An Improved Framework for Watershed Discretization and Model Calibration: Application to the Great Lakes Basin

by

Amin Haghnegahdar

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Civil Engineering

Waterloo, Ontario, Canada, 2015

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Large-scale (~103–106 km^2) physically-based distributed hydrological models have been used increasingly, due to advances in computational capabilities and data availability, in a variety of water and environmental resources management, such as assessing human impacts on regional water budget. These models inevitably contain a large number of parameters used in simulation of various physical processes. Many of these parameters are not measurable or nearly impossible to measure. These parameters are typically estimated using model calibration, defined as adjusting the parameters so that model simulations can reproduce the observed data as close as possible. Due to the large number of model parameters, it is essential to use a formal automated calibration approach in distributed hydrological modelling.

The St. Lawrence River Basin in North America contains the largest body of surface fresh water, the Great Lakes, and is of paramount importance for United States and Canada. The Lakes' water levels have huge impact on the society, ecosystem, and economy of North America. A proper hydrological modelling and basin-wide water budget for the Great Lakes Basin is essential for addressing some of the challenges associated with this valuable water resource, such as a persistent extreme low water levels in the lakes.

Environment Canada applied its Modélisation Environnementale-Surface et Hydrologie (MESH) modelling system to the Great Lakes watershed in 2007. MESH is a coupled semi-distributed land surface-hydrological model intended for various water management purposes including improved operational streamflow forecasts. In that application, model parameters were only slightly adjusted during a brief manual calibration process. Therefore, MESH streamflow simulations were not satisfactory and there was a considerable need to improve its performance for proper evaluation of the MESH modelling system. Collaborative studies between the United States and Canada also highlighted the need for inclusion of the prediction uncertainty in modelling results, for more effective management of the Great Lakes system.

One of the primary goals of this study is to build an enhanced well-calibrated MESH model over the Great Lakes Basin, particularly in the context of streamflow predictions in ungauged basins. This major contribution is achieved in two steps. First, the MESH performance in predicting streamflows is benchmarked through a rather extensive formal calibration, for the first time, in the Great Lakes Basin. It is observed that a global calibration strategy using multiple sub-basins substantially improved MESH streamflow predictions, confirming the essential role of a formal model calibration. At the next step, benchmark results are enhanced by further refining the calibration approach and adding uncertainty assessment to the MESH streamflow predictions. This enhancement was mainly achieved by modifying

the calibration parameters and increasing the number of sub-basins used in calibration. A rigorous multi-criteria comparison between the two experiments confirmed that the MESH model performance is indeed improved using the revised calibration approach. The prediction uncertainty bands for the MESH streamflow predictions were also estimated in the new experiment. The most influential parameters in MESH were also identified to be soil and channel roughness parameters based on a local sensitivity test.

One of the main challenges in hydrological distributed modelling is how to represent the existing spatial heterogeneity in nature. This task is normally performed via watershed discretization, defined as the process of subdividing the basin into manageable "hydrologically similar" computational units. The model performance, and how well it can be calibrated using a limited budget, largely depends on how a basin is discretized. Discretization decisions in hydrologic modelling studies are, however, often insufficiently assessed prior to model simulation and parameter.

Few studies explicitly present an organized and objective methodology for assessing discretization schemes, particularly with respect to the streamflow predictions in ungauged basins. Another major goal of this research is to quantitatively assess watershed discretization schemes for distributed hydrological models, with various level of spatial data aggregation, in terms of their skill to predict flows in ungauged basins. The methodology was demonstrated using the MESH model as applied to the Nottawasaga river basin in Ontario, Canada. The schemes differed from a simple lumped scheme to more complex ones by adding spatial land cover and then spatial soil information. Results reveal that calibration budget is an important factor in model performance. In other words, when constrained by calibration budget, using a more complex scheme did not necessarily lead to improved performance in validation. The proposed methodology was also implemented using a shorter sub-period for calibration, aiming at substantial computational saving. This strategy is shown to be promising in producing consistent results and has the potential to increase computational efficiency of this comparison framework.

The outcome of this very computationally intensive research, i.e., the well-calibrated MESH model for the Great Lakes and all the final parameter sets found, are now available to be used by other research groups trying to study various aspects of the Great Lakes System. In fact, the benchmark results are already used in the Great Lakes Runoff Intercomparison Project (GRIP). The proposed comparison framework can also be applied to any distributed hydrological model to evaluate alternative discretization schemes, and identify one that is preferred for a certain case.

Acknowledgements

All praise is to Allah, the Creator, the All-beneficent, and the All-merciful.

I would like to express my special gratitude to Dr. Bryan Tolson for his continuous guidance and support throughout my PhD studies. Bryan is a caring and intelligent supervisor from whom I have learnt a lot ever since I got to know him. I would also like to thank my co-advisor, Dr. Eric Soulis, for his wise and encouraging comments, which have helped me so much to step forward in my research.

I would like to thank Dr. James Craig, for his insightful comments on some parts of my research. I would also like to thank Dr. Loren S. Matott, for his continuous support on the OSTRICH optimization toolkit, and the changes he made to it exclusively for this work.

I am thankful to my defense committee members Dr. Aaron Berg from the University of Guelph, Dr. Kumaraswamy Ponnambalam, Dr. Bruce McVicar, and Dr. James Craig from the University of Waterloo, for accepting this role and providing constructive feedback.

I would like to thank Environment Canada for funding this work. The majority of this work was part a bigger project requiring close collaboration with a group of researchers from Environment Canada. I would like to particularly thank Bruce Davison, Vincent Fortin, Frank Seglenieks and Erika Klyszejko, for their consultation and for providing the data required during this work.

I am also very thankful to my friends and colleagues, Mahyar, Masoud, Saman, and Saeid, for offering their support as real friends. I would like to specially thank two of my friends, Ali and Hamed, who offered extra support during the last stages of my work.

I am sincerely grateful to my parents, Kazem and Shahnaz, who enthusiastically paved the path for me all the way to this point. This achievement was their dream as much as it was mine. I would also like to thank my sisters, Maryam and Marjan, for their support and warm wishes. I'd also like to sincerely thank my in-laws, for their support and encouragement.

Somayeh, my caring wife, showed endless love and patience during my PhD journey. She sacrificed many weekends and evenings, and selflessly, provided continuous support to me and our two sons, AmirAli and AmirMahdi, to grow and proceed. I am forever sincerely grateful to her.

Dedication

To my parents,

Kazem and Shahnaz,

To my wife,

Somayeh,

and to my sons,

AmirAli & AmirMahdi.

Table of Contents

List of Figures ix
List of Tables xii
Chapter 1 Introduction 1
1.1 Problem Statement 1
1.2 Research Objectives 3
1.2.1 Benchmark the performance of the MESH model for the Great Lakes Basin using a formal calibration strategy. 3
1.2.2 Build an enhanced MESH modelling system for the Great Lakes for streamflow predictions in ungauged basins. 3
1.2.3 Introduce a quantitative comparison framework for assessing the performance of semi-distributed hydrological models under various watershed discretization schemes. 3
1.3 Thesis Structure 4
Chapter 2 Literature Review 5
2.1 Hydrological and Land Surface Modelling 5
2.2 Discretization approaches in distributed hydrological models 6
2.3 Calibration of distributed hydrological models 11
Chapter 3 Calibrating Environment Canada's MESH Modelling System over the Great Lakes Basin: Setting a Benchmark 16
Abstract 16
3.1 Introduction 16
3.2 Methodology 19
3.2.1 Study Area: The Great Lakes Basin 19
3.2.2 MESH modelling system applied to the Great Lakes Basin 20
3.2.3 Calibration Strategy 22
3.3 Results and Discussion 28
3.4 Conclusions 34
Chapter 4 Enhancing the MESH Model by Improved Calibration Strategy and Adding Parameter Uncertainty to Model Predictions 36
4.1 Introduction 36
4.2 Methodology 38
4.2.1 Revised calibration and validation sub-basins 38
4.2.2 Revised calibration parameters 39
4.2.3 Implementation of time-saving techniques 42

4.2.4 Other changes 44
4.2.5 Uncertainty Assessment for Model Predictions 45
4.2.6 Sensitivity Analysis for the Great Lakes Basin 45
4.3 Results and Discussion 46
4.3.1 Enhanced Calibration Results 46
4.3.2 Various attributes of the MESH performance in the new calibration experiment 53
4.3.3 MESH performance for non-natural flow conditions 57
4.3.4 Uncertainty Analysis Results 59
4.3.5 Sensitivity Analysis Results 61
4.4 Conclusions 62
Chapter 5 Assessing the Performance of a Semi-distributed Hydrological Model under various Watershed Discretization Schemes 65
Abstract 65
5.1 Introduction 66
5.2 Methodology 69
5.3 Case Study: MESH model applied to the Nottawasaga River Basin 70
5.3.1 MESH Model 70
5.3.2 Study Area 71
5.3.3 Watershed Discretization Schemes 72
5.3.4 Calibration Strategy 74
5.4 Results and Discussion 79
5.4.1 Detailed assessment of Discretization Schemes 84
5.5 Conclusions 85
Chapter 6 Conclusions and Suggestions for Future Work 88
6.1 Summary and conclusions 88
6.2 General Recommendations for Modelers 89
6.3 Suggestions for Future Work 90
Appendix 92
A – MESH input files for the best parameter set from the calibration experiment of Chapter 4 92
B – Supplemental Results for Chapter 5 94
References 96

List of Figures

Figure 1. The Great Lakes Drainage Basin. 19
Figure 2. Sub-basins used in calibration and validation. Sub-basins 04213500 and 04263000 are only used in the individual strategy. All other sub-basins are used in the global strategy. 24
Figure 3. Nash-Sutcliffe values over the calibration (October 2004–September 2005) and validation (October 2005–May 2009) periods for the 10 calibration sub-basins using the global calibration strategy. 31
Figure 4. Hydrographs corresponding to (a) the worst (lowest) and (b) the best (highest) NS values, over the validation period (October 2005–May 2009), for the 10 calibration sub-basins. 32
Figure 5. Hydrographs corresponding to (a) the worst (lowest) and (b) the best (highest) Nash-Sutcliffe values for the 15 (spatial) validation sub-basins over the October 2004 to May 2009 period. 33
Figure 6. Comparison of pre-calibration and post-calibration (global strategy) hydrographs for the sub-basin with the median pre-calibration Nash-Sutcliffe value (04193500). 34
Figure 7. Comparison of pre-calibration and post-calibration (global strategy) Nash-Sutcliffe values for the calibration period (October 2004–September 2005) for the 10 calibration sub-basins 34
Figure 8. The revised set of calibration and validation sub-basins for the enhanced calibration of MESH over the Great Lakes Basin. 39
Figure 9. The NS values over the calibration period (October 04 to September 05) for the 25 calibration sub-basins and the corresponding weighted NS. 47
Figure 10. Histograms corresponding to the classification of the model performance in the new experiment with respect to the benchmark performance in a) the calibration sub-basins (temporal validation), and b) the validation sub-basins (spatial validation). 49
Figure 11. The NS values from the new and the benchmark calibration experiments for a) the 25 calibration sub-basins plus the corresponding weighted NS, over the temporal validation period (October 05 to September 09), and b) for the 14 validation sub-basins, over the spatial validation period (October 04 to September 09). 50
Figure 12. Simulated hydrographs at the best (highest NS) calibration sub-basin (04045500) for the new and the benchmark experiments. 51
Figure 13. Simulated hydrographs at the best (highest NS) validation sub-basin (02ED027) for the new and the benchmark experiments. 52
Figure 14. Range of the NS values in the 14 validation sub-basins with the crop, deciduous forest, or coniferous forest land cover fraction above 0.3. 53

Figure 15. Scatter plot of the validation NS values for a) 25 calibration sub-basins, and b) 14 validation sub-basins..... 54
Figure 16. The Great Lakes Basin map showing the validation NS value of all the calibration and validation sub-basins using a colour spectrum. Red colours indicate high performance whereas green colours indicate low performance. Sub-basins 04124000 and 04136000 (shown in black) are the two sub-basins removed from calibration due to very poor performance. 55
Figure 17. Observed and simulated hydrographs at sub-basin 04122500 with the worst NS values in calibration and temporal validation..... 56
Figure 18. Bedrock aquifer of the Great Lakes Basin the Great Lakes Basin and the approximate position of the Niagara Escarpment (Modified from Sheets et al., 2005). 57
Figure 19. The NS values over the spatial validation period (October 04 to September 09) from the new calibration experiment for the 10 sub-basins with non- natural flow conditions. 59
Figure 20. Prediction uncertainty bands for the calibration and validation sub-basins with the best performance (largest NS value) during model validation. The hydrograph corresponding to the best parameter set found by the new calibration strategy is also displayed for comparison. 60
Figure 21. The Nottawasaga River Basin near Edenvale and its sub-basins. Calibration sub-basins are labelled with “C” and validation sub-basins are labelled with “V”. 72
Figure 22. Schematic of the GRU concept in representing the sub-grid variability in land classes using the land class-based 7-GRU scheme in two given grid cells, and the relationship between various discretization schemes used in this study. The letters and numbers represent different types of GRUs. Note that the exact position of the GRU types within the grids does not matter and only their fractions are used in calculations. 74
Figure 23. Overall a) calibration, and b) validation performance of various discretization schemes, before calibration and after calibration over the main period (October 2004 to September 2007), for three calibration budgets of 24, 48 and 96 hours. Symbols indicate the median performance for the three calibration replicates. Minimum and maximum values are displayed with bars. For better display purposes, lateral positions are shifted by 3 hours and minimum y-axis was fixed at -0.2. 80
Figure 24. Individual validation performance of sub-basin a) V1, and b) V2 for various discretization schemes, before calibration and after calibration over the main period (October 2004 to September 2007), for three calibration budgets of 24, 48 and 96 hours. Symbols indicate the median performance for the three calibration replicates. Minimum and maximum values are displayed with bars. For better display purposes, lateral positions are shifted by 3 hours and minimum y-axis was fixed at -0.3. 82

Figure 25. Comparison of the overall validation results between a) the main period (October 2004 to September 2007), and b) the sub-period (October 2005 to September 2006) calibration experiments, for three calibration budgets of 24, 48 and 96 hours. Symbols indicate the median performance for the three calibration replicates. Minimum and maximum values are displayed with bars. For better display purposes, lateral positions are shifted by 3 hours and minimum y-axis was fixed at -0.2. 83
Figure 26. Simulated hydrographs and various metrics for validation sub-basin V1 for various discretization schemes a) Lumped, b) 1-GRU, c) 7-GRU, and d) 16-GRU. 84
Figure 27. Simulated hydrographs and various metrics for validation sub-basin V2 for various discretization schemes a) Lumped, b) 1-GRU, c) 7-GRU, and d) 16-GRU. 85
Figure 28. Synthetic experiment results for the overall a) calibration and b) validation model performance, and the individual validation model performance at sub-basin a) V1, and b) V2, after calibration over the main period (October 2004 to September 2007), for three calibration budgets of 24, 48 and 96 hours. Symbols indicate the median performance for the three calibration replicates. Minimum and maximum values are displayed with bars. For better display purposes, lateral positions are shifted by 3 hours and minimum y-axis was fixed at 0.5. 95

List of Tables

Table 1- Calibration parameters, their ranges and the calibrated values using the global strategy for the MESH modelling system applied to the Great Lakes Basin. 26
Table 2- Definition of various model performance metrics. 27
Table 3. Comparison of Individual Strategy (IS) vs. Global Strategy (GS) for Crop land class in calibration and validation sub-basins based on various metrics. 29
Table 4. Comparison of Individual Strategy (IS) vs. Global Strategy (GS) for Deciduous Forest land class in calibration and validation sub-basins based on various metrics. 29
Table 5. Median, Minimum and Maximum values for the four metrics defined in Table 2 over the Oct. 2004 to May 2009 period for the 15 (spatial) validation sub-basins using the global calibration strategy. 32
Table 6- Calibration parameters, their ranges and the calibrated values for the new calibration of the MESH model applied to the Great Lakes Basin. 42
Table 7. The median, minimum and maximum of the four metrics (defined in Table 2), for the new and benchmark experiments, over the temporal validation period (October 2005 to September 2009), for the 25 calibration sub-basins. 48
Table 8. The median, minimum and maximum of the four metrics (defined in Table 2), for the new and benchmark experiments, over the spatial validation period (October 2004 to September 2009), for the 14 validation sub-basins. 48
Table 9. Classification of the model performance for the new results with respect to the benchmark results. 49
Table 10. List of the sub-basins with non-natural flow conditions assessed for MESH performance. 58
Table 11. The median, minimum and maximum of the four metrics (defined in Table 2), for ten sub-basins with non-near-natural flow conditions over the spatial validation period (October 2004 to September 2009). 58
Table 12. Range of the median, minimum, and maximum for the 25 sub-basin NS values across the 8 behavioral parameter sets. 59
Table 13. Local sensitivity results for the MESH streamflow simulations to the parameters of the most common GRU type (Crop) in the Great Lakes Basin. The changes above 5% are highlighted. 63
Table 14- Characteristics of the sub-basins. 72
Table 15- The seven GRU types produced by spatial land cover data for the Nottawasaga river basin, ranked by coverage area. 73
Table 16- The sixteen GRU types produced by combining land cover and soil spatial data for the Nottawasaga river basin, ranked by coverage area. 74

Table 17- The number of parameters, simulation runtime, and the maximum number of model evaluations used in DDS for each calibration budget for each discretization scheme. 77
Table 18- Parameters and their corresponding ranges used in calibration of MESH. 78

Chapter 1
Introduction

1.1 Problem Statement

Advances in process understanding, computational capabilities, and data measurements, have inspired the increased usage of large scale physically-based distributed hydrological models for many applications, such as evaluating human impacts on global water budget (e.g., McCabe and Wolock, 2011), or determining interbasin water transfers (e.g., Islam et al., 2007). As such, these models are one of the key tools in making decisions regarding the management of water and environmental resources. These hydrological models inevitably contain a large number of parameters used in simulation of various physical processes. Many of these parameters are not measurable or nearly impossible to measure. These parameters are typically estimated using model calibration, defined as adjusting the parameters so that model simulations can reproduce the historical (observed) response data (e.g., daily streamflow) as close as possible. Without a careful model calibration and validation, model simulations can be unrealistic or misleading (e.g., Moriasi et al., 2012). Calibration strategies range from simple manual calibration to sophisticated automated calibrations. Manual calibration is very time-consuming and limited in distributed hydrological modelling studies, due to the large number of model parameters. Therefore, it is essential to adopt a formal automated calibration approach in these cases.

The St. Lawrence River Basin in North America contains the largest body of surface fresh water in the world, the Great Lakes, and is of paramount importance globally and in particular, for United States and Canada. The Lakes' water levels have huge impact on the ecosystem and economy of North America (e.g., Millerd 2005). A proper hydrological modelling and basin-wide water budget for the Great Lakes Basin is essential for addressing some of the challenges associated with this valuable water resource, such as a persistent extreme low water levels in the lakes over the past decade (e.g., Gronewold and Stow, 2014).

In 2007, Environment Canada applied its Modélisation Environnementale-Surface et Hydrologie (MESH) modelling system to the Great Lakes watershed (Pietroniro et al. 2007). MESH is a coupled semi-distributed land surface-Hydrological model developed for various water management purposes. The goal in that application was mainly to set up the coupling framework and demonstrate its potential for operational forecasts in the Great Lakes Basin, and also to identify specific requirements for different components of the MESH modelling system (Pietroniro et al. 2007). Accordingly, model parameters were only slightly adjusted during a brief manual calibration process. Therefore, MESH streamflow

simulations were not satisfactory and there was a considerable need to improve its performance for proper evaluation of the MESH modelling system. Consequently, one of the primary goals of this study is to apply proper calibration techniques to enhance the performance of MESH over the Great Lakes Basin, particularly concerning streamflow predictions in ungauged basins (PUB).

Collaborative studies between the United States and Canada have also highlighted the need for inclusion of the prediction uncertainty in modelling results, for a more effective management of the Great Lakes system (e.g., Gronewold et al., 2011). As approximations of the real world, models have uncertainty corresponding to their parameters, structure, or input data. Therefore, outcomes of modelling studies can be more effectively used by decision makers, when reported along with some assessment of uncertainty. In 2003, the International Association of Hydrological Science (IAHS) placed estimation of predictive uncertainty at the core of the decade (2003–2012) on the PUB (Sivapalan et al., 2003). Accordingly, as another contribution of this thesis, the uncertainty due to parameter estimation is evaluated for MESH streamflow predictions in the Great Lakes Basin.

The primary advantage of distributed models is in providing spatially distributed information about model outcomes. However, the main challenge remains as how to represent the existing spatial heterogeneity in nature (e.g., for land cover or soil type) in these models. For simulations at large scales, this task is normally performed via watershed discretization, here defined as the process of subdividing the basin into manageable "hydrologically similar" computational units. The model performance and how well it can be calibrated using a limited budget, largely depend on how a basin is discretized (e.g, Flügel, 1995). The more spatial details incorporated when discretizing a watershed, the more complex and computationally costly becomes the model, especially for calibration.

Discretization decisions in hydrologic modelling studies are often insufficiently scrutinized. Often, hydrologic modelling studies tend to adopt a prescribed approach or assumption guiding the discretization decision (usually developed under different circumstances), assuming it is adequate, and then proceed to model simulation and parameter calibration (e.g., Donald et al., 1995, Pietroniro et al., 1996). Few studies explicitly present an organized and objective methodology for assessing discretization schemes, particularly in the PUB context. Accordingly, another major goal of this research is to quantitatively assess watershed discretization schemes, with various level of spatial data aggregation, in terms of their skill to predict flows in ungauged basins. This analysis will also investigate if there is a tradeoff between the model performance and the computational cost for calibrating different discretization schemes in MESH.

1.2 Research Objectives

The primary research objectives of this thesis are described below:

1.2.1 Benchmark the performance of the MESH model for the Great Lakes Basin using a formal calibration strategy.

This objective intends to identify a satisfactory baseline performance of the MESH model for predicting ungauged streamflows over the Great Lakes Basin. This goal is achieved by a formal calibration (and validation) of MESH, for the first time, over the entire Great Lakes Basin. The outcomes of this objective can then serve as a reference for evaluation of relevant future studies to improve MESH performance in this region of Canada. This objective is discussed in Chapter 3 of the thesis.

1.2.2 Build an enhanced MESH modelling system for the Great Lakes for streamflow predictions in ungauged basins.

Building upon the first objective, substantial enhancement of the baseline results is targeted by further refining the calibration approach and performing an uncertainty analysis related. The outcome would be a well-calibrated MESH model in the Great Lakes Basin that can be used for predicting streamflows, predicting lake inflows for each of the Great Lakes, or ultimately, for lake level prediction. This product would then be the model recommended for future modelling studies with MESH in this region. This objective is detailed in Chapter 4.

1.2.3 Introduce a quantitative comparison framework for assessing the performance of semi-distributed hydrological models under various watershed discretization schemes.

This objective endeavors to quantitatively assess watershed discretization schemes, with various levels of spatial data aggregation, in terms of their skill to predict flows in ungauged basins. Special attention is spent to determine the added benefit of including information about land cover and soil type. In addition, the sensitivity of the results to the budget spent for model calibration is examined. Lastly, it is tested whether the same results can be obtained when the main calibration period is replaced with a sub-period. This is important because the use of a condensed calibration period can potentially save substantial amounts of time during the numerous calibration experiments required by this method. The methodology was demonstrated using the MESH model as applied to the Nottawasaga river basin, a sub-basin of the Great Lakes Basin, in South-Western Ontario, Canada. The scope of this study is not to examine the effect of grid cell size on watershed discretization and model performance.

1.3 Thesis Structure

Chapters in this thesis are organized based on the objectives discussed above; meaning that a single chapter is dedicated to each objective. When applicable, a published or submitted paper is used as a chapter, with few modifications to maintain the consistency of the thesis as a whole.

Chapter 2 provides an updated review of literature relevant to this research with a focus on the calibration of distributed hydrological models and different approaches for watershed discretization. Chapter 3 is a mirror of a published paper (Haghnegahdar et al., 2014) that highlights the methods and results for the benchmarking calibration of the MESH model over the Great Lakes Basin. Chapter 4 then discusses how the benchmark results of Chapter 3 were further enhanced by refining the calibration strategy and by introducing the parameter uncertainty in streamflow predictions. Chapter 5, based on a paper recently submitted for publication, describes a quantitative methodology for assessing various watershed discretization schemes. Finally, Chapter 6 outlines the conclusions from the studies in this thesis and provides some suggestions for related research in future.

Chapter 2
Literature Review

2.1 Hydrological and Land Surface Modelling

A hydrological model is a mathematical model used to simulate the movement and storage of the water within a certain boundary, typically a watershed. Hydrological models range from a simple single bucket conceptual model to detailed physically-based models simulating multiple processes of water flow in nature. Depending on this complexity, the required input data can also vary. Some simple models require only the rainfall as input data, whereas more sophisticated models also require, for example, temperature, pressure, wind speed and radiation data. Hydrological models have many applications such as flood forecasting, land use change impact, streamflow forecasting and water quality modelling.

Hydrologic models can also be classified as being "lumped" or "distributed". The exact definition of "lumped" versus "distributed" models varies in the literature and by the application. According to Clarke (1973), a lumped model does not consider the spatial distribution of input data or model parameters, whereas a distrusted model accounts for the spatial variability of input data. In practice, a distributed hydrologic model often refers to a model that somehow represents the spatial variability and water paths in a catchment (Kampf and Burges, 2007). A detailed classification of distributed hydrologic models is provided by Kampf and Burges (2007).

Accordingly, there is the group of physically-based distributed hydrological models. These models can be sub-divided into fully-distributed and semi-distributed models. In fully-distributed models like SMDR (Soil Moisture Distribution and Routing, Srinivasan et al., 2005), SHE (Systeme Hydrologique Europeen, Abbott et al., 1986), the exact location for each computational element is specified and the lateral transfer of water and energy fluxes are allowed between these elements throughout the model spatial domain. In contrast, although the spatial variation of parameters are somewhat represented in semi-distributed models via discretized spatial units, the exact location of each computational element within these spatial units are neglected such as in SWAT (Arnold et al., 1998), WATFLOOD (Kouwen, 1988) and TOPMODEL (Beven and Kirkby, 1979). Therefore, in semi-distributed models, routing of water fluxes between the computational elements is rather conceptualized before reaching the main channel.

A Land surface model (LSM) or Land surface Scheme (LSS) aims at mathematically simulating the transfer of water and energy between the Earth's surface, vegetation, and atmosphere. LSMs are primarily used in climate models to simulate energy exchanges between land cover and atmosphere (Pitman 2003). They are mostly one-dimensional column models that describe the water and energy pathways through

the root zone and vegetation with various levels of complexity. Sellers et al. (1997) classified LSMs into three categories: first, second and third-generation. Generally, in the first-generation models, a simple bucket model is used with soil-atmosphere interactions only. The second-generation models improve to include both soil and vegetation interactions with details, and the more sophisticated third-generation models account also for carbon flux pathways (Sellers et al., 1997).

Many studies have particularly focused on using and advancing the second-generation LSMs that contain an explicit treatment of mass and energy interactions between land, canopy and atmosphere (Pitman 2003). Examples of such models include the Simple Biosphere Model (SiB, Sellers et al., 1986), and the Canadian Land Surface Scheme (CLASS, Verseghy, 1991; Verseghy et al, 1993). CLASS was developed in response of the need to include a second-generation land surface scheme for the Canadian global circulation model (Verseghy, 1991).

Over the past two decades, there has been an extensive global effort to couple atmospheric and hydrological models using LSMs (Pietroniro et al., 2007). This connection can improve hydrological flow simulations and atmospheric predictions in both climate (Soulis et al., 2000) and weather prediction models (Benoit et al., 2000). Similar systematic efforts were made in Canada starting mid-90s (Pietroniro et al., 2007). These efforts eventually led to the creation of the Modélisation Environmentale–Surface et Hydrologie (MESH) model in 2007, by coupling the CLASS land surface scheme (Verseghy, 1991; Verseghy et al, 1993) and the WATFLOOD hydrological model (Kouwen, 1988) by Environment Canada (Pietroniro et al., 2007). Development of MESH was at the center of Hydrologic Ensemble Prediction Experiment (HEPEX) Great Lakes test bed project, aimed at improving streamflow forecast at large scales within the Great Lakes tributaries (Pietroniro at el., 2007).

2.2 Discretization approaches in distributed hydrological models

Due to data and computational constraints, application of fully-distributed models to large scale ($\sim 10^3$–10^6 km^2) hydrologic modelling is typically infeasible for operational purposes and thus, semi-distributed models are primarily relied upon in thb ese situations. However, specifying all the required model parameters with the appropriate scale, poses a challenge for semi-distributed models. These parameters are normally associated with the model approach to accommodate spatial heterogeneity (i.e. discretization) and are typically determined via model calibration. Calibration is the procedure of adjusting the parameters so that model simulations can reproduce the historical (observed) response data (e.g., daily streamflow) as close as possible.

Two fundamental decisions are to be made when discretizing a watershed for simulation with a semi-distributed model. The first one, which depends largely on the spatial resolution of the forcing data (e.g. precipitation, temperature, etc.) and/or the density of the streamflow gauges, is how to specify the size (resolution) and shape (e.g., grid cell or sub-basin) of the spatial units used in the model for channel routing. In other words, determine the spatial scale (resolution) which yields estimates of runoff rates at the specific locations of interest. Some semi-distributed models such as SWAT and LIQUID (Branger et al., 2010) divide the watershed into sub-basins (smaller sub-watersheds) while other models such as WATFLOOD and MESH divide the watershed into gridded spatial units. Some models such as Raven (Craig and Snowdon, 2011) can subdivide either way depending on the application.

The second decision in watershed discretization is concerned with subdividing the aforementioned spatial units (i.e., sub-basins or grid cells) further into elements with distinct hydrologic responses. These entities will be used by the model as the computational unit for simulation. Being rather more complex and important, this latter decision mainly determines the representation of heterogeneity in the model and thus, largely specifies the model complexity and number of (unknown) model parameters.

Many distributed watershed modelling studies have tried to find a suitable approach for spatial discretization of a basin that reasonably balances between model complexity and its applicability in practice. Wood et al. (1988) proposed the concept of Representative Elementary Area (REA) to identify characteristic spatial scales for catchment responses in terms of the runoff generation. Reggiani et al. (1998) extended this idea to the concept of Representative Elementary Watersheds (REWs) that are sub-watersheds across which same physical laws govern and are further sub-divided into several flux zones.

As an alternative to the raster-based approaches, Vivoni et al. (2004) used the triangulated irregular networks (TINs) delineated primarily based on a topographic (wetness) index (Beven and Kirkby, 1979) to be used within the TIN-based Real-Time Integrated Basin Simulator (tRIBS, Ivanov et al., 2004). Dehotin and Braud (2008) proposed the use of “hydro-landscape“ units (Winter, 2001) to address sub-catchment variability in a nested discretization approach to be implemented within the LIQUID modelling platform (Branger et al., 2010). Capable of taking into account various influential factors on hydrologic response such as slope, land use, geology, pedology and etc., the hydro-landscape units are very similar to the Hydrological Response Units (HRU, e.g., Leavesley et al., 1983; Flügel, 1995).

