

The Pennsylvania State University
The Graduate School

MODELING, PARAMETER OPTIMIZATION, AND
ECOHYDROLOGIC ASSESSMENT OF WATERSHED SYSTEMS

A Dissertation in
Civil and Environmental Engineering
by
Xuan Yu

© 2014 Xuan Yu

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

July 2014

The dissertation of Xuan Yu was reviewed and approved* by the following:

Christopher J. Duffy
Professor of Civil Engineering
Dissertation Advisor, Chair of Committee

Murali Haran
Associate Professor of Statistics

Xiaofeng Liu
Assistant Professor of Civil Engineering

Jason Kaye
Associate Professor of Soil Biogeochemistry

Peggy Johnson
Department Head, Civil and Environmental Engineering

*Signatures are on file in the Graduate School.

Abstract

Integrated watershed models describe the land-phase of hydrologic cycles by coupling processes such as canopy interception, infiltration, recharge, evapotranspiration, overland flow, vadose zone flow, groundwater flow, and channel routing. This modeling scheme serves as an important strategy for understanding the moisture redistribution processes across the watershed and river-basin landscape. For example, the Penn State Integrated Hydrologic Model (PIHM) has successfully been applied to explain the impacts of antecedent soil moisture on peak streamflow and timing. However, due to the heavy computational cost of solving integrated models with complex model structure, efficient parameter estimation for PIHM is a major computational and algorithmic challenge. The focus of this dissertation has four main themes: (1) develop an efficient calibration strategy for PIHM; (2) develop a weighted-objective calibration scheme for multi-variable distributed parameters (e.g., streamflow, water table depth, and eddy flux data); (3) test the parameter-estimation process for spatial shallow groundwater calibration of PIHM using national wetland geospatial data (National Wetland Inventory: NWI); (4) extend the capabilities of PIHM for linking vegetation dynamics from an ecosystem model and evaluating the importance of vegetation growth in water balance studies.

The temporal and geospatial complexity of the data requirements for integrated and fully coupled catchment models increases the difficulty of applying parameter optimization in real watershed applications. In this research, a new strategy known as partition calibration was proposed to enable the automatic calibration of PIHM. The concept can be thought of as a divide-and-conquer algorithm, where the parameter space is divided into two or more sub-problems that can be solved sequentially. The first partition of the parameter vector is divided according to the two dominant time-scales of catchment hydrological processes: 1) event-scale hydrologic response parameters; and 2) seasonal-scale response parameters. Once

divided, the event-scale group parameters and seasonal-scale group parameters are then calibrated sequentially. The second partition focused on the separation of the total calibration objective onto multiple targets to predict each observed hydrological variable. The informativeness of each calibration target was defined in terms of a weighted objective function. Application of the scheme suggested the use of an informativeness-based partitioning of streamflow, groundwater, and ET parameters and demonstrated that partition calibration was superior to both single-objective calibration and un-weighted average multi-objective calibration. Applications of the PIHM were found to be efficient with the partition calibration strategy. The first PIHM application involves characterization of the freshwater wetland response to climate change at seven catchments within the Susquehanna River Basin. In this case, streamflow time series and geospatial mapping of wetlands in the National Wetland Inventory (NWI) were used to calibrate the model to capture the distributed groundwater and streamflow dynamics. After calibration, the model was applied to an IPCC climate change scenario (2046-2065), and the modeling results suggested that upland groundwater levels were more sensitive to climate change than water levels of wetlands in lower parts of the catchment, as expected. In the final part of this research, long-term modeling of PIHM compared the role of fixed seasonal variation in LAI (Leaf Area Index) and a fully dynamic vegetation growth model. The community ecosystem model BIOME-BGC was linked to PIHM to test the hypothesis that default monthly LAI values are sufficient to represent long-term water balances in a catchment. By comparing model results for fixed LAI and dynamic LAI, it was demonstrated that fixed LAI is not sufficient for capturing interannual variability of forest vegetation and water flow dynamics, especially as it relates to the onset and growth of forest.

Table of Contents

List of Figures	viii
List of Tables	x
List of Symbols	xi
Acknowledgments	xii
Chapter 1	
Introduction	1
1.1 Motivation	1
1.2 Background and Scope	2
1.2.1 Watershed Hydrologic Models	2
1.2.2 Biogeochemical Processes of Hydrological Systems	5
1.2.3 A Formal Strategy for the Complex Watershed Modeling - Divide and Conquer (D&C)	7
1.3 Organization of the Dissertation	8
Chapter 2	
A Two-partition Calibration Strategy for Physics-based Fully Coupled Watershed Modeling Using National Data and the Evolutionary Algorithm	9
2.1 Introduction	9
2.2 Model Formulation and National Data Set Support	11
2.2.1 PIHM	11
2.2.2 Hydrological Processes	11
2.2.3 Watershed Description	17
2.2.4 National Data Support	18

2.2.5	A-Priori Parameters	19
2.3	Partition Calibration Strategy	21
2.3.1	Monte Carlo Sensitivity Analysis	23
2.3.2	Calibration of EG Parameters	24
2.3.3	Calibration of SG Parameters	26
2.4	Results and Discussion	27
2.4.1	Results and Validation	27
2.4.2	Strengths and Limitations of PCS	31
2.5	Uncertainty Analysis	32
2.5.1	Posterior Parameter Evaluation of One-dimensional case	32
2.5.2	Posterior Parameter Evaluation of Two-dimensional Case	32
2.6	Conclusion	35

Chapter 3

	”Informativeness” as a Quantitative Index of Weighted-objective Calibration for a Multi-state Distributed Hydrologic Model	36
3.1	Introduction	36
3.2	Proposed Methodology	38
3.2.1	Selection of Metrics According to Effectiveness	39
3.2.2	Evaluate the Informativeness of Each Calibration Target	41
3.2.3	Weighted Function for Multi-objective Optimization	42
3.3	Application Example: Catchment and Model Setup	43
3.3.1	Multiple Observation at the Catchment	43
3.3.2	Model Setup and Parameterization	44
3.3.3	Performance of Each Metric	44
3.3.4	Formulate Aggregating Objective Function	48
3.3.5	Model Calibration and Evolution of the Model Performance	48
3.4	Discussion	55
3.4.1	Performances of Metrics	55
3.4.2	The Meaning of Informativeness of Each Calibration Target in Constraining Model Parameters	55
3.4.3	Comments on the Informativeness-based Weighted Calibration	56
3.5	Conclusion	59

Chapter 4

	Chapter 4 The National Wetland Inventory as a Constraint on Wetland Hydrology Modeling for Regional Climate Change Impact Assessment	60
4.1	Introduction	60

4.2	Description of Study Area	62
4.3	Materials and Methods	66
4.3.1	Physics-based Hydrological Modeling	66
4.3.2	Climate Scenarios for SRB	67
4.3.3	Proposed Methodology	68
4.3.4	Model Data Setup and Parameter Calibration	68
4.3.5	Assessment of Climate Change Impacts on Wetland	73
4.4	Results	76
4.4.1	Hydrologic Validation	76
4.4.2	Wetland Spatial Distribution from Reanalysis	76
4.4.3	Climate Change Responses	77
4.5	Discussion	81
4.5.1	How Does Physics-based Hydrologic Modeling Explain Groundwater and Stream Water Interaction?	81
4.5.2	Differences between NWI and Hydrologic Identification	82
4.5.3	Groundwater Level Sensitivity to Climate Change	82
4.6	Conclusion and Outlook	82
Chapter 5		
Modeling Dynamic Ecosystem Processes: A Case for Improving Hydrological Predictability		84
5.1	Introduction	84
5.2	Overview of Biome-BGC	86
5.3	Biome-BGC Modeling at SSHCZO	86
5.4	Modeling Vegetation Dynamic for PIHM Simulation	91
5.5	Implications on Further Coupling of Biome-BGC and PIHM	92
5.6	Conclusion	94
Chapter 6		
Summary and Future Work		95
6.1	Summary	95
6.2	Future Work	97
Bibliography		99

List of Figures

2.1	The PIHM representation of hydrological processes and model coupling strategy.	12
2.2	PIHM processes and domain decomposition for Shale Hills.	14
2.3	Location of SSHCZO.	18
2.4	Results of regional sensitivity analysis (RSA) for parameters in PIHM.	24
2.5	Observed and simulated streamflow for the year 2009 at the outlet of SSHCZO.	28
2.6	Validation performance for 1974 experiment at SSHCZO.	30
2.7	Evolution of parameter value over time during the EG calibration runs.	33
2.8	The Monte Carlo simulation of Van Genuchten parameter α	34
2.9	The Monte Carlo simulation of Van Genuchten parameter α and β	34
3.1	Flowchart of the informativeness-based weighted-objective calibration methodology.	40
3.2	Map of the SSHCZO catchment showing the locations of streamflow station, weather station, and pressure transducers.	43
3.3	The PIHMgis tools process spatial layers to assign physical values and hydrological parameters.	45
3.4	Observed time series data for model calibration in June 2009.	47
3.5	Comparison of metrics.	49
3.6	NSE of each hydrological variables in the Monte Carlo analysis.	50
3.7	The simulated and observed streamflow, water table depth at Site A, Site B, and ET: (a) single objective calibration; (b) weighted calibration; (c) un-weighted calibration.	53
3.8	Trajectories of metrics at streamflow, site A, site B, and ET during the weighting objective function evolution.	54
3.9	Objective function values in the non-dominated solutions from CMA-ES samples.	58

4.1	Location of the 7 watersheds in the Susquehanna River Basin, and PIHM domain decomposition.	63
4.2	Climate change projection.	69
4.3	Flowchart of the hydrologic modeling for wetland assessment.	70
4.4	Model data setup processes of KC.	72
4.5	PIHM calibration objectives to capture wetland.	74
4.6	Observed and simulated streamflow (2004-2010) at each of the 7 catchments in the study.	78
4.7	NWI locations (red dots) and PIHM simulated shallow water table conditions during the reanalysis period.	79
4.8	Difference maps for the simulated groundwater depth between the climate scenario period 2046-2065 and the reference period 1979-1998.	79
4.9	Distribution of groundwater table depth responses in the wetland, upland, and the rest of the area of the watershed.	80
5.1	Biome-BGC simulated NEE compared with flux tower estimation.	89
5.2	Biome-BGC simulated inter-annual variability of max LAI of 3 dominant tree species at Shale Hills.	89
5.3	Biome-BGC simulated inter-annual variability of phenology.	90
5.4	PIHM-modeled streamflow sensitivity to LAI (default LAI and modeled LAI). Note that using the fixed-seasonal LAI tends to overestimates the peakflows later in the 1974 experiment.	92
5.5	Flow chart of hydrological coupling between PIHM and Biome-BGC.	93
5.6	Modeled spatial pattern of soil carbon at Shale Hills.	93

List of Tables

1.1	Real watershed simulated by fully-coupled physics-based hydrologic models	4
2.1	Main equations of PIHM	13
2.2	Data support for physics-based hydrologic modeling studies	20
2.3	Calibration parameters, corresponding hydrologic processes and their typical behavior as event-scale parameters group (EG) or seasonal-scale parameters group (SG)	22
2.4	Rescaling function of real parameters used in the calibration algorithm	26
3.1	Sensitive PIHM parameters and their feasible ranges.	46
3.2	Calibration results form single objective, informativeness-based weighted-average, and un-weighted average.	52
4.1	Overview of watersheds and USGS gauges.	64
4.2	Percentage of land cover classes in each watershed.	64
4.3	Mesh resolution	71
4.4	PIHM performance in daily streamflow simulation from 2004 to 2010	76
5.1	Stand characteristics at SSHCZO.	87
5.2	Eco-physiological parameters for Biome-BGC at SSHCZO.	88

List of Symbols

Acknowledgments

Foremost, I would like to express my cordial gratitude to my advisor Prof. Christopher J. Duffy for the continuous support of my study at Penn State, where I for the first time lived in a foreign nation. His enthusiasm has inspired my research aspiration. His patience has encouraged my persuasion. His open-mindedness has promoted my research with exciting collaboration. I feel extremely lucky to be his student.

I would like to thank Prof. Murali Haran, Prof. Xiaofeng Liu, and Prof. Jason Kaye not only for their time and extreme patience, but for their intellectual contributions with the research methodology and valuable feedbacks on improving the dissertation.

I thank Gopal Bhatt, my colleague and terrific friend, who has been offering creative research ideas and precious studying abroad experiences. I also thank my colleagues Yuning Shi, Wenfang Li, Lorne Leonard, Colin Duffy, Yu Zhang, and Lele Shu for their collaboration on diversity research projects.

Finally, I would like to thank my parents for their patience, and to thank my wife, Wenjuan Zheng for her love and support. I also thank my newborn baby, Annabelle Yu, for being good kid with your mother when I was studying.

Dedication

Introduction

1.1 Motivation

The terrestrial water cycle is the medium for the exchanges of energy, water and momentum through biogeospheric components: animals, plants, microorganisms, soils, rocks, and atmosphere [Smith et al.(2008)]. On the one hand, local moisture availability supports the diversity and distribution of the types of regional vegetation, and controls net primary productivity. The ecohydrological approach has been widely accepted as new methodology and solution for complex environmental systems, which helps us to improve forecasts of environmental change [Newman et al.(2006)]. However, the different time and space scale that characterizes the ecosystem and hydrological cycles is one of the major challenges for integrated modeling. Recently, with advances in computing power, a new generation of physics-based hydrologic response models attempts to simulate coupled hydrological processes over a range of spatial and temporal scales [Qu & Duffy(2007), Mirus et al.(2011)]. These comprehensive physics-based models were suggested as a strong foundation for linking research disciplines such as hydroecology and hydrogeomorphology [Loague et al.(2006)]. Nonetheless, limited research has been conducted using physics-based hydroecological models. This is largely due to the model requirement of large amounts of geospatial and temporal data, as well as the model-data computational cost. Furthermore, the calibration of such complex models greatly expands the computational requirement. Therefore, an important task for physics-based ecohydrological model development is to

design calibration tools, to simplify model application workflow, and to extend simulation results and modules for specific ecohydrological researches. It is also my expectation that fully coupled, physics-based ecohydrological models will provide an important new tool for advancing Earth system science and engineering applications.

1.2 Background and Scope

1.2.1 Watershed Hydrologic Models

Hydrological simulation has been important in the development of water resources planning and management applications in different disciplines [Praskievicz & Chang(2009), Asbjornsen et al.(2011), Barthel et al.(2012)]. Quantitative description of the hydrologic cycle is still the major task of hydrological modelers. Originated with component models, the hydrological simulation is now dominated by verity types of watershed models [Singh & Frevert(2002), Kampf & Burges(2007)]. Recently, a number of physics-based, fully coupled hydrological models have been proposed due to advances in computational power [Mirus et al.(2011)]. Such models have the advantage of understanding spatio-temporal surface-subsurface water interactions, which implicitly solve a system of equation [Loague et al.(2006)]. Specifically, physics-based models include comprehensive representation of hydrological processes and the spatial hydrological properties of a watershed. Therefore, the simulation includes different rainfall runoff generation mechanisms (Horton overland runoff, Dunne overland runoff, and subsurface flow) [Qu & Duffy(2007)], and the results are spatially heterogeneous and temporally high resolution [Mirus et al.(2011)]. These characteristics of physics-based models will not only benefit hydrological studies but also provide a strong foundation for ecological and biogeochemical research [Loague et al.(2006)].

Despite their utility, the applications of these physics-based models (Table 1.1) are limited because the models are computational expensive, because they require large amounts of geospatial data for parameterization, and because there is a lack of effective calibration strategies [Mirus et al.(2011)]. Most applications were

reported at the small experimental catchment scale. For example, the area of the R-5 catchment simulated by the Integrated Hydrology Mode (InHM) for six years, is 0.1 km^2 [Heppner et al.(2007)]. The Susquehanna-Shale Hills Critical Zone Observatory (SSHCZO) watershed is a 0.08 km^2 experimental catchment of the application site of the Penn State Integrated Hydrologic Model (PIHM) [Qu & Duffy(2007)]. A headwater area of the Sakai River of 2.5 km^2 was simulated with a spatially-distributed, physics-based model of composite dimensions, consisting of a 3-D variably-saturated groundwater flow sub-model, a 2-D overland flow sub-model, and 1-D river flow sub-model [Takeuchi et al.(2010)]. In addition, few literatures have reported the calibration strategy of such complex models. Hence, trial and error techniques for tuning the parameters according to the model sensitivity are still the primary choices to fit the runoff response [Takeuchi et al.(2010), Li et al.(2008)]. Without a comprehensive spatial calibration, even a successful application could not comprehensively demonstrate the advantage of physics-based models [Kollet & Maxwell(2008)].

Physics-based model study	Study area	Spatial resolution	Temporal resolution	Modeling duration	Validation
Atmosphere-subsurface coupling in Little Washita [Maxwell et al.(2007)]	611km ²	1km×1km / 0.5m (vertical)	1hour / 1 second / implicit	36 hours	None
HydroGeoSphere Model application in Duffins Creek watershed [Li et al.(2008)]	286.6km ²	<78000m ² / layers (vertical)	24 1 day	2 years	streamflow groundwater table elevation
Penn State Integrated Hydrologic model application in SSHCZO [Qu & Duffy(2007)]	0.078km ²	average 250m ² / layers (vertical)	2 1 second / implicit	3 months	streamflow, soil moisture, and groundwater table
The Integrated Hydrology Model (InHM) application in Laurel Creek Watershed [Jones et al.(2008)]	75km ²	29,638 nodes / layers (vertical)	20 implicit	900 hours	streamflow and hydraulic head

Table 1.1. Real watershed simulated by fully-coupled physics-based hydrologic models

1.2.2 Biogeochemical Processes of Hydrological Systems

It is often assumed that biogeochemical processes change relatively slowly in comparison to rainfall-runoff hydrological responses. If this assumption is true, it is convenient to say that the biophysical properties of soil and vegetation are static or that they can be assigned a fixed seasonal pattern during the hydrological simulation. On the other hand, we also know that forest ecosystem and catchment hydrologic response form an integrated system in which both short and long time scale processes may interact (e.g., soil moisture-groundwater redistribution), making static or fixed assumptions less likely. It seems important then that assumptions about biogeochemical and hydrological processes be tested in some way. Although calibrated parameters could lead to adequate hydrological model performance, it is safe to say that vegetation and soil biophysical contributions to hydrological processes are still inadequately understood [Brolsma & Bierkens(2007), Smucker & Hopmans(2007), Miller et al.(2010)].

Models simulating biogeochemical processes vary dramatically according to temporal scales, spatial complexity, and mechanistic detail. Based on the mechanism, there are two groups of models: the statistic empirical model and process-based model. Although empirical models are still being applied in vast majority of forest managements [Mäkelä et al.(2008), Wang et al.(2009)], growing process-based models have emerged to develop a better understanding of ecosystem functions [Smith et al.(2001), Hanson et al.(2004)]. Ecosystem models simulate from individual-level [Post & Pastor(1996)] to stand-level [Hanson et al.(2004)] to global-level [Coe et al.(2000)]. Relating to the interactions with watershed hydrologic processes, a stand-level ecosystem model is the optimum match. We therefore focused on the stand-level, process-based ecosystem models. In general, stand-level, process-based ecosystem models explicitly calculate the carbon and water fluxes at a time step varying from an hour to a month at a given point, which is then scaled on a per-square-meter basis [Hanson et al.(2004)]. The carbon, nitrogen, and water pools are stored in plants and soil, which are divided into several parts, depending on the model structure. The water cycle is usually moderately represented in a soil water balance bucket model, which is mainly driven by the meteorological data. The carbon and nitrogen fluxes are determined by eco-physiological characteristics of the plant function type (PFT) [Tague &

Band(2004)].

Due to the one-dimensional hydrologic estimation, most models demonstrate the inability to capture the day-to-day dynamics of changes of soil water content [Hanson et al.(2004)]. Therefore, the effect of soil moisture on physiological processes is not correctly represented, which impacts the model's performance of net ecosystem carbon exchange (NEE) under extreme soil water conditions [Hanson et al.(2004)]. Comparative studies suggested that models have different strengths, and improvements should be made to complete simulation of all the dynamics of the coupled water cycle and carbon cycle [Morales et al.(2005), Siqueira et al.(2006), Ichii et al.(2010)].

Coupling hydrological and biogeochemical model is one of the major solutions to understand the multi-scale of interactions between water cycle and carbon cycle [Morales et al.(2005)]. For example, an ecosystem model was linked with a hydrological model to test the sensitivity of the ecosystem model to hydrology and temperature [Wolf et al.(2008)]. The results suggested that soil moisture and soil temperature are the most sensitive factors driving carbon fluxes, particularly soil carbon emissions [Wolf et al.(2008)]. The vegetation simulation was improved in hydrologic transport model SWAT (Soil Water Assessment Tool) by implementing a field-scale plant model [Kiniry et al.(2008)]. Increasing studies showed that only complete representation of connected ecologic and hydrologic processes could comprehensively improve the understanding of the dynamics of ecohydrosystem [Yi et al.(2009)].

The Regional Hydro-Ecological Simulation System (RHESSys) is a hydro-ecologic model designed to simulate integrated water, carbon, and nutrient cycling and transport over spatially variable terrain [Tague & Band(2004)]. RHESSys explicitly couples the Biome-BGC for ecological processes and TOPMODEL for hydrological processes at a daily time scale. The model structure is based on its hierarchical landscape representation from basins to hillslopes, zones, and patches, which aggregate the fluxes of processes in different scales.

For the time scale of stand-level, processes-based models, most models use an hourly simulation time-step [Hanson et al.(2004)]. Higher temporal resolution and detailed model structure will provide better output in agreement with the observations [Amthor et al.(2001), Hanson et al.(2004)]. Among watershed

modeling studies, using physics-based distributed models has exhibited advantages in capturing multi-scale hydrological dynamics [Mirus et al.(2011)]. It is therefore necessary for future modeling studies to improve coupling strategies in order to accurately understand the ecohydrological dynamics in terrestrial water-carbon-nitrogen cycling.

1.2.3 A Formal Strategy for the Complex Watershed Modeling - Divide and Conquer (D&C)

Recent decades have seen advances in comprehensive, high-resolution geospatial data products, which have encouraged the development of distributed integrated model applications. However, as the modeling framework built up diverse processes, the model applications turned out to be a sophisticated problem involving distributed parameterization, model calibration, and model discrepancy identification, etc. An efficient solution strategy for complex watershed modeling is the divide-and-conquer strategy. The divide-and-conquer strategy solves a problem by:

1. Breaking it into sub-problems that are themselves smaller instances of the same type of problem;
2. Recursively solving these sub-problems; and
3. Appropriately combining their answers.

The real work is done piecemeal, in three different places: in the partitioning of problems into sub-problems; at the very tail end of the recursion, when the sub-problems are so small that they are solved outright; and in the gluing together of partial answers. These partial answers are held together and coordinated by the algorithm's core recursive structure. In computer science, divide and conquer (D&C) is an important algorithm design paradigm based on multi-branched recursion. A D&C algorithm works by recursively breaking down a problem into two or more sub-problems of the same (or related) type, until these become simple enough to be solved directly. The solutions to the sub-problems

are then combined to give a solution to the original problem. In the field of hydrological model calibration, many studies have suggested that it is important and efficient to decompose the calibration processes [Boyle et al.(2000), Dunn & Colohan(2010), Hay et al.(2006), Hogue et al.(2006)]. In this research, the focus is on developing a more efficient yet integrated physical representation of hydrologic processes at the catchment scale and exploring how national data sets can support such models within a testable framework. As part of this study, it was necessary to develop a new partition calibration strategy that efficiently allows multi-state geospatial/temporal parameters to be estimated. An important element of this research is the role of ecosystem variables on the hydrologic process operating in the catchment. Case studies involving wetland hydro-ecology and dynamic forest water use serve to test the hypothesis that ecosystems have a dynamic role to play in catchment hydrology.

1.3 Organization of the Dissertation

The remaining content of this dissertation is based on a serial of successive research on hydrological modeling. The chapters are organized to address individual issues, which designate four research papers.

- Chapter 2 presents a two-partition calibration strategy for physics-based, fully coupled watershed modeling using national data and an evolutionary algorithm, which has been published in *Computers and Geosciences*.
- Chapter 3 presents a three-step strategy to optimize parameters for distributed watershed modeling using weighted objectives between multiple observed variables. This paper is intended to be submitted to *Journal of Hydrologic Engineering*.
- Chapter 4 presents a spatial calibration strategy to constrain groundwater table modeling in the representation of wetland dynamics. This paper is intended to be submitted to *Hydrological Processes*.

- Chapter 5 presents a coupling study of hydrologic and terrestrial biogeochemical processes to test the common assumption that catchment runoff can be adequately simulated with a fixed seasonal forest evapotranspiration model. This paper has been accepted as a book chapter of the Geophysical Monograph Series by the American Geophysical Union.

In the end, Chapter 6 summarizes the scientific contribution of this dissertation and suggests potential improvements for future work.

A Two-partition Calibration Strategy for Physics-based Fully Coupled Watershed Modeling Using National Data and the Evolutionary Algorithm

2.1 Introduction

Physics-based, fully coupled and distributed hydrologic models seek to simulate hydrologic states in space and time with representations of hydrologic processes and parameters that have physical meanings. Ideally, the model should not require calibration if all parameters were available through experimentation, field measurements, and national data coverage sets. For example, the Soil Survey Geographic Database (SSURGO) with soil textural information was tested as a useful soil physical property for a-priori parameter estimation of distributed hydrological modeling [Anderson et al.(2006)]. Such national datasets provide a measure of spatial variations and can potentially meet the data requirements of distributed hydrological models. However, due to the high uncertainty of spatially distributed soils and estimation of geologic properties [Vereecken et al.(2010)], calibration is still an indispensable part of the physicsbased hydrologic modeling. Similarly, the National Land Cover Database (NLCD) and Global Land Cover

(UMDGLC) provide distributed vegetation information for the models. As in the previous case, the radiation, leaf-area, and canopy parameters serve as initial or a-priori parameters, and calibration is required to improve representations of the soils and vegetation information [Duan et al.(2006)]. National products for gridded climate, hydrogeologic data and digital elevation models also require processing as part of the modeling process. The scale of geospatial/temporal data necessary for modeling, and the need to simulate nonlinear, multi-state, hydrological processes over large areas led to simplified strategies such as trial-and-error techniques for parameter estimation with limited success[Ivanov et al.(2004), Qu & Duffy(2007), Du et al.(2007), Li et al.(2008), Takeuchi et al.(2010), Shih & Yeh(2011)]. In recent decades, computational methods have led to improved model-calibration frameworks, including Monte Carlo Analysis, Genetic Algorithm (GA), and Evolutionary Strategy (ES) [Tolson & Shoemaker(2007), Nicklow et al.(2010), Reed et al.(2012)]. Comparative studies between different optimization algorithms have advanced the understanding of complex models and parameter properties and benefited watershed modeling applications [Tolson & Shoemaker(2007)]. Clearly, these approaches still require very large computational resources, and the need for efficient parameterization is still an unsolved problem. Another component of any efficient strategy is a partition calibration processes [Lei et al.(2011)]. However, relatively little work on partition calibration processes has been applied to parameter estimation of physics-based, fully-coupled, distributed hydrologic models.

In this study, the physics-based, fully-coupled, distributed hydrologic model PIHM (Penn State Integrated Hydrologic Model) is analyzed by a sensitivity-based Partition Calibration Strategy (PCS) for efficient model parameter optimization. The efficiency of PCS is gained by partitioning data into groups with distinct sensitivities to model processes. The physical model uses a semi-discrete FVM (Finite Volume Method) to form the coupled equations, which include: Noah_LASM [Chen & Dudhia(2001)], 2-D overland flow, 1-D unsaturated flow and 2-D subsurface flow to streams. PIHM is an open-source, distributed hydrologic model (<http://www.pihm.psu.edu>), and has been applied to multi-scale hydrologic settings, with multiple versions adapted to hydrodynamics, transport, and landscape evolution modeling [Qu & Duffy(2007), Li & Duffy(2011)]. The

objective of this chapter is to introduce a new calibration framework for PIHM that is compatible with the intensive computational requirements of integrated catchment modeling and to future test the robustness of the method.

2.2 Model Formulation and National Data Set Support

2.2.1 PIHM

PIHM is a physics-based, fully-coupled, distributed hydrologic model. It simulates interception, throughfall, infiltration, recharge, evapotranspiration, surface runoff, groundwater flow, and channel routing in a fully coupled scheme. The spatial domain decomposition, which is represented as a quality triangular mesh, uses the triangle [Shewchuk(1997)] that is an implementation of the Delaunay triangulation algorithm. The resolution of spatial domain decomposition can be varied according to the geomorphological or hydrological characteristics of the watershed. The spatial domain decomposition can be constrained by hydrologic features such as observation point, boundary conditions, etc. [Kumar et al.(2009)]. Hydrologic equations that include partial differential equations (PDEs) for overland flow, subsurface flow, and channel routing, and ordinary differential equations (ODEs) for interception, infiltration, recharge, and evapotranspiration (ET) are assembled over each control volume. PDEs are discretized to ODEs using the finite volume method. This results in an identical local system of ODEs assigned to each model grid. The local system is referred to as the kernel. The local system of ODEs is assembled over the entire model domain to form the global system of ODEs and is solved using the SUNDIALS solver software [Cohen & Hindmarsh(1996)]. For a detailed description of the modeling approach and formulation, the reader should consult [Qu & Duffy(2007)] and [Kumar(2009)].

2.2.2 Hydrological Processes

A short introduction on the hydrological processes and corresponding model parameters is provided in this section. A brief illustration of the model coupling strategy is shown in Figure 2.2, and the main equation of PIHM is listed in Table 2.1.

Interception: Due to the vegetation and canopy cover, a fraction of precipitation is intercepted and temporally stored until it returns to the atmosphere as evaporation or passes through the canopy as throughfall or stemflow. PIHM uses a bucket model to describe the interception process:

$$\left. \frac{dh_{0I}}{dt} \right|_m = p_t - q^+ - e + \sum_{j=1}^3 q_j^s \Big|_m \quad (2.1)$$

where h_{0I} is the vegetation interception storage, P_v is the total precipitation, E_c is the evaporation from canopy interception, and P_t is the throughfall and stemflow. Subscript m represents the spatial grid, ranging from 1 to the total number of triangles.

Snow melt: The dynamic snowmelt conservation equation is given by:

$$\left. \frac{dh_{0S}}{dt} \right|_m = P_s - E_{snow} - \Delta w \Big|_m \quad (2.2)$$

where h_{0S} is the snow water equivalent storage, P_s is the solid precipitation water equivalent, E_{snow} is the evaporation directly from snow cover, and Δw is snow-melting rate, which is determined by a temperature index method.

Overland Flow: The governing equations for surface flow are the 2-D estimation of St. Venant equations [Qu & Duffy(2007)]. The equations are approximated in semidiscrete form:

$$\left. \frac{dh_1}{dt} \right|_m = p_t - q^+ - e + \sum_{j=1}^3 q_j^s \Big|_m \quad (2.3)$$

where h_1 is the shallow water depth above the ground surface and q_j^s is the normalized lateral flow rate from element to its neighbor j . The terms p_t , q^+ , and e are throughfall, infiltration, and evaporation, respectively. Subscript m represents the spatial grid, ranging from 1 to the total number of triangles.

