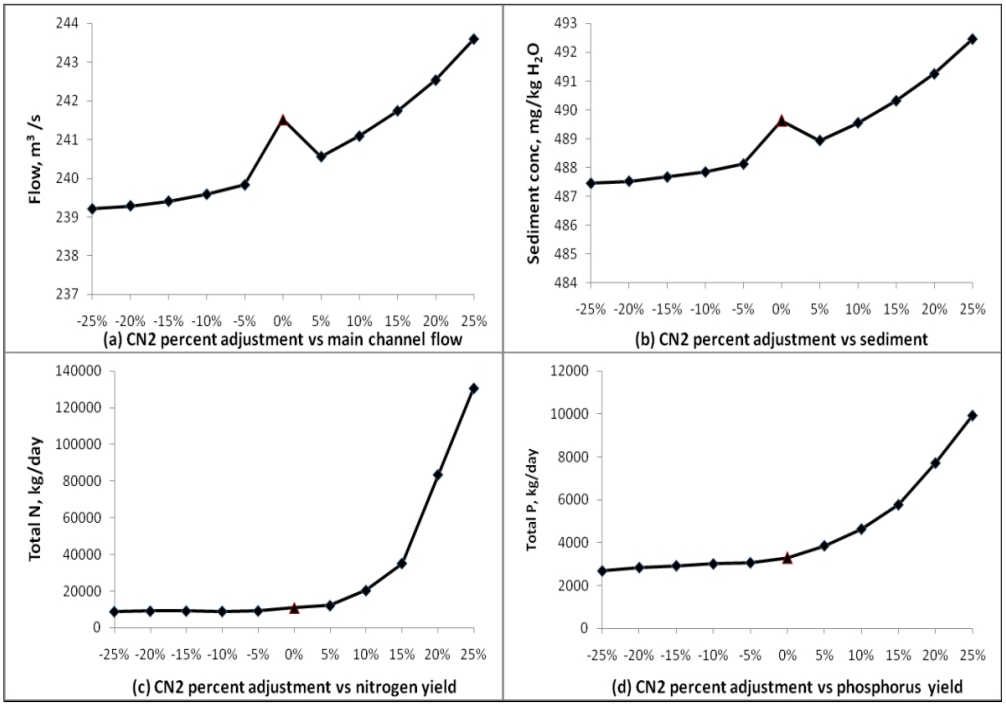

Figure 12: SCS curve number (surface runoff) scenarios

Figure 12(a) and 12(b) show a strong positive correlation between main channel flow and sediment concentration in the main channel. Figure 12(c) and 12(d) show that although higher values of CN2 significantly improve main channel flow, they would also more exponentially promote higher values of total nitrogen and total phosphorus. Typically the aim is to reduce soil erosion by surface runoff. In the case study, lowering total water delivered for irrigation would lower CN2 and thus also results in significant reductions in nutrients washed from the fields into the main channels.

USLE_C and USLE_P Scenarios

The minimum universal soil loss equation cover factor (USLE_C) and the USLE support practices factor (USLE_P) are parameters that typically affect soil erosion and hence sediment creation in the watershed. The conservation practices affecting USLE_C

and USLE_P are generally those controlling tillage operations. In the scenarios of USLE_C and USLE_P these two parameters (Figure 13) however did not have any effect on sediment concentrations in the main channel. This negligible effect to sediment concentration in main channel might be due to the weak effect of surface runoff in the watershed to significantly cause erosion.

Figure 13(a) and 13(c) show that except for the local optimum around the baseline calibrated values, reducing USLE_C linearly reduces total nutrients loadings in the main channel. Conservation practices that improve vegetative growth on the land would reduce the USLE_C by increasing nutrients uptake by plants more than nutrient deposition by foliage, thus perhaps directly reducing net nutrients yields.

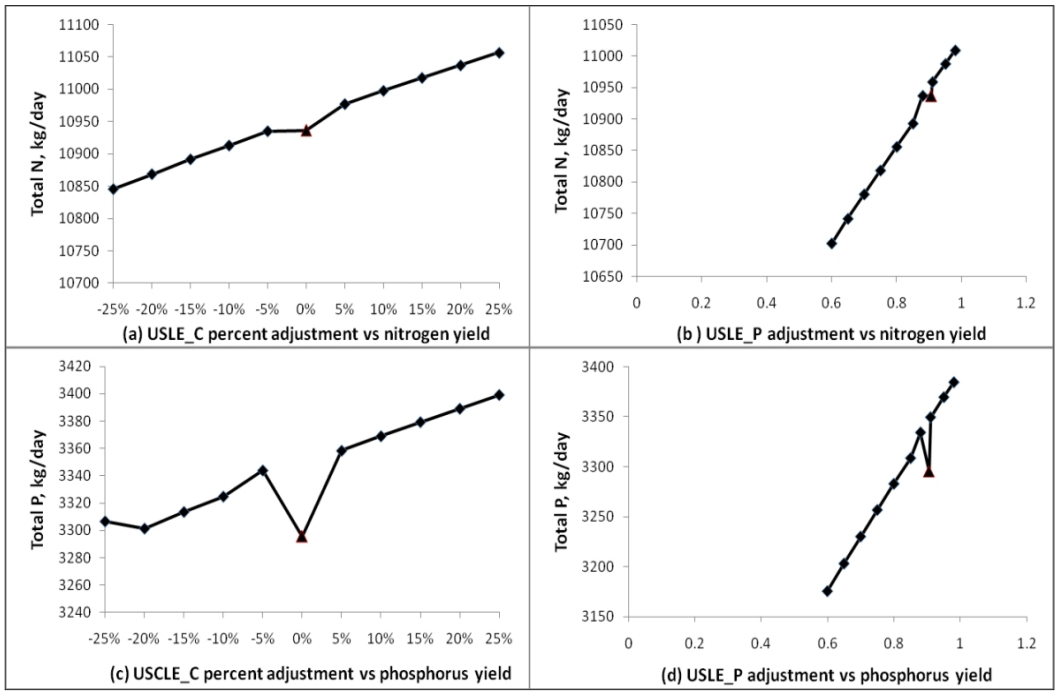

Figure 13: USLE_ C and USLE_P Vs nutrients scenarios

Figure 13(b) and 13(d) show another linear relationship between USLE_P and the total nitrogen and total phosphorus released to the main channel. Given the optimum

calibrated USLE_P value of .9065 there is sufficient ability for implementing conversation practices that desirably lower the USLE_P values such as strip-cropping, contour farming, grassed waterways and use of field borders. Although the graphs in Figure 13(b) and 13(d) show steep gradients, note that a 12% reduction in the USLE_P from .9065 to 0.80 would however only produce about 1% and 0.5% improvements in total nitrogen and total phosphorus yields respectively.

Fertilizer Application Scenarios

Figure 14 shows linear changes to total nitrogen and total phosphorus in the main channel with percent changes to amounts of fertilizer per application. Reducing nitrate fertilizer application reduces total nitrogen loadings more than reducing phosphate fertilizers would reduce total phosphorus loadings.

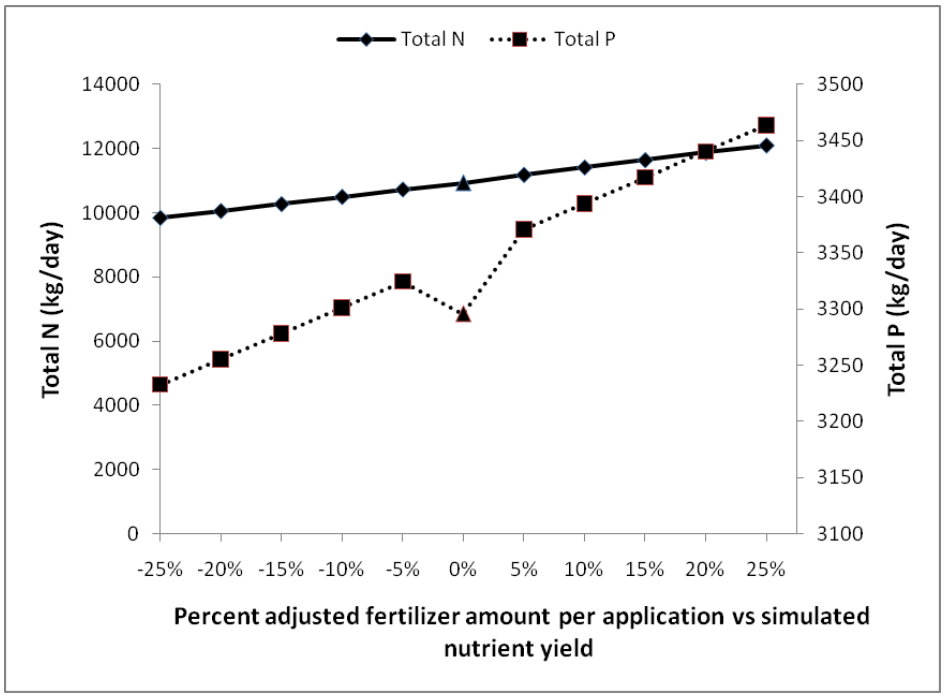

Figure 14: Fertilizer application Vs simulated nutrients yield scenarios

Using filter strips around crop fields as a conservation practice does not have any effect on the amounts of nutrients or sediment carried to the channels in the model.

Channel Scenarios

Conservation practices that affect channel conditions affect effective hydraulic conductivity (Ch_K2), Manning's value (Ch_N2), erodibility factor (Ch-Erod) and the cover factor (Ch_Cov). Among these channel scenarios, flow is affected by Ch_N2 and Ch_K2 only, as given in Figure 15(a).

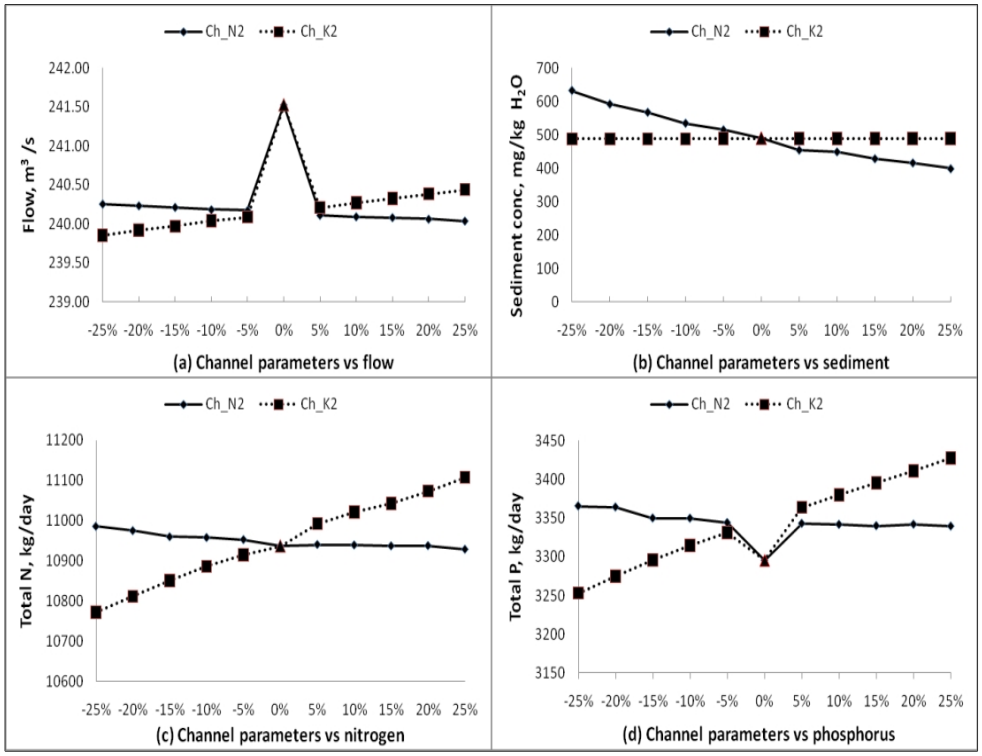

Figure 15: Percent adjusted channel parameters Vs flow, sediment and nutrients

From Figure 15(a), reducing the main channel hydraulic conductivity, Ch_K2, from a calibrated value of 45.10, results in lower flows. Reducing hydraulic conductivity of channel beds and banks reduces the water that leaves through seepage or enters the

stream from return flow. Hence lower discharges from lower conductivity imply that the main channel would experience a net gain of water by return flow from the shallow aquifers. Otherwise it would experience net loss of water (influent channel) by recharge to the shallow aquifer and the Ch_K2 graph would be reversed. If other local channels in the watershed are also net gaining channels, including irrigation canals, stabilizing the channels or making the beds and banks more impervious would potentially marginally worsen flow by reducing water gains through return flow from shallow aquifers. The Manning's channel roughness, Ch_N2, was calibrated at 0.052, which implies sparse vegetation along the banks of the main channel. Planting or encouraging more vegetation growth along the channel banks would marginally reduces main channel flow as shown in Figure 15(a). Using the earlier proposed explanation for the observed local optimality of the baseline (calibrated) values, a pronounced difference between neighborhood and calibrated responses means that both Ch_N2 and Ch_K2 are highly dependent on the values of other parameters for their effect on flow. A conflicting benefit of increasing Manning's values as shown in Figure 15(b) is that more vegetation reduces sediment concentrations in the streams. The slight increase in main channel sediment concentration with increasing effective hydraulic conductivity is so small (after the decimal point) it is not perceptible on the scale of the graph in Figure 15(b) and should be considered negligible.

In contrast, Figure 15(c) and 15(d) show a significant reduction of total nitrogen and total phosphorus yields with reductions in hydraulic conductivity. A speculative explanation of this significant reduction in nutrients loadings with reduction in conductivity is that shallow aquifer water, which might have higher concentrations of

nutrients, mixes less with channel water. Although we did not monitor the prediction of groundwater nutrients of the model, explanations for channel hydraulic conductivity as well as for irrigation scenarios suggest that the watershed aquifers in fact typically contain higher concentrations of leached nutrients than the streams.

As expected, higher Manning's roughness values, perhaps because of more aquatic vegetation and hence more nutrient uptake, reduce the amount of total nitrogen and phosphorus yields in the main channel. The local optimization of the calibrated baseline value of Figure 15(d) shows that both Ch_N2 and Ch_K2 are highly dependent on interactions with other parameters for predicting total phosphorus. Other values of Ch_N2 and Ch_K2 away from the calibrated value result in noticeable changes in the predictions of phosphorus loadings of the model.

The calibrated channel cover and erodibility factors of 1.000 and 0.9968 mean that the main channel has very little to no cover and is highly erodible respectively. Figure 16 shows a similar gradient linear trend of the effect to sediment concentration in main channel of adjusting the cover and erodibility factors. Values of Ch_Cov and Ch_Erod close to zero are for high vegetative cover and very low erodibility of the main channel respectively. From the baseline value in Figure 16, there is a lot of room for reasonable reductions in sediment concentrations and related sediment yields through reducing channel erodibility and improving vegetative cover.

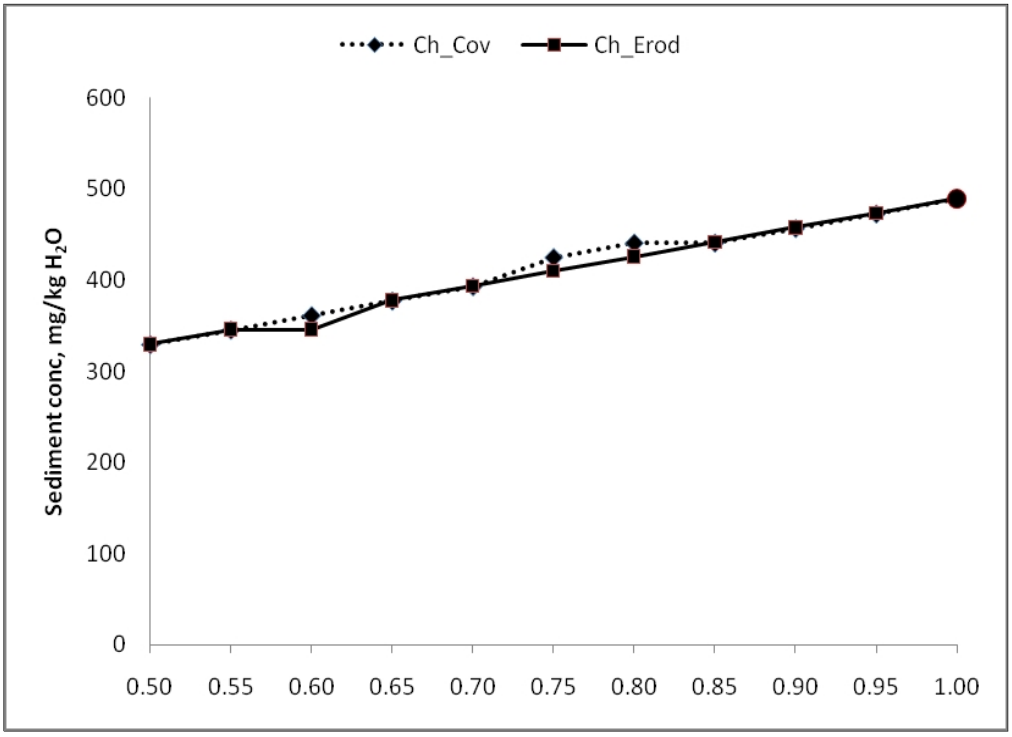

Figure 16: Average sediment concentration in main channel, Ch_Cov, Ch_Erod

There were no predicted effects of Ch_Cov and Ch_Erod on total nitrogen and total phosphorus yields in the main channel. This is interesting since if Ch_Cov represents channel plant cover then one would expect more plants to reduce the nutrients loads as suggested for higher Ch_N2. Perhaps while this might be so, organic nitrogen and phosphorus that result from decaying foliage balances out the plant nutrients uptake in the long run.

Summary and Conclusion

A new way for reducing bias in the selection of parameter values representing different conservation practice scenarios is to use a range of, instead of single, values for each conservation practice. Besides reducing selection bias, using a range of parameter values also shows the trends and sensitivities of parameters and hence the extent to which

related conservation practices might have to be implemented before desirable results can be expected. Using this range-based approach to scenario analysis the following points were deduced about the case study watershed to the impact of multiple proposed conservation practices.

To a limit, irrigation systems that deliver smaller amounts of water more frequently counter-intuitively results in lower discharges than less frequent application of larger amounts of water. This could be due to frequent water application promoting more evapotranspiration while less frequent but larger amounts of water delivery promote more return flows and also more sediment into the channels.

The baseline model shows that there has been a steady increase in sediment concentration in the Yellowstone River. The sediment concentration is more correlated to discharge than precipitation, and hence implies it is potentially generated by flow processes in the channel. Channel conservation practices, such as controlling channel erodibility, cover and Manning's roughness can positively reduce sediment creation. Practicing zero tillage reduces sediment concentrations but promotes the washing away of nitrogen and phosphorus into the channels. Also, conservation practices that reduce SCS curve numbers, CN2, would reduce sediment and by reducing surface runoff and hence discharges. While reducing CN2 from baseline values do not reduce nutrients yields into the channels, increasing CN2 would exponentially increase nutrients loadings.

The main source of total nitrogen and phosphorus is fertilizer application, which produces the highest concentrations in March, the beginning of the planting season. Improving biological mixing in the watershed would reduce nutrients concentrations in the river without increasing sediment concentrations. Reducing the channels' effective

hydraulic conductivity (say, by making channel beds impervious) linearly reduces nutrients loadings and discharges, but not sediment concentrations. Lower flows for more impervious beds and banks imply that watershed channels are net gainers of water by return flow from shallow aquifers. That is, more impervious surfaces prevent more return flow into channels and hence reduce discharge. It also implies that return flow typically has more nutrients than channel water, hence the significant rise in nutrients loadings with more return flow. Field filter strips and crop rotation did not have significant reductions to nutrients loadings in the channels. The limited effect of field strips again perhaps indicates that nutrients and sediment are not typically lost from the fields to the rivers through surface runoff.

Such a complete description of the effects of different scenarios can only be obtained by testing for more than one value of a parameter for each conservation practice. A drawback in using ranges of parameter values for each conservation practice being studied is that the number of scenario runs and the analysis required multiplies. The localized optimum observed in this study for several parameters suggests that considering parameter interactions by use of factorial designs could additionally enhance the approach. However, distributed models for large watersheds have heavy data requirements and typically contain many parameters worsening the computational burden of any sensitivity analysis, calibration and scenario analysis. Sensitivity analysis is critical in determining which parameters and inputs affect model output the most. Parameters and inputs that do not significantly affect model behavior are excluded from calibration and uncertainty analysis, with significant time gains. Confidence intervals from uncertainty analysis are useful in determining whether different model scenarios

being studied are significantly different or are otherwise indistinguishable from random error. Overall, the utility of these models is undeniable but model efficiency and usefulness would be enhanced by attempts to improve computational efficiency.

Contribution of Authors and Co-Authors

Manuscript in Chapter 3:
A Case Study Examining Factors Affecting the Performance of Response Surface Modeling During Bayesian Optimization and Uncertainty Analysis of Hydrologic Models

Author: Able Mashamba
Contributions: Overall concept development, experimental design, models development and assessment, results and manuscript write-up.

Co-author: Lucy Marshall
Contributions: Study supervision and oversight and provision of the Australian Water Balance case study model and data.

Manuscript Information

Able Mashamba and Lucy Marshall

Journal of Environmental Modeling and Software

Status of manuscript (check one):

___Prepared for submission to a peer-reviewed journal

__✓_Officially submitted to a peer-reviewed journal

___Accepted by a peer-reviewed journal

___Published in a peer-reviewed journal

Publisher: Elsevier Science

Date of submission: June 1, 2010.

CHAPTER 3

A CASE STUDY EXAMINING FACTORS AFFECTING THE PERFORMANCE OF RESPONSE SURFACE MODELING DURING BAYESIAN OPTIMIZATION AND UNCERTAINTY ANALYSIS OF HYDROLOGIC MODELS

Abstract: The use of response surface methodology (RSM) for approximating objective functions during the optimization of complex time expensive models has been well researched in environmental modeling and specifically hydrologic modeling (Ye, Li et al. 2000; Buhmann 2003; Regis and Shoemaker 2004; Regis and Shoemaker 2007; Bliznyuk, Ruppert et al. 2008). Generally, successful optimization algorithms based on RSM approximated surfaces are provided as stand-alone RSM-optimization implementations that are compared to existing alternatives. However, there has not been any specific effort to compare the underlying factors that affect the performance of different RSM based optimization and uncertainty assessment algorithms. A study of factors affecting the performance of optimization on RSM fitted surfaces would help analysts with applying effective RSM-based optimization approaches for computationally time-expensive model calibration and analysis in operational modeling and not only in research. We studied the effect on Bayesian optimization performance of (i) the approximation function used and the type of model or response function approximated, (ii) the number of points used in the RSM curve fitting, (iii) the size of the region from which fitting points are sampled and (iv) the size of normal error in the observed-data. We considered multiple typical hydrologic models, using synthetic and real data. The performance of an RSM fitting scenario was measured as the Euclidean distance between the optimum point obtained by Bayesian optimization on the RSM fitted surface and the

optimum point obtained by Bayesian optimization directly on the actual model. We maintain that a shorter distance implies better overall optimization performance than a longer distance. Bayesian Markov chain Monte Carlo (MCMC) optimization over a fitted response surface created by the spline based radial basis function (RBF) generally improved parameter estimates over using the computationally simpler second order multiple linear regression (MLR) except in cases where the simulated model had similarities in structure. Generally, increasing the number of points used in RSM fitting improves the accuracy of fit and hence parameter optimization. However, results suggest that beyond some value, more RSM points produce diminished improvements, or may even begin to negatively affect the performance of RSM approximation for some situations. Reducing the parameter sampling domains consistently improved the RSM fitting and Bayesian optimization. From the results, we infer that fitting fewer points on a smaller range gives better results than fitting many more points on a larger region. Finally, the size of the Bayesian residual error (which may be analogous to observation error or model structural error) significantly affects the performance of RSM methodologies.

Keywords: Response surface methodology, radial basis function, Bayesian optimization, uncertainty analysis, hydrologic modeling

Introduction

Computer modeling of hydrologic systems is becoming standard in hydrologic science (Sivakumar 2008). It is now feasible to model physically-based as well as spatially and temporally distributed hydrologic processes and interactions. Even so, such

complex models often need calibration, in addition to the estimation of predictive uncertainty desirable for many predictive models (Yang, Herath et al. 2000). Optimizing such models and quantifying their predictive uncertainty is still often time consuming. Practically, a large number of model runs is often needed for optimization and uncertainty assessment. These models however have sufficient complexity that make a large number of runs often time infeasible (Sahoo, Ray et al. 2006). As such, a number of approaches have been proposed that reduce the computational time needed to calibrate and estimate predictive uncertainty for these computationally time expensive models. Three prominent approaches (in hydrologic analysis in particular) for reducing computational time during model calibration and uncertainty analysis are the use of more efficient optimization algorithms (Vrugt, Gupta et al. 2003; Ter Braak 2006; Tolson and Shoemaker 2007), the use of surrogate models that approximate the responses of actual models (Regis and Shoemaker 2004; Regis and Shoemaker 2007; Bliznyuk, Ruppert et al. 2008) and multi-node parallel processing (Vrugt, Gupta et al. 2003; Muttil, Liong et al. 2007). Within hydrologic modeling, the development or optimization of fast global search algorithms used for calibration may be reaching diminishing returns (Andr´eassian, Lerat et al. 2007; Sivakumar 2008; Yang, Reichert et al. 2008). Parallel processing has the potential for significant computational speed gains in complex hydrologic modeling as processing networks become more available. An attractive way for reducing computation time on single or serial processing machines is by using fast surrogate models to approximate the responses of computationally time-expensive hydrologic models during calibration and uncertainty analysis. Response surface methodology (RSM) is a collection of statistical tools used for fitting an approximate

function to the responses of an actual model, for suitably selected values of the explanatory variables, with a view to using this fitted function in interpolating responses at some other input values of the explanatory variables.

In this paper, we investigate four factors that affect the performance of Bayesian calibration and uncertainty analysis when using simple RSM approximation for hydrologic models. We are interested in how the (i) approximation function and the type of model or response function approximated, (ii) the number of points used in the actual curve fitting, (iii) the size of the region from which fitted points are sampled, and (iv) the size of observed data error, affect the accuracy of Bayesian calibration and uncertainty assessment when using RSM models. The size of residual model error affected the performance of the Bayesian Markov chain Monte Carlo, MCMC, algorithm in finding theoretical optima, and therefore could also affect the performance of response surface fitting.

To our knowledge there has not been a review of these factors and their impact on hydrologic or time series model specification presented in one paper. We feel that before choosing specific algorithms, it may be informative for analysts to consider how the different RSM factors and levels selected may affect the performance of the resulting Bayesian optimization and uncertainty assessment. The RSM approximating functions reviewed are the second order least-squares multiple linear regression (MLR) and the spline based radial basis functions (RBF), which were selected for their ease of implementation and common use in RSM fitting (Buhmann 2003; Regis and Shoemaker 2004; Regis and Shoemaker 2007; Bliznyuk, Ruppert et al. 2008).One case study and three synthetic simulation models were used for this study. The case study model is a

simple water balance model that uses precipitation, evapotranspiration and discharge data for a catchment in Australia to simulate runoff. The first synthetic model is a multivariate quadratic function that has simple curvature. A cosine function was used to represent basic time-series output, and a synthetic water balance hydrologic model chosen for its comparatively many number of parameters that could provide a more complex response surface. Finally we implemented a well established conceptual rainfall runoff model with real data to investigate issues related to real model errors.

Although the functions assessed in this paper are not computationally expensive, results from this study will help inform future efforts in setting up effective RSM approaches for use with distributed hydrologic models. This is because recent efforts in global RSM fitting (Buhmann 2003; Regis and Shoemaker 2004; Regis and Shoemaker 2007; Bliznyuk, Ruppert et al. 2008) assume prior knowledge of some highest posterior density (HPD) region, which typically requires prior optimization. In Bayesian inference, the HPD region is equivalent to a statistical confidence interval or credible interval. However, for computationally expensive function evaluations, this prior optimization adds to the total computational burden. Rather than determining the HPD prior to fitting, using an extended sampling domain and local fitting is an alternative that Regis and Shoemaker (2004) use with evolutionary optimization. It would be interesting to learn how local and global RSM fitting compare, for use with Bayesian MCMC for cases where there is prior knowledge of the HPD parameter sampling region.