HRUs are perhaps the most commonly used approach to incorporate landscape heterogeneity in watershed modelling. The exact definition of HRUs depends on the model and the application (Haverkamp et al., 2002). Flügel (1995) defines HRUs as “distributed, heterogeneously structured entities having a common climate, land use and underlying pedo-topo-geological associations controlling their

hydrological dynamics". In the case of SWAT, HRUs are sub-basin divisions characterized by a unique combination of land cover, soil and land management (e.g., Neitsch et al., 2005). Although HRUs are mostly defined as subunits of a sub-basin, they may also be delineated based on a grid too.

Kouwen et al. (1993) introduced the Grouped Response Unit (GRU) approach in WATFLOOD model and described it as the aggregation of multiple land classes (or possibly other attributes) into a single unit in order to increase computational efficiency. Pietroniro and Soulis (2003) found the GRU concept a suitable approach for semi-distributed hydrological modelling in large basins because it accounts for the physics while it is operationally simple. In most WATFLOOD or MESH model applications, the GRUs are defined as subdivisions (not necessarily contiguous) of a Cartesian grid cell rather than a topographically-defined sub-basin (e.g., Pietroniro et al., 2007).

When dealing with hydrological models, it is very well recognized that a number of parameters are unknown and have to be determined through calibration. However, modelers typically overlook the fact that the most appropriate discretization scheme is also unknown and that it could also be determined as part of the model calibration process. Instead, modelers tend to use a single discretization decision (perhaps for simplicity or operational considerations) that is already developed from a combination of subjective user input, modeler experience, available data format, and built-in model constraints. Consequently, it is desirable to develop guidelines for identifying the appropriate watershed discretization that is assessed and less subjective when calibrating a model.

Few studies, however, explicitly present an organized and objective methodology for defining and assessing discretization schemes. Ideally, a scheme should find the balance between spatial aggregation level and model complexity. Some of the important studies include the work of Flügel (1995), Haverkamp et al. (2002), Dehotin and Braud (2008).

Flügel (1995) presented a fairly detailed procedure to delineate HRUs for regional scale hydrological modelling in a manner that seeks to preserve the three-dimensional heterogeneity of the basin. The procedure involves a step-by-step GIS-based map overlay analysis to specify HRUs according to physiographic properties such as topography, soil type, geology and land use. As a result, 23 HRUs were delineated by his method in the River Brol basin (216 km^2) in Germany.

Haverkamp et al. (2002) introduced a rather objective and statistical methodology for determining the appropriate level of spatial aggregation (discretization) based on the entropy function (e.g., Krasovskaia, 1997). Conducting experiments with SWAT, they found a threshold value for the number of sub-watersheds (and HRUs) beyond which model performance is not improved any further. On this basis,

they developed the SUbwatershed Spatial Analysis Tool (SUSAT) within the SWAT framework for finding an appropriate level of discretization. Haverkamp et al. (2002) assessed the quality of the discretization decisions in terms of how well the model can be calibrated to available monitoring data. Unfortunately, this approach to quality assessment does not necessarily reflect the quality of the discretization decision in terms of predictions in ungauged basins.

Dehotin and Braud (2008) proposed a flexible three-level discretization strategy for implementation in a modular modelling platform called LIQUID (Branger et al., 2010). In their method, the spatial discretization can become finer adaptively according to modelling objectives or numerical requirements. They use a very similar concept to HRU, called "hydro-landscape" units (Winter, 2001) to address sub-catchment variability. Hydro-landscape units in their approach are determined in a very similar manner to that of Flügel (2005) for HRU delineation, except that they basically employ a supervised classification technique to preserve possible small areas that are hydrologically influential based on the chosen processes and could have been neglected otherwise.

Consequently, presumably after following one of these suggested procedures, one obtains a set of non-overlapping HRUs (or similarly GRUs) characterized by a unique combination of attributes considered important for characterizing the hydrological response.

Even in these approaches there exist some inevitable subjective decisions that have to be made such as:

1. Determination of the most influential hydrological factors which is a common fundamental decision across all approaches (e.g., delineation of GRUs/HRUs on the basis of land cover, land use, soil type, terrain attributes, etc.).
2. Dividing each factor into different categories, again a common fundamental decision (e.g., dividing slope into low, moderate, high).
3. Deciding on how to merge some HRUs in order to keep the number of HRUs manageable for modelling purposes in the method introduced by Flügel (1995) (e.g., merging small agricultural areas (less than 0.2%) into rangeland and slopes <10% into the gradient class of 0-10%).
4. The user-specified accepted level of information loss (e.g., 5%) in the approach developed by Haverkamp et al. (2002).
5. Defining the size and shape of the neighbourhood window (e.g. ellipsis neighbourhood window) or the reference zones used in classification in the procedure proposed by Dehotin and Braud (2008).

It is worth emphasizing that the goal here is not to eliminate all these subjective decisions, which is of course impossible, but rather to make them less subjective to the extent possible and quantify their impact on model performance and efficiency particularly in the ungauged areas.

Moreover, many studies, including the aforementioned ones, report the model success in terms of the calibration quality without measuring its performance in spatial validation which could be a big concern. The reason for measuring model performance in ungauged areas is because this criteria better reflects the ability of the discretization scheme in capturing the natural heterogeneity. In other words, the more HRUs/GRUs represents the real physics of the watershed the model hydrologic response would be more consistent over the entire basin including ungauged areas. In contrast, if HRUs/GRUs are poorly delineated, the resultant parameter set is effectively compensating for the errors induced from improper delineation during the calibration process. Therefore, even with apparently reasonable simulation results in the gauged basin, it is unlikely that these compensatory parameter values will port very well to an ungauged basin.

Similar discretization issues are commonly confronted when using the MESH model. In particular, for the case study of the MESH model calibration in the Great Lakes Basin, Neff and Nicholas (2005) note that a total of 32% of the land areas draining into the Great Lakes were ungauged as of 1992. Kouwen et al. (1993) advocate building GRUs primarily based on land cover type, since this property dominantly controls the hydrologic response. Subsequently, other studies also adopted same strategy to delineate GRUs in WATFLOOD (e.g., Donald et al., 1995, Pietroniro et al., 1996) and MESH (e.g., Pietroniro et al., 2007; MacLean, 2009). However, using MESH, other researchers (e.g., Davison et al., 2006, Dornes et al., 2008) have found that topography can also impact the hydrologic response in northern climates and have thus explicitly incorporated slope and aspect into the GRU definition in addition to land cover. Haghnegahdar et al. (2010) also showed that the current GRU discretization used in WATFLOOD is not optimal based on the results obtained for WATFLOOD calibration in the Great Lakes Basin. They observed inconsistent results between model performances in some stations when WATFLOOD was calibrated in one station individually versus the case when it was calibrated globally for all stations. Unfortunately, currently there is no general framework for determining GRUs and there is very little flexibility to evaluate alternative discretization approaches for WATFLOOD and MESH modelers. Accordingly, it is desirable to know and quantify the value of adding more spatial attributes into GRU definition in MESH.

Flügel (1995) suggested detailed field investigations to assess the appropriateness of HRU delineation (particularly that the variation of hydrological process dynamics within a single HRU must be small

compared to the dynamics between different HRUs), this work is not typically possible in most situations, especially at regional scales. Some researchers only consider the first level of watershed discretization in a semi-distributed modelling approach, which is, applying a conceptual (lumped) model in a distributed way to individual grid cells or sub-basins and neglecting the heterogeneity within a grid or sub-basin. For example, Boyle et al. (2001) found improvement in model performance when a semi-distributed version of the U.S. National Weather Service (NWS) Sacramento Soil Moisture Accounting (SAC-SMA) model (Burnash et al., 1973) was compared with the original lumped version in the Blue River watershed (1227 km^2) located in southern Oklahoma. They reported that improvements are clearly associated with the spatially distributed model input (rainfall) and routing approach, but "surprisingly, there is no improvement associated with the distribution of the surface characteristics" (i.e. soil parameters). Applying same model and strategy to the Illinois River basin at Watts, Oklahoma (1645 km^2) Ajami et al. (2004) also concluded that although the semi-distributed models are better for providing spatially distributed information (e.g., at interior points), "the resulting improvement in simulation capability at the outlet, compared to the lumped model is not yet significant to justify adoption of semi-distributed model." Pokhrel and Gupta (2010) followed the same discretization approach and applied a lumped parameter model in the same basins and the Baron Fork River basin in the Oklahoma-Arkansas border. Although Pokhrel and Gupta (2010) reported that their distributed calibration methodology achieves considerable performance improvements at the basin outlet, they also noted that the strategy resulted in degraded predictions for the interior gauges that were not explicitly calibrated to. Specifically, Pokhrel and Gupta (2010) reported that prior parameter estimates are found to give much better performance at the interior points (treated as ungauged) and thus, they concluded that the spatial information has not been properly exploited. Pokhrel and Gupta (2010) did not suggest a framework for testing how to exploit that spatial information. It is not clear if conclusions from these studies would be the same if the heterogeneity within a grid or sub-basin is further accounted for using HRUs or GRUs approaches. Alternatively, a novel framework is proposed here by which, the assessment of watershed discretization schemes, with various levels of complexity, can be conducted in a quantitative and systematic way, primarily by considering the model performance for the parameter transferability to ungauged areas. This method is described in Chapter 5.

2.3 Calibration of distributed hydrological models

Regional (large scale) distributed land surface-hydrological models are being increasingly used for the purpose of hydrologic and climatic simulations or forecasts (e.g., Flügel, 1995; Benoit et al., 2000; Arnold & Fohrer, 2005; Samaniego et al., 2010). These models contain parameters that are impossible or

very difficult to measure. These parameters are typically estimated using model calibration, defined as adjusting the parameters so that model simulations can reproduce the historical (observed) response data (e.g., daily streamflow) as close as possible. Due to the model and thus calibration complexity, particularly for distributed environmental models, automated calibration approaches are normally adopted for parameter estimation as opposed to manual calibration. Some researchers suggest a combination of both approaches to enhance the calibration results (e.g., Boyle et al., 2001).

Automatic calibration techniques can be divided into two main categories: optimization-based and uncertainty-based. More traditionally used, optimization-based calibration is facilitated via an optimization algorithm that searches for the global "optimal" parameter set that can best simulate the historical observations. A large number of such optimization algorithms have been applied in environmental modelling practice like the Genetic Algorithm (Wang, 1991), the Shuffled Complex Evolution (SCE) algorithm (Duan et al., 1992, 1993) and Dynamically Dimensioned Search (DDS) algorithm (Tolson and Shoemaker, 2007). A newer generation of calibration algorithms aims at providing a range or probability distribution (instead of a single value) for parameters to account for uncertainty in estimation that can be further translated into model predictions too. Examples of such algorithms include Generalized Likelihood Uncertainty Estimation, GLUE (Beven and Binley, 1992), Sequential Uncertainty FItting (SUFI-2) (Abbaspour et al., 2004) and various methods using Markov Chain Monte Carlo (MCMC) algorithms (e.g., Kuczera and Parent, 1998; Vrugt et al., 2003).

The SCE algorithm (Duan et al. 1992, 1993) developed at the University of Arizona (SCE-UA) is one of the most commonly used global optimization algorithm for watershed calibration (e.g., Sorooshian et al., 1993; Gan and Biftu, 1996; Kuczera, 1997). SCE is a probabilistic population-based evolutionary type of algorithm designed to find the globally optimal set of model parameters. In particular, many hydrologic modelling studies with the SWAT (e.g., Rouhani et al., 2009; Abbaspour et al., 2007) and SAC-SMA models (e.g., Duan et al., 1994; Gupta et al., 1998; Hogue et al., 2000) tend to use the SCE-UA for parameter calibration. Tolson and Shoemaker (2007) introduced DDS as an efficient and simple single-solution stochastic heuristic global search algorithm. DDS was designed to find "good" global solutions (as opposed to globally optimal solutions) within the modeler's specified time frame. Using a number of calibration case studies with the SWAT2000 model (Neitsch et al., 2001), Tolson and Shoemaker (2007) report that DDS outperforms SCE in the context of distributed watershed model automatic calibration and is more efficient when 10 or more parameters should be calibrated.

DDS has been applied to a number of hydrological modelling studies in Canada. Dornes et al. (2008) used DDS in a step-wise calibration procedure in north-eastern Canada to evaluate parameter

transferability of the MESH model to ungauged basins. MacLean (2009) applied DDS to calibrate MESH simultaneously to snow water equivalent and streamflow in two study areas in the United States and Canada. DDS was also used by Seglenieks (2009) to calibrate WATFLOOD over all of Canada. Haghnegahdar et al. (2010) also calibrated WATFLOOD model using DDS for more than 100 stations across the Great Lakes Basins. In this research, DDS will be adopted as the calibration algorithm.

It is widely recognized that often a single objective measure is not adequate to obtain a proper parameter set for simulation and a multi-objective calibration is strongly recommended to reduce the error in the results (e.g., Yapo et al., 1998; Madsen, 2003). Two major approaches are often adopted to implement a multi-objective calibration. One simple strategy is to use a weighted average to aggregate multiple objectives into a single objective (e.g., Madsen, 2003 and Tolson and Shoemaker, 2008). This strategy is adopted in this research as well by using a weighted average Nash-Sutcliffe coefficient (Nash and Sutcliffe, 1970) of streamflows in several stations as the objective function. Another more sophisticated strategy is to use a multi-objective optimization algorithm such as Evolutionary Multi-objective Optimizers or EMOs (e.g., Tang et al., 2006). These algorithms approximate the entire non-dominated set that allows the modeler to evaluate the resultant tradeoffs and select a solution they deem to best balance all objectives. The Multiple Objective Shuffled Complex Evolution Metropolis global search algorithm (MOSCEM) is commonly applied for multi-objective hydrologic model calibration (e.g., Vrugt et al., 2003; Pokhrel and Gupta, 2010).

The calibration of regional scale physically-based distributed hydrological models is normally very computationally costly due to long model runs and large number of parameters involved. Accordingly, a substantial body of research has been devoted to find methods for reducing the calibration computational demand and increasing its efficiency. One approach commonly used to reduce the number of parameters in distributed model calibration is "regularization". It is a mathematical strategy that normally helps to reduce the search space dimensionality by including additional information often in the form of `constraints on parameters (e.g., Pokhrel et al., 2008 and Tonkin and Doherty, 2005). The so-called "multiplier" method is a very common and simple practice in spatially distributed watershed modelling for regularization. It implies applying a scalar multiplier to each prior parameter field (e.g., vegetation parameters across all GRUs) and calibrating this multiplier (e.g., Bandaragoda et al., 2004; Yatheendradas et al., 2008). Pokhrel and Gupta (2010) compare the multiplier method to alternative spatial regularization approaches. In contrast, MacLean (2009) calibrated the MESH model without spatial regularization and as a result ended up calibrating close to 80 model parameters.

Alternatively, Metamodels (surrogate models) have been widely used in various disciplines including parameter estimation in environmental modelling (e.g., Khu and Werner, 2003; Mugunthan and Shoemaker, 2006; Zhang et al., 2009). Metamodelling approaches focus on approximating the original complex model through examining the relationship between model responses and a selected number of parameter sets. The computationally expensive distributed hydrologic model is replaced with a drastically more efficient surrogate model that is fitted to approximate the hydrologic model. Razavi et al. (2012) provided a thorough review of surrogate modelling in water resources field.

Many researchers have proposed efficient optimization algorithms attempting to reduce computational cost of calibration. In general, these algorithms are designed to find "good" or "near-optimal" solutions within a limited time frame as opposed to algorithms who claim to find "optimal" solution given unlimited (huge) time budget. Examples of such optimization-based algorithms are Dynamically Dimensioned Search (DDS, Tolson and Shoemaker, 2007) and Stepwise Linear Search (SLS, Kuzmin et al., 2008). Moreover, DDS Approximation of Uncertainty (DDS-AU, Tolson and Shoemaker, 2008) and limited-memory Markov Chain Monte Carlo (MCMC, Kuczera et al., 2010) can be classified as efficient uncertainty-based calibration methods.

The growth in computing capabilities, in general, and the parallel computing networks, in particular, has led to reduced calibration times. For instance, this fact allows for more efficient implementations of uncertainty-based calibration techniques like GLUE that are normally very computationally demanding (e.g., Brazier et al., 2000). Moreover, some optimization-based algorithms such as parallel SCE-UA (e.g., Vrugt et al., 2006), parallel GA (e.g., Cheng et al., 2005), and parallel PSO (e.g., Schutte et al., 2004) are proposed in which a parallel search strategy is adopted versus the regular serial approach resulting in considerably higher calibration efficiencies. In a population-based parallel optimization algorithm, the solutions in the each generation to be evaluated are distributed over multiple processors so that they can be evaluated simultaneously.

A limited number of studies have tried to increase savings by selectively avoiding unbeneficial model simulations (i.e. model runs that predictably result in a poorer performance with respect to a certain threshold). Gray and Kolda (2006) consulted stored simulation history (caching) to avoid repeated runs in optimizations. Ostfeld and Salomons (2005) specified a threshold ("hurdle race") to terminate model runs at a certain point in simulation in a calibration experiment. Razavi et al. (2010) formulized a "deterministic model preemption" approach that objectively dismisses model runs that will definitely result in a poorer model performance compared to the previous best solution in calibration. The core idea in this approach is to terminate a model run as soon as it is recognized as being implausible in order to

save computation time. Deterministic preemption, as demonstrated by Razavi et al. (2010), involves terminating a model run only when it is known with certainty that the candidate parameter set is so poor that it will have no influence on the optimization algorithm search trajectory. Accordingly, a deterministic preemption threshold is objectively defined based on the optimization algorithm being used. In the DDS algorithm, this threshold is the value of the objective function for the current best solution. During subsequent model runs evaluating candidate parameter sets, the model performance is constantly monitored during each model evaluation. Once the objective function value exceeds the preemption threshold in a given evaluation, then that run is identified as implausible and thus, terminated immediately. Consequently, the remainder of the simulation time will be saved in the computation budget. Razavi et al. (2010) report deterministic preemption savings ranging from 14% to 96% across multiple calibration studies depending on the choice of optimization algorithm. Razavi et al. (2010) note that the deterministic model preemption approach is only applicable to objective functions with monotonically degrading quality, such as Sum of Squared Errors (SSE), as the simulation progresses. This preemption technique was used in Chapter 4 and Chapter 5 of this thesis during the calibration experiments performed.

Chapter 3
Calibrating Environment Canada's MESH Modelling System over the Great Lakes Basin: Setting a Benchmark

This chapter is a mirror of the following published article in the Atmosphere-Ocean journal. References are unified at the end of the thesis.

Haghnegahdar, A., Tolson, B.A., Davison, B., Seglenieks, F.R., Klyszejko, E., Soulis, E.D., Fortin, V., Matott, L.S. (2014), Calibrating Environment Canada's MESH Modelling System over the Great Lakes Basin, Atmosphere-Ocean,Vol. 52, Iss. 4, 2014.

Abstract

This paper reports on recent progress towards improved predictions of the land surface-hydrological modelling system MESH (Modélisation Environmentale–Surface et Hydrologie) via its calibration over the Laurentian Great Lakes Basin. Accordingly, a "global" calibration strategy is utilized in which parameters for all land class types are calibrated simultaneously to a number of sub-basins and then validated in time and in space. Model performance was evaluated based on four performance metrics, including the Nash-Sutcliffe (NS) coefficient and simulated versus observed hydrographs. Results from two calibration approaches indicate that in the model validation mode, the global strategy generates preferred results over an alternative calibration strategy, referred to as the "individual" strategy, in which parameters are calibrated to a single sub-basin with a dominant land type individually and then validated in another sub-basin with the same dominant land type. The global calibration strategy was relatively successful despite the high problem dimensionality (51 model parameters calibrated) and relatively small number of model evaluations (1000 parameter sets evaluated per trial) used in the automatic calibration procedure. NS values for spatial validation range from 0.10 to 0.72 with a median of 0.41 for the 15 sub-basins considered. Results also confirm that a careful model calibration and validation is unavoidable before any application of the model.

3.1 Introduction

Since the start of the predictions in ungauged basins (PUB) decade (2003-2012) by the International Association of Hydrological Science (IAHS) in 2003 (Sivapalan et al., 2003) a very large number of studies have focused on various approaches for PUB such as statistical methods, conceptual modelling and distributed modelling (e.g., Razavi & Coulibaly, 2013; Spence et al., 2013). Statistical methods and conceptual modelling for PUB normally consist of testing and comparing various regionalization

techniques such as spatial proximity, regression-based, and physical similarity approaches (e.g., Samuel et al., 2011; Zhang et al., 2008). However, when distributed physically-based models are used where parameters are tied to spatial units with land cover, soil or other characteristics, the regionalization strategies can be replaced with more global calibration experiments, especially for regional scales as is explained later in this paper.

Regional (large scale) distributed land surface-hydrological models are being increasingly utilized for the purpose of hydrologic and climatic simulations or forecasts (e.g., Flügel, 1995; Benoit et al., 2000; Arnold & Fohrer, 2005; Samaniego et al., 2010). Although some of the parameters in distributed models are readily available from field observations, many of the parameters are not measurable or are extremely hard to obtain. These parameters are typically estimated using model calibration (sometimes called inverse modelling). Model calibration is the process of tuning model parameters so that model simulations can closely replicate the historical observations. Without a careful model calibration and validation, model simulations can be very unrealistic and/or misleading. Calibration strategies range from simple manual calibration to sophisticated automated calibration. Although manual calibration provides insights about the modelling process and model structure, it is very time-consuming and limited in distributed environmental modelling due to a large number of parameters. Therefore, automated calibration approaches are often preferred over manual calibration in these cases.

In order to address different water resources management issues, Environment Canada (EC) is working on developing a modelling system called MESH (Modélisation Environmentale–Surface et Hydrologie) in which a land surface scheme is coupled with a hydrological model. MESH was applied to the Great Lakes Basin for the first time in 2007 (Pietroniro et al., 2007). Their study was mainly intended to set up the coupling framework and demonstrate its potential for operational forecasts in the Great Lakes Basin, and to identify specific requirements for different components of the MESH modelling system. Accordingly, model parameters were adopted from similar modelling applications at much smaller scales in the region and were only slightly adjusted during a brief manual calibration focused on the routing parameters with minimal changes to the parameters corresponding to vegetation or soil characteristics. Therefore, a more thorough calibration effort using automated optimization techniques was needed over the Great Lakes Basin to properly evaluate the ongoing improvement of the MESH modelling system.

In this paper, we provide the calibration and validation results from a series of model calibration experiments to improve MESH streamflow simulation in the Great Lakes region. Accordingly, a “global” calibration strategy is adopted in which model parameters are adjusted simultaneously in a number of

Great Lakes sub-basins. The results are further compared with an alternative calibration strategy ("individual") at some selected sub-basins since both approaches have been previously suggested or used in MESH-related studies. Case study and model details along with the calibration approach are described in the methodology section. Then the results and related discussion are presented followed by a conclusion.

This study is the first of its kind to report on extensive calibration and validation results (at many locations) for MESH over a large basin. Only few studies have used MESH in the past. Dornes et al. (2008) employed a step-wise calibration strategy with MESH to promote the transferability of vegetation parameters based on landscape similarity as opposed to basin attributes. Dupont et al. (2012) linked MESH with a 3-D global ocean model, NEMO (Nucleus for European Modelling of the Ocean), to simulate a number of hydrodynamic properties of the Great Lakes including lake levels variation, ice concentrations, and lake surface temperatures. Deacu et al. (2012) conducted a series of experiments with MESH to improve Net Basin Supply (NBS) predictions for the Great Lakes by modifying estimations of precipitation, evaporation, and runoff. Using the MESH predecessor, WATCLASS (Soulis et al., 2000), Davison et al. (2006) showed that much better snow cover and snow melt results can be obtained when the modelling process is modified according to the characteristics governing the snowmelt process such as windswept tundra and drifts. Only in Davison et al. (2006) and Dornes et al. (2008) was the MESH model calibration (and not necessarily validation) process reported on and these studies all used a much smaller basin (63 km^2) than the Great Lakes as well as fewer sub-basins (one or two sub-basins at most).

The Great Lakes system has paramount importance locally, regionally and globally. Thus, many agencies endeavor to provide operational forecasting tools at various temporal and spatial scales within the Great Lakes region. The Great Lakes Environmental Research Laboratory (GLERL) of the National Oceanic and Atmospheric Administration (NOAA) in the United States provides hourly forecasts of water level displacements for the Great Lakes for the next 48-hr through the Great Lakes Coastal Forecasting System (GLCFS). It also predicts monthly lake inflows, outflows and lake levels one to nine months in advance through its Advanced Hydrologic Prediction System (AHPS). Environment Canada is working towards similar type of forecasting systems and this research will contribute towards building such a system by providing an improved estimation of the lake inflows.

3.2 Methodology

3.2.1 Study Area: The Great Lakes Basin

Located in North America between Canada and the United States, the Great Lakes are composed of five interconnecting large lakes namely, Superior, Michigan, Huron, Erie, and Ontario (Figure 1). The entire watershed system has a drainage area of nearly 1 million square kilometers and includes one small lake called St. Clair (between Lake Huron and Lake Erie), four connecting channels (St. Marys, St. Clair, Detroit, and the Niagara Rivers) and the St. Lawrence River. The entire drainage basin along with the drainage areas associated with each lake and the St. Lawrence River are shown in Figure 1 below.

Figure 1. The Great Lakes Drainage Basin.

3.2.2 MESH modelling system applied to the Great Lakes Basin

MESH is a semi-distributed physically-based land surface-hydrological modelling system developed by Environment Canada for water resources management applications (Pietroniro et al., 2007; Deacu et al., 2012). MESH has evolved from WATCLASS (Soulis et al., 2000), which was a combination of the routing module of WATFLOOD model (Kouwen et al., 1993) and the CLASS (Canadian Land Surface Scheme) model (Verseghy, 1991; Verseghy et al., 1993). The interface coupling these two models retained the vertical moisture budget in CLASS, and added horizontal flow (baseflow, interflow, and overland flow) calculations to it in order to generate the total streamflow that can then be routed by WATFLOOD (Soulis et al., 2000).

WATFLOOD is a semi-distributed hydrological model. The watershed is discretized using grid cells and thus, can efficiently use remotely sensed and radar data (Kouwen et al., 1993). It is considered a semi-distributed model (as opposed to a fully distributed model) since it does not explicitly allow for the lateral exchange of water between grid cells, and there must be a river within every single grid cell to transfer the water (streamflow) from one cell to another. In order to represent the sub-grid heterogeneity, grid-cells can be further subdivided into GRUs (Grouped Response Units) (Kouwen et al., 1993) that are the computational units in the model with each one having a distinct hydrological response. GRUs are often loosely defined to aggregate multiple attributes (e.g., soil and vegetation characteristics) into a single unit. This approach increases computational efficiency in distributed models by reducing the number of model parameters and allowing for parameter transferability between the same GRUs at different locations. Pietroniro & Soulis (2003) found the GRU approach a suitable compromise for semi-distributed hydrological modelling in large basins because it is simple to implement while respecting the physics to allow for different hydrological responses. In many previous WATFLOOD, WATCLASS, and MESH applications (e.g., Kouwen et al., 1993; Soulis et al., 2000; Pietroniro & Soulis, 2003; Kouwen et al., 2005; Bingeman et al., 2006; Pietroniro et al., 2007), GRUs are delineated only based on the land cover type that is assumed to have the dominant contribution to the hydrological response.

CLASS is a physically-based land surface scheme that simulates the vertical energy and water fluxes for vegetation, soil, and snow for each GRU. The soil moisture content and its vertical movement in the soil layers (three in this study), is represented by Richard's Equation in CLASS. The horizontal flow for each GRU consists of three components in MESH: overland flow, interflow, and baseflow. The overland flow is calculated using the Manning's approximation of the kinematic wave velocity. The interflow, which accounts for the downward hill slope flow, is estimated from the bulk saturation of each soil layer calculated at each time step. Baseflow is simply treated as any water that percolates out of the bottom of

the soil column in a GRU and is immediately added to streamflow. The total amount of surface runoff (as well as other fluxes) is then calculated for each grid cell by the areal average of GRUs within that grid cell. Then water is routed by the WATFLOOD routing scheme between the grid cells and across the river network for the entire basin using the Manning's formula combined with a simple application of the continuity equation.

Due to its advanced land surface scheme, the meteorological forcing data requirements of MESH are higher than hydrological models that rely on less physically-based algorithms such as temperature-index snowmelt models. The MESH model requires forcing data that include precipitation, air temperature, wind speed, incoming short wave radiation, incoming long wave radiation, specific humidity, and barometric pressure. The archived gridded forecasts for the forcing data (excluding precipitation) are produced by the Regional Deterministic Prediction System (RDPS) of EC, which is based on the regional configuration of the Global Environmental Multiscale-Numerical Weather Prediction (GEM-NWP) model (Mailhot et al., 2006). The spatial resolution of the RDPS for the time period in question is 0.1375 degree (about 15 km) and the model time step is 7.5 minutes. This was reinterpolated to a spatial scale of 0.1667 degree (about 18 km) and a temporal resolution of 1 hour. Although changes to the spatial and temporal resolution of the RDPS did not occur for the time period of the study, other changes may have an impact on the quality of the forcing data. These changes include improvements to the surface temperature forecasts in January, 2005; improvements to the quantity of precipitation forecast in summer situations in July, 2005; assimilation of new satellite data in December, 2005; various changes to the regional forecast system in July, 2007; improvements to the assimilation system in May, 2008; improvements to the data assimilation and forecast system in March, 2009; extension into the Middle Atmosphere of the forecast system in June, 2009; and updates to the data assimilation system in December, 2009. The gridded precipitation data are obtained from the Canadian Precipitation Analysis (CaPA, Mahfouf et al., 2007), which uses a statistical interpolation method to combine ground-based point observations with the gridded GEM precipitation forecasts as a background field. The observational database used by CaPA consisted of surface synoptic reports of 6-hour precipitation accumulation from the Meteorological Service of Canada's historical weather and climate archives. These archives include data from four networks across North America (SYNOP, METAR, SHEF, and RMCQ). The SYNOP network contains ~300 manned and ~750 automated weather stations maintained by Environment Canada and partner organizations. The METAR network contains over 1300 stations that are typically located at airports. The SHEF network is an American cooperative network of more than 11,000 volunteers who collect observations in the United States. The RMCQ network is a cooperative network of private companies and provincial agencies in Quebec, Canada containing about 200 stations. CaPA is used

instead of GEM precipitation because, as an assimilation of observations with GEM precipitation, it is considered to be the most accurate precipitation product available at the spatial scale of this study.