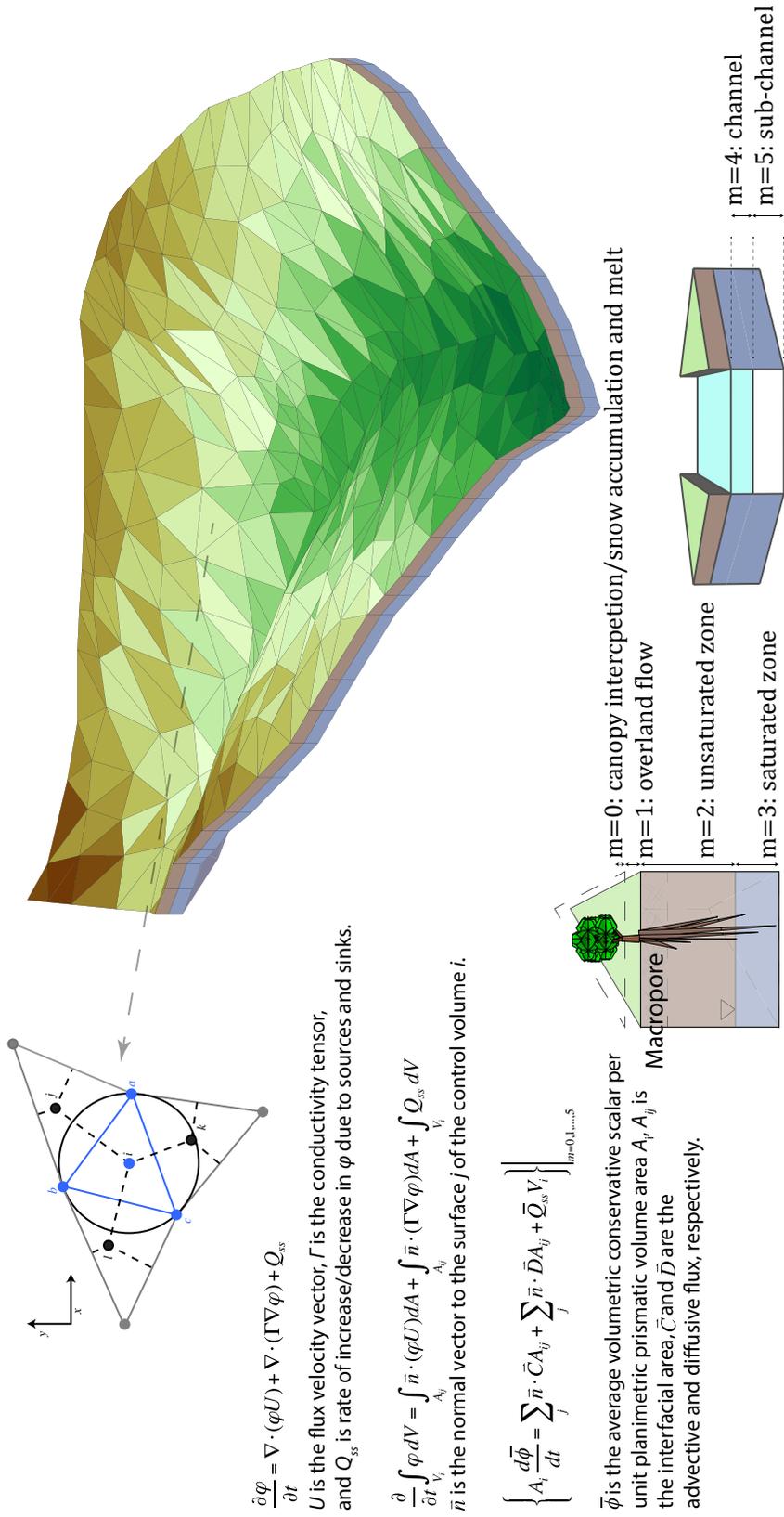


Figure 2.1. The PIHM representation of hydrological processes and model coupling strategy. The triangle mesh represents the spatial domain decomposition of a real watershed: Shale Hills Watershed in central Pennsylvania. The equations demonstrate a semidiscrete, finite-volume formulation for each computational grid. The equations are applied to each process and then assembled over the whole domain. Related details can be found in [Qu & Duffy(2007)] and [Kumar(2009)].

Process	Governing equationmodel	Original governing equation	Semi-discrete form
Interception	Bucket model	$\frac{dh}{dt} = P_v - E_c - P_t$	$\frac{dh_{0I}}{dt} = P_v - E_c - P_t$
Snowmelt	Temperature index model	$\frac{dh}{dt} = P_s - E_{snow} - \Delta w$	$\frac{dh_{0S}}{dt} = P_s - E_{snow} - \Delta w$
Evapotranspiration	Penman-Monteith approach	$ET_0 = \frac{\Delta(R_n - G) + \rho_a C_p \frac{(\varepsilon_s - \varepsilon_a)}{r_a}}{\Delta + \gamma(1 + \frac{r_s}{r_a})}$	$ET_0 = \frac{\Delta(R_n - G) + \rho_a C_p \frac{(\varepsilon_s - \varepsilon_a)}{r_a}}{\Delta + \gamma(1 + \frac{r_s}{r_a})}$
Overland flow	St. Venant equation	$\frac{\partial h}{\partial t} + \frac{\partial(uh)}{\partial x} + \frac{\partial vh}{\partial y} = q$	$\frac{dh_1}{dt} = p_t - q^+ - e + \sum_{j=1}^3 q_j^s$
Unsaturated flow	Richards equation	$C(\Psi) \frac{\partial \Psi}{\partial t} = \nabla \cdot K(\Psi) \nabla(\Psi + Z)$	$\theta_s \frac{dh_2}{dt} = q^+ - q^0$
Groundwater flow	Richards equation		$\theta_s \frac{dh_g}{dt} = q^0 + \sum_{j=1}^3 q_j^g$
Channel flow	St. Venant equation	$\frac{\partial h}{\partial t} + \frac{\partial(uh)}{\partial x} = q$	$\frac{dh_{4,5}}{dt} = p - e + \sum_{j=1}^2 (q_l^s + q_l^g) + q_{in}^c - q_{out}^c$

Table 2.1. Main equations of PIHM

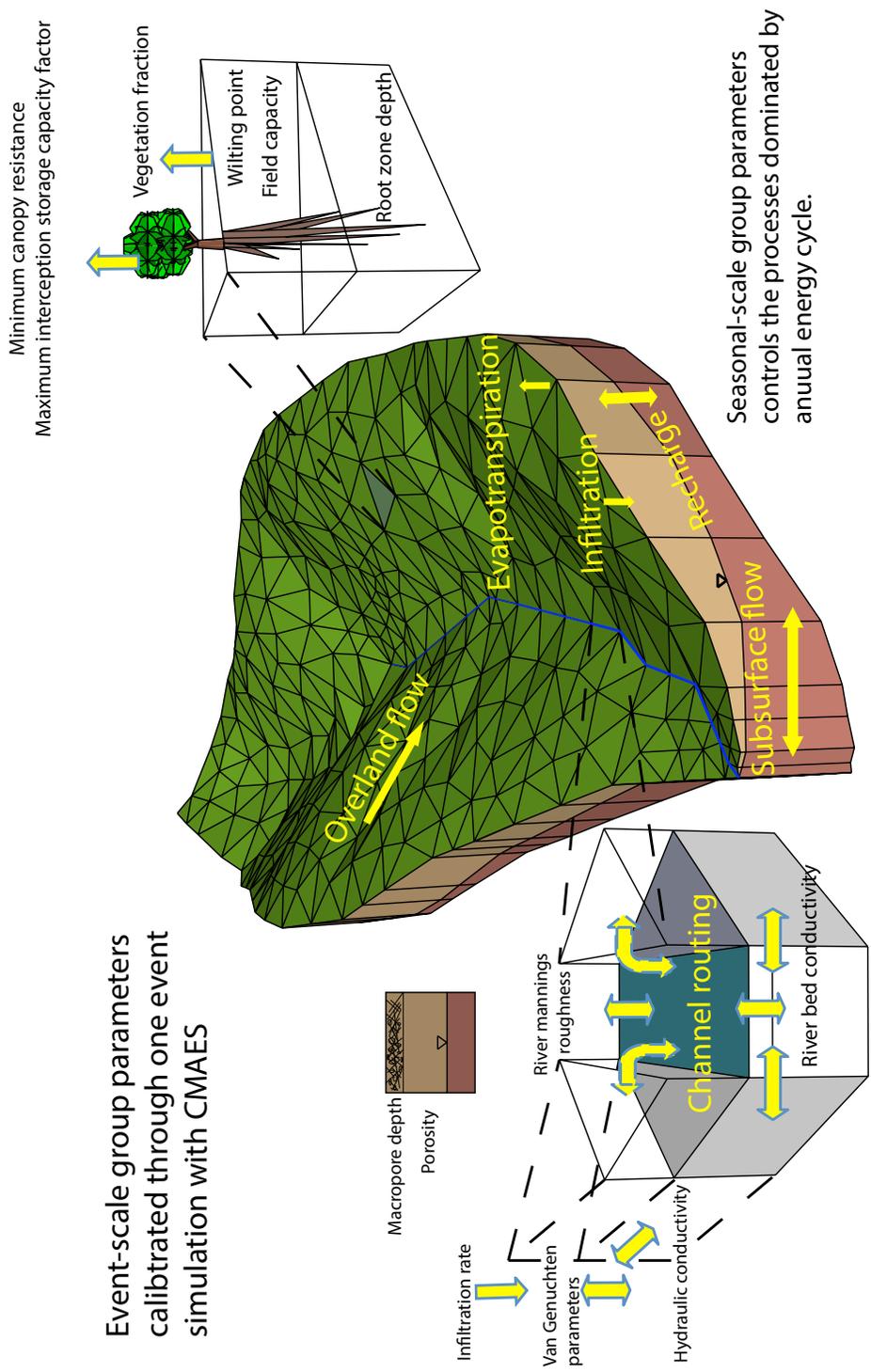


Figure 2.2. PIHM processes and domain decomposition for Shale Hills. The triangles represent the watershed domain, linear segments in blue represent channels.

Subsurface Flow: The model assumes that each subsurface layer can have both unsaturated and saturated storage components [Qu & Duffy(2007)]. With the estimation (only vertical flow in unsaturated zone) and integration, the balance equations are formed:

$$\theta_s \frac{dh_2}{dt} \Big|_m = q^+ - q^0 \Big|_m \quad (2.4)$$

$$\theta_s \frac{dh_3}{dt} \Big|_m = q^0 + \sum_{j=1}^3 q_j^g \Big|_m \quad (2.5)$$

where θ_s is the moisture content, h_2 is the unsaturated storage depth, h_3 is the groundwater depth, q^0 is flux between unsaturated-saturated zone [Kumar(2009)], and q_j^g is the normalized lateral groundwater flow rate from element i to its neighbor j . Here [van Genuchten(1980)] formulation was used in discretized form to improve the computation performance [Qu & Duffy(2007)].

Channel Routing: The same semi-discrete finite volume approach is applied to the 1-D estimation of Saint Venant equations [Qu & Duffy(2007)]:

$$\frac{dh_{4,5}}{dt} \Big|_k = p - e + \sum_{j=1}^2 (q_l^s + q_l^g) + q_{in}^c - q_{out}^c \Big|_k \quad (2.6)$$

where $h_{4,5}$ is the depth of water in the channel and beneath the channel, p and e are precipitation and evaporation from the channel segment respectively, and q_l^s and q_l^g are the lateral surface flow and groundwater interaction with the channel respectively from each side of the channel. The upstream and downstream flow for each channel segments are q_{in}^c and q_{out}^c respectively. Subscript k represents the channel segment, ranging from 1 to the total number of channel segments.

Evapotranspiration (ET): The total evaporation is the sum of evaporation from canopy interception (e_c), transpiration from vegetation (e_t), and evaporation from soil (e_s). The Penman-Monteith approach is used for the calculation of the potential evaporation:

$$ET_0 = \frac{\Delta(R_n - G) + \rho_a C_p \frac{(\varepsilon_s - \varepsilon_a)}{r_a}}{\Delta + \gamma(1 + \frac{r_s}{r_a})} \quad (2.7)$$

Here ET_0 refers to potential evapotranspiration, R_n is net radiation at the vegetation surface, G is soil heat flux density, $\varepsilon_s - \varepsilon_a$ represents the air vapor pressure deficit, and ρ_a is the air density, and C_p is the specific heat of the air. Δ is the slope of the saturation vapor pressure-temperature relationship, γ is the psychrometric constant, and r_s, r_a are the surface and aerodynamic resistances. The ET calculation equations are adapted from Noah-LSM [Chen & Dudhia(2001)] for computing the actual evapotranspiration:

$$e_c = \sigma_f e_p \left(\frac{W_c}{S}\right)^{0.5} \quad (2.8)$$

$$e_t = \sigma_f e_p B_c \left[1 - \left(\frac{W_c}{S}\right)^{0.5}\right] \theta \quad (2.9)$$

$$e_s = (1 - \sigma_f) \beta e_p \quad (2.10)$$

where σ_f refers to vegetation fraction, W_c is the intercepted canopy water content, S is the maximum canopy capacity, B_c is a function of canopy resistance, and β is calculated by:

$$\beta = \frac{\theta - \theta_w}{\theta_{ref} - \theta_w} \quad (2.11)$$

where θ_{ref} is field capacity and θ_w is wilting point.

2.2.3 Watershed Description

The Susquehanna-Shale Hills Critical Zone Observatory (SSHCZO) is funded by the National Science Foundation under the Critical Zone Observatory Program (CZO). The site is located in the Ridge-and-Valley physiographic province in Central Pennsylvania. The upland site has an area of 8 hectares. The SSHCZO has been the focus of several interdisciplinary studies, in biogeochemistry, ecology, geomorphology, meteorology, hydrology, and pedology [Brantley(2008)], which provide a unique and well-documented watershed to benefit cross-disciplinary science [Anderson et al.(2008)] and intensive data for model testing. The site has a long history of experimental research going back to the the 1970s when a series of irrigation experiments were conducted at this site by the Forest Hydrology group at The Pennsylvania State University in 1974 [Lynch(1976)]. The experiment provided insight into the physical mechanisms of runoff and streamflow

generation of SSHCZO and revealed the effects of the antecedent soil moisture on the runoff peak and timing. The hydrological processes involving the irrigation experiment were reproduced by a numerical fully-coupled, physics-based model [Qu & Duffy(2007)]. Spatial hydropedologic heterogeneity of SSHCZO was studied by year-round soil moisture monitoring across the watershed [Lin(2006)]. The SSHCZO is a rapid erosion-cut, deep V-shaped valley watershed with an underlying Rose Hill shale geology layer [Lynch(1976)]. Within the forested watershed, an ephemeral stream flows into Shavers Creek (185 km²), which eventually discharges into the Juniata River, the second largest tributary of the Susquehanna River Basin (Figure 2.3).

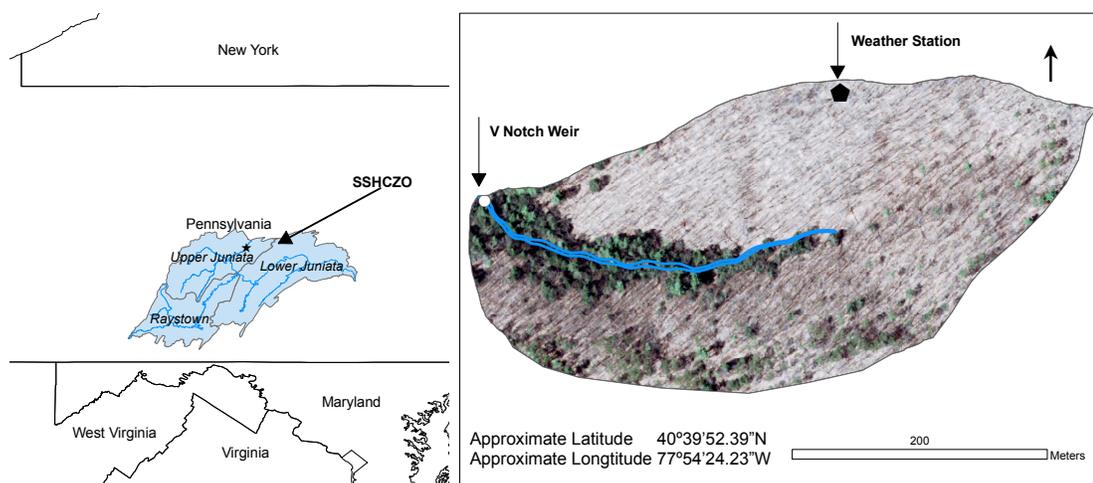


Figure 2.3. Location of SSHCZO.

The left figure shows the position of SSHCZO in the Juniata River watershed, and the right figure shows the stream and shape of SSHCZO.

2.2.4 National Data Support

Data support is one of the most important issues related to the model applications, and the PIHM model development uses national data products (Table 2.2). GIS tools have been developed to process the national geospatial/temporal data into model parameters, and the reader is referred to [Bhatt(2012)] for details of this process. Additional data for calibration and validation of such models is desirable and is available from selective research networks: National Critic

Zone Observatories (CZOs), Long Term Ecological Research (LTER) Network, US Forest Service Watershed Condition Framework, and USDA Experimental Watersheds. With respect to river-basin scale, predicting and managing the national surface and groundwater resources in the United States requires seamless and fast access to the essential geo-spatial/geo-temporal data necessary for physics-based numerical models as well as access to data fusion tools. A basic cyber infrastructure is currently available for the necessary data for all HUC-12 watersheds in the continental US [Leonard & Duffy(2013)] The site (www.hydroterre.psu.edu) includes national watershed, soils, climate, digital elevation data as well as land cover and river basin hydrographic data sets. The current prototype includes accessibility and scalability of virtualized services supporting essential national geospatial/temporal data sufficient for numerical watersheds. The national data products include 30m DEM, the National Land Data Assimilation System (NLDAS-2) as atmospheric forcing, soils survey (SSURGO, STATSGO), and land cover (NLCD 2001, UMDGLC). The data have been processed for the HUC-12 (12 digit Hydrologic Unit Codes) watershed/stream network, and soil hydraulic properties. In addition, PIHM teams are coordinating this effort with NSF-supported communities (CZO, LTER, CUAHSI) and EU colleagues (Soiltrec) with a goal of interoperable data sets and watershed models for the essential terrestrial geospatial data necessary for supporting numerical watershed prediction at high resolution (20-100m). In this study, raw data of SSHCZO was demonstrated to understand the workflow from the real watershed to calibrated numerical modeling.

2.2.5 A-Priori Parameters

The national data provides consistency in model application across any watershed in the United States. However, incorporating the physical data layers into the hydrologic model requires intensive data development, topology definitions, and projection of parameters to individual model grids. A tightly-coupled GIS interface to PIHM called PIHMgis, which is an open-source, platform independent, and extensible framework, enables synthesis of model parameters using national data

Watershed	Reference	Data support	Modeling scale
V-catchment	[diGiammarco et al.(1996)]	Numerical experiment	Numerical scale
Shale Hills	[Yu et al.(2013)]	National CZO	Catchment scale
Lysina	[Yu et al.(2014b)]	LTER network	Catchment scale
Caspar Creek	[Carr et al.(2013)]	US Forest Service Station	Catchment scale
Walnut Gulch	[Renard et al.(2008)]	USDA Watershed Research	Catchment scale
Experimental Watershed			
Little Juniata River	[Yu et al.(2013)]	National Hydrography Dataset	HUC-12 → national
Juniata River	[Bhatt et al.(2013)]	HydroTerre [Leonard & Duffy(2013)]	HUC-12 → national

Table 2.2. Data support for physics-based hydrologic modeling studies

products [Bhatt et al.(2006)]. The tight coupling between GIS and the model is achieved by developing a shared data-model and hydrologic-model data structure with a-priori parameter estimates from the national dataset for PIHM simulation. Soil hydraulic properties are obtained by applying pedotransfer functions (PTF) [Wösten et al.(2001)] to textural classes in the national soils data, SSURGO, and STATSGO. Each soil class contains soil texture data with the proportion of sand, silt, clay, and organic matter and the bulk density. PTF is applied to generate hydraulic properties including conductivity, porosity, and van Genuchten water-retention parameters [van Genuchten(1980)]. Estimates of vegetation parameters are available from the NLDAS that provide season dynamics of each land-cover type. Vegetation parameters such as seasonal LAI, stomatal resistance, and surface roughness are projected into the model according to the spatial vegetation class map of NLCD or UMDGLC.

2.3 Partition Calibration Strategy

In this section, an efficient calibration strategy is developed with an analysis of parameter sensitivity to hydrological processes. Sensitive parameters are partitioned into 2 groups, which generally are sensitive to hydrologic events and to seasonal response variables (Table 2.3). The calibration strategy is then used to estimate parameters through a sequential process with respect to each group.

Parameter	Hydrological processes	Group
Matrix conductivity (horizontal)	Subsurface flow	EG
Matrix conductivity (vertical)	Subsurface flow	EG
Macropore conductivity (horizontal)	Subsurface flow	EG
Macropore conductivity (vertical)	Subsurface flow	EG
Infiltration rate	Infiltration	EG
Macropore depth	Subsurface flow	EG
Porosity	Subsurface flow, recharge	EG
Van Genuchten parameter α	Subsurface flow, recharge	EG
Van Genuchten parameter β	Subsurface flow, recharge	EG
River Mannings roughness	Subsurface flow, recharge	EG
River bed conductivity (horizontal)	Channel routing	EG
River bed conductivity (vertical)	Channel routing	EG
Root zone depth	Channel routing	EG
Vegetation fraction	Transpiration, evaporation	SG
Field capacity	Transpiration, evaporation	SG
Wilting point	Evaporation	SG
Maximum interception storage capacity factor	Evaporation	SG
Minimum canopy resistance	Interception loss	SG
	Transpiration	SG

Table 2.3. Calibration parameters, corresponding hydrologic processes and their typical behavior as event-scale parameters group (EG) or seasonal-scale parameters group (SG)

2.3.1 Monte Carlo Sensitivity Analysis

The objective function based in root-mean-square error (RMSE) was selected to analyze the goodness of fit between observations and simulations. The SSHCZO and a simple hydrological event (precipitation event: from 2009 June 15th to 2009 June 30th) were selected for Monte Carlo experiments. Because of the large number of non-behavioral simulations (parameter set leads to convergence failure of the solver), 100,000 Monte Carlo samples were tested, and only 7,000 simulations were selected as behavioral results. Regional sensitivity analysis (RAS) was applied for each parameter [Demaria et al.(2007)]. The Monte Carlo simulations were ranked according to the value of their objective function and then equally divided into ten bins. The first bin contained the best 10% of the behavioral simulations, the second bin the next best 10%, and so forth. For each bin, the normalized objective function values were plotted as a cumulative distribution function of the parameter value. The curves for each parameter are listed in each panel in Figure 2.4. A straight one-to-one line suggested the insensitivity of the parameter to the event-scale hydrologic response, whereas different curves in each bin suggested a high sensitivity of the parameter.

The most sensitive parameters are hydraulic conductivities and soil porosity, which is also found by [Shi et al.(2014a)]. The van Genuchten parameters presented variation between each cumulative curve, indicating their sensitivity to the event-scale hydrological processes. The root zone depth, field capacity, wilting point, minimum canopy resistance, and vegetation presented straight lines, suggesting that the response of a single hydrological event is insensitive to their values. As a result of the sensitivity analysis, the total parameter space was partitioned into two groups: the EG, or hydrologic "Event Group," and SG, or the "Seasonal Group," parameters (Table 2.2).

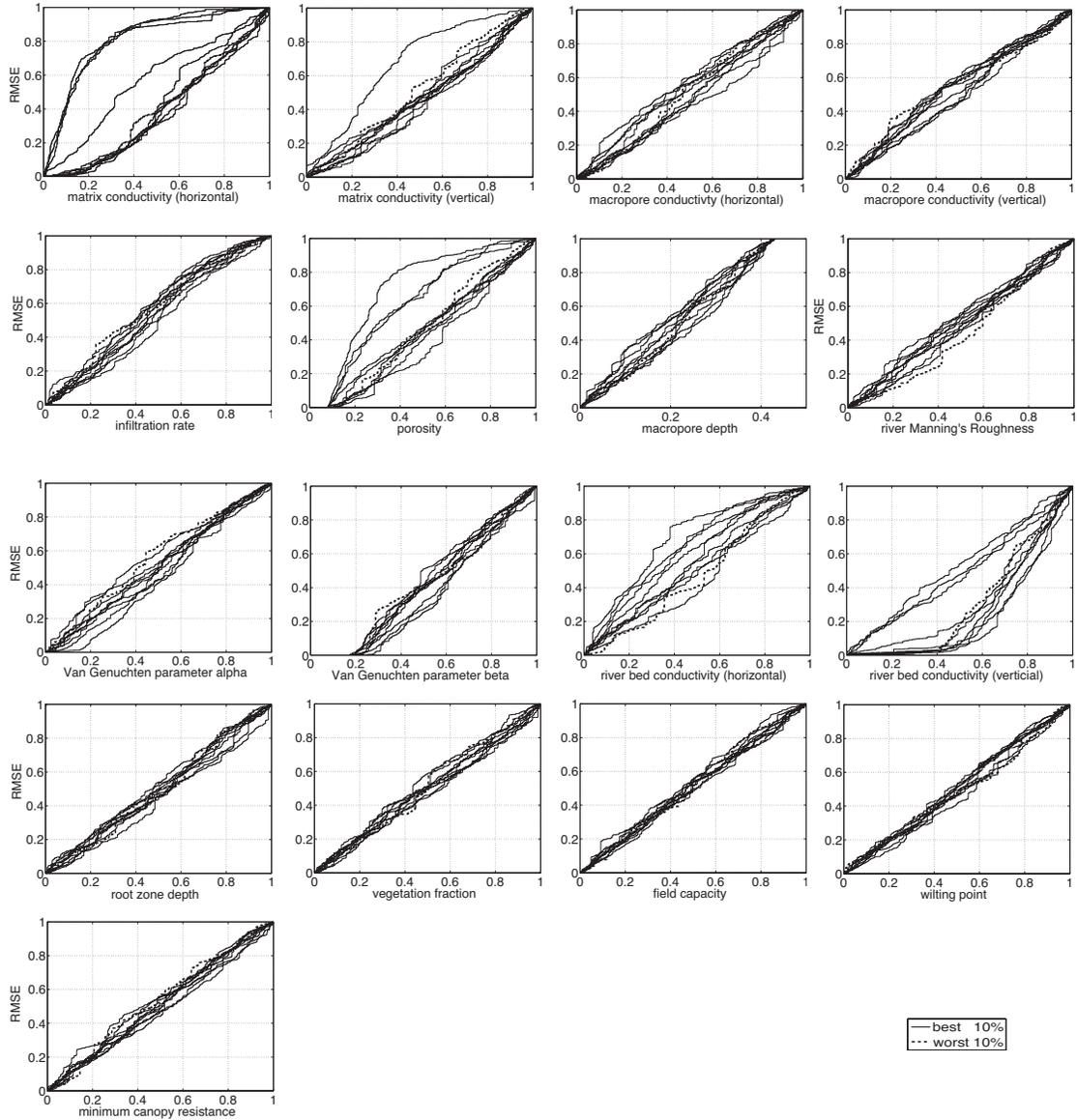


Figure 2.4. Results of regional sensitivity analysis (RSA) for parameters in PIHM. A straight one-to-one line suggested the insensitivity of the parameter to the event-scale hydrologic response, whereas different curves in each bin suggested the sensitivity of the parameter.

2.3.2 Calibration of EG Parameters

The computational cost of PIHM simulations necessitates an assessment of the minimum number of Monte Carlo runs for EG parameter calibration. This is especially important for large-scale watershed modeling. Here we tested the

CMA-ES [Hansen(2006)]. The CMA-ES was initially applied in groundwater remediation design where it was shown to be an effective strategy [Bayer & Finkel(2004)]. In particular, CMA-ES is suited for complex problems without any problem-dependent parameters or for problems with a modest number of objective function evaluations [Hansen(2006)]. Comparative studies have concluded that CMA-ES is an efficient algorithm and less computationally expensive than other optimization algorithms [Hansen et al.(2003), Li & Heinemann(2007)]. According to [Bayer & Finkel(2004)], CMA-ES has the advantage in terms of efficiency and outperforms the Simple Genetic Algorithms (SGAs) with problem dimensions and the optimality of the objective function value. Therefore, CMA-ES could be an ideal fit for the calibration of computationally intensive models.

The CMA-ES is an evolutionary algorithm for difficult non-linear, non-convex optimization problems in a continuous domain. CMA-ES is a rank-based (η , λ) evolution strategy in which the best of the offspring form the next parent generation. It generates new population members by sampling from a probability distribution that is constructed during the optimization process. I provide a short overview of CMA-ES. The source code and detailed description are available online (<https://www.lri.fr/~hansen/cmaesintro.html>).

When I optimize the fitness $f : \mathbf{R}^n \rightarrow \mathbf{R}$, where the n is the dimension of the problem, CMA-ES follows a randomized black box search scenario:

1. Initialize the distribution parameters θ^0 .
2. For generation $g=0, 1, 2, \dots$:
 - a. Sample λ independent points from distribution $P(x - \theta^{(g)}) = x_1, \dots, x_\lambda$.
 - b. Evaluate the sample x_1, \dots, x_λ on f .
 - c. Update parameters $\theta^{(g+1)}$ based on the best performers $x_1, \dots, x_\eta (f(x_1) \leq f(x_2) \leq f(x_\eta))$.

Break, if termination criterion met.

In CMA-ES, the search distribution to be estimated is a multivariate normal distribution $N(m, \delta^2 C)$. m is the mean of the current population. The covariance matrix $\delta^2 C$ is used to guide the search towards optimized parameter space. I use the recommended default equation to determine the population size.

$$\lambda = 4 + \lfloor 3 \ln(n) \rfloor \quad \text{and} \quad \mu = \left\lfloor \frac{\lambda}{4} \right\rfloor \quad (2.12)$$

Parameter	Scale	Function
Conductivity	(0.01,100)	$10^{(0.5-x) \times 4}$
Porosity	(0,0.7)	$x \times 0.7$
Macropore depth	(0, <i>bedrockdepth</i>)	$x \times \textit{bedrockdepth}$
Mannings roughness	(0.1,10)	$10^{(0.5-x) \times 2}$
Van Genuchten parameter (α, β)	(0,5)	$x \times 5$

Table 2.4. Rescaling function of real parameters used in the calibration algorithm x represents the generated parameters, which range from 0 to 1. The rescaling functions project the generated parameters to physically acceptable parameter values for PIHM.

where n is the search space dimension (here the number of EG parameters to be calibrated in PIHM is 12).

The CMA-ES was implemented using Message Passing Interface (MPI), necessary for distributed modeling with PIHM. The main processor is responsible for the parameter population, recombination, and selection. At the beginning of each generation, the parameters are sent to different processors. Each processor rescales the parameters and input to the independent PIHM simulation. The rescaling function listed in Table 2.4 ensured the physical meaning of each parameter. The main processor gathers the objective function values after all the processors finish the PIHM simulation and objective function evaluation. The fitness function (cost function or objective function) is defined as the root mean square error (RMSE) for predicted versus actual observations. The CMA-ES stops when the defined maximum number of generations is reached.

2.3.3 Calibration of SG Parameters

A series of studies have been conducted for parameter sensitivity of Land Surface Model [Kato et al.(2007), Prihodko et al.(2008), Rosero et al.(2010)], which provide guidance for calibration of SG parameters. Root zone depth, which controls the availability of water for transpiration, is critical for transpiration. Vegetation fraction can balance the transpiration between evaporation. Wilting point controls the lower limit of soil evaporation, while higher field capacity values tend to decrease soil evaporation [Rosero et al.(2010)]. Due to near-linear responses, these SG parameters could be calibrated according to previous studies of Land Surface

Models.

2.4 Results and Discussion

2.4.1 Results and Validation

The unstructured mesh was constructed from a 0.5m LiDAR digital elevation model to construct the unstructured mesh. Soil and vegetation parameters were obtained from SSURGO and NLCD, as described earlier. The spatial soil class map was updated with field survey [Lin(2006)]. The domain of 8.4 was spatially discretized into 535 triangles and 20 linear segments to represent the stream. The hourly observation from the year 2009 was used for the calibration, and a validation period was chosen from a previous experiment [Lynch(1976)]. The same event in Monte Carlo experiment was selected as the calibration period for the optimization of the EG parameters.

The calibration proceeds sequentially by first estimating EG parameters, and then the SG parameters were determined by trial and error that estimated the annual water budget after the EG parameter calibration. In this sequence, the calibration process provided the most consistent calibration results. After calibration, running the model for the entire year demonstrated that large precipitation events during the summer dry period (July-August) and the fall wet period (September-October) were well simulated (Figure 2.5). Note that the antecedent moisture (soil moisture and shallow groundwater) has a strong affect on the flood peak and this is well represented in the model results. The NashSutcliffe model efficiency coefficient (NSE) [Nash & Sutcliffe(1970)] of 3-hour streamflow was 0.88 in 2009.