Methods for Model Optimization and Uncertainty Assessment in Hydrologic Modeling

Bayesian Inference and Markov Chain Monte Carlo

Bayesian inference is an approach to statistical inference in which all forms of uncertainty are expressed in terms of probability (Gelman, Carlin et al. 2004). A Bayesian approach to model parameter uncertainty assessment starts with the formulation of a model that best describes the system under study. A prior distribution over the unknown parameters of the model is formulated, which is meant to capture our beliefs about the system before seeing the system data. After observing some system data, we apply Bayes' Rule to obtain a posterior distribution for the unknown model parameters, which takes account of both the prior and the data. From this posterior distribution we can compute predictive distributions for future observations. Markov Chain Monte Carlo sampling is the main computational implementation of Bayesian inference (Gelman, Carlin et al. 2004). Classical statistical methods for predictive uncertainty analysis typically rely on large numbers of model runs to create acceptable confidence intervals on model variables and parameters under study. However, for complex hydrologic models, the large number of model runs would require significant run-times (Mugunthan and Shoemaker 2006). For such computationally time-expensive models, Bayesian uncertainty assessment methods have been found to converge to an optimum faster than classical statistics based methods, without necessarily being more effective (Hong, Strawderman et al. 2005). While Bayesian methods are arguably more inefficient than non-Bayesian evolutionary algorithms for global optimization in higher dimensional spaces such as found in hydrologic models, they are effective for uncertainty analysis (Hastings 1970; Bates and Campbell 2001; Huard and Mailhot 2006; Kavetski, Kuczera

et al. 2006; Marshall, Nott et al. 2007; Huard and Mailhot 2008). A number of Bayesian optimization and uncertainty assessment algorithms have been developed and successfully tested in hydrologic models (Kaheil, Gill et al. 2006; Kuczera, Kavetski et al. 2006; Yang, Reichert et al. 2007; Huard and Mailhot 2008). These algorithms, such as the Adaptive Metropolis, AM, (Haario, Saksman et al. 2001; Marshall, Nott et al. 2007; Smith and Marshall 2008) use MCMC sampling to implement Bayesian optimization and uncertainty assessment. MCMC sampling uses three main concepts; (i) Markov chains, (ii) Monte Carlo (pseudo-random) sampling and (iii) Bayes' Theorem (Renard, Garreta et al. 2006). Running a Bayesian Markov chain after its convergence is equivalent to sampling from an underlying but unknown joint posterior probability distribution of the model parameters. The histograms of sampled values after convergence are in fact approximations of the parameters' uncertainty distributions (Huard and Mailhot 2008). Thus Bayesian inference gives more uncertainty information than just the confidence intervals of classical statistical inference.

Say we cast a simulation model as Equation 3.1;

$$y_{obs,t} = f_t(X, \theta) + \varepsilon_t \tag{3.1}$$

where $y_{obs,t}$ is the the observed or system data at time step *t*, f_t is the simulated model, θ is the unknown model parameters, X is the model input (forcing) data, and ε is an error term representing the model residual error (Bates and Campbell 2001; Kuczera, Kavetski et al. 2006).

Equation 3.2 shows the MCMC setup for implementing Bayes' rule in sampling the limiting posterior distribution in a Metropolis Hastings (Metropolis, Rosenbluth et al.

1953; Hastings 1970) Markov chain at the i^{th} iteration (Gaganis and Smith 2001; Gelman, Carlin et al. 2004);

$$\frac{P\left(\theta^{i} \mid y\right)}{P\left(\theta^{i-1} \mid y\right)} = \frac{P\left(\theta^{i-1} \mid \theta^{i}\right)P\left(y \mid \theta^{i}\right)P\left(\theta^{i}\right)}{P\left(\theta^{i} \mid \theta^{i-1}\right)P\left(y \mid \theta^{i-1}\right)P\left(\theta^{i-1}\right)} \tag{3.2}$$

where θ^{i} is the current parameter value and θ^{i-1} is the proposed parameter value. The left hand side of Equation 3.2 is the ratio of the joint posteriors at states *i* (numerator) and $i-1$ (denominator) and is the probability of accepting the proposed parameter state. $P(\theta^{i} \mid \theta^{i-1})$ denotes the joint proposal probability densities for transitioning from state $i-1$ to i. For a symmetric random walk MCMC chain, $P(\theta^{i} \mid \theta^{i-1}) = P(\theta^{i-1} \mid \theta^{i})$, a Gaussian multivariate distribution is typically used as the proposal distribution (Gelman, Carlin et al. 2004). The $P(y|.)'s$ are the joint likelihoods at the two states, i and $i-1$ and the P(.) are the joint priors of parameters θ at the two iteration steps i and $i-1$.

The five main decisions in implementing Bayesian MCMC uncertainty analysis scheme are; (i) choosing a prior distribution, (ii) deciding on the proposal scheme, (iii) choosing an appropriate likelihood function, (iv) determining the warm up period and (v) finding a high posterior density (HPD) region from which to sample values (Jing Yang; Gaganis and Smith 2001; Balakrishnan, Roy et al. 2003; Gelman, Carlin et al. 2004; Kavetski, Kuczera et al. 2006).

A good choice of the prior distribution can help quicken chain convergence (Gelman, Carlin et al. 2004). However, when a model-specific prior is not known, hydrologic modelers often use a default non-informative uniform continuous distribution for a prior (Smith and Marshall 2008). This works since large datasets often used in

hydrologic modeling quickly correct the effects of poor priors (Gelman, Carlin et al. 2004).

The proposal distribution is a jump or perturbation distribution used for sampling trial parameter values or input data under study (Gelman, Carlin et al. 2004). It provides the transition mechanism of the Markov chain. A proposal distribution that is closer to the often unknown limiting posterior distribution converges faster than a dissimilar proposal (Gelman, Carlin et al. 2004). A popular Bayesian MCMC algorithm, the Adaptive Metropolis, AM (Haario, Saksman et al. 2001), starts with a multivariate random normal proposal distribution whose initial covariance matrix is defined by the user based on prior model knowledge. As the MCMC chain progresses, the AM updates the covariance matrix of the proposal distribution based on the accepted historical values of the MCMC chain.

The likelihood function, P(Y), considered in the studies in this paper, Equation 3.3, is based on assuming normal error and is widely used for characterizing parameter or input error in hydrologic modeling (Bates and Campbell 2001; Gaganis and Smith 2001; Hong, Strawderman et al. 2005; Huard and Mailhot 2006; Marshall, Nott et al. 2007). This is because Gaussian multivariate functions are readily integrable for any number of dimensions and are generally considered a valid depiction of a random error probability from an inferential point of view (Jaynes and Bretthorst 2003).

$$P(Y) = P(y \mid \Theta) = (2\pi\sigma^2)^{-n/2} \cdot \prod_{t=1}^{n} \exp\left\{ -\frac{(y_{obs,t} - f_t)^2}{2\sigma^2} \right\} \tag{3.3}$$

Where y, $y_{obs,t}$ and f_t are the observed data and the time-series values of system observed data and simulated model output at time step t respectively. For example, these could be discharge, or concentrations of sediment, nutrient, pesticide, bacteria or metals

or any other responses under study. The variance, σ^2, is an estimate of the variance of the model error and n is the number of the observed time-series data. In practice, the error variance is not known before-hand and has to be sampled for and optimized together with the other unknown model parameters (Gelman, Carlin et al. 2004). Useful variations of the Gaussian likelihood function have also been successfully used in hydrology. A notable example is the auto-correlated Gaussian error likelihood function (Romanowicz, Beven et al. 1994; Romanowicz, Beven et al. 1996). The auto-correlated Gaussian error likelihood is used where output data are not independent. Experts differ on the improvements brought about by modifying the Gaussian likelihood to account for autocorrelation (Gupta, Thiemann et al. 2003).

Proper likelihood functions are supported by statistical assumptions on the nature of observed output data and error. Using simplified or generalized 'likelihood functions' compromises the statistical soundness of Bayesian analysis (Mantovan and Todini 2006; Mantovan, Todini et al. 2007). For instance, objective functions have been extensively used to replace the proper conditional probabilistic likelihoods in hydrology (Aronica, Bates et al. 2002; Beven, Smith et al. 2006; Andr´eassian, Lerat et al. 2007; Beven, Smith et al. 2007; Mantovan, Todini et al. 2007; Montanari 2007; Beven, Smith et al. 2008). Unfortunately, the simplified or generalized objective functions are not based on any justifiable statistical assumptions about the nature of error. Hence Bayesian data analysis on MCMC sampling schemes based on objective function or other informal likelihoods have been as asserted as being statistically incoherent (Beven, Smith et al. 2006; Hall, O'Connell et al. 2006; Mantovan, Todini et al. 2007).

The Markov chain warm-up or burn-in period is the number of states the posterior Markov chain transitions through before the influence of the initial values of the parameters or inputs used becomes negligible (Gelman, Carlin et al. 2004). Identifying a burn-in period helps in evaluating the performance of the MCMC in chain mixing and convergence (Gelman, Carlin et al. 2004). During the burn-in period, posterior densities are not updated (Thiemann, Trosset et al. 2001). Gelman et al (2004) propose a general rule of discarding half of the MCMC chain as burn-in period. Finally, an MCMC may sample from the entire parameter space or only from the response surface as mapped by the 'good' enough parameters from a prior calibration process (Balakrishnan, Roy et al. 2003). For some complex hydrologic models, MCMC sampling from the entire parameter space may be time-prohibitive, hence the need to sample from the HPD region that tightly bounds the global optimum (Balakrishnan, Roy et al. 2003).

Response Surface Methodology Approximation

Response surface methodology (RSM) is a collection of statistical and mathematical techniques useful for developing, improving, and optimizing processes through the exploration of relationships between several explanatory variables and one or more response variables (Myers and Montgomery 1995).

Consider the model defined in Equation 3.1, and suppose it describes a hydrologic system, such as channel flow. The term y is the response of the system, which in this case would be observed channel discharge. The function or model f represents the average system relationships between inputs and outputs, which may be poorly known or difficult to evaluate under all conditions given the complexity of f. X is a vector of known and controllable (input) variables, such as precipitation and initial soil moisture

content. θ is a vector of actual system parameters or constants of the poorly known function or model and ε represents the natural variability or error of the system. The long term distribution of ε is often assumed to be a normal with a mean of zero and variance, σ^2 (i.e. Equation 3.1), such that the expectation of the hydrologic system $E(y) = \hat{y} = E[f(X,\theta)] + E(\varepsilon)$, is just, $\hat{y} = f(X,\hat{\theta})$ where $\hat{\theta}$ is the unbiased estimator of the model or function *f* parameters. However, since f is poorly known or difficult to evaluate, we often have to approximate for f in order to feasibly characterize the hydrologic system, for a given set of inputs (Myers and Montgomery 1995). With known observed discharge y and input X, estimating *f* is basically estimating for the $\hat{\theta}$ parameters. RSM provides a sequential approach to the approximation of the response, $\hat{y} = f(X,\hat{\theta})$, of a model or function with some surrogate or fitted function, $\hat{y}' = f' = f(X,\hat{\theta}')$.

The first step towards successful RSM approximation involves screening off all explanatory variables that do not cause a statistically significant change to the response variable(s) of the study. In modeling, finding the important explanatory variables is known as sensitivity analysis while in other studies this involves designing and conducting factorial experiments. The second step involves establishing the domains or ranges of values that must be taken by the sensitive parameters vector, θ, to give near-optimum response. This near optimum region is much smaller than the operability or feasible region of the parameters. In Bayesian sampling schemes, this near optimum region is called the high posterior density region. Prior optimization routines and knowledge may be used to establish these narrow domains of near optimality. The third step is to approximate for the true response in this narrow near-optimum region either using regression, interpolation or emulation functions.

For model optimization and uncertainty analysis, the response variable of interest is generally the likelihood $P(Y)$ as given in Equation 3.3, not channel discharge y. Estimating likelihood functions helps with conditioning or calibrating the parameters of a model based on historical observed data so as improve the capability of the model to predict some immediate future response. With the increasing use of complex hydrological models, such as distributed models, computing an objective function such as a likelihood, $P(Y)$, can be particularly time-consuming especially if the models are being conditioned to long time-series y data.

Popular RSM approximation methods include the least squares second order multiple linear regression, the radial basis interpolation functions (Myers and Montgomery 1995; Regis and Shoemaker 2004) and the use of neural networks (Jin, Olhofer et al. 2000). The second order approximating function $P(Y')$ is developed by finding the approximate $\hat{\beta}$ coefficients of Equation 3.4 so that the sum of squared residuals between $P(Y)$ and $P(Y')$ is minimized;

$$P(Y') = \hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i x_i + \sum_{i=1}^{k} \hat{\beta}_{ii} x_{ii}^2 + \sum_{i<j} \sum_{j} \hat{\beta}_{ij} x_{ij}^2 \quad (3.4)$$

where the x's are the elements of the parameter vector θ. The attraction of the second order regression model is that it is one of the simplest functions that can represent curvature (Myers and Montgomery 1995; Regis and Shoemaker 2004). Since most optimum points are turning points, with some curvature, fitting a second order polynomial to the HPD region is often a good and reasonable approximation (Myers and Montgomery 1995; Regis and Shoemaker 2004). Standard RSM literature provides readily accessible further detail on the least squares method of locally fitting a second order regression model (Myers and Montgomery 1995).

A radial basis approximation function, Equation 3.4, is a generalized kriging (interpolation) real-valued function of the form based only on the norm of **k** points with known responses from a given point **x**, whose response needs to be approximated (Buhmann 2003; Regis and Shoemaker 2004).

$$P(Y') = \sum_{i=1}^{k} \omega_i \, \phi(\|x - x_i\|) + \; p(x) \tag{3.5}$$

Typically, the norm $\|\cdot\|$ is some function of the Euclidean distance such that $\phi(r) = r^2 \log r \, , r > 0$ and $\phi(0) = 0$ (thin plate spline), $\phi(r) = r^3$ (cubic), $\phi(r) = \sqrt{r^2 + \gamma^2}$ (multi-quadratic), or $\phi(r) = e^{-\gamma r^2}, \; r \geq 0$ and γ is a positive constant (Gaussian) (Regis and Shoemaker 2004). The ω_i is a weight showing how the Euclidean distance for point x_i from a sampled point x contribute to the approximation of response at point x. The tail of the radial basis function, $p(x)$, is often linear with respect to the explanatory variables, but higher order polynomials could also be used. The coefficients ω_i's and p_j's in the radial basis approximation function $\hat{y}'$ are found by solving for a system of linear algebra using previously evaluated responses for the x_i points. Note the change in notation for the RHS of the second order approximating function, for which x_i represents a specific value of the i^{th} parameter while the RHS of the radial basis uses x_i to represent a point on the multidimensional Cartesian coordinate system of parameters that give an evaluated likelihood. The mathematical setup of the system of linear algebra equations to obtain the coefficients is readily available from current papers in the area (Buhmann 2003; Regis and Shoemaker 2004; Regis and Shoemaker 2007; Bliznyuk, Ruppert et al. 2008).

Current research that uses RSM methods within hydrologic modeling has been in model calibration and, more usefully, uncertainty analysis of computationally time expensive models (Regis and Shoemaker 2004; Regis and Shoemaker 2007; Bliznyuk, Ruppert et al. 2008). The RSM method as applied to Bayesian optimization and uncertainty analysis within modeling can be summarized in the following steps (Ye, Li et al. 2000; Buhmann 2003; Regis and Shoemaker 2004; Regis and Shoemaker 2007; Bliznyuk, Ruppert et al. 2008);

1. Calibrate model using a global optimization routine (Evolutionary Algorithm, Dynamically Dimensioned Search, Greedy search, etc) and observed data, y, so as to approximate the HPD region, the near optimum region of the parameters of the model (Bliznyuk, Ruppert et al. 2008).
2. Sample some RSM design points, x, at which the response or objective function of the expensive model is evaluated often using symmetric Latin hypercube sampling design, SLHD, from the HPD region for unbiased sampling (Ye, Li et al. 2000; Regis and Shoemaker 2004).
3. Evaluate the objective function on the selected RSM design points.
4. Fit a second order, or nth order, least squares multiple linear regression function or a radial basis function to the best responses of a subset of the RSM design points evaluated in 3. As much as all, or just the minimum nearest neighbors may be used for fitting.
5. Run a Bayesian MCMC optimization and uncertainty (or any other optimization) routine, evaluating on the fitted function for N number of cycles.

6. Evaluate the actual response for the final Bayesian MCMC parameter set after the N cycles in 5 and append to the RSM design points.
7. Repeat 4 to 6 until there are no further systematic improvements to the responses being evaluated. Bayesian MCMC convergence to the optimum.
8. Use the parameter values drawn after Bayesian convergence to characterize the uncertainty and optimality of the model.

In design points sampling, step 2, Regis and Shoemaker (2004, 2007), Gutmann (2001) and Bliznyuk et al (2008) describe RSM approximation over a narrow parameter sampling domain containing the global optimum. Regis and Shoemaker (2004) initially sample design points using Latin hypercube sampling from the feasible domain and then later use adaptive sampling from the HPD region as their local fitting scheme progresses. On the other hand, Bliznyuk, Ruppert et al (2008) use an involved approach for design points sampling. Some of their design points come from points evaluated during the global optimization phase falling within some determined critical region. Other design points come from points evaluated during the development of the critical region (typically some 95% confidence interval) around the optimum set of parameter points. Finally, a *maximin* design is used to further provide additional evaluated points that are sufficiently far apart (Euclidian distance) from those so far selected and from each other, within the critical region.

As yet, there has not been any approach presented that assumes ignorance of the HPD region during RSM fitting for optimization of time – expensive models. The step 1 prior optimization may in fact involve a significant number of actual expensive model runs, which may also need to be avoided, if possible. Also, there has not been any

detailed review of potential factors that affect the performance of RSM fitting during optimization and uncertainty assessment.

Factors that have been indirectly considered, in different research - for choosing an RSM procedure, include the average performance of a fitting function when applied to different approximated models and functions, the locality and number of the points used in fitting, and the RSM points sampling routine used (Regis and Shoemaker 2004), but not the size of model error in the data. One relevant study, by Regis and Shoemaker (2004), studies the second order and spline radial basis functions with the objective of showing that fitting $(d+1)(d+2)/2$ nearest neighboring points at a time, from RSM design points sampled using the symmetric Latin hypercube sampling, could effectively be used for optimization and uncertainty analysis of time-expensive functions.

Methods

We adopted the general methodology of Bliznyuk et al (2008), following the findings of Regis and Shoemaker (2004) on the usefulness of RSM methodology for hydrologic models. The four main standard steps used in the RSM likelihood fitting (Equation 3.2) experiments were; (i) Sample a desired number of RSM design points within a predetermined parameter space. We preferred using symmetric Latin hypercube sampling of the design points for this report as used for initial design points sampling in similar research (Ye, Li et al. 2000; Regis and Shoemaker 2004), leaving future work to address in more depth adaptive sampling of design points for these types of models. (ii) Evaluate the likelihoods (responses) of the actual model or function, for each of the sampled RSM parameter points. (iii) Compute a surrogate or approximate function that maps the sampled RSM points to the evaluated likelihoods (responses) as compared to

the actual function and (iv) Use an MCMC algorithm to optimize for the approximate likelihood (response) using the fitted function alone until convergence.

Setting up the Experimental Factors

To identify factors that affect the performance of RSM approximation, computer experiments were set-up to examine the impact of the following four factors; (i) the approximation function used and the type of model or response function approximated, (ii) the number of points used in the actual curve fitting, (iii) the size of the region from which fitting points are sampled and (iv) the size of normal observed data error. Details of these experimental settings follow below.

For the approximating functions, we tested the performance of the second order least squares multiple linear regression (MLR) and the spline radial basis (RBF) functions. The MLR is ideal as one of the simplest non-linear function that can represent curvature while radial basis functions are popular general interpolation methods (Buhmann 2003; Regis and Shoemaker 2004; Bliznyuk, Ruppert et al. 2008). A simple linear p(x) tail, Equation 3.4, was used for the spline radial basis function.

For experimental fitting, we used three hypothetical models having three different properties and one case study water balance model. The first synthetic function has multi-variable quadratic curvature to represent simple non-linearity. The second and more complex model is a cosine time series function; to represent non-linear cyclic output typical of time series model output. The third is a synthetic lumped hydrologic model chosen for its relatively many dependent parameters, which is common with hydrologic models. The case study is a water balance model using data from an Australian sub-basin. For ease of reference, the quadratic, cosine and lumped flow models is referred to FX2,

FCOS and FLO models respectively. The case study model is referred as the AWBM, for the Australian water balance model.

The first function to be experimented on, using RSM approximation scenarios, was a quadratic function (FX2) with two input variables, x_1 and x_2 and two parameters or coefficients to be optimized for, as shown in Equation 3.6. A second order least squares multiple linear regression fit is expected to perform well on such a quadratic function. To test for the common effect of variable dependency found in most hydrologic models, five years of maximum and minimum daily temperatures for a case study watershed were used as the explanatory variables x_1 and x_2.

$$f(x_1, x_2) = 2x_1^2 - 0.5x_2^2 + Ax_1x_2 - Bx_1 - 20 \tag{3.6}$$

The objective was to investigate how well the different RSM approximation scenarios, coupled to MCMC optimization and uncertainty analysis, rediscover the coefficients A and B in the function Equation 3.6, as well as the normal error variances, $N(0, \sigma^2)$, introduced to each set of hypothetical observed data used in the experiments to represent potential data error or residual model error. Theoretical optimum parameter values of A and B, parameter sampling ranges and MCMC starting values for two classes of sampling ranges (extended and high posterior density) and three sizes of observed data error (Level 1, Level 2 and Level 3) are shown in Table 7. Theoretical optimum values (aka true values) are the parameter settings that used in creating the hypothetical observed data used in experiments for the FX2, FCOS and FLO models.

To test for the performance of RSM fitting to time series models, a cosine function (FCOS), Equation 3.7, was used.

$$f(t) = U_1 \cos\left(\frac{U_2 \pi t}{180}\right), \quad 0 \leq t \leq 12 \tag{3.7}$$

This time series function, f(t), used values of t from 0 to 12 at 0.025 intervals, while the theoretical optimum values of U_1 and U_2 values to be rediscovered were 180 and 30 respectively. Similar to the quadratic function scenarios already described, 3 sizes of normal error (Levels 1, 2 and 3) were introduced to approximated function responses making three hypothetical observed datasets, which were used in optimization and uncertainty analysis. Selecting the default (Level 2) or initial normal error variance for the three approximated models was somewhat arbitrary, with the only requirement that the error should not dominate the output. Since the effect of observed data error was considered for testing midway into experimentation, the other two sizes were selected by subtracting or adding half of the default error to make Level 1 and Level 3 error sizes respectively.

Table 8 shows the true and starting values of U_1, U_2 and of observed data errors used and their respective sampling limits for both the EXT and HPD categories. A slightly more complex response surface with more dimensions was obtained from a synthetic lumped hydrologic model (FLO) shown in Figure 17 that tracks daily precipitation, evapotranspiration, surface runoff, percolation, revaporation, lateral flow and return flow to predict channel flow. The α, β, δ, φ, κ, λ, μ are the proportion factors. Processes evt, srf, ltf, rvp, per, rcg, rtf and dsc are evapotranspiration, surface runoff, lateral flow, revaporation, percolation, recharge, return flow and discharge respectively. The min_process represent minimum amount of water required before process can occur while rtf_min is minimum base flow. FLO was designed to represent more complex

system behavior such as thresholds, and potential parameter interactions or insensitivities. The five parameters α, β, δ, λ and min_srf were calibrated for while the rest were fixed.

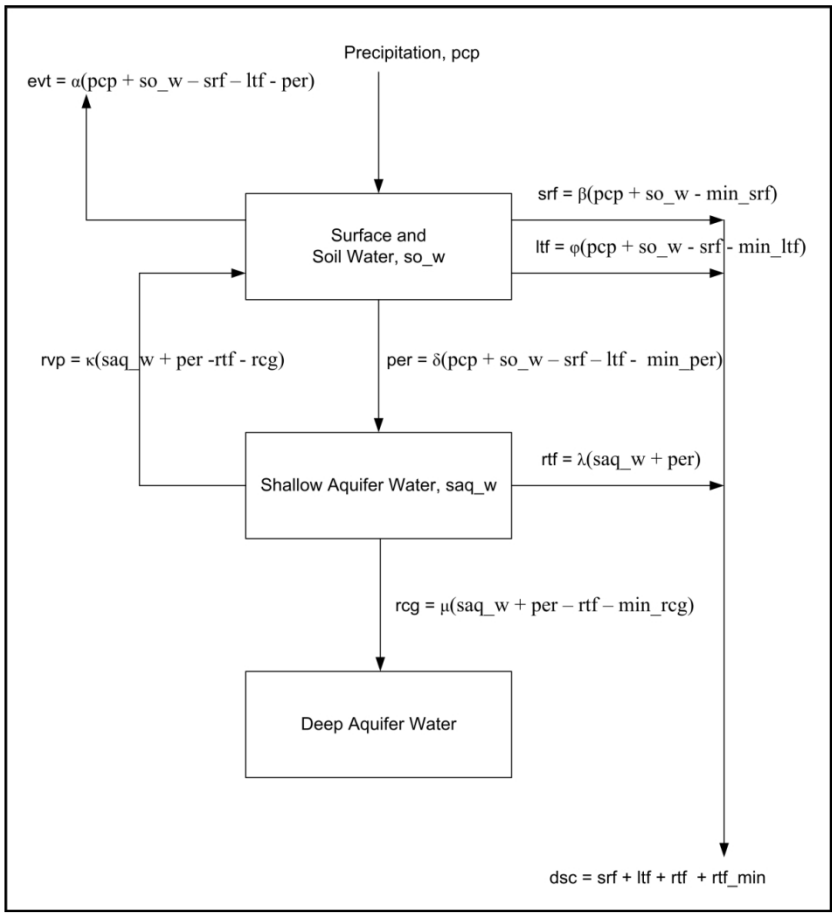

Figure 17: Schematic view of the FLO bucket model.

Based on previous uncertainty analysis, the four most sensitive parameters (α, β, δ, λ) and one less sensitive one (min_srf) were selected for Bayesian calibration and uncertainty experimentation while the other 10 were held constant. Five years of daily precipitation data for the case study watershed, the Lower Yellowstone River basin, was used as input to the hypothetical model, with all its 15 parameters set at specific values. For hypothetical observed data, the simulated flow (in mm) is adjusted by a random

normally distributed error $N(0, \sigma^2)$, as shown in Table 9. Table 9 shows the experimental settings that were used for the FLO model. Finally, the case study hydrologic model for this study was the Australian water balance model (AWBM), Figure 18 developed in the early 1990s (Boughton 1993; Boughton and Carroll 1993; Boughton 2004) considered to be one of the most widely used models in Australia with research by a number of researchers (Sharifi and Boyd 1994; Cakers and Yu 1998; Cheung and Yu 1999; Bates and Campbell 2001; Marshall, Nott et al. 2004). Observed discharge for the main channel of the watershed was used for assessing the performance of the experimental arrangements under study. Slightly more than ten years of daily observed data were used, from the period towards the end of 1988 to end of 1998.

Figure 18 shows a schematic of the AWBM case study model and the parameters that were used in experimentation. The parameters S, BS, BFI and K_S are surface water storage capacity, current baseflow storage, baseflow infiltration index and baseflow recession index respectively. The parameters S, BFI and K_S are considered variable for experimentation. The processes P, E, QE, QS, QR, QB and Q are precipitation, evapotranspiration, excess water available for surface flow, surface flow, recharge, baseflow and channel discharge respectively. Direction of water flow is shown by arrows and baseflow storage capacity is unlimited. The parameters S, BFI and K_S are considered variable for experimentation.

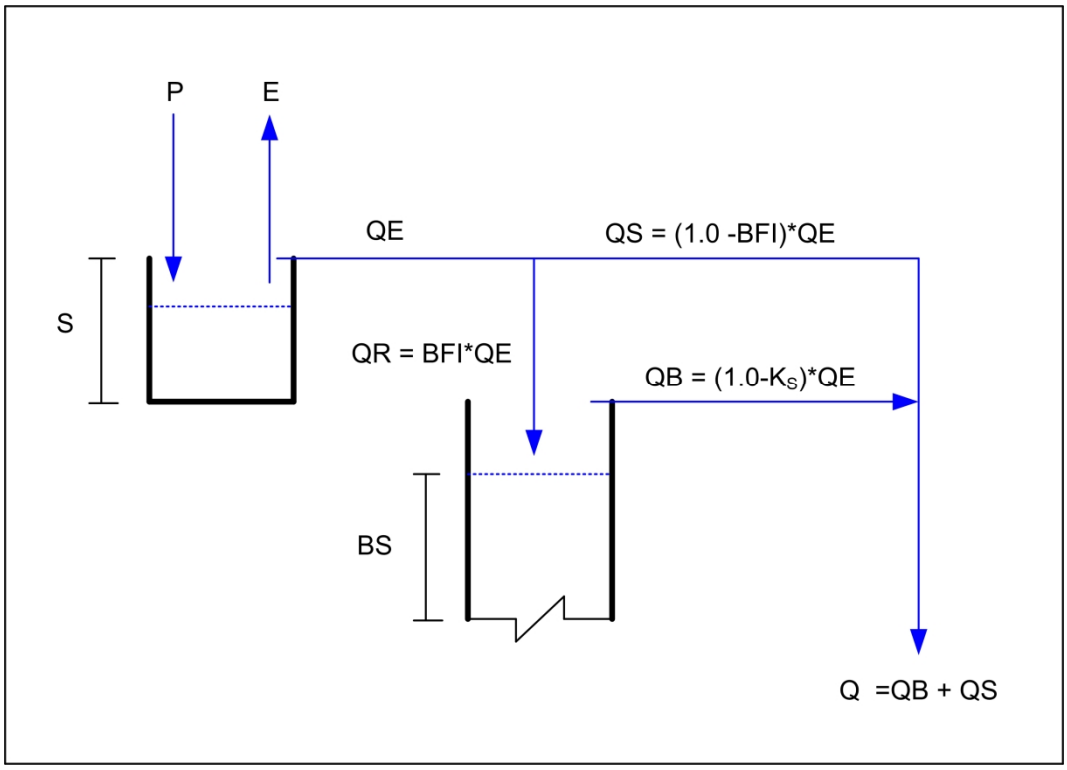

Figure 18: Schematic view of the AWBM case study bucket model.