The MESH configurations remained largely unchanged relative to the original application of MESH to the Great Lakes Basin as described in Pietroniro et al. (2007). The drainage area of approximately 920,000 km^2 was modeled with 4026 grid cells of 1/6 degree size. Each cell was further sub-divided into up to seven land class based GRUs namely: crop, grass, deciduous forest, coniferous forest, mixed forest, water and impervious. For routing purposes, MESH can use up to five river classes, which were newly identified in this study for the entire basin on the basis of the nature of the streamflow and natural boundaries.

This research is based on MESH version 1.3.003, which utilizes CLASS version 3.5. Numerical experiments were performed using an Intel Core i7-920 2.67 GHz workstation running a 64-bit version of the Windows 7 operating system with 12 gigabytes of installed memory and a quad-core processor. Model simulations were performed using hourly numerical time steps and one simulation took approximately 27 minutes to simulate daily flows at 10 sub-basins for 16 months (used in model calibration). This relatively long runtime was one of the main factors in deciding on the maximum number of model runs during calibration as explained later. Moreover, MESH is not coded to take advantage of parallel computing. Parallel optimization techniques were also not considered in the current study.

3.2.3 Calibration Strategy

Hydrologic models should be calibrated and validated before they can be used for any research or practical purposes. However, no universal guidelines exist for such procedures particularly in the field of environmental modelling (Moriasi et al., 2012). Consequently, the choice of calibration and validation strategy is subjective and numerous methods exist ranging from simple manual calibration to sophisticated automated calibration techniques or a combination of both.

Similarly, various calibration strategies have been used in the past for WATFLOOD or MESH-related studies among which two strategies were identified as the top two alternatives based on model developers' opinion and previous studies. The first strategy (e.g., Bingeman et al., 2006), hereafter called the "global" strategy, was used as the primary approach in this paper for the purpose of model calibration and validation. In the global strategy, all parameters for all GRUs (land class types) are calibrated simultaneously considering all calibration sub-basins and then validated in space and time for other flow gauges. Ricard et al. (2013) suggested a similar strategy for calibrating a semi-distributed model at a large

scale over the St. Lawrence River basin in Quebec, Canada. In the second strategy, called "individual" strategy hereafter, parameters associated with a certain GRU type (land class), are calibrated to a single sub-basin dominated by that GRU and then validated for another sub-basin with the same dominant GRU. This latter strategy was tested against the global strategy only for two GRUs at selected gauges. Details explained below refer to the global strategy in general, unless otherwise expressed.

The calibrated models were tested in validation over both time and space. Temporal model validation is essentially the classic split-sample test defined by Klemeš (1986) which splits the observations at a single site between those used for calibration and those used for validation. Spatial model validation is based on two approaches. The first approach is to check validation performance at locations upstream of calibration gauges. The second approach, which included 12 of 15 spatial validation sub-basins, was essentially the proxy-basin test as defined by Klemeš (1986) where validation sub-basins were limited to basins that were spatially independent from calibration sub-basin drainage areas. It should be noted that the only difference between our application of the split sample test and proxy-basin test from Klemeš (1986) is that our tests were one-way tests while Klemeš (1986) advocates these should be two-way tests (switching time periods or locations between calibration and validation). Two-way tests would generate two calibration solutions.

Out of the hundreds of hydrologic sub-basins available through the Water Survey Canada and the U.S. Geological Survey (USGS) for the Great Lakes basin, 10 sub-basins were selected for calibration and temporal validation (total area of 50,243 km^2), and another 15 for model validation in both space and time (total area of 49,559 km^2). Drainage areas of the selected sub-basins ranged approximately from 470 km^2 to 16400 km^2. A multi-step procedure was followed for selecting calibration and validation sub-basins. First, only sub-basins with drainage areas larger than two grid cells (nearly 450 km^2) were considered. These sub-basins were further filtered based on the criteria that flow data should also be available for the period of study. Finally, sub-basins with completely regulated flows were eliminated from consideration. The sub-basins selected by the aforementioned procedure are shown in Figure 2. Note that there are four pairs of these sub-basins that have overlapping drainage areas as seen in Figure 2: a) 04191500 and 04193500, b) 04237500 and 04232000, c) 04116000 and 04114000 and d) 02GA018 and 02GB018. In each of these pairs (except case d), the upstream gauge is used as a validation location.

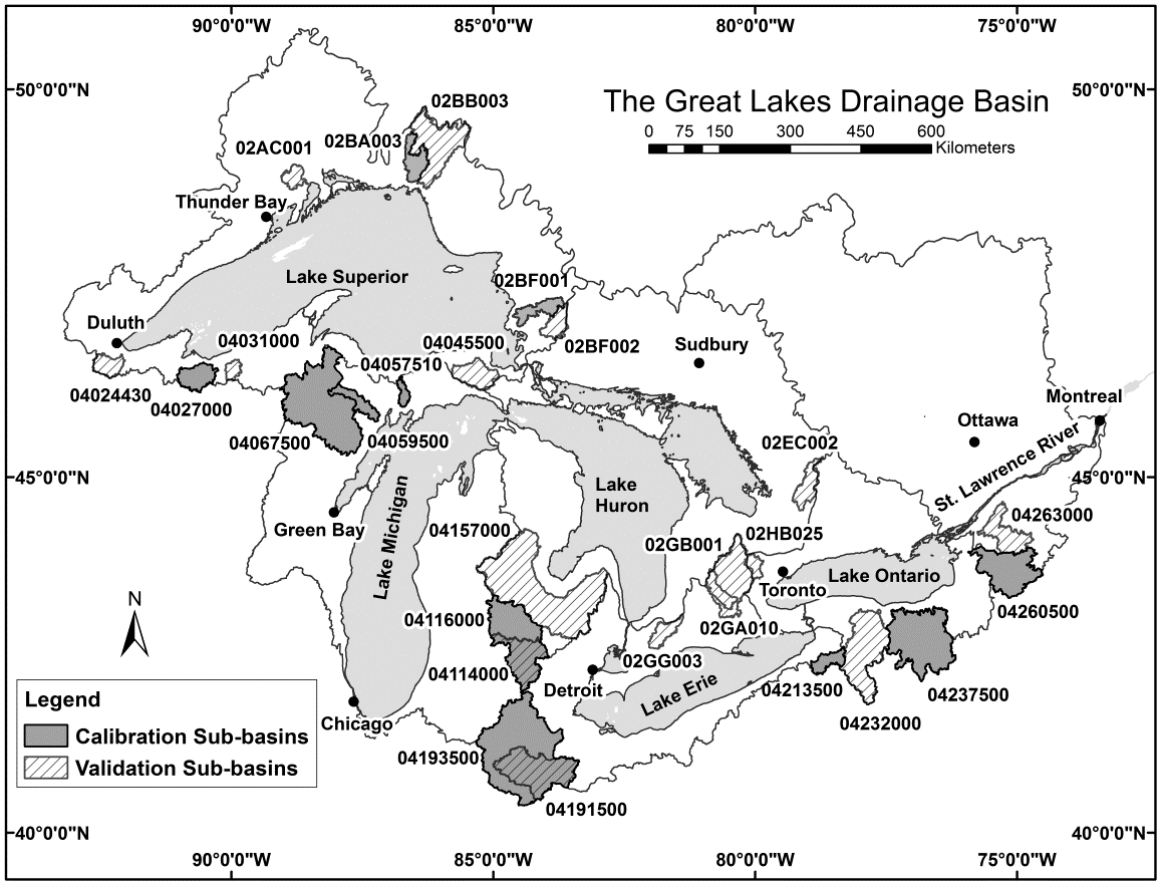

Figure 2. Sub-basins used in calibration and validation. Sub-basins 04213500 and 04263000 are only used in the individual strategy. All other sub-basins are used in the global strategy.

The choice of calibration and validation time periods is limited by the availability of the preferred precipitation forcing data (CaPA as described above). Accordingly, after excluding a four month simulation initialization (i.e. warm-up) period, October 2004 to September 2005 was selected for model calibration, and October 2005 to May 2009 was selected for temporal model validation. For the selected calibration sub-basins, the low, high, and average daily flow rates during the calibration period were considered reasonably representative of the entire study period. In other words, extreme events and outlier data points were not found in the selected calibration data. The starting and ending months for these periods were selected at the end of the spring snow-melt period. It is acknowledged that the calibration result can always be sensitive to the choice of calibration period and its length (Gharari et al., 2013). A relatively small calibration period was selected due to computational limitations and the quality of this choice is implicitly evaluated based on validation performance.

The calibration parameters and their corresponding ranges were specified using a combination of expert opinion from MESH developers and previous modelling studies (e.g., Dornes et al., 2008; Davison et al., 2006; Bingeman et al., 2006). Table 1 below lists the selected parameters based on the land class-based GRU definition in MESH. The majority of these parameters (one to nine in Table 1) are related to physiography, vegetation, and soil characteristics and are calibrated for five GRU types being crop, grass, deciduous forest, coniferous forest and mixed forest. Parameters associated with water and impervious GRUs were not calibrated.

One of the calibration parameters is the GRU overland slope. This parameter is used in the calculation of overland flow and interflow in MESH. Although this slope could be assigned based on DEMs for each pixel, given the large pixel size (approximately 15 km by 15 km) and thus correspondingly large GRUs within the grid, we decided to treat these GRU type specific slopes as an effective parameter to be calibrated. Davison et al. (2006) and Dornes et al. (2008) also followed this approach and linked slopes to the land class type. As for the soil constituent percentages (percent sand, silt, and clay), only sand and clay content were calibrated and the silt content was calculated by difference to ensure the sum of soil constituent component percentages added to 100%.

The soil permeable depth or depth to bedrock (SDEP) was calibrated as one parameter that is constant across all GRU types in the global calibration strategy. This parameter varied with sub-basin for the individual calibration strategy, which gave this strategy more flexibility than the global strategy. The channel roughness coefficient (WFR2) is an effective parameter encompassing the combined effects of Manning's coefficient and channel width to depth ratio. In the global strategy, five channel roughness coefficients (WFR2) were calibrated corresponding to the five river classes identified across the Great Lakes Basin as explained earlier. Consequently, a total of 51 parameters were calibrated for the global strategy. In the case of individual strategy, depending on the size and location of the sub-basins, one or two channel roughness coefficient(s) were calibrated resulting in a total of 12 and 11 calibration parameters for GRU1 (crop) and GRU3 (deciduous forest), respectively. Other required parameters in the model were set according to previous studies or values suggested by MESH model developers for this case study.

Given the case study scale, this study requires calibration to multiple sites and thus the calibration problem is inherently multi-objective. The goal of the study was to calibrate MESH deterministically and hence identify a single best calibration solution. There are multiple calibration approaches demonstrated in the literature to achieve this goal. Some authors address the problem in two steps by 1) posing and solving a multiple-criteria optimization problem to generate a 'Pareto set' of non-dominated solutions and

then 2) applying some decision-making procedure to select their preferred single solution. See for example Gupta et al. (1998) or Efstratiadis & Koutsoyiannis (2010) for a description of this approach. As described in Gupta et al. (2009), other authors prefer to address the problem in a single step by defining an aggregated objective (based on multiple calibration metrics or based on multiple calibration sites) and then applying a single-objective optimization algorithm to yield a single calibration result (Lindström, 1997; Bergström et al., 2002; Madsen, 2003; Parajka et al., 2005; Young, 2006; Rode et al., 2007; Marcé et al., 2008; Wang et al., 2009; Safari et al., 2012).

Table 1- Calibration parameters, their ranges and the calibrated values using the global strategy for the MESH modelling system applied to the Great Lakes Basin.

Parameter	**Description**	**Range**	**Calibrated Values for Crop, Grass, Deciduous, Coniferous, Mixed Forest**
ROOT	Annual maximum rooting depth of vegetation category [m]	1.0 - 3.5	1.096, 2.142, 2.353, 2.382, 1.299
QA50	Reference value of incoming shortwave radiation [W m^{-2}]	30.0 - 50.0	49.209, 48.878, 35.206, 49.906, 46.534
RSMN	Minimum stomatal resistance of vegetation category [s m^{-1}]	50 - 300	290.940, 240.893, 269.698, 299.457, 288.279
DRN	Soil drainage index	0.0 - 1.0	0.403, 0.490, 0.409, 0.217, 0.723
DDEN	Drainage density [km/km^2]	2.0 - 100	24.018, 8.773, 11.053, 48.288, 3.521
XSLP	The average overland slope of a given GRU [%]	0.0001 - 0.1* (0.04 for crop)	0.039, 0.046, 0.016, 0.010, 0.084
WFCI	Saturated surface soil conductivity [m/s]	0.0 - 1.02	0.040, 0.533, 0.135, 0.866, 0.529
SAND	Percentage sand content [%]	0.0 - 50.0	47.654, 6.991, 4.807, 19.283, 38.306
CLAY	Percentage clay content [%]	0.00 - 50.0	33.001, 23.334, 15.235, 8.233, 28.773
SDEP	Soil permeable depth [m]	0.35 - 4.1	4.095
WFR2	WATFLOOD channel roughness coefficient	0.02 - 2.0	0.698, 0.187, 0.415, 0.024, 0.823 (for the 5 river classes)

In our study, we chose the second approach and calibrated to the multiple sites by optimizing an aggregated objective function. The main benefit to this approach is that it is more computationally efficient than attempting to generate multiple Pareto set solutions. Hourly MESH simulated flows were aggregated to average daily simulated flows and were calibrated and validated against observed daily average streamflows using a weighted sum of Nash-Sutcliffe coefficients of efficiency (Nash & Sutcliffe, 1970). The weights were assigned proportional to the mean measured flow values of sub-basins over the

calibration period. This choice of weights clearly favors larger basins and was adopted because one end goal of this model is to forecast total lake inflows and hence we deemed fitting larger basins (contributing more flow into lake) more important than fitting smaller basins. It was observed that the model performance across the validation sub-basins shows no relation to these weights (i.e. drainage areas). It is acknowledged that there exist alternative choices for the objective function and other criteria can be added for a possible gain in model assessment (e.g., Legates & McCabe, 1999; Moriasi et al., 2007). However, the Nash-Sutcliffe coefficient of efficiency still stands as one of the most widely used criterion and was considered a reasonable choice. All validation results, and the calibration results when the global strategy is compared to the individual strategy, are further assessed with three other performance metrics, in addition to the Nash-Sutcliffe coefficient of efficiency. Each metric is defined in Table 2.

Table 2- Definition of various model performance metrics.

Metric Equations	**Notes**
Nash-Sutcliffe Coefficient of Efficiency: $NS = 1 - \frac{\sum(Q_{obs} - Q_{sim})^2}{\sum(Q_{obs} - \overline{Q_{obs}})^2}$	Nash and Sutcliffe (1970) Range: -∞ to 1, where 1 is best, unitless
Volumetric Efficiency: $VE = 1 - \frac{\sum\|Q_{obs} - Q_{sim}\|}{\sum Q_{obs}}$	Criss & Winston (2008) Range: 0 to 1, where 1 is best, unitless
Root Mean Squared Error: $RMSE = \sqrt{\mathrm{N}^{-1}\sum(Q_{obs} - Q_{sim})^2}$	Range: 0 to +∞, where 0 is best, units of Q
Percent Bias: $PBIAS = \frac{\sum(Q_{obs} - Q_{sim})}{\sum Q_{obs}} * 100$	Gupta et al. (1999) Range: -∞ to 100%, where 0 is best, units of %
Q_{obs}: Observed flow, Q_{sim}: Simulated flow, $\overline{Q_{obs}}$: Mean of observed flows, N: Number of flow observations	

The Dynamically Dimensioned Search (DDS) algorithm (Tolson & Shoemaker, 2007) was selected as the automatic calibration tool for this study. DDS is well suited for optimization problems with a large number of calibration parameters, such as distributed hydrologic model calibration and example applications of DDS for calibrating distributed models include Wallner et al.(2012), White et al. (2011), Matott et al. (2011), and Clark et al. (2008). DDS was designed specifically for automatic calibration of

hydrologic models and the algorithm is able to rapidly converge to a good calibration solution and easily avoids poor local optima (Tolson and Shoemaker, 2007).

DDS is a direct (derivative-free) optimization method that stochastically searches around the best solution (decision variable values) identified so far. Candidate solutions are compared to the current best solution to determine if an update to the best solution is required. DDS starts by searching globally and transitions to a more local search as the number of objective function evaluations approaches the user-specified maximum. The adjustment from global to local search is achieved by dynamically and probabilistically reducing the number of dimensions in the search neighborhood (i.e. the set of decision variables modified from their best value). Since DDS, like most global optimizers, is stochastic due to the use of random numbers, optimization results can vary between optimization trials.

DDS is very simple to use as it requires only one parameter (perturbation parameter) that was set to 0.2 by default as suggested by Tolson and Shoemaker (2007). The maximum number of model simulations (objective function evaluations) also has to be specified in DDS as the stopping criteria and was set to 1000 in this study based on the computational burden imposed by the model as discussed earlier. This number also seems to work well according to some previous studies with MESH (e.g., Dornes et al., 2008; Razavi et al., 2010). The model calibration was conducted using DDS as implemented in OSTRICH, which is a model-independent calibration and optimization tool (Matott, 2005) consisting of a number of popular optimization algorithms including DDS.

Our final best global parameter set is the outcome of a series of calibration experiments that evolved over time. Initially, the global strategy was run for one optimization trial with an initial set of parameters and ranges according to some past studies (e.g., Dornes et al., 2008; Davison et al., 2006; Bingeman et al., 2006). Then, calibration parameters and their ranges were modified through further consultation with model developers. The final refined global and individual calibration experiments in the series utilized the parameters and corresponding ranges listed in Table 1 and was repeated twice for the global experiment and for each individual calibration case (two independent optimization trials from different initial solutions). The calibrated parameter set was taken to be the best solution of the multiple trials for the calibration period and these parameters are reported in Table 1 and were applied to generate the validation results.

3.3 Results and Discussion

In this section, first we compare the two calibration strategies for select sub-basins dominated by the crop and deciduous forest GRU types. Then, detailed results for the global calibration strategy are presented

for the set of ten calibration and 15 validation sub-basins. In each case, results are presented using the performance metrics of Table 2 and various select hydrographs.

Comparison results are presented for sub-basins dominated by one GRU type in Table 3 (crop) and Table 4 (deciduous forest). In Table 3, performance metrics for the calibration sub-basin and validation sub-basins are given for crop dominated sub-basins. For the crop dominated sub-basins, when parameters are calibrated based on the individual strategy (IS), the calibration NS values are higher than the global strategy (GS) values. However, validation results at the four sub-basins show NS values that are either very similar for both IS and GS (two sub-basins) or better for GS (sub-basins 02GA010 and 02GB001). PBIAS results for validation sub-basins are all reasonably close to 0 with both strategies (ranging from -6% to 15%). A similar pattern is observed for sub-basins dominated by deciduous land type as indicated in Table 4. Calibration metrics for IS are better than or similar to those of GS. However, unlike calibration period, performance results for NS, VE and RMSE are better for GS in validation especially for NS. Although similar, IS shows a slightly better PBIAS result (21%) in comparison to GS (28%).

Table 3. Comparison of Individual Strategy (IS) vs. Global Strategy (GS) for Crop land class in calibration and validation sub-basins based on various metrics.

Sub-basin ID (% crop)		**NS**		**VE**		**RMSE (m^3/s)**		**PBIAS (%)**	
		IS	**GS**	**IS**	**GS**	**IS**	**GS**	**IS**	**GS**
Calibration	04193500 (96)	0.79	0.67	0.55	0.36	150	188	4	-1
Validation	02GG003 (98)	0.62	0.58	0.44	0.39	14	15	-6	4
Validation	04191500 (98)	0.54	0.58	0.30	0.31	113	109	11	8
Validation	02GA010 (97)	0.13	0.38	0.23	0.46	17	14	1	15
Validation	02GB001 (89)	-0.23	0.33	0.25	0.52	76	57	2	14

Table 4. Comparison of Individual Strategy (IS) vs. Global Strategy (GS) for Deciduous Forest land class in calibration and validation sub-basins based on various metrics.

Sub-basin ID (% Deciduous Forest)		**NS**		**VE**		**RMSE (m^3/s)**		**PBIAS (%)**	
		IS	**GS**	**IS**	**GS**	**IS**	**GS**	**IS**	**GS**
Calibration	04213500 (96)	0.63	0.43	0.62	0.50	16	20	10	12
Validation	04263000 (96)	0.20	0.41	0.51	0.59	39	34	21	28

In summary, the numerical results in Table 3 and Table 4 show the individual strategy is somewhat worse than the global strategy at three of the five validation basins (02GA010, 02GB001 and 04263000).

For the other validation basins, results are fairly similar for the suite of metrics considered. Even if these comparative results were to be interpreted that the global and individual strategies were of similar overall quality, the global strategy has the advantage that the entire basin is calibrated – parameter values in ungauged basins are a direct result. With the individual approach this is not the case since there are other GRUs that require parameter values and, in addition, subjective decisions are required in order to select which basins are defined as the individual calibration sub-basins for each GRU type. Therefore, the global calibration strategy is used for the purpose of predictions in ungauged basins for the remainder of results.

Although utilizing multiple calibration sub-basins for the individual strategy might improve results, the identification of sub-basins clearly dominated by a single GRU type for all GRU types becomes more challenging as many sub-basins are not clearly dominated by a single GRU type (i.e. 50% to 80% is one GRU type). Another major factor contributing to the failure of the individual strategy could be the definition of GRUs solely based on land cover types in this study. This deficiency can clearly result in degradation of the model predictions when parameters are transferred to a sub-basin where the same land cover type sits on a different soil texture. It is possible that GRUs defined on the basis of other factors (e.g., soil type) may yield more transferable model parameters in the individual strategy.

The complete results for the global calibration strategy are presented below in two parts: calibration sub-basins and validation sub-basins. Ten sub-basins covering all land classes were used for model calibration (October 2004 to September 2005) as well as temporal validation (October 2005 to May 2009) and 15 other sub-basins were used for the spatial model validation on the entire period (October 2004 to May 2009).

Figure 3 depicts the NS values corresponding to the calibration and validation periods for the ten calibration sub-basins. Since the objective function used in calibrating the model was the weighted NS of the calibration sub-basins, this value is also provided for calibration and validation periods. The weighted NS is 0.67 and 0.41 over the calibration and validation periods, respectively. The initial weighted NS over the calibration period prior to this calibration was improved from -0.3 to 0.67. After calibration, all NS values in the calibration period, and all but one NS value in the validation period are positive and the worst is a NS of -0.01. For the calibration period, the average NS value is 0.57 and ranges from 0.31 to 0.79. For the validation period, the average NS value is 0.37 ranging from -0.01 to 0.59. Figure 4 shows the hydrographs corresponding to the worst (lowest) and the best (highest) NS values of the 10 calibration sub-basins over the validation period. Figure 4a for sub-basin 04116000 shows that the model greatly overpredicts peak flows (especially in Oct 2008). Close inspection of the drainage area for sub-basin 04116000 revealed that the flow is impacted by upstream reservoirs and should be removed from the

calibration sub-basin list in further calibration. Calibration period results for this sub-basin do not exhibit the same degree of model overpredictions as the validation period.

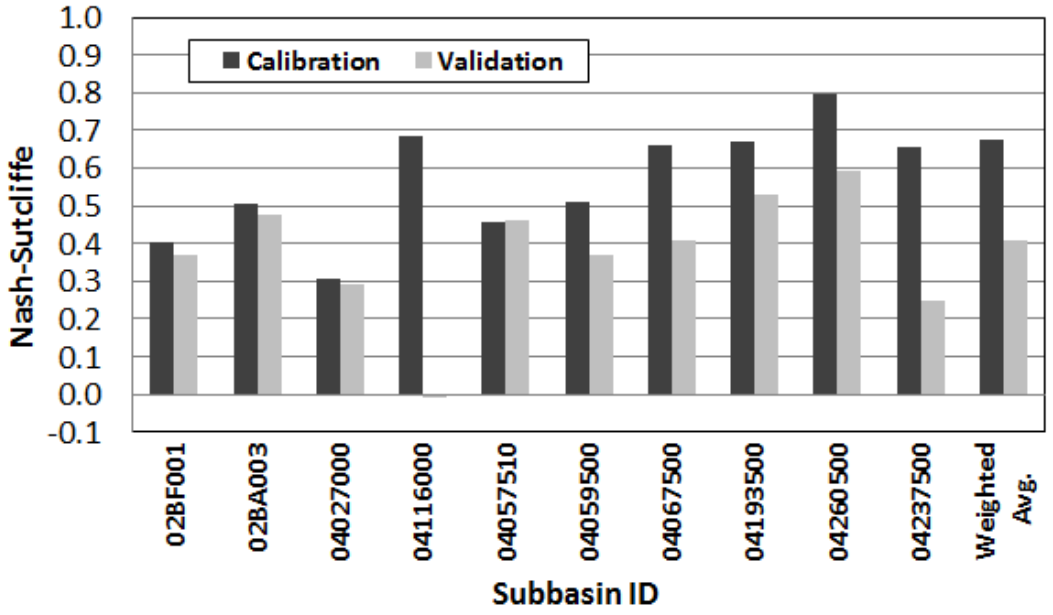

Figure 3. Nash-Sutcliffe values over the calibration (October 2004–September 2005) and validation (October 2005–May 2009) periods for the 10 calibration sub-basins using the global calibration strategy.

MESH performance, according to the four metrics defined in Table 2, is summarized in Table 5 for the 15 (spatial) validation sub-basins distributed across the basin. These metrics are calculated for the October 2004 to May 2009 period and indicate reasonably good overall performance of the model. All the NS values are positive and range from 0.10 to 0.72 with a median of 0.41. VE ranges from 0.31 to 0.62 with a median of 0.52. RMSE changes from 7 m^3/s to 128 m^3/s, and has a median of 18 m^3/s. PBIAS results indicate a tendency of the model to slightly overpredict flows (ranges from 8 % to 54% with a median of 21%).

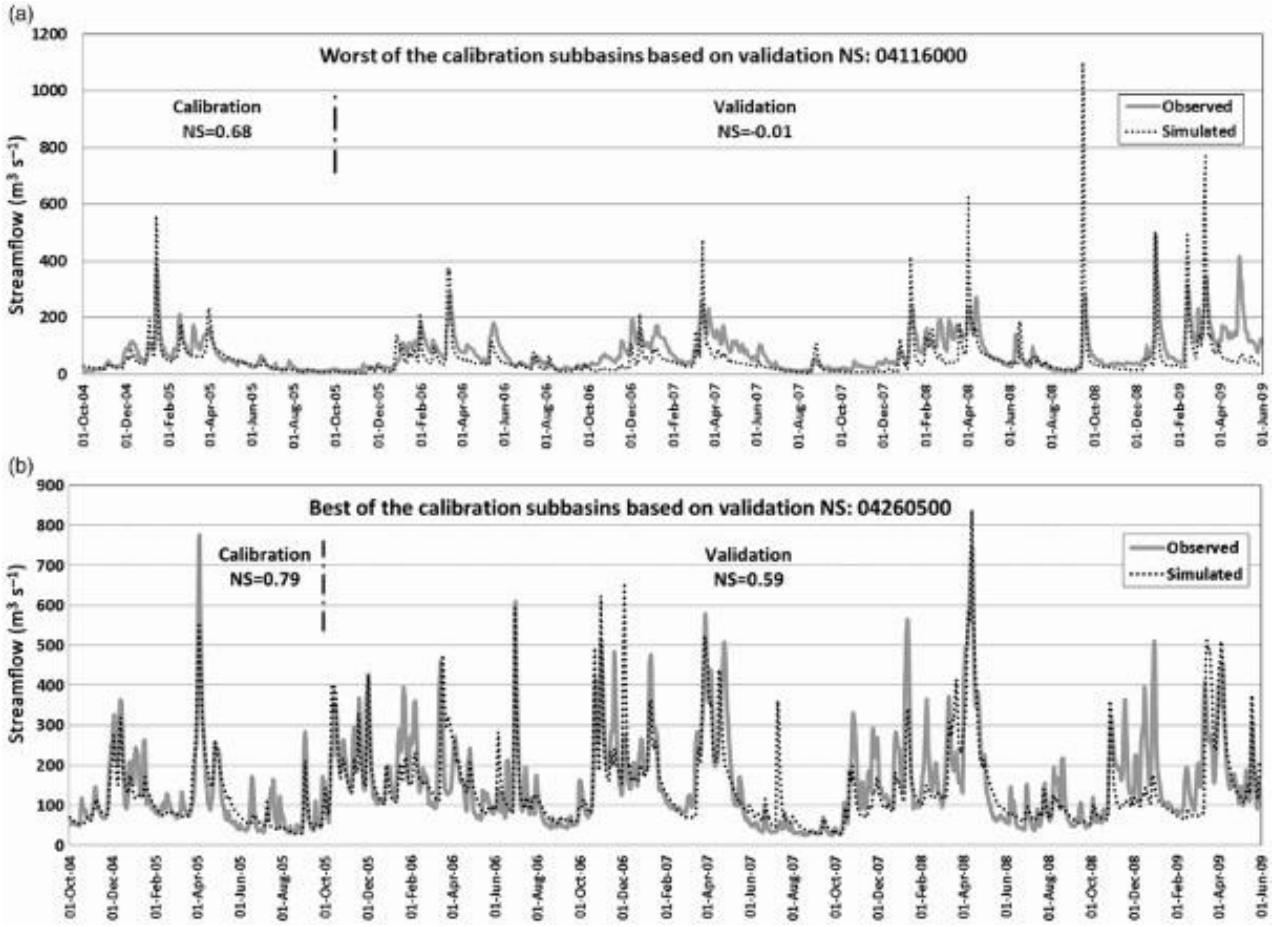

Figure 4. Hydrographs corresponding to (a) the worst (lowest) and (b) the best (highest) NS values, over the validation period (October 2005–May 2009), for the 10 calibration sub-basins.

Table 5. Median, Minimum and Maximum values for the four metrics defined in Table 2 over the Oct. 2004 to May 2009 period for the 15 (spatial) validation sub-basins using the global calibration strategy.

	NS	VE	RMSE (m^3/s)	PBIAS (%)
MEDIAN	0.41	0.52	18	21
MIN	0.10	0.31	7	8
MAX	0.72	0.62	128	54

Hydrographs corresponding to the worst and the best NS values for the spatial validation sub-basins are plotted in Figure 5. The worst NS value (0.10) is for sub-basin 04024430 and Figure 5a shows that peak flows are both severely overpredicted and underpredicted. In contrast, Figure 5b for the best sub-

basin (02EC002, NS=0.72) shows temporal trends and peak flows in the observations are all simulated closely. Low flows are either underestimated (in the worst sub-basin) or overestimated (in the best sub-basin).