The validation was performed using the artificial irrigation experiment [Lynch(1976), Qu & Duffy(2007)], with the historical measurements of 15-minute precipitation, temperature data and the artificial irrigation rates (Figure 2.6). The parameters estimated from 2009 were adequate for the validation period even though the validation period was several decades earlier. The departure of model results from observed streamflow is likely explained by the lack of

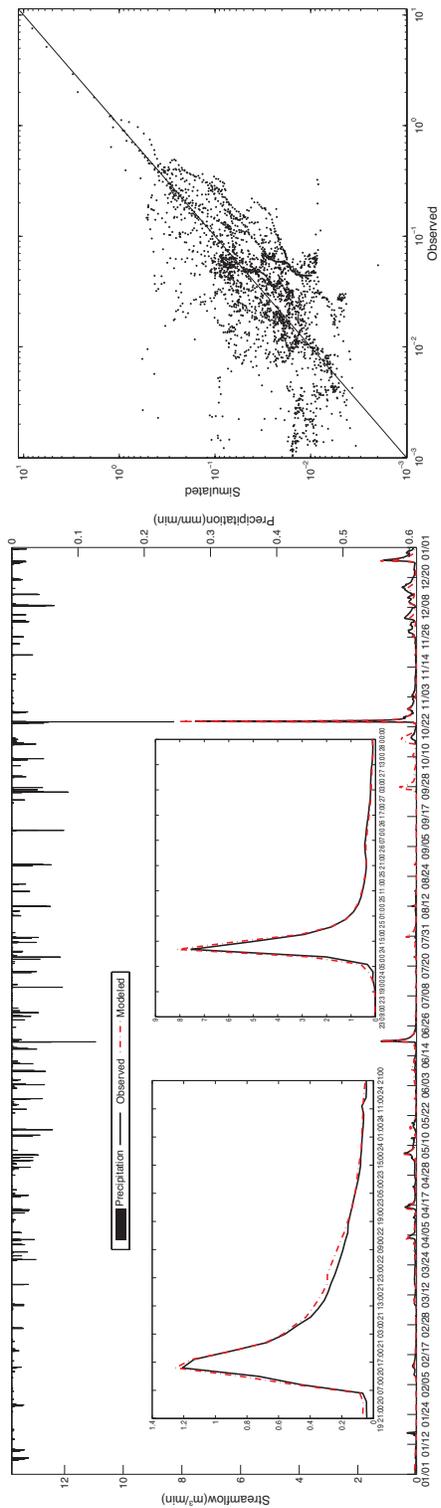


Figure 2.5. Observed and simulated streamflow for the year 2009 at the outlet of SSHCZO.

historical meteorological data (relative humidity, wind speed, and solar radiation) and vegetation dynamic. The flooding events were underestimated in July and overestimated in August, which suggests the seasonal Leaf Area Index pattern used in the model could not apply to the growth stage of the forest 30 years earlier(1970s). The research of biogeochemical modeling will evaluate biomass forest growth for long-term hydrologic reanalysis (Chapter 5).

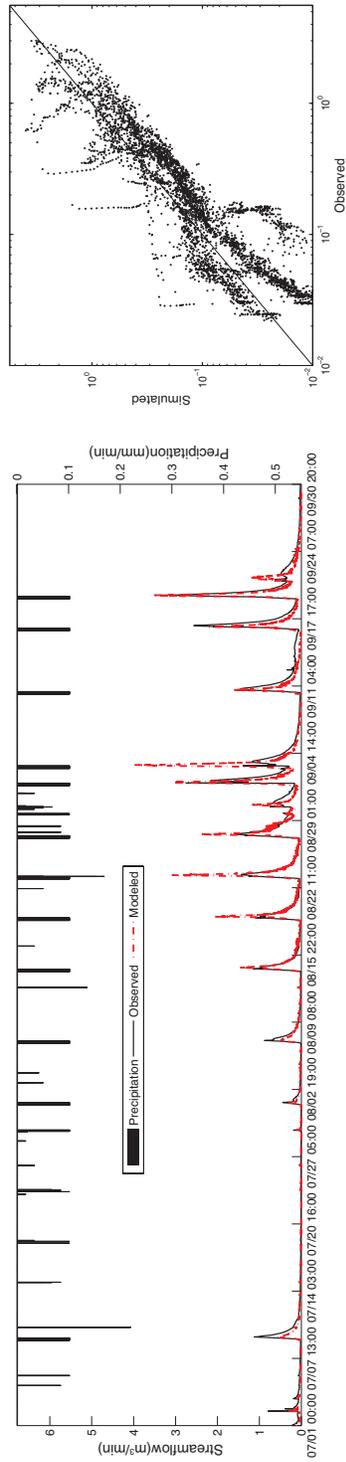


Figure 2.6. Validation performance for 1974 experiment at SSHCZO.

From July to September 1974, a series of equal artificial rainfall events (0.64 cm/h for 6 hours) were applied to the entire watershed. The 2009-data calibrated parameters reproduced the hydrological responses in 1974.

2.4.2 Strengths and Limitations of PCS

Partition calibration generally attempts to decompose the vector of unknown model parameters into groups that in some way allow for greater efficiency in the estimation process. The question is, how does one define the groups? For example, certain parameters can be estimated using baseflow relaxation periods during cool seasons or when vegetation is dormant. The strategy in this case partitioned the watershed into sub-basins and allowed for efficient parameter estimation.

In this study, the parameter space was partitioned according process time scales in PIHM. Integrated models such as PIHM implicitly couple multiple hydrological time scales, ranging from minutes to decades. Our approach was to use Monte Carlo simulation to assess the sensitivity and time scales of all parameters in the model. These simulations suggested a natural grouping of parameters in two categories: 1) EG made up of hydrologic parameters; and 2) SG of energy-related parameters. These 2 groups formed the basic partitioning strategy and greatly improved the efficiency of the estimation. The results of the three watershed applications suggested the event-scale calibration efficiently enabled automated parameter estimation for PIHM. We expect this approach is a potential solution for calibration of other complex, spatially distributed models.

One serious limitation is the lack of availability of a-priori estimates and ranges of parameters. At this stage PIHM team is developing a prototype data service for soils, hydrogeology, landuse/cover, topography, and climate from national data sources (<http://www.hydroterre.psu.edu> [Leonard & Duffy(2013)]). However, there is a national need for community support of high resolution geospatial data for distributed modeling in the future. Another limitation of the method is that the single objective approach we are using may not always be sufficient, especially for the integrated models with prediction of different hydrological processes. We expect that the approach can be easily extended to a weighted objective strategy or a multi-objective approach with multiple hydrological state and flux measurements. In Chapter 3, the calibration strategy is further developed for the case of multiple constraints.

2.5 Uncertainty Analysis

The evolution of the parameter value over time was plotted from three independent calibration runs of EG calibration (Figure 2.7). The figure suggested Van Genuchten parameter α converged very quickly to an optimum in the three runs. The infiltration rate, macropore depth, porosity, and Van Genuchten parameter β converged less quickly, and with greater error. The remaining parameters did not sensitive to the model result, which randomly fell into some values. The different results from three independent runs demonstrated remarkable uncertainties of the calibrated parameters. Monte Carlo analysis is often used for the quantification of uncertainty of environmental models [Liu & Gupta(2007), Bastina et al.(2013)]. Statistics of model output were found by generating many simulations according to certain distributions of parameter uncertainty. Here, we evaluated the posterior distribution of model parameters.

2.5.1 Posterior Parameter Evaluation of One-dimensional case

The Monte Carlo sampling of Van Genuchten parameter α was tested by fixing the rest of model parameters. The calibrated Van Genuchten parameter α was consistent with the Monte Carlo simulation results.

2.5.2 Posterior Parameter Evaluation of Two-dimensional Case

Further, the case of two parameters was considered. Both Van Genuchten parameter α and β were sampled at each Monte Carlo simulation. Local optimal zones (RMSE<60) were shown (Figure 2.9). This explained the different optimization results in the three independent calibration runs (Figure 2.7).

The above Monte Carlo analysis was supported by a high-speed parallel computing environment. It cost a huge volume of computation resources for the study at Shale Hills, a 0.08km² experimental catchment. However, for larger application

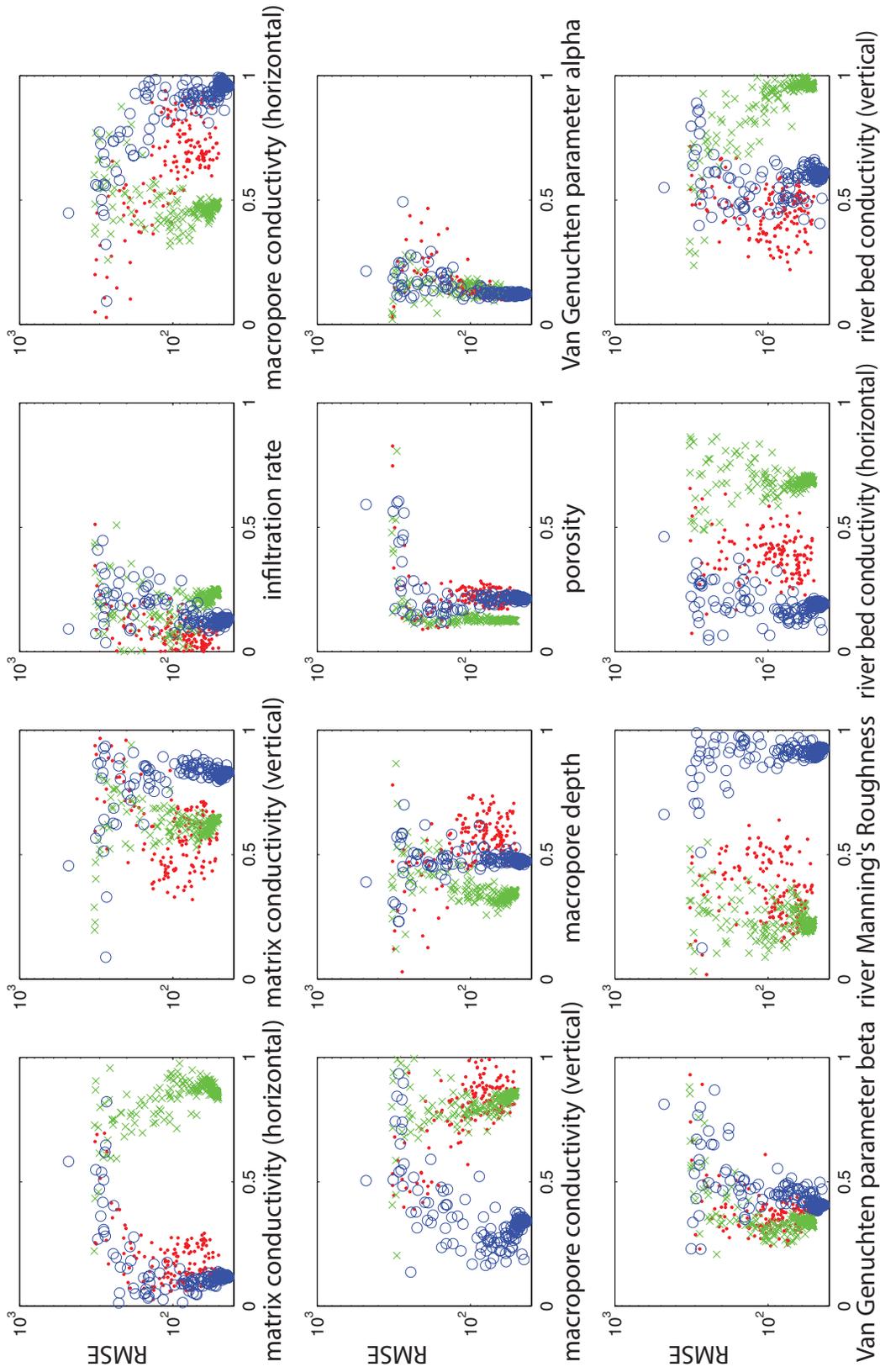


Figure 2.7. Evolution of parameter value over time during the EG calibration runs. There are there independent runs, which are represented by different marker type and color.

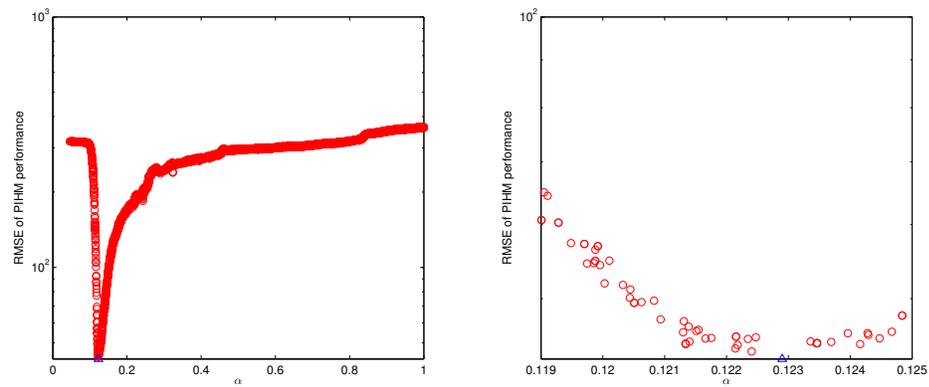


Figure 2.8. The Monte Carlo simulation of Van Genuchten parameter α . The \circ represents the PIHM RMSE of each Van Genuchten parameter α . The \triangle represents the calibrated Van Genuchten parameter α .

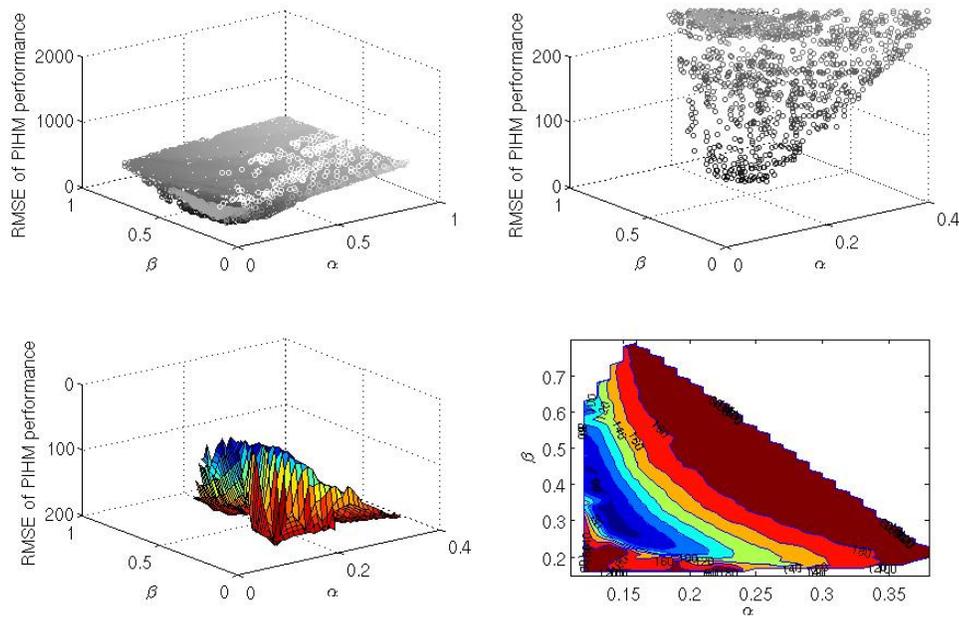


Figure 2.9. The Monte Carlo simulation of Van Genuchten parameter α and β .

and higher spatial resolution studies, it may not be possible to obtain enough samples for Monte Carlo-based approaches. To save the number of model runs, Bayesian emulation has been proved as an efficient approach for many complex environmental models [OHagan(2006), Bhat et al.(2010), Stone(2011)]. Future study should focus on the model uncertainty quantification of all other parameters

by Bayesian emulation.

2.6 Conclusion

Physics-based, distributed hydrologic models are data intensive, with a large computational cost. There is limited literature and tools available for parameter sensitivity analysis and calibration of such models. An important asset is the availability of national geospatial data products that can support distributed models. Generally, these are underused or have limited accessibility to support rapid implementation of watershed models. The study demonstrated that national data products could serve as an a-priori estimate of critical parameters of distributed model applications. Parameter sensitivity analysis partitioned the parameter space into two parts, which enabled the application of the CMA-ES optimization algorithm.

A general conclusion of this study is that a sensitivity-based, 2-scale partition for parameter calibration provides a useful way to isolate parameters in a coupled, multi-processes modeling approach. At the scale of first-order catchment, the vegetation parameters should be improved for long-term modeling. Validation over an extended period did not show evident degradation in the model performance, which reflects the robustness of the calibrated parameters.

The partition calibration strategy has been applied not only at SSHCZO, but also at the international CZO Lysina Catchment in the Czech Republic with success [Yu et al.(2014b)]. The corresponding tutorial webpage is available at http://www.organicdatascience.org/index.php/PIHM_calibration_using_evolutionary_algorithms.

”Informativeness” as a Quantitative Index of Weighted-objective Calibration for a Multi-state Distributed Hydrologic Model

3.1 Introduction

Distributed hydrologic models supported by geospatial information on soil, geology, topography, and vegetation data products can provide valuable information about the watershed hydrologic cycle. However numerical simulation of the multi-state, multi-process system is structurally complex and computationally intensive and typically involve a high level of parameterization [Foglia et al.(2009)]. These complications present a major difficulty in real watershed modeling applications, which require advanced expertise in modeling schemes and computation as well as experience in parameter calibration [Fang et al.(2013), Foglia et al.(2009), Hansen et al.(2013)].

Current research is attempting to advance our understanding of the spatial and temporal variability of watershed processes by implementing a new generation of monitoring networks, including real-time observations of hydrologic stores and fluxes such as soil moisture, water table depth, streamflow, and latent heat flux

[Kerkez et al.(2012), Morin et al.(2012), Reba et al.(2011), Zreda et al.(2012)]. The use of such geospatial and geo-temporal observations can be potentially beneficial for resolving multi-state water stores and fluxes across the watershed or river basin. However, constraining models with multi-state data increases the complexity of model validation and testing. Of course, the benefit of distributed models is that they preserve the important heterogeneities of the catchment and provide spatially variable storage and flux predictions rather than averaged watershed behavior. Where predictions of multiple states are required, an automated calibration procedure that includes multiple geo-spatial and geo-temporal observations, as well as a multi-objective calibration strategy, becomes increasingly important [Fang et al.(2013), Foglia et al.(2009), Khu et al.(2008), Li et al.(2010), Stisen et al.(2011), Rientjes et al.(2013), Hsie et al.(2014), Shi et al.(2014b)].

Another critical aspect of the model calibration is the selection of appropriate metrics in assessing model performances for a set of model parameters. Based on those model performance metrics, it is necessary to formulate objective function for use in conjunction with a suitable optimization algorithm. The particular model performance metrics provide quantitative assessment of the model's accuracy in reproducing catchment behavior and testing measurable criteria for evaluating the model performance [Krause et al.(2005)]. Recently, a number of papers have compared the performances and effectiveness of different assessment metrics [Dawson et al.(2007), Krause et al.(2005)]. Their conclusions did not ideally favor any particular assessment metric, but rather point out that each criteria has specific pros and cons, and that the model context is necessary for model calibration and evaluation [Krause et al.(2005)]. That is, the metrics are fundamental in developing a calibration strategy for a specific model and purpose.

The spatial heterogeneity of watershed hydrological cycles has been intensively documented using a network of observations and has been explained by the integrated modeling strategies. [Frei et al.(2009)] recognized the importance of structural aquifer heterogeneity leading to spatial patterns in river–aquifer exchange by a process-based distributed watershed model. [Naranjo et al.(2013)] demonstrated that heterogeneity and anisotropy have a strong influence on the mean residence time in the riparian zone. At the Shale Hills Watershed (<http://www.czo.psu.edu>), hydrological modeling studies have been carried out

in the scheme of a fully-coupled modeling [Li(2008), Qu & Duffy(2007)] supported by high-resolution geo-spatial, geo-temporal observations. The physics-based, fully-coupled, distributed hydrologic model Penn State Integrated Hydrologic Model (PIHM) was used to explain the effects of the antecedent soil moisture condition of the watershed on the runoff generation mechanism [Qu & Duffy(2007)]. [Yu et al.(2013)] implemented Covariance Matrix Adaptation Evolution Strategy (CMA-ES) to the calibration of PIHM to support hydrologic reanalysis of rainfall-runoff response of the watershed over the continuous simulation of three decades [Yu et al.(2014a)]. In order to understand the other catchment behaviors, for e.g. surface/subsurface flow interactions and watershed heterogeneity between riparian zone and upland, it is desirable to calibrate and validate model predictions against multiple observed hydrologic states and fluxes in the watershed.

In this context, the aim of this paper is to describe the development and implementation of an informativeness-based, weighted-objective calibration strategy to support distributed integrated hydrological modeling applications with model constraints against multiple measurements. In this study, the strategy has been used in conjunction with PIHM. First, the comparison of multiple efficiency metrics was conducted to determine their effectiveness on PIHM calibration. Second, correlations of PIHM performance at multiple observation sites, including streamflow, water table depth, and ET (evapotranspiration), were explored by Monte Carlo analysis. We define informativeness according to the performance correlation. Finally, the weighted objective function was formulated based on the informativeness from the previous step, and the model parameters were calibrated to simulate the multi-state watershed response. The goal is to reduce the uncertainty in simulated spatial dynamics of hydrologic processes constrained by multiple observation sites.

3.2 Proposed Methodology

Watershed models attempt to capture the response of the hydrologic processes and to predict hydrological state variables and fluxes with a coupling modeling framework. Multiple objectives can be synthesized corresponding to a number of

the available observations for e.g. soil moisture, water table depth, streamflow, and evapotranspiration. Although calibrating a numerical model that involves representation of hydrologic processes in a fully coupled scheme, a single objective function could be possibly satisfied by multiple unique combinations of parameters resulting in almost comparable model efficiency. A desirable outcome of the optimization process would be a unique combination of calibrated model parameters that satisfy multiple objectives pertaining to different hydrologic states and fluxes. Under ideal conditions, the objectives should be non-conflicting. However, due to errors in model structure and uncertainty in a-priori parameters and observed data, the objectives of fitting different hydrological responses simultaneously are usually not satisfied. Traditionally, optimization problems involving multiple and conflicting objectives have been solved by combining the objectives into a scalar function and, next, solving the equivalent single-optimization problem to identify the best-compromise solution [Efstratiadis & Koutsoyiannis(2010)]. However, the selection of weights and formulation of a weighted-objective function is usually subject to modelers experiences may lead to unsatisfactory solutions [Rientjes et al.(2013)]. Here, we define informativeness as the indicator of the physical weights of each target in the calibration. In this context, "informativeness" attempts to resolve both model structure error and observation uncertainty. The highest level of informativeness suggests that the fitting of that observed variable is the priority of calibration. Finally, I solve the multi-objective optimization problem for a physics-based, fully coupled, computationally intensive model with a three-step strategy informativeness-based weighted-objective calibration strategy. The proposed three-step strategy includes:

1. evaluate different metrics of goodness-of-fit criteria for the model;
2. calculate the "informativeness" weighting for each calibration target, which determines the weight in the formulation of objective function; and
3. formulate the weighted objective function and run the optimization algorithm to calibrate the model.

The calibration strategy is illustrated in Figure 3.1.

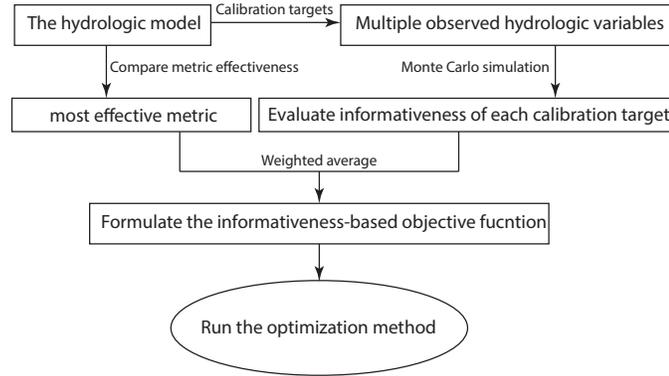


Figure 3.1. Flowchart of the informativeness-based weighted-objective calibration methodology.

3.2.1 Selection of Metrics According to Effectiveness

The first task is to select the metrics to evaluate the goodness-of-fit between simulated watershed response and observed data. Most studies select the metrics from statistical estimation theory, including the root mean square error (RMSE), NashSutcliffe model efficiency coefficient (NSE), and Pearson product-moment correlation coefficient (R) [van Werkhoven et al.(2008), van Werkhoven et al.(2009)]. Each metric has its own advantages and could be ideal in terms of fulfilling particular needs of an certain application [Dawson et al.(2007)]. RMSE records the level of overall agreement between the observed and modeled datasets; it is a non-negative metric that has no upper bound, where for a perfect model the result would be zero. Studies have demonstrated that RMSE is more sensitive to peaks values and higher magnitude events [Dawson et al.(2007)]. To constrain for low flow values, a logarithmical transformation of RMSE, log-RMSE, could be an appropriate solution [Wagener et al.(2009)]. NSE is sensitive to differences in both the observed and modeled means and variances [Dawson et al.(2007)]. NSE usually ranges from 0.0 (poor model) to 1.0 (perfect model), but negative scores are also permitted. R is an indicator used to evaluate the linear dependence of observed and simulated results, which demonstrate an overall agreement between the observed and modeled datasets. It is intended to range from -1 to 1. In some cases, conjunctive use of multiple evaluation metrics is necessary to better constrain the model performances [Li et al.(2010), Wagener et al.(2009)]. In this study, I implemented Covariance Matrix Adaptation Evolution Strategy (CMA-

ES) in the calibration of the physics-based, fully coupled, distributed hydrologic model Penn State Integrated Hydrologic Model (PIHM). We used the convergence rates to test the effectiveness of three metrics: RMSE, NSE, R. These metrics are defined:

$$RMSE = \sqrt{\frac{1}{t} \sum_{i=1}^t (O_i - P_i)^2} \quad (3.1)$$

$$NSE = 1 - \frac{\sum_{i=1}^t (O_i - P_i)^2}{\sum_{i=1}^t (O_i - \bar{O})^2} \quad (3.2)$$

$$R = \frac{\sum_{i=1}^t (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^t (O_i - \bar{O})^2} \sqrt{\sum_{i=1}^t (P_i - \bar{P})^2}} \quad (3.3)$$

where t is the total number of time steps in the calibration period, O is the observed value, and P is the predicted value at any time step.

I applied one metric in the CMA-ES based optimization procedure, and at the same time monitored the values of the other two metrics to evaluate the effectiveness of each metric in the model calibration. The metric with highest effectiveness would be selected in following study. For the details of the implementation of CMA-ES to PIHM parameter optimization, the reader should consult [Yu et al.(2013)].

3.2.2 Evaluate the Informativeness of Each Calibration Target

I considered 3 kinds of common hydrological observations: streamflow, water table depth, and latent heat flux. Streamflow is often the most informative variable in watershed modeling and it is the major target of most simulations. Streamflow gauges measure level (stage), average velocity and discharge at a stream cross-section. The water table depth below land surface or elevation is also a valuable variable, reflecting the saturated storage of groundwater, and is observed through piezometers. The latent heat flux, or the atmospheric flux of moisture from evaporation and transpiration, is directly observed through the eddy covariance instrument positioned above the canopy. I applied uniform random Monte Carlo

sampling (precisely 100,000 samples) to explore the parameter space. Following that the model performance were evaluated at each observation site using the most efficient metric. We plotted the objective function values between each pair of variables to assess the correlation of model performance at each observation site. High correlation implies the characteristic of non-confliction between the two objectives. We argue that high correlation suggests a favorable informativeness, which should be assigned a high weight in the aggregating function. Conversely, the less correlated variable suggests that the variable is less informative and should be assigned a low weight. The primary objective of most hydrological models is the prediction of streamflow. We assign the informativeness of streamflow as 1. The informativeness of other variables is defined by the correlation between the variable and streamflow.

3.2.3 Weighted Function for Multi-objective Optimization

A multi-objective calibration involves the simultaneous optimization of model residuals with respect to a vector of model parameters X [Gupta et al.(1998)], which can be stated as

$$\min E(\theta) = \{e_i(\theta), \dots, e_m(\theta)\}, \theta \in \Theta \quad (3.4)$$

where the goal is to find values for θ (a set of model parameters) within the feasible parameters space Θ that minimize all of the model residuals $e_i(\theta)$, $i = 1, 2, 3, \dots, m$ at different calibration targets. Here, we apply an aggregation scheme to solve the multi-objective optimization problem. The informativeness strategy described above is used to determine the weights between each calibration target. Next the multi-objective functions of prediction of different variables are aggregated into one single objective with appropriate weighting:

$$e(\theta) = \sum_{i=1}^m \omega_i \times e_i(\theta) \quad (3.5)$$

where ω is the weight for the model residual at each calibration target. The weighted function is then used to constrain the model with observations from available monitoring sites.

3.3 Application Example: Catchment and Model Setup

3.3.1 Multiple Observation at the Catchment

The study area, the 20-acre Shale Hills Watershed (Figure 3.2), is characterized by relatively uniform side slopes with swales on both sides of the stream [Lin(2006)]. Over several decades, the Shale Hills Watershed has been used as an experiment field in a series of hydrological studies [Nutter(1964), Lynch(1976), Qu & Duffy(2007)]. Recently, the watershed has been part of the Critical Zone Observatory (CZO) project supported by U.S. National Science Foundation, as Susquehanna-Shale Hills CZO (SSHCZO).

For high-resolution, fully coupled hydrological modeling, sub-daily climate

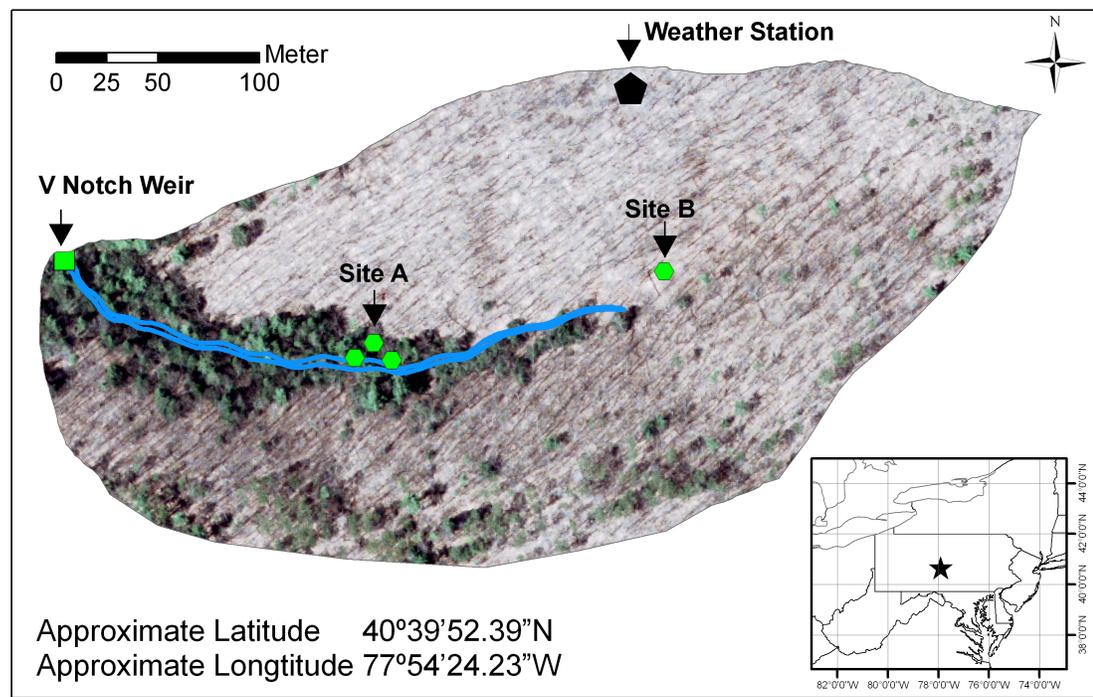


Figure 3.2. Map of the SSHCZO catchment showing the locations of streamflow station, weather station, and pressure transducers.

data are required for precipitation, air temperature, relative humidity, incoming shortwave radiation, and wind speed. These data are available from the weather station at the north ridge with 10-minute frequency [Duffy(2012)].

The soil hydraulic parameters for the van Genuchten model [van Genuchten(1980)], including parameters to describe the inverse of air-entry suction (α), pore size distribution (β), saturated hydraulic conductivity (K_{sat}), saturated water content (h_{sat}), and residual water content (h_{res}), are derived from the field data set of [Lin(2006)].

For the land surface, parameters such as leaf area index (LAI) and roughness length were projected from NLDAS vegetation parameters. Notably the spatial land cover pattern was determined by tree species and density [Yu et al.(2014a)].

3.3.2 Model Setup and Parameterization

We applied PIHMgis [Bhatt et al.(2006)], a tightly-coupled GIS interface to PIHM to set up the modeling at SSHCZO. The procedures are illustrated in Figure 3.3. The 1m DEM [Guo(2010)] was applied to decompose the watershed into 535 triangles and 20 linear segments of stream channels. The tree survey data [Eissenstat(2008)] was used to spatially parameterize the land cover at SSHCZO. The soil classes [Lin(2006)] were also projected on each computational unit of PIHM.