Tables 7 to 10 show the experimental settings that were used in the case study models FX2, FCOS, FLO and AWBM. In Table 10, there are no different levels of observed data error to be tested for, as in other tables for synthetic model settings. Also, instead of using true values for performance comparisons, the case study scenarios use the parameter values obtained by direct MCMC optimization.

Table 7: Experimental settings for the quadratic function (FX2)

Property	Parameter					Type of sampling range
	A	*B*	*Error variance level*			
Size of error			1	2	3	
True value	6.00	-80.00	2.50	5.00	7.50	
Starting value	5.50	-78.00				
Lower bound	5.50	-85.00	1.00	3.00	3.00	Extended region, EXT
Upper bound	7.00	-74.00	8.00	10.00	10.00	
Lower bound	5.80	-81.00	1.00	4.00	5.50	Reduced region, HPD
Upper bound	6.40	-78.00	4.00	7.00	8.50	

Table 8: Experimental settings for the cosine function (FCOS)

Property	Parameter					Type of sampling range
	U1	*U2*	*Error variance level*			
Size of error			1	2	3	
True value	180.00	30.00	5.00	10.00	15.00	
Starting value	178.00	32.00				
Lower bound	170.00	25.00	3.00	8.00	13.00	Extended region, EXT
Upper bound	190.00	35.00	7.00	12.00	17.00	
Lower bound	178.00	28.00	4.00	9.00	14.00	Reduced region, HPD
Upper bound	183.00	32.00	6.00	11.00	16.00	

Table 9: Experimental settings for the synthetic lumped hydrologic model (FLO)

Property	Parameter								Type of sampling range
	alpha	*beta*	*delta*	*lambda*	*min_srf*	*Error variance level*			
Size of error						1	2	3	
True value	0.80	0.30	0.50	0.60	1.50	0.25	0.50	0.75	
Starting value	0.75	0.40	0.50	0.65	1.45				
Lower bound	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	Extended region, EXT
Upper bound	1.00	1.00	1.00	1.00	5.00	0.80	0.80	0.80	
Lower bound	0.60	0.20	0.40	0.50	1.00	0.00	0.40	0.40	Reduced region, HPD
Upper bound	0.90	0.50	0.70	0.80	2.50	0.40	0.80	0.80	

Table 10: Experimental settings for the case study lumped hydrologic model (AWBM)

Property	Parameter				Type of sampling Range
	S	*BFI*	K_S	*Error variance*	
Direct MCMC optimization, Mean value	210.8821	0.4882	0.9307	35.0022	
Starting value	209.00	0.4907	0.9005	36.5587	
Lower bound	50.00	0.00	0.00	10.00	Extended region, EXT
Upper bound	400.00	1.00	1.00	100.00	
Lower bound	150.00	0.30	0.50	30.00	Reduced region, HPD
Upper bound	250.00	0.60	1.00	40.00	

The second order MLR, requires a minimum of $(d+1)(d+2)/2$ RSM points for full characterization while RBF functions with a linear tail require only $d+1$ points (Regis and Shoemaker 2004), where d is the number of dimensions or actual model parameters under study. Thus the minimum RSM points required during MLR

approximation of the FLO model with 6 dimensions is 24. For this research, the numbers of RSM design points used in fitting the approximate functions were 50, 200, 600 and 900. These points cover for the minimum RSM points required as well as allow for the exploration of using excess points. Other arbitrarily chosen values could have been used as long as they satisfied the minimum number of RSM points required.

The size of the region from which the RSM points are sampled determines the size of surface that must be fitted at a time, for global fitting. Larger surfaces can be more difficult to approximate and hence result in poorer fits than smaller ones. This is because it is more difficult to find approximate functions that can account for the mixed curvature likely to be included in a large surface. In RSM approximation literature, the region from which fitted RSM points are sampled is assumed to be in the HPD region (Regis and Shoemaker 2004; Bliznyuk, Ruppert et al. 2008). Regis and Shoemaker (2004) and Bliznyuk et al (2008) fit RSM functions to 'nearest neighbors' to implicitly avoid using points that are spread out over a large region, which would result in a poor fit. To test how the locality of RSM points used for fitting affects fitting accuracy, RSM points are sampled from two sizes of regions, the reduced (HPD) and extended (EXT) regions, respectively. For our study the reduced (HPD) region is bounded by the parameter ranges found from prior converged direct MCMC optimization. Finding the HPD region using MCMC optimization would not be possible for expensive models. The extended region, which contains the HPD sampling domain, was based on our prior knowledge of the physical values the parameters can take.

Complete experimentation would include tests for 2 approximating functions, 4 approximated models, 2 sampling spaces, 4 quantities of RSM fitted points, 3 levels of

observed data error and 2 sampling strategies, resulting in 384 experiments. Due to time constraints there was need to significantly reduce the total number of experiments. However the different numbers of factor levels being investigated made factorial design unwieldy. Experiments were thus conducted based on the Balanced Incomplete Block Design, BIBD (Montgomery 2004), framework which gave a total of 150 experiments.

Adaptive Metropolis (MCMC) Uncertainty Analysis Settings

In applying Bayesian inference to the models and the fitted response surfaces, we used the adaptive Metropolis algorithm adapted from Haario et al (2001). The starting parameters, error, and variance-covariance matrix were manually set. Echoing previous studies, the rate of convergence of the algorithm was found to be sensitive to the starting values (Smith and Marshall 2008). Starting values closer to global optima result in good chain mixing and faster convergence than those further away.

A critical property of the AM algorithm is how the variance-covariance matrix of the transition distribution, used for multivariate random normal sampling of the next values in the MCMC chain, improves with increasing chain length. The starting variance-covariance matrix was defined as a diagonal matrix of proposed parameter variances (Haario, Saksman et al. 2001; Marshall, Nott et al. 2004; Smith and Marshall 2008). This is a simplification which assumes no correlation between parameters. During the algorithm execution, accepted parameter and error values after an initial chain length, set at 5% of the stopping chain length, were used to generate the next variance-covariance matrix. This data-based second variance-covariance matrix now includes the correlation between parameters and error found from the initial accepted values.

Subsequent variance-covariance values are updated adaptively using a history of accepted parameter values. A multivariate sampling scale was set so as to obtain acceptance rates between 25% and 45%, as recommended by Gelman et al (2004). Based on preliminary runs to convergence and on execution times, the stopping criterion was set at 100,000 MCMC iterations. The warm-up period, which is removed from analysis, is half the MCMC chain (Gelman, Carlin et al. 2004).

The Euclidean distance, a common measure of proximity performance (Lumijärvi, Laurikkala et al. 2004; Borkowski and Piepel 2009) was the main evaluation criteria used for this research. For all situations, the Bayesian Adaptive Metropolis (Haario, Saksman et al. 2001) algorithm was used as the optimization algorithm of choice assuming that all observed data were normally and independently distributed. To block variability, all the other factors such as random number seeds, starting values, processors input datasets and Bayesian Adaptive Metropolis algorithm settings, except scaling factors, were kept constant.

Results and Discussions

To minimize the volume of results presented, the mean distances of the parameter values at maximum posterior, from the theoretical optimum points, are the main performance measures reported on.

Effect of Size of Normal Observed Data Error

The results in Figure 19 are average Euclidean distances from the theoretical optimum for the MLR and RBF approximation experiments on the FLO model. These are results from computer experiments in the reduced (HPD) region and averaged over all

(50, 200, 600 and 900) RSM fitting points. MCMC direct, MLR and RBF refers to the results from optimizing directly on the FLO model using the MCMC algorithm, and from optimizing using the fitted models MLR and RBF, respectively.

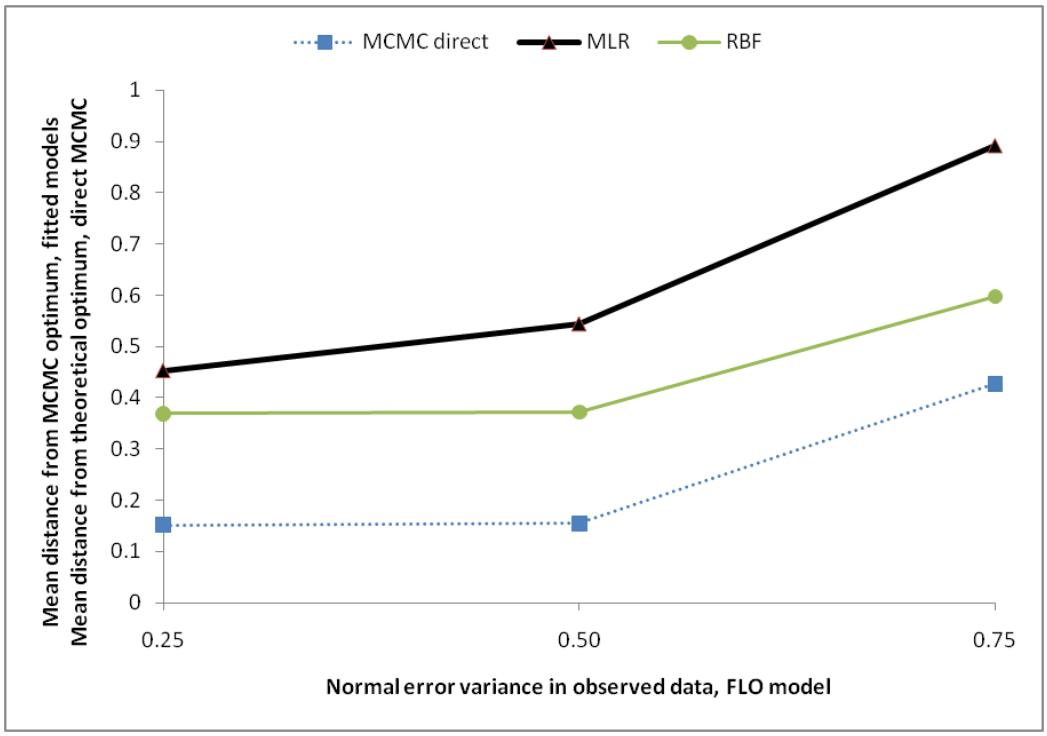

Figure 19: Effect of observed-data normal error size for the hydrologic bucket (FLO) model.

The MCMC direct line shows the performance of direct MCMC optimization with distance computed from theoretical optimum while the MLR and RBF lines have distances computed from the optimum MCMC points.

Figure 19 shows that increasing normal error in the observed data reduces the performance of the MCMC optimization on the FLO model (based on theoretical optimum) and also reduces the performance of MCMC optimization on the fitted models. The rate of performance degradation when using MLR fitting is greater than when the

MCMC algorithm is applied to the FLO model directly. This suggests that the MLR fitting gets worse with increasing observed data normal error, even after isolating the performance of the MCMC algorithm. On the other hand, the rate at which optimization performance declines when using RBF fitting is slightly less than when optimizing directly for the model using MCMC, which implies that the RBF fitting has a smoothening effect to the size of normal error in the observed data. For brevity, results for FCOS and FX2, which were similar to those of AWBM and FLO models, are excluded.

Performance of the Approximating Functions

The performance of the MLR and RBF function approximations for MCMC optimization of the FLO, FCOS, FX2 and AWBM models are summarized in Figure 20. This time, the distances presented for the scenarios when using the MLR and RBF fitting functions are from the optimum parameter values obtained by direct MCMC optimization to enable assessment of the appropriateness of the response surface for each model outside of the MCMC results. The case study AWBM also has no theoretical optimum values to assess the fitting the performances from, as is the case with the synthetic models. Performances in experiments in the reduced (HPD) region and for all the number of RSM approximating points and using only the default (Level 2) size of normal error variance for the synthetic observed data are averaged and presented in Figure 20.

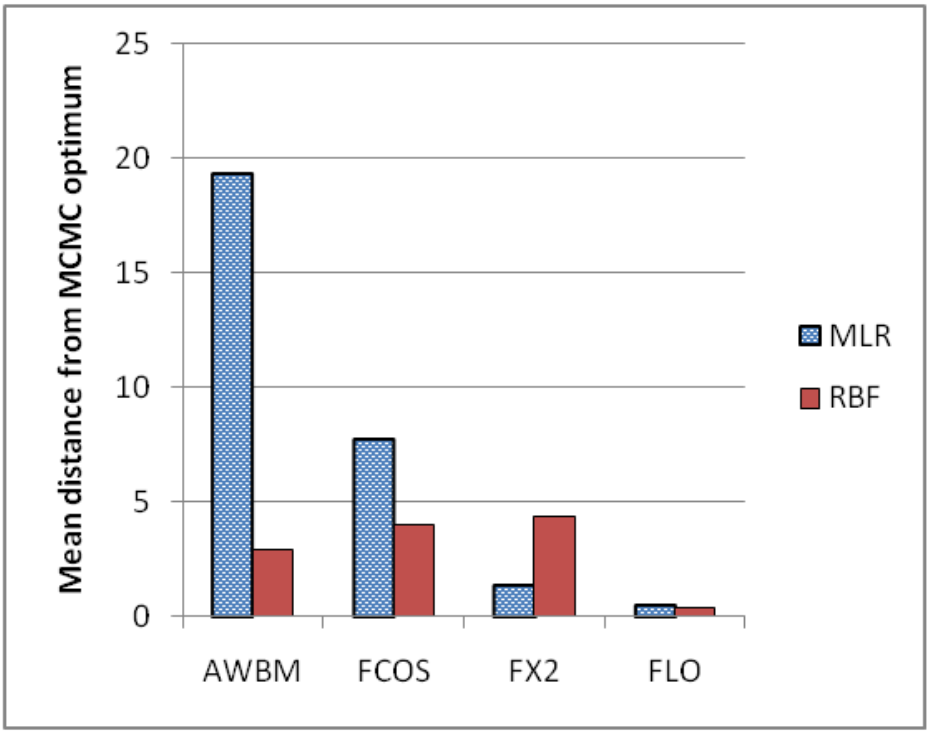

Figure 20: Mean distance from MCMC optima for the AWBM, FCOS, FX2, and FLO models using MLR and RBF approximation

Note that the mean distances in Figure 20 do not mean that the approximating models were performing better for FLO than for AWBM as each model has a different number of dimensions and uses different data. Instead, they show the respective performances of each approximating function (radial basis function, RBF, vs. multiple linear regression, MLR) for each approximated model. Figure 20 shows that RBF scenarios obtained closer parameter approximates, on average, for AWBM, FLO and FCOS models. For the case study model, AWBM, the MLR was about 6.6 times farther from the MCMC optimum values than was RBF. However, for the FX2 model, the RBF's mean distance of 4.370 from the MCMC optimum was about 3.2 times farther than MLR's 1.351. To demonstrate the superior performance of MLR for the FX2 model, Figure 21 shows that MLR approximation performed better than the exact model when

mean distances a measured from the theoretical optimum. This superior performance is generally expected, given that both the approximated and approximating functions are of the same second order polynomial family.

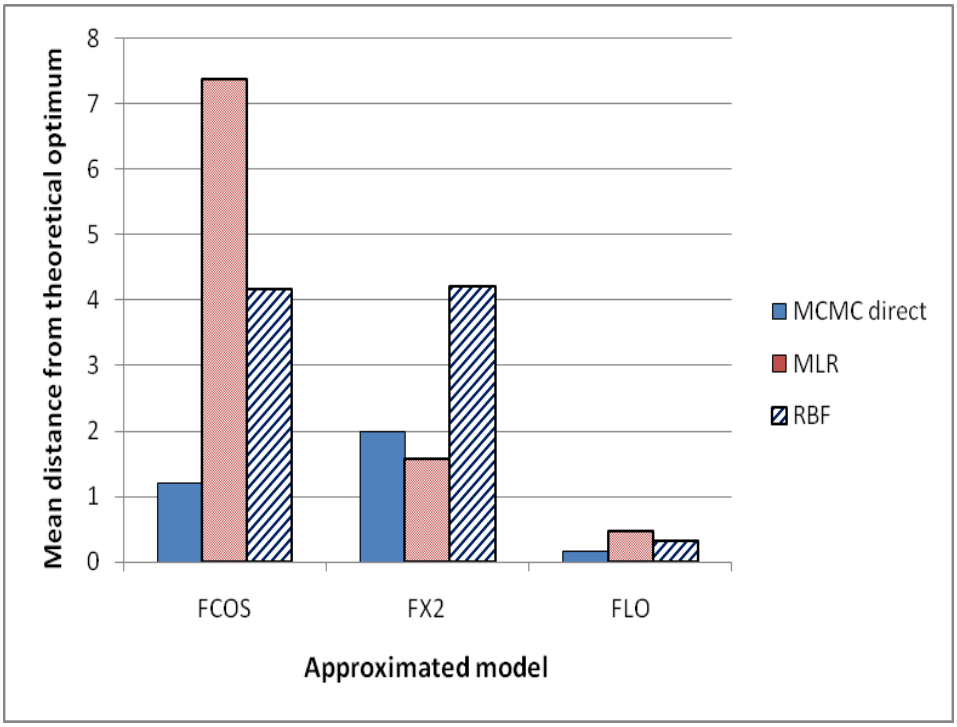

Figure 21: Mean distance from theoretical optima for the FCOS, FX2, and FLO models using MLR and RBF approximation and direct MCMC optimization.

Mean distance data for Figure 21 suggest that overall, using MLR for approximation produced values that were about 2.48 times farther from the MCMC optimum values than when using RBF approximation. By considering the synthetic cases only, and thus basing our distances from the theoretical optima, Figure 21 shows that RBF was only about 4.6 times farther than optimizing for the actual models, while MLR is about 7 times farther off the mark, when totaled for the three approximated models.

Effect of the Number of Fitted RSM Points

To investigate how the number of fitted RSM points affects the performance of the fitting functions, results from two sizes of sampling domains (HPD and EXT) were investigated for each test model. Four settings of fitted points (50, 200, 600 and 900) were used in the analysis. The results are given in Figure 22. MLR and RBF refer to MCMC optimization on fitted second order multiple linear regression and on the fitted radial basis functions respectively.

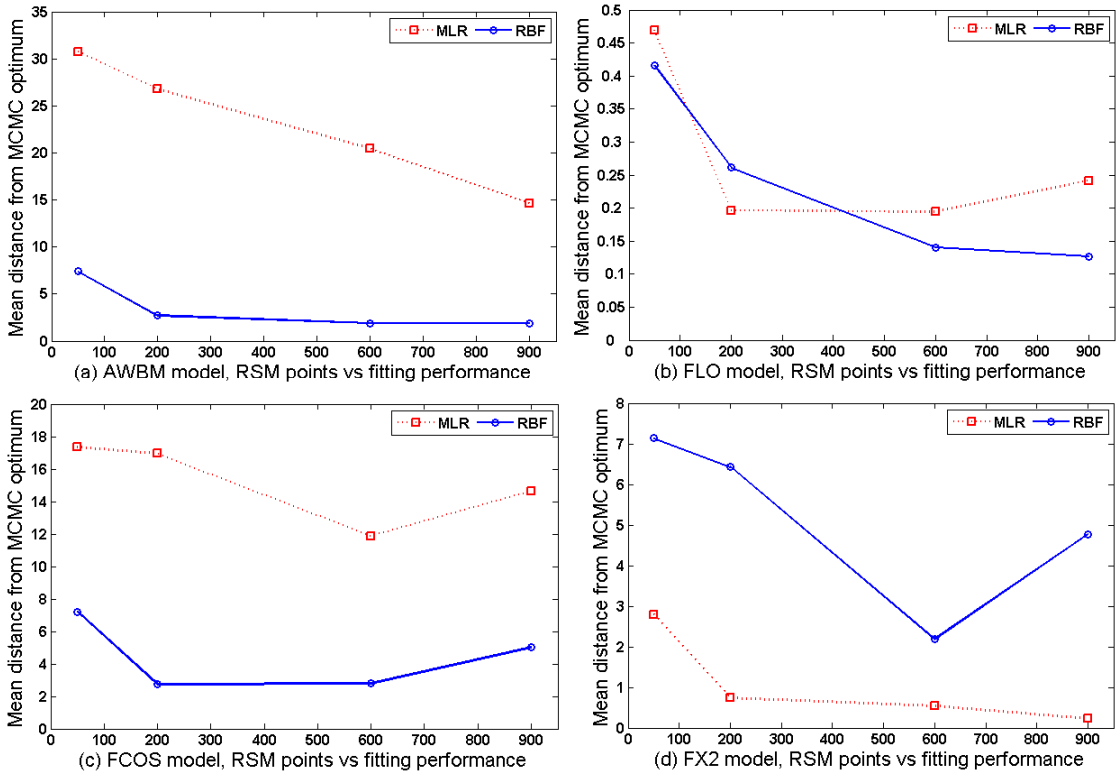

Figure 22: Optimization performance versus number of RSM points (50, 200, 600, and 900) used for fitting.

Figure 22(a) of the case study AWBM model shows that for both the MLR and RBF fitting scenarios, the higher the number of RSM points used for fitting the closer the optimum points get to the MCMC optimum, and hence the better the fit. The percentage

improvement on the accuracy of the MLR and RBF fitting scenarios with MCMC search was 33% and 62% for an 18 fold increase in RSM points (from 50 to 900 points), respectively. Figure 22(a) also shows diminishing parameter accuracy improvements as the number of RSM points used for fitting increases.

Figure 22(b) shows that the performance of MCMC optimization on the fitted MLR function seems to suffer slightly from increasing RSM points beyond 200. Such deterioration in performance, with increasing number of RSM fitted points, is also observed in Figure 22(c) and 22(d). The deterioration in performance as RSM increase beyond a certain number may be explained by two reasons. The first explanation could be that the higher numbers of RSM points used for fitting provide points that are too close together, making the algorithm's matrix factorization and inversion less stable, as the matrix becomes more singular with increasing similar points. A more likely explanation is that the contrived response surfaces of the synthetic models, which are sharply peaked at the global optima and relatively flat everywhere else in the experimental domain, result in the majority of sampled points lying in the featureless region of the response surface. Thus increasing RSM fitted points increases the distortion caused by points lying in the insensitive and featureless region of the response surface, and hence weakening the overall performance of the algorithm more.

Overall Figure 22(a - d) shows that increasing the number of RSM fitting points improves the performance of fitting with diminishing returns. Results in Figure 22 may also be interpreted as showing that there possibly is some optimum number of RSM fitting points dependent on the model properties and parameter space complexity, for which function approximation produces the best fitting performance. Further

experimentation using a greater range of RSM numbers and more case study models may be warranted.

Effect of the Size of Sampling Space

The summarized results presented are average performances, based on distance from MCMC optimum, over all the numbers of RSM points used for fitting. Figure 23 shows that, as expected, reducing the sampling space (from EXT to HPD) improves optimization accuracy for FLO, FX2 and FCO for all scenarios. The mean distances from the MCMC optima, in Figure 23, should not be compared across FLO, FX2 and FCOS since the values of the parameters determine their scales. The distances should only be used as relative quantities within each approximated model. The EXT and HPD are the extended and reduced high posterior density sampling regions for the test models AWBM, FCOS, FX2 and FLO.

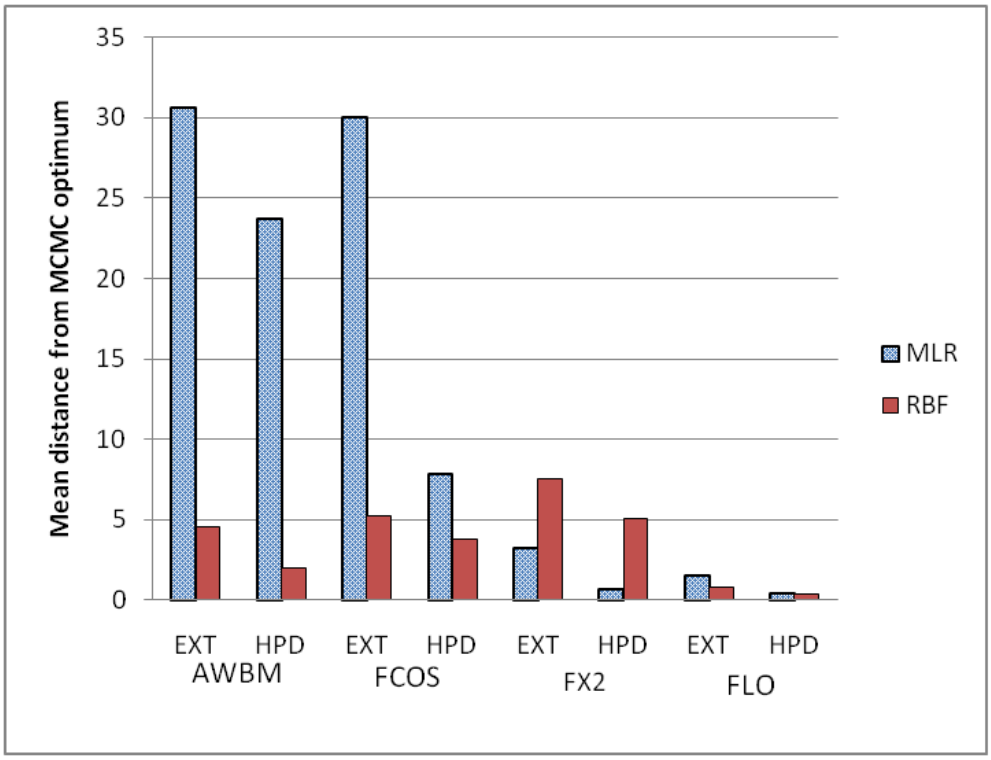

Figure 23: Performance of optimization on fitted MLR and RBF functions based on size of sampling region.

Reducing the parameter sampling ranges by as much as 70 percent resulted in significant improvement in optimization performance for the MLR and RBF fitted functions across all models. Although some of the resulting gains in accuracy are not very high, the fact that there was always some improvement in accuracy from reducing the experimental domain strongly advises on the need to have a reduced HPD region before RSM fitting for both MLR and RBF surface approximations.

Representative Maximum Posterior Parameters

Tables 11 to 14 summarize maximum posterior parameters, excluding the error variance estimates, for some representative experimental settings for each of the three approximating and approximated functions for a single error size.

Table 11: FLO model maximum posterior parameters after MCMC convergence for four experimental settings

Experimental setting	Parameters					Distance
	alpha	beta	delta	lambda	min_srf	
True values	0.8	0.3	0.5	0.6	1.5	
MCMC	0.7909	0.3201	0.4767	0.5879	1.5974	0.10326
MCMC + MLR + HPD + 200 RSM points	0.8660	0.2983	0.6000	0.5001	1.2001	0.3381
MCMC + RBF + HPD + 600 points	0.8480	0.2948	0.5922	0.5000	1.4449	0.1544

The first experimental setting (MCMC + FLO + HPD), in Table 11, is when the MCMC algorithm was acting on the FLO model directly using the HPD region, hence has no fitted RSM points. Since the fitted points were the only actual function evaluations during the MCMC algorithm's 100,000 runs for the RSM approximation experiments, the results were notably close to when AM was evaluating the actual FLO model 100,000 times. Also the reduced (HPD) region provided all the best maximum posterior parameters in the FLO experimental settings.

Table 12: FCOS model maximum posterior parameters after MCMC convergence for four experimental settings

Experimental setting	Parameters		Distance
	U1	U2	
True values	180.0000	30.0000	
MCMC	180.6225	30.0841	0.6282
MCMC + MLR + HPD + 600 points	183.2105	30.3441	3.2289
MCMC + RBF + HPD + 200 points	179.7626	30.4308	0.4919

Table 12 details results for the FCOS model when using MCMC directly in comparison to two different RSM case studies. The results show that some approximation setups performed marginally better than when optimization was directly on the FCOS model.

Table 13: FX2 model maximum posterior parameters after MCMC convergence for four experimental settings

Experimental setting	Parameters		Distance
	A	B	
True values	6.0000	-80.0000	
MCMC	5.9998	-80.0000	0.0002
MCMC + MLR + EXT + 900 points	6.0319	-80.1945	0.1971
MCMC + RBF + HPD + 900 points	6.0002	-80.1033	0.1033

Table 13 details results for the FX2 model. The table shows that if we exclude the error variance approximation from performance estimates (distances), the best of MLR is slightly poorer at optimizing for parameters compared the best of RBF approximation or of direct optimization on the FX2 model.

Table 14: AWBM model maximum posterior parameters after MCMC convergence for four experimental settings

Experimental setting	Parameter values			Distance from MCMC
	S	BFI	K_S	
MCMC	210.8821	0.4882	0.9307	
MCMC+MLR +HPD + 900 points	203.2037	0.5306	0.9999	7.6788
MCMC+RBF+ EXT + 900 points	209.0013	0.4906	0.9006	1.8810

Finally, Table 14 shows that the distance from the MCMC optimum, for the AWBM case study model, was created mostly from the S parameter. Further

investigation showed that parameter S, the surface storage capacity, was the least sensitive of the three.

Summary and Conclusions

This study aimed at investigating how (i) the approximation function used and the type of model or response function approximated, (ii) the number of points used in the actual curve fitting, (iii) the size of the region from which fitting points are sampled, and (iv) the size of normal observed data error affect the performance of RSM approximation for Bayesian MCMC calibration and uncertainty analysis.