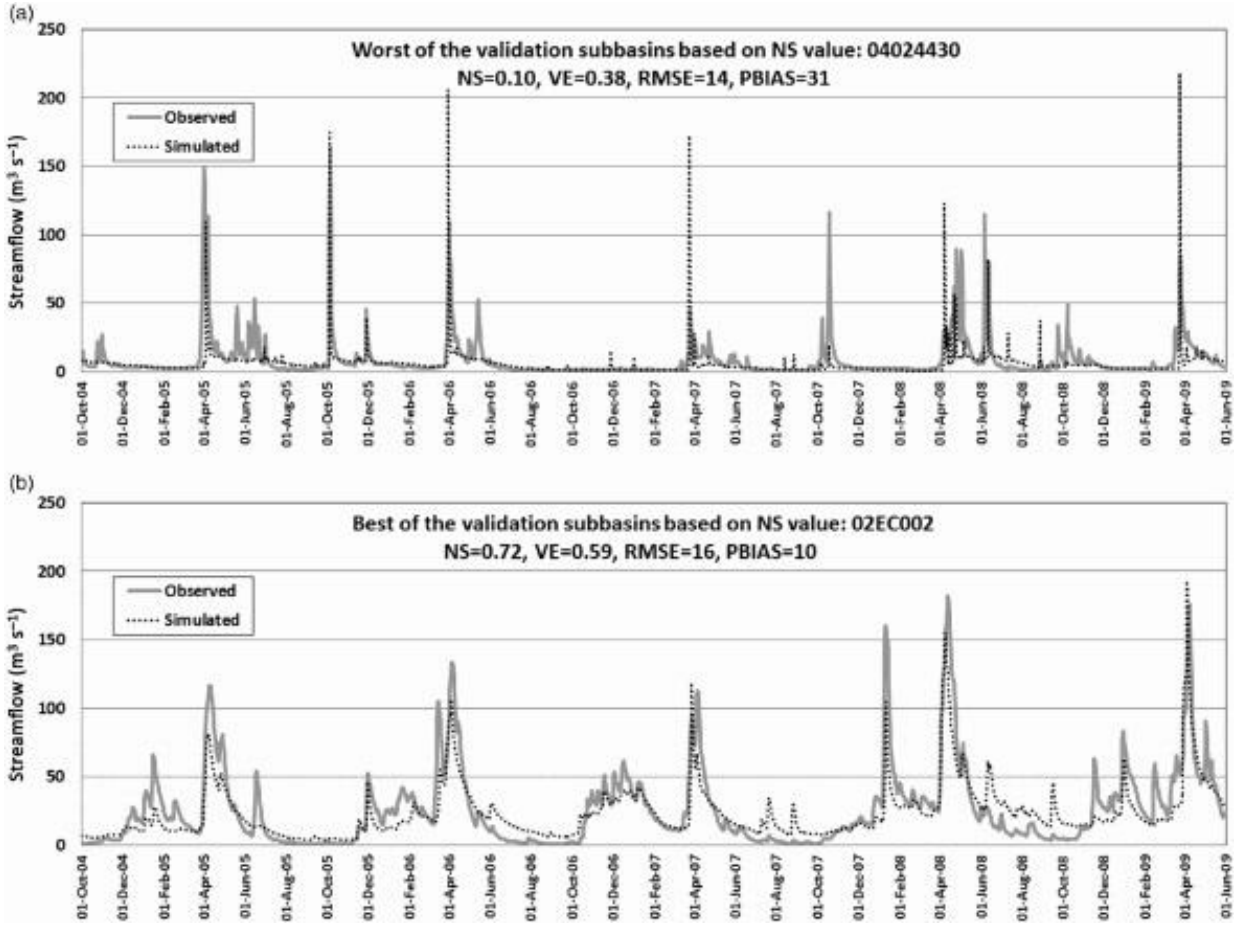

Figure 5. Hydrographs corresponding to (a) the worst (lowest) and (b) the best (highest) Nash-Sutcliffe values for the 15 (spatial) validation sub-basins over the October 2004 to May 2009 period.

In order to evaluate the improvement of the model due to calibration, we compared the pre-calibration and post calibration hydrographs and performance statistics. Pre-calibration hydrographs are defined as the solution where all the parameters are assigned the midpoint of their ranges used in calibration (see Table 1 for ranges). Figure 6 is one example calibration period hydrograph showing this comparison for the sub-basin showing the median quality NS pre-calibration value. Results show that the pre-calibration hydrograph is not at all as flashy as the observed data and the calibrated hydrograph does a much better, although far from perfect, job simulating timing and magnitude of flow events. Figure 7 compares

calibration period NS values for pre-calibration and calibrated hydrographs for all calibration sub-basins and shows that for all but one of the ten sub-basins, the calibrated NS values are significantly improved over pre-calibration results, where six sub-basins have a negative NS value.

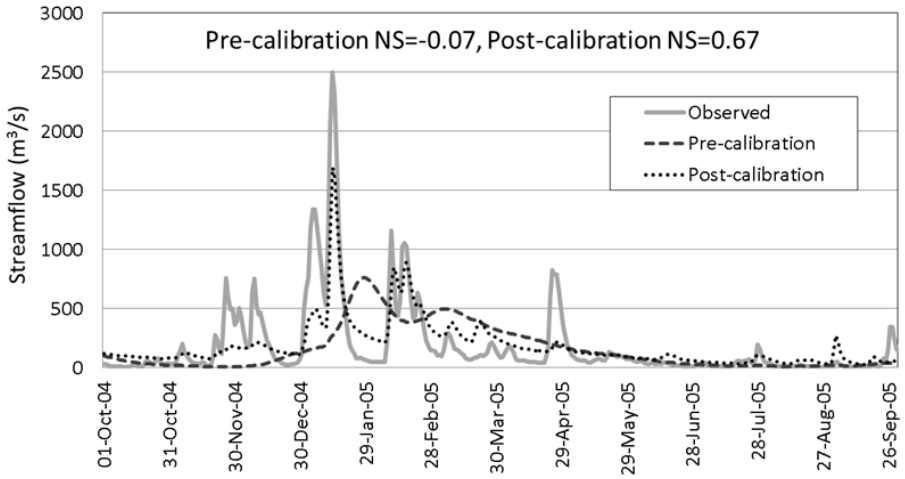

Figure 6. Comparison of pre-calibration and post-calibration (global strategy) hydrographs for the sub-basin with the median pre-calibration Nash-Sutcliffe value (04193500).

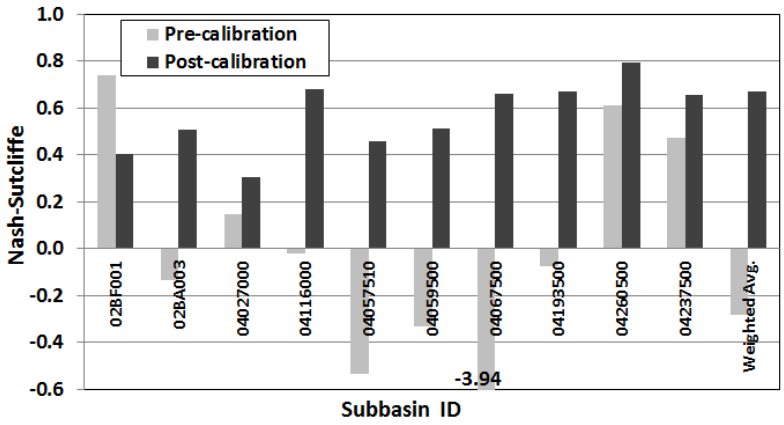

Figure 7. Comparison of pre-calibration and post-calibration (global strategy) Nash-Sutcliffe values for the calibration period (October 2004–September 2005) for the 10 calibration sub-basins.

3.4 Conclusions

The MESH modelling system over the Great Lakes Basin, was calibrated and validated (both in time and space) against measured streamflows using the Nash-Sutcliffe coefficient, three other performance

metrics, and simulated versus observed hydrographs as the model assessment criteria. The global calibration strategy, in which a number of sub-basins containing all land classes are calibrated simultaneously, generate preferred results over the strategy in which one single sub-basin with a certain land type is calibrated individually. This result is consistent with findings of Ricard et al. (2013) who calibrated a semi-distributed model in the St. Lawrence River basin, Quebec, Canada. Insufficient definition of GRUs solely based on land cover types also can negatively impact the predictions results when parameters are transferred to other sub-basins. It was also observed that, as expected, model performance substantially improved after calibration compared to precalibration hydrographs, indicating the vital role of a reasonably thorough model calibration and validation before any application. Future work to improve the MESH modelling system for the Great Lakes involves further modifying calibration strategy as well as modifying the GRU definition to account for other factors such as soil type and topography that contribute to hydrologic response of the watershed.

Chapter 4
Enhancing the MESH Model by Improved Calibration Strategy and Adding Parameter Uncertainty to Model Predictions

This Chapter is not submitted for publication yet. It is, however, intended for submission to the Canadian Water Resources journal. To avoid redundant material with respect to Chapter 3, repetitive materials such as the case study and model description are removed and only the new methodology and results are presented.

4.1 Introduction

Large scale hydrological models have been used in a variety of water and environmental resources management applications such as for assessing human impacts on the global water budget (e.g., McCabe and Wolock, 2011), and calculating interbasin water transfers (e.g., Islam et al., 2007). The St. Lawrence River Basin in North America contains the largest body of surface fresh water, The Great Lakes, and is of paramount importance globally and in particular for United States and Canada. A proper hydrological modelling and basin-wide water budget for the Great Lakes Basin allows for better addressing some of the challenges associated with this valuable water resource. Persistent extreme low lake levels are one of the critical challenges facing the Great Lakes (e.g., Gronewold and Stow, 2014), which has huge impact on the ecosystem and economy of North America (e.g., Millerd 2005).

The United States National Oceanic and Atmospheric Administration (NOAA) and Environment Canada (EC) are experimenting with various modelling systems for better understanding and forecasting the dynamics of the Great Lakes system. For example, the Great Lakes Environmental Research Laboratory (GLERL) of NOAA in the United States provides hourly forecasts of water level displacements for the Great Lakes for the next 48 hours through the Great Lakes Coastal Forecasting System (GLCFS). It also predicts monthly lake inflows, outflows and lake levels one to nine months in advance through its Advanced Hydrologic Prediction System (AHPS).

Environment Canada applied its Modélisation Environnementale—Surface et Hydrologie (MESH) model to the Great Lakes watershed (Pietroniro et al. 2007) for various water management purposes. For example, Dupont et al. (2012) linked MESH with a 3-D global ocean model, NEMO (Nucleus for European Modelling of the Ocean), to simulate a number of hydrodynamic properties of the Great Lakes including lake level variation, ice concentrations, and lake surface temperatures. Deacu et al. (2012) conducted a series of experiments with MESH to improve Net Basin Supply (NBS) predictions for the

Great Lakes by modifying estimates of precipitation, evaporation, and runoff. Gronewold and Fortin (2012) emphasized the importance of improving Great Lakes basin-wide runoff estimates through a bi-national collaboration. The improved runoff simulations result in improved lake inflow estimation that is one of the main components needed in calculating the Great Lakes' water levels.

Gronewold et al. (2011) highlighted the importance of including prediction uncertainty for the lakes' water levels forecasts based on the assessment of uncertainty in the hydrological results of the hydrological models. Despite all advances made over the past decades in understanding and simulating various natural processes in models, they are still far from reality and subject to different levels of uncertainty. Therefore, results from modelling studies are more effective when accompanied by some sort of uncertainty estimation. The International Association of Hydrological Science (IAHS) placed "estimation of predictive uncertainty, and its subsequent reduction" at the core of the decade (2003–2012) on Predictions in Ungauged Basins (PUB) in 2003 (Sivapalan et al., 2003).

In Chapter 3, as the first extensive effort to calibrate MESH for the Great Lakes Basin, benchmark results were obtained for the MESH model performance to predict streamflows in ungauged basins. The baseline results were quite promising and were immediately used in some other studies such as the Great Lakes Runoff Intercomparison Project (GRIP, Fry et al., 2014). As MESH continued to progress over time, so did the strategy in this research. Further investigations identified certain areas where there was still room for further enhancing the MESH streamflow predictions with respect to the baseline results presented in the previous chapter. Accordingly, as another contribution of this research, a new calibration experiment is conducted using an improved calibration approach. These improvements are highlighted in this chapter, and are primarily focused on refining the number of sub-basins and the parameters used in calibration. In order to show the new calibration experiment is improved over the old experiment in Chapter 3, the results between the two experiments are compared in this chapter. As another contribution of this research, prediction uncertainty associated with estimation of the model parameters is also presented in this chapter. The extent of the case study in this work requires substantial computational effort. Therefore, some strategies to reduce the calibration time are also discussed here. Finally, a local sensitivity test is conducted after calibration to identify the most influential parameters of MESH in the Great Lakes Basin.

It is expected that the final calibrated MESH model along with the uncertainty assessment reported in this work can provide an enhanced and effective tool for other researchers studying various aspects of the Great Lakes System.

4.2 Methodology

4.2.1 Revised calibration and validation sub-basins

One of the major revisions in the new calibration experiment was the selection of the calibration and validation sub-basins. This new procedure led to an increased number of sub-basins that have near-natural flow conditions, and were distributed around each lake and across the entire basin.

In Chapter 3, only sub-basins with "completely" regulated flows were excluded from calibration and validation. This criterion was changed by including only basins with near-natural flow conditions. It was believed that this change would better represent the modelling problem at hand, which did not include accounting for flow regulation or other non-natural basin features. Accordingly, in Canada, only gauging stations were considered that are classified as "natural" in the HYDAT database of the Water Survey Canada. In US, only sub-basins identified as "reference" gauges based on the Geospatial Attributes of Gages for Evaluating Streamflow (GAGES-II, Falcone, 2011) data set of the U.S. Geological Survey (USGS) were considered. The GAGES-II data set defines "reference" watersheds as those in which "hydrologic conditions are least disturbed by human influences".

Then, since one ultimate usage of the resultant model would be in forecasting lake inflows, multiple sub-basins were selected around each of the lakes as well as in the St. Lawrence sub-watershed. Consequently, 25 calibration and 14 validation sub-basins were selected for the new experiment (compared with 10 calibration and 15 validation sub-basins in the baseline calibration). The total drainage area used in calibration and validation was approximately 30,650 km^2 and 11,520 km^2, respectively. The size of the new sub-basins ranged from nearly 230 km^2 to 4220 km^2 for calibration, and from 230 km^2 to 1800 km^2 for validation. These sub-basins are shown in Figure 8.

It is noteworthy that initially, a total of 27 sub-basins were selected for calibration. However, two of them (04124000 and 04136000) located adjacent to each other in Michigan (between Lake Michigan and Lake Huron) showed very poor performance after MESH was calibrated for 1000 model evaluations, degrading the overall performance across the model as well. A number of measures were taken to resolve this issue including conducting longer calibration experiments with 2000 and 5000 model evaluations. None of these measures worked. The poor performance in these sub-basins is perhaps caused by small lakes upstream of the flow gauges that control the flow to some degree. Other possible causes could be errors in the model structure or forcing data used in MESH. These two sub-basins were eliminated from the calibration process in this study.

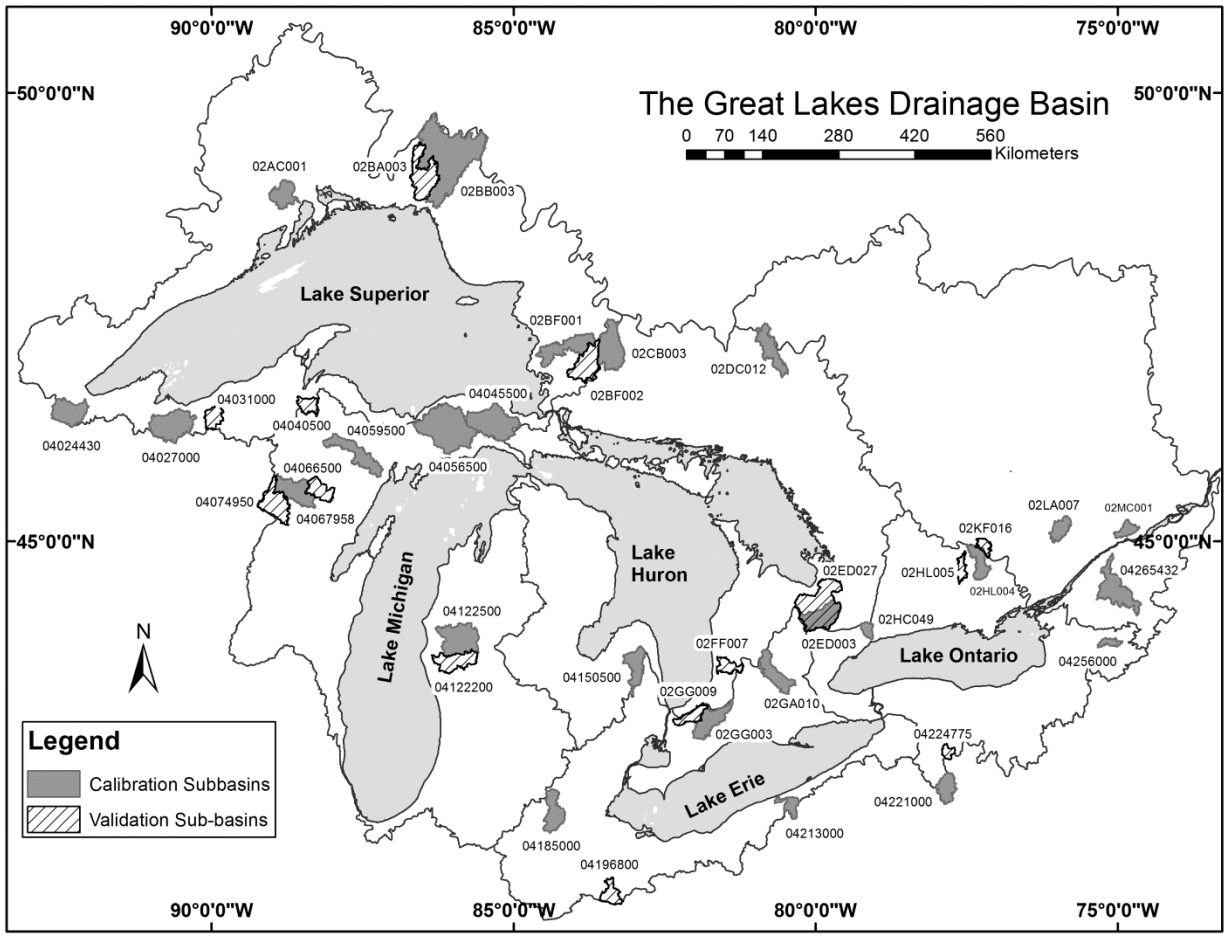

Figure 8. The revised set of calibration and validation sub-basins for the enhanced calibration of MESH over the Great Lakes Basin.

4.2.2 Revised calibration parameters

Another major change compared with the benchmark study in the last chapter was in the use of revised calibration parameters and their ranges. These changes were mainly informed based on local manual sensitivity analysis using a sub-basin scale the case study of Chapter 5, updated information on the MESH wiki page (https://wiki.usask.ca/display/MESH/Welcome+to+the+Standalone+MESH+Wiki!), and further literature search (details are explained below). The local manual sensitivity test involved varying the parameters between their minimum and maximum feasible values at a couple of points, and inspecting the change in simulated hydrographs. Subsequently, the most influential parameters are identified and used in calibration as shown in Table 6. In addition, a post-calibration local sensitivity test was also performed for the Great Lakes Basin case study as explained in section 4.2.6.

Note that due to the nature of the approach adopted by Environment Canada in developing MESH as a community-based model, a wiki page is dedicated to MESH and used by model users as the primary source of information for MESH and is updated from time to time. However, it is recognized that many of the instructions found on the MESH wiki page originates from the CLASS technical documentation (Verseghy, 2011). The refined calibration parameters and their ranges are shown in Table 6 and the modifications are outlined in the five paragraphs below.

As suggested by instructions on the MESH wiki page, not all vegetation parameters are independent of one-another. Therefore, they should not be calibrated altogether to save on computation effort. Accordingly, from the two previous calibration parameters, RSMN (minimum stomatal resistance of vegetation category) and QA50 (coefficient governing the response of stomata to light), QA50 was removed from the calibration, since it was found to have a lesser impact on streamflow simulations. Also, ranges for RSMN were further narrowed down based on the suggested values for each vegetation type on the MESH wiki page. Along the same line, from the two coefficients governing the stomatal resistance to vapour pressure deficit, VPDA and VPDB, VPDA was found to have a larger impact on flow simulations and was added for calibration in the new study. VPDA ranges were assigned according to the work by Dornes et al. (2008). Ranges for the annual maximum rooting depth of vegetation category (ROOT) were also adjusted according to the information available on the Ontario Ministry of Agriculture Food and Rural Affairs (http://www.omafra.gov.on.ca/CropOp/en/general_agronomics/irrigation.html).

The soil drainage index (DRN), previously calibrated in Chapter 3, was excluded from the calibration and fixed at 1. This DRN value enables MESH to simulate the drainage process using soil physical properties, whereas a 0 value would shut off the vertical drainage completely. A value between 0 and 1.0 simulates impeded drainage where the calculated drainage is multiplied by this value. DRN=1 was deemed a better choice to simulate the drainage process based on soil physics. Another parameter that was also eliminated in the new calibration experiment was the average overland slope of a given GRU (XSLP). Instead, the internal slope value for each gird cell was calculated using Digital Elevation Model (DEM) data and used for all GRUs within a given cell. This approach, which is inspired by available measurements, was deemed to be more consistent with observed DEM data compared to the previous strategy in which slope was calibrated for each GRU as an effective parameter.

Multiple changes were also applied to the parameters related to soil properties. The upper range for sand and clay percentages, increased from 50% to 100% to allow more flexibility for calibrating sand and clay percentages. Organic matter was kept constant at 5% and the silt percentage was calculated as the remainder by the model, so they all add to 100%. Accordingly, the MESH input flag corresponding to

these percentages (SOILFLAG) was set equal to 4, which adjusts percentages proportionally, when soil constituent percentages add up to more than 100%. Furthermore, the horizontal surface saturated hydraulic conductivity (WFCI), was not calibrated independently anymore in the new experiment. It was calibrated indirectly through a newly introduced calibration parameter, RATIO, defined as the ratio of horizontal to vertical saturated hydraulic conductivity. The RATIO parameter is not a MESH parameter. Rather, it is a calibration parameter introduced in this study to tie horizontal and vertical saturated hydraulic conductivity together for a given soil type. Vertical saturated hydraulic conductivity is calculated using percent sand in MESH. Then, instead of calibrating the horizontal saturated hydraulic conductivity independently, this RATIO parameter is used to obtain the horizontal saturated hydraulic conductivity from the vertical one. The ranges for this parameter were assigned to the values in Todd (1980). This revised strategy where soil hydraulic conductivities are tied to the soil type results in a more physically meaningful representation of the soil characteristics in the model, compared with the previous strategy where horizontal hydraulic conductivity was calibrated independently.

Another distinct change to the set of calibration parameters is the addition of three parameters corresponding to the infiltration process with snow effect. These parameters include the limiting snow depth below which coverage is less than 100% (ZSNL), the maximum water ponding depth for snow-free areas (ZPLG), and the maximum water ponding depth for snow-covered areas (ZPLS). The ranges for these parameters were assigned based on studies from MacLean (2009) and Dornes et al. (2008). Drainage density (DDEN), soil permeable depth (SDEP), and channel roughness coefficient (WFR2) did not change compared with the baseline study of the previous chapter.

Parameters 1 to 11 in Table 6 are tied to land cover type and, similar to Chapter 3, were calibrated for the five land covers of crop, grass, deciduous forest, coniferous forest and mixed forest. The last parameter in Table 6, channel roughness coefficient (WFR2), was calibrated for all five river classes in MESH. Subsequently, the total number of calibration parameters was increased from 51 in Chapter 3 to 60 in this chapter. Non-calibrated parameters related to soil or land cover types were assigned according to the suggested values in the CLASS manual version 3.5 (Verseghy, 2011).

Table 6- Calibration parameters, their ranges and the calibrated values for the new calibration of the MESH model applied to the Great Lakes Basin.

NO.	Parameter	Description	Range	Calibrated Values for Crop, Grass, Deciduous, Coniferous, Mixed Forest
1	ROOT	Annual maximum rooting depth of vegetation category [m]	(0.2, 1.0) crop and grass (1, 3.5) all forest types	0.374, 0.669, 2.316, 2.730, 2.955
2	RSMN	Minimum stomatal resistance of vegetation category [s m^{-1}]	(60,110) crop (75,125) grass (100,150) all forest types	99.017, 109.498, 106.689, 185.967,148.055
3	VPDA	Vapour pressure deficit coefficient	(0.5,1)	0.754, 0.630, 0.662, 0.944, 0.783
4	SDEP	Soil permeable (Bedrock) depth [m]	(0.35,4.10)	1.504, 1.470, 0.586, 4.095, 0.440
5	DDEN	Drainage density [km/km^2]	(2,100)	79.41, 72.88, 68.15, 78.77, 27.76
6	SAND	Percent sand content [%]	(0,100)	83.6, 6.5, 48.097, 57.8, 20.0
7	CLAY	Percent clay content [%]	(0,100)	12.0, 2.6, 48.201, 0.2, 75.8
8	RATIO	The ratio of horizontal to vertical saturated hydraulic conductivity	(2,100)	34.836, 79.518, 49.896, 97.051,98.735
9	ZSNL	Limiting snow depth below which coverage is less than 100% [m]	(0.05,1)	0.703, 0.701, 0.967, 0.731, 0.520
10	ZPLS	maximum water ponding depth for snow-covered areas [m]	(0.02,0.15)	0.079, 0.142, 0.136, 0.073, 0.024
11	ZPLG	maximum water ponding depth for snow-free areas [m]	(0.02,0.15)	0.122, 0.129, 0.126, 0.040, 0.101
12	WFR2	Channel roughness factor	(0.02,2)	1.862, 1.696, 0.127, 0.629, 0.360

4.2.3 Implementation of time-saving techniques

In order to reduce the computational demand of the calibration process, two methods were used in the new calibration experiments: model preemption (Razavi et al., 2010) and GRU polishing. GRU polishing involves removing GRU types from each grid cell with area coverage fractions below a certain threshold. The removed GRU type areas are redistributed proportionally on the remaining GRU types in the grid cell. The idea is that GRUs with areas below a certain threshold will have very small impact on streamflow simulations. This strategy was applied with 5% and 10% thresholds. After assessing the differences in hydrographs induced by these thresholds, the 5% threshold was deemed appropriate. It resulted in a reduction of about 10% in a single simulation time, while the associated changes in hydrographs remained very small (the average deviation from the highest peak in calibration sub-basins was less than 1 m^3/s). This saving was equivalent to 2 minutes per each single model simulation, or more than 33 hours for a calibration with 1000 model runs, using an Intel Core i7-920 2.67 GHz workstation

running a 64-bit version of the Windows 7 operating system with 12 gigabytes of installed memory and a quad-core processor. Consequently, a polished version of the 7 land class-based GRU types were used for all new calibration experiments in this chapter.

As in Chapter 3, a weighted sum of the Nash-Sutcliffe (NS) coefficient of efficiency (Nash and Sutcliffe, 1970) for daily flows was used, according to Equation (3), as the calibration objective and the main criterion to evaluate and compare model performance. The weights were assigned proportional to the long-term average of daily flows at each sub-basin, in a way that they add to one.

$$NS_{weighted} = \sum W_i \cdot NS_i = \sum W_i \cdot \left(1 - \frac{SSE_i}{n_i \cdot Var_i}\right) = 1 - \sum \frac{W_i \cdot SSE_i}{n_i \cdot Var_i} \qquad 1$$

Here, i denotes the sub-basin index, W_i is the weight for sub-basin i, SSE_i is the sum of squared errors between simulated and observed flows over the calibration period for sub-basin i, Var_i is the variance of observations sub-basin i, and n_i is the number of observations over the calibration period for sub-basin i.

As another strategy to save on calibration time, it was desired to use the deterministic model preemption (Razavi et al., 2010). According to Razavi et al. (2010), deterministic model preemption involves dismissing model runs that will definitely result in a worse model performance (an objective function value) compared to the current best calibration solution. They reported a saving of nearly 50%, when model preemption was applied for calibration of MESH to a watershed of approximately 55 km^2 using a sum of square error calibration objective. The pre-emption concept requires using a monotonically increasing objective function, such as Sum of Squared Errors (SSE), during calibration (Razavi et al., 2010). Accordingly, a transformed objective function (Z) that is monotonic with increasing simulation timestep and rank equivalent to Equation (1) was defined for model the calibration process as shown in Equation 4 below.

$$Minimize\ Z = \sum \frac{W_i \,.\, SSE_i}{n_i\, Var_i} \qquad 2$$

Model preemption requires that Z is computed after each day of the calibration period is simulated in MESH with the Var_i values computed over the whole calibration period (i.e., constants) and the SSE_i values computed only over the simulated days so far. In other words, model preemption requires that SSE_i values are computed as a function of simulation time step. The model is then preempted only if Z is

computed to be larger than the current best objective function value found by DDS. Due to the nature of the DDS optimization algorithm used in this study, this transformation would not change the calibration results at all (compared to maximizing Equation 1). DDS is a rank-based optimization algorithm in that only the relative rank of the candidate solutions influences the search result (as opposed to the absolute magnitude of objective function differences between candidate solutions), and the exact same answer would be obtained with or without this transformation.

The model preemption implementation in MESH was not coded to track time steps simulated before preemption and hence, the preemption savings could not be estimated using the deterministic preemption introduced by Razavi et al. (2010). Furthermore, actual runtime differences of calibration experiments with and without preemption were not available because only experiments with preemption were conducted.

4.2.4 Other changes

The two other changes made in the new calibration of MESH over the Great Lakes Basin included extending the warm-up (model initialization) period and adjusting the simulated drainage areas of the flow gauges were. In order to attenuate, more than before, the impact of the presumed initial conditions, the model initialization period was increased from 12 months to 16 months. Moreover, due to the resolution of the grid cells used to discretize the watershed (about 15 km), the modeled drainage areas in MESH will be sometimes different than the reported areas by EC. In order to circumvent this deficiency, the gauge coordinates in MESH were adjusted as suggested by Pietroniro at al. (2007). This strategy still would not necessarily result in a match between real and modeled areas, so, MESH simulated flows at each gauge were additionally multiplied by a scale factor when calculating the objective function. This scale factor is simply the ratio of the EC drainage area to the modeled drainage area and ranged from 0.85 to 1.33 for the 25 calibration sub-basins, and 0.76 to 1.46 for the 14 validation sub-basins used in the new calibration experiment.

Finally, a couple of changes were also made to the model version and the input forcing data. Firstly, a new set of input forcing data was used as they became available through Environment Canada in late 2013. This new forcing data corrected an error in the forcing data used in Chapter 3. Secondly, in accordance with the ongoing development of the MESH modelling system, MESH version 1.3.006 replaced the older version 1.3.003 in the new study. The major difference between the two versions was the capability of reading in the gridded slope values in the new version. As explained in section 0, MESH version 1.3.006 also was modified in this work to facilitate the implementation of model preemption during the calibration process.