In the model, soil hydraulic properties are used to adjust parameters during the calibration. The pre-calibration estimation of each parameter is presented in Table 3.1. The range of each parameter is estimated from the physical meaning. We focused on the model calibration targets, including the streamflow at the outlet [Duffy(2010b)], the water table depths at riparian zone [Duffy(2010a)] and upland [Lin(2010)], and the total evapotranspiration (ET) of the watershed [Davis(2010)] (Figure 3.2). The observed data was resampled into hourly time series for calibration targets (Figure 3.4).

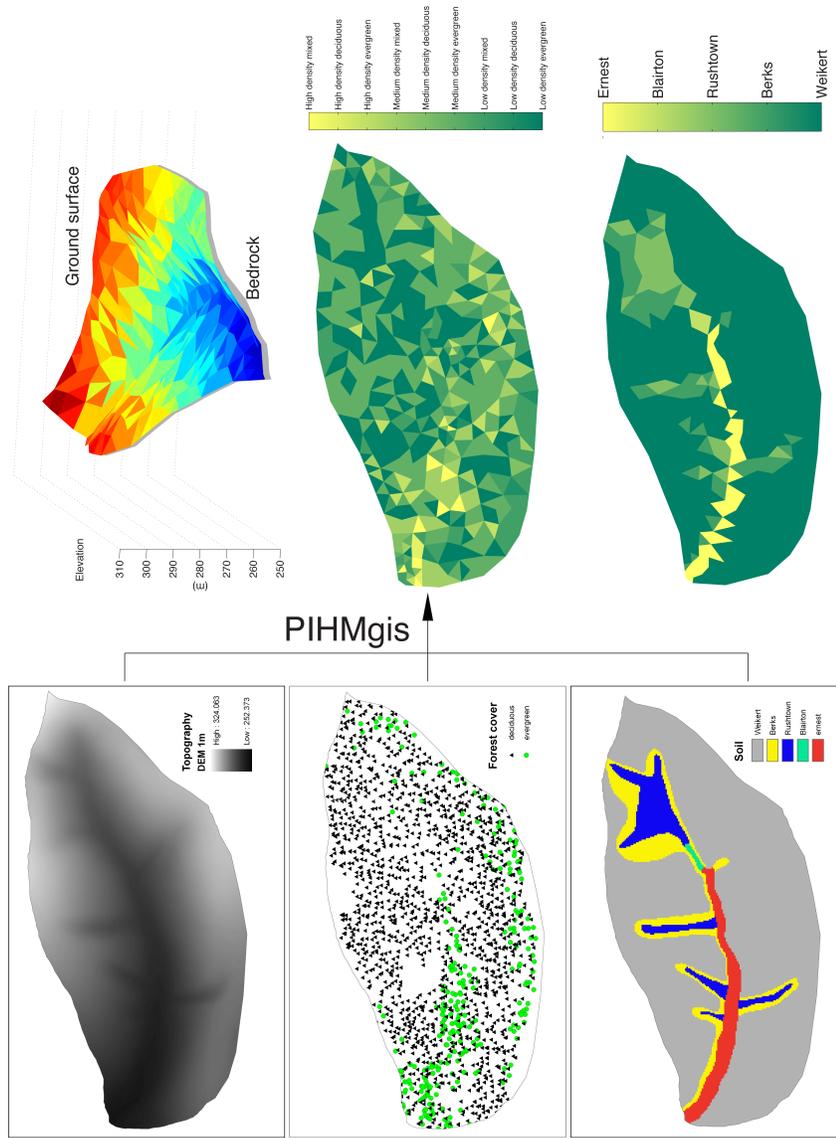


Figure 3.3. The PIHMgis tools process spatial layers to assign physical values and hydrological parameters. The 1m DEM was used to decompose the watershed into 535 triangles and 20 linear segments of stream channels (upper pair). The tree survey data was used to spatially parameterize the land cover types (middle pair). The soil classes were projected on each triangle (lower pair).

Parameter	Hydrological processes	Pre-calibration estimation	Range
Matrix conductivity	Subsurface flow	Pedotransfer functions; field data *	2 order (multiply by 0.01 ~ 100)
Macropore conductivity	Subsurface flow	100 times of matrix conductivity *	2 order (multiply by 0.01 ~ 100)
Topsoil conductivity	Infiltration	Pedotransfer functions; field data	2 order (multiply by 0.01 ~ 100)
Macropore depth	Subsurface flow	Estimated from root system	0 ~ bedrock depth
Porosity	Subsurface flow	Pedotransfer functions from soil texture; field data *	0 ~ 1
Air-entry suction α	Subsurface ow, recharge	Pedotransfer functions from soil texture	1 order (multiply by 0.1 ~ 10)
Pore size distribution β	Subsurface ow, recharge	Pedotransfer functions from soil texture	1 order (multiply by 0.1 ~ 10)
River bed conductivity	Channel routing	Hard coded to be 1.0 (vertical) and 0.1 lateral	2 order (multiply by 0.01 ~ 100)
River Mannings roughness	Channel routing	m/day * [Dingman(2002)]	1 order (multiply by 0.1 ~ 10)

Table 3.1. Sensitive PIHM parameters and their feasible ranges.

* parameters have both vertical and lateral values

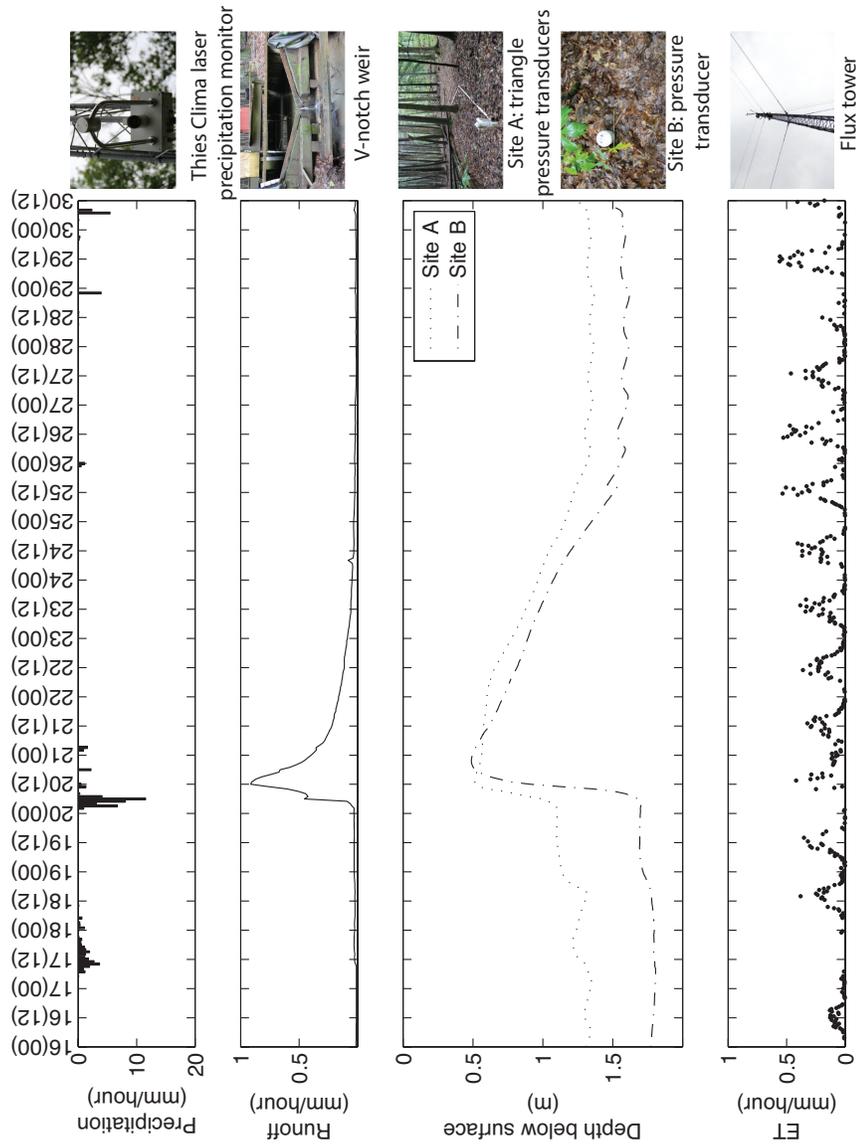


Figure 3.4. Observed time series data for model calibration in June 2009.

Top x axis is time in June 2009 in the format of date(hour). The precipitation is monitored by Thies CLIMA Laser Precipitation Monitor at the weather station. The streamflow is monitored by the notch. The water table depth at site A is observed by a Druck pressure transducer

CS420-L. The water table depth at site B is observed by a 0.5 m Odyssey Capacitance Water Level Recorder. The latent heat flux is measured with a LI-COR LI-7500 CO₂/H₂O Analyzer and then is converted into ET.

3.3.3 Performance of Each Metric

To evaluate the performance of each metric, we conducted three single-objective optimization experiments. We targeted the outlet streamflow by applying different model evaluation metrics: (a) RMSE; (b) 1-NSE; and (c) 1-R. In each experiment, the evolutions of all three metrics were plotted in Figure 3.5. All of the metrics were converging to a minimum value when we constrained RMSE and 1-NSE (Figure 3.5 (a) and (b)). The constraining of 1-R did not lead to the minimization of RMSE and 1-NSE, especially after the 50th generation in the algorithm (Figure 3.5(c)). The constraining of RMSE and 1-NSE had a similar effect in the single-objective optimization processes. Because of the dimensionlessness of NSE, we decided to use NSE as the metric for the calibration with multiple variables.

3.3.4 Formulate Aggregating Objective Function

We did Monte Carlo analysis to assess the informativeness of each calibration target. For each simulation, the NSE of each target was determined for streamflow, water table monitoring site A and site B, and ET. In each case, the NSE ranging between 0 and 1 was plotted (Figure 3.6).

From Figure 3.6, I can see that the highest correlated variables were the streamflow and water table monitored at site A. The interpretation is that a satisfactory simulation of water table monitoring at site A is necessary for the prediction of streamflow. The low correlation suggested conflicting model error between the two variables. The informativeness of each target was calculated according to Pearson product-moment correlation coefficient (R) in Figure 3.6 . Therefore, we formulated the weighted-objective function as:

$$e(\theta) = \frac{1-NSE_{streamflow}+0.567 \times (1-NSE_{siteA})+0.280 \times (1-NSE_{siteB})+0.011 \times (1-NSE_{ET})}{1+0.567+0.280+0.011} \quad (3.6)$$

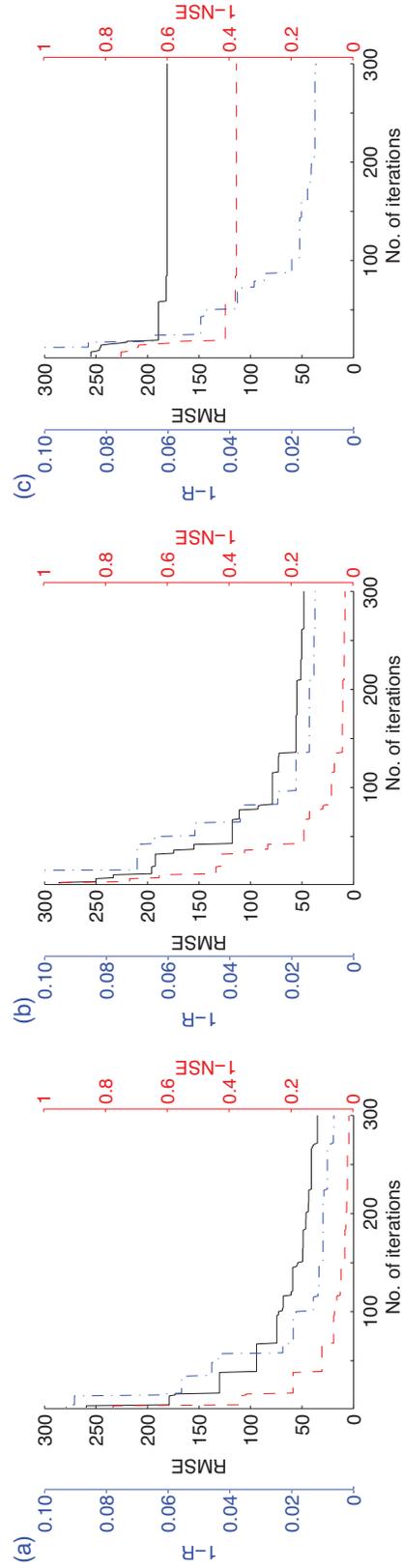


Figure 3.5. Comparison of metrics.

The trajectories were plotted in three cases with different objective functions: (a) RMSE, (b) 1-NSE, (c) 1-R. (a) and (b) suggest that it is efficient to use RMSE and 1-NSE constraining the model. (c) suggests that minimizing 1-R does not improve the model performance according to the evaluation from RMSE and 1-NSE.

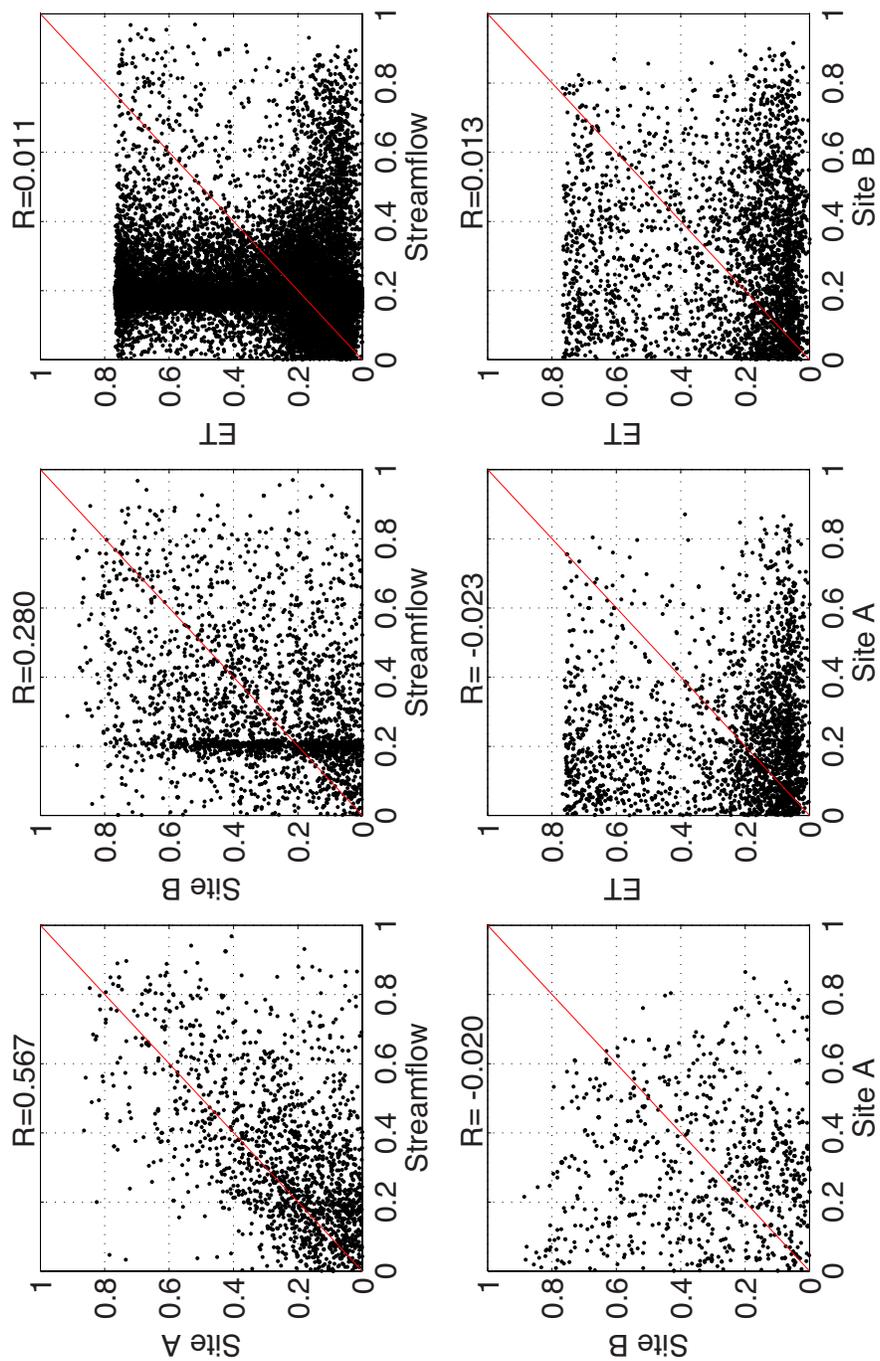


Figure 3.6. NSE of each hydrological variables in the Monte Carlo analysis.

3.3.5 Model Calibration and Evolution of the Model Performance

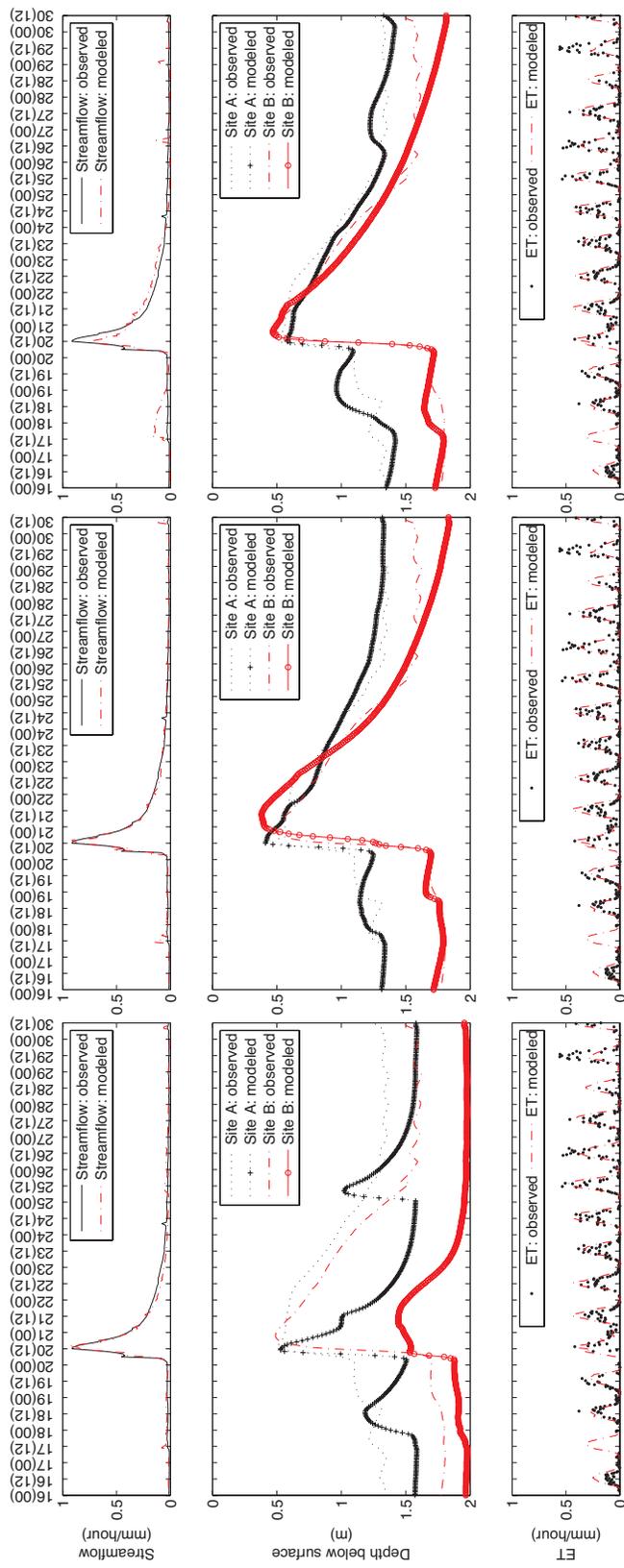
Applying the informativeness-based, weighted objective function, we calibrated PIHM for hourly streamflow, water table depth, and ET. The weighted-objective calibration improved the prediction of water table depths and ET without significant degradation of streamflow prediction (Figure 3.7, Table 3.2). However, without informativeness, the un-weighted averaged calibration demonstrated strong conflicting between the model performance of streamflow and groundwater table and was inclined to the prediction of the groundwater table at Site B and ET.

We plotted the metrics of streamflow, groundwater table at site A, groundwater table at site B, and ET during the evolution of the weighted objective function (Figure 3.8). The trajectories of the metrics 1-NSE converged to a minimum when the weighted objective e was optimized. The general trend of each 1-NSE demonstrated the non-conflicting characteristic in ideal circumstances. The fluctuation suggested that mild conflict existed due to model structure and observation error. The 1-NSE at each site converged to less than 0.1, except for that of ET, which implied the model discrepancy in the representation of the observed latent heat flux [Shi et al.(2013)].

	Single-objective		Informativeness-based weighted-average		Un-weighted average	
	NSE	Weight*	NSE	Weight*	NSE	Weight*
Streamflow	0.977	1	0.916	1.000	0.880	1
Groundwater table: Site A	-0.867	0	0.923	0.567	0.877	1
Groundwater table: Site B	-0.927	0	0.866	0.280	0.914	1
ET	0.457	0	0.460	0.011	0.593	1

Table 3.2. Calibration results from single objective, informativeness-based weighted-average, and un-weighted average.

* Weight is the ω in Equation 3.5.



(a) Single objective calibration
 (b) Informativeness-based weighted-average
 (c) Un-weighted average

Figure 3.7. The simulated and observed streamflow, water table depth at Site A, Site B, and ET: (a) single objective calibration; (b) weighted calibration; (c) un-weighted calibration.
 Top x axis is time in June 2009 in the format of day(hour).

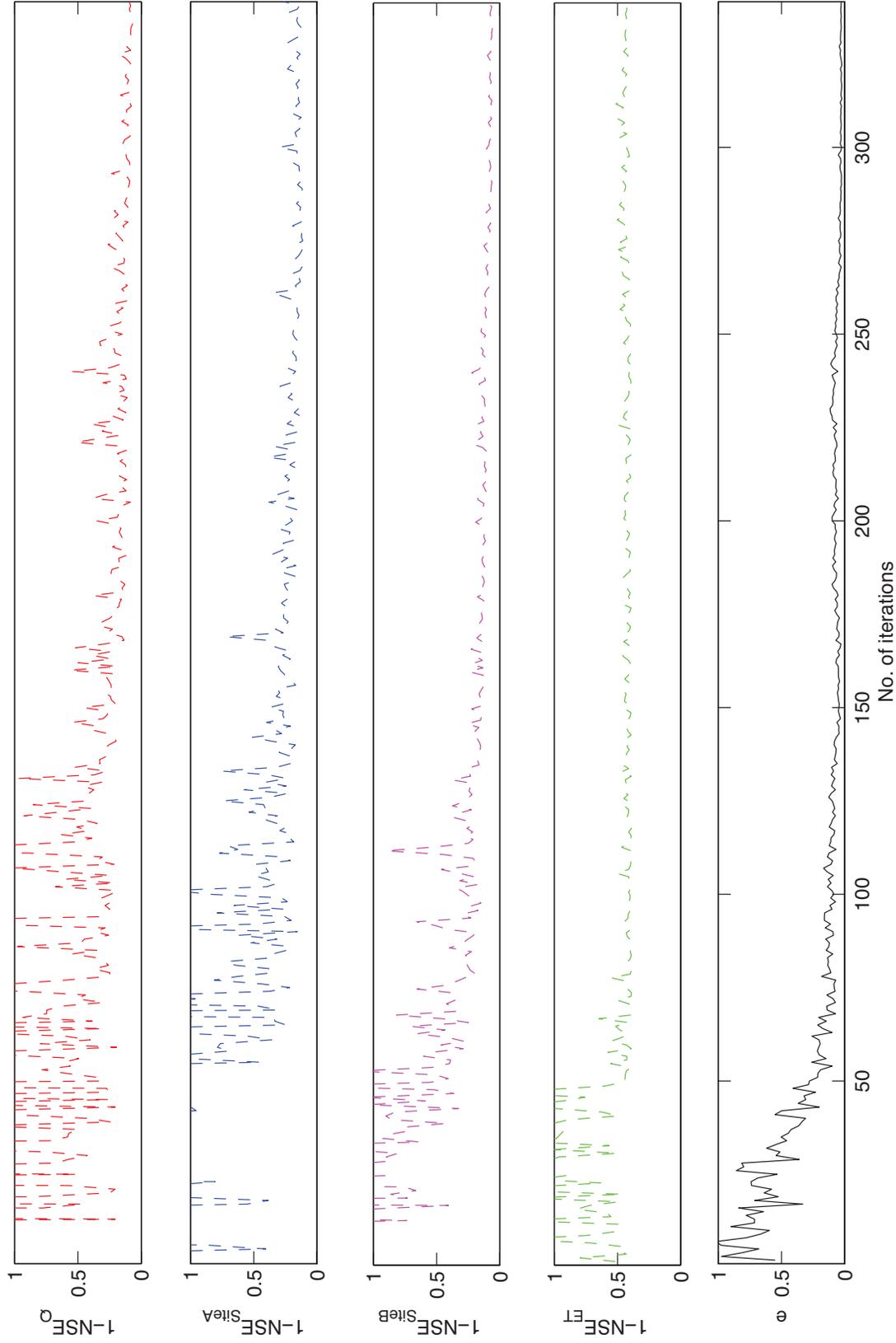


Figure 3.8. Trajectories of metrics at streamflow, site A, site B, and ET during the weighting objective function evolution.

3.4 Discussion

3.4.1 Performances of Metrics

The performance of three common metrics - RMSE, NSE, and R - was used to evaluate the residual error between observed and modeled time series. RMSE is dependent on the scale of the dataset that is being analyzed and limited in comparison model performances on the same catchment and dataset [Dawson et al.(2007)]. An advantage of NSE is that it is dimensionless. No matter the scale of catchment and hydrological property of a dataset, good model performances can be detected when the value of NSE is close to 1. R is also a dimensionless metrics; however, it is insensitive to additive and proportional differences between the observed and modeled datasets [Dawson et al.(2007)]. In the comparison experiment, it was demonstrated as the stop of RMSE and 1-NSE after the 50th generation of minimizing the value of 1-R. Similar functions of RMSE, NSE, and R were found in [Dawson et al.(2007)] and [Abrahart et al.(2011)]. [Dawson et al.(2007)] compared 18 metrics of four sets of hypothetical flow forecasting model outputs, and the result indicated that the best performance of R did not score well when compared to the metrics of RMSE and NSE. RMSE and NSE both reached satisfactory performance with one set of hypothetical flow forecasting model outputs. In addition, the dimensionless NSE is convenient for the aggregating of objective functions [Rozos et al.(2004), Dung et al.(2011)]. Dimensional metrics require normalization processes before aggregation [Li et al.(2010)].

3.4.2 The Meaning of Informativeness of Each Calibration Target in Constraining Model Parameters

The implementation of distributed and fully coupled environmental models clearly increases the the amount of observational data required to constrain the model [Stisen et al.(2011)]. For PIHM, the calibration strategy suggests that streamflow is strongly dependent on the water table depth within the riparian zone (Site A). In Figure 3.6, it can be seen that the model performance at streamflow and Site A behaved very much in tandem with each other: i.e., changing a parameter set will

either increase or decrease the NSE in most cases. However, the model performance of water table depth at Site B was weakly correlated with the performance of streamflow. The low correlation of model performance between ET and other variables suggested that the model performance for a event does not significantly affect the general behavior of the season-scale performance [Yu et al.(2013)].

The selection of weighting coefficients for hydrological models has been reported in several multi-objective calibration studies [Rozos et al.(2004), Li et al.(2010), Dung et al.(2011)]. It was found that even though NSE was a dimensionless metric, a direct incorporation of these metrics of different measurements into the objective function might not be theoretically reasonable [Rozos et al.(2004)]. In one study, the comparison experiment showed that uniform weighted calibration performed worse than the result of weighted case [Dung et al.(2011)]. Different optimal parameter sets can be obtained by changing the weighting coefficients. [Rozos et al.(2004)] followed a hybrid strategy based on a combination of automatic and manual methods by adjusting the weights according to previous optimization results. [Dung et al.(2011)] developed a multi-site water level calibration of an inundation model. The weight of NSE at each gauging station was assigned according to the inundation impact. Expert knowledge was required to subjectively assign the weights. In this study, I developed an informativeness concept, which reflects the model-coupling scheme of different processes and enables consideration of the spatial constraints of the watershed. It was argued that the quantification of the weighting coefficients could be obtained by the informativeness of each calibration target. The informativeness-based strategy for formulating the weighted objective function avoids subjective judgements and could be easily adapted by other integrated models. Noticeably, informativeness could also be used as a threshold for the selection of calibration targets. When the informativeness is too high, which means the two calibration targets are highly correlated, it is redundant to incorporate both targets, while a low value of informativeness suggests the model's incapability of reproducing both targets at the same time.

3.4.3 Comments on the Informativeness-based Weighted Calibration

In this paper, I used the weighted average function to realize multiple constraints of the model from observed variables at different hydrologic processes. With the constraint of multi-measurements, the PIHM modeling results significantly enhanced the prediction of water table depth. The key aims of distributed modeling schemes are to reproduce multiple moisture fluxes and to reflect the spatial heterogeneities of the hydrological mechanisms [Kim et al.(2012)]. The weights selection determines the search preference for high-weight targets. I obtained the non-dominated solutions from CMA-ES samples in the weighted average calibration and un-weighted average calibration (Figure 3.9). Both calibrations attempt a compromise between all objective functions. However, the un-weighted calibration sacrificed the model's performance for streamflow to improve prediction of ET and the groundwater level at Site B, which generated hydrologically unacceptable solutions. Clearly, informativeness-based weighted calibration avoided the situations of unfavorable compromise.

For the aggregation approaches, multi-objective optimization problems can be resolved by the non-dominated sorting. This method often generates a large number of Pareto optimal sets. Usually, a further step of selection is necessary to obtain a smaller number of Pareto fronts as physically sound solutions [Khu et al.(2008)]. Also, selecting a final set of parameters from the non-dominated solutions could be another challenge (Figure 3.9). The pre-defined weights decide the final selection of the solution. Comparison between aggregation and non-dominated approaches suggested that both of the methods can find parameter values with overall good performance [Huang(2014)]. We argue that the informativeness represents a preference for ordering for each calibration target in the hydrological model. The weighted average aggregation method saved computation cost and avoided the selection process from a large number of Pareto-optimal solutions, each of which are efficient for computationally expensive models.

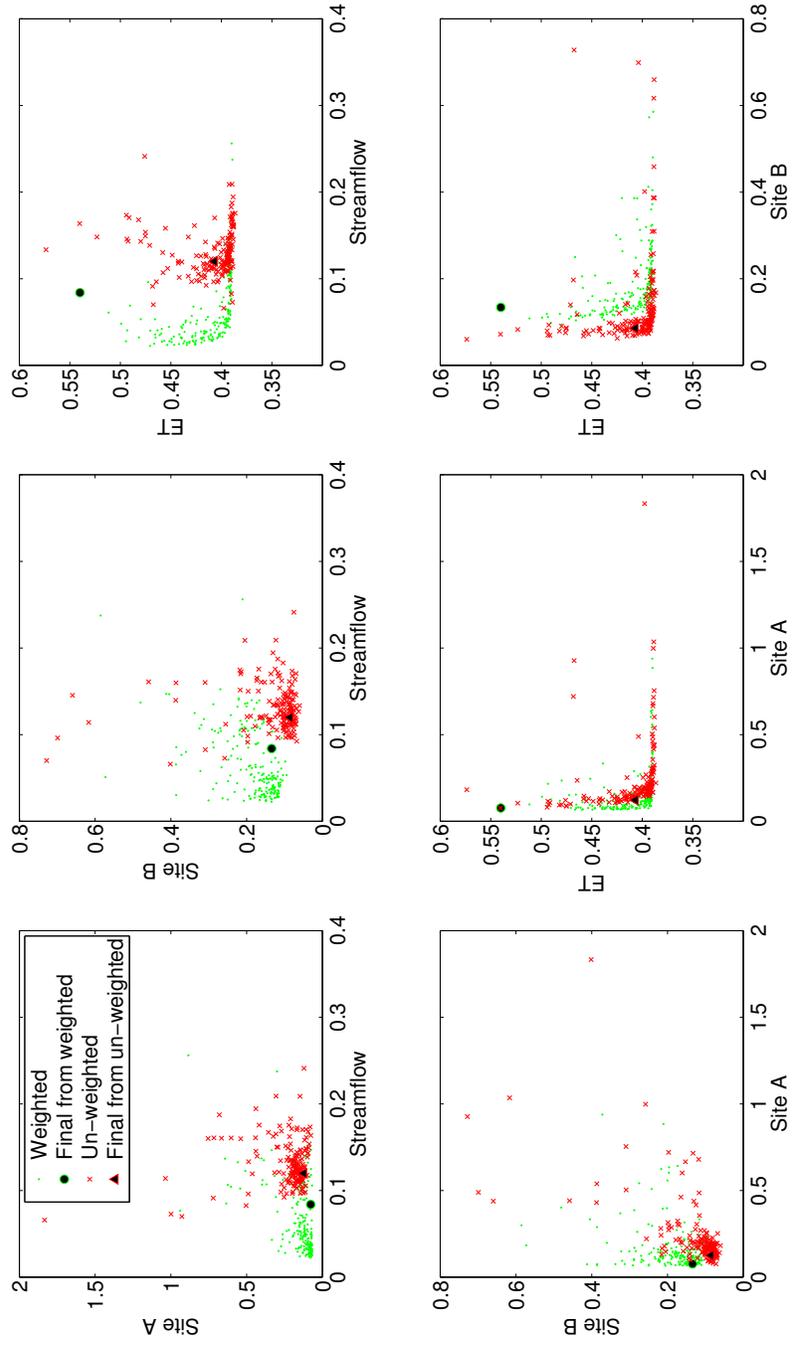


Figure 3.9. Objective function values in the non-dominated solutions from CMA-ES samples.