The main finding of this study is that the performance of fitting and optimization was clearly negatively correlated to the size of sampling domains of the parameters and error variance used. In all the experimental settings, reducing the parameter and error variance sampling ranges improved the RSM fitting and hence Bayesian optimization results. More importantly, fitting fewer points on a smaller range gave better results than fitting many more points on a larger region. That is, a smaller neighborhood of points fitted better than higher densities of points over a much larger sampling domain. We posit that this is likely due to the response surface of typical hydrologic models, where the likelihood is sharply peaked over a small portion of the domain (Beven and Freer 2001). Hence despite the computational burden in locating the HPD region or in designing a robust sampling strategy, the HPD region may be critical for acceptable performance of RSM fitting for use with Bayesian optimization and uncertainty assessment.

Using MCMC to optimize over an approximate response surface created by the spline based RBF function performed better overall than the second order multiple linear

regression (MLR) for the case study (AWBM) discharge, the synthetic bucket discharge (FLO) and the cosine (FCOS) models . However, fitting a second order MLR function resulted in MCMC optimization that was 3.4 times better than for the RBF fitting scenarios, for the multivariate second order function (FX2). The overall marginal superiority of MLR over RBF in approximating for FX2 was likely because MLR and FX2 were same order polynomials. Even then, the best of RBF performed better than the best of MLR when used in optimizing for FX2, if error variance estimation is excluded.

Generally, increasing the number of points used in fitting improves the accuracy of fit and hence parameter optimization. However, it is likely that beyond some value, more RSM points result in diminishing improvements or may even begin to negatively affect the performance of approximation. A likely explanation for this is that the contrived response surfaces of the synthetic models, which are sharply peaked at the global optima and relatively flat everywhere else in the experimental domain, result in the majority of sampled points lying in the featureless region of the response surface. Thus increasing RSM fitted points increases the distortion caused by points lying in the insensitive and featureless region of the response surface, and hence weakening the overall performance of the algorithm more.

While the computational time of the algorithms was not a focus of this study, the average clock times for RBF, MLR and direct Bayesian MCMC over the test models were 309.32, 313.42 and 3587.23 seconds respectively. These average times were from fitting MLR and RBF functions based on 50 to 900 actual model runs followed by 100,000 runs on resultant approximate response surface versus optimizing for the direct models using 100,000 actual model runs. RSM approximation gave 11 fold time savings

and 100 times fewer actual function evaluations compared to when using direct model evaluation. The results show that the loss of parameter information from using RSM approximation can easily be acceptable considering the potential computation time gains, especially for computationally expensive models.

The size of the normal error in the observed data significantly affects the performance of second order multiple linear regression (MLR) and radial basis function (RBF) fitting performance. In non-synthetic studies, this variable is akin to the model residual error and/or observation error. MLR performs worse than RBF as the normal error in observed data increases. Parameter optimization for the FLO model using MCMC directly was also sensitive to changes in size of observed data error variance than when optimizing for the fitted MLR and RBF functions.

Poor algorithm performance caused by the presence of a large variance of normal error may be slightly dampened by using a radial basis approximating function during Bayesian uncertainty assessment, resulting in some smoothing effect of the optimization surface. Very few actual function evaluations used with approximation can give very close optimization results compared to when using large actual function evaluations without approximation. The use of few evaluations during RSM approximation would be beneficial for computationally time-expensive models, such as distributed hydrologic models.

Based on these findings, for the models considered in this study RSM approximation for Bayesian optimization performed best when the fitted to the HPD region, with design points that are not too many or too few, when model or data error is not excessive and when using the thin plate spline radial basis function as the

approximating function over the second order function. That is, the number of design points is less critical than where they are sampled from. Sampling from the HPD results in better performance of RSM fitting for Bayesian optimization than sampling from an extended region, as might happen when there is poor prior knowledge of the HPD region before fitting.

Contribution of Authors and Co-Authors

Manuscript in Chapter 4:
Bayesian Contrained Optimization and Uncertainty Analysis Using Radial Basis Random Local Fitting

Author: Able Mashamba
Contributions: Overall concept development, experimental design, models development and assessment, results and manuscript write-up.

Co-author: Lucy Marshall
Contributions: Study supervision and oversight and provision of the Australian Water Balance case study model and data.

Manuscript Information

Able Mashamba and Lucy Marshall

Journal of Stochastic Environmental Research and Risk Assessment

Status of manuscript (check one):

_✓_ Prepared for submission to a peer-reviewed journal

___ Officially submitted to a peer-reviewed journal

___ Accepted by a peer-reviewed journal

___ Published in a peer-reviewed journal

Publisher: Springer

CHAPTER 4

BAYESIAN CONSTRAINED OPTIMIZATION AND UNCERTAINTY ANALYSIS USING RADIAL BASIS RANDOM LOCAL FITTING

Abstract: This study describes improvements made by randomized local radial basis function fitting compared to global fitting during constrained Bayesian model calibration and uncertainty analysis as applied to hydrologic models. Random local fitting involves using a randomly selected few of the previously evaluated neighbor points for response surface methodology (RSM) fitting. Global fitting involves using all the previously evaluated design points for radial basis function (RBF) fitting (Bliznyuk, Ruppert et al. 2008; Mashamba and Marshall 2010).

In hydrologic modeling, distributed models are often computationally expensive to automatically calibrate and assess for uncertainty due to their many parameters and the resulting large numbers of model runs required (Wooldridge, Kalma et al. 2001; Wagener and Wheater 2006; Muttil, Liong et al. 2007; Rouhani, Willems et al. 2007). The use of radial basis function global fitting and Bayesian inference optimization has been shown to be effective for automated hydrologic model calibration and uncertainty analysis within a constrained highest posterior density (HPD) region (Bliznyuk, Ruppert et al. 2008). However, for effective global fitting, prior actual model evaluations are often required to locate the parameter HPD region (Bates and Campbell 2001; Gelman, Carlin et al. 2004; Regis and Shoemaker 2004; Regis and Shoemaker 2005; Regis and Shoemaker 2007; Bliznyuk, Ruppert et al. 2008), which adds to the computational burden.

To reduce excessive prior actual model evaluations, larger parameter sampling domains may be used to represent limited prior knowledge of the parameter bounds. However, extended parameter sampling domains contain greater surface complexities that significantly degrade the quality of global surface fitting and hence the optimization and uncertainty analysis based on the RSM. One case study and 3 theoretical models were used to test the quality of calibration and uncertainty analysis results obtained from the two fitting approaches for HPD and extended sampling domains.

The closeness (Euclidian distance) of the optimum points obtained through fitting from those obtained by actual model evaluations during Bayesian Markov chain Monte Carlo (MCMC) was the main performance measure for calibration. Means and standard deviations of the convergent parameter posterior distributions were also used as measures of performance of the MCMC uncertainty analysis on fitted responses.

The randomized local fitting routine adopted for Bayesian model calibration performed better overall, and especially for the extended sampling domains than global fitting. Although the global and random local fitting schemes were tested using only 400 previously evaluated points from SLHD sampling (for the four test models), the comparison of the two fitting performances is expected to remain the same under some other more elaborate adaptive sampling schemes. Also, a more robust DRAM (Haario, Laine et al. 2006) may replace the basic adaptive MCMC algorithm to improve the parameter distributions during uncertainty analysis on fitted models.

Keywords: Radial basis function, random local fitting, global fitting, response surface methodology, design points, Markov chain Monte Carlo.

Introduction

Research in the use of the response surface methodology (RSM) for approximating computationally expensive functions (Jones, Schonlau et al. 1998; Booker, Dennis et al. 1999; Bjorkman and Holmtrom 2000; Gutmann 2001(a); Regis and Shoemaker 2004; Regis and Shoemaker 2007) can be and has been useful for the calibration and uncertainty analysis of distributed hydrologic models (Mugunthan, Shoemaker et al. 2005; Mugunthan and Shoemaker 2006; Bliznyuk, Ruppert et al. 2008; Mashamba and Marshall 2010). The RSM approach involves the development of approximate functions (aka surrogate models) that map input variables to some response variable(s) or objective function(s) of an unknown or difficult to evaluate original or exact function. The resulting surrogate functions are typically much faster to evaluate than the original or exact function while providing comparable approximate responses within pre-determined experimental domains (Regis and Shoemaker 2007; Bliznyuk, Ruppert et al. 2008).

In hydrologic modeling, distributed models are often computationally expensive to automatically calibrate and assess for uncertainty due to their many parameters and the resulting large numbers of model runs required (Wooldridge, Kalma et al. 2001; Wagener and Wheater 2006; Muttil, Liong et al. 2007; Rouhani, Willems et al. 2007). Thus RSM function such as multiple linear regression, radial basis functions and other interpolation functions (Powell 1996; Marazzi and Nocedal 2002; Bliznyuk, Ruppert et al. 2008) set up to map how the model responds to changing parameter values have been used in finding approximate parameter values that give minimum model residuals.

When the interest is in automated model calibration or function optimization only, RSM functions can be directly evaluated to give the global optimum of the fitted model (Conn, Scheinberg et al. 1997; Booker, Dennis et al. 1999; Bjorkman and Holmtrom 2000; Powell 2000; Gutmann 2001(a); Marazzi and Nocedal 2002; Regis and Shoemaker 2007). Instead of direct RSM function optimization, search algorithms can also be used to find the optimum on the fitted function (Regis and Shoemaker 2004; Regis and Shoemaker 2005; Bliznyuk, Ruppert et al. 2008).

When uncertainty analysis is as important as model optimization, as in an increasing number of hydrologic modeling scenarios (Tyagi and Haan 2001; Balakrishnan, Roy et al. 2003; Moradkhani, Hsu et al. 2005; Arabi, Govindaraju et al. 2007; Beven, Smith et al. 2007; Liu and Gupta 2007; Blasone, Madsen et al. 2008), an approach initially using search algorithms over the RSM approximate function may be used (Regis and Shoemaker 2004; Regis and Shoemaker 2005; Bliznyuk, Ruppert et al. 2008). In such a setup, the RSM method approximates the model objective function (such as sum of squared residual or the likelihood) while a local or global search algorithm finds the optimum parameters optimizing the objective function for that approximated surface (Regis and Shoemaker 2004; Regis and Shoemaker 2005; Bliznyuk, Ruppert et al. 2008).

The use of RSM approximation methods and Bayesian inference optimization (Bliznyuk, Ruppert et al. 2008) has been shown to be effective for automated hydrologic model calibration and uncertainty analysis within a constrained region that is highly likely to contain the true optimum (Bliznyuk, Ruppert et al. 2008). Prior model evaluations are often required to locate this tightly constrained domain called the highest

posterior density (HPD) region that has a high likelihood of containing the true optimum model parameters (Bates and Campbell 2001; Gelman, Carlin et al. 2004; Regis and Shoemaker 2004; Regis and Shoemaker 2005; Regis and Shoemaker 2007; Bliznyuk, Ruppert et al. 2008). Generally this involves rough-cut global optimization to determine the HPD followed by RSM approximation of the model residual and Bayesian fine-tuning local optimization within the HPD (Gelman, Carlin et al. 2004). However finding the HPD region by automated calibration routines uses actual model evaluations, which come with increased computational overhead. It would be desirable to utilize an extended parameter sampling region with little or no prior model evaluations for sampling RSM fitting points and Bayesian optimization. This research demonstrates that it is possible to use a much larger parameter sampling domain than the HPD, perhaps representing poorer prior knowledge about the HPD from fewer model evaluations, but still successfully use RSM fitting and Bayesian optimization and uncertainty analysis.

From earlier work by Mashamba and Marshall, in review, Bayesian calibration performs better at optimizing over a more extended parameter sampling domain than does RSM global fitting over the same-sized domain. We use an algorithm heavily adapted from concepts by Regis and Shoemaker (2004) and Bliznyuk et al (2008) to enhance the performance of RSM radial basis function fitting with initial parameter design points sampled from an extended region (EXT) multiple times larger than, and containing, the HPD region. Although this study used adaptive Bayesian MCMC (Haario, Saksman et al. 2001), spline radial basis functions (Powell 1992; Powell 1996; Booker, Dennis et al. 1999; Powell 2000; Gutmann 2001(a); Powell 2002; Buhmann 2003; Regis and Shoemaker 2004; Regis and Shoemaker 2005; Regis and Shoemaker

2007; Bliznyuk, Ruppert et al. 2008), symmetric Latin hypercube design sampling (SLHD) (Ye et al. 2000), and non-increasing actual function evaluations, the concept of local fitting of randomly selected nearest neighbors can be used with other search algorithms, approximating functions and sampling schemes for constrained model optimization and uncertainty analysis.

There are three main differences between the random local function approximation algorithm proposed here for optimization and uncertainty analysis and a local approximation algorithm by Regis and Shoemaker (2004). First, the proposed algorithm uses a randomly selected decreasing number of nearest neighbors for local fitting while Regis and Shoemaker's 2004 algorithm uses a fixed (d+1)(d+2)/2 minimum number of neighbors for curve fitting. Random local fitting allows for escaping possible entrapments in local fitting subspaces that have localized optima. Decreasing the numbers of neighbors for RSM fitting as the local Bayesian search progresses allows for a shift from global to local optimum fitting. Second, the proposed algorithm uses a single Markov chain Monte Carlo (MCMC) chain while Regis and Shoemaker's (2004) evolutionary algorithm requires multiple chains, from which to select the best estimates, to be used for actual costly function evaluations. Finally, Regis and Shoemaker's (2004) algorithm updates the fitted RSM function after approximating the objective functions of each chain once only. This is because the genetic algorithm approach used by Regis and Shoemaker (2004) requires the selection of 'fittest survivors' after only a few generations or iterations, which are then used for updating the RSM fitting points. On the other hand, the proposed MCMC based algorithm updates points used for fitting after a user defined number of iterations, to reduce excessive fitting computations.

The main concept from Bliznyuk et al (2008) adopted in the proposed algorithm is the use of a Bayesian inference (Metropolis-Hastings) algorithm with RSM fitting calibration and uncertainty analysis in environmental modeling. The Bliznyuk et al (2008) algorithm relies on prior optimization using Powell's (2002) UOBYQA derivative free trust region optimization algorithm to help locate the reduced HPD parameter sampling domain. The Unconstrained Optimization by Quadratic Approximation (UOBYQA) algorithm starts off with a minimum number,$(d+1)(d+2)/2$, of previously evaluated objective function points required for quadratic curve fitting. In the UOBYQA algorithm, the distance between the overall exact and latest fitted (quadratic) optima is used to adjust the jump step during the search. A shorter distance means a shorter jump step (since close to the trust region) and conversely a longer distance means a larger jump step (being away from the trust region). Based on the UOBYQA, Bliznyuk et al (2008) proposed the HPD region as being made from design points sampled close to the overall exact optimum, after an initial space filling design made from the minimum number of points. Once the HPD is located, all the previously evaluated design points are simultaneously used in the RSM function approximation. We call this, global fitting in this study. The proposed method (on the other hand) uses only a randomly selected fraction of the available design points for fitting, which we refer to as random local fitting.

To evaluate the performance of random local fitting versus global fitting during Bayesian Markov chain Monte Carlo (MCMC) model calibration and uncertainty analysis, a number of experiments and performance measures were used. A case study and three synthetic hydrologic-biased models were calibrated by using both a random

locally and globally fitted spline radial basis function coupled to an adaptive MCMC calibration routine. An HPD parameter sampling region was determined a priori from MCMC calibration using the actual models and an extended (EXT) region was arbitrarily proposed to be thrice larger than and to contain the HPD for all four of the models. The scale of this extended region is on par with the feasible parameter ranges for similar hydrologic models. The same RSM design points (400 actual model evaluations) were used for the two RSM fitting approaches and all the other experimental factors were held constant. Thus the Euclidean distance between the optimum parameters found from the global and random local RSM fitting schemes, and the optimum parameters from evaluating the exact model during Bayesian MCMC calibration would in fact be a relative measure of the accuracy of each fitting scheme. Shorter distances infer better accuracy of RSM approximation when all the other experimental factors are invariant. Descriptive statistics were used to characterize the MCMC convergent parameter distributions from both fitting schemes for the four models.

Radial Basis Function Model

The radial basis interpolation function has been extensively studied by Powell (1992, 1996, 2000) and Buhmann (2005), and used by Gutman (2001a), Regis and Shoemaker (2004, 2005, 2007), Bliznyuk et al (2008) in optimization and uncertainty analysis. A radial basis function (RBF), $S_n(x)$, is a real-valued function whose value depends only on the distance of x from the origin or from some given point called a center so that given n distinct points $x_1, \ldots, x_n \in R^d$, where the function values $f(x_1), \ldots, f(x_n)$, are known, the RBF is an interpolant of the form as in Equation 4.1.

$$S_n(x) = \sum_{i=1}^{k} \omega_i \, \phi(\|x - x_i\|) + \; p(x), \; x \, \epsilon \, R^d \tag{4.1}$$

Where $\|.\|$ is the Euclidean norm, $\omega_i \, \epsilon \, R$ for i = 1, …, n, $p \, \epsilon \, \Pi_{\mathrm{m}}^{\mathrm{d}}$ (a linear space of polynomials in d variables of degree less than or equal to m), and ϕ is either $\phi(r) = r^2 \log r \, , r > 0$ and $\phi(0) = 0$ (thin plate spline), $\phi(r) = r^3$ (cubic), $\phi(r) = \sqrt{r^2 + \gamma^2}$ (multi-quadratic), or $\phi(r) = e^{-\gamma r^2}, \; r \geq 0$ where γ is a positive constant (Gaussian) (Regis and Shoemaker 2004). The ω_i is a weight showing how the Euclidean distance for point x_i from a sampled point x contribute to the approximation of response at point x. The tail of the radial basis function, $p(x)$, is often linear with respect to the explanatory variables, but higher order polynomials could also be used. The thin plate spline and the quadratic functions have been reported to have superior fitting performance especially for applications similar to the case studies presented here (Buhmann 2003; Regis and Shoemaker 2004; Regis and Shoemaker 2005; Bliznyuk, Ruppert et al. 2008).

To find the coefficients ω_i and of tail function $p(x)$ one proceeds as follows; Select a particular ϕ. Define the matrix $\Phi \, \epsilon \, R^{nxn}$ by $\Phi_{\mathrm{ij}} := \; \phi(\|x_i - x_j\|), i, j = 1, \dots, n$. Define m_ϕ to -1 if ϕ is Gaussian, 0 if ϕ is multi-quadratic and 1 if ϕ is cubic or thin plate spline. Let $m \; \geq m_\phi$ and let $\widehat{m}$ be the dimension of the linear space $\Pi_{\mathrm{m}}^{\mathrm{d}}$, where $\widehat{m} = \binom{m+d}{d}$. Also let $p_1, \dots, p_{\widehat{m}}$ be a basis of $\Pi_{\mathrm{m}}^{\mathrm{d}}$, and define the matrix $P \, \epsilon \, R^{nxn}$ as follows: $P_{ij} := p_j(x_i), \; i = i, \dots, n; \; j = 1, \dots, \widehat{m}$. The coefficients of the RBF function, $S_n(x)$, that interpolates the points $\big(x_1, f(x_1)\big), \dots . , \big(x_n, f(x_n)\big)$ are obtained by solving the Equation 4.2 matrix system;

$$\begin{pmatrix} \Phi & P \\ P^T & 0 \end{pmatrix} \begin{pmatrix} \omega \\ c \end{pmatrix} = \begin{pmatrix} F \\ 0 \end{pmatrix} \tag{4.2}$$

where $F = (f(x_1), \ldots, f(x_n)^T$, $\omega = (\omega_1, \ldots, \omega_n)^T \epsilon R^n$ and $c = (c_1, \ldots, c_{\hat{m}})^T \epsilon R^{\hat{m}}$. The coefficient matrix system in Equation 4.2 is invertible if and only if the matrix *P* defined above has full column rank (Powell 1992). Therefore as long as long as $n \geq \hat{m}$ the matrix in Equation 4.2 above will be invertible and remain so with the addition of new data points that are distinct from the previous ones (Regis and Shoemaker 2004).

MCMC Approach

Markov chain Monte Carlo (MCMC) methods are a class of Bayesian inference algorithms for sampling from an unknown probability distribution based on constructing a Markov chain (Markov 1906) that has the desired distribution as its equilibrium distribution. The state of the chain stabilizes after a large number of steps and is then used as a sample from the unknown desired distribution.

By Bayes theorem, given a joint prior density $P(\Theta)$, the joint posterior density after Y observations is as given in Equation 4.3;

$$P(\Theta|Y) = \frac{P(Y|\Theta).P(\Theta)}{P(Y)} \tag{4.3}$$

where $P(Y|\Theta)$ is the joint likelihood and $P(Y)$ is a (difficult to compute) marginal probability density of data. In hydrologic modeling, Θ and Y are typically vectors of model parameters and observed response respectively (Gelman, Carlin et al. 2004; Huard and Mailhot 2006; Marshall, Nott et al. 2007; Smith and Marshall 2008). Setting up a Markov chain and computing the ratios of the current and previous chain states based on Equation 4.3 avoids the usually intractable computation of $P(Y)$ (Gelman, Carlin et al. 2004). To hasten the convergence of a Markov chain based on Equation 4.3, the likelihood and /or the prior must be closer to the unknown posterior. Often though,

uninformative uniform continuous distributions suffice as priors in hydrologic modeling where the time-series observed dataset Y is often large enough to overwhelm poor estimation of the prior (Gelman, Carlin et al. 2004; Marshall, Nott et al. 2004). Multivariate random normal probability distributions are popularly used (Haario, Saksman et al. 2001; Gelman, Carlin et al. 2004; Marshall, Nott et al. 2004; Marshall, Nott et al. 2007; Smith and Marshall 2008) as the Markov chain transition probability in the Metropolis random walk versions of the MCMC algorithms. Details of different MCMC algorithms can be found in standard Bayesian inference texts.

Methods

The experimental methods used for this study are based on an earlier analysis (Mashamba and Marshall 2010), which examined factors affecting the performance of Bayesian calibration and uncertainty analysis using RSM fitting. To evaluate the performance of global versus random local fitting of RSM points during Bayesian MCMC calibration and uncertainty analysis, one case study and three synthetic models were used. The test models were a cosine and a quadratic function, a synthetic storage-based model of the land phase hydrology and a case study Australian water balance model. The cosine function provides non-linear periodic time series behavior, as might be observed with environmental data such as channel discharge. The second order polynomial function possesses the simplest form of curvature, quadratic curvature. The synthetic storage-based model for discharge and the Australian water balance case study model provide more model complexity, as well as being similar to other water balance hydrologic models.

For each of the four test models, two sizes of parameter sampling domains are used for fitting response surfaces. That is, parameter values used for fitting were either sampled from the high posterior density (HPD) or from extended (EXT) sampling domains. The HPD sampling domains were the ranges of parameter values after convergence obtained from prior direct Bayesian MCMC calibration and uncertainty analysis runs. The extended (EXT) parameter sampling domains are arbitrarily thrice bigger than, and inclusive of, the HPD region. While characterizing the HPD regions requires prior model knowledge, which itself usually takes prior model evaluations, the EXT region represents poorer prior knowledge of such tight sampling domains.

The main measure of RSM fitting performance was the Euclidean distance (Borkowski and Piepel 2009) between the optimum parameters found by using RSM approximation, for the global and random local fitting schemes, and the optimum parameters from evaluating the exact model during Bayesian MCMC calibration. Shorter distances infer better accuracy of RSM approximation, for either of the two fitting schemes, when everything else is invariant. Thus all the other experimental factors such as the Bayesian MCMC method, starting parameter values, sampling and number of RSM points and spline RBF fitting function were kept uniform across the global and local random fitting schemes. Other performance measures that were monitored included the processing times, and other descriptive statistics of the parameters produced by the two fitting scenarios.

Test Models

Equations 4.4 and 4.5 are for the quadratic (FX2) and cosine (FCOS) models, two of the test functions used for experimentation. The overall objective in using these

synthetic cases was to investigate how well the different fitting scenarios (global Vs random local fitting), coupled to MCMC optimization and uncertainty analysis, rediscovered the coefficients *A* and *B* in Equation 4.4 and U_1 and U_2 in Equation 4.5 used in creating the synthetic observed responses.

$$f(x_1, x_2) = 2x_1^2 - 0.5x_2^2 + Ax_1x_2 - Bx_1 - 20 \tag{4.4}$$

$$f(t) = U_1 \cos\left(\frac{U_2 \pi t}{180}\right), \quad 0 \leq t \leq 12 \tag{4.5}$$

To introduce the effect of variable dependency found in most environmental models, five years of maximum and minimum daily temperatures for a case study watershed were used as the explanatory variables x_1 and x_2 of Equation 4.4. The Equation 4.5 time series function, $f(t)$, used values of t from 0 to 12 at 0.025 intervals, while the theoretical optimum values of U_1 and U_2 values were specified as 180 and 30 respectively.

Complex response surfaces with more dimensions that have higher parameter dependency came from a synthetic lumped hydrologic flow model (FLO) and a case study Australian water balance model (AWBM). The FLO synthetic hydrologic flow model described processes of water storage and transfer between atmosphere, land surface, soil, shallow aquifer, deep aquifer and channels. These hydrological processes include evapotranspiration, surface runoff, lateral flow, revaporation, percolation, recharge, return flow and channel discharge. Out of 15 model parameters only the most sensitive 5 were selected for study while the rest were held constant. Five years of daily precipitation data for a case study watershed, the Lower Yellowstone River basin, was used as input to the hypothetical model, with all its 15 parameters set at specific values.

The hypothetical observed data of the synthetic case came from model output corrupted by a random normally distributed error, $N(0, 0.5)$.

Figure 24 shows a schematic of the AWBM case study model and the parameters that were used in experimentation. The AWBM model was developed in the early 1990s (Boughton 1993; Boughton and Carroll 1993; Boughton 2004) and is considered as one of the most widely used models in Australia (Boughton 2004) with studies by a number of researchers (Sharifi and Boyd 1994; Cakers and Yu 1998; Cheung and Yu 1999; Marshall, Nott et al. 2004). Slightly more than ten years (end of 1988 to 1998) of daily observed climate and discharge data from an Australian case study watershed studied by Marshall et al (2004) were used.

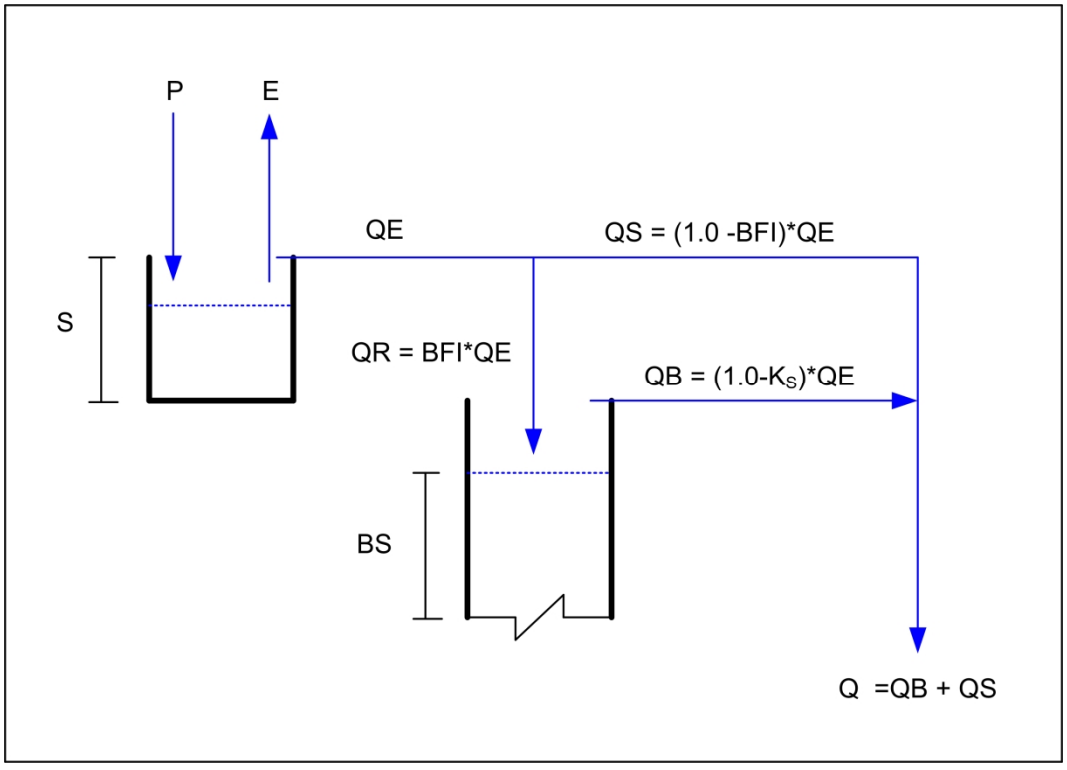

Figure 24: Schematic view of the AWBM case study bucket model.

The parameters S, BFI and K_S are surface water storage capacity, baseflow infiltration index and baseflow recession index respectively. The parameters S, BFI and K_S are considered variable for experimentation. The processes P, E, QE, QS, QR, QB and Q are precipitation, evapotranspiration, excess water available for surface flow, surface flow, recharge, baseflow and channel discharge respectively. Direction of water flow is shown by arrows and baseflow storage capacity is unlimited.

The Random Local Fitting Algorithm

The random local fitting algorithm is presented in six general but explicit steps below. The words in italics are algorithm parameters, those in bold serve as key instructive verbs and words in parentheses generally denote options that were used for this study. For instance, this study used symmetric Latin hypercube sampling of design points for RSM fitting. Other designs such as adaptive sampling could be used without changing the performance comparisons between global and local fitting.