4.2.5 Uncertainty Assessment for Model Predictions

Uncertainties in model predictions stem either from model parameters, model structure, or data (Tolson and Shoemaker, 2008). In this work, only the uncertainty associated with estimating model parameters is considered and it is assessed using an efficient uncertainty-based calibration approach called Dynamically Dimensioned Search-Approximation of Uncertainty (DDS-AU, Tolson and Shoemaker, 2008). DDS-AU was shown to require orders of magnitude less computational time for quantifying prediction uncertainty, compared with the widely-used Generalized Likelihood Uncertainty Estimation (GLUE, Beven and Binley, 1992) technique (Tolson and Shoemaker, 2008). Essentially, DDS-AU involves performing several independent DDS optimization trials of relatively small number of model evaluations. Each trial starts from a different random solution and thus, ends at a different final parameter set. Consequently, DDS-AU can find a number of "behavioral" solutions (parameter sets) used in estimation of the uncertainty bounds in predictions. Behavioral solutions are identified based on a subjective performance threshold. In the simplest application of DDS-AU, only each final best parameter set is assessed as being a possible behavioral solution. In this study, twenty calibration trials were used each with 200 model evaluations. The parameter sets for which all NS values across all 25 calibration sub-basins (excluding the worst performing sub-basin) were positive were considered as the behavioral parameter sets.

4.2.6 Sensitivity Analysis for the Great Lakes Basin

In order to find the sensitivity of the MESH streamflow simulations to parameters after calibration, a manual local sensitivity test is performed from the best parameter values found in the new calibration experiment. In this test, the sensitivity of simulated maximum peak flow and total volume to MESH parameters is examined over the calibration period in the 25 calibration sub-basins. All the parameters corresponding to the most common GRU (land) type in the region, crop, are considered in the analysis plus the five channel roughness parameters. Parameters are perturbed, one at a time, below and above their value by 20% of their range. If the upper or lower perturbed value for a given parameter was outside its feasible range (e.g., 0% or 100% for percent sand), then the maximum or minimum feasible range value is used instead. The percent change for the maximum peak flow and the total volume associated with each parameter perturbation is then calculated for each calibration sub-basin. The average, minimum, and maximum percent changes across all sub-basins are also calculated for each parameter perturbation. The lower and upper values for the saturated surface soil conductivity (WFCI) are indirectly calculated by perturbing the RATIO parameter. The results from this analysis can provide useful information for MESH modelers about the relative importance of each parameter, and the selection of calibration parameters.

4.3 Results and Discussion

Firstly, the results of the new (enhanced) calibration strategy are presented and compared with the benchmark calibration results of Chapter 3. Then, results of the uncertainty analysis are presented followed by the sensitivity analysis results.

4.3.1 Enhanced Calibration Results

In the new calibration experiment, similar to the calibration effort in the baseline calibration, three calibration experiments were performed with 1000 model evaluations each. Each experiment was started from a different initial point (parameter set). The parameter set associated with the best calibration performance was then selected as the "best" parameter set and used to generate model predictions for the temporal and spatial model validation. As in Chapter 3, the model was calibrated from October 2004 to September 2005 and validated until September 2009. Likewise, the final calibrated parameter set from Chapter 3 (referred to as "benchmark" parameter set hereafter) was also applied to the new calibration and validation sub-basins (under the revised MESH model setup described in this chapter) to generate comparative temporal and spatial model validation results. This experiment and its associated results will be referred to as "benchmark" results or experiment hereafter. The validation model performance under each parameter set was evaluated and contrasted for these two experiments using the four metrics defined in Table 2 of Chapter 3 namely, the Nash-Sutcliffe Coefficient of Efficiency (NS), Volumetric Efficiency (VE), Root Mean Squared Error (RMSE), and Percent Bias (PBIAS). In this section, the results of the new model calibration are first presented for the calibration sub-basins over the calibration period. Then, the validation results are shown for both the new and the benchmark experiments.

Figure 9 indicates the individual NS values over the calibration period (October 04 to September 05) for the 25 sub-basins used in calibration. Sub-basins are shown in the order of their locations from the west around Lake Superior to the east around Lake Ontario and in St. Lawrence River basin. The weighted sum of NS values is also shown as the overall calibration performance. The overall NS value is above 0.4 and the individual NS values range from -0.61 to 0.85 with a median of 0.53. Overall, model calibration performance is reasonably high for all sub-basins, except sub-basin 04122500. A discussion on this poor performance is provided at the end of this section.

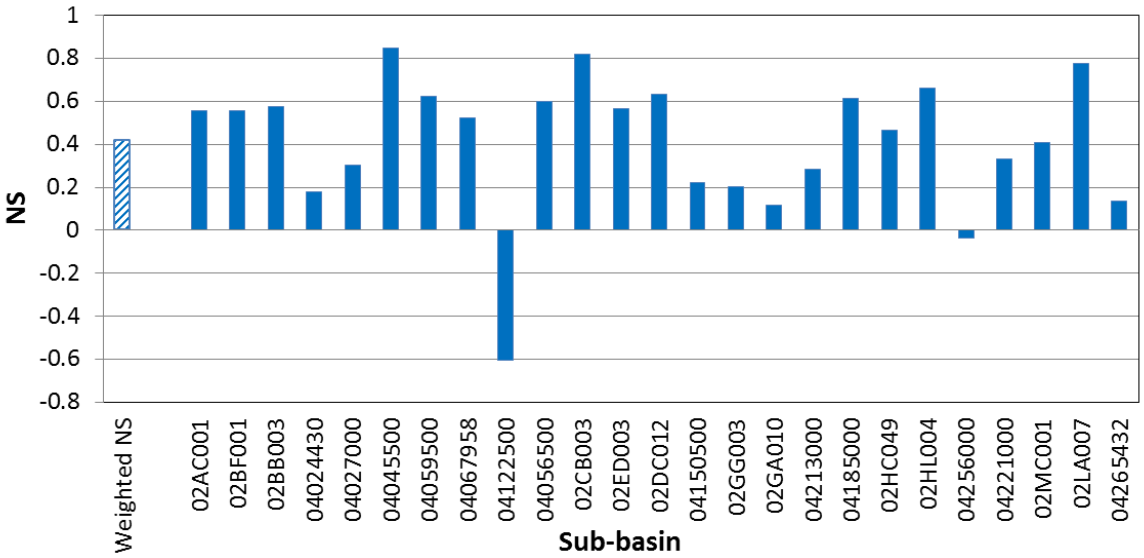

Figure 9. The NS values over the calibration period (October 04 to September 05) for the 25 calibration sub-basins and the corresponding weighted NS.

In order to conduct a robust comparison between the new and the benchmark experiment, multiple interpretations of the validation results are presented here. This includes comparing the validation model performance overall and for each sub-basin individually. Table 7 summarizes the model performance based on the four metrics, defined in Table 2 of Chapter 3, during the temporal validation period (October 2005 to September 2009) for the 25 calibration sub-basins. As seen in this table, overall, the new experiment indicates an improved performance for temporal validation compared with the benchmark results. The median, minimum and maximum NS and PBIAS values are all better for the new results, except for the minimum PBIAS. VE and RSME values are very close for the two experiments.

Similar results are presented in Table 8 for the spatial validation of the model in the 14 validation sub-basins. Again, the results confirm that overall, model performance is improved compared with the benchmark results. The median performance is higher for all metrics in the new experiment. The minimum and maximum values of the metrics also show a better or comparable performance for the new experiment, except for the minimum NS and maximum PBIAS. Again, VE and RSME values are very close for both experiments.

Table 7. The median, minimum and maximum of the four metrics (defined in Table 2), for the new and benchmark experiments, over the temporal validation period (October 2005 to September 2009), for the 25 calibration sub-basins.

	Benchmark Experiment				**New Experiment**			
	NS	**VE**	**RMSE (m^3/s)**	**PBIAS (%)**	**NS**	**VE**	**RMSE (m^3/s)**	**PBIAS (%)**
Median	0.19	0.40	14	35	0.30	0.45	15	26
Min.	-0.90	0.19	4	5	-0.65	0.17	5	-14
Max.	0.69	0.64	50	50	0.79	0.66	54	39

Table 8. The median, minimum and maximum of the four metrics (defined in Table 2), for the new and benchmark experiments, over the spatial validation period (October 2004 to September 2009), for the 14 validation sub-basins.

	Benchmark Experiment				**New Experiment**			
	NS	**VE**	**RMSE (m^3/s)**	**PBIAS (%)**	**NS**	**VE**	**RMSE (m^3/s)**	**PBIAS (%)**
Median	0.12	0.37	14	34	0.29	0.47	9	21
Min.	-0.75	0.06	3	13	-1.09	0.05	3	6
Max.	0.51	0.57	19	44	0.66	0.56	19	49

Next, the two experiments were compared for each sub-basin individually, based on a scoring framework introduced here. RMSE was removed from this analysis to avoid double counting information, because RSME is rank equivalent to NS. In each sub-basin, the three metrics were contrasted one by one between the two experiments. For a given metric, if the new experiment was dominant, a score of +1 was assigned, and if the benchmark experiment was dominant, a score of -1 was assigned to the model performance. If the poorer value was within the 10% of the better value for a certain metric, then the model performance was considered equal for both experiments and a value of 0 was assigned to it. Subsequently, the sum of scores across all three metrics determined the degree by which the model performance was upgraded or degraded in each sub-basin. The model performance for the new experiment was classified between “much better” to “much worst” with respect to the benchmark results as seen in Table 9. A “much better” score (+3) shows that the new experiment performed better based on all the three metrics in a given sub-basin, whereas a “much worse” score (-3) indicates that the benchmark results were dominant for all the three metrics.

Table 9. Classification of the model performance for the new results with respect to the benchmark results.

Score	Class
3	Much Better
2	Better
1	Slightly Better
0	Equal
-1	Slightly Worse
-2	Worse
-3	Much Worse

Histograms corresponding to the application of this classification scheme are presented in Figure 10a and 10b. Figure 10a shows the result of the temporal model validation for the 25 calibration sub-basins. As seen in this figure, model performance is better for the new experiment in more than 50% of the sub-basins. In three sub-basins, model performs much better compared to the benchmark results. The new results are worse only in about 30% of the sub-basins. If we further consider the "slightly better", "equal", and "slightly worse" classes as equal performance, then in only 2 (8%) calibration sub-basins the benchmark results are better. Figure 10b, depicts similar results for the 14 spatial validation sub-basins. The performance improvement of the new results is even stronger during the spatial validation of the model where more than 70% of the sub-basins have a better performance in the new experiment. Further grouping of the three "slightly better", "equal", and "slightly worse" classes into the "equal" class, leaves no validation sub-basin for which the benchmark results are better.

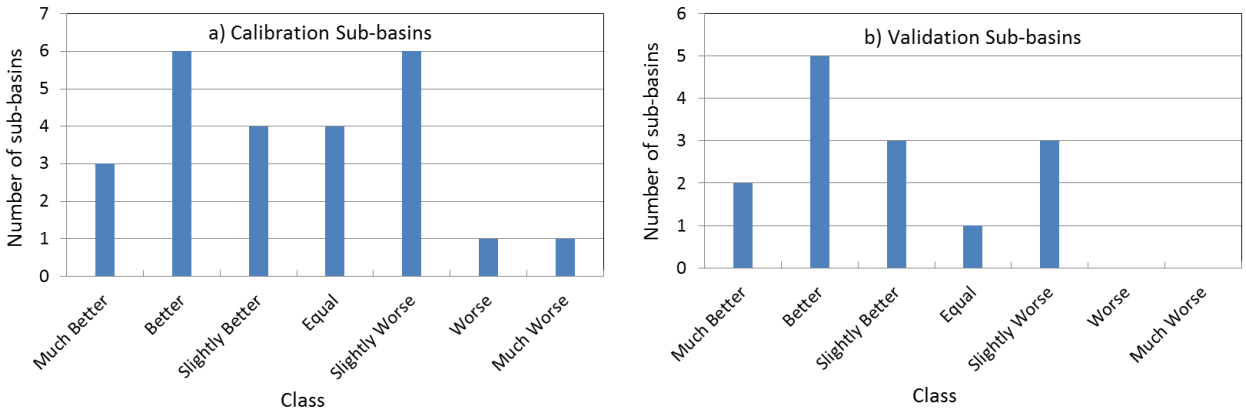

Figure 10. Histograms corresponding to the classification of the model performance in the new experiment with respect to the benchmark performance in a) the calibration sub-basins (temporal validation), and b) the validation sub-basins (spatial validation).

In summary, various interpretations of the results confirm that the MESH model performance indeed improved with the revised calibration strategy compared to the benchmark results. Individual validation NS values from both calibration experiments are shown in Figure 11 for both calibration and validation sub-basins. As seen in Figure 11a, the calibration sub-basin 04045500 has the highest validation NS value in both experiments. In spatial validation, sub-basin 02ED027 has the best NS value in both the new and benchmark experiments.

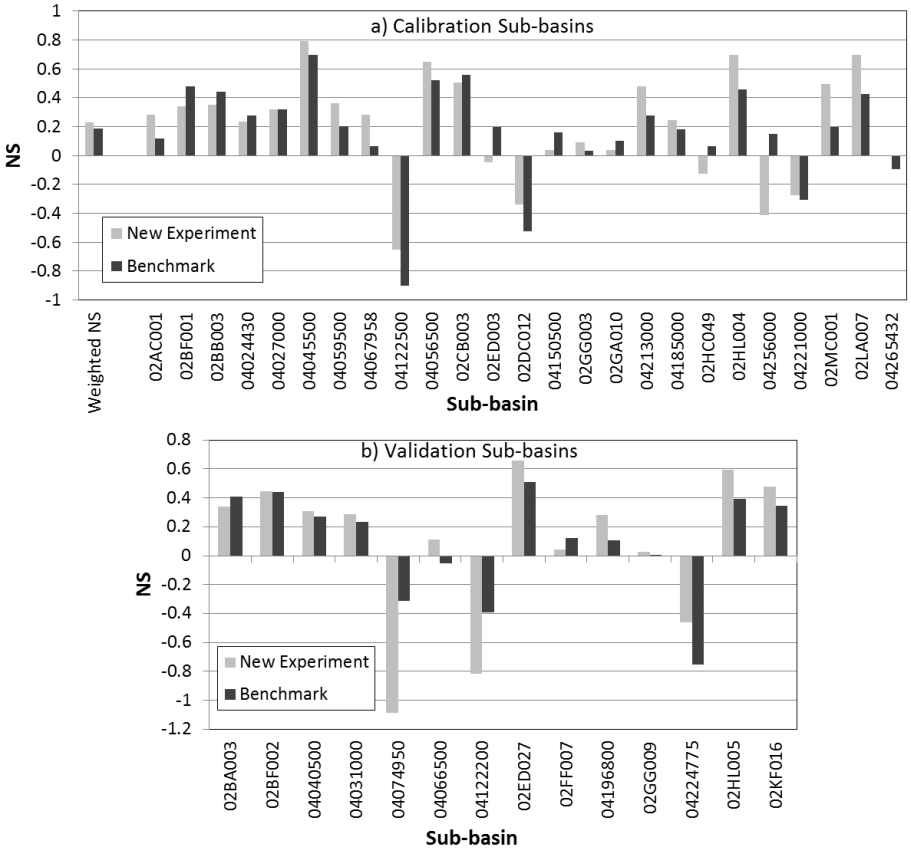

Figure 11. The NS values from the new and the benchmark calibration experiments for a) the 25 calibration sub-basins plus the corresponding weighted NS, over the temporal validation period (October 05 to September 09), and b) for the 14 validation sub-basins, over the spatial validation period (October 04 to September 09).

Figure 12 displays simulated hydrographs at the best (highest NS) validation sub-basin for both experiments. As seen in this figure, the simulated hydrograph in the new experiment clearly better matches the observed peak flows compared with the benchmark hydrograph.

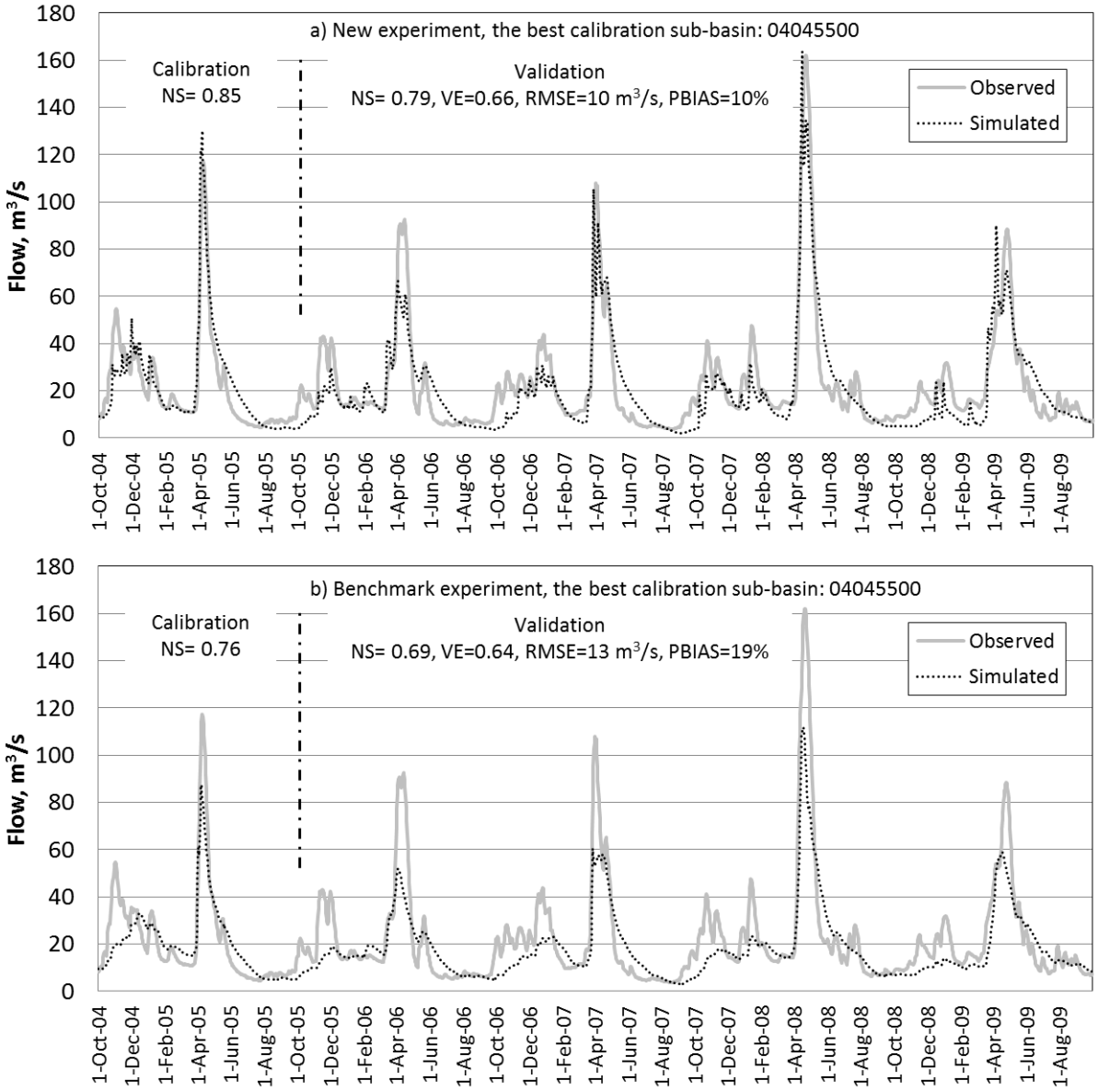

Figure 12. Simulated hydrographs at the best (highest NS) calibration sub-basin (04045500) for the new and the benchmark experiments.

Figure 13 displays simulated hydrographs at the best (highest NS) validation sub-basin for both experiments. As seen in this figure, the simulated hydrograph in the new experiment is much better in matching the higher observed peak flows compared with the benchmark hydrograph. However, the benchmark hydrograph is more successful in simulating the base flows.

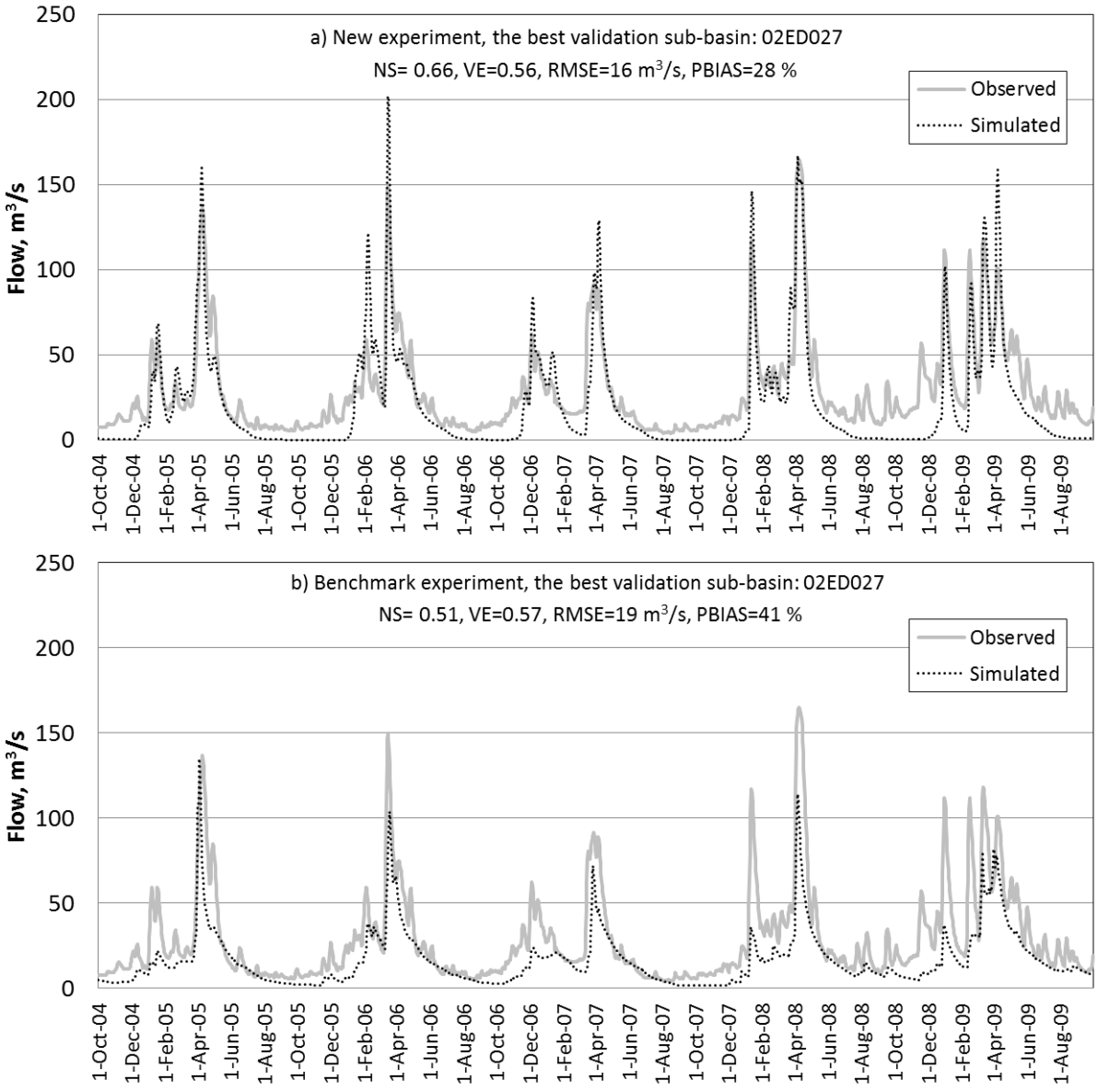

Figure 13. Simulated hydrographs at the best (highest NS) validation sub-basin (02ED027) for the new and the benchmark experiments.

4.3.2 Various attributes of the MESH performance in the new calibration experiment

Further investigations were conducted to explore the relationship between validation MESH performance in the new calibration experiment and various sub-basin characteristics such as land cover type, size, and location. The validation period model performance was considered as the NS values over the temporal validation period (October 05 to September 09) for calibration sub-basins, and as the NS values over the spatial validation period (October 04 to September 09) for the validation sub-basins. The corresponding results are discussed below.

The correlation between the validation period model performance and the GRU (land) type was examined in the new calibration experiment. No correlation was found between the NS values and a particular GRU (land) type. It is worth noting that either crop, or coniferous forest, or deciduous forest had the dominant coverage areas across all calibration and validation sub-basins. Figure 14 shows the range of the NS values in the 14 validation sub-basins with the crop, deciduous forest, or coniferous forest land cover fraction above 0.3.

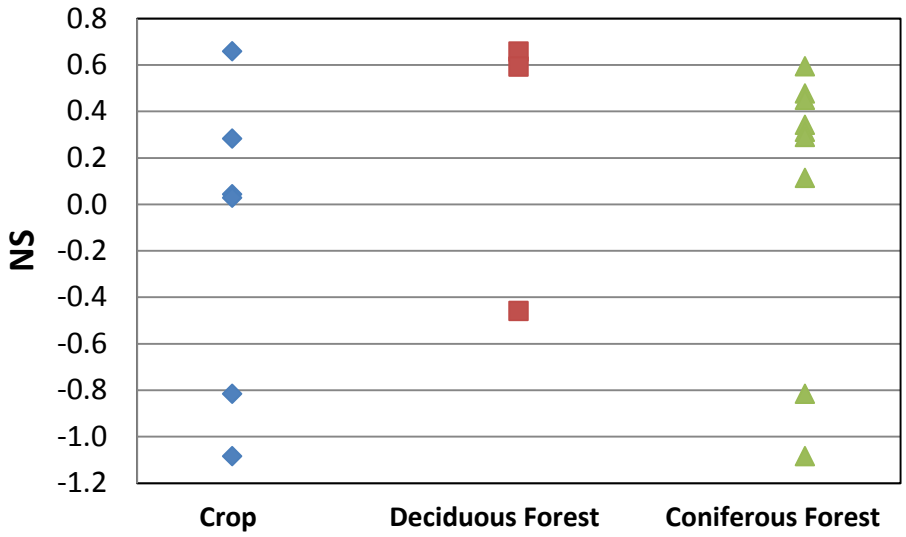

Figure 14. Range of the NS values in the 14 validation sub-basins with the crop, deciduous forest, or coniferous forest land cover fraction above 0.3.

Furthermore, the relationship between the model performance (based on NS values) and the size of the sub-basins in the new experiment was investigated. As seen from the results (Figure 15), there is no

correlation between the size of the sub-basins and the validation model performance. Same result (not shown) was observed for the calibration model performance (i.e. NS value over the calibration period).

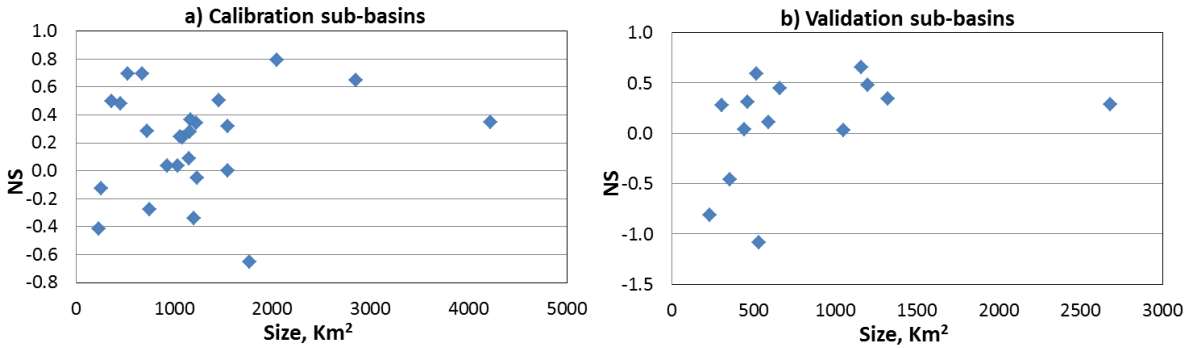

Figure 15. Scatter plot of the validation NS values for a) 25 calibration sub-basins, and b) 14 validation sub-basins.

In order to examine the relationship between the location of sub-basins and model performance, all the sub-basins along with their grouped validation NS values are plotted on a map shown in Figure 16. NS values are classified into four groups. Green colour displays the poor performing sub-basins with negative NS values, and the red colour depicts the best performing sub-basins with NS greater than 0.5. As seen, sub-basins around Lake Superior and to the north of Lake Ontario indicate a higher performance in general. The regions to the north of Lake Erie and in particular, in between Lake Michigan and Lake Huron have poor performance. As an example, the simulated hydrograph associated with sub-basin 04122500 (the poorest performing sub-basin during model calibration) is shown in Figure 17. As seen in this figure, the timing of the flow values are simulated relatively well by MESH. However, the main problem is that the baseflow (mainly coming from the interflow in MESH) is substantially underestimated by MESH. Even calibrating MESH individually to only this sub-basin did not resolve this issue. Similar behavior was also observed at this sub-basin when the benchmark parameter set was used for simulation. Therefore, it was concluded that the issue was not caused by the calibration strategy. Another adjacent validation sub-basin (04122200) in that area to the east of Lake Michigan also showed similar performance and behavior. This is interesting because the two sub-basins (04124000 and 0413600 shown in black in Figure 17) that were previously eliminated from the calibration due to their poor performance (as explained in section 4.2.1) are also from the same region and show similar behavior. Further investigation using the surficial geology map of the Great Lakes Basin (map not shown) did not indicate any correlation between model performance and soil types, particularly in this region.