3.5 Conclusion

This paper explores the informativeness-based, weighted-objective calibration of a physics-based, fully-coupled hydrological model PIHM using observed streamflow, water table depth, and eddy-flux tower data. The calibration strategy application in Shale Hills Watershed, suggests the following.

1. The metric of NSE can be efficiently used in the assessment of PIHM performance.
2. The informativeness provided a useful framework for objectively determining weights between each calibration target. Here, the correlation of model performance at each target was used for the evaluation of informativeness. Results suggest a reasonable balance in streamflow, water table depth and ET was achieved with this weighting strategy.
3. The comparison between single-variable optimization and weighted-average optimization suggested that targeting only on streamflow could hardly predict the spatial subsurface flow processes. Distributed hydrologic process modeling relies on multi-variable constraints across the domain.

Chapter 4 The National Wetland Inventory as a Constraint on Wetland Hydrology Modeling for Regional Climate Change Impact Assessment

4.1 Introduction

One of the recurring themes of research at the environmental resources management is the understanding and sustaining of wetlands [Smardon(2009), Tiner(2002)]. Our limited knowledge of the geospatial extent, behavior, and classification of wetlands has caused difficulties in wetland identification and spatial mapping [Tiner(2002), Wardrop et al.(2007a)]. From a hydrological view, it is generally recognized that shallow groundwater is a defining feature of wetlands and can exert a strong control on plant and animal life, as well as on soil development [Council(1995)]. The National Research Council [Council(1995)] concluded, wetland hydrology should be considered to be saturation within 1 ft. of the soil surface for 2 weeks or more during growing season in most years (about every other year on average).

The National Wetlands Inventory (NWI) is the most readily available data source for mapping the location and spatial extent of wetlands throughout the United States [Martin et al.(2012)], though it has limitations. A number of investigators have attempted to critique and improve the NWI database [Johnston & Meysembourg(2002), Kudray & Gale(2000), Maxa & Bolstad(2009), Ozesmi & Bauer(2002)]. These efforts to improve wetland maps and the underlying database have used remotely sensed, geospatial information [Maxa & Bolstad(2009)] as well as geomorphological, biophysical, and hydrologic observations [Wardrop et al.(2007b)] to establish and map similar classes on a watershed basis. It is generally agreed that quantitative hydrologic characteristics are among the most important factors in wetland identification, delineation, classification, and mapping [Johnson et al.(2004), Wardrop et al.(2007b)].

Clearly, geospatial classification of wetlands and their relation to shallow groundwater-stream conditions across the watershed provide a valuable tool for resource assessment, and the close relation of shallow groundwater suggests that wetlands may also play a useful role in watershed modeling studies.

Traditional watershed modeling has had a tendency to focus primarily on streamflow simulation [Pyzoha et al.(2008), Su et al.(2000), Yuan et al.(2011)], with limited attention paid to understanding spatial patterns of streamflow [Grayson et al.(2002)]. More recently, physics-based hydrological models have begun to demonstrate the capability of utilizing geospatial data for landscape, vegetation, soil, and geology as a basis for a more complete mapping of groundwater-stream dynamics across the watershed [Mirus & Loague(2013)]. A question we attempt to answer here is, can distributed watershed models utilize the National Wetlands Inventory as a data source for constraining surface-groundwater conditions [Lu et al.(2009), Min & Wise(2009), Scibek & Allen(2006)]?

Within the Chesapeake Bay watershed, climate variability has and is expected in the future to influence environmental changes in the Bay itself [Najjar et al.(2010), Neff et al.(2000), Rogers & McCarty(2000)]. Statistic analysis suggested that a 3°C increase in mean annual temperature may be associated with a strong draught in mid-Atlantic United States. Although increases in precipitation may mitigate part or the likely water shortage [Huntington(2003)], it is also reasonable to assume that freshwater wetlands would be susceptible to

climate change under these conditions [Najjar et al.(2010), Wardrop et al.(2007a)]. Located within the Chesapeake Bay watershed, the Susquehanna River Basin encompasses an area of 27, 501 square miles, of which 76 percent is in Pennsylvania, 23 percent in New York, and 1 percent in Maryland. Over 90 percent of the basin is underlain by sedimentary rock strata largely undisturbed in the Appalachian Plateau Province but convoluted and eroded in the valley and Ridge Province. The principle rivers are the Susquehanna (444 miles), West Branch (228 miles), and Juniata (86 miles). Approximately 90% of the Upper Chesapeake Bay and 50% of the entire Chesapeake Bay freshwater inputs are from the Susquehanna. Groundwater is by far the largest store of water in the Susquehanna River Basin, serving 50% of the water users in the basin. Groundwater recharge, discharge (baseflow), and shallow groundwater storage are highly sensitive to climate conditions [Fan et al.(2013), Green et al.(2011), Kløve et al.(2013), Taylor et al.(2013)]. The timing and magnitude of drought conditions depend on the space and time scales of groundwater storage. Recently, the assessment of climate change impacts and management practices on the watershed resources was the focus of a comprehensive multidisciplinary study [Brooks & Wardrop(2013), Zhang et al.(2010)]. Another important study is the NSF-funded Susquehanna-Shale Hills Critical Zone Observatory (SSHCZO), an experimental site devoted to integrated mathematical modeling and testing studies of environmental change impact [Banwart et al.(2011)]. In this paper, we examine the response of shallow groundwater conditions of NWI wetlands for 7 mesoscale watersheds of the Susquehanna River basin, ranging in size from 163 km² to 902 km². The objectives of the study were to: 1) calibrate and validate a physics-based, spatially distributed hydrologic model utilizing NWI data as a constraint on the groundwater table over each basin; 2) use the model to extract distributed information for the near-surface hydrologic response for the 2004-present period; and 3) examine the sensitivity of shallow groundwater table depth at NWI sites under the IPCC climate change scenario [Najjar et al.(2009)] using the validated model.

4.2 Description of Study Area

This study is based on 7 mesoscale watersheds across SRB (Figure 4.1). The characteristics of watersheds are listed in Table 4.1 and Table 4.2.

(a) KC

Kettle Creek is a 74.7-km-long tributary of the West Branch Susquehanna River, situated in the mountainous high plateau section of the Allegheny Plateau province of north central Pennsylvania. Overall, the land cover in the watershed is categorized as 85% forested, 13% agriculture. The historic surface and underground coal mines have seriously affected the water quality of the downstream part of the stream before it discharges to the main stem of the West Branch of the Susquehanna (Love et al., 2005). Many isolated wetlands are found within the mined land [Tiner(2003)].

(b) YWC

Young Womans Creek is another tributary of the West Branch Susquehanna River, encompassing the area of 120 km². YWC situated in the Allegheny Plateau province of north central Pennsylvania, next to KC. Heavy forest covers 98.8% of the watershed. Both KC and YWC consist of many very deep, steep-sloped valleys that are separated by narrow, flat to sloping uplands.

(c) LJR

The Little Juniata River is a tributary of the Juniata River, the second-largest tributary of the Susquehanna River. It is formed by the confluence of several short streams. It flows northeast in the Logan Valley at the foot of Brush Mountain. The land cover of the LJR is 72.0% forest, and 25.6% agricultural. The LJR is located at the transition between the Valley and Ridge and Appalachian plateau physiographic provinces.

(d) SC

Shavers Creek is located southeast of LJR. The land use in the valleys of SC is mixed, primarily, with farming and small towns and significant amounts of forests that cover the run along the ridges. Both LJR and SC consist of numerous long, narrow mountain ridges separated by narrow to wide valleys. Competent sandstones occur at the crests of the ridges, with relatively softer shales and siltstones occurring in most of the valleys.

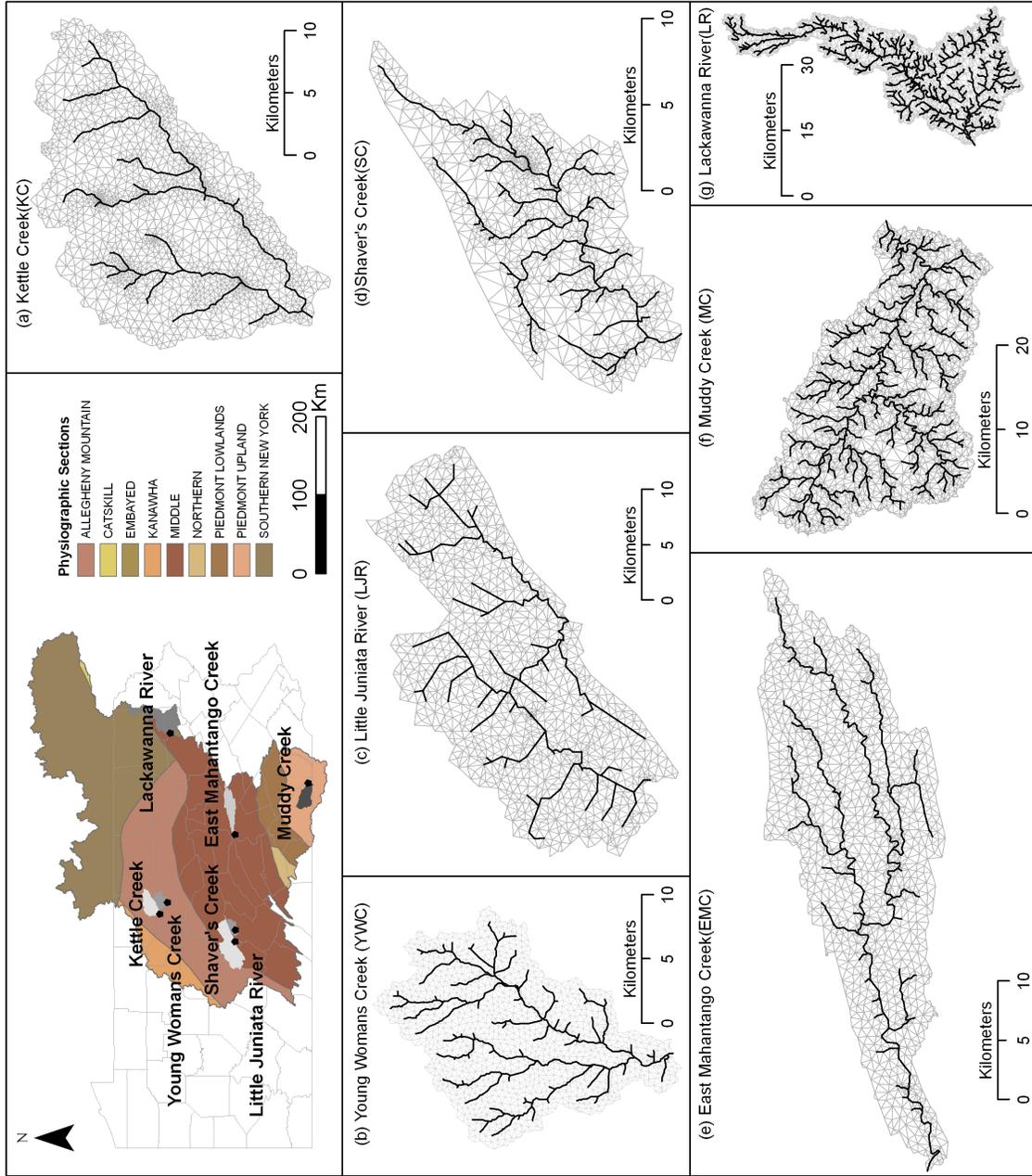


Figure 4.1. Location of the 7 watersheds in the Susquehanna River Basin, and PIHM domain decomposition.

Watershed	Name	Modeling area (km ²)	USGS gauge ID	LAT	LON	Drainage area (km ²)
KC	Kettle Creek	355.44	01544500	41.4758	-77.8261	352.2
YWC	Young Womans Creek	230.70	01545600	41.3894	-77.6911	119.7
LJR	Little Juniata River	843.26	01558000	40.6125	-78.1408	569.8
SC	Shaver's Creek	162.71	01558500	40.6106	-78.0072	120.2
EMC	East Mahantango Creek	422.37	01555500	40.6111	-76.9122	419.6
MC	Muddy Creek	360.50	01577500	39.7725	-76.3161	344.5
LR	Lackawanna River	902.03	01536000	41.3592	-75.7447	859.9

Table 4.1. Overview of watersheds and USGS gauges.

USGS Land Use class ID	Use Description	Wetland	Forest	Development area	Agricultural Land Use	Type
		91, 92	41, 42, 43	21, 22, 23	81, 82, 85	
KC		0.06	85.32	0.12	13.98	Forest
YWC		0.01	98.78	0.10	0.61	Forest
LJR		0.06	72.03	1.75	25.61	Forest
SC		0.27	70.64	0.16	28.39	Forest
EMC		0.23	53.76	1.27	43.06	Herbaceous
MC		0.47	33.06	0.39	65.68	Herbaceous-Forest
LR		1.35	70.09	14.58	9.20	Urban-Herbaceous

Table 4.2. Percentage of land cover classes in each watershed.

(e) EMC

East Mahantango Creek is located in east-central Pennsylvania, consisting of low to moderately high linear ridges and linear valleys. The land cover is characterized by being predominantly forested at ridge tops with agriculture dominating the valley floors.

(f) MC

Muddy Creek is a 27.7 km tributary of the SR in York County, PA, which represents an intensive agricultural watershed. The topography is characterized by broad, gently rolling hills and valleys representative of the Piedmont Physiographic region.

(g) LR

Lackawanna River is the largest tributary to the North Branch of the Susquehanna River in Northeastern Pennsylvania. The LR watershed forms a northern extension of the Appalachian Ridge and Valley Physiographic Province. The presence of anthracite coal has had the most significant impact on the present day LR. The coalfields of the Lackawanna Valley were developed between the 1820s and 1850s, and ended around the 1960s. In addition to mine land, the land cover is predominantly agricultural and forest.

4.3 Materials and Methods

4.3.1 Physics-based Hydrological Modeling

PIHM is a physics-based, fully coupled, spatially distributed, hydrologic model. It simulates the terrestrial water cycle, including interception, throughfall, infiltration, recharge, evapotranspiration, overland flow, unsaturated soil water, groundwater flow, and channel routing, in a fully coupled scheme [Qu & Duffy(2007)]. Evapotranspiration is calculated using the Penman-Monteith approach adapted from Noah_LSM [Chen & Dudhia(2001)]. Overland flow is described in 2-D approximation of St. Venant equations. Movement of moisture in unsaturated zones is assumed to be vertical, which is modeled using Richards equation. The model assumes that each subsurface layer can have both unsaturated and saturated storage components. Balance equations of the unsaturated and

saturated zones are formed in a fully coupled way. The channel routing is modeled using 1-D approximation of St. Venant equations. Spatially, the modeling domain is decomposed into Delaunay triangles. The resolution of triangular mesh allows users to customize according to the geomorphological or hydrological characteristics of the watershed. Also, the triangles can be constrained by point observations (e.g., streamflow, groundwater level, soil moisture, LAI) and the watershed boundary conditions [Kumar(2009)]. The model resolves hydrological processes for land surface energy, overland flow, channel routing, and subsurface flow, governed by a partial differential equation (PDE) system. The system is discretized on the triangular mesh and on projected prisms from canopy to bedrock. The model also includes canopy interception, evapotranspiration (ET), infiltration, and recharge within the fully-coupled system. PIHM uses a semi-discrete, finite-volume formulation for solving the system of coupled PDEs, resulting in a system of ordinary differential equations (ODE) representing all processes within the prismatic control volume. The local system is assembled for the model domain, and the global ODE system is solved using the CVODE implicit solver [Cohen & Hindmarsh(1994)]. Detailed descriptions of the modeling theory and mathematical formulation can be found at the PIHM website (<http://www.pihm.psu.edu/>) and associated publications [Kumar(2009), Qu & Duffy(2007)].

The basic modeling strategy is to calibrate the model using national datasets for streamflow (USGS), soil hydraulic properties (USDA), climate (NLDAS-2), land cover (NLCD) and the NWI data. The latter is used to constrain the shallow groundwater level across each study domain within the calibration. Once calibrated, the model for each catchment is then used to carry out an IPCC climate scenario to project the likely dynamics of wetland groundwater level change under a warming climate (2045-2065). Advantages of the modeling strategy for simulating wetland hydrology are: 1) the space-time patterns of surface and groundwater levels can be evaluated from the climate scenario and compared to historical conditions across each basin; 2) using the PIHMgis tool, the model can extract hydrologic performance and detect change anomalies at each NWI site across the basin; and 3) the distributed model and GIS tool can be used as a climate change assessment tool. An important result of the study is that high resolution, physically based models can be used for establishing a scientific basis for the evaluation of

environmental change and the impact on ecosystem services.

4.3.2 Climate Scenarios for SRB

The climate scenarios used in this study were generated in the EPA-Global Change Research Program [Najjar et al.(2009)], based on IPCC scenarios, which were used in assessing the large-scale impacts of climate change on the physical, chemical, and biological processes in the SRB [Najjar et al.(2010)]. The scenarios were developed by averaging the regional (downscaled) output of seven IPCC global climate models. All model scenarios were for the period 2046-2065 under six greenhouse-gas scenarios considered. The IPCC annual precipitation and temperature are plotted in Figure 4.2.

4.3.3 Proposed Methodology

In principal, a distributed watershed hydrologic model attempts to simulate the hydrological processes for state variables and fluxes within an integrated or coupled modeling framework. Of particular focus here is the coupled surface and groundwater response across the catchment. Limited spatial groundwater-table depth data were available, and the NWI maps were used to constrain the water table during calibration (e.g., wetlands are defined as having a water table within 30 cm of the surface for at least part of the year). USGS streamflow data provides another constraint in the optimization process. The procedure includes the following steps:

1. assign a-priori PIHM modeling parameters from national data sets for land cover, soils, hydrogeology, and topography;
2. use the reanalysis forcing data to calibrate the model by following the method of using streamflow time serials and NWI spatial maps; and
3. conduct a climate change assessment for wetland response under downscaled IPCC climate forcing scenarios.

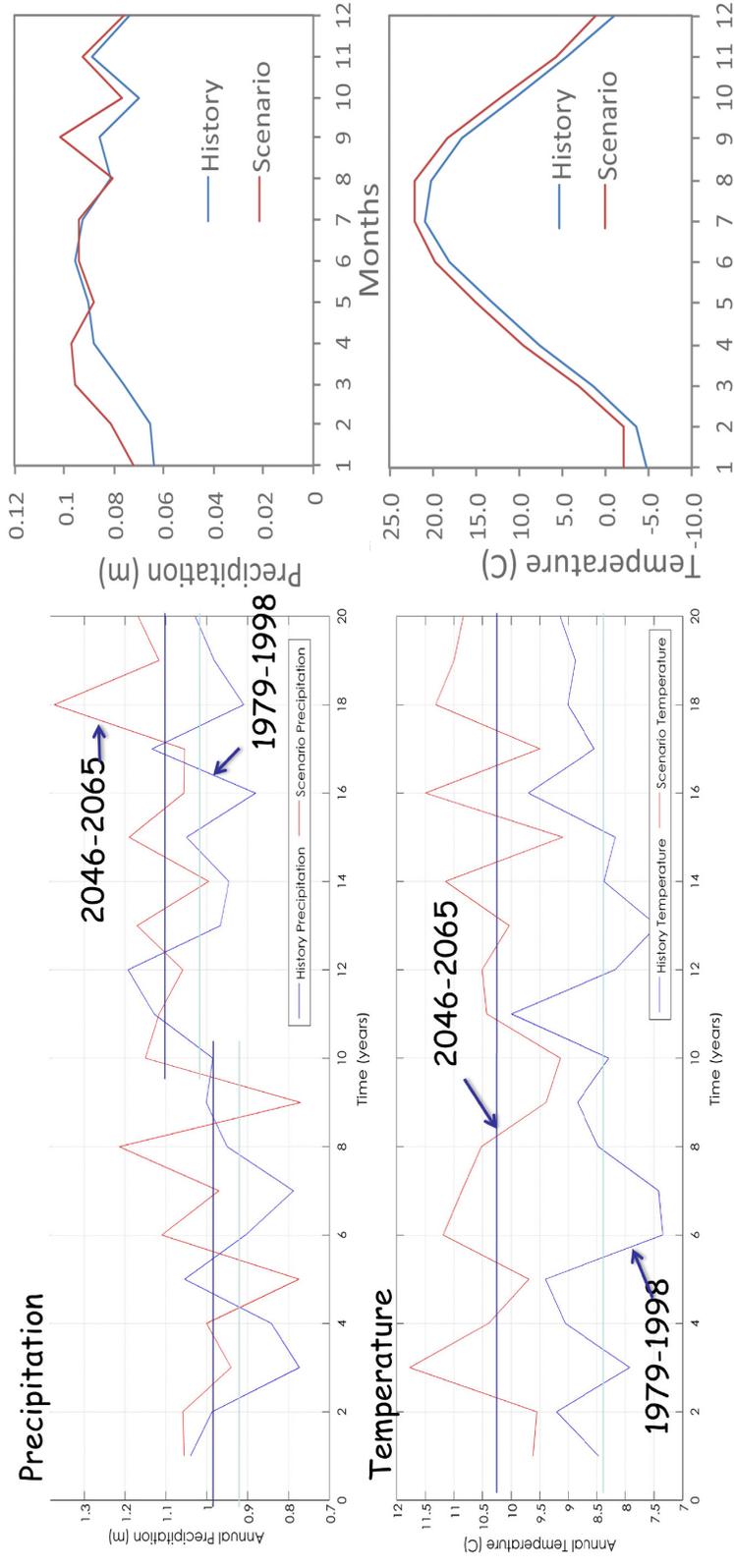


Figure 4.2. Climate change projection.

The overall approach is illustrated in Figure 4.3. We only considered the climate change scenario in this study. A future paper will explore the competing impacts of land-use change and climate change scenarios.

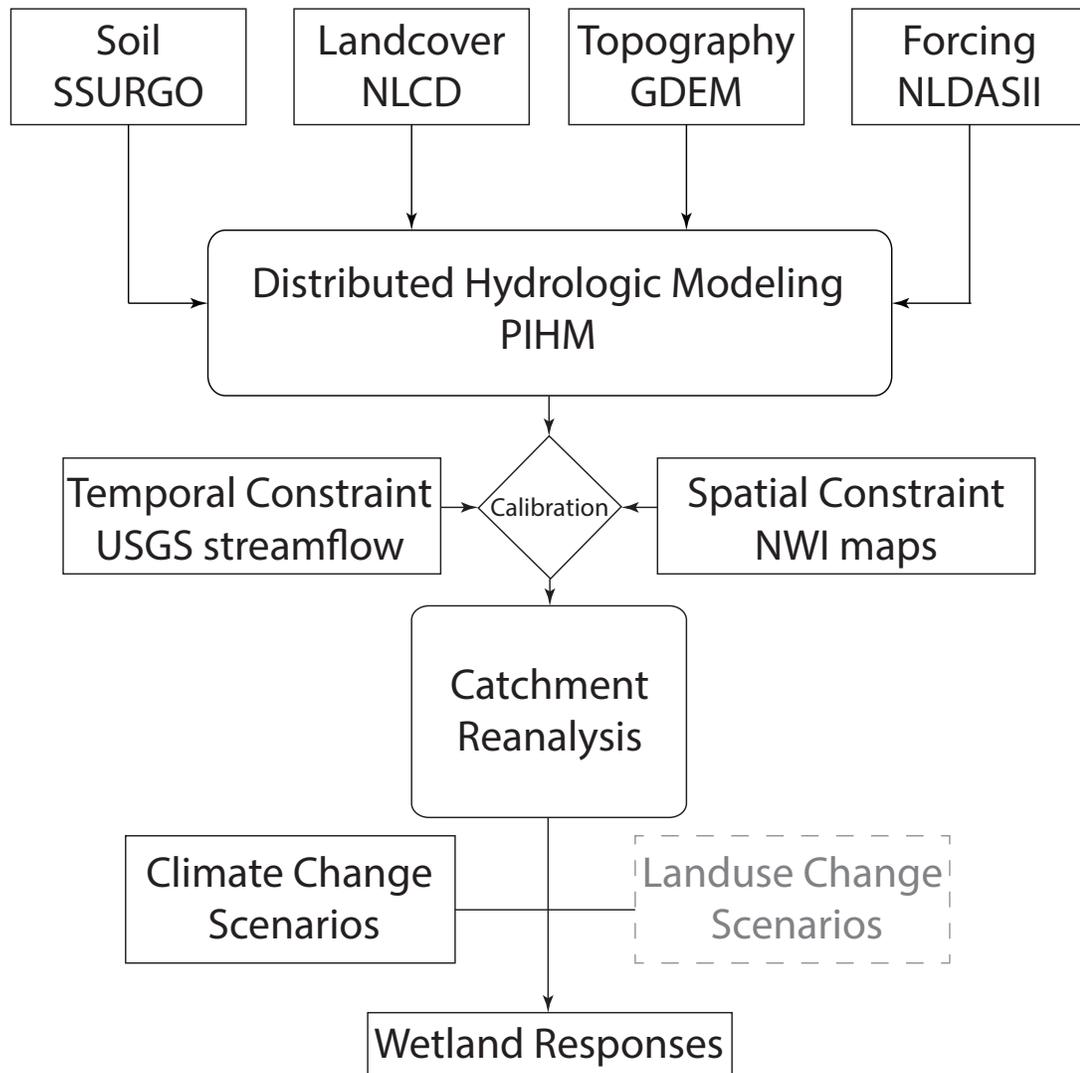


Figure 4.3. Flowchart of the hydrologic modeling for wetland assessment.

4.3.4 Model Data Setup and Parameter Calibration

The Penn State Integrated Hydrologic Model (PIHM) was implemented at each watershed. The model data input included spatial information of landcover, soil, and geology, which were derived from national databases [Yu et al.(2013)]. The depth of shallow subsurface flow was assumed to be a uniform 5 meters below ground surface as an initial estimate. PIHMgis [Bhatt(2012)] is utilized to construct the unstructured mesh and to assign parameters to each mesh element. The model data preparation processes steps are illustrated in Figure 4.4. The soil, vegetation, and topography parameters were estimated by overlapping national data with the model mesh domain. Also the meteorological forcing from NLDAS II was assigned to each mesh in the domain. The resolution of mesh domain is listed in Table 4.3.

Watershed	Modeling area (km ²)	Number of triangles in the mesh domain	Numbers of channels in the mesh domain	Spatial modeling resolution (km ²)
KC	355.44	3098	342	0.115
YWC	230.70	3172	651	0.073
LJR	843.26	2089	264	0.404
SC	162.71	1986	414	0.082
EMC	422.37	2606	509	0.162
MC	360.50	4779	1399	0.075
LR	902.03	5355	1521	0.168

Table 4.3. Mesh resolution

In previous studies, PIHM was manually calibrated [Kumar et al.(2013)] based on a relaxation experiment designed to capture drying soil matrix and macro hydraulic properties (porosity and conductivity) and the baseflow recession to the stream reach. In the relaxation experiment, the soil is set to saturated conditions at the beginning of model simulation. The input is set to zero precipitation, and the model is run until the streamflow approaches zero. The relaxation curve in the simulation is compared to the observed streamflow during the summer drought period (July-August). More recently, [Yu et al.(2013)] have devised a calibration strategy that sequentially estimates hydrologic parameters (e.g., van Genuchten parameters) based on short-term hydrological events and longer time-

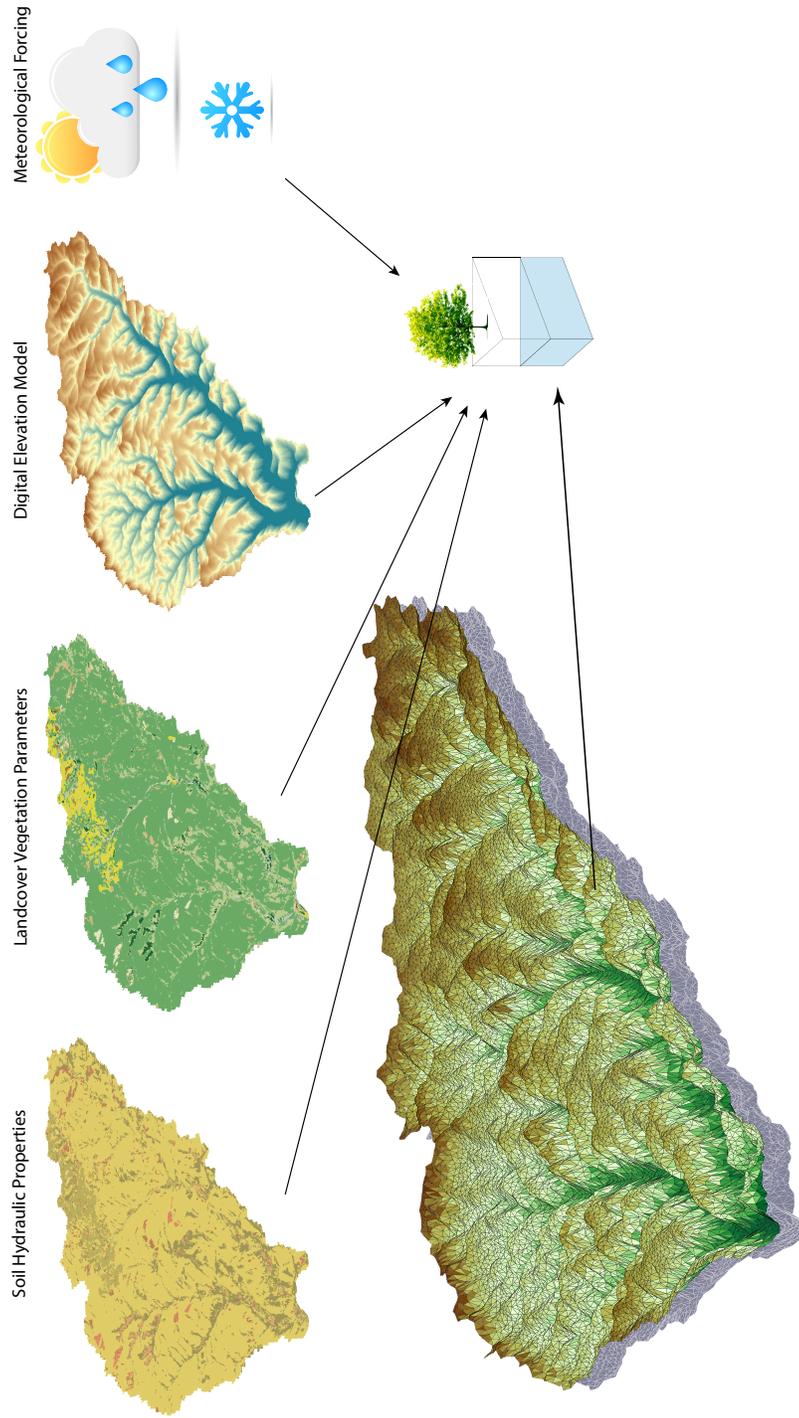


Figure 4.4. Model data setup processes of KC.

scale processes associated with seasonal ET, which control the transpiration, infiltration, and recharge processes. The method reduces the computational effort in model calibration by using an evolutionary algorithm Covariance Matrix Adaptation Evolution Strategy (CMA-ES) and is now the main strategy for PIHM parameter optimization [Yu et al.(2013)]. Formally, the calibration process creates partitions of PIHM parameters into two groups: the first group of parameters generally describes hydrologic processes influenced by hydrologic events and is calibrated by CMA-ES, while the second group of parameters is largely influenced by seasonal changes in energy and is calibrated using annual data. In this study, we use both streamflow and groundwater table depth as estimated by NWI to constrain the model. The NWI map was overlain on the model domain, and NWI wetlands were identified for each triangle. For each triangle that included a wetland, the average of simulated groundwater depth was constrained to be less than 0.3 meters below the land surface. Also, daily streamflow time series were used to calibrate the soil and subsurface hydrologic parameters. The calibration objective function was formulated as:

$$e = 1 - NSE_{streamflow} + \sqrt{\frac{1}{n} \sum_{i=1}^n (GW_i - 0.3)^2} \quad (4.1)$$

where e was the objective function, $NSE_{streamflow}$ was the Nash-Sutcliffe model efficiency coefficient NSE [Nash & Sutcliffe(1970)] of streamflow, n was the number of triangles with minimum groundwater table less than 0.3 meters below land surface at NWI map locations, GW_i was the corresponding minimum groundwater depth below land surface. The event-based calibration time period [Yu et al.(2013)] was for a single flooding event during growing season (from May to Sep) in 2004, and the seasonal scale calibration period [Yu et al.(2013)] was the entire year of 2004.