1. **Set**:

   - total number of MCMC iterations, $maxchain$,
   - number of MCMC iterations before updating each fitted model, run_len,
   - number of MCMC iterations after which updating of fitted models ceases, $upchain$,
   - number of design points to be sampled from the experimental domain, $design\_pts$,
   - number of model dimensions for optimization, d,

- counters: MCMC iteration counter, $i = 0$; RSM update counter, $j = run\_len$,
- minimum number of fitting points, $min\_pts = (d+1)(d+2)/2$,
- parameters limits for the sampling domain, *(HPD or EXT)*.
- logarithm of likelihood function at MCMC iteration i, $loglikelihood[i]$
- current MCMC point (made from d elements) at ith iteration, $curr\_pt[i]$

2. **Sample** $design\_pts$ within experimental parameter sampling domain using space filling designs or adaptive sampling. (In this case study, symmetric Latin hypercube design (SLHD) sampling scheme was used)
3. **Evaluate** actual model MCMC likelihoods for each of the design points ($design\_pts$) in Step 2.
4. **Test** If MCMC iteration counter, $i \leq upchain$ AND RSM update counter, $j == run\_len$ THEN
    a. **Reset** RSM update counter, $j = 0$.
    b. **Reset** $rsm\_pts = U(min, max)$, (where rsm_pts is the number of design points used for RSM fitting) such that:
        i. **Test** IF MCMC iteration, $i \leq upchain/4$, THEN $rsm\_pts = U(\lceil 0.75 * design\_pts \rceil,\ design\_pts)$, ELSE
        ii. **Test** IF MCMC iteration, $i \leq upchain/2$, THEN $rsm\_pts = U(\lceil 0.5 * design\_pts \rceil,\ \lceil 0.75 * design\_pts \rceil)$, ELSE
        iii. **Test** IF MCMC iteration, $i \leq 3 * upchain/4$, THEN $rsm\_pts = U(\lceil 0.25 * design\_pts \rceil,\ \lceil 0.5 * design\_pts \rceil,$
        iv. ELSE $rsm\_pts = U(min\ \_pts,\ \lceil 0.25 * design\_pts \rceil)$.

c. **Test** IF MCMC iteration, $i == 0$, THEN

i. **Set** the point with the best evaluated objective function as the MCMC algorithm's starting point,

ELSE

ii. **Set** the point with the best approximated objective function from the latest RSM run as the MCMC algorithm's starting point.

Where each RSM run is from $j = 1, \dots, run\_len$.

d. **Find** rsm_pts nearest evaluated points (Euclidian distance) to the current MCMC algorithm starting point from Step 4(c) and **fit** a selected RSM approximating function on those points. (The spline RBF approximation function was tested in this case study*)*

5. **Evaluate** MCMC approximate likelihood of current point using the locally fitted (RBF) function

a) **Test** IF $loglikelihood[i] >= loglikelihood[i-1]$, THEN

i. **Accept** current point, $curr\_pt[i]$, and its $loglikelihood[i]$ as the latest MCMC values,

ELSE

Test IF $U(0,1) >= \frac{loglikelihood[i]}{loglikelihood[i-1]}$, THEN

ii. **Accept** current point,$curr\_pt[i]$,and its $loglikelihood[i]$ as the latest MCMC values,

ELSE

iii. **Reject** current point, and its log likelihood and keep the previous MCMC values: $curr\_pt[i]$ = $curr\_pt[i-1]$ and

$loglikelihood[i] = loglikelihood[i-1]$ as the latest MCMC values,

b) **Update** MCMC iteration counter, $i = i + 1$; RSM update counter, $j = j + 1$.

6. **Test** IF MCMC iteration counter, $i \leq upchain$, THEN

a. **Go to** Step 4,

ELSE

b. **Test** IF MCMC iteration counter, $i < maxchain$, THEN

c. **Go to** Step 5

d. ELSE **STOP**.

The overall aim of the approach presented here is to randomly adapt the design points used in the fitting process such that only the most informative design points are used in the fitting process. By randomly selecting the number of points we avoid potential sampling issues such as fitting to localized optima. The 0.25, 0.5 and 0.75 rsm_pts sampling increments (to design points) were selected so that the process remained computationally efficient whilst still providing a minimum number of fitting points.

For this study, the $maxchain$, $upchain$ and $design\_pts$ were set to 60000, 20000 and 400 respectively. From experimenting with various scenarios and models, the length of the MCMC chain within which RSM updates occurs, $upchain$, may not be more than a third of the total chain, $maxchain$, to allow for MCMC convergence in the remaining chain without disturbances in response surface caused by re-fitting. For

algorithm stability under various scenarios, the $upchain$ and run_len were set so that there are only about a hundred RSM local function fittings in total.

Only results from the final section of the total MCMC chain, the final local approximation, were used for Bayesian inference. Using the same local fitting function for the last part of the MCMC chain allows the MCMC algorithm to converge to the final local optimum to produce statistically coherent Bayesian output.

The algorithm for global fitting differs from the random local fitting algorithm above, in that in Step 4(d), it proceeds by fitting all the *design_pts* at once. This means it then uses (without updating), the resulting approximate response surface function for the entire MCMC algorithm iterations. While the actual likelihood function could have been evaluated during Step 4, we deliberately avoided doing so in order to maintain the same number of actual function evaluations for the global as well as the random local fitting algorithm.

Experimental Setups

We used Haario et al's (2001) adaptive Metropolis algorithm (AM) for Bayesian calibration and uncertainty analysis. To enable fair comparison, the AM settings were kept similar for both global and random local fitting algorithms scenarios. A property of the AM algorithm is how the variance-covariance matrix of the proposal distribution, used for multivariate random normal sampling of the next values in the MCMC chain, improves with increasing chain length.

The starting variance-covariance matrix was defined as a diagonal matrix of proposed parameter variances (Haario, Saksman et al. 2001; Marshall, Nott et al. 2004; Smith and Marshall 2008). This is a simplification which assumes no correlation between

parameters. During the algorithm execution, accepted parameter and error values after an initial chain length, set at 5% of the stopping chain length, were used to generate the next variance-covariance matrix. This data-based second variance-covariance matrix now includes the correlation between parameters and error found from the initial accepted values. Subsequent variance-covariance values are updated adaptively using a history of accepted parameter values. A multivariate sampling scale was set so as to obtain acceptance rates between 25% and 45%, as recommended by Gelman et al (2004). Based on preliminary runs to convergence and on execution times, the stopping criterion was set at 100,000 MCMC iterations. The warm-up period, which is removed from analysis, is half the MCMC chain (Gelman, Carlin et al. 2004).

The Euclidean distance, a common measure of proximity performance (Lumijärvi, Laurikkala et al. 2004; Borkowski and Piepel 2009) was the main evaluation criteria used for this research. For all situations, the Bayesian Adaptive Metropolis (Haario, Saksman et al. 2001) algorithm was used as the optimization algorithm of choice assuming that all observed data were normally and independently distributed. To block variability, all the other factors such as random number seeds, starting values, processors input datasets and Bayesian Adaptive Metropolis algorithm settings, except scaling factors, were kept constant.

Results and Discussions

Optimum convergent model parameters were obtained using an adaptive Metropolis (Haario, Saksman et al. 2001) routine on the four actual (also known as original, exact) models from 5 different initial parameter sets each. Global and random local fitting schemes were used to approximate substitute surrogate models in subsequent

AM calibration experiments. The mean Euclidian distances between the optimum convergent parameters obtained using actual and surrogate models were taken as measures of performance of the two fitting schemes (global Vs random local) since all the other critical experimental settings were held constant. Shorter distances infer closer accuracy of AM calibration on a particular RSM approximation scheme than longer distances. There were no significant computational time differences between the two fitting scenarios for all the four models given that each approach uses the same number of design points.

Figure 25(a) shows that random local fitting consistently gave better averaged parameter calibration accuracy than global fitting in the extended region, for all the four models. This is reasonable since globally fitting a model over an extend domain produces poorer response surface approximation than approximating smaller regions at a time as done within the random local fitting scheme. That is, an extended sampling domain includes greater response surface complexity, which makes global fitting difficult, than smaller domains. Figure 25(b) shows that when the HPD region is known, possibly through prior optimization and actual model evaluations, using global fitting over this region performs slightly better on the FCOS, AWBM and FX2 models than random local fitting. This is because the HPD region likely has lower surface complexity and hence using more fitting points is better than using fewer neighborhood points that only partially interpolate the HPD at a time.

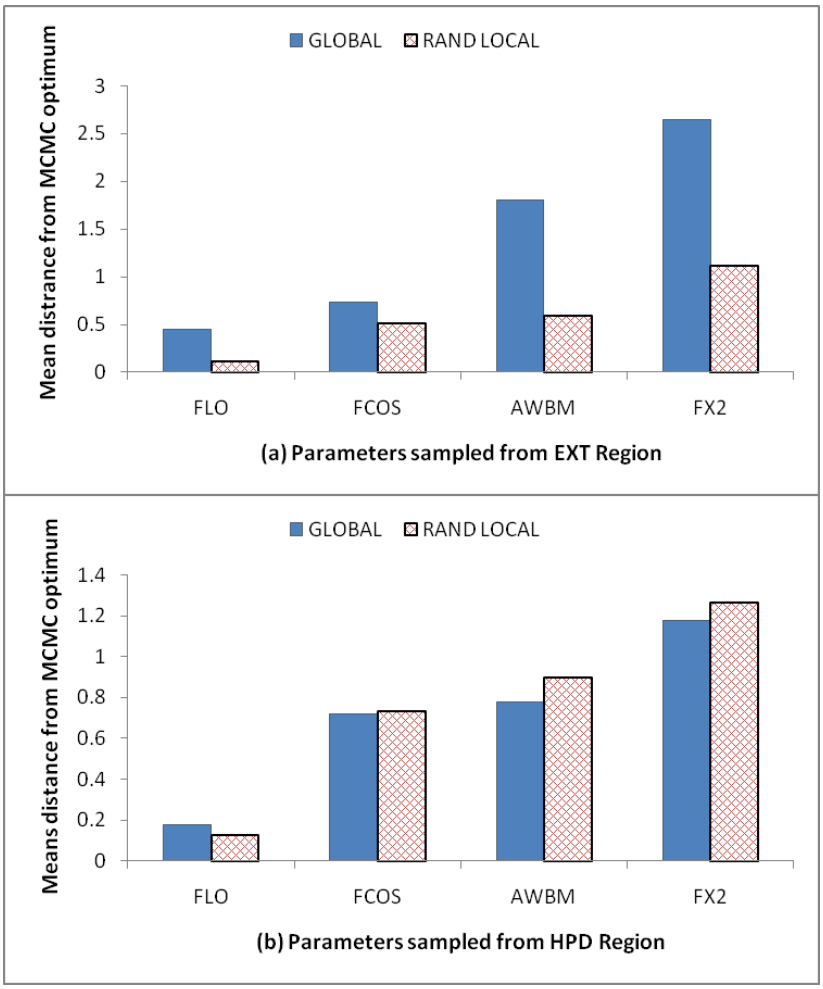

Figure 25: Global vs random local fitting

Figure 26(a) more clearly shows how the performance of global fitting deteriorates as the fitting region increases. Figure 26(b) shows that random local fitting performs about 30% better in the extended region than in the HPD region. This is potentially because in the HPD region, using smaller neighborhood points results in points that are numerically similar and hence in the poorer performance of the matrix inversion and factorization algorithms used. Even then, Figure 26(b) is better interpreted as showing good performance of random local fitting in the EXT region than poor performance in the HPD region. This is because the poorer HPD performance of the

random local fitting has already been shown to be only slightly worse than the best performance of global fitting in Figure 26(b).

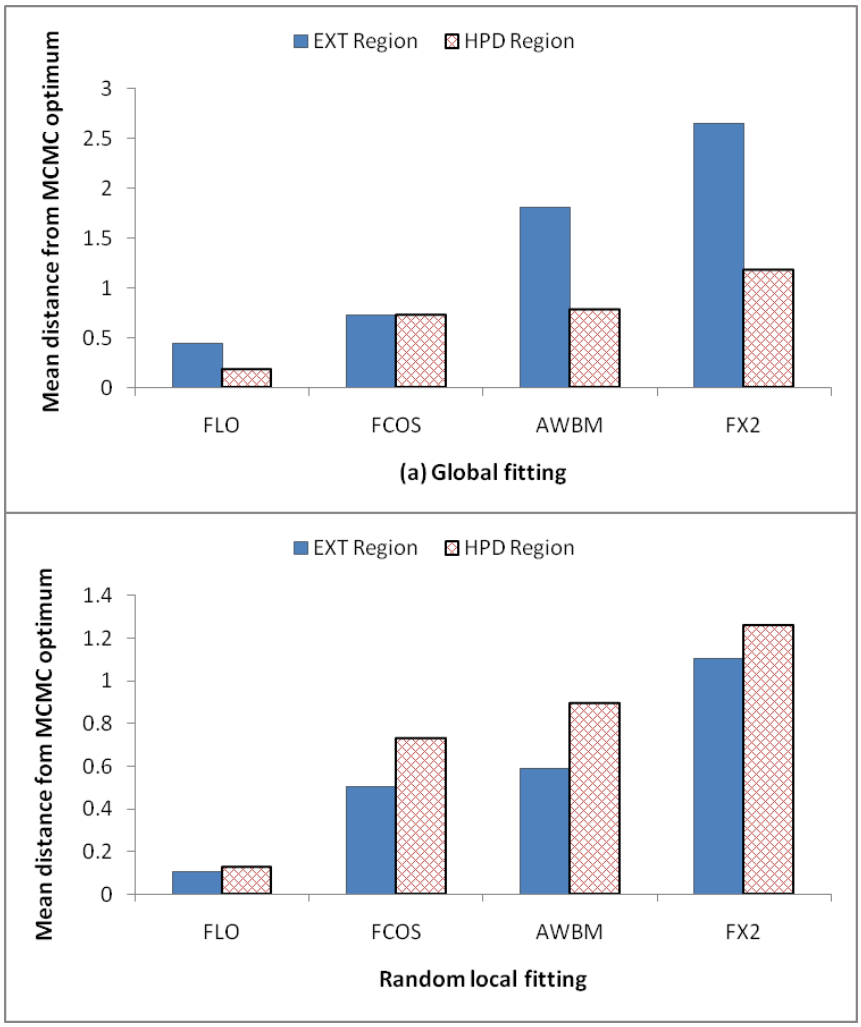

Figure 26: Extended vs high posterior density region

Figure 27 shows the drop in performance suffered by most of the global fitting for larger sampling domains compared to the other experimental scenarios. Conversely, random local fitting within the EXT region shows the best performance of all scenarios without excessive worsening of performance in the HPD region. Thus the average

stability of performance of random local fitting in the HPD and EXT region is better than that of global fitting in the same sampling domains.

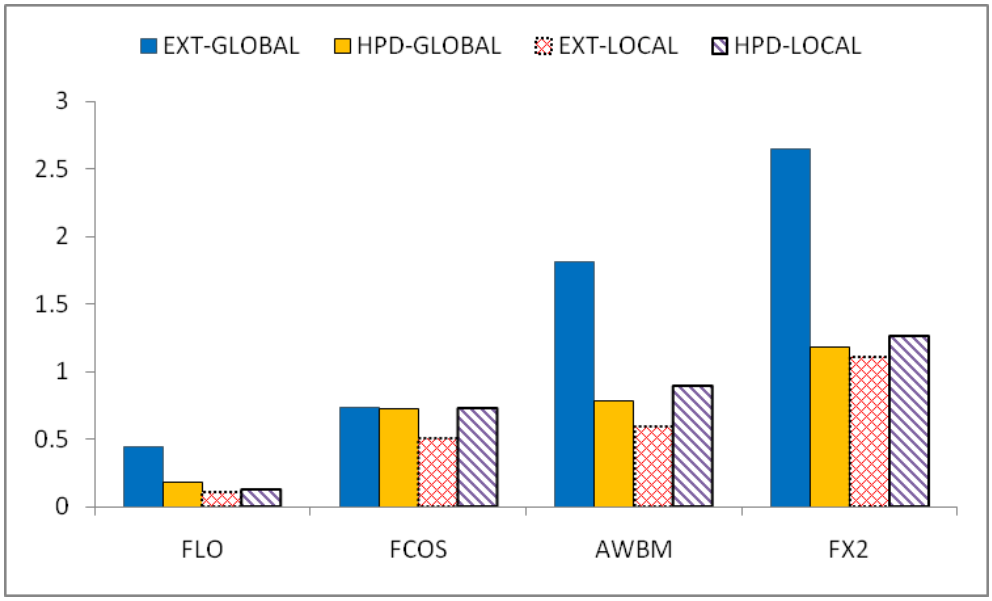

Figure 27: Overall Global Vs random local fitting performance

Tables 15 and 16 show representative results of means and standard deviations of parameter values for the AWBM and FLO models. The posterior distribution mean values are from the second half (30,000 iterations) of the MCMC chain. The mean parameter values were obtained by using the MCMC algorithm on the radial basis (400 actual model evaluations) are mostly within less than 20% of the mean values obtained by 60,000 MCMC actual model evaluations. However, the parameter posterior standard deviations from the fitted scenarios are much narrower than those from exact model evaluation. That is, using the Adaptive Metropolis on fitted scenarios resulted in narrow parameter posterior distributions.

Table 15: Sample convergent parameter means and (standard deviations) for the AWBM model

Scenario		Parameter		
		S	BFI	Ks
Exact AWBM model		210.0513 (1.5683)	0.4884 (0.0093)	0.9302 (0.0067)
Global fitted model	HPD	210.4774 (0.1795)	0.5810 (0.0321	0.9394 (0.0136)
	EXT	210.1229 (1.5785)	0.4880 (0.0094	0.9303 (0.0066)
Random local fitted model	HPD	209.9855 (0.0038)	0.4904 (0.0001)	0.9217 (0.0002)
	EXT	210.0684 (0.0152)	0.4964 (0.0001)	0.9295) (0.0006)

Table 16: Sample convergent parameter means and (standard deviations) for the FLO model

Scenario		Parameter				
		Alpha	Beta	Delta	Lambda	Min_srf
Exact AWBM model		0.8000 (0.0148)	0.3015 (0.0260	0.4954 (0.0272)	0.6007 (0.0208)	1.5153 (0.1866)
Global fitted model	HPD	0.7386 (0.0026)	0.2934 (0.0020)	0.4067 (0.0029)	0.7409 (0.0024)	1.4178 (0.0057)
	EXT	0.4997 (0.0030)	0.4436 (0.0014)	0.5144 (0.0030)	0.2556 (0.0030)	1.8285 (0.0039)
Random Local fitted model	HPD	0.7688 (0.0002)	0.3380 (0.0010)	0.4457 (0.0011)	0.6215 (0.0019)	1.4461 (0.0017)
	EXT	0.7523 (0.0022)	0.3289 (0.0031)	0.4582 (0.0020)	0.6812 (0.0049)	1.5480 (0.0074)

Improvements to the spread of the parameter distributions for fitted scenarios can potentially be obtained by improving the performance of the MCMC algorithm. For instance, using delayed rejection, DRAM (Haario, Laine et al. 2006), has been shown to

give improved robustness to the Adaptive Metropolis (Haario, Saksman et al. 2001) algorithm particularly in sampling the tails of the posterior distribution (Haario, Laine et al. 2006; Smith and Marshall 2008).

Conclusions

This paper studies the improved performance of local random fitting compared to global fitting over some experimental domain when using radial basis function approximation and Bayesian MCMC methods for calibration and uncertainty analysis. When the highest posterior density (HPD) region is known, fitting all the previously evaluated points (global fitting) to the radial basis function produced a 30% more accurate calibration results than local random fitting averaged over all the four test models. However, when the parameter sampling domain is much larger than the HPD region, fitting randomly selected local points performs about 144% better than global fitting. Overall, local random fitting performs about 60% better than global fitting, based on the Euclidian distance measure. We conclude that the global fitting scheme is highly sensitive to the size of the parameter sampling domain used. Larger parameter sampling domains result in poorer global fitting whereas random local fitting may even improve. This makes random local fitting ideal for situations where limited prior knowledge and model evaluations loosely locate a parameter sampling region much larger than the HPD. This could help in reducing the number of prior expensive model runs required and thus lower the total computational burden during calibration and uncertainty analysis. Although the global and random local fitting schemes were tested using only 400 previously evaluated points from SLHD sampling (for the four test models), the

comparison of the two fitting performances is expected to remain the same under some other more elaborate adaptive sampling schemes.

The results also suggest that the approach may underestimate the potential uncertainty of the true model but markedly narrow parameter posterior distributions. We theorize this may be due to underestimation of the RSM uncertainty (as RBF parameters are assumed exact) or due to the smoothing effect of the RSM around the parameters optima.A more robust DRAM (Haario, Laine et al. 2006) may replace the basic adaptive MCMC algorithm to improve the parameter distributions during uncertainty analysis on fitted models.

Contribution of Authors and Co-Authors

Manuscript in Chapter 5:
Bayesian Uncertainty Analysis of the Distributed Hydrology Soil-Vegetation Model Using Radial Basis Functions

Author: Able Mashamba
Contributions: Data collection and collation, experimental design, models development and assessment, results and manuscript write-up.

Co-author: Lucy Marshall
Contributions: Study supervision and oversight

Manuscript Information

Able Mashamba and Lucy Marshall

Journal of Environmental Modeling and Software

Status of manuscript (check one):

_✓_ Prepared for submission to a peer-reviewed journal

___ Officially submitted to a peer-reviewed journal

___ Accepted by a peer-reviewed journal

___ Published in a peer-reviewed journal

Publisher: Springer

CHAPTER 5

BAYESIAN UNCERTAINTY ANALYSIS OF THE DISTRIBUTED HYDROLOGY SOIL-VEGETATION MODEL USING RADIAL BASIS FUNCTIONS

Abstract: This manuscript demonstrates the feasibility of using radial basis approximating functions in Bayesian optimization and uncertainty analysis of the Distributed Hydrology Soil-Vegetation model (DHSVM). DHSVM is a physically based model that models the dynamics of hydrology, vegetation, soil and topographical interactions to predict sediment yields, water balance and channel discharge. The model still requires some parameter calibration that is typically performed by drawing from literature or expert knowledge of the case study sub-basin under study. In cases where such knowledge is missing, automatic calibration would be useful if it were not computationally expensive. In this case study we explore the use of surrogate models that avoid having to run the 'expensive' model during the many optimization iterations that may be necessary makes automatic calibration feasible. DHSVM is applied to the Tenderfoot Creek Experimental Forest (TCEF) in central Montana. Automatic Bayesian Markov chain Monte Carlo (MCMC) optimization and uncertainty assessment of 8 model parameters using a fitted spline radial basis function (RBF) requiring only 300 actual model runs improved the Nash Sutcliffe efficiency (NSE) on main channel flow from 0.83 after manual calibration to 0.89. We present the predictive and parameter uncertainty assessment for the case study DSHVM obtained using MCMC on the radial basis function.

Keywords: Radial basis function, Bayesian Markov chain Monte Carlo, uncertainty analysis, parameter calibration.

Introduction

Climate and land use change and their impact on precipitation patterns, land-phase hydrology, flooding, water scarcity and the resulting need to better manage water resources is attracting continued research in hydrologic modeling (Brennan 2010; Li, Zhong et al. 2010; Qureshi, Schwabe et al. 2010). Increased computing power now permits modelers to use more complex physically-based distributed modeling paradigms that have more realistic detail and descriptive power than the traditional lumped or basin-averaged approaches (Beven and Binley 1992; Beven 2000; Beven and Feyen 2002). These distributed hydrologic models describe the effect of spatial heterogeneity and locality of catchment processes to fully represent sub-basin hydrology. Although many distributed models (such as the Distributed Hydrology Soil-Vegetation Model (DHSVM) (Wigmosta, Vail et al. 1994; Wigmosta, Nijssen et al. 2002) and MikeSHE by the Danish Hydraulics Institute) are physically based, they often need some type of parameter calibration. It is often feasible for hydrologists to manually specify many of the physically based parameters, such as hydraulic conductivity or porosity, since these can be measured from the sub-basin or researched from literature; typically however most operational models require some form of model calibration.

An increasingly important aspect of modeling watershed processes is also recognizing the uncertainty associated with models and subsequent model outputs, and being able to assess the deficiencies in any model simulations. Addressing uncertainty can give model users a better understanding of the likelihood or risk associated with particular outcomes. In recent years, watershed modeling has evolved to take advantage of computational advances in Monte Carlo methods, Bayesian inference, and data

assimilation to better characterize the uncertainty associated with model predictions. The operational nature of many watershed models (e.g. for water resources management or prediction of the impacts of long term climate variability) has made this particularly important. The biggest hurdle facing widespread adoption of process based models and uncertainty methods is the computational burden associated with model simulations and analysis of model outputs. The large amounts of parameters that characterize distributed models often make complete manual calibration difficult. On the other hand, the automated optimization and uncertainty assessment of spatially distributed models is considered computationally expensive (Regis and Shoemaker 2007; Bliznyuk, Ruppert et al. 2008) . Potential solutions to this computational complexity problem in hydrologic models such as parallel processing and use of fast surrogate or meta-models are on-going research objectives.

Among distributed models, DHSVM has gained wide acceptance and use in climate change scenario analysis (Leung and Wigmosta 1999), forest management (Storck, Bowling et al. 1998) and land use and land cover studies (Thanapakpawina, Richeyb et al. 2007) as they relate to hydrologic processes. DSHVM has its roots in the hydrology-vegetation model (Wigmosta, Vail et al. 1994) that originally modeled canopy interception, evapotranspiration and snow accumulation and melt, as well as runoff generation via the saturation excess mechanisms. Model inputs include Geographic Information System (GIS) datasets describing watershed boundaries, meteorology, watershed topography, soils data, vegetation, stream and road networks and initial watershed states (e.g. snow water equivalent and soil water content or storage). The main outputs of DHSVM include channel flow, surface runoff, water balance, sediment yields,

evapotranspiration, soil storage, road and channel interception, canopy storage, snow water equivalent, and other snow and soil hydrology. Channel flow can be used to predict flooding or critical low-flows; sediment yields indicate extent of soil erosion and river siltation; higher canopy storage increases evapotranspiration; snow characteristics affect surface runoff and channel wetting and recession patterns; and watershed storage affects base-flow and vegetation. As with other fully distributed modeling paradigms, there have been few published routines for automatic calibration and predictive uncertainty analysis for the DHSVM model (Yao and Yang 2009).

One such study by Yao and Yang (2009) used an evolutionary algorithm to automate the calibration of 5 parameters of DHSVM for a 4623 km^2 case study watershed in China with daily observed data for an unspecified calibration period and unspecified spatial resolution. This approach reportedly took on average 61 hours to solve for optimum parameter values on unspecified computing hardware. However, reducing the time-step to hours and increasing the spatial resolution can dramatically increase the calibration time. For instance, we calculated it would take over 25 days (608.3 hr) to execute 1000 DHSVM calibration iterations using 1 year of observed data at 6-hourly time-steps for the 22.99 km^2 Tenderfoot Creek Experimental Forest (TCEF) sub-basin at 10 meter spatial resolution on an Intel Core 2 Quad with 2 x 2.83 GHz processor speed and 4MB RAM 64-bit desktop machine. Such increase in calibration time for reduced time-steps and increasing spatial resolution makes the use of ordinary automated calibration routines impractical. It is thus desirable to reduce the automated calibration time required for such computationally expensive models. Other than parallel processing, substituting expensive model runs with a faster fitted response surface model (RSM) is a

feasible way of reducing long automatic calibration times. A suitable measure of model error or performance is computed for a carefully selected and limited set of parameter values (design points) using the expensive actual model simulations. From the resulting parameter-residual set, RSM approximating functions, such as the radial basis, kriging or multiple linear regression models, can be developed to replace further expensive model evaluations. The performance of RSM fitting decreases as the number of parameters or dimensions increases. Thus it is important to perform model sensitivity analysis to eliminate insensitive parameters and thus minimize the total number of fitting dimensions.

To the best of our knowledge there is no published research that uses RSM fitting for automated calibration and uncertainty assessment of the DHSVM model or of other distributed hydrologic models in general. Thus the main objective of this research is to use radial basis function fitting for Bayesian calibration and uncertainty analysis of a case study DHSVM model. The purpose of this is to demonstrate that RSM fitting could be successfully used for automated calibration of DHSVM or other similar distributed models. While a Bayesian MCMC algorithm is used for uncertainty analysis, other uncertainty routines may be used. A secondary objective is to perform automated sensitivity analysis of the parameters affecting stream channel discharge. As mentioned above, sensitivity analysis is useful for pruning insensitive parameters so as to minimize the fitting dimensions in the RSM model.

The DHSVM Model and Parameter Selection

DHSVM development and application started in the 1990s (Wigmosta, Vail et al. 1994). DHSVM simulates the dynamics of hydrology, vegetation, soil and topographical interactions to predict sediment yields, water balance and storages, and channel discharge. DHSVM has been widely used for various hydrologic analyses, water resources management, forest management research (Wissmar, Raymond et al. 2004), land use studies (Thanapakpawina, Richeyb et al. 2007), flood forecasting (Gao, LÜ et al. 2003) and climate change scenario analyses (Wiley and Palmer 2008). DSVHM is written in C computer language and runs from a computer command line. The source code DHSVM is available for public distribution without a license.