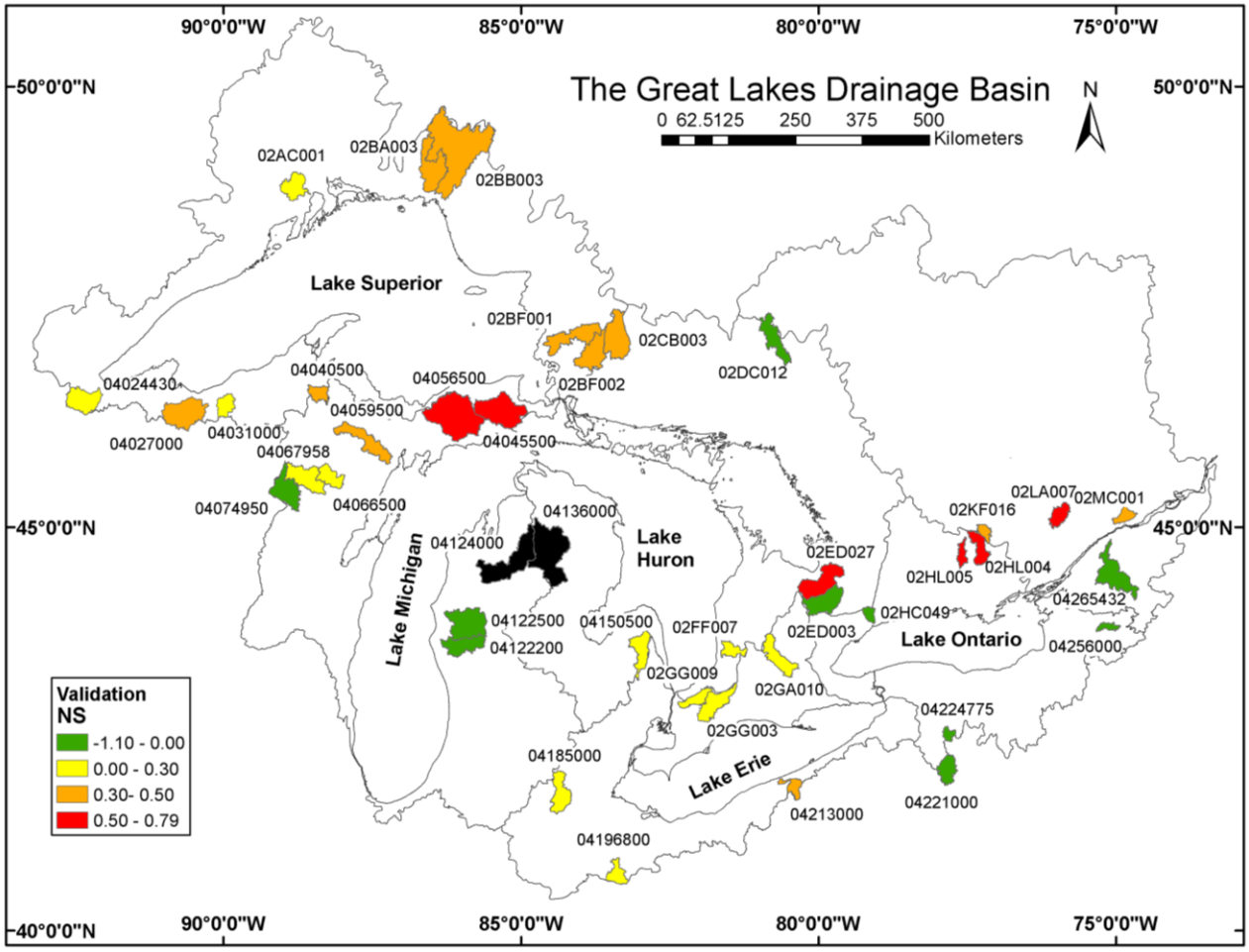

Figure 16. The Great Lakes Basin map showing the validation NS value of all the calibration and validation sub-basins using a colour spectrum. Red colours indicate high performance whereas green colours indicate low performance. Sub-basins 04124000 and 04136000 (shown in black) are the two sub-basins removed from calibration due to very poor performance.

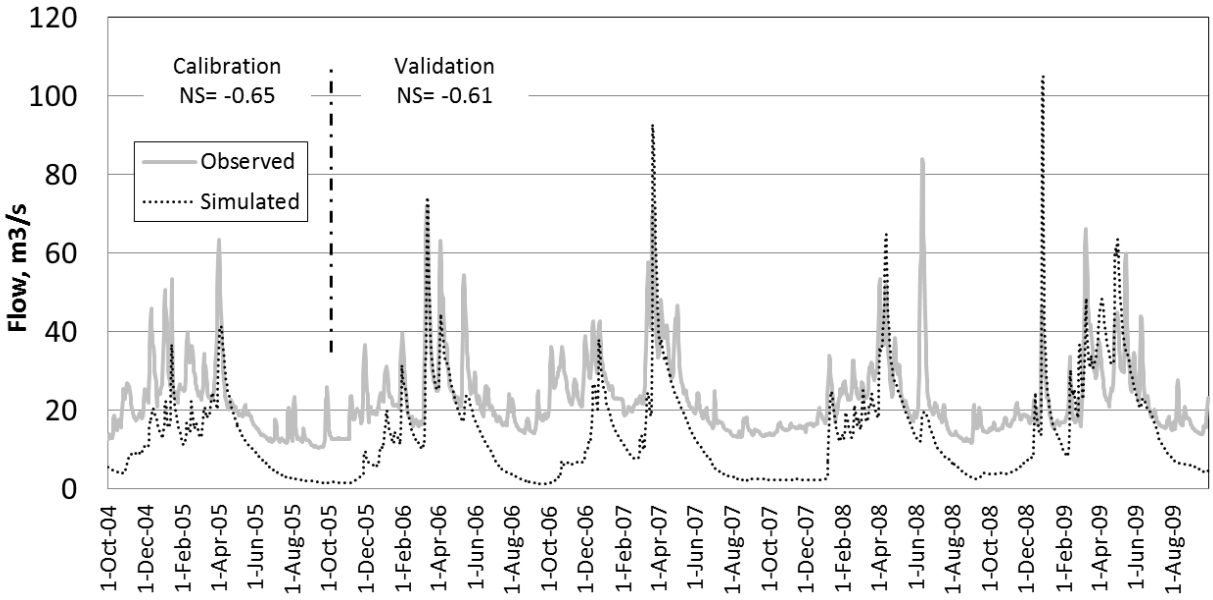

Figure 17. Observed and simulated hydrographs at sub-basin 04122500 with the worst NS values in calibration and temporal validation.

To further investigate this issue, a bedrock geology map of the Great Lakes Basin was examined as shown in Figure 18. The approximate location of the Niagara Escarpment is also shown on this figure. Interestingly, comparing Figure 16 and Figure 18, reveals that almost all the sub-basins with a higher performance (NS>0.3, shown in orange and red in Figure 16) fall to the north the Niagara Escarpment. On the other hand, the region between Lake Michigan and Lake Huron with poor performing sub-basins has a unique geology consisting of poorly permeable rocks, shale and sandstone aquifers. This indicates that the land-cover based GRU types used in this study are perhaps not sufficient to model the baseflow and interflow for this region.

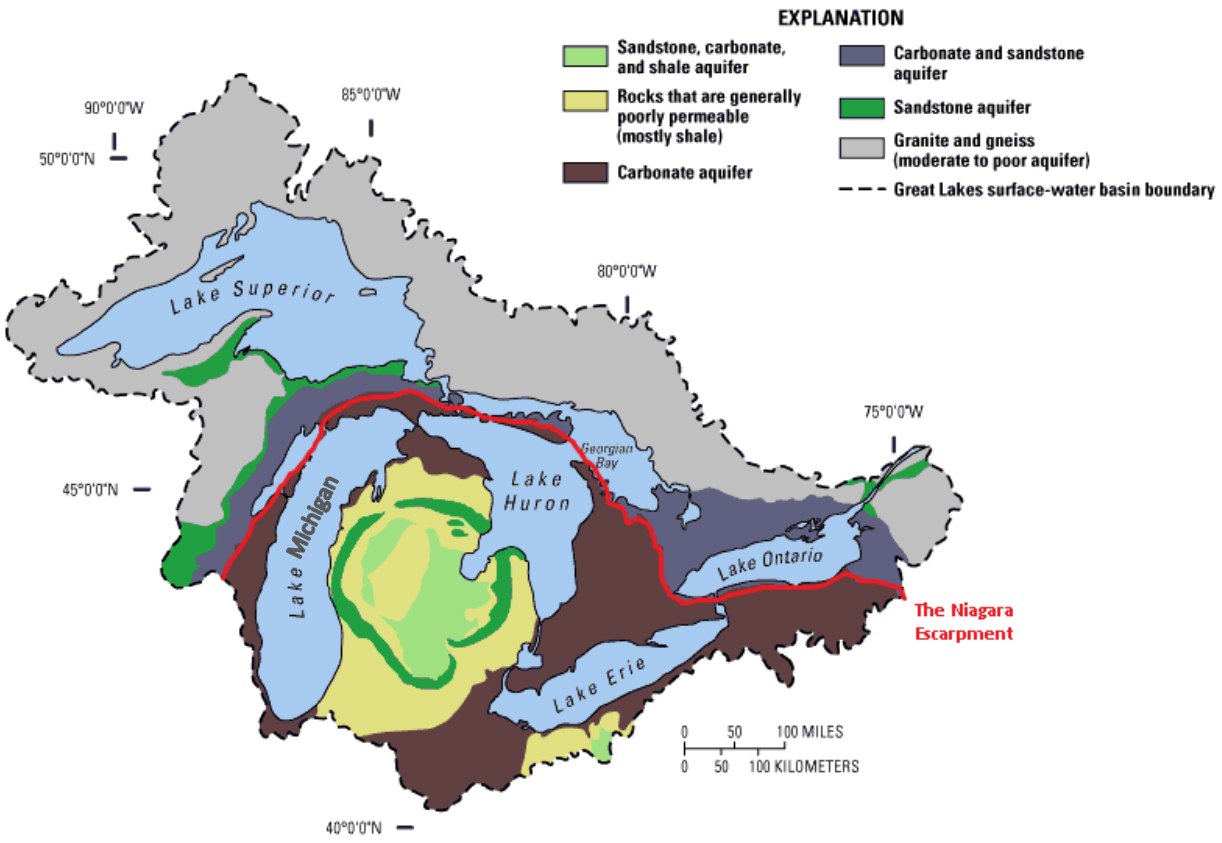

Figure 18. Bedrock aquifer of the Great Lakes Basin the Great Lakes Basin and the approximate position of the Niagara Escarpment (Modified from Sheets et al., 2005).

4.3.3 MESH performance for non-natural flow conditions

Since a considerable portion of the Great Lakes Basin consists of sub-basins that have flow conditions that are not near-natural, the MESH model performance is also evaluated for a sample of ten such sub-basins. These sub-basins are classified as "non-reference" by the USGS GAGES-II data set, or as "regulated" by the Water Survey Canada. A variety of factors cause the flow condition at these sub-basins to be non-natural such as being downstream of cities or the presence of upstream powerplants, dams, or lakes. Table 10 lists the selected sub-basins along with the corresponding remarks on flow conditions. Table 11 shows the median, minimum and maximum of the four metrics for these ten sub-basins over the entire simulation period (October 2005 to September 2009). As expected, the model performance is poor in these sub-basins. The maximum NS value is 0.14 and the rest are negative. Figure 19 depicts the individual NS values for these ten sub-basins. It is acknowledged that MESH performance could be

assessed similarly in many other non-natural sub-basins, but it is assumed that the results would not be that different.

Table 10. List of the sub-basins with non-natural flow conditions assessed for MESH performance.

Sub-basin ID	Name (Area, km^2)	Remarks (from USGS or Water Survey Canada)
04193500	Maumee River at Waterville OH (16409.4)	Many upstream cities
04157000	SAGINAW RIVER AT SAGINAW, MI (14327.2)	Minimum flows affected by wind direction and seiche on Saginaw Bay, 20.3 mi downstream. Considerable diversion through metropolitan area of Saginaw.
04067500	MENOMINEE RIVER NEAR MC ALLISTER, WI (10155.0)	Flow regulated by powerplants and by Michigamme Reservoir, and Peavy Pond on Michigamme River, and by many smaller reservoirs above station.
04249000	OSWEGO RIVER AT LOCK 7, OSWEGO NY (13209.4)	Prior to 1933 and subsequent to 1972, flow in Oswego (Barge) Canal not included. A large amount of natural storage and some artificial regulation is afforded by the many large lakes and the Erie (Barge) and Oswego (Barge) Canal systems in the river basin.
04260500	BLACK RIVER AT WATERTOWN NY (4842.9)	Flow regulated by Stillwater Reservoir (see station 04256500), Fulton Chain of Lakes, and other reservoirs. Extensive diurnal fluctuation at low and medium flow caused by mills and powerplants in and above Watertown. During canal season, water is diverted
04237500	SENECA RIVER AT BALDWINSVILLE NY (8110.1)	Since November 1949, a large amount of natural storage and some artificial regulation is afforded by many large lakes and the Erie (Barge) Canal system in the river basin. Large diurnal fluctuations at low and medium flows caused by powerplants upstream
04119000	GRAND RIVER AT GRAND RAPIDS, MI (12741.2)	Moderate diurnal fluctuation at low and medium flow caused by powerplants
04121970	MUSKEGON RIVER NEAR CROTON, MI (6050.8)	Flow completely regulated by Croton Dam 1,000 ft upstream.
02GB001	GRAND RIVER AT BRANTFORD (5200.5)	Regulated
02HB025	CREDIT RIVER AT NORVAL (644.84)	Regulated

Table 11. The median, minimum and maximum of the four metrics (defined in Table 2), for ten sub-basins with non-near-natural flow conditions over the spatial validation period (October 2004 to September 2009).

	NS	VE	RMSE (m^3/s)	PBIAS (%)
Median	-0.55	0.34	127	38
Min.	-1.75	0.21	9	-5
Max.	0.14	0.45	294	48

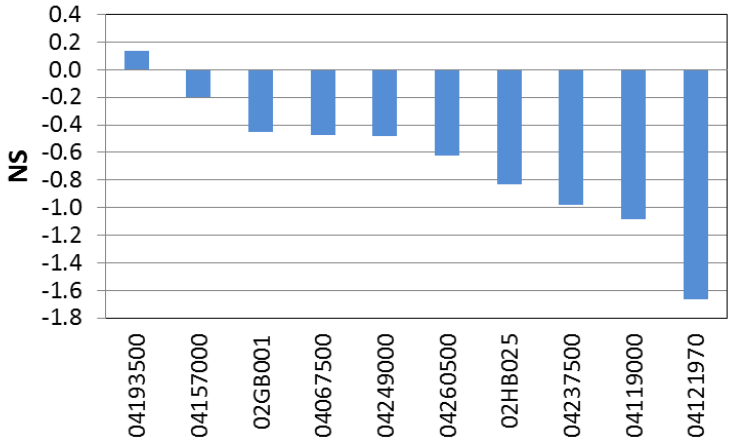

Figure 19. The NS values over the spatial validation period (October 04 to September 09) from the new calibration experiment for the 10 sub-basins with non- natural flow conditions.

4.3.4 Uncertainty Analysis Results

In order to find the prediction uncertainty band for the MESH predations in the Great Lakes Basin, 20 DDS calibration trials with 200 model evaluations were performed with different initial solutions (i.e., random seeds). Next, the parameter sets where all NS values were positive across all calibration sub-basins were considered "behavioral" and used in generating the uncertainty bands for flow predictions. In doing so, the 04122500 sub-basin was not considered because it constantly had a very poor performance as was discussed earlier under the calibration results in this chapter. Consequently, eight parameter sets were identified as the behavioral sets. The median, minimum, and maximum NS values of the 25 calibration sub-basins are calculated for each of these parameter sets and the associated ranges across all eight behavioral sets are shown in Table 12.

Table 12. Range of the median, minimum, and maximum for the 25 sub-basin NS values across the 8 behavioral parameter sets.

	Min	**Max**
Median	0.41	0.51
Min	0.00	0.11
Max	0.86	0.91

Flows were then simulated by MESH using these eight sets for all calibration and validation sub-basins over the entire period (October 04 to September 09). The prediction uncertainty band in a given day was calculated as the difference between the maximum and the minimum of the simulated flow values of all the 8 sets on that day. Figure 20 shows the uncertainty bands for the best (i.e. maximum NS value) calibration (Figure 20a) and validation (Figure 20b) sub-basins. The uncertainty band is fairly narrow for the calibration sub-basin but wider in the validation sub-basin. The observed hydrograph frequently falls out of the bands in those cases.

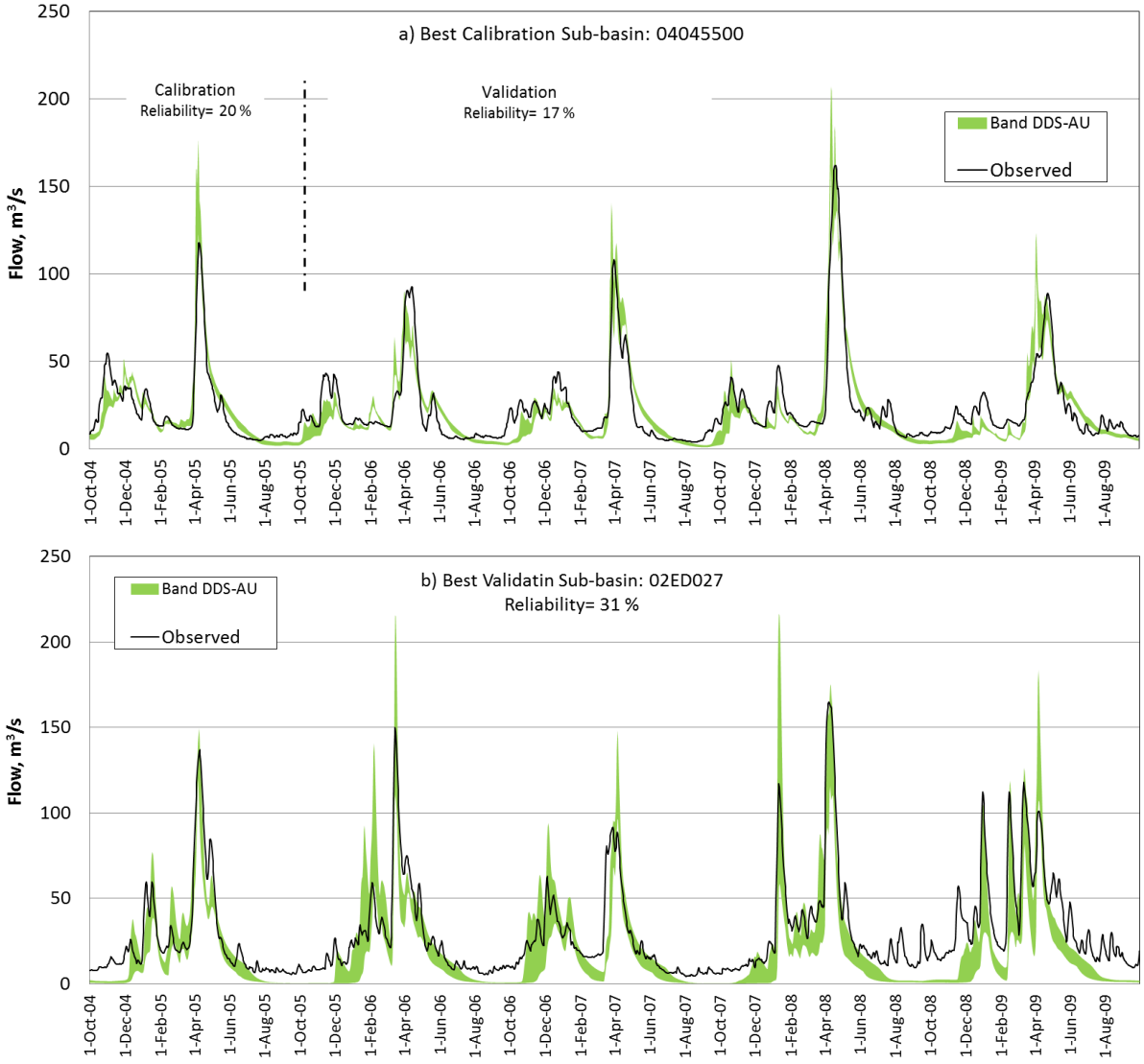

Figure 20. Prediction uncertainty bands for the calibration and validation sub-basins with the best performance (largest NS value) during model validation. The hydrograph corresponding to the best parameter set found by the new calibration strategy is also displayed for comparison.

The prediction uncertainty bands were quantitatively evaluated using the "reliability" metric defined by Yadav et al. (2007). The reliability metric measures the percentage of the observed flows that are within the uncertainty band over the prediction period. For the 25 calibration sub-basins, the reliability metric was calculated over the temporal validation period (October 05 to September 09) and ranged from 12 % to 65 % with a median of 25%. For the 14 validation sub-basins during the spatial validation period (October 04 to September 09) reliability values ranged from 11% to 38% with a median of 19%.

These reliability values are low. This indicates that the uncertainty analysis was not sufficient. The number of short DDS runs was not large enough for sufficiently exploring the entire parameter space and thus, finding more behavioral sets. The reliability can increase when more behavioral sets are used. Ideally, one would like to obtain a narrow band with high reliability values covering most of the measured flows.

4.3.5 Sensitivity Analysis Results

Table 13 shows the local sensitivity results for the MESH streamflow simulations to the parameters of the most common GRU type (Crop) in the Great Lakes Basin, as explained in section 4.2.6. Parameters are ranked based on the magnitude of the percent change in the peak or volume of the 25 calibration sub-basins. Soil permeable depth or depth to bedrock (SDEP) is by far the most influential parameter, followed by the saturated surface soil conductivity (WFCI) and percent sand. All of these parameters are associated with a change larger than 60% in the maximum peak flow. Four river channel roughness values and percent clay are at the next level causing a change more than 30% in the peak or volume. The annual maximum rooting depth (ROOT), the channel roughness factor of the third river class (WFR3), and the limiting snow depth below which coverage is less than 100% (ZSNL) produce changes between 6% and 24%. The changes associated with the rest of the parameters are all relatively small and below 5%.

Moreover, it is observed that the results of this post-calibration sensitivity test at the Great Lakes Basin scale matches fairly well with the results of the local sensitivity test at the sub-basin scale that was presented in Table 6 as the final calibration parameters in the new experiment. All the influential parameters identified here (number 1 to 11 in Table 13) were also included in the calibration parameters as shown in Table 6. Some of the parameters included in the calibration were found to have very small impact on the MESH simulations in the post-calibration sensitivity analysis. These parameters include DDEN, VPDA, RSMN, ZPLS, and ZPLG. This observation could be a result of the sensitivity test being

local, and it is still believed that these parameters are influential in MESH simulations overall. A global sensitivity test in future can be performed to confirm this hypothesis.

Overall, both sensitivity tests revealed that the soil characteristics including soil permeable depth (depth to bedrock), soil hydraulic conductivity, and soil constituent percentages, as well as the channel roughness coefficients are among the most influential parameters in MESH model. Ideally, these parameters should be obtained directly from measured data; however, when data is not available, they should be included in the calibration. The annual maximum rooting depth of the vegetation category is also another important parameter that should be calibrated if cannot be estimated from the available or measured data.

4.4 Conclusions

This chapter reported on the contributions of this research in enhancing the MESH streamflow predictions with respect to the benchmark results reported in chapter 3. Accordingly, the calibration strategy was largely refined particularly, with respect to the sub-basins and model parameters used in calibration. Also, due to the significant computational demand of this work, two strategies were employed for increased calibration efficiency. In addition, a simple uncertainty assessment was conducted to further enhance the outcomes of the research in this chapter. A local sensitivity test was also performed after MESH was calibrated for the Great Lakes Basin.

A rigorous comparison between the new and the benchmark experiments confirmed that the MESH model performance indeed improved by the revised calibration strategy compared with the benchmark results. This improvement was a combined result of all the changes made in the calibration strategy. Therefore, it was impractical to isolate the individual impact of each change. However, two factors are believed to have contributed the most to this improvement. The first one is the increased number of the calibration sub-basins that had near-natural flow conditions and were better distributed across the entire Great Lakes Basin. The second one is a more careful selection of the calibration parameters that also takes into account parameter dependencies. Although successful results were attained, it was observed that the Great Lakes Basin geology can be correlated to model performance. Accordingly, alternate GRU definition considering geology of the basin should be considered in future studies. At the minimum, for example, GRU types can be delineated by also considering their location as being to the north or south of the Niagara Escarpment. This would reflect the impact of the Canadian Shield in the modelling study.

Table 13. Local sensitivity results for the MESH streamflow simulations to the parameters of the most common GRU type (Crop) in the Great Lakes Basin. The changes above 5% are highlighted.

Rank	Parameter	Description		Lower Value % Change Peak	Lower Value % Change Volume	Upper Value % Change Peak	Upper Value % Change Volume
1	SDEP*	Soil permeable (Bedrock) depth [m]	Average	39	3	10	1
			Max	177	13	38	4
2	WFCI*	Saturated surface soil conductivity [m/s]	Average	3	0	9	0
			Max	16	0	65	1
3	SAND*	Percent sand content for all layers [%]	Average	5	4	8	1
			Max	35	15	63	3
4	WFR2*	River channel roughness factor for river class 2	Average	6	0	2	0
			Max	46	2	16	1
5	WFR4*	River channel roughness factor for river class 4	Average	0	0	3	0
			Max	6	0	41	1
6	WFR1*	River channel roughness factor for river class 1	Average	7	0	2	0
			Max	35	1	9	0
7	WFR5*	River channel roughness factor for river class 5	Average	3	0	6	0
			Max	18	1	31	1
8	CLAY*	Percent clay content for all layers [%]	Average	3	3	4	9
			Max	24	10	23	30
9	ROOT*	Annual maximum rooting depth of vegetation category [m]	Average	1	5	2	6
			Max	12	21	24	18
10	WFR3*	River channel roughness factor for river class 3	Average	0	0	1	0
			Max	2	0	19	0
11	ZSNL*	Limiting snow depth below which coverage is less than 100% [m]	Average	3	0	2	1
			Max	13	2	6	2
12	DRN	Soil drainage index	Average	1	1	0	0
			Max	5	2	2	1
13	DDEN*	Drainage density [km/km2]	Average	1	0	1	0
			Max	3	0	3	0
14	Organic	Percent organic content for all layers [%]	Average	0	0	0	0
			Max	2	0	3	1
15	VPDA*	Vapour pressure deficit coefficient	Average	0	1	0	0
			Max	2	2	1	2
16	LAMX	Annual maximum plant area index of vegetation category	Average	0	1	0	0
			Max	1	2	1	2
17	RSMN*	Minimum stomatal resistance of vegetation category [s m^{-1}]	Average	0	0	0	0
			Max	2	2	1	2
18	PSGA	Soil moisture suction coefficient	Average	0	0	0	0
			Max	1	0	2	0
19	PSGB	Soil moisture suction coefficient	Average	0	0	0	0
			Max	1	0	1	0
20	CMAS	Annual maximum canopy mass for vegetation category [kg m^{-2}]	Average	0	0	0	0
			Max	1	0	1	0
21	LNZ0	Natural logarithm of maximum vegetation roughness length	Average	0	0	0	0
			Max	1	0	1	0
22	ALVC	Average visible albedo of vegetation category when fully-leafed	Average	0	0	0	0
			Max	1	0	0	0
23	VPDB	Vapour pressure deficit coefficient	Average	0	0	0	0
			Max	1	0	0	0
24	QA50	Reference value of incoming shortwave [W m^{-2}]	Average	0	0	0	0
			Max	1	1	1	1
25	GRKF	Fraction of the saturated soil horizontal conductivity at depth of 1 m to the saturated horizontal conductivity at the surface	Average	0	0	0	0
			Max	1	0	1	0
26	ALIC	Avg. near-IR albedo of vegetation fully-leafed	Average	0	0	0	0
			Max	1	0	0	0
27	MANN	The Manning's coefficient used for overland flow	Average	0	0	0	0
			Max	0	0	0	0
28	ZPLS*	Max. water ponding depth for snow-covered areas	Average	0	0	0	0
			Max	0	0	0	0
29	ZPLG*	Max. water ponding depth for snow-free areas	Average	0	0	0	0
			Max	0	0	0	0

* indicates the calibrated parameters in the new experiment

The GRU polishing strategy allowed for a relatively large (~10%) saving in calibration time. The model preemption implementation in MESH was not coded to track time steps simulated before preemption and hence, the preemption savings could not be estimated.

The uncertainty analysis in this study also identified eight behavioral parameter sets that were used to estimate the uncertainty bands for the MESH flow predictions. However, the low reliability values revealed that this analysis is not sufficient and more short DDS runs are required for improved results. As a secondary observation, it was seen that it is perhaps better to split a certain calibration budget, between the DDS-AU (uncertainty-based DDS optimization) methodology and the deterministic DDS approach. This way, one can also estimate parameter uncertainty on top of obtaining a "best" set. The outcomes of this computationally intensive work, i.e., the behavioral and the best parameter sets found, can be used separately or together in future relevant studies that, for example, try to estimate the best prediction and/or the uncertainty ranges for the lake inflows or the lake levels using the MESH model.

Lastly, the local sensitivity test revealed that the soil characteristics including soil permeable depth (depth to bedrock), soil hydraulic conductivity, and soil constituent percentages, as well as the channel roughness coefficients are the most influential parameters in MESH model. These parameters should be included in MESH calibration if they cannot be obtained from real measurements.

Chapter 5

Assessing the Performance of a Semi-distributed Hydrological Model under various Watershed Discretization Schemes

This chapter is a mirror of the following article submitted recently to the journal of *Hydrological Processes*, with minor changes to increase its consistency with the body of the thesis. These changes include updating table and figure numbers, and adding a small section 5.4.1 for extra supporting results. References are unified at the end of the thesis.

Haghnegahdar, A., Tolson, B. A., Craig, J. R., Paya, K. T. (2014), Assessing the Performance of a Semi-distributed Hydrological Model under various Watershed Discretization Schemes (submitted to *Hydrological Processes*)

Abstract

Physically-based distributed and semi-distributed hydrological models have become one of the primary tools for water resources studies and management over the past decades owing to increased computational capabilities and advances in data measurements. Representation of the existing heterogeneity in nature still remains one of the main challenges in these models, and is accomplished primarily via watershed discretization, subdividing watersheds into hydrologically similar land parcels. Discretization decisions in distributed modelling studies are often ad hoc and determined with little or no quantitative analysis to support these decisions. In this work, we present a quantitative methodology for assessing alternative watershed discretization schemes in terms of their corresponding model performance in ungauged basins. The effect of the time spent for calibrating each scheme is considered as part of the assessment. Here, these schemes differ in how they represent landscape attributes, and range from a simple lumped scheme to more complex ones by adding spatial land cover and then soil information. The methodology was demonstrated using the Modélisation Environmentale–Surface et Hydrologie (MESH) model as applied to the Nottawasaga river basin in Ontario, Canada. Results reveal that model performance in ungauged basins depends upon the location of the validation sub-basin (i.e. upstream or downstream) with respect to the calibration sub-basins. Also, using a more complex scheme did not necessarily lead to improved performance in validation, when constrained by calibration budget. Therefore, the calibration budget also should be considered as a factor in the assessment process. This methodology was also implemented using a shorter sub-period for calibration, which leads to substantial computational saving. Results of the sub-period test were promising and consistent particularly when sufficient budget is spent to calibrate the

model. Other strategies utilized for reducing the computational burden of the proposed analyses are also discussed in this study.

5.1 Introduction

Physically-based distributed and semi-distributed hydrological models have become one of the primary tools for water resources studies and management over the past decades, owing to increased computational capabilities and advances in data measurements. In general, these models can be divided into fully-distributed and semi-distributed models. In fully-distributed models like Soil Moisture Distribution and Routing (SMDR, Srinivasan et al., 2005), or Systeme Hydrologique Europeen (SHE, Abbott et al., 1986), the exact location for each computational element is specified and the lateral transfer of water and energy fluxes are allowed between these elements throughout the model spatial domain. In contrast, for semi-distributed models such as Soil and Water Assessment Tool (SWAT, Arnold et al., 1998), WATFLOOD (Kouwen, 1988) and Modélisation Environmentale–Surface et Hydrologie (MESH, Pietroniro et al., 2007), although the spatial variation of parameters is represented via computational elements within discretized spatial units, the exact location of each element within these spatial units are neglected.