As a result of the CMA-ES calibration process and the NWI constraints, the spatial and temporal model for each watershed was adequately calibrated by constraining the groundwater table pattern with the NWI wetlands for each catchment. The statistical criteria used to evaluate PIHM performances included wetland prediction percentage and mean error (ME), Pearsons Correlation

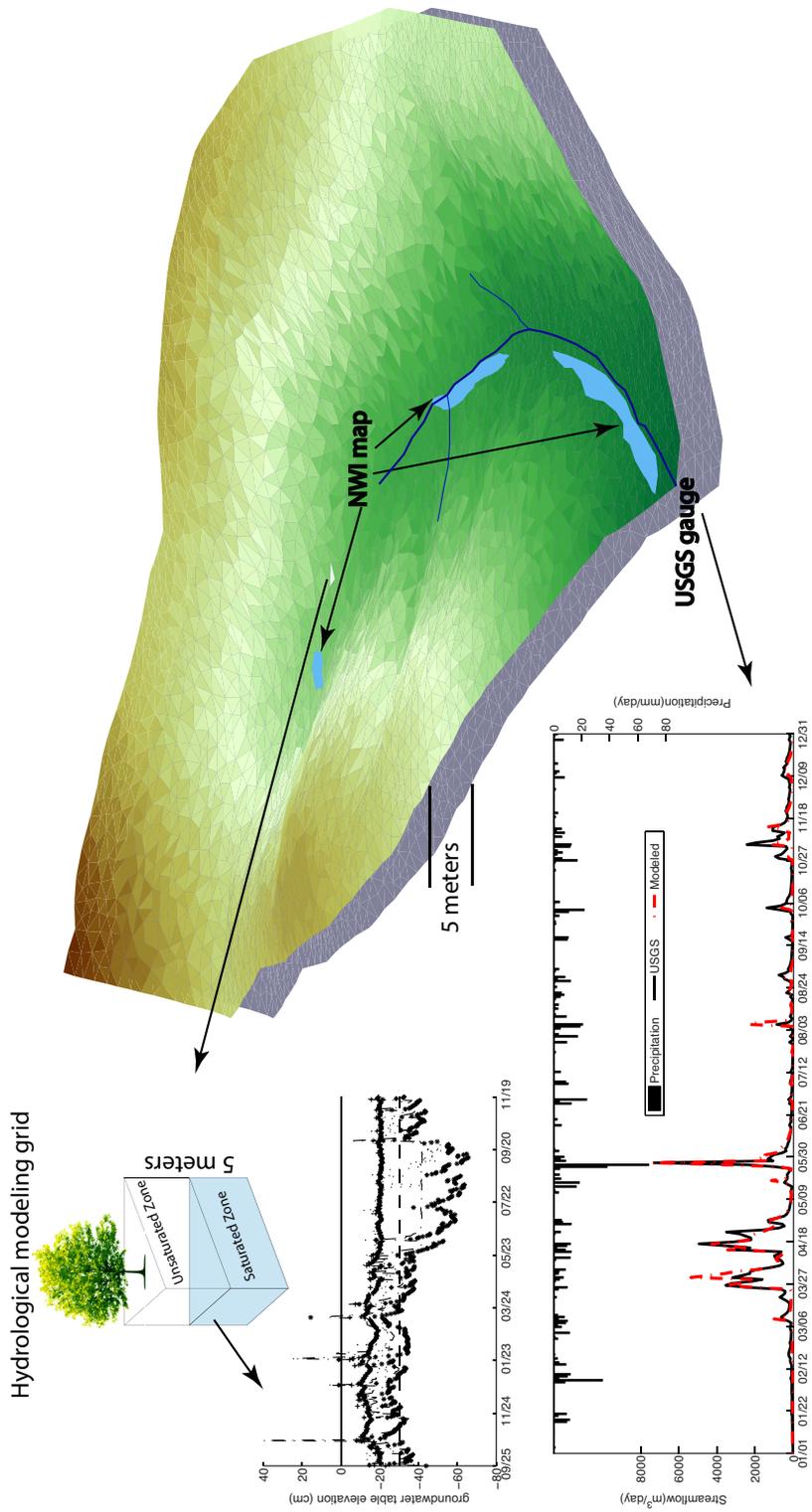


Figure 4.5. PIHM calibration objectives to capture wetland. The average groundwater table depth at each triangle of the computational mesh with NWI-mapped locations, where the water table was constrained to be less than 0.3 meters. The streamflow was also calibrated with USGS stream flow observations.

Coefficient (R), and NSE [Nash & Sutcliffe(1970)].

4.3.5 Assessment of Climate Change Impacts on Wetland

The integrated model and calibrated physical parameters developed in this paper can be used to assess the possible impacts of global climate change on the function of wetlands in the Susquehanna River Basin (SRB). It is critical to note that regional drivers such as climate change scenarios involve large uncertainties including future societal priorities, which can be linked as part of complex environmental models. Here we focused on the sensitivity of groundwater systems to changes in critical input of climate forcing. It is important to consider such impacts in the analysis because of the ever-increasing influence of humans on the physical and biological environment.

4.4 Results

4.4.1 Hydrologic Validation

The simulated and observed streamflow were plotted in Figure 4.6 and model performance was listed in Table 4.4, including daily streamflow ME, R, and NSE. ME is commonly used to evaluate the average systematic error among the simulated and the observed values. Positive values of ME indicate model under-estimation, while negative values correspond to over-estimation. R is a measure of the strength of the association between observed and predicted values and may take any value between -1 and 1. NSE varies from minus infinity to 1.0, with higher values indicating better agreement. The simulated daily streamflow dynamics were calibrated with USGS observations where available.

Watershed	Mean (ME)	error	Pearsons Correlation Coefficient (R)	Nash-Sutcliffe coefficient of efficiency (NSE)
KC	-3.47		0.72	0.41
YWC	10.64		0.74	0.54
LJR	-8.37		0.83	0.68
SC	N/A		N/A	N/A
EMC	18.46		0.60	0.33
MC	N/A		N/A	N/A
LR	-8.37		0.68	0.24

Table 4.4. PIHM performance in daily streamflow simulation from 2004 to 2010

4.4.2 Wetland Spatial Distribution from Reanalysis

According to the simulated daily groundwater table depth for the period 2004 to 2010, we evaluated the wetland area according to simulated depth-to-groundwater-table dynamics (groundwater table depth was less than 0.3 meters for at least 2 weeks from May to September). The spatial constraint during calibration had impacts throughout the entire reanalysis period (1979-present). In general, using historical climate reanalysis to force the model, along with the NWI constraint on the depth to water table, improved the model performance. Figure 4.7 shows the model results for depth to groundwater less than 0.3 meters and the NWI locations. In almost all cases, the model adequately simulates the expected water level of the wetlands, though a few very small NWI wetlands were not identified from the PIHM calibration.

4.4.3 Climate Change Responses

The future climate in Pennsylvania for the 21st century is projected to increase in both temperature and precipitation, although the relationship is complex. In this study, we explore the prospect that climate change in this region may result in changing groundwater table responses that might affect wetlands across the Susquehanna River basin. The approach is to use the Reanalysis period (1979-present) to initialize the model for each basin and to simulate the IPCC climate projection for each watershed and to use present results for the period 2045-2065.

Figure 4.8 illustrates how the climate change projections may induce changes in the groundwater level and thus the wetlands for each basin in our study. The distributions of groundwater level responses were classified into wetland, upland, and the others to examine the spatial heterogeneity of climate change impacts (Figure 4.9). Specifically, the triangles identified as wetlands are selected as the wetland group. And then, considering all the rest of the triangles, if the drainage area above the triangle is less than 10% of the whole watershed, the triangle is classified as the upland group. Last, the remaining triangles are classified as the other group.

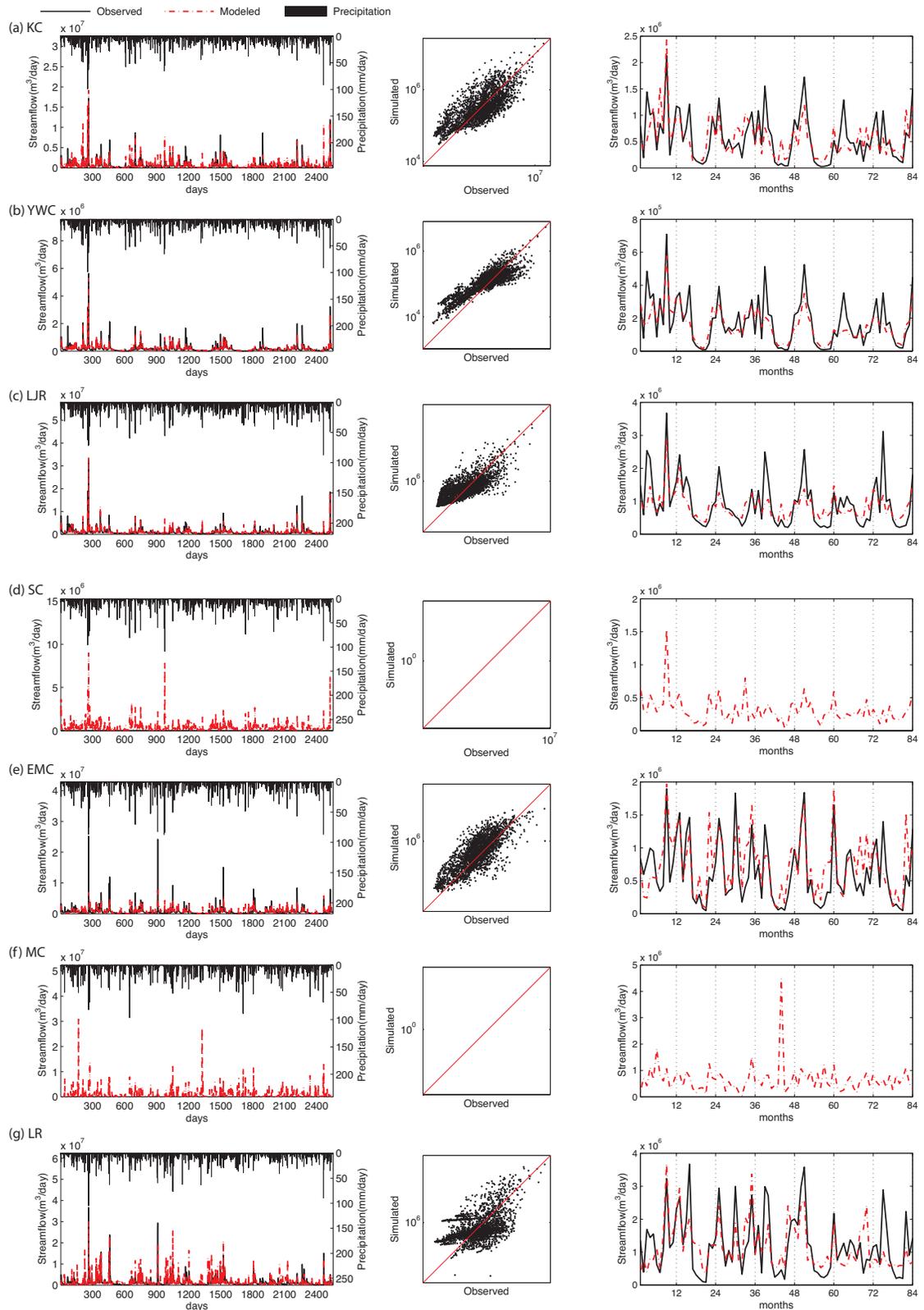


Figure 4.6. Observed and simulated streamflow (2004-2010) at each of the 7 catchments in the study.

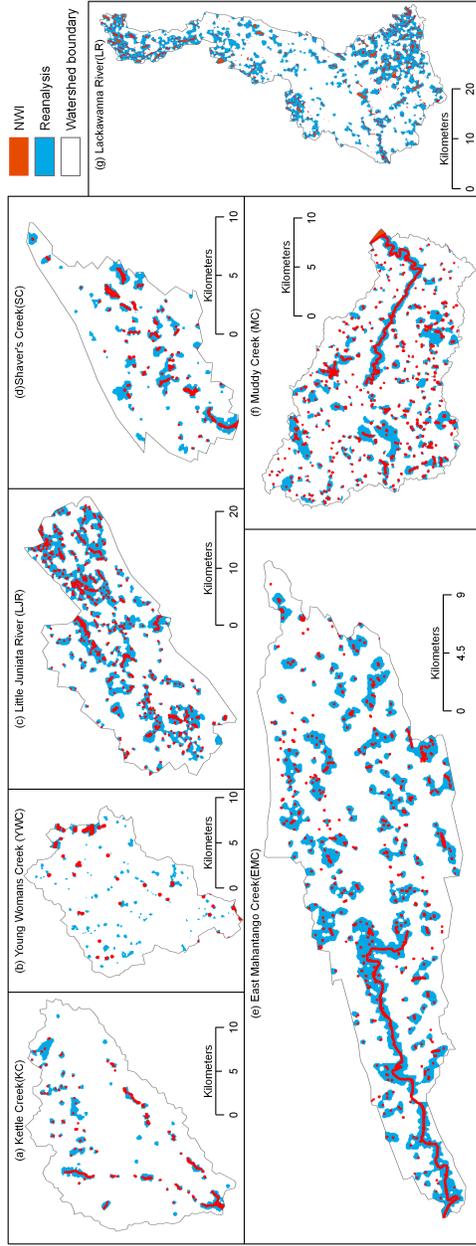


Figure 4.7. NWI locations (red dots) and PIHM simulated shallow water table conditions during the reanalysis period.

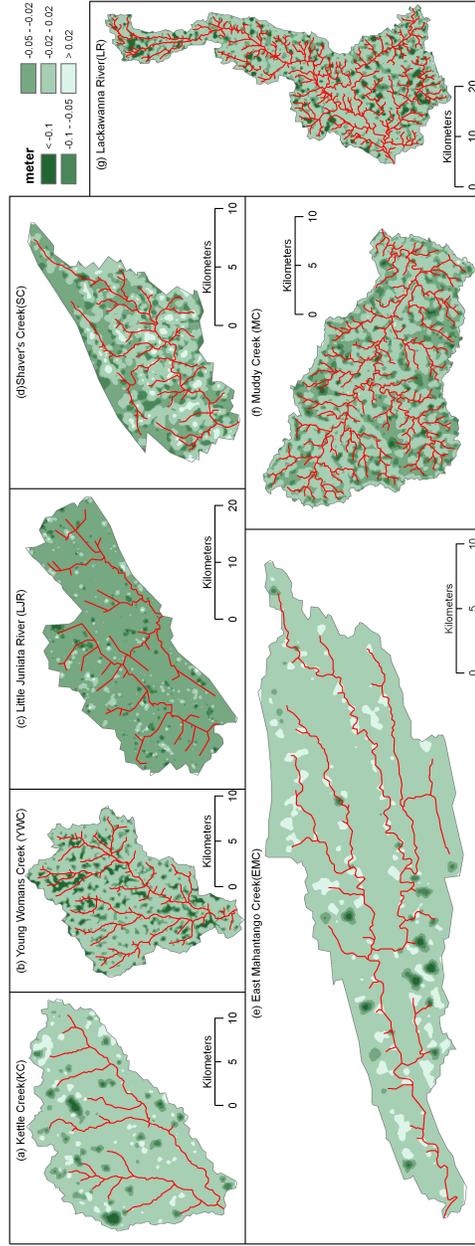


Figure 4.8. Difference maps for the simulated groundwater depth between the climate scenario period 2046-2065 and the reference period 1979-1998.

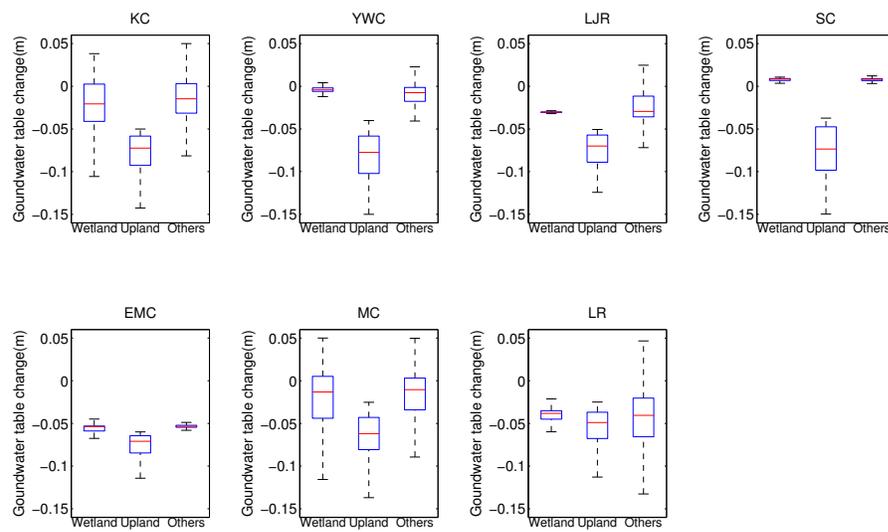


Figure 4.9. Distribution of groundwater table depth responses in the wetland, upland, and the rest of the area of the watershed.

A negative change indicates the region is dryer under the climate change projection scenario.

4.5 Discussion

In this study, national wetland geospatial data (NWI) was utilized for wetland hydrology simulation across 7 watersheds in SRB. The study developed historical simulations (1979-present) resulting in an online GIS watershed data set and wetland database containing information about the physical hydrological dynamics in each wetland area. The objective of the second part of this work was to assess the impact of climate change using the IPCC climate projections (2045-2065) on all wetlands in the NWI located within the 7 catchments. We expect that the future wetland projections developed in this study will also be subject to human-induced activities related to changes in land cover and land use. Simulation with scenarios that include land cover-use dynamics is an important area of future study.

4.5.1 How Does Physics-based Hydrologic Modeling Explain Groundwater and Stream Water Interaction?

Physics-based hydrologic modeling resolves the whole watershed into landscape mesh elements. The water movement equations govern the fully coupled, dynamic hydrological processes across the entire domain, providing insight into the spatial sources of runoff. By considering groundwater table dynamics in the calibration process, the model can better simulate surface and near-surface hydrological process [Ebel & Loague(2006)]. Additionally, spatial information from the NWI map further refines model performance.

Another assumption used in the model development was the uniform subsurface layer for all the watersheds. A 5-meter depth was assumed as the bottom boundary of shallow groundwater flow throughout the whole watershed. Other studies have suggested that most of the active flow zone in the SRB is relatively shallow ([Fan et al.(2013)] suggest less than 10 meters). The assumed 5-meter depth may be a minimum but proved to be sufficient to handle the recharge and baseflow processes in the watershed.

4.5.2 Differences between NWI and Hydrologic Identification

The NWI maps are prepared through conventional photointerpretation techniques as well as through field checks [Tiner(2002)]. Studies have shown that if an NWI map indicates the presence of a wetland in a given area, it is highly likely that a wetland is located there [Johnston & Meysembourg(2002), Kudray & Gale(2000), Maxa & Bolstad(2009)]. It has also been shown that unmapped wetlands, particularly in favorable landscape positions such as along stream in narrow valleys or in depressions, do exist [Tiner(2002)]. Catchment reanalysis data preserved the spatial groundwater table depth pattern and also provided a likely mapping for exploring missing wetlands in the NWI, though the model results from reanalysis may also tend to overestimate the existence of wetlands. The conclusion then is that both the model and the NWI maps tend to err more by omission than by commission [Tiner(2002)].

4.5.3 Groundwater Level Sensitivity to Climate Change

Average groundwater level response to climate change was subtle due to the combined effects of increased precipitation and increased evapotranspiration. The advantage of distributed modeling is that it enables an analysis of spatial location: wetland, upland, and the others. In general, the model results suggest that upland groundwater levels and thus upland wetlands are more sensitive to the IPCC climate change projections. It will be important in the future to expand the research to include land-use change as well as the overall impact on water and ecosystem services.

4.6 Conclusion and Outlook

In this study, the first mesoscale-scale watershed wetland assessment of groundwater table dynamics under climate change using a physics-based hydrological model PIHM is presented for high-resolution geospatial data, climate

reanalysis, and IPCC projections. A special focus of the study was on national data sources, geospatial model behavior constraints, wetland hydrology interpretation, and watershed hydrological reanalysis. This assessment is especially important for data and model applications in the environment of national requirements [Duffy et al.(2011)]. Furthermore, temporal and spatial wetland dynamics modeling using hydrological modeling-driven method provides an alternative for wetland hydrology research.

The wetland maps in these watersheds were generated through physics-based, distributed hydrological modeling. To summarize, shallow groundwater system modeling was able to adequately reproduce the temporal and spatial wetland hydrology dynamics in the model as well as the near surface hydrological processes and channel streamflow processes. The shallow groundwater demonstrated complex responses to the climate change, where upland groundwater levels decrease significantly under the climate scenario 2046-2065.

The inherent uncertainty related to meteorological forcing, watershed physical representation, bedrock depth, and the climate scenario itself remains an important area of future work. In addition, land-use dynamics should be incorporated in the modeling framework. In order to improve our model representations, it will be necessary to have higher-resolution data with improved strategies for parameterization.

Modeling Dynamic Ecosystem Processes: A Case for Improving Hydrological Predictability

5.1 Introduction

Terrestrial water and nutrient cycling are generally described as interacting hydrological and biogeochemical processes, with various assumptions about the degree and rate of process coupling. For example, biogeochemical processes are often assumed to be relatively slow in comparison to rainfall-runoff dynamics, and vegetation is often assumed to be fixed or to have only seasonal changes in the energy-water conditions and negligible interannual variability. These assumptions are adequate for predictions during rain events where short-term changes in vegetation are unlikely. However, for longer-term hydrologic predictions, the role of vegetation and soil biophysical contributions may play a more dominant role [Brolsma & Bierkens(2007), Miller et al.(2010), Smucker & Hopmans(2007)].

The current state of the Penn State Integrated Hydrologic Model (PIHM) was developed for a seasonally fixed canopy that uses the National Land Cover Data (NLCD, [Homer et al.(2007)]) and climate forcing reference parameters [NLDAS(1999)]. The basic parameters used for the calculation of vegetation water use are: the leaf area index (LAI) and the roughness length (RL). Two additional

parameters, the field capacity and wilting point, are also estimated to calculate the vegetation water use. Note that LAI is a measure of canopy biomass used to predict photosynthetic primary production and evapotranspiration and is a reference tool for plant growth. Although the fixed-season strategy used in PIHM is simple and captures some of the important seasonal dynamics necessary for water balance studies, the approach has several drawbacks. Namely: 1) the approach does not capture annual variations in the emergence (phenology) and senescence of forest and plant growth during the growing season; and 2) fixed LAI cannot by definition capture inter annual or longer time scale vegetation variations from growth and senescence. Both of these factors may be important to long term water balances in catchments.

At Shale Hills, during streamflow observations the 1990s drought, late spring and summer flow in the outlet weir completely dried up and did not recover until late in the fall, even though the stream flowed all summer during a study in the 1970s [Lynch & Corbett(1985)]. This difference might be explained by a reduction in precipitation during the 1990s or by a significant inter annual, or even decadal variability of tree water use. In fact, both of these explanations can be tested by implementing a dynamic vegetation model, rather than using a fixed seasonal pattern of vegetation parameters. [Wolf et al.(2008)] linked an ecosystem model with a hydrological model to test the sensitivity of the ecosystem model to hydrology and temperature. The results demonstrated that soil moisture and soil temperature are the most sensitive driving factors of carbon fluxes, particularly of soil carbon emissions. [Kiniry et al.(2008)] implemented a field-scale plant model to improve hydrologic transport modeling results of SWAT (Soil Water Assessment Tool). [Peng et al.(2013)] coupled a vegetation model with a distributed hydrologic model to assess future ecohydrological responses of climate change. These studies demonstrated the role of dynamic vegetation modeling in the study of hydrologic cycles.

Previous chapters discussed a method to constrain parameters for watershed modeling. This chapter links the physics-based watershed model (PIHM) and an ecophysiological model (Biome-BGC) to gain insight into the dynamic role of vegetation for hydrologic modeling. The chapter deals with the long-term changes in the water cycle by vegetation and will (a) demonstrate how biogeochemical

models and vegetation dynamics might resolve the inter annual hydrologic variability of the catchment and (b) evaluate how dynamic vegetation regulates the interaction between water and carbon cycles.

5.2 Overview of Biome-BGC

Biome-BGC is a one-dimensional model describing the carbon (C), nitrogen (N), and water (H₂O) states and fluxes of a plant functional type (PFT). The PFTs generally include: evergreen needle leaf forest (ENF), shrub, deciduous broad leaf forest (DNF), deciduous Broad leaf forest (DBF), C3 grass (C3G), and C4 grass (C4G). In this study we used Biome-BGC version 4.1.2, provided by Peter Thornton at the National Center for Atmospheric Research (NCAR) and by the Numerical Terradynamic Simulation Group (NTSG) at the University of Montana. The processes simulated in Biome-BGC include photosynthesis, respiration (autotrophic and heterotrophic), evapotranspiration, decomposition, the final allocation of photosynthetic assimilate and mortality [Running & Hunt(1993)]. Biome-BGC first models the phenology of the systems based on the input meteorological data. Then the rest of the processes are simulated at daily and annual scales. Daily processes related to water flux include interception, evaporation, transpiration, infiltration, snow melt, and outflow. Daily processes related to the carbon cycle include photosynthesis, leaf, stem, and root respiration, and respiration of soil and leaf litter [Running & Hunt(1993)]. The model simulates annual processes including storage and allocation of free carbon, and allocation and loss of free nitrogen. The allocation of carbon content and metabolism is parameterized by the ratio between leaves, stems, and roots, and depends on the specific plant. Free nitrogen enters the soil and litter. Then the mineralized N is lost from the system either through leaching when there is outflow or through bulk denitrification at a constant rate.

Common name	Scientific name	Species	Stem number	Total basal area(m^2)
Sugar maple	<i>Acer saccharum</i>	DBF	122	8.40
Pignut hickory	<i>Carya glabra</i>	DBF	122	7.75
Mockernut	<i>Carya tomentosa</i>	DBF	178	13.31
Eastern white pine	<i>Pinus strobus</i>	ENF	59	7.02
Virginia pine	<i>Pinus virginiana</i>	ENF	96	7.39
White oak	<i>Quercus alba</i>	DBF	282	26.77
Chestnut oak	<i>Quercus prinus</i>	DBF	558	47.50
Red oak	<i>Quercus rubra</i>	DBF	287	39.25
Eastern hemlock	<i>Tsuga canadensis</i>	ENF	164	15.76
Other			249	25.08

DBF: deciduous broadleaf forest; ENF: evergreen needleleaf forest.

Data were retrieved from [Eissenstat(2008), Meinzer et al.(2012)]

Table 5.1. Stand characteristics at SSHCZO.

5.3 Biome-BGC Modeling at SSHCZO

Vegetation at SSHCZO is mixed forest at mature status. Major species include *Quercus prinus*, *Quercus rubra*, *Quercus alba*, *Tsuga canadensis*, *Carya tormentosa*, *Acer saccharum*, *Carya glabra*, *Pinus strobus*, and *Pinus virginiana* [Meinzer et al.(2012), Naithani et al.(2013)]. The tree survey results are shown in Table 5.1. The parameters (Table 5.2) for Biome-BGC modeling are obtained from [White et al.(2000)]. The modeling results are compared with flux tower estimated NEE (Figure 5.1). The modeling meteorological data is from 1904 to 2013. The histogram of phenology and vegetation dynamics demonstrates significant inter-annual variability (Figure 5.2, Figure 5.3).

Parameter	Chestnut Oak	Red Oak	White Oak	Unit
transfer growth period as fraction of growing season	0.2	0.2	0.2	
litter-fall as fraction of growing season	0.2	0.2	0.2	
annual leaf and fine root turnover fraction	1	1	1	year ⁻¹
annual live wood turnover fraction	0.54	0.54	0.54	year ⁻¹
annual whole-plant mortality fraction	0.003	0.003	0.003	year ⁻¹
annual fire mortality fraction	0.001	0.001	0.001	year ⁻¹
new fine root C:new leaf C	0.1	0.1	0.1	
new stem C:new leaf C	1.32	1.32	1.36	
new live wood C:new total wood C	0.15	0.15	0.279	
new coarse root C:new stem C	0.1	0.1	0.1	
current growth proportion	0.6	0.6	0.6	
C:N of leaves	35	33.1	27.2	kgCkgN ⁻¹
C:N of leaf litter	98.7	61	62.5	kgCkgN ⁻¹
C:N of fine roots	42	42	42	kgCkgN ⁻¹
C:N of live wood	50	50	50	
C:N of dead wood	742	479	451	
leaf litter labile proportion	0.39	0.308	0.324	
leaf litter cellulose proportion	0.36	0.425	0.474	
leaf litter lignin proportion	0.25	0.267	0.202	
fine root labile proportion	0.3	0.3	0.3	
fine root cellulose proportion	0.45	0.45	0.45	
fine root lignin proportion	0.25	0.25	0.25	
dead wood cellulose proportion	0.76	0.75	0.76	
dead wood lignin proportion	0.24	0.25	0.24	
canopy water interception coefficient	0.001	0.001	0.001	LAI ⁻¹ d ⁻¹
canopy light extinction coefficient	0.4	0.4	0.4	
all-sided to projected leaf area ratio	2	2	2	
canopy average specific leaf area	19.9	26.2	20.4	m ² kgC ⁻¹
ratio of shaded SLA: sunlit SLA*	2	2	2	
fraction of leaf N in Rubisco	0.2	0.2	0.2	
maximum stomatal conductance	0.006	0.006	0.006	ms ⁻¹
cuticular conductance	0.00001	0.00001	0.00001	ms ⁻¹
boundary layer conductance	0.005	0.005	0.005	ms ⁻¹
leaf water potential: start	-0.3	-0.3	-0.2	MPa
leaf water potential: complete	-2	-2.2	-2	MPa
VPD: start	169	169	169	Pa
VPD: complete	2100	2100	2100	Pa

Table 5.2. Eco-physiological parameters for Biome-BGC at SSHCZO.

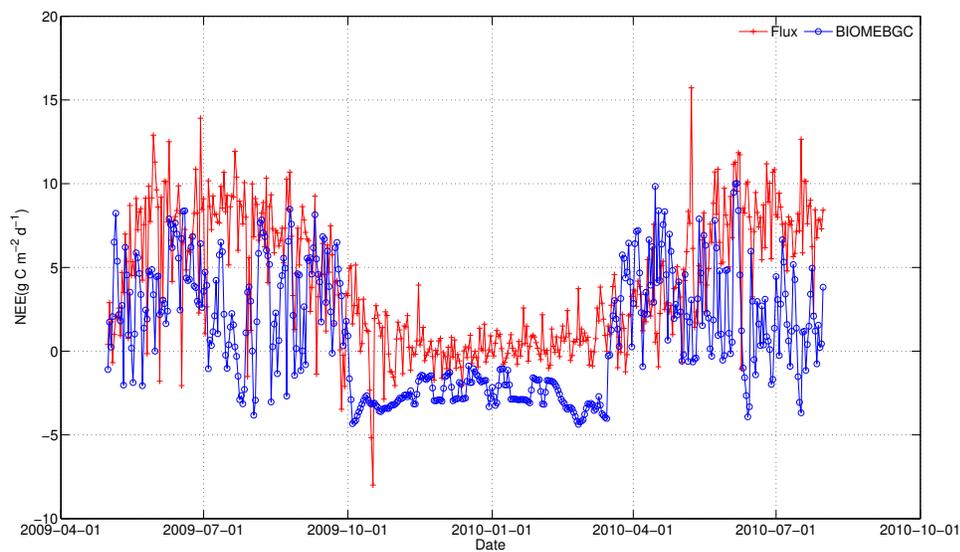


Figure 5.1. Biome-BGC simulated NEE compared with flux tower estimation. The dynamics of NEE in spring are well captured by Biome-BGC.