DHSVM is a fully distributed model that is typically used to represent spatial scales of 10-200 meter grid cells and time steps of 1-24 hours. The model primarily uses saturation excess to model surface runoff, although a basic infiltration excess paradigm can also be used (Cuo, Lettenmaier et al. 2008). Darcy's law is used to model infiltration and unsaturated lateral flow (Yao and Yang 2009). DHSVM also includes a sophisticated snow accumulation and release model, and physically-based relationships for canopy interception, Penman-Monteith evapotranspiration and surface energy balance.

Figure 28 summarizes the DSHVM inputs and outputs. Main data types used for the model simulation include meteorological time series as forcing data, initial catchment state boundary conditions, and spatial data including those regarding soils, vegetation, stream networks and topography. These spatial data are typically prepared via geographic information system (GIS) software.

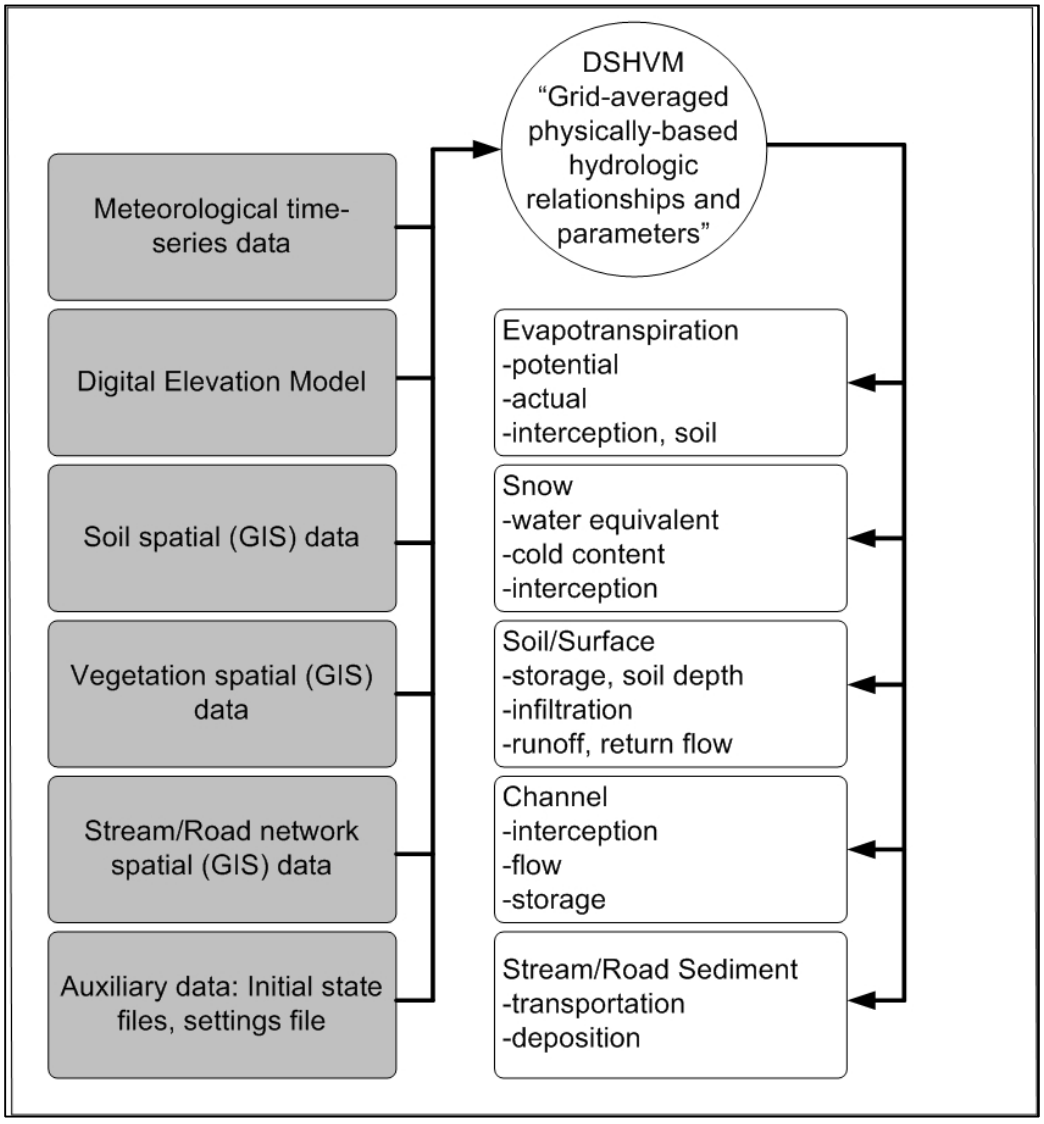

Figure 28: Summary of DHSVM inputs and outputs

Various DHSVM parameters are considered sensitive for model simulations depending on the purpose of the modeling exercise and the catchment characteristics. In terms of catchment-wide constants or characteristics, rain and snow leaf area index multipliers and snow water capacity could be considered important parameters to target for calibration when simulated discharge is the output of interest as they affect snow accumulation and evapotranspiration. For soil properties the model provides 16 default

soil types each with 14 parameters that can take different numerical values for each soil type and/or layer. Of these, lateral hydraulic conductivity, exponential decrease, porosity and field capacity were the 4 soil parameters we considered potentially sensitive to main channel flow. DHSVM vegetation is represented by 27 numerical parameters for each of the 20 model vegetation types. Of these vegetation parameters, we considered overstory leaf area index, minimum stomatal resistance, snow interception efficiency, moisture threshold, radiation attenuation and vapor pressure deficit as the 6 vegetation parameters potentially sensitive to discharge. Thus in general, DSHVM has about 13 relatively sensitive parameters that may be calibrated when channel discharge is the targeted simulation output.

Yao and Yang (2009) propose that the 5 most sensitive parameters of these 13 are lateral conductivity, exponential decrease, field capacity, porosity and minimum stomatal resistance. As with all distributed models, since parameters take on different values for different soil and vegetation types in the DHSVM model, the final number of distinct parameter values would be usually much higher than the original thirteen or Yao and Yang's (2009) five parameters.

Radial Basis Approximation

A radial basis functions (RBF) is one of the more popular approaches to approximating the response surfaces of difficult to evaluate or computationally expensive black box functions (Buhmann 2003; Regis and Shoemaker 2004; Mugunthan, Shoemaker et al. 2005; Regis and Shoemaker 2005; Regis and Shoemaker 2007; Bliznyuk, Ruppert et al. 2008). Other methods often applied may include multiple linear

regression (Regis and Shoemaker 2005), kriging (Davis and Ierapetritou 2009) and the more involved neural networks (Zorzettoa, Maciel Filho et al. 2001). Together these approximation or curve fitting approaches belong to the response surface methodology (RSM). Response surface methodology (RSM) is a collection of statistical and mathematical techniques useful for developing, improving, and optimizing processes through the exploration of relationships between several explanatory variables and one or more response variables (Myers and Montgomery 1995).

The radial basis interpolation function was extensively studied by Powell (1992, 1996, 2000) and Buhmann (2005), and used by Gutman (2001a), Regis and Shoemaker (2004, 2005, 2007), Bliznyuk et al (2008) in optimization and uncertainty analysis of computationally expensive black-box functions. In hydrology, the automated optimization and uncertainty assessment of distributed models is considered computationally intensive. Thus approximating response surfaces of model residual functions to changes in parameters would allow for faster parameter optimization and uncertainty assessment by limiting the number of actual expensive model simulations needed. To our knowledge no published research exists where response surface models have been used in the optimization and uncertainty analysis of the DHSVM.

A radial basis function (RBF), $S_n(x)$, is a real-valued function whose value depends only on the distance x from the origin or from some given point (called a center).Given n distinct points $x_1, \ldots, x_n \ \epsilon R^d$ whose function values $f\ (x_1), \ldots,$ $f\ (x_n)$, are known, the RBF then is an interpolant of the form as in Equation 5.1;

$$S_n(x) = \sum_{i=1}^{k} \omega_i \, \phi(\|x - x_i\|) + \ p(x), \ x \epsilon R^d \tag{5.1}$$

where $\|.\|$ is the Euclidean norm, $\omega_i \epsilon R$ for i = 1, …, n, $p \epsilon \Pi_m^d$ (a linear space of polynomials in d variables of degree less than or equal to m), and ϕ may take the forms $\phi(r) = r^2 \log r, r > 0$ and $\phi(0) = 0$ (thin plate spline), $\phi(r) = r^3$ (cubic), $\phi(r) = \sqrt{r^2 + \gamma^2}$ (multi-quadratic), or $\phi(r) = e^{-\gamma r^2}$, $r \geq 0$ and γ is a positive constant (Gaussian) (Regis and Shoemaker 2004). The ω_i is a weight showing how the Euclidean distance for point x_i from a sampled point x contributes to the approximation of the response at point x. The tail of the radial basis function, $p(x)$, is often linear with respect to the explanatory variables, but higher order polynomials could also be used. The thin plate spline and the quadratic functions have been reported to have good fitting performance for applications to complex models (Buhmann 2003; Regis and Shoemaker 2004; Regis and Shoemaker 2005; Bliznyuk, Ruppert et al. 2008). Further detail on how to obtain the ω_i and $p(x)$ radial basis coefficients from a system of linear algebra matrices can be obtained from recent RSM literature (Regis and Shoemaker 2004; Regis and Shoemaker 2005; Regis and Shoemaker 2007; Bliznyuk, Ruppert et al. 2008).

There are several implementation insights and recommendations for RSM methods that are worth highlighting. First, RSM methods require that the regressand (the model residual) be sensitive to the regressor variables (the distributed model parameters) for the response surface approximation to exist (Regis and Shoemaker 2007). This is because changes to parameters ought to affect the output for an approximate relationship between inputs and outputs to be established. Second, our prior research (Mashamba and Marshall 2010) shows that as the sampling ranges of regressor variable values increase, the accuracy of fitting or approximation rapidly deteriorates. Thus the regressor variables have to be sampled from minimum domains that contain the posterior distribution for the

target regressand. Finally, while space filling sampling designs such as symmetric Latin hypercube can be and have been used for regressor variable selection, adaptive parameter sampling schemes have been favored in practice (Regis and Shoemaker 2004; Regis and Shoemaker 2005; Bliznyuk, Ruppert et al. 2008). Adaptive sampling allows parameter values and hence the fitting points to be selected near the global optimum thus minimizing the number of actual model evaluations needed for smooth curve approximation around the optimum.

Bayesian MCMC and RSM

Bayesian Markov chain Monte Carlo uncertainty assessment is based on setting up a Markov chain (Markov 1906) on Bayes' conditional probability Equation 5.2 so that the chain converges to a limiting joint posterior distribution of the parameters under study (Gelman, Carlin et al. 2004).

$$P(\Theta|Y) = \frac{P(Y|\Theta).P(\Theta)}{P(Y)} \tag{5.2}$$

Where Y and Θ represent vectors or matrices of observed data and model parameters under study respectively. First, to use Equation 5.2 for setting up a Markov chain we assume a specific initial (prior) probability distribution of parameters, $P(\Theta)$. A continuous uniform distribution can be used as $P(\Theta)$ when no better joint prior parameters distribution can be inferred(Gelman, Carlin et al. 2004). Second, the probability distribution for observing data (Y) given an unknown joint probability distribution of parameters (the likelihood function) $P(Y|\Theta)$ has to be inferred. The inference of the likelihood function is based on the assumption about the variation of

error in observed data (Gelman, Carlin et al. 2004). Thus if one assumes that error is normally distributed, then the likelihood function ($P(\mathrm{Y}|\Theta)$) is as given in Equation 5.3 (Haario, Saksman et al. 2001; Marshall, Nott et al. 2004; Haario, Laine et al. 2006; Huard and Mailhot 2006; Marshall, Nott et al. 2007; Huard and Mailhot 2008; Smith and Marshall 2008);

$$P(Y \mid \Theta) = (2\pi\sigma^2)^{-n/2} \cdot \prod_{i=1}^{n} \exp\left\{-\frac{(y_i - \hat{y}_i)^2}{2\sigma^2}\right\} \tag{5.3}$$

where y_i and $\hat{y}_i$ are the observed (real system) and predicted (simulated) responses for each time-step $i = 1 \ldots n$ where n is the size of the observed dataset. The variance of the model error (residual), σ^2, is an extra parameter to be found together with the Θ parameter. A Markov chain can now be set on the joint posterior distribution of parameters ($P(\Theta|Y)$) as given in Equation 5.4;

$$\frac{P\left(\Theta^{j} \mid Y\right)}{P\left(\Theta^{j-1} \mid Y\right)} = \frac{P\left(\Theta^{j-1} \mid \Theta^{j}\right) \cdot P\left(Y \mid \Theta^{j}\right) \cdot P\left(\Theta^{j}\right)}{P\left(\Theta^{j} \mid \Theta^{j-1}\right) \cdot P\left(Y \mid \Theta^{j-1}\right) \cdot P\left(\Theta^{j-1}\right)} \tag{5.4}$$

where $P\left(\Theta^{j} \middle| \Theta^{j-1}\right)$ and $P\left(\Theta^{j-1} \middle| \Theta^{j}\right)$ represent joint proposal (perturbing or transition) distributions for the Markov chain parameters changing from previous $(j-1)$ values to current (j) values and vice versa respectively. The index $j = 1, \ldots, m$, where m is the maximum MCMC chain length after convergence. This Markov chain is ergodic and should eventually converge to a limiting posterior that is approximately equal to the unknown joint posterior distribution of the model parameters under study (Markov 1906; Gelman and Rubin 1992; Haario, Saksman et al. 2001; Gelman, Carlin et al. 2004). Once such a distribution is found, parameter and hence predictive uncertainty can be constructed. The ratio form of the Markov chain in Equation 5.4 is simply meant to avoid

having to compute the total marginal distribution of the observed data $P(Y)$, which is difficult or sometimes intractable (Gelman, Carlin et al. 2004).

The MCMC requires the evaluation of the posterior distribution for each step $j = 1, \dots, m$. But as Equation 5.3 shows, this requires simulating for the time-steps $i = 1, \dots, n$ for each j^{th} MCMC chain. If the simulation period is long (large n) or if the model computation is intense for each i then iterating through the MCMC chains can be computationally intensive. To reduce the computational burden, approximate response functions such as the radial basis function can be used to replace actual model simulation during MCMC method. For Bayesian MCMC, our objective function of interest is the likelihood function as given in Equation 5.3. Thus the RSM (such as the radial basis function) is used to approximate the likelihood function and thus quicken the traversal of the MCMC chain (Bliznyuk, Ruppert et al. 2008; Mashamba and Marshall 2010).

Methods

Case Study

The DHSVM model for this research was setup to simulate the discharge of the Tenderfoot Creek Experimental Forest (TCEF) watershed. The TCEF watershed encompasses 3693 hectares (9,125 acres) on the Lewis and Clark National Forest in Meagher Country, Montana. Further details on the watershed characteristics are available from the College of Forestry and Conservation 2007 Metadata Report from the University of Montana and from Mincemoyer and Birdsall (2006) and Jencso, McGlynn et al. (2009). The main channel of TCEF, the Tenderfoot Creek, is a headwater tributary of the Missouri River that drains the Little Belt Mountains of central Montana in a

westerly direction as shown in Figure 29. TCEF lies within the north, west, south and east bounding coordinates of 46.957, -110.907, 46.891 and -110.832 decimal degrees respectively. Slopes average 8^{o}, although steeper 25^{o} to 30^{o} hillsides occur along the main canyon and upper headwalls. Watershed elevation ranges from 1,991 m at the watershed outlet to 2,420 m at the highest point on the watershed boundary. The TCEF watershed is heavily forested with few access roads. Lodgepole pine (Pinus contorta) and mixed lodgepole pine with Engelmann spruce (Picea engelmannii) and subalpine fir (Abies lasiocarpa) stands are the dominant forest types and occupy about 95% (3,514 ha) of the experimental forest. Wet, irregularly shaped meadows are interspersed throughout the forest but only cover approximately 3.5% (125 ha), of the area. High, exposed hillsides with scree slopes and some dry grass communities make up the remaining 1.5% (54 ha) of the research area.

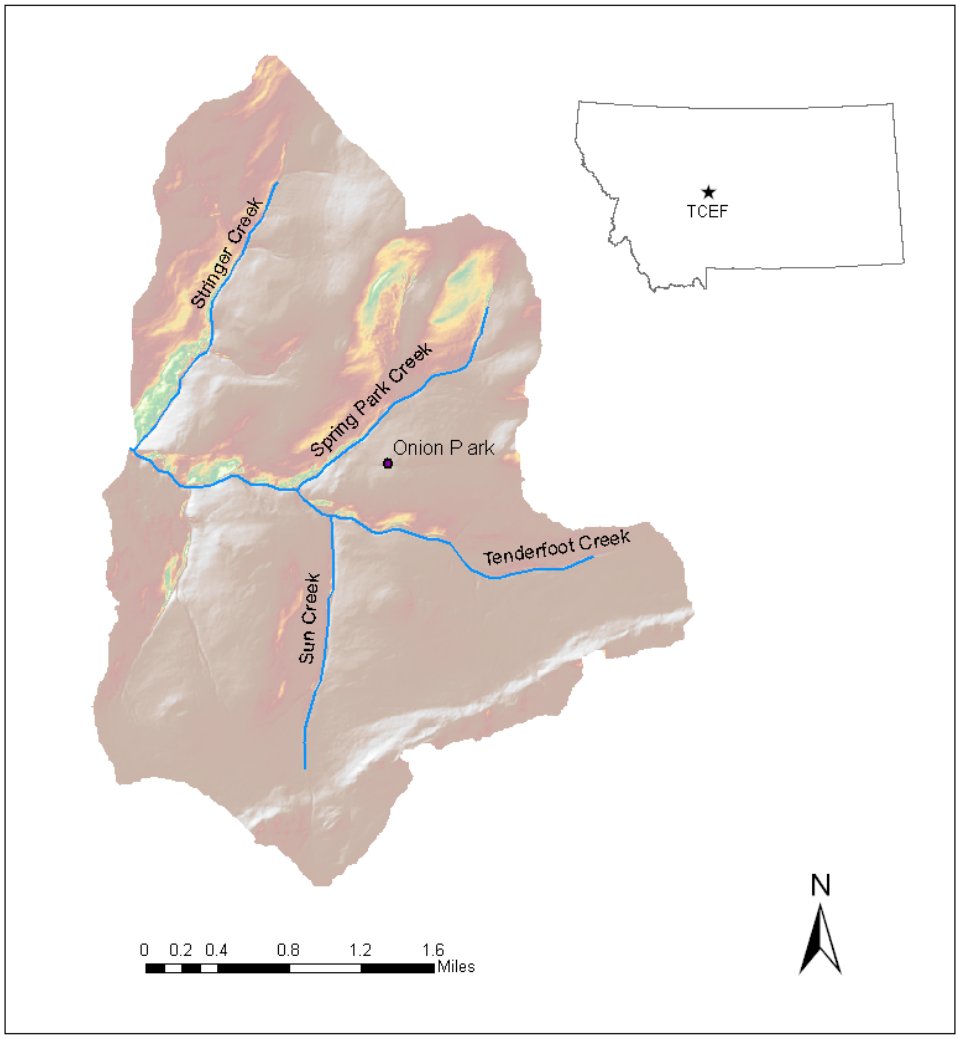

Figure 29: Tenderfoot Creek Experimental Forest

The most extensive soil groups are loamy skeletal, mixed Typic Cryochrepts and clayey, mixed Aquic Cryoboralfs. Rock talus slopes are prominent on the perimeter of the landscape, but rock outcrops are confined chiefly to areas adjacent to main stream channels. Soils in the grassland parks range from well to poorly drained and with depth ranging from 0.5 to 2 meters. Seeps and springs are common over the entire Experimental Forest. The TCEF watershed has a montane continental climate and the elevation-weighted mean annual precipitation of 800 mm. Almost 70% of the annual precipitation

falls as snow, which accumulates in the watershed between November and May. The annual runoff occurs in response to snowmelt between May and early June, creating a sharply peaked hydrograph. Base flow begins in mid July and persists through April. Summer rainfall is limited, generally occurring as brief, high intensity thunderstorms, and contributes little to the total runoff. TCEF is a suitable watershed for DHSVM discharge modeling because of the availability of climate and discharge data at high temporal and spatial resolution. Climate and discharge data is obtained from the Onion Park Snotel (1008) research site located within the watershed.

Random Local Fitting Algorithm

The radial basis function approximation used as the surrogate model for DHSVM in the Bayesian MCMC calibration and uncertainty analysis in this research depends heavily on a fitting approach called the random local fitting (RLF) algorithm (summarized in Figure 30). RLF improves the fitting performance of previously evaluated parameter sets (design points), compared to other local or global fitting methods. This is because RLF starts by global fitting, which helps in locating the region of global optimality, before eventually fitting to a local region, which improves the fitting. Fitting to a small region (local fitting) has been shown in previous research (Mashamba and Marshall 2010) to improve overall approximation by reducing the response surface complexity, compared to global fitting. Use of randomly selected sizes of the neighbor points helps with a fitting routine that can escape from local optima.

In the algorithm following, the words in italics represent RLF parameters, those in bold serve as key actions and those in brackets generally denote options that were used for this study. For instance, this study used symmetric Latin hypercube sampling for

selecting design points used in RSM fitting. Other designs such as adaptive sampling could be used together with the RLF algorithm below.

7. **Set**:
    - total number of MCMC iterations, $maxchain$,
    - number of MCMC iterations before updating each fitted model, run_len,
    - number of MCMC iterations after which updating of fitted models ceases, $upchain$,
    - number of design points to be sampled from the experimental domain, $design\_pts$,
    - number of model dimensions for optimization, d,
    - counters: MCMC iteration counter, $i = 0$; RSM update counter, $j = run\_len$,
    - minimum number of fitting points, $min\_pts = (d+1)(d+2)/2$,
    - parameters limits for the sampling domain, $(HPD\ or\ EXT)$.
    - logarithm of likelihood function at MCMC iteration i, $loglikelihood[i]$
    - current MCMC point (made from d elements) at ith iteration, $curr\_pt[i]$
8. **Sample** $design\_pts$ within experimental parameter sampling domain using space filling designs or adaptive sampling. (In this case study, symmetric Latin hypercube design (SLHD) sampling scheme was used)
9. **Evaluate** actual model MCMC likelihoods for each of the design points ($design\_pts$) in Step 2.
10. **Test** If MCMC iteration counter, $i \leq upchain$ AND RSM update counter, $j == run\_len$ THEN

a. **Reset** RSM update counter, $j = 0$.

b. **Reset** $rsm\_pts = U(min, max)$, (where rsm_pts is the number of design points used for RSM fitting) such that:

i. **Test** IF MCMC iteration, $i \leq upchain/4$, THEN

$rsm\_pts = U(\lceil 0.75 * design\_pts \rceil,\ design\_pts)$, ELSE

ii. **Test** IF MCMC iteration, $i \leq upchain/2$, THEN

$rsm\_pts = U(\lceil 0.5 * design\_pts \rceil,\ \lceil 0.75 * design\_pts \rceil)$, ELSE

iii. **Test** IF MCMC iteration, $i \leq 3 * upchain/4$, THEN

$rsm\_pts = U(\lceil 0.25 * design\_pts \rceil,\ \lceil 0.5 * design\_pts \rceil,$

iv. ELSE $rsm\_pts = U(min\ \_pts,\ \lceil 0.25 * design\_pts \rceil)$.

c. **Test** IF MCMC iteration, $i == 0$, THEN

i. **Set** the point with the best evaluated objective function as the MCMC algorithm's starting point,

ELSE

ii. **Set** the point with the best approximated objective function from the latest RSM run as the MCMC algorithm's starting point.

Where each RSM run is from $j = 1, \ldots, run\_len$.

d. **Find** rsm_pts nearest evaluated points (Euclidian distance) to the current MCMC algorithm starting point from Step 4(c) and **fit** a selected RSM approximating function on those points. (The spline RBF approximation function was tested in this case study*)*

11. **Evaluate** MCMC approximate likelihood of current point using the locally fitted (RBF) function

a) **Test** IF $loglikelihood[i] >= loglikelihood[i-1]$, THEN

i. **Accept** current point, $curr\_pt[i]$, and its $loglikelihood[i]$ as the latest MCMC values,

ELSE

Test IF $U(0,1) >= \frac{loglikelihood[i]}{loglikelihood[i-1]}$, THEN

ii. **Accept** current point,$curr\_pt[i]$,and its $loglikelihood[i]$ as the latest MCMC values,

ELSE

iii. **Reject** current point, and its log likelihood and keep the previous MCMC values: $curr\_pt[i]$ = $curr\_pt[i-1]$ and $loglikelihood[i]$ = $loglikelihood[i-1]$ as the latest MCMC values,

b) **Update** MCMC iteration counter, $i = i + 1$; RSM update counter, $j = j + 1$.

12. **Test** IF MCMC iteration counter, $i \leq upchain$, THEN

a. **Go to** Step 4,

ELSE

b. **Test** IF MCMC iteration counter, $i < maxchain$, THEN

c. **Go to** Step 5

d. ELSE **STOP**.

For this study, the $maxchain$, run_len, $upchain$, and $design\_pts$ were set to 20000, 100, 5000 and 300 respectively. This combination of values was chosen because of how it improved the performance of the entire algorithm (Figure 30). Other values

could be and perhaps should be tried for different models. Only the results from the second half of the total MCMC chain (last 10,000 iterations) were used for estimating the posterior distribution of parameters. Using the same local fitting function for the last part of the MCMC chain allows the MCMC algorithm to converge to the final local optimum to produce statistically coherent Bayesian output. The algorithm for global fitting differs from the random local fitting algorithm above, in that instead of Step 4, it proceeds by fitting all the $design\_pts$ at once. Thus global fitting always uses all the currently evaluated points for surrogate model approximation during the entire MCMC algorithm iterations. The use of a fixed number of neighboring design points (Regis and Shoemaker 2004) for radial basis function approximation is susceptible to being trapped on local optimum fitting if the sampling domain is large and complex enough, which is avoided by using randomly variable number of neighboring design points. Step 4 deliberately avoids exact evaluation of the current parameter set in order to minimize the number of computationally expensive actual function evaluations. Reducing the initial number of design points would allow Step 4 to add exact function evaluation of current approximately optimum parameter sets as a form of adaptive sampling.

Experimental Setup

DHSVM was setup for our case study watershed using climate and 10 meter grid Geographic Information System (GIS) data for the Lower Tenderfoot Creek watershed. A single six hourly time-step simulation needed 6 seconds per simulated day on an Intel Core 2 Quad with 2 x 2.83 GHz processor speed and 4MB RAM 64-bit desktop machine.

As previously noted, the DHSVM parameters that can be calibrated are related to catchment soil properties, vegetation properties, and climatic constants such as rainfall

and snowmelt leaf area index multipliers, and snow threshold and water capacity. Most of the approximated parameter values were set based on literature research and knowledge of the case study watershed. As DHSVM is actively used in hydrologic modeling research, thirteen unique parameters were identified as most sensitive to discharge (The DHSVM model and parameter selection). Note that depending on the number of soil and vegetation types identified in the model application the number of actual parameters to be optimized would increase. Of these thirteen parameters, five were specifically considered for automatic calibration by (Yao and Yang 2009). Thus, for this case study, lateral hydraulic conductivity, exponential decrease, field capacity, and minimum stomatal resistance were selected for automatic calibration while all other parameters were calibrated manually. These parameters generally are expected to control the characteristics of the wetting and recession periods of the hydrograph. As noted, these 4 parameters would have different values for different soil and vegetation types. Thus to minimize the total number of distributed parameters for automatic calibration only the two most significant soils and vegetation types were automatically adjusted for. Hence effectively, 8 distinct distributed parameters were used to automatically calibrate the DHSVM for discharge.

Table 17 shows the sampling domain of the parameters for our case study. The sampling domain is obtained from the literature and represents limited prior knowledge of the parameter value sets of the model or system. Of the 300 previously evaluated (design points), the random local fitting algorithm (RLF) was used to fit between 300 and 55 nearest neighbors during surrogate model approximation. A minimum of 55 was selected based on the tradition of using the minimum $(d + 1) * (d + 2)$ required to

characterize a second order polynomial, the simplest non-linear function where d equals 9; the number of model parameters used in the fitting function. The number of points used for fitting a surrogate model was limited to a maximum of 300 to prevent stack overflows.

Table 17: Parameter sampling domain

Parameter	LB	UB	Default
Lateral conductivity for loam soil, K_6	0.0001	0.2	0.01
Exponential decrease for loam soil, ExD_6	1.0	5.0	3.0
Field capacity for loam soil, FC_6	0.14	0.43	0.29
Lateral conductivity for silty clay soil, K_11	0.0001	0.20	0.01
Exponential decrease for silty clay soil, ExD_11	1.0	5.0	3.0
Field capacity for silty clay soil, FC_11	0.25	0.43	0.37
Minimum stomatal resistance for evergreen needle leaf, Rmin_1	200.0	1000.0	666.6
Minimum stomatal resistance for grassland, Rmin_10	200.0	1000.0	200.0

Figure 30 illustrates the automated Bayesian MCMC calibration and uncertainty assessment for the DHSVM model using radial basis function approximation. Initially the sampling scheme involves symmetric Latin hypercube sampling of 300 design points. The lower shaded region of Figure 30 shows how these 300 parameter design points are firstly evaluated (simulated) using DHSVM (Route 1). The RLF algorithm uses the exact likelihoods calculated from the DSHVM runs and fits random local points into the approximating radial basis function RSM(β) model. Route 2 represents subsequent updates to the RSM(β) model during MCMC chains by use of new neighborhood points (Step 4 and 5 in the algorithm described previously). When not updating the RSM(β) model, Route 3 is used to approximate the Bayesian likelihood (step 6(c) in the RLF algorithm). After the initial SLHD sampling, the subsequent sampling schemes based on

the multivariate random normal distribution of the MCMC. The top region of Figure 30 represents the Bayesian MCMC calibration and uncertainty assessment routine.