Due to data and computational constraints, application of fully-distributed models to large (regional) scale (~10^3–10^6 km^2) watersheds can be infeasible for operational purposes. Therefore, semi-distributed models are primarily relied upon in these situations. However, specifying all the required model parameters at the appropriate scale, poses a challenge for semi-distributed models. These parameters are normally associated with the model approach to accommodate spatial heterogeneity and are typically determined via calibration.

Dealing with landscape heterogeneity in semi-distributed models is normally performed through "watershed discretization" that consists of two fundamental steps. The first step, which depends largely on the spatial resolution of the forcing data (e.g., precipitation, temperature, etc.) or the density of the flow gauges, is how to specify the size (resolution) and shape (e.g., grid cell or sub-basin) of the spatial units used in the model. The second step in watershed discretization is concerned with subdividing the aforementioned spatial units (i.e. sub-basins or grid cells) further into elements with distinct hydrologic responses. These entities will be used by the model as the computational unit for simulation. This latter decision mainly determines the representation of heterogeneity in the model and thus, largely specifies the model complexity and number of (unknown) model parameters. Hereafter, when the term "watershed discretization" or "discretization" is used it refers to this second step.

The main challenge is to identify the best method for discretization of the watershed while maximizing the model reliability and utility. Model performance and how well it can be calibrated, often with a limited budget, largely depend on how a basin is discretized in the first place. If done adequately, calibrated model parameters also generate useful predictions in ungauged basins (PUB).

Multiple studies have tried to find a suitable approach for basin discretization which reasonably balances model complexity and model practicality. Wood et al. (1988) proposed the concept of Representative Elementary Area (REA) to identify characteristic spatial scales for catchment responses in terms of the runoff generation. Reggiani et al. (1998) extended this idea to the concept of Representative Elementary Watersheds (REWs), which are sub-watersheds across which the same physical laws govern. Vivoni et al. (2004) proposed the use of Triangulated Irregular Networks (TINs), delineated primarily based on the topographic (wetness) index (Beven and Kirkby, 1979), for use in the TIN-based Real-Time Integrated Basin Simulator (tRIBS, Ivanov et al., 2004) model. Dehotin and Braud (2008) employed the concept of "hydro-landscape" units (Winter, 2001) to address sub-catchment variability, using a nested discretization approach implemented within the LIQUID modelling platform (Branger et al., 2010). Capable of taking into account various influential factors on hydrologic response such as slope, land use, geology and pedology, the hydro-landscape units are very similar to the concept of Hydrological Response Units (HRU, e.g., Leavesley et al., 1983; Flügel, 1995).

HRUs are perhaps the most commonly used approach to incorporate landscape heterogeneity in distributed watershed modelling. The exact definition of HRUs depends on the model and the application (Haverkamp et al., 2002). Flügel (1995) defines HRUs as "distributed, heterogeneously structured entities having a common climate, land use and underlying pedo-topo-geological associations controlling their hydrological dynamics". In the case of SWAT, HRUs are sub-basin divisions characterized by a unique combination of land cover, soil and land management (e.g., Neitsch et al., 2005).

Similarly, Kouwen et al. (1993) introduced the Grouped Response Unit (GRU) approach in WATFLOOD model and described it as the aggregation of multiple land classes, or possibly other landscape attributes, into a single computational element which is aggregated in order to increase computational efficiency. In most WATFLOOD or MESH model applications, the GRUs are defined as subdivisions (not necessarily contiguous) of a Cartesian grid cell rather than a topographically-defined sub-basin (e.g., Mekonnen et al., 2014; Dornes et al. 2008; 2014; Pietroniro et al., 2007).

It is well recognized that a number of parameters of semi-distributed models are unknown and have to be determined through calibration. However, modelers typically overlook the fact that the most appropriate discretization scheme is also unknown, and that it too could be determined as part of the

model calibration process. The more details incorporated into a model, the more complex and computationally costly it becomes, especially for calibration. In some cases excessive complexity may degrade model performance in validation. Consequently, there will be a trade-off between model performance and computational cost for different discretization schemes, and it remains the modeler's choice to appropriately adopt the discretization that best meets the requirements of the system under study. This crucial decision should be given extra care as it has significant impact on the model calibration and predictions.

Among numerous studies that consider watershed discretization strategies, few explicitly present an organized and objective methodology for determining an appropriate discretization scheme, one which presumably finds the balance between spatial aggregation level and model complexity. Flügel (1995) presented a fairly detailed procedure to delineate HRUs for regional scale hydrological modelling, in a manner that seeks to preserve the three-dimensional heterogeneity of the basin. This procedure involves a step-by-step GIS-based map overlay analysis to specify HRUs according to physiographic properties such as topography, soil type, geology and land use. Haverkamp et al. (2002) introduced a statistical methodology for determining the appropriate level of spatial aggregation (discretization) based on the entropy function (e.g., Krasovskaia, 1997). Conducting experiments with SWAT, they found a threshold value for the number of sub-watersheds (and HRUs) beyond which model performance is not improved any further. Dehotin and Braud (2008) proposed a flexible three-level discretization strategy for implementation in a modular modelling platform called LIQUID (Branger et al., 2010). In their approach, the spatial discretization can become finer adaptively, according to modelling objectives or numerical requirements. Determination of the hydro-landscape units in their approach is similar to the HRU delineation procedure of Flügel (1995); except that they employ a supervised classification technique to preserve possible small areas that are hydrologically influential based on the chosen processes, and could have been neglected otherwise. In a recent relevant study, Petrucci and Bonhomme (2014) compared various discretization schemes (and model structures) for water quality and quantity models in a semi-distributed urban model for a 2.3 km^2 catchment in the suburbs of Paris. All of these studies report the model success in terms of the calibration or temporal validation quality, without validating its performance spatially at other gauges. This is a cause for concern because ungauged model predictions better reflect the capability of a discretization scheme to correctly capture the relevant spatial heterogeneity found in nature. Moreover, they do not consider the potential impacts of the calibration strategy, such as calibration duration, on their results.

The primary goal of this proposed research is to quantitatively assess watershed discretization schemes, with various level of spatial data aggregation, in terms of their skill to predict flows in ungauged basins. Special attention is spent to determine the added benefit of including more finely resolved information about land cover and soil type. In addition, the sensitivity of the results to the budget spent for model calibration is examined. Lastly, it is tested whether the same results can be obtained when the main calibration period (e.g., 3 years) is replaced with a sub-period (e.g., 1 year). This is important because the use of a condensed calibration period can potentially save substantial amounts of time during the numerous calibration experiments required by this method. The methodology is demonstrated using the MESH model as applied to the Nottawasaga river basin in South-Western Ontario, Canada.

5.2 Methodology

Here we propose a comparison framework by which different watershed discretization schemes with various complexity (spatial information) levels are compared in terms of their performance, mainly in ungauged validation basins, as well as their computational budget. First, multiple candidate watershed discretization schemes are selected. Then, each one is calibrated to the same record, for a finite budget. Finally, each scheme is validated at hypothetical ungauged basin(s), using appropriate metrics.

Any candidate discretization scheme can be assessed using this methodology, ranging from a simple lumped scheme to one that is built using all available spatial information. In between these extremes, one may choose to test the effectiveness of a scheme which is commonly used in practice, as has been done here. In practice, it is suggested to start with the simplest scheme, and add more complexity sequentially until additional discretization ceases to improve model quality. Modellers can then stop assessing more complex schemes based on their subjective evaluation of the model performance.

A fixed calibration budget (clock time) was considered as the maximum budget available to calibrate each scheme. Sensitivity of the results to the choice of calibration budget is also explored in this study by considering three different calibration budgets of one, two and four days. To account for the stochastic nature of the search algorithm with parameter estimation during calibration, multiple replicates should be used to evaluate each scheme. In this study, three replicates were used to calibrate and validate each scheme per each calibration budget. As a result, a total of 36 calibration runs (3 replicates per 4 discretization schemes per 3 calibration budgets) were performed for application of the comparison framework in this study.

In order to test whether calibrating to a sub-period leads to similar results as calibrating to the original period, the comparison methodology was applied to a calibration sub-period (e.g., one year) with the

highest correlation to the main calibration period. In this study, the second year was identified as the suitable sub-period and subsequently, another 36 calibration runs were conducted when testing the sub-period experiment.

5.3 Case Study: MESH model applied to the Nottawasaga River Basin

5.3.1 MESH Model

MESH is a land surface-hydrological model, developed by Environment Canada (Pietroniro et al., 2007) for various water resources management applications mostly at large scales (e.g., Deacu et al., 2012; Dupont et al., 2012; Haghnegahdar et al., 2014). MESH evolved from WATCLASS (Soulis et al., 2000), which was a combination of the routing module of WATFLOOD (Kouwen et al., 1993) and the land surface scheme CLASS (Verseghy, 1991; Verseghy et al., 1993).

WATFLOOD is a semi-distributed hydrological model initially developed for flood forecasting. It has been widely used for various water resources applications (e.g., Stadnyk et al., 2014; Roberts et al., 2012; Dorner et al., 2006). It emphasizes the use of remotely sensed and radar data and thus, operates on a domain divided into rectangular grid cells (Kouwen et al., 1993). Grid cells are further subdivided into the Grouped Response Units (GRUs, Kouwen et al., 1993), to accommodate the spatial heterogeneity within each grid cell. The land surface scheme, CLASS, simulates the energy and water balances of the soil, snow, and vegetation canopy for each GRU type in each grid cell. The amount of surface runoff (as well as other fluxes) is then calculated for each grid cell by the weighted area average of the existing GRUs within each cell. Then water is routed by WATFLOOD between the grid cells and across the river network for the entire basin. The overland flow in each GRU is calculated using the Manning's approximation of the kinematic wave velocity. The interflow, which accounts for the downward hillslope flow, is estimated from the bulk saturation of each soil layer calculated at each time step. Baseflow is simply treated in MESH as any water that percolates out of the bottom of the soil column in a GRU, and is sent to the stream network immediately.

Pietroniro & Soulis (2003) found the GRU concept suitable for semi-distributed hydrological modelling in large basins, because it is simple to implement while respecting the physics. Employing the GRU concept in WATFLOOD and subsequently MESH, facilitates the calibration of the model on the whole domain at once, resulting in significant time savings (Pietroniro et al., 2007).

The MESH model requires input forcing data of precipitation, air temperature, wind speed, incoming short wave radiation, incoming long wave radiation, specific humidity and barometric pressure. Except for precipitation, the archived gridded forecasts for the forcing data were produced by the Regional

Deterministic Prediction System (RDPS) of EC. This is reliant upon the regional configuration of the Global Environmental Multiscale-Numerical Weather Prediction (GEM-NWP) model (Mailhot et al., 2006). The data was available at a spatial resolution of 0.1667 degree and a temporal resolution of 1 hour. For the precipitation, the gridded data from the Canadian Precipitation Analysis (CaPA, Mahfouf et al., 2007) is used. This data is produced by combining different types of precipitation observations with the archived GEM gridded precipitation forecasts. In the absence of reliable snowfall observations over such large areas, using the GEM background field is considered to be the best option available for estimating solid precipitation.

Version 1.3.006 of MESH was used in this work with a minor modification. The only change to the model was the implementation of model preemption (Razavi et al., 2010) based upon the sum of squared errors (SSE) metric, as described later in the "Calibration Strategy" section.

5.3.2 Study Area

The study is conducted on the Nottawasaga River basin in South-Western Ontario, Canada. It is located between Georgian Bay and Lake Simcoe, and drains into the Georgian Bay. This watershed was selected because it consists of several sub-basins with near-natural flow condition and complete record for the time period under study. These attributes allow for multi-site calibration and validation of the model, which enables a better assessment of the model performance. Furthermore, this basin was deemed to have sufficient variability in land cover and soil type for the analyses within the context of this study. This watershed and the corresponding sub-basins used in this analysis are shown in Figure 21. Table 14 indicates the characteristics of these sub-basins.

The entire drainage area of nearly 2700 km^2 was discretized into 20 grid cells of about 15 by 15 km. This resolution was mainly dictated by the highest available resolution of the input forcing data. Consequently, as seen in Table 14, the modeled drainage areas in MESH will be sometimes different than the reported areas by EC. In order to circumvent this deficiency, the gauge coordinates in MESH were adjusted as suggested by Pietroniro at al. (2007). This strategy still would not necessarily result in a match between real and modeled areas, so, MESH simulated flows at each gauge were additionally multiplied by a scale factor when calculating the objective function. This scale factor is simply the ratio of the EC drainage area to the modeled drainage area, and is reported in Table 14 for each sub-basin.

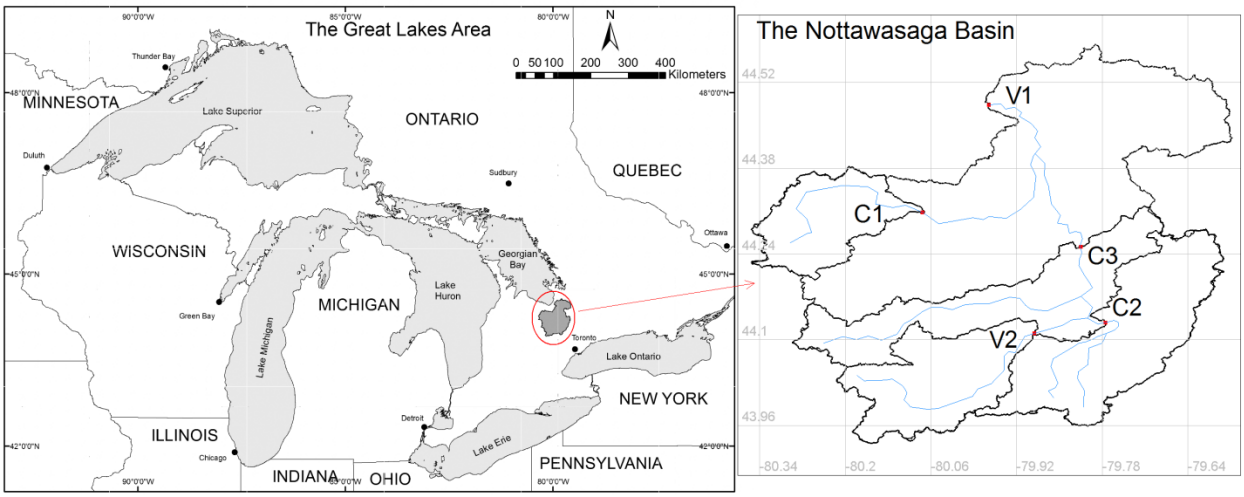

Figure 21. The Nottawasaga River Basin near Edenvale and its sub-basins. Calibration sub-basins are labelled with "C" and validation sub-basins are labelled with "V".

Table 14- Characteristics of the sub-basins.

	Label	EC ID	EC Area (km^2)	Adjusted Coordinates in MESH?	Modeled Area in MESH (km^2)	Scale Factor	Weight
Calibration	**C1**	**02ED015**	239	Yes	246.2	0.971	0.590
	C2	**02ED029**	461	No	419.6	1.099	0.199
	C3	**02ED003**	1180	No	1271.1	0.928	0.211
Validation	**V1**	**02ED027**	2680	No	2672.3	1.003	0.891
	V2	**02ED101**	334	Yes	358.0	0.933	0.109

5.3.3 Watershed Discretization Schemes

Four different discretization schemes were considered here as shown in Figure 22: lumped, 1-GRU, 7-GRU and 16-GRU. These schemes all use an identical grid number and resolution, but have various levels of complexity in terms of the number or type of GRUs they use to subdivide grid cells. The lumped scheme is the simplest scheme, where only one single GRU type is used for all grid cells across the entire watershed. It therefore, has the lowest simulation runtime. Other schemes are understood better if the 7-GRU scheme is defined first. Using the land cover map from Ontario Ministry of Natural Resources, land cover data was subdivided into seven classes of Cropland, Grassland, Deciduous forest, Coniferous forest,

Mixed forest, Exposed land (or impervious) and Water. These seven land cover-based GRU types were used to define MESH GRUs for the 7-GRU scheme. In this scheme, each grid cell contains up to seven GRU types per grid cell that are solely defined based on land type, with no recognition of the impact of varying soil types. These GRU types are shown in Table 15. It is noteworthy that this land-cover-only scheme is commonly adopted by WATFLOOD and MESH modelers for regional scale studies (e.g., Kouwen et al., 1993; Soulis et al., 2000; Pietroniro et al., 2007, Haghnegahdar et al., 2014).

Table 15- The seven GRU types produced by spatial land cover data for the Nottawasaga river basin, ranked by coverage area.

GRU NO.	GRU Type	Area Covered (%)
1	Cropland	57.1
2	Deciduous Forest	14.6
3	Grassland	12.5
4	Mixed Forest	6.7
5	Coniferous Forest	5.9
6	Water	2.6
7	Impervious	0.5

In the 1-GRU scheme, each grid cell gets assigned 100% to only the dominant GRU type, out of the seven possible types in the 7-GRU scheme. This 1-GRU scheme may be preferred option because it has same efficiency as the lumped scheme, but still allows for some land cover variability. It also runs much faster than the 7-GRU scheme making it more practical to be coupled with large scale atmospheric models. Note that the 1-GRU scheme has the same number of GRU types, and thus calibration parameters, as the 7-GRU scheme. Their only difference lies in the number of possible GRU types per grid cell. The 7-GRU scheme can have up to seven GRU types in each cell; whereas, out of these seven types, the 1-GRU scheme has only one GRU type per grid cell.

Finally, the 16-GRU scheme has the highest level of complexity since it contains up to sixteen GRU types per grid cell. These GRUs are formed by adding the soil spatial data to the land cover classes of the 7-GRU scheme. At the first step, soil data from the Land Information Ontario, was subdivided into eight classes of Clay Loam, Clay, Gravelly, Impermeable, Organic, Rock, Sandy and Silty. Then, for both land cover and soil type classes, the ones with less than 10% of area coverage in the study area were reclassified as the dominant class type, which was Cropland (~56%) for land cover and Sandy Soil (~47%) for soil type. Finally, the remaining land cover and soil classes were merged to construct sixteen new GRU types shown in Table 16.

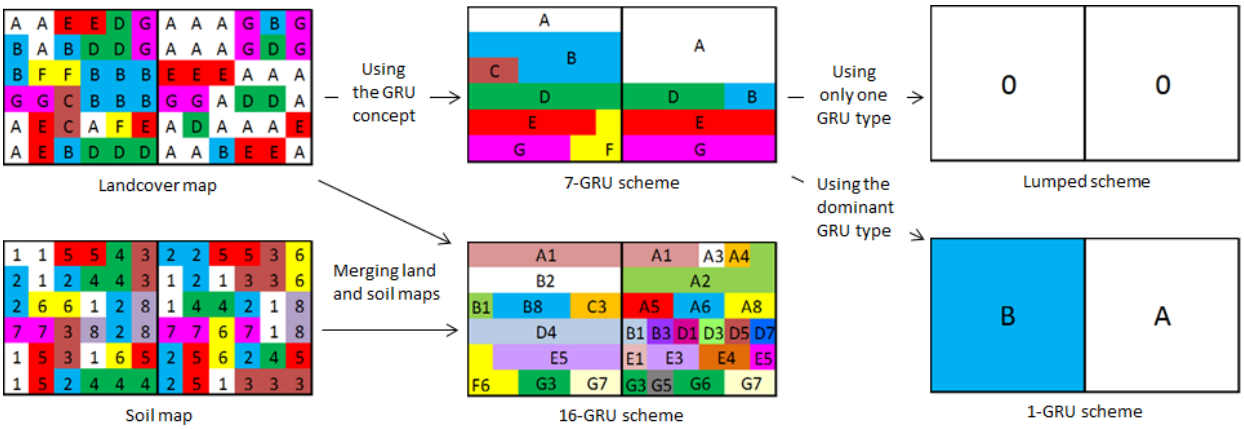

Figure 22. Schematic of the GRU concept in representing the sub-grid variability in land classes using the land class-based 7-GRU scheme in two given grid cells, and the relationship between various discretization schemes used in this study. The letters and numbers represent different types of GRUs. Note that the exact position of the GRU types within the grids does not matter and only their fractions are used in calculations.

Table 16- The sixteen GRU types produced by combining land cover and soil spatial data for the Nottawasaga river basin, ranked by coverage area.

GRU NO.	GRU Type	Area Covered (%)	GRU NO.	GRU Type	Area Covered (%)
1	Cropland, Sandy	36.3	9	Grassland, Silty	3.7
2	Cropland, Silty	18.3	10	Deciduous Forest, Organic	2
3	Deciduous Forest, Sandy	8.7	11	Mixed Forest, Organic	1.3
4	Grassland, Sandy	7.4	12	Mixed Forest, Silty	1.2
5	Cropland, Clay Loam	7.3	13	Grassland, Clay Loam	0.7
6	Mixed Forest, Sandy	4.2	14	Deciduous Forest, Clay Loam	0.5
7	Cropland, Organic	4.1	15	Grassland, Organic	0.3
8	Deciduous Forest, Silty	3.9	16	Mixed Forest, Clay Loam	0.1

5.3.4 Calibration Strategy

As shown in Figure 21 and Table 14, three sub-basins (C1, C2, C3) are used for calibration and two (V1, V2), presumably "ungauged", for spatial validation. The calibration period runs for three years from 1

October 2004 to 30 September 2007, and is preceded by a one year warmup period that begins on 1 October 2003.

Generally, any objective function can be used in this methodology. Here, a weighted sum of the Nash-Sutcliffe (NS) coefficient of efficiency (Nash and Sutcliffe, 1970) for daily flows is used, according to Equation (3), as the main criterion to evaluate and compare model performance. The weights were assigned based on the long-term average of daily flows at each sub-basin, in a way that they add to one.

$$NS_{weighted} = \sum W_i \cdot NS_i = \sum W_i \cdot \left(1 - \frac{SSE_i}{n_i \cdot Var_i}\right) = 1 - \sum \frac{W_i \cdot SSE_i}{n_i \cdot Var_i} \quad (3)$$

Here, i denotes the sub-basin index, W_i is the weight for sub-basin i, SSE_i is the sum of squared errors between simulated and observed flows over the calibration period for sub-basin i, Var_i is the variance of observations sub-basin i, and n_i is the number of observations over the calibration period for sub-basin i.

Due to the computation burden of the methodology, it was desired to use the model preemption strategy (Razavi et al., 2010) to save calibration time. According to Razavi et al. (2010), model preemption involves dismissing model runs that will definitely result in a poorer model performance compared to the current best solution in calibration. They reported a saving of nearly 50%, when model preemption was applied for calibration of MESH to a watershed of approximately 55 km^2. It, however, requires using a monotonically increasing objective function, such as Sum of Squared Errors (SSE), during calibration (Razavi et al., 2010). Accordingly, a transformed equivalent SSE objective function (Z) was used during the calibration process as shown in Equation 4 below.

$$Minimize\ Z = \sum \frac{W_i \cdot SSE_i}{n_i\, Var_i} \quad (4)$$

As seen from the above equations, minimizing a weighted sum of SSE values is equivalent to maximizing a weighted sum of NS values. Due to the nature of the optimization algorithm used in this study, the Dynamically Dimensioned Search (DDS, Tolson and Shoemaker, 2007), this transformation would not change the calibration results at all. DDS is a rank-based optimization algorithm in that only the relative rank of the candidate solutions influences the search result (as opposed to the absolute

magnitude of objective function differences between candidate solutions), and the exact same answer would be obtained with or without this transformation.

DDS is a stochastic global search algorithm that is designed for optimization problems with many decision variables. DDS has been successfully applied to a number case studies before, and has shown to be efficient for finding good solutions in complex models with limited calibrated budget (Tolson and Shomaker, 2007; Dornes et al., 2008; Haghnegahdar et al., 2014). Razavi et al. (2010) classified DDS as an "ideal" algorithm for model preemption strategy. These attributes make DDS appropriate for this comparison framework since the calibration time is substantial; however, any other optimization algorithm may be used. The model calibrations were conducted using DDS as implemented in OSTRICH, which is a model-independent calibration and optimization tool (Matott, 2005) consisting of a number of popular optimization algorithms including DDS. The maximum number of model simulations (objective function evaluations) also has to be specified in DDS as the stopping criteria. This number is different for each discretization scheme, and was calculated based on the total calibration time divided by a nominal runtime for one MESH simulation under each scheme.

MESH simulations are performed for each GRU present in a given grid cell. Therefore, the 1-GRU scheme runs much faster than the 7 and 16-GRU schemes. Accordingly, another simple and yet efficient strategy implemented in this work to reduce simulation runtimes was GRU polishing. This strategy involves removing GRUs from each grid cell with area coverage fractions below a certain threshold. The idea is that GRUs with areas below a certain threshold will have very small impact on streamflow simulations. This strategy was applied to the 7 and 16-GRU cases, with 5% and 10% thresholds. After assessing the differences in hydrographs induced by these thresholds, 5% was deemed appropriate. For both 7 and 16-GRU cases, this resulted in a substantial reduction of nearly 23% and 46% in simulation time, respectively, while the associated changes in simulated hydrographs remained very small (average deviation from the highest peak in calibration sub-basins was less than 5 m^3/s). Consequently, a polished version of the 7 and 16-GRU schemes were used for performing calibration experiments.

Simulation runtime and the maximum number of model evaluations in DDS are reported in Table 17, for the three calibration budgets of one, two and four days. Model simulations were performed at hourly numerical time steps for a four year period (one warmup year plus three calibration years). The simulation runtime in each case was measured using an Intel Core i7-920 2.67 GHz workstation, running a 64-bit version of the Windows 7 operating system, with 12 gigabytes of installed memory, and a quad-core processor. In order to account for the stochastic nature of the optimization algorithm, each calibration run was repeated three times, with three different initial solutions, for each case. To maintain consistency

between all cases, these three initial solutions were intentionally kept identical for each scheme when calibrated for different budgets. The initial solutions were also kept unchanged between the schemes with identical calibrated parameters (i.e. 1-GRU and 7-GRU).

Table 17- The number of parameters, simulation runtime, and the maximum number of model evaluations used in DDS for each calibration budget for each discretization scheme.

		Lumped	1GRU	7GRU (5% polished)	16GRU (5% polished)
Number of calibrated parameters		12	34	34	40
Runtime for Oct.03 to Sep.07 simulation (sec)		49	51	137	169
Maximum no. of model evaluations for calibration budget of	**4 days**	7053	6776	2523	2045
	2 days	3527	3388	1261	1022
	1 day	1763	1694	631	511

The calibration parameters and their corresponding ranges were specified using a combination of expert opinion, a local sensitivity analysis, and suggested values from previous relevant studies (e.g., Dornes et al., 2008; Davison et al., 2006; MacLean, 2009; Haghnegahdar et al., 2014). Table 18 lists these parameters and their corresponding ranges. The first eleven parameters in this table are GRU dependent, and are calibrated for each GRU type that is chosen for calibration. The last parameter (channel roughness) is linked to river class types. Pietroniro et al. (2007) delineated five river classes based on the geomorphology of the rivers in the Great Lakes Basin. According to that same classification, Nottawasaga river basin contains only one river class type, and only a single channel roughness is used.

Out of the seven land classes considered in this basin, crop (~57%), deciduous forest (~15%) and grass (~13%) classes, each cover more than 10% of the total area (see Table 15), and were considered as three GRU types for calibration in 1 and 7-GRU schemes. Consequently, nearly 85% of the watershed was calibrated in these cases. For the 16-GRU scheme with added soil properties, only the five GRU types with the coverage areas above 5% of the entire watershed were considered for calibration. As seen in Table 16, these GRU types are cropland-sandy, cropland- silty, deciduous forest-sandy, grassland-sandy, and cropland-clay loam. As a result, nearly 80% of the watershed was calibrated. In the case of the lumped scheme, only a single GRU type, which was assumed to be cropland, is calibrated. Consequently, as shown in Table 17, the total number of calibrated parameters was 12, for the lumped scheme, 34 for the 1 and 7-GRU schemes, and 40 for the 16-GRU scheme.

Table 18- Parameters and their corresponding ranges used in calibration of MESH.

NO.	Parameter	Description	(Lower Bound, Upper Bound)
1	ROOT	Annual maximum rooting depth of vegetation category (m)	(0.2, 1.0) (1, 3.5) for Deciduous forest
2	RSMN	Minimum stomatal resistance of vegetation category (s m^{-1})	(60,110) crop (75,125) grass (100,150) deciduous forest
3	VPDA	Vapour pressure deficit coefficient (used in stomatal resistance formula)	(0.5,1)
4	SDEP	Soil permeable (Bedrock) depth (m)	(0.35,4.10)
5	DDEN	Drainage density	(2,100)
6	SAND	Percent sand of all soil layers (%)	(0,100) In 16-GRU scheme: (30,65) for Clay Loam, (85,100) for Sandy soil, and (0,20) for Silty soil
7	CLAY	Percent clay of all soil layers (%)	(0,100) In 16-GRU scheme: (30,40) for Clay Loam, (0,10) for Sandy soil, and (0,15) for Silty soil
8	RATIO	The ratio of horizontal to vertical saturated hydraulic conductivity	(2,100)
9	ZSNL	Limiting snow depth below which coverage is less than 100% (m)	(0.05,1)
10	ZPLS	Maximum water ponding depth for snow-covered areas (m)	(0.02,0.15)
11	ZPLG	Maximum water ponding depth for snow-free areas (m)	(0.02,0.15)
12	WFR2	Channel roughness factor	(0.02,2)

The sand and clay percentages, as seen in Table 18, range from 0 to 100% for all schemes except for the 16-GRU scheme where ranges are further narrowed down for each soil type, based on the standard USDA soil texture triangle. Organic matter was kept constant at 5%, and the silt percentage was calculated as the remainder by the model, such that all the soil constituents added to 100%. Accordingly, the soil flag in MESH was used, which adjusts percentages proportionally when soil constituent percentages add up to more than 100%. The RATIO parameter is not a MESH parameter. Rather, it is a calibration parameter introduced in this study to tie horizontal and vertical saturated hydraulic conductivity together for a given soil type. Vertical saturated hydraulic conductivity is calculated using percent sand in MESH. Then, instead of calibrating the horizontal saturated hydraulic conductivity independently, this RATIO parameter is used to obtain the horizontal saturated hydraulic conductivity from the vertical one. In other words, the horizontal saturated hydraulic conductivity, which is used by the interflow module in MESH, is calibrated indirectly in this study.

Non-calibrated parameters related to soil or land cover types were assigned according to the instructions on the CLASS manual version 3.5, and the values found from a recent calibration of the entire Great Lakes Basin (Haghnegahdar et al., 2014).