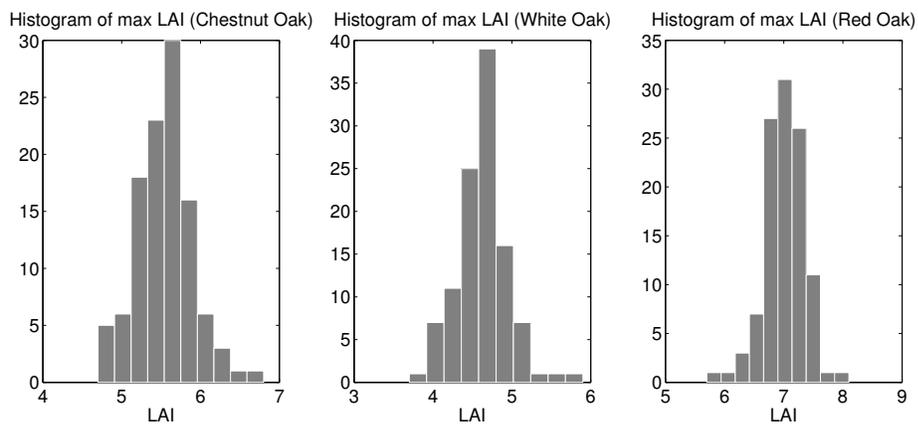


Figure 5.2. Biome-BGC simulated inter-annual variability of max LAI of 3 dominant tree species at Shale Hills. The results suggest that due to meteorological variability from 1904 to 2013, the annual max LAI has significant variability.

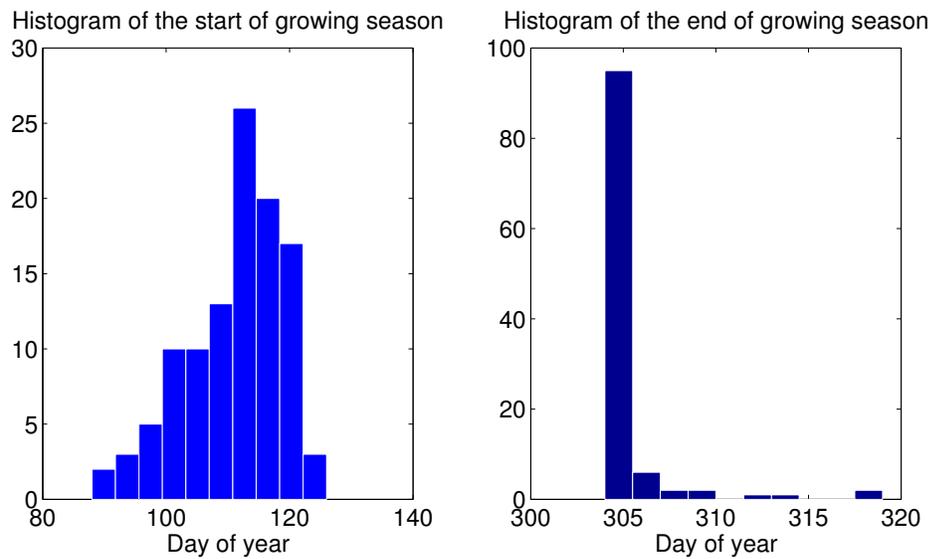


Figure 5.3. Biome-BGC simulated inter-annual variability of phenology. For deciduous woody plants, the start of growing season is determined by the precipitation and temperature, and the end of growing season is determined by the day length and temperature. The results in 1904 to 2013 suggested that the start of growing season has significant variability; however, the end of growing season is the 305th day of the year in most years.

5.4 Modeling Vegetation Dynamic for PIHM Simulation

The Forest Hydrology group at The Pennsylvania State University conducted a series of hydrologic irrigation experiments at The Shale Hills in 1974 [Lynch & Corbett(1985), Lynch(1976)]. The objectives of the experiments were to determine the physical mechanisms of runoff and streamflow generation at the upslope forested watershed and to evaluate the effects of antecedent soil moisture on the runoff peak and timing. Recently, a program of research using Earths Critical Zone Observatories (CZOs) has been initiated, and Shale Hills is one the CZOs: the Susquehanna-Shale Hills Critical Zone Observatory (SSHCZO), which focuses on hydrologic flow paths and timescales, as well as the regolith formation and ecosystem dynamics within a small, forested catchment. The modeling studies mainly focused on the seasonal scales, such as land surface energy processes [Shi et al.(2013), Shi et al.(2014b)] and antecedent moisture condition [Qu & Duffy(2007)]. The inter-annual variability of vegetation dynamics could be resolved by nudging vegetation parameters according the water budget of the modeling period. For the long-term modeling, the dynamic vegetation simulation will resolve the inter-annual variability of plant growth in way of physical meaning. Here, the continuous simulation of PIHM from 1974 to 2013 was conducted to test the role of dynamic vegetation in the modeling of long-term hydrological cycles.

By default, PIHM uses NLDAS monthly vegetation parameters [NLDAS(1999)]: based on UMD classification [Hansen et al.(2000)]. Then, the vegetation parameters are mapped to NLCD 2006 classification [Bhatt(2012)]. Clearly, Biome-BGC simulated results imply that vegetation phenology is dependent on species and the surrounding environment (e.g., soil temperature, day length). Hence, I linked Biome-BGC and PIHM through vegetation parameters to test the impact of dynamic vegetation parameters in PIHM simulation.

PIHM simulations were forced by default LAI and Biome-BGC simulated LAI to simulate hydrologic response of the irrigation experiment in 1974 [Lynch(1976)]. The result suggested that phenology could lead to major errors in simulated peak flow of PIHM (Figure 5.4).

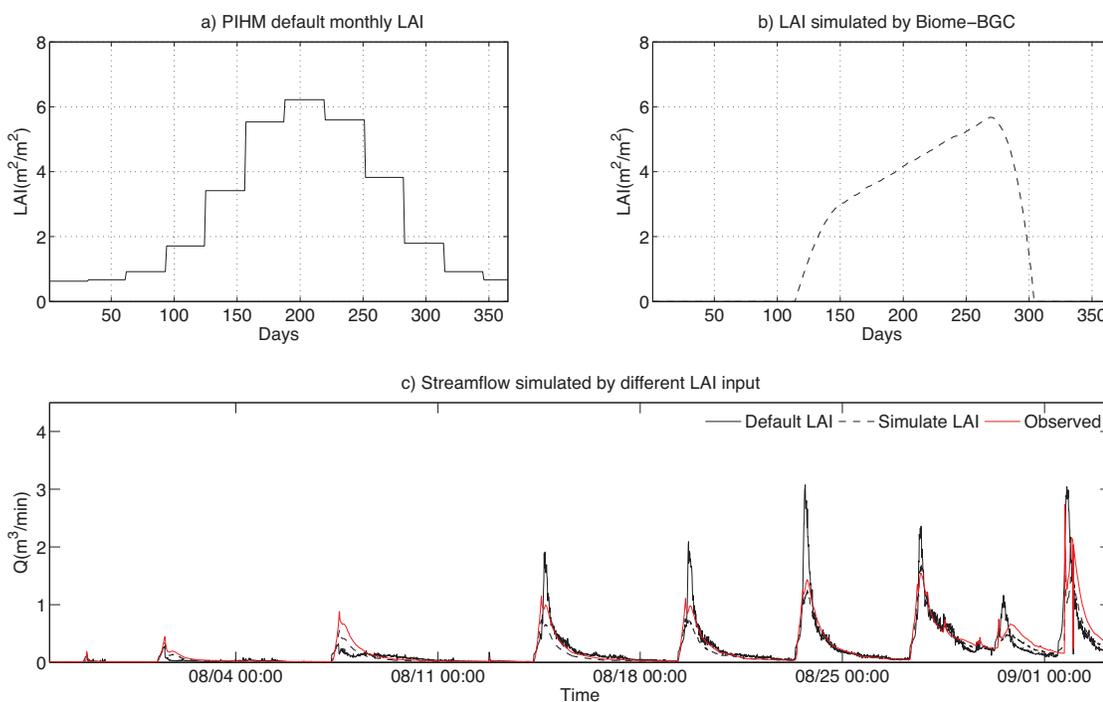


Figure 5.4. PIHM-modeled streamflow sensitivity to LAI (default LAI and modeled LAI). Note that using the fixed-seasonal LAI tends to overestimates the peakflows later in the 1974 experiment.

5.5 Implications on Further Coupling of Biome-BGC and PIHM

Further coupling schemes should include hydrological processes coupling (Figure 5.5) and lateral redistribution of carbon and nitrogen. The water storage and fluxes have impacts on the vegetation growth, and the lateral routing of subsurface flow plays a key role in the redistribution of nutrients. The challenge is to resolve multi-scale ecosystem processes appropriately in a hydrologic framework. Here, we used the distributed water cycle from the simulation results of PIHM to force the modeling processes of Biome-BGC. The results showed the spatial controls of vegetation, soil properties and topography on the soil carbon (Figure 5.5).

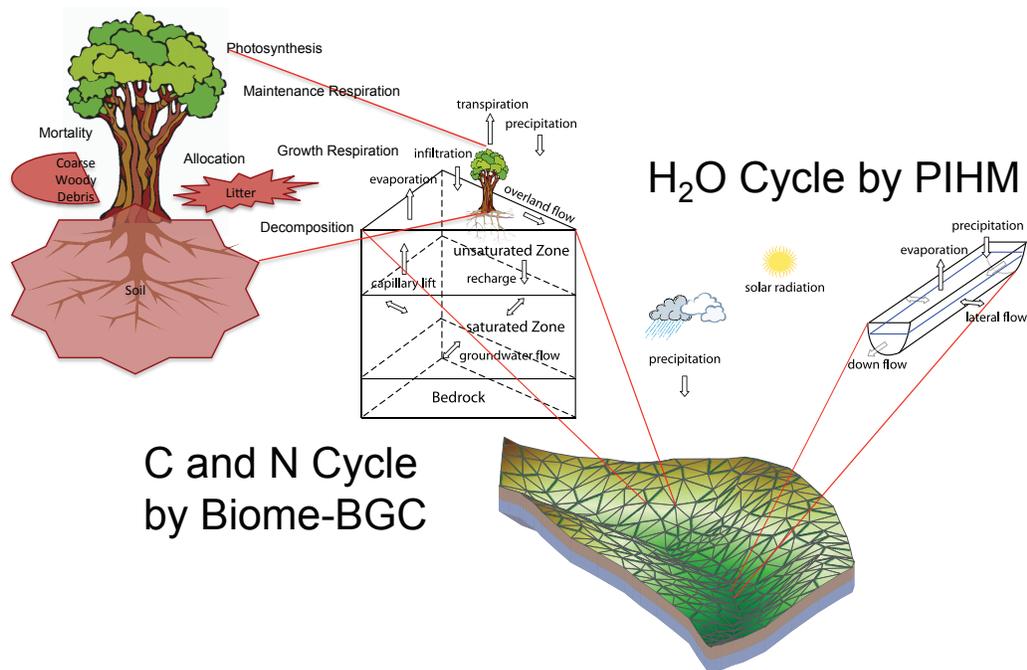


Figure 5.5. Flow chart of hydrological coupling between PIHM and Biome-BGC.

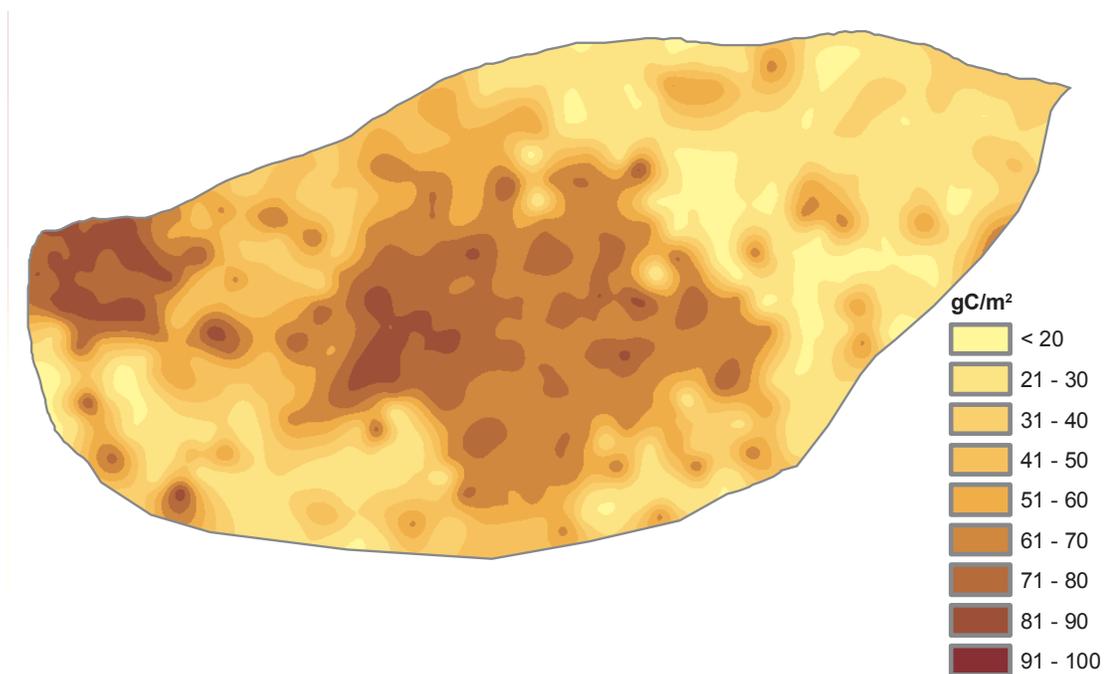


Figure 5.6. Modeled spatial pattern of soil carbon at Shale Hills.

5.6 Conclusion

In this study, the PIHM and Biome-BGC models were linked through vegetation dynamics and tested at a headwater catchment. The vegetation dynamics at Shale Hills were simulated through Biome-BGC, and then the simulated daily LAI was used in the long-term hydrologic simulation by PIHM. Simulated peakflow and low-flows were improved in the linked model. The result of streamflow suggested that inter annual vegetation dynamics and succession could have significant impacts on watershed hydrology. This test was successful in demonstrating the importance of dynamic vegetation in watershed models. In addition, the spatial pattern of the carbon dynamic was modeled by coupling the distributed water cycle with Biome-BGC. Further coupling of hydrologic processes needs to include the important effects of multi-scale ecosystem processes that emphasize key connections between watershed hydrology and vegetation dynamics. First, spatial and temporal variability in soil water has impact on vegetation growth. Second, the lateral water movement determines the nutrient redistribution within the watershed, and future work should fully couple this factor in the PIHM Transport code. Third, the vegetation dynamics in Biome-BGC could improve the plant-water-use calculation in PIHM, improving the peak-flow simulation and related hydrologic predictions.

Summary and Future Work

6.1 Summary

Three key objectives have been achieved in this research: (a) a new strategy for estimation of parameters in the physics-based, fully-coupled watershed model PIHM; (b) application of the parameter estimation strategy to watershed and wetland modeling for the climate reanalysis period 1979-2009 and IPCC climate projection period 2045-2065; (c) an assessment of the importance of dynamic LAI in PIHM on the runoff response in the Shale Hills CZO.

Integrated environmental models seek to simulate coupled environmental processes and to assess management practices. As the model complexity and parameter space increase, the calibration of such models presents increased challenges to users and requires considerable modeling experience. In this research, different calibration cases demonstrate that incorporation of available data in calibration can effectively increase the simulation results of multi-scale, multi-process environmental systems, which will have important implications for watershed management.

The calibration of PIHM used a time-scale partition strategy: event-scale group and seasonal time-scale group. The event-based group (EG) parameters were calibrated first using the CMA-ES optimization algorithm targeting runoff responses, followed by the seasonal group (SG) calibration targeting water budget. A general conclusion of this study is that the 2-scale partition for parameter estimation according to the dominant time scales of the system provides a useful way to isolate dynamic processes and integrate detailed behavior within the model.

Validation was extended without degrading the model performance.

Physics-based models simulate not only discharge rate at outlet but also spatial patterns of soil moisture, evapotranspiration, and groundwater table. In Chapter 3, the weighting strategy between multiple calibration targets is illustrated. First, the informativeness of each calibration target is determined by the model-performance correlation. Second, the weighted objective function is formulated based on the informativeness from the previous step. Third, the weighted function is used to constrain the model with multiple observations. The application at SSHCZO suggested a satisfactory compromise between streamflow, water table depth and ET was achieved with the weighting strategy.

Another spatial data application is the freshwater wetland modeling study in Chapter 4. This study demonstrates the use of spatial maps for modeling near-surface wetland hydrology. The model was constrained not only by the time serial of streamflow, but also by the spatial pattern of the groundwater table, which utilized the information from the National Wetland Inventory map. Expectedly, the wetland derived from a simulated spatial groundwater pattern was consistent with the NWI. Based on the well-calibrated modeling system, climate change impacts on shallow groundwater were evaluated. The simulation results suggested heterogeneous responses to climate change: the decrease of upland groundwater level will be more significant under the climate scenario 2046-2065 than the decrease of wetland groundwater level.

Chapter 5 is the long-term modeling study to test the capability of the vegetation module. The default PIHM vegetation dynamics were represented by monthly parameters repeated year by year. This method of representation could cause model discrepancy for the inter-annual variability of vegetation dynamics. Hence, the dynamic vegetation processes was reproduced by an ecosystem process model: Biome-BGC. The simulation, by linking Biome-BGC and PIHM, suggested improved streamflow prediction during the growing season.

In addition to the four papers presented in this thesis, the C code package PIHM-PCS is developed. PIHM-PCS targets PIHM parameter optimization with both the single objective of streamflow and weighted objective of multiple observations. Parameterization and calibration make up a large part of the hydrological modeling literature. PIHM-PCS demonstrates an idea to simplify the

calibration according to the model and data. The processes-and-parameters partition divides calibration into several steps in a sequential workflow. When the simulation target contains different kinds of measured variables, the informativeness-based, weighted-objective function provides an idea of partitioning the simulation target into quantitative evaluation of each variable within the target. The study case at SSHCZO demonstrated a framework used to calibrate a high-resolution, distributed hydrologic simulation of a catchment where multiple observations of hydrological variables are recorded to constrain the understanding of hydrological processes. Currently, the PIHM-PCS has been applied in many other watershed modeling studies, and the results have demonstrated the robustness of PIHM-PCS.

6.2 Future Work

In Chapter 2, a method was developed for distributed parameter estimation that takes advantage of time-scale separation in governing processes. It was found that parameters could be divided into one group that was sensitive to hydrologic events and another group controlling seasonal energy dynamics. However, in some cases, different combinations of parameters could generate the same runoff responses. Future study should evaluate parameter sensitivity and uncertainty.

The weighting strategy presented in Chapter 3 provides an efficient method for handling multiple constraints of model behavior. The informativeness was defined to understand the importance of observation in constraining model parameters, and the model performance correlations are used for the calculation of informativeness. The results suggested an informativeness order: streamflow > downslope groundwater table > upslope groundwater table > evapotranspiration. To further improve the applicability of informativeness, a study of information conflict and redundancy should be considered when the model is constrained by increasing observation.

Modeling shallow groundwater dynamics for wetland study demonstrated the need for a distributed watershed model in wetland hydrologic research. The focus of the study was on data availability, model behavior constraints, wetland hydrologic interpretation, and watershed hydrological reanalysis. It remains unknown if we could confidently improve the NWI by the PIHM predicted wetland area and how

to understand the assessment results of the groundwater table responses to the future climate. Uncertainty quantification is a critical component in the description and attribution of the wetland hydrology.

Chapter 5 demonstrates the importance of vegetation dynamics and the improved hydrologic response when linking Biome-BGC and PIHM. In this study, it was shown that the vegetation dynamics is an important element in both hydrologic cycles. In the future, a fully coupled vegetation-biogeochemical-watershed model will support hydro-ecological predictions, including diversity distributions of the types of regional vegetation, and control net primary productivity and carbon cycling. On the other hand, moisture distribution in the water cycle is significantly impacted by vegetation via hydrological processes, such as interception and evapotranspiration. There is a need for the development of a coupled, dynamic vegetation module for understanding the water and carbon cycles.

Bibliography

- [Abrahart et al.(2011)] Abrahart, R. J., Mount, N. J., Ab Ghani, N., Clifford, N. J. & Dawson, C. W. (2011) DAMP: a protocol for contextualising goodness-of-fit statistics in sediment-discharge data-driven modelling. *Journal of Hydrology* 409(3):596–611.
- [Amthor et al.(2001)] Amthor, J. S., Chen, J. M., Clein, J. S., Frohling, S. E., Goulden, M. L., Grant, R. F., Kimball, J. S., King, A. W., McGuire, A. D. & Nikolov, N. T. (2001) Boreal forest CO₂ exchange and evapotranspiration predicted by nine ecosystem process models: Intermodel comparisons and relationships to field measurements. *J. Geophys. Res* 106(33):623–633.
- [Anderson et al.(2006)] Anderson, R. M., Koren, V. I. & Reed, S. M. (2006) Using SSURGO data to improve sacramento model a priori parameter estimates. *Journal of Hydrology* 320(1-2):103–116.
- [Anderson et al.(2008)] Anderson, S. P., Bales, R. C. & Duffy, C. J. (2008) Critical zone observatories: Building a network to advance interdisciplinary study of earth surface processes. *Mineralogical Magazine* 72(1):7.
- [Asbjornsen et al.(2011)] Asbjornsen, H., Goldsmith, G. R., Alvarado-Barrientos, M. S., Rebel, K., Van Osch, F. P., Rietkerk, M., Chen, J., Gotsch, S., Tobon, C. & Geissert, D. R. (2011) Ecohydrological advances and applications in plant-water relations research: a review. *Journal of Plant Ecology* 4(1-2):3–22.
- [Banwart et al.(2011)] Banwart, S., Bernasconi, S. M., Bloem, J., Blum, W., Brandao, M., Brantley, S., Chabaux, F., Duffy, C., Kram, P., Lair, G., Lundin, L., Nikolaidis, N., Novak, M., Panagos, P., Ragnarsdottir, K. V., Reynolds, B., Rousseva, S., Ruiter, P., Gaans, P., Riemsdijk, W., White, T. & Zhang, B. (2011) Soil processes and functions in critical zone observatories: hypotheses and experimental design. *Vadose Zone Journal* 10(3):974–987.

- [Barthel et al.(2012)] Barthel, R., Reichenau, T. G., Krimly, T., Dabbert, S., Schneider, K. & Mauser, W. (2012) Integrated modeling of global change impacts on agriculture and groundwater resources. *Water Resources Management* pp. 1–23.
- [Bastina et al.(2013)] Bastina, L., Cornforda, D., Jonesa, R., Heuvelinkb, G. B., Pebesmac, E., Staschc, C., Nativid, S., Mazzettid, P. & Williamsa, M. (2013) Managing uncertainty in integrated environmental modelling: The uncertweb framework. *Environmental Modelling & Software* 39:116134.
- [Bayer & Finkel(2004)] Bayer, P. & Finkel, M. (2004) Evolutionary algorithms for the optimization of advective control of contaminated aquifer zones. *Water Resources Research* 40(6):W06506.
- [Bhat et al.(2010)] Bhat, M., Haran, M. & Goes, M. (2010) Computer model calibration with multivariate spatial output: a case study. In: Chen, M. H., Muller, P., Sun, D., Y. K. & Dey, D. (eds.), *Frontiers of statistical decision making and bayesian analysis*. Springer-Verlag, New York.
- [Bhatt(2012)] Bhatt, G. (2012) A distributed hydrologic modeling system: framework for discovery and management of water resources. Ph.D. thesis, Pennsylvania State University.
- [Bhatt et al.(2006)] Bhatt, G., Kumar, M., Kamath, R. M. & Duffy, C. J. (2006) PIHMgis: a "Tightly coupled" GIS framework for integrated hydrologic modeling. In: *AGU Fall Meeting Abstracts*. vol. 1, p. 0418.
- [Bhatt et al.(2013)] Bhatt, G., Yu, X., Duffy, C., Leonard, L. & Kumar, M. (2013) Development of multi-scale, multi-state application of physics-based distributed hydrologic model in the chesapeake bay watershed. In: *22nd Biennial Conference of The Coastal And Estuarine Research Federation*, San Diego, CA.
- [Boyle et al.(2000)] Boyle, D. P., Gupta, H. V. & Sorooshian, S. (2000) Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods. *Water Resources Research* 36(12):3663–3674.
- [Brantley(2008)] Brantley, S. L. (2008) Observations emerging from a network of critical zone observatories: Shale weathering at the susquehanna-shale hills observatory. In: *2008 Joint Meeting of The Geological Society of America, Soil Science Society of America, American Society of Agronomy, Crop Science Society of America, Gulf Coast Association of Geological Societies with the Gulf Coast Section of SEPM*.

- [Brolsma & Bierkens(2007)] Brolsma, R. J. & Bierkens, M. F. P. (2007) Groundwater-soil water-vegetation dynamics in a temperate forest ecosystem along a slope. *Water resources research* 43(1):W01414.
- [Brooks & Wardrop(2013)] Brooks, R. P. & Wardrop, D. (2013) *Mid-Atlantic Freshwater Wetlands: Advances in Wetlands Science, Management, Policy, and Practice*. Springer London, Limited.
- [Carr et al.(2013)] Carr, A. E., Loague, K. & VanderKwaak, J. (2013) Hydrologic response simulations for the north fork of Caspar creek: second growth, clearcut, new growth, and cumulative watershed effect scenarios. *Hydrological Processes* .
- [Chen & Dudhia(2001)] Chen, F. & Dudhia, J. (2001) Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. part I: Model implementation and sensitivity. *Monthly Weather Review* 129(4):569–585.
- [Coe et al.(2000)] Coe, M. T., Lenters, J. D., Young-Molling, C. & Ramankutty, N. (2000) Testing the performance of a dynamic global ecosystem model: Water balance, carbon balance, and vegetation structure. *Global Biogeochemical Cycles* 14(3):795–825.
- [Cohen & Hindmarsh(1994)] Cohen, S. & Hindmarsh, A. (1994) CVODE user guide, Lawrence Livermore National Laboratory technical report UCRLMA118618. Tech. rep.
- [Cohen & Hindmarsh(1996)] Cohen, S. D. & Hindmarsh, A. C. (1996) CVODE, a stiff/nonstiff ODE solver in C. *Computers In Physics* 10(2):138–143.
- [Council(1995)] Council, N. R. (1995) *Wetlands characteristics and boundaries*. National Academy Press.
- [Davis(2010)] Davis, K. (2010) CZO dataset: Shale hills - flux tower (2009-2010).
- [Dawson et al.(2007)] Dawson, C., Abraham, R. & See, L. (2007) HydroTest: a web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. *Environmental Modelling & Software* 22(7):1034–1052.
- [Demaria et al.(2007)] Demaria, E. M., Nijssen, B. & Wagener, T. (2007) Monte Carlo sensitivity analysis of land surface parameters using the variable infiltration capacity model. *Journal of Geophysical Research* 112(D11):D11113.

- [diGiammarco et al.(1996)] diGiammarco, P., Todini, E. & P., L. (1996) A conservative finite elements approach to overland flow: the control volume finite element formulation. *Journal of Hydrology* 175:267–291.
- [Dingman(2002)] Dingman, S. L. (2002) *Physical hydrology*. Prentice Hall Upper Saddle River, New Jersey, USA.
- [Du et al.(2007)] Du, J., Xie, S., Xu, Y., Xu, C. & Singh, V. P. (2007) Development and testing of a simple physically-based distributed rainfall-runoff model for storm runoff simulation in humid forested basins. *Journal of Hydrology* 336(3-4):334–346.
- [Duan et al.(2006)] Duan, Q., Schaake, J., Andreassian, V., Franks, S., Goteti, G., Gupta, H. V., Gusev, Y. M., Habets, F., Hall, A. & Hay, L. (2006) Model parameter estimation experiment (MOPEX): an overview of science strategy and major results from the second and third workshops. *Journal of Hydrology* 320(1):3–17.
- [Duffy et al.(2011)] Duffy, C., Leonard, L., Bhatt, G., Yu, X. & Giles, L. (2011) Watershed reanalysis: Towards a national strategy for model-data integration. In: *e-Science Workshops (eScienceW)*, 2011 IEEE Seventh International Conference on. IEEE, pp. 61–65.
- [Duffy(2010a)] Duffy, C. J. (2010a) CZO dataset: Shale hills - groundwater depth data (2009-2012) .
- [Duffy(2010b)] Duffy, C. J. (2010b) CZO dataset: Shale hills - streamflow data (2006-2012) .
- [Duffy(2012)] Duffy, C. J. (2012) CZO dataset: Shale hills - precipitation (2006-2012) .
- [Dung et al.(2011)] Dung, N. V., Merz, B., Brdossy, A., Thang, T. D. & Apel, H. (2011) Multi-objective automatic calibration of hydrodynamic models utilizing inundation maps and gauge data. *Hydrology and Earth System Sciences* 15(4):1339–1354.
- [Dunn & Colohan(2010)] Dunn, S. M. & Colohan, R. J. E. (2010) Developing the snow component of a distributed hydrological model: a step-wise approach based on multi-objective analysis. *Journal of Hydrology* 223(1):1–16.
- [Ebel & Loague(2006)] Ebel, B. A. & Loague, K. (2006) Physics-based hydrologic-response simulation: Seeing through the fog of equifinality. *Hydrological Processes* .

- [Efstratiadis & Koutsoyiannis(2010)] Efstratiadis, A. & Koutsoyiannis, D. (2010) One decade of multi-objective calibration approaches in hydrological modelling: a review. *Hydrological Sciences Journal* 55(1):58–78.
- [Eissenstat(2008)] Eissenstat, D. (2008) Czo dataset: Shale hills - vegetation (2008) - tree survey. pp. <http://criticalzone.org/shale-hills/data/dataset/2648/>.
- [Fan et al.(2013)] Fan, Y., Li, H. & Miguez-Macho, G. (2013) Global patterns of groundwater table depth. *Science* 339(6122):940–943.
- [Fang et al.(2013)] Fang, X., Pomeroy, J. W., Ellis, C. R., MacDonald, M. K., DeBeer, C. M. & Brown, T. (2013) Multi-variable evaluation of hydrological model predictions for a headwater basin in the canadian rocky mountains. *Hydrol. Earth Syst. Sci* 17:1635–1659.
- [Foglia et al.(2009)] Foglia, L., Hill, M. C., Mehl, S. W. & Burlando, P. (2009) Sensitivity analysis, calibration, and testing of a distributed hydrological model using error-based weighting and one objective function. *Water Resour. Res.* 45:W06427.
- [Frei et al.(2009)] Frei, S., Fleckenstein, J. H., Kollet, S. J. & Maxwell, R. M. (2009) Patterns and dynamics of river-aquifer exchange with variably-saturated flow using a fully-coupled model. *Journal of Hydrology* 375(3):383–393.
- [Grayson et al.(2002)] Grayson, R. B., Blöschl, G., Western, A. W. & McMahon, T. A. (2002) Advances in the use of observed spatial patterns of catchment hydrological response. *Advances in Water Resources* 25(8):1313–1334.
- [Green et al.(2011)] Green, T. R., Taniguchi, M., Kooi, H., Gurdak, J. J., Allen, D. M., Hiscock, K. M., Treidel, H. & Aureli, A. (2011) Beneath the surface of global change: Impacts of climate change on groundwater. *Journal of Hydrology* 405(3):532–560.
- [Guo(2010)] Guo, Q. (2010) Susquehanna shale hills critical zone observatory: Leaf off survey .
- [Gupta et al.(1998)] Gupta, H. V., Sorooshian, S. & Yapo, P. O. (1998) Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resources Research* 34(4):751–763.
- [Hansen et al.(2013)] Hansen, A. L., Refsgaard, J. C., Christensen, B. S. B. & Jensen, K. H. (2013) Importance of including small-scale tile drain discharge in the calibration of a coupled groundwater-surface water catchment model. *Water Resour. Res.* 49:585–603.