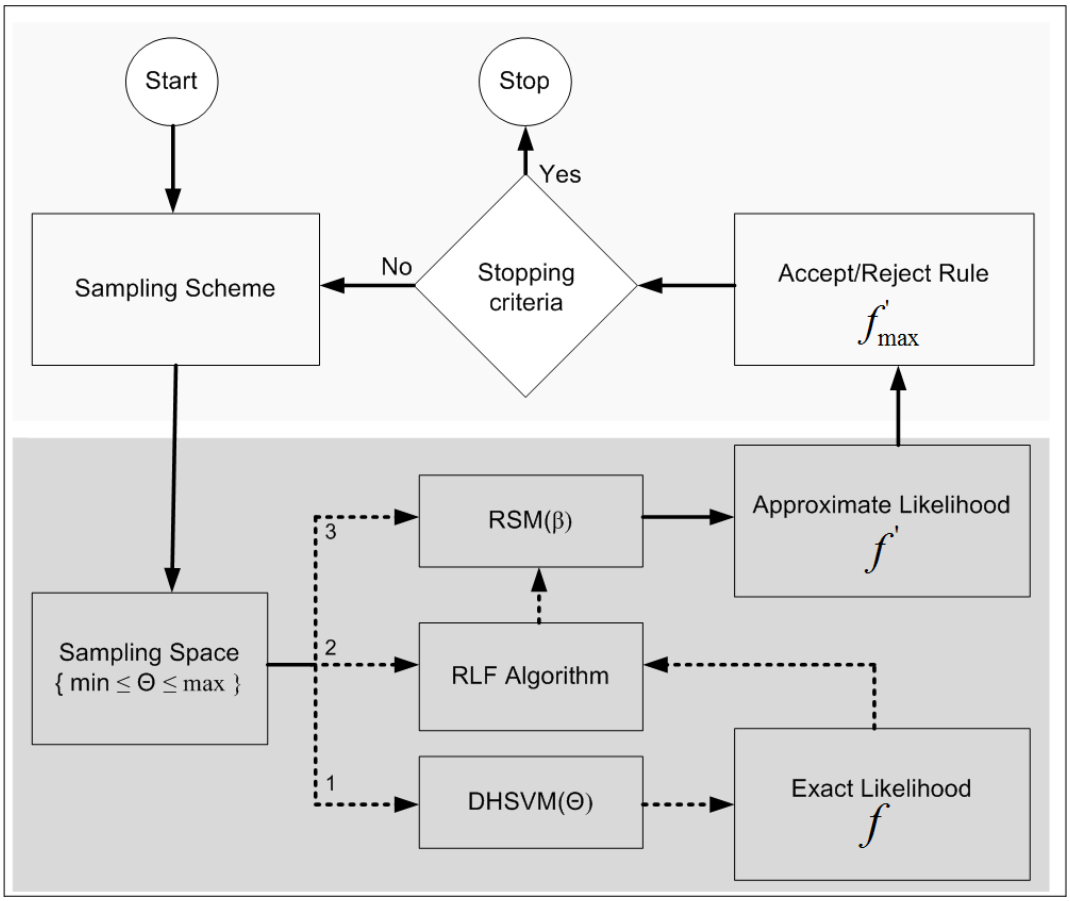

Figure 30: Schematic of algorithm interaction

Thus in general the automatic calibration and uncertainty analysis is setup by replacing the computationally expensive DHSVM evaluation with a faster spline-based radial basis surrogate model during the MCMC iterations. Except for the DHSVM model which is coded in C, the implementation schema shown in Figure 30 was coded in Intel FORTRAN 90. We used 20000 iterations of Haario et al's (2001) adaptive Metropolis algorithm (AM) for Bayesian MCMC calibration and uncertainty analysis on the surrogate radial basis function model made from 300 actual DHSVM model runs for the

calibration period March to September, 2008. A property of the AM algorithm is how the variance-covariance matrix of the proposal distribution, used for multivariate random normal sampling of the next values in the MCMC chain, improves with increasing chain length. The starting variance-covariance matrix was defined as a diagonal matrix of proposed parameter variances (Haario, Saksman et al. 2001; Marshall, Nott et al. 2004; Smith and Marshall 2008). This is a simplification which assumes no dependency between parameters. During execution, accepted parameter and error variance values were used to generate the next variance-covariance matrix after an initial MCMC chain length, set at 5% of the stopping chain length. This data-based second variance-covariance matrix now includes the correlation between parameters and error found from the initial accepted values. Subsequent variance-covariance values are updated adaptively using a history of accepted parameter values. A multivariate sampling scale was set so as to obtain acceptance rates between 25% and 45%, as recommended by Gelman et al (2004). Based on preliminary runs to convergence and on execution times, the stopping criterion was set at 20,000 MCMC iterations. The warm-up period, which is removed from analysis, is half the MCMC chain (Gelman, Carlin et al. 2004). We compared the model simulations from the point of highest posterior density to those estimated via manual calibration. We also consider the error of the fitting function and the uncertainty associated with the model simulations.

Results and Discussions

Due to the high (10 meter) spatial resolution used for the DHSVM model and the 6-hour time step, we limited the calibration period to 7 months (March – September

2008), so as to minimize the calibration time while capturing the spring runoff period and hydrologic dynamics of the entire year. A further two years, 2006 and 2007, were used for model assessment. The 300 DHSVM runs used for fitting the likelihood during MCMC calibration took about 4.46 days, running at 6 seconds per day. By comparison, 20000 fitted runs using the response surface model only take about 3 minutes on average. Typically, 30 meter grids are considered reasonable resolution for distributed modeling (VanShaar, Haddeland et al. 2002). Using 10 meter resolution for this research was however desirable in connection with further work using the same model and given the total catchment size.

Performance of the Radial Basis Function

Figure 31 shows how the (exact) logarithm of the Bayesian MCMC DHSVM likelihood compares to its (fitted) radial basis approximation. The linear scatter plot made from the 300 previously evaluated points shows how well the radial basis function is able to reproduce the log likelihood of the DHSVM model within the parameter sampling space.

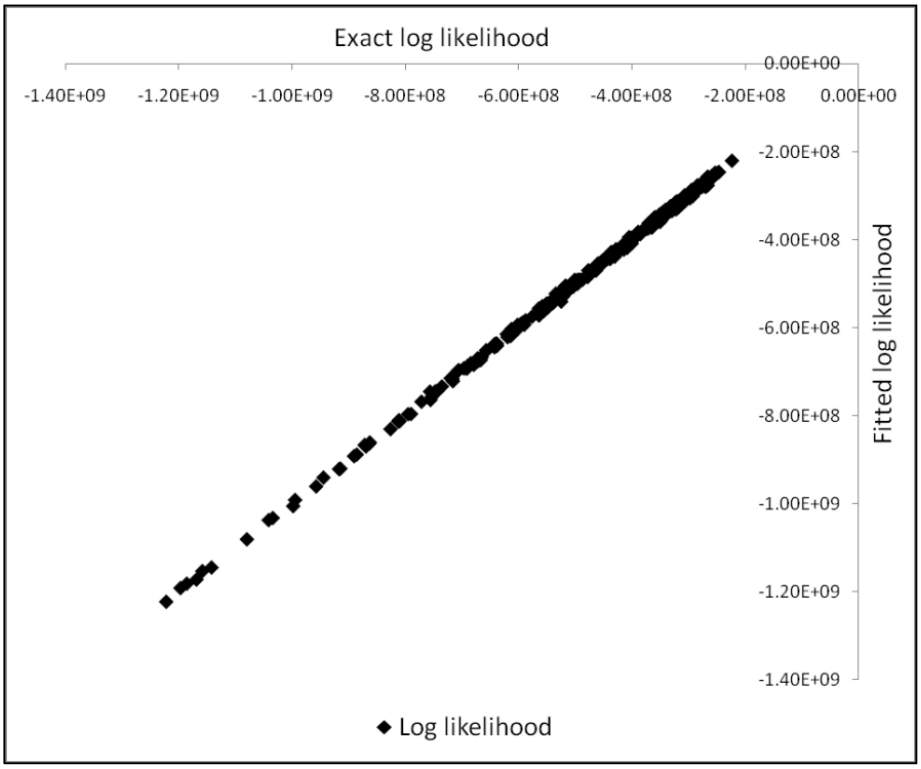

Figure 31: Exact versus fitted log likelihood

Figure 31 provides confidence in the approximation method and overall results of this case study. This is so because all the subsequent fittings were conducted within the domain from which the 300 design points were sampled. The large variation of the logarithmic likelihood in Figure 31 also shows that the parameters selected for the study were sensitive and thus useful for use in Bayesian MCMC analysis.

Bayesian Optimization Performance

Figure 32 shows three hydrographs of Tenderfoot Creek during the analysis period, April – September 2008. The observed, simulated (manual calibration) and simulated (automated calibration) hydrographs show relatively very low base flow followed by snow-melt driven peak flow from early May to early July. Automated

Bayesian MCMC calibration on a fitted radial basis response function improved the simulated discharge from a Nash Sutcliffe Efficiency (NSE) (Nash and Sutcliffe 1970) of 0.83, after manual calibration, to 0.89. The high prior knowledge of the TCEF watershed coupled to the uncomplicated features of the hydrographs enabled a manual calibration with an NSE of 0.83, which could be considered satisfactory (Eckhardt, Fohrer et al. 2005). Nonetheless, the optimized simulation shows improved recession characteristics (with a faster falling recession period) and a slower wetting up period.

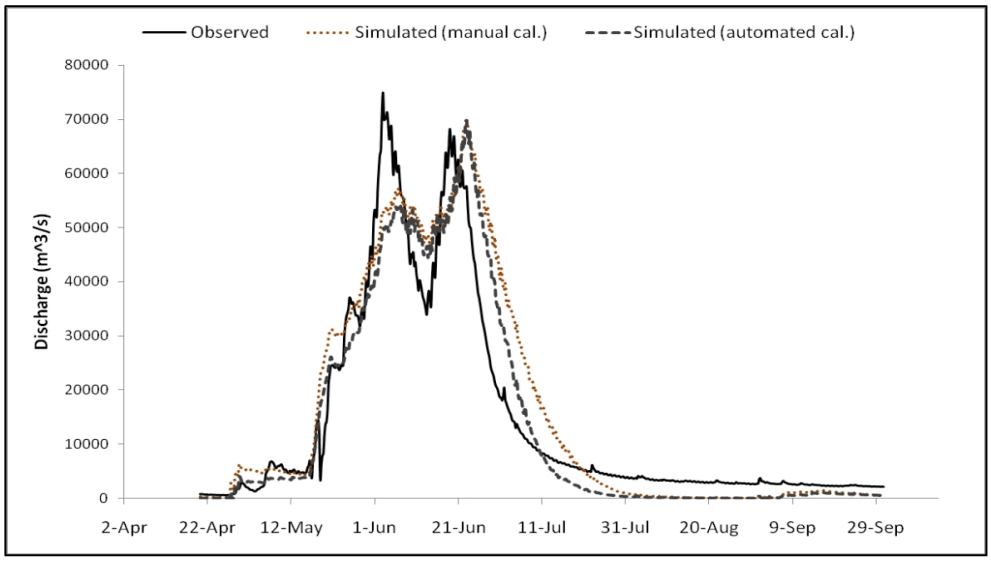

Figure 32: Observed Vs simulated discharge for Tenderfoot Creek for the year 2008

Figure 33 shows the observed, simulated (manual calibration) and simulated (automated calibration) hydrographs for the Tenderfoot Creek from March to September of 2006 (6a) and 2007(6b) validation periods. The 3 peak flows correspond to the May-June snowmelt periods for the two years.

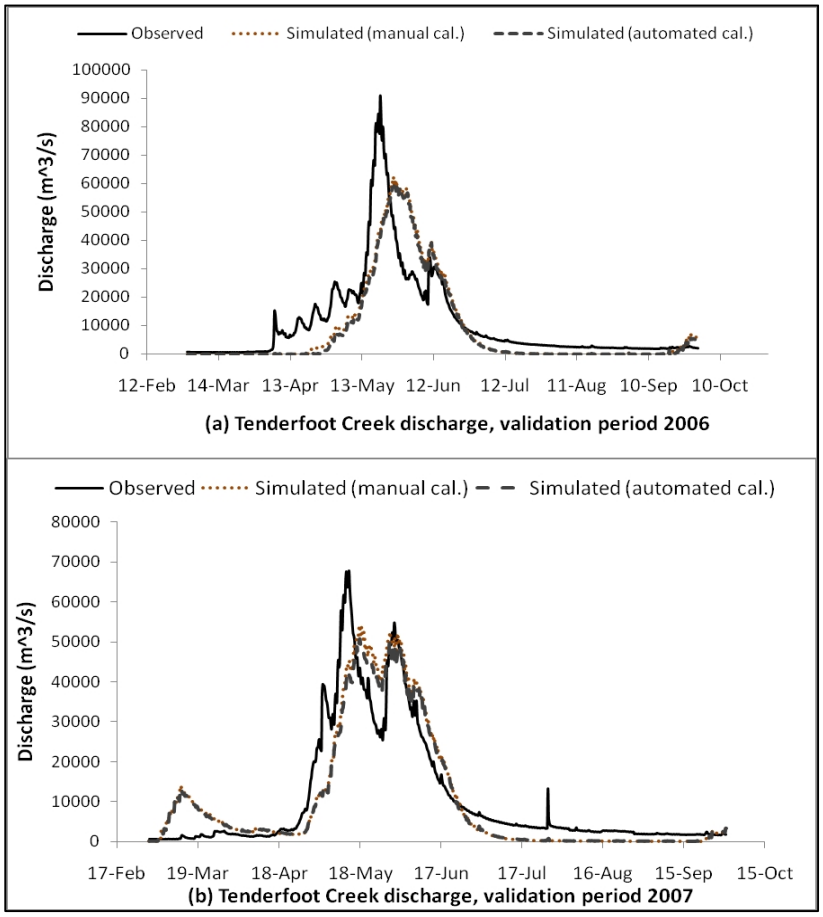

Figure 33: Observed Vs simulated discharge for Tenderfoot Creek for the validation period 2006-2007

The NSE numbers for the simulated discharge after manual and automated calibration are 0.796 and 0.765 respectively. This supports the view that automated calibration slightly over-fitted the model parameters to the calibration period in Figure 32, and thus performing slightly poorer over other periods. Over-fitting of model parameters to a calibration period can occur regardless of the calibration method. The model assessment supports the view that the model calibration and robustness would be improved by a multi-year calibration.

Bayesian Uncertainty Analysis

Table 18 shows the optimization and uncertainty assessment results summary of the 8 sensitive parameters that were studied. All the other DHSVM parameters were manually calibrated from existing knowledge of the TCEF watershed. Table 18 results were compiled from the converged last half (10000 iterations) of the MCMC chain. The standard deviations and 95% confidence intervals (CI) represent the posterior distribution of the 'Bayesian MCMC on fitted radial basis function' scheme. The parameter posterior distributions are relatively narrow, indicating the likely limited uncertainty and high sensitivity of he selected parameters.

Table 18: Uncertainty and calibration summary

Parameter	95% CI		Std dev	Mean	Optimal
	Min				
Lateral conductivity for loam soil, K_6	0.141578	0.141579	3.55E-06	0.141579	0.141582
Exponential decrease for loam soil, ExD_6	1.046850	1.046854	9.94E-05	1.046852	1.046937
Field capacity for loam soil, FC_6	0.295196	0.295201	1.21E-04	0.295199	0.295306
Lateral conductivity for silty clay soil, K_11	0.019528	0.019529	1.85E-05	0.019528	0.019512
Exponential decrease for silty clay soil, ExD_11	1.888893	1.888934	1.04E-03	1.888914	1.888701
Field capacity for silty clay soil, FC_11	0.413995	0.414003	2.01E-04	0.413999	0.413700
Minimum stomatal resistance for evergreen needle leaf, Rmin_1	379.98574	380.02519	1.01E+00	380.00546	381.63633
Minimum stomatal resistance for grassland, Rmin_10	236.05224	236.07605	6.07E-01	236.06415	236.28526

Figure 34 shows 6 soil parameter traces (parameter values sampled via the MCMC algorithm), for the converged last half of the MCMC chain. K, ExD and FC refer to the lateral conductivity, exponential decrease and field capacity respectively. The

numbers 6 and 11 which refer to different soil types within the configuration file of DHSVM: loam and silty clay soils. Note that the parameter samples appear to have converged to the true posterior distribution for the catchment data. Although not included to minimize clutter, the vegetation parameters share the same pattern as displayed for the soil parameters.

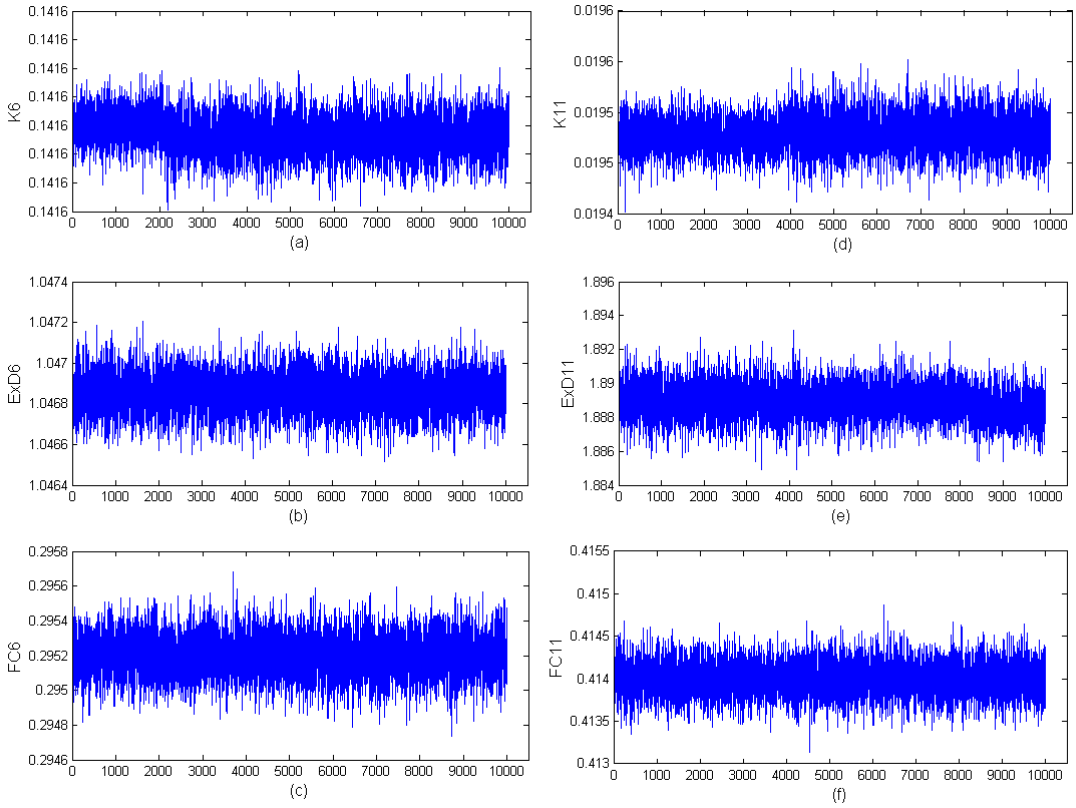

Figure 34: Soil parameter traces

Figure 35 shows the corresponding posterior distributions of the 6 soil parameters. The parameters are uni-modal and appear symmetric.

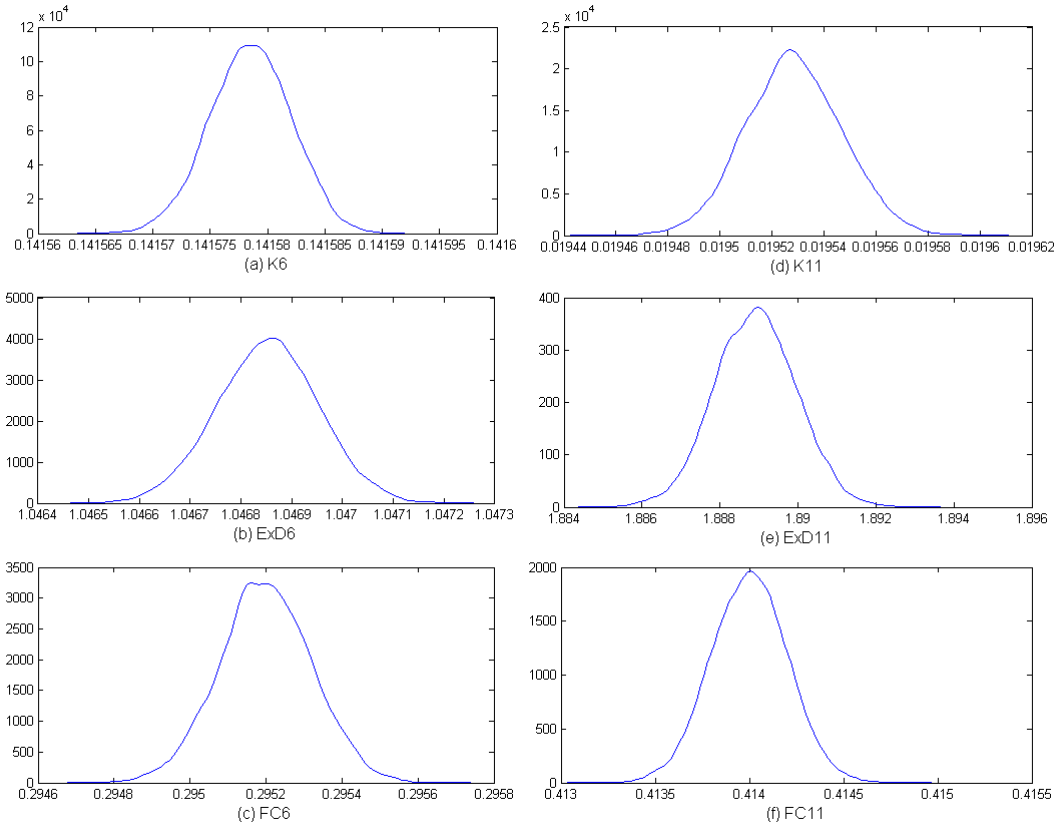

Figure 35: Soil parameter probability distributions

In previous studies (Mashamba and Marshall 2010) where fitting and exact analysis could be compared, RLF fitting tended to produce narrower distributions than direct Bayesian optimization and uncertainty assessment. This is often due to the reduction in the fitting region as a result of using neighboring points only. Global fitting sometimes help in widening the parameter probability distributions by widening the fitting region.

Figure 36 shows the predictive uncertainty for the flow of Tenderfoot Creek during the melt period of 2008. The lower and upper bounds are created by parameter sets with the best and worst model residual of the 300 points.

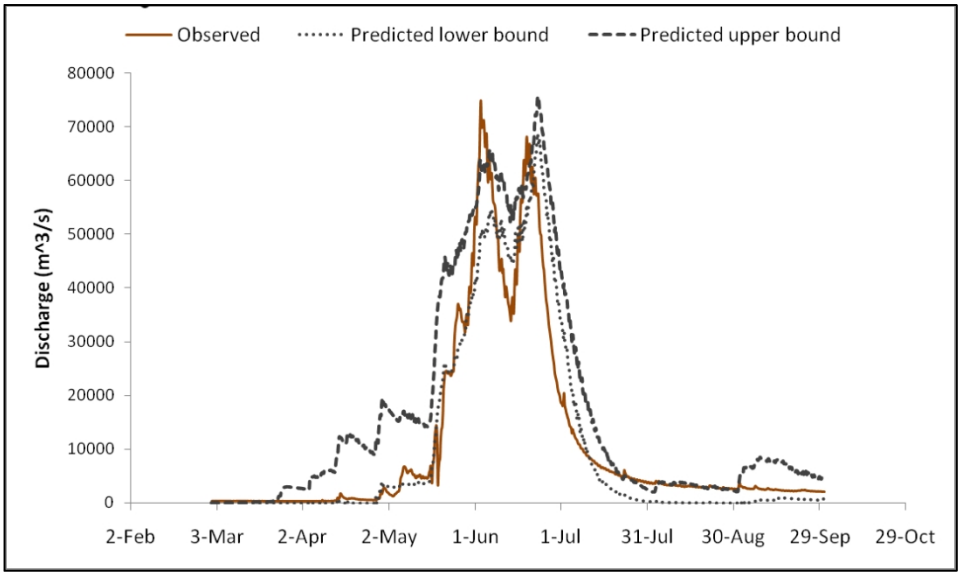

Figure 36: Predictive uncertainty for 2008 Tenderfoot Creek discharge

It is clear from Figure 36 that within the global experimental domain, the model tends to over predict the wetting period (rising limb) and under predict the recession (falling limb) of the Tenderfoot Creek discharge. The bigger size of the range between lower and upper bounds (Figure 36) means that the parameters chosen for this study affect wetting more than recession in the case study.

Conclusion

The main goal of this case study was to demonstrate the feasibility of Bayesian Markov chain Monte Carlo calibration and analysis of uncertainty for the DHSVM, a fully distributed hydrologic model. MCMC methods typically require thousands of actual model evaluations, which often make their use with computationally expensive spatially distributed hydrologic models infeasible. To reduce high computation, a spline radial basis approximation function replaced actual model evaluations during MCMC

uncertainty analysis. The improvement of the model simulation after automatic calibration demonstrates a strong potential for automated Bayesian optimization of the DHSVM model using radial basis surrogate models. To compensate for a limited prior knowledge of the highest posterior density, the random local fitting algorithm was successfully used to improve the quality of fit and thus hasten the MCMC search.

The automated calibration tested in this research used basic space-filling sampling in an extended parameter domain. Only 300 actual model evaluations were needed, which took 4.64 days, for use in building surrogate models during the rest of the 20000 approximations, which took 3 minutes on average. Further reductions of actual model evaluations may be possible by using adaptive sampling instead of relying on space filling designs alone.

While the random local fitting algorithm improves response surface fitting in extended regions, its tendency to reduce parameter uncertainty distributions may warrant the use of global fitting within reduced sampling domains. Thus RLF may be used to locate the region of optimality and a fixed global fitting scheme may be used to create wider uncertainty distributions. For this research, implementing the fixed global fitting algorithm to a limited domain did not widen the parameter distribution. Finally, radial basis fitting and Adaptive Metropolis MCMC algorithms seem to demand significant computer memory as the number of parameters and fitted points increases. Thus it is always necessary to minimize the number of automatically calibrated parameters even with the use of surrogate models.

CHAPTER 6

CONCLUSION

Physically based and/or spatially distributed hydrologic models are increasingly being used in climate change scenario analysis, flood forecasting, studies of water scarcity, pollution and riparian ecological studies and in agricultural, settlement and civil works planning and design. Spatial distribution in modeling allows for the identification of localized critical hydrological processes such as flooding or water pollution in a large watershed, while still allowing for basin-averaged output. Physical process based modeling improves visualization and hence communication, conceptualization and creditability of the model output, all of which are critical to decision makers.

However, the ongoing concerns of distributed models are their large numbers of parameters and input data requirements, increased structural complexity and high computational demand when compared to basin-averaged models. Large input data requirements limit where distributed hydrologic models can be used since fewer watersheds have complete measured datasets required for modeling. Where prior knowledge and expertise exist about a watershed, distributed models benefit from manual optimization.

On the other hand limited knowledge about a watershed or model structure forces modelers to resort to automated optimization and uncertainty assessment. Optimization involves adjusting model parameters manually or by automated search routines so that the difference between simulation output and available observed data (sum of squared residual) is minimized. Uncertainty analysis is about quantifying the range of values that

model parameters and output can acceptably take and still be considered to be correctly representing the real system in structure and output. Uncertainty assessment is important because it is equivalent to finding a statistical measure of risk associated with making decisions based on the output of a model. While model optimization seeks to find the ideal set of parameter values that best fit the real system response, uncertainty analysis establishes an ensemble of acceptable parameter values describing the real system response.

Among hydrologic modelers, there is increasing consensus that uncertainty analysis is more important than calibration. This is because instead of a single optimum set, there often exists a significant number of parameter combinations that give near optimum model simulations. Modelers are thus interested in identifying these parameter sets as a better way of model analysis than identifying only the best set. Among the different ways of finding near optimum parameter sets that make uncertainty distributions and intervals, Bayesian inference is arguably be the most statistically valid. As such it is the approach for both uncertainty assessment and calibration that was used in this study.

Large numbers of parameters and increased simulation time associated with distributed models make automated calibration and uncertainty assessment time consuming or even infeasible. There are two ways to reduce computation time during automated calibration and uncertainty analysis. One approach, a technological or hardware solution, involves parallel processing of the optimization and uncertainty analysis routines. Computer networks and multi-core desktop computing favors this brute-force solution. A second method features the conceptual or software solution involving the use of surrogate models (models of models) that simulate or evaluate faster

than the distributed model during optimization and uncertainty assessment. The work presented in the chapters of this dissertation described conceptualization, experimentation, and implementation of the response surface methodology to automated optimization and uncertainty assessment of distributed hydrologic models.

The first manuscript documented work on scenario analysis of a case study irrigation watershed (the Buffalo Rapids) by use of a semi-distributed Soil and Water Assessment Tool (SWAT) model. Native automated calibration routines were used to condition the SWAT model to available observed data. In undertaking this model assessment the problem of long computation times during calibration of such complex models became apparent. Although this study is concerned with best management scenario analysis, long calibration times set the foundation for further work in the subsequent manuscripts. The scenario analysis conducted for Buffalo Rapids used one factor at a time experimental designs to link specific model parameters to best watershed management practices. More robust factorial designs were infeasible due to the different levels of parameters (factors) that had to be considered. Further, the categorical nature of several model parameters (such as irrigation delivery systems) in addition to numerical parameters (such as irrigation water amounts per delivery) complicated typical approaches to scenario analysis. A novel idea introduced in this scenario analysis involved investigating how parameter changes might translate to best management practices. This is a departure from the existing scenario analysis approach which focuses on best management practices and how they might change the model parameter values.

The second manuscript is an investigation and review of the use of response surface methodology for automated Bayesian Markov chain Monte Carlo (MCMC)

calibration and uncertainty assessment of models. Three synthetic models and an Australian Water Balance model were used in the study of factors that affect the performance of Haario's (2001) Adaptive Metropolis calibration and uncertainty assessment of models using the response surface methodology. In the study, we investigated how (i) the approximation function used and the type of model or response function approximated, (ii) the number of points used in the actual curve fitting, (iii) the size of the region from which fitting points are sampled, and (iv) the size of normal observed data error affect the performance of RSM approximation for Bayesian MCMC calibration and uncertainty analysis. The main findings of the study were that the performance of response surface methodology and MCMC was negatively correlated to the size of sampling domains of the parameters and error variance used. Reducing the parameter and error variance sampling ranges improved the RSM fitting and hence Bayesian optimization results. Fitting fewer points on a smaller range gave better results than fitting many more points on a larger region. The radial basis function RSM method performed better than the second order multiple linear regression fit. Generally, increasing the number of points used in RSM fitting improves the accuracy of fit and hence parameter optimization. However, it is likely that beyond some value, more RSM points result in diminishing improvements or may even begin to negatively affect the performance of approximation. Based on these findings, for the models considered in this study RSM approximation for Bayesian optimization performed best when the fitted to the high posterior density (HPD) region, with design points that are not too many or too few, when data error is not excessive and when using the thin plate spline radial basis function as the approximating function over the second order function. Overall, it was

concluded that the number of design points is less critical than where they are sampled. Sampling from the HPD results in better performance of RSM fitting for Bayesian optimization than sampling from an extended region, as might happen when there is poor prior knowledge of the HPD region before fitting.

The third manuscript is based on findings from the second. One of the main findings in the second manuscript was how the size of the parameter sampling domain affects the performance of RMS fitting and MCMC uncertainty assessment. The best RSM fitting and MCMC uncertainty assessment performance is when parameters are sampled from a limited region of high posterior density (HPD). However the parameter HPD region is established mostly on the basis of expertise and prior automated search. Prior automated search to establish an HPD region from which RSM fitting values can be sampled increases the computational burden since the search is conducted using the actual computationally expensive model. Existing RSM fitting paradigms for distributed model calibration and uncertainty analysis and for the optimization of other computationally expensive black-box functions assume prior knowledge of the HPD region.

To minimize the need for prior knowledge of the HPD region, we developed an automated RSM fitting procedure that performs well in extended parameter sampling regions. The new approach is called the random local fitting (RLF) algorithm. RLF uses an adaptive and randomly decreasing number of neighboring design points (previously evaluated parameter sets) to fit a response surface (the radial basis function) of the likelihood function of the MCMC algorithm that is then used during calibration and uncertainty analysis. When the highest posterior density (HPD) region is known, fitting

all the previously evaluated points (global fitting) to the radial basis function produced a 30% more accurate calibration results than local random fitting averaged over all the four test models. However, when the parameter sampling domain is much larger than the HPD region, fitting randomly selected local points performs about 144% better than global fitting. Overall, local random fitting performs about 60% better than global fitting, based on the Euclidian distance measure.

We conclude that the global fitting scheme is highly sensitive to the size of the parameter sampling domain used. Larger parameter sampling domains result in poorer global fitting. However, random local fitting in the larger sampling domain resulted in comparable performance to sampling from the HPD region. This makes random local fitting ideal for situations where limited prior knowledge and model evaluations loosely locate a parameter sampling region much larger than the HPD. This could help in reducing the number of prior expensive model runs required and thus lower the total computational burden during calibration and uncertainty analysis. Although the global and random local fitting schemes were tested using only 400 previously evaluated points from SLHD sampling (for the four test models), the comparison of the two fitting performances should be the same under some other more elaborate adaptive sampling schemes.

Finally, the fourth manuscript is a feasibility study of the implementation of MCMC calibration and uncertainty assessment using radial basis random local fitting (RLF) for a Distributed Hydrology Soil Vegetation model (DSHVM) of the Tenderfoot Creek Experimental Forest case study. Only 300 design points were used to approximate 20,000 simulations during the analysis. Calibration and predictive and parameter

uncertainty analysis results from the study showed improved model simulations even though no prior HPD region from which to sample design points was assumed. Overall, the work in these manuscripts provides concepts, algorithms, experiments and a demonstration of algorithm solutions to the problem of long computational times during the Bayesian optimization and uncertainty analysis of spatially distributed models.

REFERENCES CITED

Andr´eassian, V., J. Lerat, et al. (2007). "What is really undermining hydrologic science today?" Hydrological Processes 21(20): 2819-2822.

Arabi, M., J. R. Frankenberger, et al. (2007). "Representation of Agricultural Conservation Practices with SWAT." Hydrological Processes 22(16): 3042–3055.

Arabi, M., R. S. Govindaraju, et al. (2007). "A probabilistic approach for analysis of uncertainty in the evaluation of watershed management practices." Journal of Hydrology 333(2-4): 459-471.

Arnold, J. G. and N. Fohrer (2005). "SWAT2000: current capabilities and research opportunities in applied watershed modelling." Hydrological Processes 19(3): 563-572.

Arnold, J. G., R. Srinivasan, et al. (1998). "Large area hydrologic modelling and assessment part I: model development." Journal of American Water Resources Association 34(1): 73–89.

Aronica, G., P. D. Bates, et al. (2002). "Assessing the uncertainty in distributed model predictions using observed binary pattern information within GLUE." Hydrological Processes 16(10): 2001-2016.

Balakrishnan, S., A. Roy, et al. (2003). "Uncertainty reduction and characterization for complex environmental fate and transport models: An empirical Bayesian framework incorporating the stochastic response surface method." Water Resources Research 39(12): SBH 8-1 - SBH 8-13.

Bates, B. and E. Campbell (2001). "A Markov Chain Monte Carlo Scheme for Parameter Estimation and Inference in Conceptual Rainfall-Runoff Modeling." Water Resour. Res. 37(4): 937-947.

Bertalanffy, Ludwig von (1968). "General Systems Theory: Foundadtions, Development, Applications." Ed.: George Braziller, New York

Beven, K. (2000). "On the future of distributed modelling in hydrology." Hydrological Processes 14(16-17): 3183-3184.

Beven, K. and J. Feyen (2002). "The Future of Distributed Modelling." Hydrological Processes 16(2): 169-172.

Beven, K.J. and J. Freer, (2001a). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems, Journal of Hydrology, 249, 11–29

Beven, K. and J. Freer (2001). "A dynamic TOPMODEL." Hydrological Processes 15(10): 1993-2011.

Beven, K., P. Smith, et al. (2006). "Comment on “Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology” by Pietro Mantovan and Ezio Todini." Journal of Hydrology 338(3-4): 315-318.

Beven, K., P. Smith, et al. (2007). "Comment on "Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology" by Pietro Mantovan and Ezio Todini." Journal of Hydrology 338(3-4): 315-318.

Beven, K. J. (2001). Rainfall-Runoff Modelling, The Primer, John Wiley & Sons.

Beven, K. J., P. J. Smith, et al. (2008). "So just why would a modeller choose to be incoherent?" Journal of Hydrology 354(1-4): 15-32.

Beven, K. J. a. and A. M. Binley (1992). "The future of distributed models: model calibration and uncertainty prediction." Hydrological Processes 6: 279–298.

Bjorkman, M. and K. Holmtrom (2000). "Global optimization of costly non-convex functions using radial basis functions." Optimization and Engineering 1(4): 373-397.

Black, P. E. (1996). Watershed Hydrology, CRC.

Blasone, R.-S., H. Madsen, et al. (2008). "Uncertainty assessment of integrated distributed hydrological models using GLUE with Markov chain Monte Carlo sampling." Journal of Hydrology 353(1-2): 18-32.

Bliznyuk, N., D. Ruppert, et al. (2008). "Bayesian Calibration and Uncertainty Analysis for Computationally Expensive Models Using Optimization and Radial Basis Function Approximation." Journal of Computational and Graphical Statistics 17(2): 270–294.

Booker, A. J., J. E. Dennis, et al. (1999). "A rigorous framework for optimization of expensive functions by surrogates." Structural Optimization 17(1): 1-13.

Borkowski, J. J. a. and G. F. Piepel (2009). "Uniform Designs for Highly Constrained Mixture Experiments." Journal of Quality Technology 49(1): 35-47.

Boughton, W. C. (1993). A hydrograph-based model for estimating the water yield of ungauged catchments. Proceedings of the 1993 Hydrology and Water Resources Conference, Institution of Engineers, Australia, National Conference Publication no. 93/14.

Boughton, W. C. (2004). "The Australian water balance model." Environmental Modelling and Software 19(10): 943-956.

Boughton, W. C. and D. G. Carroll (1993). A simple combined water balance/flood hydrograph model. Proceedings of the 1993 Hydrology and Water Resources Conference, Australia, National Conference Publication no. 93/14.

Brennan, D. (2010). "Economic potential of market-oriented water storage decisions: Evidence from Australia." Water Resour. Res. 46(W08537).

Buhmann, M. D. (2003). Radial Basis Functions: Theory and Implementations, Cambridge Unversity Press.

Cakers, U. a. and B. Yu (1998). "Using a water balance model—AWBM—to assess likely hydrological response to a significant decrease of rainfall in south-west Western Australia." Journal of Water Resources 2(2): 67–75.

Cheung, L. a. and B. Yu, 1999 (1999). Three conceptual stores in AWBM. Are they really necessary? Proceedings of the Water99 Joint Congress, Institution of Engineers, Australia.

Conn, A. R., K. Scheinberg, et al. (1997). "Recent progress in unconstrained nonlinear optimization without derivatives." Mathematical Programming, Series A 79(3): 397-414.

Cuo, L., D. P. Lettenmaier, et al. (2008). "Hydrologic prediction for urban watersheds with the Distributed Hydrology–Soil–Vegetation Model." Hydrological Processes 22: 4205–4213.

Davis, E. and M. Ierapetritou (2009). "A kriging-based approach to MINLP containing black-box models and noise." Industrial and Engineering Chemistry Research 47(16): 6101-6125.

Duan, Q. Y., V. K. Gupta, et al. (1993). "Shuffled complex evolution approach for effective and efficient global minimization." Journal of Optimization Theory and Applications 76(3): 501-521.

Dunn, S. M., J. Freer, et al. (2008). "Conceptualization in catchment modelling: simply learning?" Hydrological Processes 22(13): 2389-2393.

Eckhardt, K., L. Breuer, et al. (2003). "Parameter uncertainty and the significance of simulated land use change effects." Journal of Hydrology 273(1-4): 164-176.

Eckhardt, K., N. Fohrer, et al. (2005). "Automatic model calibration." Hydrological Processes 19(3): 651-658.

Efstratiadis, A., I. Nalbantis, et al. (2007). "HYDROGEIOS: A semi-distributed GIS-based hydrological model for disturbed river basins." Hydrology and Earth System Sciences 4: 1947-1998.

Gaganis, P. and L. Smith (2001). "A Bayesian Approach to the Quantification of the Effect of Model Error on the Predictions of Groundwater Models." Water Resour. Res. 37(9): 2309-2322.

Gao, Y., S. LÜ, et al. (2003). "Simulation of rainfall-runoff and watershed convergence process in the upper reaches of Heihe River Basin, July 2002 " Science in China Series D: Earth Sciences 47(1): 1-8.

Gassman, P. W., M. R. Reyes, et al. (2007). "The Soil and Water Assessment Tool: Historical Development, Applications, and Future Research Directions." Transactions of the ASABE 50(4): 211-1250.

Gelman, A., J. B. Carlin, et al. (2004). Bayesian Data Analysis, Chapman & Hall/CRC.

Gelman, A. and D. B. Rubin (1992). "Inference from iterative simulation using multiple sequences." Statistical Science 7: 457-511.

Gupta, H., M. Thiemann, et al. (2003). "Reply to comment by K. Beven and P. Young on ‘‘Bayesian recursive parameter estimation for hydrologic models’’." Water Resources Research 39(5): 11-17.

Gutmann, H. M. (2001(a)). "A radial basis function method for global optimizaiton." Journal of Global Optimization 19(3): 201-227.

Haario, H., M. Laine, et al. (2006). "DRAM: Efficient adaptive MCMC." Statistics and Computing 16: 239-249.

Haario, H., E. Saksman, et al. (2001). "An adaptive Metropolis algorithm
" Bernoulli 7: 223–242.

Hall, J., E. O'Connell, et al. (2006). "On not undermining the science: coherence, validation and expertise. Discussion of Invited Commentary by Keith Beven " Hydrological Processes 20: 3141-3146.

Hansen, J. R., J. C. Refsgaard, et al. (2007). "Problems with heterogeneity in physically based agricultural catchment models." Journal of Hydrology 342(1-2): 1-16.

Hastings, W. K. (1970). "Monte Carlo Sampling Methods Using Markov Chains and Their Applications." Biometrika 57(1): 97-109.

Helton, J. C., J. D. Johnson, et al. (2006). "Survey of sampling-based methods for uncertainty and sensitivity analysis." Reliability Engineering and System Safety 91: 1175–1209.

Hong, B., R. L. Strawderman, et al. (2005). "Bayesian estimation of input parameters of a nitrogen cycle model applied to a forested reference watershed, Hubbard Brook Watershed Six." Water Resour. Res. 41(3): 1-16.

Huard, D. and A. Mailhot (2006). "A Bayesian perspective on input uncertainty in model calibration: Application to hydrological model “abc”." Water Resour. Res. 42(7): 1-14.

Huard, D. and A. Mailhot (2008). "Calibration of hydrological model GR2M using Bayesian uncertainty analysis." WATER RESOURCES RESEARCH 44(W02424).

Jayakrishnan, R., R. Srinivasan, et al. (2005). "Advances in the application of the SWAT model for water resources management." Hydrological Processes 9: 749–762.

Jaynes, E. T. and G. L. Bretthorst (2003). Probability Theory: The Logic of Science. New York, Cambridge University Press.

Jencso, K. G., B. L. McGlynn, et al. (2009). "Hydrologic connectivity between landscapes and streams: Transferring reach- and plot-scale understanding to the catchment scale " Water Resource Research 45(W04428): 16pp.

Jin, Y., M. Olhofer, et al. (2000). On evolutionary optimization with approximate fitness functions. Proceedings of Genetic Evolutionary Computation Conference, Las Vegas, NV.

Jing Yang, P. R., 1 and Karim C. Abbaspour1 "Bayesian Uncertainty analysis in distributed hydrologic modeling; A case study in the Thur River basin (Switzerland)."

Jones, D. R., M. Schonlau, et al. (1998). "Efficient global optimizaiton of expensive black box functions." Journal of Global Optimization 13(4): 455-492.

Kaheil, Y. H., M. K. Gill, et al. (2006). "A new Bayesian recursive technique for parameter estimation." Water Resour. Res. 42(8): 1-12.

Kavetski, D., G. Kuczera, et al. (2006). "Bayesian analysis of input uncertainty in hydrological modeling: 2. Application." Water Resour. Res. 42(3): 1-10.

Kirkpatrick, A., L. S. Browning, et al. (2006). Irrigating with limited water supplies: A practical guide to choosing crops well suited to limited irrigation, MSU-Extension Service, EB169: 29.

Kuczera, G., D. Kavetski, et al. (2006). "Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters." Journal of Hydrology 331: 161– 177.

Lenhart, T., K. Eckhardt, et al. (2002). "Comparison of two different approaches of sensitivity analysis " Physics and Chemistry of the Earth, Parts A/B/C 27(9-10): 645-654.

Leung, L. R. and M. S. Wigmosta (1999). "Potential climate change impacts on mountain watersheds in the pacific northwest." Journal of American Water Resources Association 35(5): 1463–1471.

Li, X., S. Zhong, et al. (2010). "Hydroclimate and variability in the Great Lakes region as derived from the North American Regional Reanalysis." Journal of Geophysical Research 115(D12104).

Liu, Y. and H. V. Gupta (2007). "Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework." Water Resour. Res. 43(7): 1-18.

Lumijärvi, J., J. a. Laurikkala, et al. (2004). A Comparison of Different Heterogeneous Proximity Functions and Euclidean Distance. MEDINFO 2004, Amsterdam, IOS Press.

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

Mantovan, P., E. Todini, et al. (2007). "Reply to comment by Keith Beven, Paul Smith and Jim Freer on "Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology"." Journal of Hydrology 338(3-4): 319-324.

Marazzi, M. and J. Nocedal (2002). "Wedge trust region methods for derivative free optimization." Mathematical Programming, Series A 91: 289-305.

Markov, A. A. (1906). "Rasprostranenie zakona bol'shih chisel na velichiny, zavisyaschie drug ot druga " Izvestiya Fiziko-matematicheskogo obschestva pri Kazanskom universitete 2(15): 135-156.

Marshall, L., D. a. Nott, et al. (2004). "A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modeling." Water Resour. Res. 40(2): 1-11.

Marshall, L., D. a. Nott, et al. (2007). "Towards dynamic catchment modelling: a Bayesian hierarchical mixtures of experts framework." Hydrological Processes 21: 847–861

Mashamba, A. and L. Marshall (2010). A case study examining factors affecting the performance of response surface modeling during bayesian optimization and uncertainty analysis of hydrologic models (In Review), Environmental Modeling and Software.

Metropolis, N., A. W. Rosenbluth, et al. (1953). "Equations of State Calculations by Fast Computing Machines." Journal of Chemical Physics 21(6): 1087-1092.

Mincemoyer, S. A. and J. L. Birdsall (2006). "Vascular flora of the Tenderfoot Creek Experimental Forest, Little Belt Mountains, Montana." Madrono 53(3): 211-222.

Montanari, A. (2007). "What do we mean by "uncertainty"? The need for a consistent wording about uncertainty assessment in hydrology." Hydrological Processes 21(6): 841-845.

Montanari, A. and A. Brath (2004). "A stochastic approach for assessing the uncertainty of rainfall-runoff simulations." Water Resour. Res. 40(1): 1-11.

Montgomery, D. C. (2004). Design and Analysis of Experiments, John Wiley & Sons.

Moradkhani, H., K. L. Hsu, et al. (2005). "Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter." Water Resour. Res. 41(5): 1-17.

Moreda, F., V. Koren, et al. (2006). "Parameterization of distributed hydrological models: learning from the experiences of lumped modeling." Journal of Hydrology 320(1-2): 218-237.

Mugunthan, P. and C. A. Shoemaker (2006). "Assessing the impacts of parameter uncertainty for computationally expensive groundwater models." Water Resour. Res. 42(10): 1-15.

Mugunthan, P., C. A. Shoemaker, et al. (2005). "Comparison of Function Approximation, Heuristic and Derivative-Based Methods for Automatic Calibration of Computationally Expensive Groundwater

Bioremediation Models." Water Resour. Res. 41(W11427).

Muleta, M. K. and J. W. Nicklow (2005). "Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model." Journal of Hydrology 306(1-4): 127-145.

Muttil, N., S.-y. a. Liong, et al. (2007). A parallel shuffled complex evolution model calibrating algorithm to reduce computational time. MODSIM 2007 International Congress on Modelling and Simulation, Christchurch, New Zealand.

Myers, R. H. a. and D. C. Montgomery (1995). Response Surface Methodology, Process and Product Optimization Using Designed Experiments, John Wiley and Sons, Inc.

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

Pappenberger, F., K. Beven, et al. (2005). "Uncertainty in the calibration of effective roughness parameters in HEC-RAS using inundation and downstream level observations." Journal of Hydrology 302(1-4): 46-69.

Powell, M. J. D. (1992). The theory of radial basis function approximation in 1990. Advances in Numerical Analysis, Oxford University Press.

Powell, M. J. D. (1996). A review of algorithms for thin plate spline interpolation in two dimensions. Advanced topics in multivariate approximation, River Edge, NJ, World Scientific Publishing.

Powell, M. J. D. (2000). "UOBYQA: unconstrained minimization without derivatives." Mathematical Programming, Series A 92: 555-582.

Powell, M. J. D. (2002). " On trust region methods for unconstrained minimization without derivatives." Mathematical Programming, Series A 97: 605-623.

Qureshi, M. E., K. Schwabe, et al. (2010). "Environmental water incentive policy and return flows." Water Resour. Res. 46(W04517).

Regis, R. G. and C. A. Shoemaker (2005). "Constrained global optimization using radial basis functions." Journal of Global Optimization 31: 153-171.

Regis, R. G. a. and C. A. Shoemaker (2004). "Local Function Approximation in Evolutionary Algorithms for the Optimization of Costly Functions." IEEE Transactions On Evolutionary Computation 8(5): 490-505.

Regis, R. G. a. and C. A. Shoemaker (2007). "Improved strategies for radial basis function methods for global optimization." Journal of Global Optimization 37: 113–135.

Renard, B., V. Garreta, et al. (2006). "An application of Bayesian analysis and Markov chain Monte Carlo methods to the estimation of a regional trend in annual maxima." Water Resour. Res. 42(12): 1-17.

Romanowicz, R., K. J. Beven, et al. (1994). Evaluation of predictive uncertainty in nonlinear hydrological models using a Bayesian approach Statistics for the Environment 2: Water-Related Issues, New York, John Wiley.

Romanowicz, R., K. J. Beven, et al. (1996). Bayesian calibration of flood inundation models. Floodplain Processes, New York, John Wiley.

Rouhani, H., P. Willems, et al. (2007). "Parameter estimation in semi-distributed hydrological catchment modelling using a multi-criteria objective function." Hydrological Processes 21: 2998–3008.

Sahoo, G. B., C. Ray, et al. (2006). "Calibration and validation of a physically distributed hydrological model, MIKE SHE, to predict streamflow at high frequency in a flashy mountainous Hawaii stream." Journal of Hydrology 327(1-2): 94-109.

Santhi, C., R. Srinivasan, et al. (2006). "A modeling approach to evaluate the impacts of water quality management plans implemented in a watershed in Texas " Environmental Modelling and Software 21(8): 1141-1157.

Selker, J. (2002). "Breaking the cycle of futility in Hydrosciences." Hydrological Processes 16(3): 743-744.

Sharifi, F. a. and M. J. Boyd (1994). A comparison of the SFB and AWBM rainfall–runoff models. Proceedings of the Water Down Under Conference, Institution Engineers Australia.

Sieber, A. and S. Uhlenbrook (2005). "Sensitivity analyses of a distributed catchment model to verify the model structure." Journal of Hydrology 310(1-4): 216-235.

Sivakumar, B. (2008). "The more things change, the more they stay the same: the state of hydrologic modelling." Hydrological Processes 22(21): 4333-4337.

Sivakumar, B. (2008). "Undermining the science or undermining Nature?" Hydrological Processes 22(6): 893-897.

Smith, T. a. and L. Marshall (2008). "Bayesian methods in hydrologic modeling: A study of recent advancements in Markov chain Monte Carlo techniques." Water Resources Research 44(4).

Soulsby, C., C. Neal, et al. (2008). "Catchment data for process conceptualization: simply not enough?" Hydrological Processes 22(12): 2057-2061.

Storck, P., L. Bowling, et al. (1998). "Application of a GIS-based distributed hydrology model for prediction of forest harvest effects on peak stream flow in the Pacific Northwest." Hydrological Processes 12(6): 889–904.

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 6: 239–249.

Thanapakpawina, P., J. Richeyb, et al. (2007). "Effects of landuse change on the hydrologic regime of the Mae Chaem river basin, NW Thailand " Journal of Hydrology 334(1-2): 215-230.

Thiemann, M., M. Trosset, et al. (2001). "Bayesian recursive parameter estimation for hydrologic models." Water Resources Research 37(10): 2521–2535.

Tolson, B. A. and C. A. Shoemaker (2007). "Dynamically dimensioned search algorithm for computationally efficient watershed model calibration." WATER RESOURCES RESEARCH 43(W01413).

Tyagi, A. and C. Haan (2001). "Uncertainty Analysis Using Corrected First-Order Approximation Method." Water Resour. Res. 37(6): 1847-1858.

V´azquez, R. F., L. Feyen, et al. (2002). "Effect of grid size on effective parameters and model performance of the MIKE-SHE code." Hydrological Processes 16(2): 355-372.

van Griensven, A. and T. Meixner (2006). "Methods to quantify and identify the sources of uncertainty for river basin water quality models." Water Science and Technology 53(1): 51-59.

van Griensven, A., T. Meixner, et al. (2006). "A global sensitivity analysis tool for theparameters of multivariable catchment models." Journal of Hydrology 324(1-4): 10-23.

VanShaar, J. R., I. Haddeland, et al. (2002). "Effects of land-cover changes on the hydrological response of interior Columbia River basin forested catchments." Hydrological Processes 16(13): 2499–2520.

Vogel, R. M. and A. Sankarasubramanian (2003). "Validation of a watershed model without calibration." Water Resour. Res. 39(10): SWC 7-1 - SWC 7-9.

Vrugt, J. A., H. V. Gupta, et al. (2003). "Effective and efficient algorithm for multiobjective optimization of hydrologic models." Water Resour. Res. 39(8): SWC 5-1 - SWC 5-19.

Vrugt, J. A., H. V. Gupta, et al. (2003). "A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters." Water Resour. Res. 39(8): SWC 1-1 - SWC 1-16.

Wagener, T. and H. S. Wheater (2006). "Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty." Journal of Hydrology 320(1-2): 132-154.

Whitaker, A., Y. Alila, et al. (2003). "Application of the distributed hydrology soil vegetation model to Redfish Creek, British Columbia: model evaluation using internal catchment data." Hydrological Processes 17(2): 199-224.

Wigmosta, M. S., B. Nijssen, et al. (2002). The Distributed Hydrology Soil Vegetation Model Mathematical Models of Small Watershed Hydrology and Applications, Littleton, CO, Water Resource Publications.

Wigmosta, M. S., L. W. a. Vail, et al. (1994). "A distributed hydrology-vegetation model for complex terrain." Water Resour. Res. 30(6): 1665-1679.

Wiley, M. W. and R. N. Palmer (2008). "Estimating the impacts and uncertainty of climate change on a municipal water supply system." Journal of Water Resources Planning and Management 134(3): 239-246.

Winchell, M., R. Srinivasan, et al. (2007). ArcSWAT interface for SWAT2005 – User's Guide. Blackland Research Center, Texas Agricultural Experiment Station and Grassland, Soil and Water Research Laboratory, USDA Agricultural Research Service, Temple, Texas.

Wissmar, R. C., K. T. Raymond, et al. (2004). "Effects of Changing Forest and Impervious Land Covers on Discharge Characteristics of Watersheds." Environmental Management 34(1): 91-98.

Wooldridge, S., J. Kalma, et al. (2001). "Parameterisation of a simple semi-distributed model for assessing the impact of land-use on hydrologic response." Journal of Hydrology 254(1-4): 16-32.

Yang, D., S. Herath, et al. (2000). "Comparison of different distributed hydrological models for characterization of catchment spatial variability." Hydrological Processes 14(3): 403-416.

Yang, J., P. Reichert, et al. (2008). "Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China." Journal of Hydrology 358(1-2): 1-23.

Yang, J., P. Reichert, et al. (2007). "Hydrological modelling of the Chaohe Basin in China: Statistical model formulation and Bayesian inference." Journal of Hydrology 340(3-4): 167-182.

Yao, C. and Z. Yang (2009). "Parameters optimization on DHSVM model based on a genetic algorithm." Frontiers of Earth Science in China 3(3): 374-380.

Ye, K. Q., W. a. Li, et al. (2000). "Algorithmic Construction of Orthogonal Symmetric Latin Hypercube Designs." Journal of Statistical Plannning and Inference 90: 145-159.

Zheng, Y. and A. A. Keller (2006). "Understanding parameter sensitivity and its management implications in watershed-scale water quality modeling." Water Resources Research 42(5): 1-14.

Zorzettoa, L. F. M., R. Maciel Filho, et al. (2001). "Processing modelling development through artificial neural networks and hybrid models " Computers and Chemical Engineering 24(2-7): 1355-1360.