- [Hansen et al.(2000)] Hansen, M., DeFries, R., Townshend, J. & Sohlber, R. (2000) Global land cover classification at 1km resolution using a decision tree classifier. *International Journal of Remote Sensing* 21:1331–1365.
- [Hansen(2006)] Hansen, N. (2006) The CMA evolution strategy: a comparing review. *Towards a new evolutionary computation* pp. 75–102.
- [Hansen et al.(2003)] Hansen, N., Müller, S. D. & Koumoutsakos, P. (2003) Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). *Evolutionary Computation* 11(1):1–18.
- [Hanson et al.(2004)] Hanson, P. J., Amthor, J. S., Wullschleger, S. D., Wilson, K. B., Grant, R. F., Hartley, A., Hui, D., Hunt, J., Johnson, D. W. & Kimball, J. S. (2004) Oak forest carbon and water simulations: model intercomparisons and evaluations against independent data. *Ecological Monographs* 74(3):443–489.
- [Hay et al.(2006)] Hay, L. E., Leavesley, G. H., Clark, M. P., Markstrom, S. L., Viger, R. J. & Umemoto, M. (2006) Step wise, multiple objective calibration of a hydrologic model for a snowmelt dominated basin. *Journal of the American Water Resources Association* 42(4):877–890.
- [Heppner et al.(2007)] Heppner, C. S., Loague, K. & VanderKwaak, J. E. (2007) Long term InHM simulations of hydrologic response and sediment transport for the r 5 catchment. *Earth Surface Processes and Landforms* 32(9):1273–1292.
- [Hogue et al.(2006)] Hogue, T. S., Gupta, H. & Sorooshian, S. (2006) A user-friendly approach to parameter estimation in hydrologic models. *Journal of Hydrology* 320(1):202–217.
- [Homer et al.(2007)] Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., Herold, N., McKerrow, A., VanDriel, J. N. & Wickham, J. (2007) Completion of the 2001 national land cover database for the counterminous united states. *Photogramm. Eng. Remote Sens.* 73(4):337–341.
- [Hsie et al.(2014)] Hsie, M., Yan, S. & Pan, N. (2014) Improvement of rainfall-runoff simulations using the runoff-scale weighting method. *Journal of Hydrologic Engineering* in press.
- [Huang(2014)] Huang, Y. (2014) Multi-objective calibration of a reservoir water quality model in aggregation and non-dominated sorting approaches. *Journal of hydrology* 510:280–292.

- [Huntington(2003)] Huntington, T. G. (2003) Climate warming could reduce runoff significantly in new england, usa. *Agricultural and Forest Meteorology* 117(3):193201.
- [Ichii et al.(2010)] Ichii, K., Suzuki, T., Kato, T., Ito, A., Hajima, T., Ueyama, M., Sasai, T., Hirata, R., Saigusa, N. & Ohtani, Y. (2010) Multi-model analysis of terrestrial carbon cycles in japan: limitations and implications of model calibration using eddy flux observations. *Biogeosciences* 7(7):2061–2080.
- [Ivanov et al.(2004)] Ivanov, V. Y., Vivoni, E. R., Bras, R. L. & Entekhabi, D. (2004) Catchment hydrologic response with a fully distributed triangulated irregular network model. *Water Resources Research* 40(11):W11102.
- [Johnson et al.(2004)] Johnson, W. C., Boettcher, S. E., Poiani, K. A. & Guntenspergen, G. (2004) Influence of weather extremes on the water levels of glaciated prairie wetlands. *Wetlands* 24(2):385398.
- [Johnston & Meysembourg(2002)] Johnston, C. A. & Meysembourg, P. (2002) Comparison of the wisconsin and national wetlands inventories. *Wetlands* 22(2):386405.
- [Jones et al.(2008)] Jones, J. P., Sudicky, E. A. & McLaren, R. G. (2008) Application of a fully-integrated surface-subsurface flow model at the watershed-scale: A case study. *Water Resources Research* 44(3):W03407.
- [Kampf & Burges(2007)] Kampf, S. K. & Burges, S. J. (2007) A framework for classifying and comparing distributed hillslope and catchment hydrologic models. *Water Resources Research* 43:W05423.
- [Kato et al.(2007)] Kato, H., Rodell, M., Beyrich, F., Cleugh, H., van Gorsel, E., Liu, H. & Meyers, T. P. (2007) Sensitivity of land surface simulations to model physics, land characteristics, and forcings, at four CEOP sites. *Journal of the Meteorological Society of Japan. Ser. II* 85:187–204.
- [Kerkez et al.(2012)] Kerkez, B., Glaser, S. D., Bales, R. C. & Meadows, M. W. (2012) Design and performance of a wireless sensor network for catchment-scale snow and soil moisture measurements. *Water Resources Research* 48(9):W09515.
- [Khu et al.(2008)] Khu, S. T., Madsen, H. & di Pierro, F. (2008) Incorporating multiple observations for distributed hydrologic model calibration: An approach using a multi-objective evolutionary algorithm and clustering. *Advances in Water Resources* 31(10):1387–1398.

- [Kim et al.(2012)] Kim, J., Warnock, A., Ivanov, V. Y. & Katopodes, N. D. (2012) Coupled modeling of hydrologic and hydrodynamic processes including overland and channel flow. *Advances in Water Resources* 37:104–126.
- [Kiniry et al.(2008)] Kiniry, J. R., MacDonald, J. D., ARMEN, R. K., Watson, B., Putz, G. & ELLIE, E. P. (2008) Plant growth simulation for landscape-scale hydrological modelling. *Hydrological Sciences Journal* 53(5):1030–1042.
- [Kløve et al.(2013)] Kløve, B., Ala-Aho, P., Bertrand, G., Gurdak, J. J., Kupfersberger, H., Kværner, J., Muotka, T., Mykrä, H., Preda, E., Rossi, P., Uvo, C. B., Velasco, E. & Pulido-Velázquez, M. (2013) Climate change impacts on groundwater and dependent ecosystems. *Journal of Hydrology* .
- [Kollet & Maxwell(2008)] Kollet, S. J. & Maxwell, R. M. (2008) Capturing the influence of groundwater dynamics on land surface processes using an integrated, distributed watershed model. *Water Resour. Res* 44(2):W02402.
- [Krause et al.(2005)] Krause, P., Boyle, D. P. & Bse, F. (2005) Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences* 5(5):89–97.
- [Kudray & Gale(2000)] Kudray, G. M. & Gale, M. R. (2000) Evaluation of national wetland inventory maps in a heavily forested region in the upper great lakes. *Wetlands* 20(4):581587.
- [Kumar(2009)] Kumar, M. (2009) Toward a hydrologic modeling system. Ph.D. Dissertation The Pennsylvania State University:251pp.
- [Kumar et al.(2009)] Kumar, M., Bhatt, G. & Duffy, C. J. (2009) An efficient domain decomposition framework for accurate representation of geodata in distributed hydrologic models. *International Journal of Geographical Information Science* 23(12):1569–1596.
- [Kumar et al.(2013)] Kumar, M., Marks, D., Dozier, J., Reba, M. & Winstral, A. (2013) Evaluation of distributed hydrologic impacts of temperature-index and energy-based snow models. *Advances in Water Resources* 56:77–89.
- [Lei et al.(2011)] Lei, X., Tian, Y., Liao, W., Bai, W., Jia, Y. W., Jiang, Y. Z. & Wang, H. (2011) Development of an AutoWEP distributed hydrological model and its application to the upstream catchment of the miyun reservoir. *Computers & Geosciences* .
- [Leonard & Duffy(2013)] Leonard, L. & Duffy, C. (2013) Essential terrestrial variable data workflows for distributed water resources modeling. *Environmental Modelling & Software* 50:85–96.

- [Li & Heinemann(2007)] Li, C. & Heinemann, P. H. (2007) A comparative study of three evolutionary algorithms for surface acoustic wave sensor wavelength selection. *Sensors and actuators B: Chemical* 125(1):311–320.
- [Li et al.(2008)] Li, Q., Unger, A. J. A., Sudicky, E. A., Kassenaar, D., Wexler, E. J. & Shikaze, S. (2008) Simulating the multi-seasonal response of a large-scale watershed with a 3D physically-based hydrologic model. *Journal of hydrology* 357(3-4):317–336.
- [Li(2008)] Li, S. (2008) Integrated modeling of multi-scale hydrodynamics, sediment and pollutant transport. Ph.D. thesis, The Pennsylvania State University.
- [Li & Duffy(2011)] Li, S. & Duffy, C. J. (2011) Fully coupled approach to modeling shallow water flow, sediment transport, and bed evolution in rivers. *Water Resources Research* 47(3):W03508.
- [Li et al.(2010)] Li, X., Weller, D. E. & Jordan, T. E. (2010) Watershed model calibration using multi-objective optimization and multi-site averaging. *Journal of Hydrology* 380(3):277–288.
- [Lin(2006)] Lin, H. (2006) Temporal stability of soil moisture spatial pattern and subsurface preferential flow pathways in the shale hills catchment. *Vadose Zone Journal* 5(1):317.
- [Lin(2010)] Lin, H. (2010) CZO dataset: Shale hills - hydrogeologic properties (2007-2010) - water table .
- [Liu & Gupta(2007)] Liu, Y. & Gupta, H. V. (2007) Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. *Water Resources Research* 43:W07401.
- [Loague et al.(2006)] Loague, K., Heppner, C. S., Mirus, B. B., Ebel, B. A., Ran, Q., Carr, A. E., BeVile, S. H. & VanderKwaak, J. E. (2006) Physics-based hydrologic-response simulation: Foundation for hydroecology and hydrogeomorphology. *Hydrological Processes* 20(5):1231–1237.
- [Lu et al.(2009)] Lu, J., Sun, G., McNulty, S. G. & Comerford, N. B. (2009) Sensitivity of pine flatwoods hydrology to climate change and forest management in florida, usa. *Wetlands* 29(3):826836.
- [Lynch(1976)] Lynch, J. A. (1976) Effects of antecedent soil moisture on storm hydrographs. Ph.D. thesis, The Pennsylvania State University.
- [Lynch & Corbett(1985)] Lynch, J. A. & Corbett, E. S. (1985) Source area variability during peakflow: a function of antecedent soil moisture content. In: *Watershed management in the eighties*. ASCE.

- [Mäkelä et al.(2008)] Mäkelä, A., Pulkkinen, M., Kolari, P., Lagergren, F., Berbigier, P., Lindroth, A., Loustau, D., Nikinmaa, E., Vesala, T. & Hari, P. (2008) Developing an empirical model of stand GPP with the LUE approach: analysis of eddy covariance data at five contrasting conifer sites in europe. *Global Change Biology* 14(1):92–108.
- [Martin et al.(2012)] Martin, G. I., Kirkman, L. K. & Hepinstall-Cymerman, J. (2012) Mapping geographically isolated wetlands in the dougherty plain, georgia, usa. *Wetlands* 32(1):149160.
- [Maxa & Bolstad(2009)] Maxa, M. & Bolstad, P. (2009) Mapping northern wetlands with high resolution satellite images and lidar. *Wetlands* 29(1):248260.
- [Maxwell et al.(2007)] Maxwell, R. M., Chow, F. K. & Kollet, S. J. (2007) The groundwater-land-surface-atmosphere connection: Soil moisture effects on the atmospheric boundary layer in fully-coupled simulations. *Advances in Water Resources* 30(12):2447–2466.
- [Meinzer et al.(2012)] Meinzer, F. C., Woodruff, D. R., Eissenstat, D. M., Lin, H. S., Adams, T. S. & McCulloh, K. A. (2012) Above- and belowground controls on water use by trees of different wood types in an eastern us deciduous forest. *Tree Physiology* 33:345–356.
- [Miller et al.(2010)] Miller, G. R., Chen, X., Rubin, Y., Ma, S. & Baldocchi, D. D. (2010) Groundwater uptake by woody vegetation in a semiarid oak savanna. *Water Resources Research* 46(10).
- [Min & Wise(2009)] Min, J.-H. & Wise, W. R. (2009) Simulating shortcircuiting flow in a constructed wetland: the implications of bathymetry and vegetation effects. *Hydrological Processes* 23(6):830841.
- [Mirus & Loague(2013)] Mirus, B. B. & Loague, K. (2013) How runoff begins (and ends): Characterizing hydrologic response at the catchment scale. *Water Resources Research* 49(5):2987–3006.
- [Mirus et al.(2011)] Mirus, B. B., Ebel, B. A., Heppner, C. S. & Loague, K. (2011) Assessing the detail needed to capture rainfall-runoff dynamics with physics-based hydrologic-response simulation. *Water Resources Research* 47(null):W00H10.
- [Morales et al.(2005)] Morales, P., Sykes, M. T., Prentice, I. C., Smith, P., Smith, B., Bugmann, H., Zierl, B., Friedlingstein, P., Viogy, N., Sabaté, S., Sánchez, A., Pla, E., Gracia, C. A., Sitch, S., Arneth, A. & Ogee, J. (2005) Comparing and evaluating process-based ecosystem model predictions of carbon and

water fluxes in major european forest biomes. *Global Change Biology* 11(12):2211–2233.

- [Morin et al.(2012)] Morin, S., Lejeune, Y., Lesaffre, B., Panel, J.-M., Poncet, D., David, P. & Sudul, M. (2012) A 18-yr long (1993–2011) snow and meteorological dataset from a mid-altitude mountain site (col de porte, france, 1325 m alt.) for driving and evaluating snowpack models. *Earth System Science Data Discussions* 5(1):29–45.
- [Naithani et al.(2013)] Naithani, K. J., Baldwin, D. C., Gaines, K. P., Lin, H. & Eissenstat, D. M. (2013) Spatial distribution of tree species governs the spatio-temporal interaction of leaf area index and soil moisture across a forested landscape. *PLOS ONE* 8(3):e58704.
- [Najjar et al.(2009)] Najjar, R., Patterson, L. & Graham, S. (2009) Climate simulations of major estuarine watersheds in the mid-atlantic region of the us. *Climatic Change* 95(1-2):139–168.
- [Najjar et al.(2010)] Najjar, R. G., Pyke, C. R., Adams, M. B., Breitburg, D., Hershner, C., Kemp, M., Howarth, R., Mulholland, M. R., Paolisso, M. & Secor, D. (2010) Potential climate-change impacts on the chesapeake bay. *Estuarine, Coastal and Shelf Science* 86(1):1–20.
- [Naranjo et al.(2013)] Naranjo, R. C., Pohll, G., Niswonger, R. G., Stone, M. & Mckay, A. (2013) Using heat as a tracer to estimate spatially distributed mean residence times in the hyporheic zone of a riffle-pool sequence. *Water Resources Research* 49:3697–3711.
- [Nash & Sutcliffe(1970)] Nash, J. E. & Sutcliffe, J. V. (1970) River flow forecasting through conceptual models part i—a discussion of principles. *Journal of hydrology* 10(3):282–290.
- [Neff et al.(2000)] Neff, R., Chang, H., Knight, C. G., Najjar, R. G., Yarnal, B. & Walker, H. A. (2000) Impact of climate variation and change on mid-atlantic region hydrology and water resources. *Climate Research* 14(3):207–218.
- [Newman et al.(2006)] Newman, B. D., Wilcox, B. P., Archer, S. R., Breshears, D. D., Dahm, C. N., Duffy, C. J., McDowell, N. G., Phillips, F. M., Scanlon, B. R. & Vivoni, E. R. (2006) Ecohydrology of water-limited environments: a scientific vision. *Water Resources Research* 42(6):W06302.
- [Nicklow et al.(2010)] Nicklow, J., Reed, P., Savic, D., Dessalegne, T., Harrell, L., Chan-Hilton, A., Karamouz, M., Minsker, B., Ostfeld, A. & Singh, A. (2010) State of the art for genetic algorithms and beyond in water resources planning and management. *Journal of Water Resources Planning and Management* 136:412.

- [NLDAS(1999)] NLDAS (1999) Mapped monthly vegetation data. p. <http://ldas.gsfc.nasa.gov/nldas/web/web.veg.monthly.table.html>.
- [Nutter(1964)] Nutter, W. L. (1964) Determination of the head-discharge relationship for a sharp-crested compound weir and a sharp-crested parabolic weir. Ph.D. thesis, Pennsylvania State University.
- [Ozesmi & Bauer(2002)] Ozesmi, S. L. & Bauer, M. E. (2002) Satellite remote sensing of wetlands. *Wetlands ecology and management* 10(5):381–402.
- [OHagan(2006)] OHagan, A. (2006) Bayesian analysis of computer code outputs: A tutorial. *Reliability Engineering & System Safety* 91(10-11):12901300.
- [Peng et al.(2013)] Peng, H., Jia, Y., Qiu, Y., Niu, C. & Ding, X. (2013) Assessing climate change impacts on the ecohydrology of the jinghe river basin in the loess plateau, china. *Hydrological Sciences Journal* 58(3):651–670.
- [Post & Pastor(1996)] Post, W. M. & Pastor, J. (1996) LINKAGES—an individual-based forest ecosystem model. *Climatic Change* 34(2):253–261.
- [Praskievicz & Chang(2009)] Praskievicz, S. & Chang, H. (2009) A review of hydrological modelling of basin-scale climate change and urban development impacts. *Progress in Physical Geography* 33(5):650.
- [Prihodko et al.(2008)] Prihodko, L., Denning, A. S., Hanan, N. P., Baker, I. & Davis, K. (2008) Sensitivity, uncertainty and time dependence of parameters in a complex land surface model. *Agricultural and Forest Meteorology* 148(2):268–287.
- [Pyzoha et al.(2008)] Pyzoha, J. E., Callahan, T. J., Sun, G., Trettin, C. C. & Miwa, M. (2008) A conceptual hydrologic model for a forested carolina bay depressional wetland on the coastal plain of south carolina, usa. *Hydrological Processes* 22(14):2689–2698.
- [Qu & Duffy(2007)] Qu, Y. & Duffy, C. J. (2007) A semidiscrete finite volume formulation for multiprocess watershed simulation. *Water Resources Research* 43:W08419.
- [Reba et al.(2011)] Reba, M. L., Marks, D., Seyfried, M., Winstral, A., Kumar, M. & Flerchinger, G. (2011) A long-term data set for hydrologic modeling in a snow-dominated mountain catchment. *Water Resources Research* 47(7):W07702.
- [Reed et al.(2012)] Reed, P. M., Hadka, D., Herman, J. D., Kasprzyk, J. R. & Kollat, J. B. (2012) Evolutionary multiobjective optimization in water resources: The past, present, and future. *Advances in Water Resources* .

- [Renard et al.(2008)] Renard, K. G., Nichols, M. H., Woolhiser, D. A. & Osborn, H. B. (2008) A brief background on the u.s. department of agriculture agricultural research service walnut gulch experimental watershed. *Water Resources Research* 44:W05S02.
- [Rientjes et al.(2013)] Rientjes, T. H. M., Muthuwatta, L. P., Bos, M. G., Booiij, M. J. & Bhatti, H. A. (2013) Multi-variable calibration of a semi-distributed hydrological model using streamflow data and satellite-based evapotranspiration. *Journal of Hydrology* 505:276–290.
- [Rogers & McCarty(2000)] Rogers, C. E. & McCarty, J. P. (2000) Climate change and ecosystems of the mid-atlantic region. *Climate Research* 14(3):235–244.
- [Rosero et al.(2010)] Rosero, E., Yang, Z. L., Wagener, T., Gulden, L. E., Yatheendradas, S. & Niu, G. Y. (2010) Quantifying parameter sensitivity, interaction, and transferability in hydrologically enhanced versions of the noah land surface model over transition zones during the warm season. *J. Geophys. Res* 115:D03106.
- [Rozos et al.(2004)] Rozos, E., Efstratiadis, A., Nalbantis, I. & Koutsoyiannis, D. (2004) Calibration of a semi-distributed model for conjunctive simulation of surface and groundwater flows/Calage dun modle semi-distribu pour la simulation conjointe dcoulements superficiels et souterrains. *Hydrological Sciences Journal* 49(5).
- [Running & Hunt(1993)] Running, S. W. & Hunt, E. R. (1993) Generalization of a forest ecosystem process model for other biomes, biome-bgc, and an application for global-scale models. *Scaling Physiological Processes: Leaf to Globe* pp. 141–158.
- [Scibek & Allen(2006)] Scibek, J. & Allen, D. M. (2006) Modeled impacts of predicted climate change on recharge and groundwater levels. *Water Resources Research* 42(11).
- [Shewchuk(1997)] Shewchuk, J. R. (1997) Delaunay refinement mesh generation. Tech. rep., DTIC Document.
- [Shi et al.(2013)] Shi, Y., Davis, K. J., Duffy, C. J. & Yu, X. (2013) Development of a coupled land surface hydrologic model and evaluation at a critical zone observatory. *Journal of Hydrometeorology* 14:1401–1420.
- [Shi et al.(2014a)] Shi, Y., Davis, K. J., Zhang, F. & Duffy, C. J. (2014a) Evaluation of the parameter sensitivities of a coupled land surface hydrologic model at a critical zone observatory. *Journal of Hydrometeorology* 15:279–299.

- [Shi et al.(2014b)] Shi, Y., Davis, K. J., Zhang, F., Duffy, C. J. & Yu, X. (2014b) Parameter estimation of a physically based land surface hydrologic model using the ensemble kalman filter: A synthetic experiment. *Water Resources Research* 50(1):706–724.
- [Shih & Yeh(2011)] Shih, D. S. & Yeh, G. T. (2011) Identified model parameterization, calibration, and validation of the physically distributed hydrological model WASH123D in taiwan. *Journal of Hydrologic Engineering* 16:126.
- [Singh & Frevert(2002)] Singh, V. P. & Frevert, D. K. (2002) Mathematical modeling of watershed hydrology. *Mathematical models of large watershed hydrology* pp. 1–22.
- [Siqueira et al.(2006)] Siqueira, M. B., Katul, G. G., Sampson, D. A., Stoy, P. C., Juang, J. Y., McCarthy, H. R. & Oren, R. (2006) Multiscale model intercomparisons of CO₂ and H₂O exchange rates in a maturing southeastern US pine forest. *Global Change Biology* 12(7):1189–1207.
- [Smardon(2009)] Smardon, R. C. (2009) *Sustaining the Worlds Wetlands: Setting Policy and Resolving Conflicts*. Springer.
- [Smith et al.(2001)] Smith, B., Prentice, I. C. & Sykes, M. T. (2001) Representation of vegetation dynamics in the modelling of terrestrial ecosystems: comparing two contrasting approaches within european climate space. *Global Ecology and Biogeography* 10(6):621–637.
- [Smith et al.(2008)] Smith, J. W. N., Bonell, M., Gibert, J., McDowell, W. H., Sudicky, E. A., Turner, J. V. & Harris, R. C. (2008) Groundwater-surface water interactions, nutrient fluxes and ecological response in river corridors: Translating science into effective environmental management. *Hydrological Processes* 22(1):151–157.
- [Smucker & Hopmans(2007)] Smucker, A. J. & Hopmans, J. W. (2007) Preface: Soil biophysical contributions to hydrological processes in the vadose zone. *Vadose Zone Journal* 6(2):267–268.
- [Stisen et al.(2011)] Stisen, S., McCabe, M. F., Refsgaard, J. C., Lerer, S. & Butts, M. B. (2011) Model parameter analysis using remotely sensed pattern information in a multi-constraint framework. *Journal of Hydrology* 409(1):337–349.
- [Stone(2011)] Stone, N. (2011) *Gaussian process emulators for uncertainty analysis in groundwater flow*. Ph.D. Dissertation University of Nottingham:199pp.

- [Su et al.(2000)] Su, M., Stolte, W. J. & Van der Kamp, G. (2000) Modelling canadian prairie wetland hydrology using a semidistributed streamflow model. *Hydrological Processes* 14(14):2405–2422.
- [Tague & Band(2004)] Tague, C. L. & Band, L. E. (2004) RHESSys: regional hydro-ecologic simulation system-an object-oriented approach to spatially distributed modeling of carbon, water, and nutrient cycling. *Earth Interactions* 8(19):1–42.
- [Takeuchi et al.(2010)] Takeuchi, J., Kawachi, T., Imagawa, C., Buma, N., Unami, K. & Maeda, S. (2010) A physically based FVM watershed model fully coupling surface and subsurface water flows. *Paddy and Water Environment* 8(2):145–156.
- [Taylor et al.(2013)] Taylor, R. G., Scanlon, B., Doll, P., Rodell, M., van Beek, R., Wada, Y., Longuevergne, L., Leblanc, M., Famiglietti, J. S., Edmunds, M., Konikow, L., Green, T. R., Chen, J., Taniguchi, M., Bierkens, M. F. P., MacDonald, A., Fan, Y., Maxwell, R. M., Yechieli, Y., Gurdak, J. J., Allen, D. M., Shamsudduha, M., Hiscock, K., Yeh, P. J.-F., Holman, I. & Treidel, H. (2013) Ground water and climate change. *Nature Clim. Change* 3(4):322–329.
- [Tiner(2002)] Tiner, R. W. (2002) *Wetland indicators: A guide to wetland identification, delineation, classification, and mapping*. CRC Press.
- [Tiner(2003)] Tiner, R. W. (2003) Geographically isolated wetlands of the united states. *Wetlands* 23(3):494–516.
- [Tolson & Shoemaker(2007)] Tolson, B. A. & Shoemaker, C. A. (2007) Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research* 43(1):W01413.
- [van Genuchten(1980)] van Genuchten, M. T. (1980) A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Sci. Soc. Am. J* 44(5):892–898.
- [van Werkhoven et al.(2008)] van Werkhoven, K., Wagener, T., Reed, P. & Tang, Y. (2008) Characterization of watershed model behavior across a hydroclimatic gradient. *Water Resources Research* 44(1):W01429.
- [van Werkhoven et al.(2009)] van Werkhoven, K., Wagener, T., Reed, P. & Tang, Y. (2009) Sensitivity-guided reduction of parametric dimensionality for multi-objective calibration of watershed models. *Advances in Water Resources* 32(8):1154–1169.

- [Vereecken et al.(2010)] Vereecken, H., Weynants, M., Javaux, M., Pachepsky, Y., Schaap, M. G. & van Genuchten, M. T. (2010) Using pedotransfer functions to estimate the van Genuchten-Mualem soil hydraulic properties: A review. *Vadose Zone Journal* 9(4):795–820.
- [Wagener et al.(2009)] Wagener, T., van Werkhoven, K., Reed, P. & Tang, Y. (2009) Multiobjective sensitivity analysis to understand the information content in streamflow observations for distributed watershed modeling. *Water Resources Research* 45(2).
- [Wang et al.(2009)] Wang, Y. P., Trudinger, C. M. & Enting, I. G. (2009) A review of applications of model-data fusion to studies of terrestrial carbon fluxes at different scales. *Agricultural and Forest Meteorology* 149(11):1829–1842.
- [Wardrop et al.(2007a)] Wardrop, D. H., Kentula, M. E., Jensen, S. F., Stevens, D. L., Hychka, K. C. & Brooks, R. P. (2007a) Assessment of wetlands in the upper Juniata watershed in Pennsylvania, USA using the hydrogeomorphic approach. *Wetlands* 27(3):432–445.
- [Wardrop et al.(2007b)] Wardrop, D. H., Kentula, M. E., Stevens, D. L., Jensen, S. F. & Brooks, R. P. (2007b) Assessment of wetland condition: an example from the upper Juniata watershed in Pennsylvania, USA. *Wetlands* 27(3):416–431.
- [White et al.(2000)] White, M. A., Thornton, P. E., Running, S. W. & Nemani, R. R. (2000) Parameterization and sensitivity analysis of the Biome-BGC terrestrial ecosystem model: net primary production controls. *Earth Interactions* 4(3):1–85.
- [Wolf et al.(2008)] Wolf, A., Blyth, E., Harding, R., Jacob, D., Keup-Thiel, E., Goettel, H. & Callaghan, T. (2008) Sensitivity of an ecosystem model to hydrology and temperature. *Climatic Change* 87(1):75–89.
- [Wösten et al.(2001)] Wösten, J. H. M., Pachepsky, Y. A. & Rawls, W. J. (2001) Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics. *Journal of Hydrology* 251(3):123–150.
- [Yi et al.(2009)] Yi, S., McGuire, A. D., Harden, J., Kasischke, E., Manies, K., Hinzman, L., Liljedahl, A., Randerson, J., Liu, H. & Romanovsky, V. (2009) Interactions between soil thermal and hydrological dynamics in the response of Alaska ecosystems to fire disturbance. *J. Geophys. Res.* 114:G02015.
- [Yu et al.(2013)] Yu, X., Bhatt, G., Duffy, C. & Shi, Y. (2013) Parameterization for distributed watershed modeling using national data and evolutionary algorithm. *Computers & Geosciences* 58:80–90.

- [Yu et al.(2014a)] Yu, X., Duffy, C. J., Jason, K., Wade, C., Bhatt, G. & Shi, Y. (2014a) Reanalysis of water and carbon cycle models at a critical zone observatory. In: Remote Sensing of the Terrestrial Water Cycle. p. in press.
- [Yu et al.(2014b)] Yu, X., Lamačová, A., Duffy, C., Krám, P., Hruška, J., White, T. & Bhatt, G. (2014b) Modeling the long term water yield effects of forest management in a norway spruces forest. *Hydrological Sciences Journal* .
- [Yuan et al.(2011)] Yuan, L.-R., Xin, P., Kong, J., Li, L. & Lockington, D. (2011) A coupled model for simulating surface water and groundwater interactions in coastal wetlands. *Hydrological Processes* 25(23):3533–3546.
- [Zhang et al.(2010)] Zhang, Z., Dehoff, A. D., Pody, R. D. & Balay, J. W. (2010) Detection of streamflow change in the susquehanna river basin. *Water resources management* pp. 1–18.
- [Zreda et al.(2012)] Zreda, M., Shuttleworth, W. J., Zeng, X., Zweck, C., Desilets, D., Franz, T., Rosolem, R. & Ferre, T. P. A. (2012) COSMOS: the COsmic-ray soil moisture observing system. *Hydrol. Earth Syst. Sci.* 16(4079).

Vita

Xuan Yu

EDUCATION

PhD, Civil Engineering, May 2014, The Pennsylvania State University, University Park, PA
ME, Hydrology and Water Resources, April 2009, China Institute of Water Resources and Hydropower Research, Beijing, China
BS, Hydrology and Water Resources, June 2006, China University of Geosciences (Beijing), Beijing, China

EXPERIENCE

2009 - 2013, Research Assistant, Department of Civil and Environmental Engineering, The Pennsylvania State University, University Park, PA
2007 - 2009, Research Assistant, Department of Water Resources, China Institute of Water Resources and Hydropower Research, Beijing, China

SELECTED PAPERS

Christopher Duffy, Lorne Leonard, Gopal Bhatt, Xuan Yu, and Lee Giles. Watershed Reanalysis: Towards a National Strategy for Model-Data Integration. In e-Science Workshops (eScienceW), 2011 IEEE Seventh International Conference on, 2011.

Xuan Yu, Gopal Bhatt, Christopher Duffy, Yuning Shi, Parameterization for Distributed Watershed Modeling Using National Data and Evolutionary Algorithm, Computers & Geosciences, 2013.

Yuning Shi, Kenneth J. Davis, Christopher Duffy, Xuan Yu, Development of a Coupled Land Surface Hydrologic Model and Evaluation at a Critical Zone Observatory, Journal of Hydrometeorology, 2013.

Xuan Yu, Anna Lamačová, Christopher Duffy, Pavel Krám, Jakub Hruška, Tim White, and Gopal Bhatt, Modeling long term water yield effects of forest management in a Norway spruce forest. Hydrological Sciences Journal, 2014.

Xuan Yu, Christopher Duffy, Doug C. Baldwin, Henry Lin, The role of macropores and multi-resolution soil survey datasets for distributed surface-subsurface flow modeling, Journal of Hydrology, 2014.

Xuan Yu., Christopher Duffy, Kaye Jason, Crow Wade, Gopal Bhatt, Yuning Shi, Reanalysis of water and carbon cycle models at a critical zone observatory, in: Lakshmi, V., Remote Sensing of the Terrestrial Water Cycle, Chapman Remote Sensing monograph, AGU, Washington, D. C., 2014.

Yuning Shi, Kenneth J. Davis, Fuqing Zhang, Christopher J. Duffy, and Xuan Yu, Parameter Estimation of a Physically-Based Land Surface Hydrologic Model Using the Ensemble Kalman Filter: A Synthetic Experiment, Water Resources Research, 2014.