5.4 Results and Discussion

Results are presented for two cases: the main (3-year) calibration period (October 2004 to September 2007) and the sub-period calibration (October 2005 to September 2006). In both of these scenarios, validation performance is evaluated at ungauged basins for the main 2004-2007 period. Both calibration and validation performance is assessed in terms of NS. For each case, the overall model performance for calibration and validation modes are shown. In validation, which is the primary assessment criteria in this study, model performance is also shown for each validation sub-basin individually (individual NS values). Since each calibration experiment was repeated three times, each graph shows the median, minimum and maximum performance of these three runs for each case. Furthermore, NS values associated with each initial solution, indicating non-calibrated model performance, are also plotted in each figure. For display purposes, points are shifted by 3 hours relative to each other on the time axis.

Figure 23 shows the overall calibration and validation performance of various discretization schemes, after calibration over the main period (October 2004 to September 2007), for three calibration budgets (1, 2 and 4 days) and for the initial solutions without any calibration. Performance is indicated in terms of the weighted sum of NS values for the calibration and validation sub-basins. In the calibration, as seen in Figure 23a, especially prior to any calibration, and also for each calibration budget, performance increases when a more complex scheme is used. There is one exception to this trend when going from 7-GRU to 16-GRU scheme, where model performances are comparable (for 1 and 2 day) or a bit higher (for 4 day) for the 7-GRU scheme. This could be because the number of model evaluations was higher for the 7-GRU scheme than the 16-GRU scheme (as shown in Table 17), thus, providing the search algorithm with a higher chance to find a better solution. Unsurprisingly, model performance levels are very poor without any calibration, especially for the lumped scheme. They significantly improve even using a one day calibration budget. Beyond a one day budget, there is only slight increase in model performance with more calibration effort.

In validation (Figure 23b), performance increases with added complexity to the schemes, prior to calibration and for all calibration budgets, except for one case. It increases from nearly 0.4 for the lumped schemes to 0.6 for the 16-GRU scheme in all calibration budgets. The exception to this improvement is where the 7-GRU scheme performs slightly poorer than the 1-GRU scheme when calibrated for one day only. However, the 1-GRU scheme shows a higher variation in performance compared with the 7-GRU

scheme. Variance of validation results is greatest for the lumped scheme, confirming it is a poor scheme for prediction in ungauged basins. Similar to what was seen with the calibration results, model performance improves significantly even after a one day calibration time and beyond this budget there is only a slight gain in performance with further increase of the calibration effort.

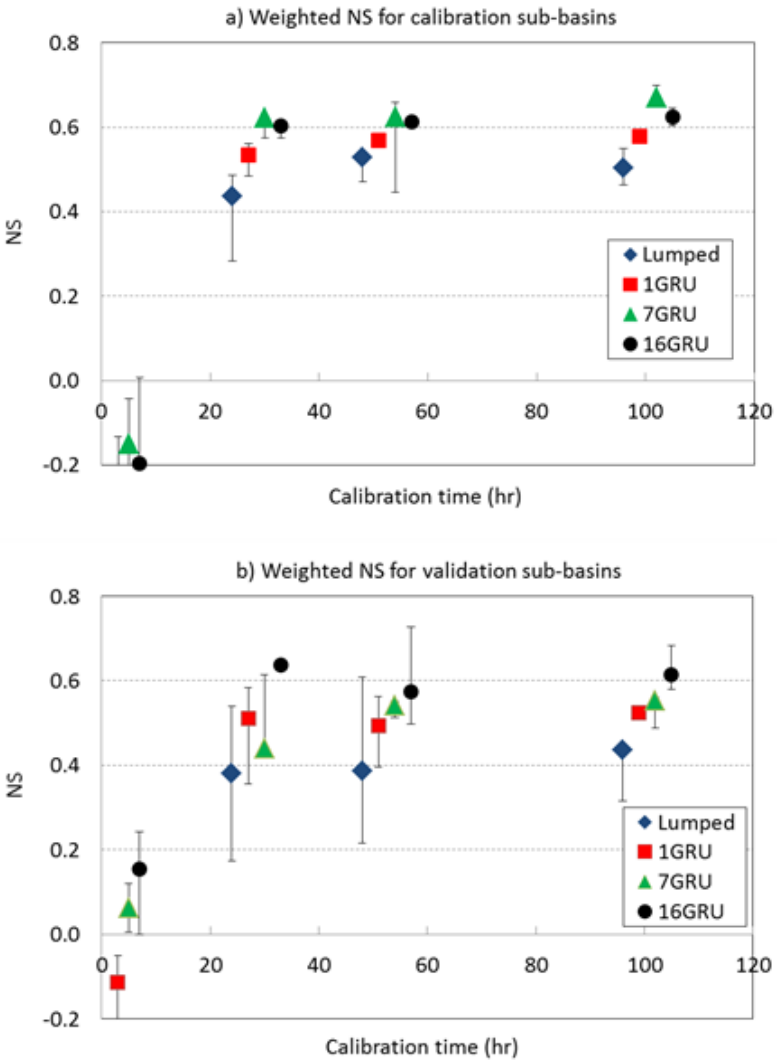

Figure 23. Overall a) calibration, and b) validation performance of various discretization schemes, before calibration and after calibration over the main period (October 2004 to September 2007), for three calibration budgets of 24, 48 and 96 hours. Symbols indicate the median performance for the three calibration replicates. Minimum and maximum values are displayed with bars. For better display purposes, lateral positions are shifted by 3 hours and minimum y-axis was fixed at -0.2.

Figure 24 shows the individual validation performance of each sub-basin (V1 and V2) for various discretization schemes, after calibration over the main period (October 2004 to September 2007), for three calibration budgets (1, 2 and 4 days) and prior to any calibration. As expected, for sub-basin V1 (Figure 24a), which is the main and largest sub-basin with a weight of 0.892 (see Table 14), validation results are very similar to the overall validation performance (Figure 23b). Performance after calibration ranged from nearly 0.4 to 0.6, and the 16-GRU scheme always has the highest performance. This pattern is not surprising because almost 90% of the overall weighted NS score is contributed by this sub-basin.

Figure 24b shows a very different pattern for sub-basin V2. In general, NS values are lower compared to the V1 sub-basin, going from around -0.2 to for the lumped scheme to as high as almost 0.6 for the 16-GRU scheme in the four day calibration case. This behavior could be explained by the fact that the V2 sub-basin is an interior sub-basin of the calibration sub-basins, whereas the V1 sub-basin is located downstream of all the calibration sub-basins. Therefore, unlike the V2 sub-basin, most of the flow at sub-basin V1 is already calibrated, leading to higher NS values. Furthermore, in contrasting the results in Figure 23a and Figure 24b it can be confirmed that the overall calibration performance is not indicative of the interior ungauged basin performance. Similar behavior was reported by Pokhrel and Gupta (2010) where they see degraded predictions for the interior gauges that were not explicitly calibrated to. Moving to a more complex discretization scheme always increases model performance in sub-basin V2 except for two cases. In these two cases, the 7-GRU scheme shows comparable (for one day calibration) or a little higher performance (for two day calibration) in relation to the 16-GRU scheme. This trend could indicate that adding spatial soil information while only adding a few more parameters, can increase performance as long as the model is calibrated long enough. The lumped scheme constantly performs very poorly in this sub-basin having a NS value of less than or equal to zero for all calibration budgets. It again has the greatest variance followed by the 7-GRU scheme.

To investigate the feasibility of using truncated calibration time period for assessing discretization choices, these tests were repeated for when the model is calibrated to a sub-period of one year only (second year here). The results of these tests are intended to determine whether a reduced period for calibration, which saves considerable computational cost, can yield similar results to the main 3-year calibration period. Since the sub-period calibration results were effectively same as those for the main period, they are not presented. Moreover, focus is placed only on comparing the overall model performance in ungauged basins and not on the individual sub-basins.

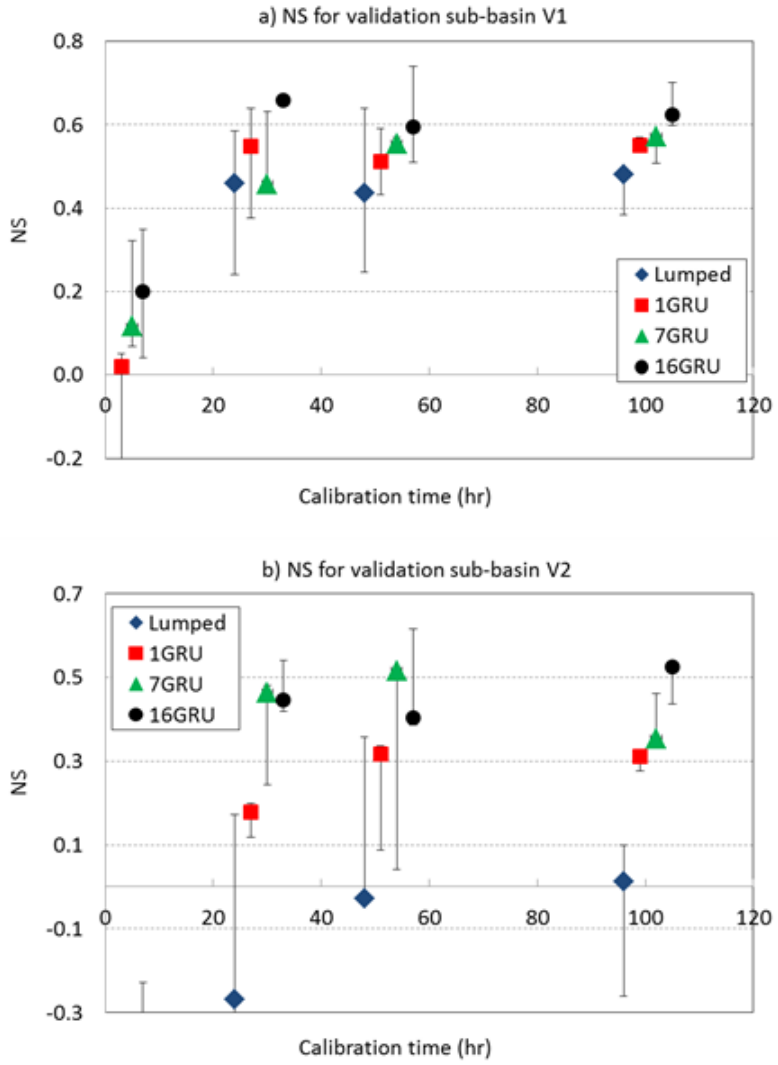

Figure 24. Individual validation performance of sub-basin a) V1, and b) V2 for various discretization schemes, before calibration and after calibration over the main period (October 2004 to September 2007), for three calibration budgets of 24, 48 and 96 hours. Symbols indicate the median performance for the three calibration replicates. Minimum and maximum values are displayed with bars. For better display purposes, lateral positions are shifted by 3 hours and minimum y-axis was fixed at -0.3.

Figure 25 compares the overall validation results between the main period (October 2004 to September 2007) and the sub-period calibration experiments, for three calibration budgets of 24, 48 and 96 hours. Overall, results are consistent between the two experiments and thus, similar information can be obtained from both of them. In both scenarios, the model performance improves with a more complex discretization scheme for all calibration budgets. The only exception to this trend in the main period calibration case is between the 1-GRU and the 7-GRU schemes for the one day calibration time. Again,

the 16-GRU had the best performance regardless of the budget spent for calibration. Also, the greatest variance is seen for the lumped scheme followed by the 1-GRU scheme in both experiments. The NS values are lower for the sub-period experiment, compared to the main period scenario, especially for the lumped and the 1-GRU schemes. This fact is because in the sub-period experiment, the model was calibrated to one year only, but was validated against all three years (i.e., it is a validation both in space and time). Similar comparison results (not shown) were observed for the individual validation sub-basins between the two experiments.

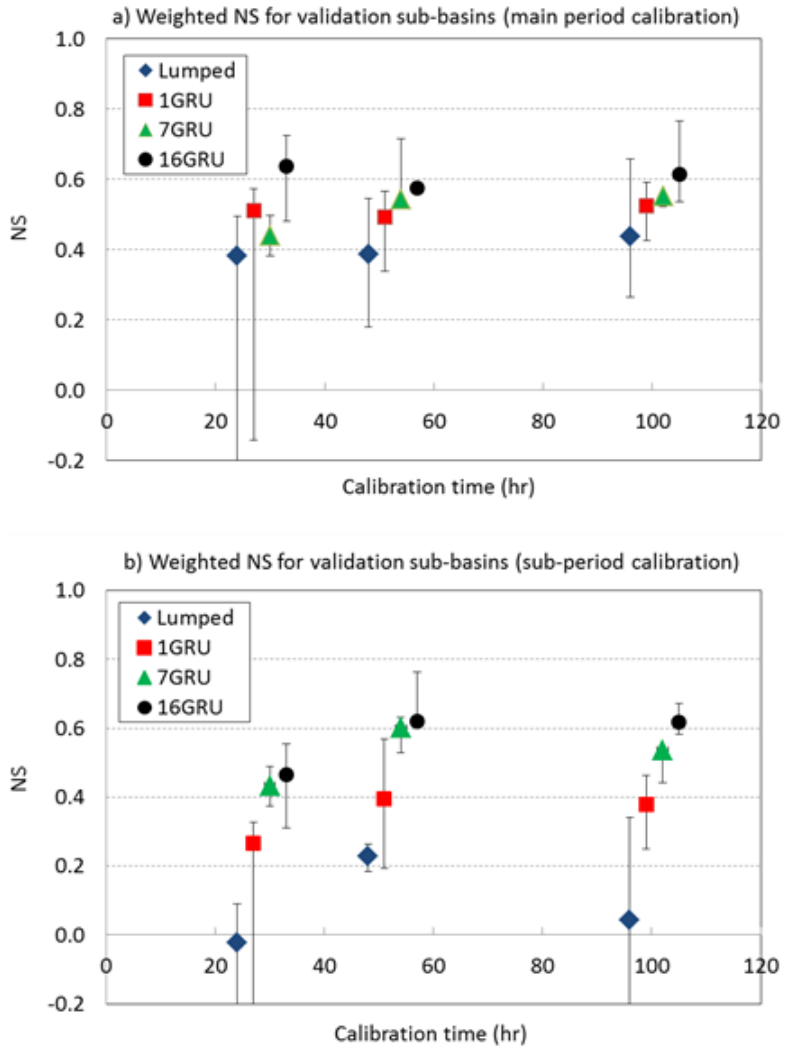

Figure 25. Comparison of the overall validation results between a) the main period (October 2004 to September 2007), and b) the sub-period (October 2005 to September 2006) calibration experiments, for three calibration budgets of 24, 48 and 96 hours. Symbols indicate the median performance for the three calibration replicates. Minimum and maximum values are displayed with bars. For better display purposes, lateral positions are shifted by 3 hours and minimum y-axis was fixed at -0.2.

5.4.1 Detailed assessment of Discretization Schemes

As seen in Figure 24, in both validation sub-basins, model performance (i.e., NS value) increases with using a more complex scheme when model is calibrated for four days. In order to further examine this finding, these schemes are also compared using hydrographs and the additional metrics defined in Table 2. Accordingly, Figure 26 and Figure 27 display the simulated hydrographs and the values of these metrics for validation sub-basin V1 and V2, respectively. These results are associated with the best overall validation performance for each scheme after four days of calibration.

As seen in Figure 26, these results confirm that overall the model performance in sub-basin V1 increases with the scheme complexity when model is calibrated sufficiently for four days. All the metrics improve with increased complexity of the schemes except for PBIAS. The 1-GRU scheme has the lowest PBIAS and the 7-GRU scheme has the largest PBIAS. The simulated hydrographs also improve according to the scheme complexity, particularly in simulating low flows and lower peak flows. The highest peak flow gets more overestimated in more complex schemes.

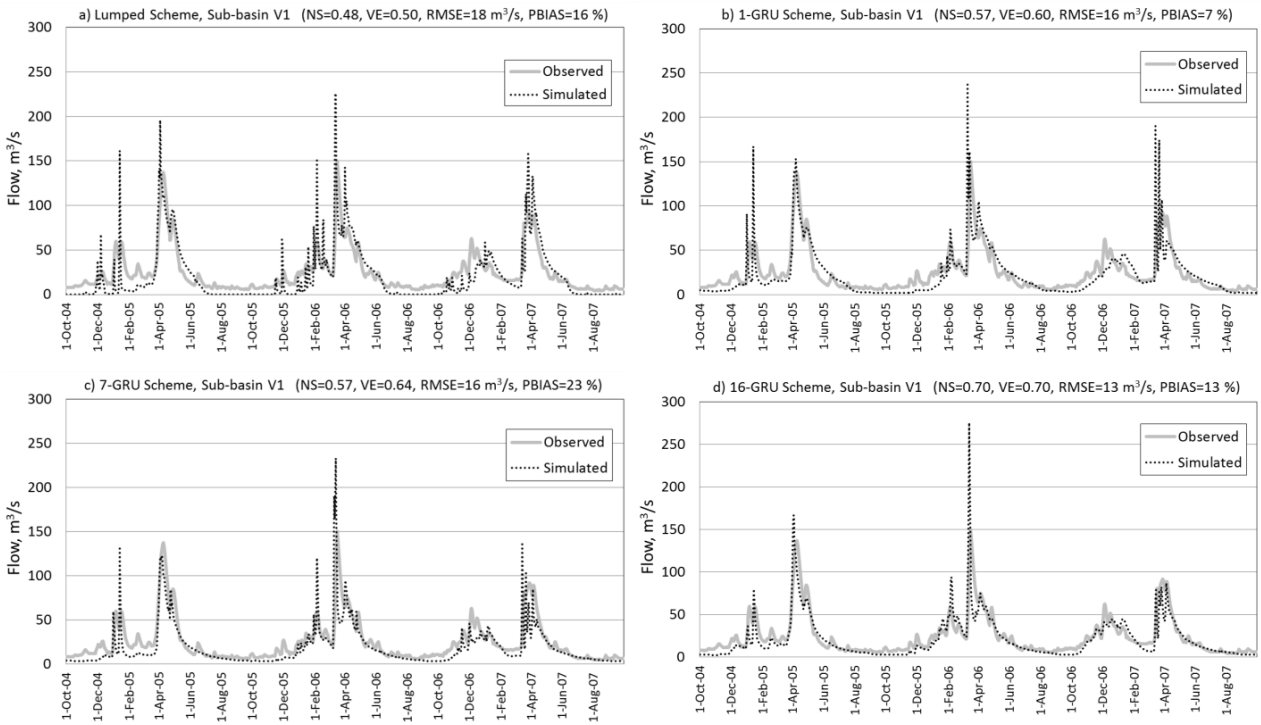

Figure 26. Simulated hydrographs and various metrics for validation sub-basin V1 for various discretization schemes a) Lumped, b) 1-GRU, c) 7-GRU, and d) 16-GRU.

As seen in Figure 27, these results also confirm that overall, similar to the sub-basin V1, the model performance in sub-basin V2 also increases with the scheme complexity when model is calibrated sufficiently for four days. All the metrics improve with increased complexity of the schemes except for PBIAS. Again, the 1-GRU scheme has the lowest PBIAS and the 7-GRU scheme has the largest PBIAS. The simulated hydrographs also improve according to the scheme complexity, particularly in simulating low flows and lower peak flows. The highest peak flow gets more overestimated in more complex schemes. Consequently, it is observed that consistent conclusions can be drawn when the validation model performance is assessed using multiple criteria compared with when NS is only used. The proposed comparison methodology was also applied using synthetic experiments. The corresponding results are presented in Appendix A.

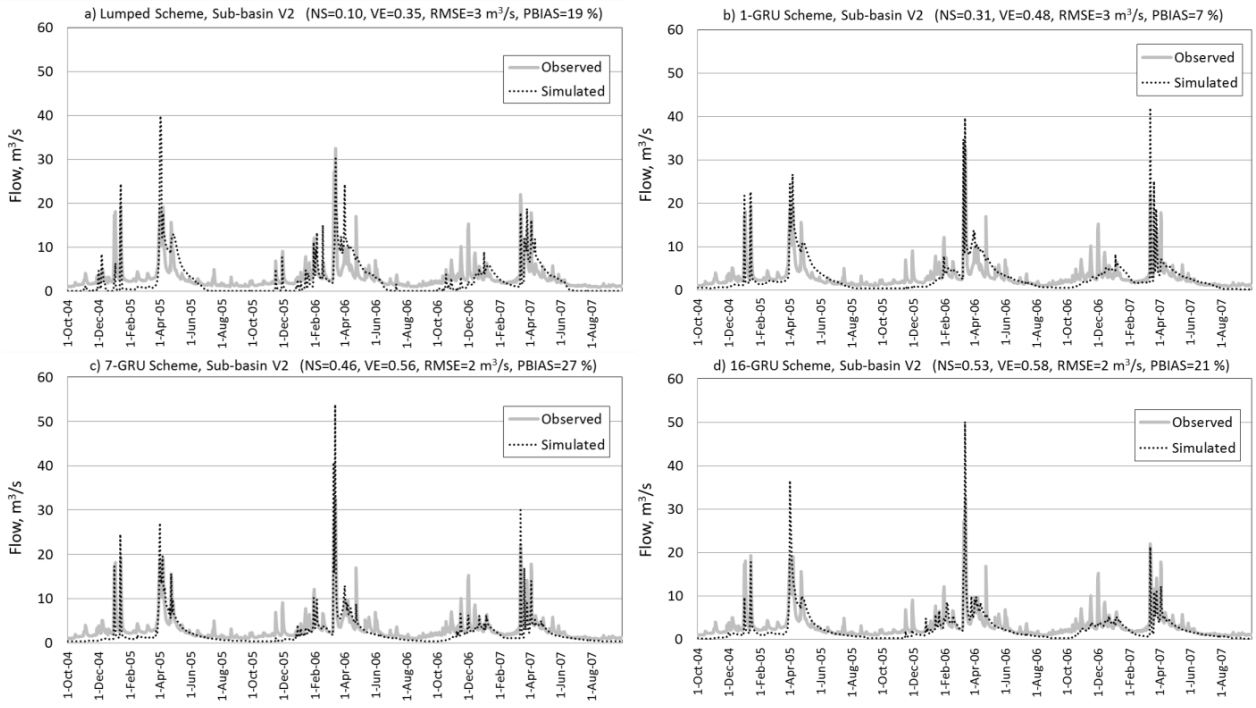

Figure 27. Simulated hydrographs and various metrics for validation sub-basin V2 for various discretization schemes a) Lumped, b) 1-GRU, c) 7-GRU, and d) 16-GRU.

5.5 Conclusions

In this study, we introduced a quantitative methodology to assess and compare watershed discretization schemes in terms of their accuracy in predicting flows in ungauged basins. The effect of adding more land cover and soil information to watershed discretization was evaluated by applying a semi-distributed land

surface-hydrological model, MESH, to the Nottawasaga River basin in South-Western Ontario, Canada. Moreover, impact of the calibration budget on the results was investigated. Finally, it was tested whether the same results can be observed when model is calibrated to a sub-period as opposed to the full calibration period. In each case, model was calibrated in three sub-basins and validated in two other sub-basins, one downstream and one upstream of the calibration sub-basins.

The proposed methodology systematically provides quantitative assessment of the model performance under various watershed discretization schemes in the context of prediction in ungauged basins. Results confirm that calibration period model performance is not indicative of its performance for validation in ungauged basins and therefore, is not sufficient for assessing alternative discretization schemes. It is also concluded that the model predictions in ungauged basins does not necessarily improve with using a more complex scheme when the time spent on calibration is insufficient. So, the calibration budget also plays an important role in determining the better scheme. The methodology also allows for identification of a poor watershed discretization. It was also observed that with no calibration, model performance improved when a more complex scheme with added land cover and then soil information was used.

Results revealed that model performance depends upon the location of the validation sub-basin (i.e., upstream or downstream) with respect to the calibration sub-basins. For both the basin as a whole and the downstream sub-basin, using both land cover and soil data to define GRU types increases model performance. The most complex scheme (16GRU), containing both land cover and soil spatial data, consistently had the highest performance regardless of the calibration time. However, when only land cover data are used, using only the dominant GRU type per cell has similar performance compared with the scheme that has up to seven GRU types per cell. Moreover, it is concluded that the lumped scheme should be avoided for the purpose of PUB, because it constantly performs poorly with the greatest performance variability. For the upstream interior validation sub-basin, adding land cover data and using more GRU types per grid cell largely increases model performance. In contrast, adding soil information to GRU definition improves the model performance only when it is calibrated long enough to account for the increased number of parameters. Model performance is also shown to be lower, in general, at the interior upstream sub-basin compared to the downstream sub-basin.

The idea of applying the methodology by calibrating to a sub-period for increased efficiency is very promising. It can produce consistent discretization assessments if one spends sufficient budget to calibrate the problem in hand. Alternatively, in order to reduce the computation burden of this methodology, one can use other techniques such as in Razavi and Tolson (2014), proposed for efficiently calibrating to a sub-period and yet obtaining same performance as the main period.

Certainly, there are limitations associated with the methodology proposed here. The first and most important is the massive time investment required to complete the experiments due to the computational burden of numerous calibration iterations. This limitation may become prohibitive for large scale studies. Therefore, any strategy that would make the calibration strategy more efficient can greatly benefit this comparison framework. Examples of such strategies that were used in this study include: adopting an efficient optimization algorithm, polishing the watershed discretization, using model preemption during calibration, and calibrating to a sub-period.

The findings of this work are limited to the case study and specific modelling and calibration details used. However, the proposed comparison methodology is general and applied in future studies, using any distributed model, discretization scheme or calibration strategy. These types of quantitative assessments can provide valuable information for modellers in assessing their model discretization choices and improving their predictive performance.

Chapter 6
Conclusions and Suggestions for Future Work

This final chapter provides a summary and conclusions of the research contributions discussed in in Chapters 3 to 5. For each chapter, primary conclusions related to the main research objectives are highlighted first, followed by other secondary findings. Suggestions for related future work are provided at the end.

6.1 Summary and conclusions

One of the primary goals of this research was to build an enhanced MESH modelling system for streamflow predictions over the Great Lakes Basin. This major contribution was achieved in two steps associated with the first two objectives of this thesis. In Chapter 3, MESH performance in predicting streamflows was benchmarked through a formal calibration, for the first time in the Great Lakes Basin. Then in Chapter 4, benchmark results were enhanced by further refining the calibration strategy and adding uncertainty assessment to the MESH streamflow predictions.

Chapter 3 highlights the calibration efforts to benchmark the performance of MESH in the Great Lakes Basin. Two calibration strategies were compared as part of this work. It is observed that a global calibration strategy using multiple sub-basins generates preferred results over an alternate strategy in which one single sub-basin with a single land type is calibrated individually. Also, as expected, the model performance substantially improves after calibration, confirming the essential role of a formal model calibration and validation. The MESH results found from the global calibration strategy in this study can be considered as the benchmark results for MESH performance in the Great Lakes Basin. These results are currently being used in other related studies in this region such as the Great Lakes Runoff Intercomparison Project (GRIP, Fry et al., 2014).

Chapter 4 describes the efforts made to improve upon the benchmark results in Chapter 3, by further refining the calibration strategy. This was mainly achieved by modifying the calibration parameters and selecting an increased number of calibration sub-basins with near-natural flow conditions. A rigorous multi-criteria comparison between the new and the benchmark experiments confirmed that the MESH model performance is indeed better in the revised calibration strategy. The prediction uncertainty bands associated with the parameter estimation were also calculated as another important enhancement in this chapter. Consequently, eight more behavioral parameter sets are identified for MESH, in addition to the best parameter set. A local sensitivity test was also conducted after MESH was calibrated for the Great

Lakes Basin. This sensitivity test showed that the parameters associated with soil properties and channel roughness are the most influential parameters in MESH. As the outcome of this very computationally intensive work, all these parameter sets can be used in other studies aiming, for example, at simulating lake inflows and lake levels for the Great Lakes region. The MESH input files for the best parameter set found in the revised calibration experiment of Chapter 4 can be found in Appendix 2.

Chapter 5 presents another major contribution of this research by introducing a quantitative comparison framework for assessing the validation performance of semi-distributed hydrological models under various watershed discretization schemes. The methodology was demonstrated using the MESH model as applied to the Nottawasaga river basin in Ontario, Canada. In this study, the schemes differed from a simple lumped scheme to more complex ones by adding spatial land cover and then spatial soil information. Results reveal that when constrained by calibration budget, using a more complex scheme did not necessarily lead to improved performance in validation. It is also observed that model performance in ungauged basins depends upon whether the validation sub-basin is located upstream or downstream of the calibration sub-basins. The proposed methodology was also implemented using a shorter sub-period for calibration, aiming at substantial computational saving. This strategy is shown to be promising in producing consistent results and has the potential to increase computational efficiency of this comparison framework. A secondary observation in this study was that the model performance variation due to the stochastic nature of the DDS optimization algorithm can be significant. Therefore, findings from this type of study are more reliable when calibration experiments are performed using multiple replicates.

6.2 General Recommendations for Modelers

As a result of the modelling and calibration efforts conducted in this thesis, which involved overcoming many pitfalls and repeating many experiments, below is a list of general recommendations for modelers:

a) Model settings and data are so important. Model structure, the way it is set up for a case study, and data used in modelling are so important. They can control model simulation and calibration, and potentially resulting in misleading findings. Each of these components can contain error. Prior to performing any calibration, modelers should make every effort to find errors in the code, data, and model structure, and fix them.

b) Calibration is essential. Once a modeler is confident in the model structure, model setup, and data used, he/she should proceed to the calibration process. Performing a formal auto-calibration, particularly for complex models with a large number of parameters, is essential for

obtaining good results. A through calibration should include uncertainty analysis for more effective usage for decision makers. Also, watershed discretization scheme should be part of the calibration process and carefully selected and assessed. A poor watershed discretization can negatively impact the results no matter how advance a model is. In the case of the MESH model, GRUs should be defined carefully and independently using the suitable attributes (e.g., land cover, soil properties) for a given case study.

c) Use a "guided calibration" to improve model performance. Calibration is tricky; you can get the right answer for the wrong reason. Therefore, while calibration is essential, it has to be guided by the modeler to the extent possible. This will lead to more robust and rightful results at the end. Accordingly, a modeler should provide as much information to lead the calibration. For example, a modeler should avoid calibrating independently highly correlated parameters; or should narrow the calibration parameter ranges as much as possible based on available data from literature or measured data; or when possible, should use real measurements for a given parameter (e.g., sand and clay percentages) in place of calibrating it.

6.3 Suggestions for Future Work

The recommendations for future work are presented in two categories: one related to the calibration of MESH over the Great Lakes Basin, and the other related to the methodology used in assessing watershed discretization schemes. With regards to the first category these recommendations are as follows:

d) Calibrate MESH for each of the Great Lakes sub-watersheds individually. Although a well-calibrated MESH model is now built for the entire Great Lakes Basin, it is expected that the individual calibration of the model for each lake can enhance streamflow predictions and can take advantage of measured levels as a secondary calibration target.

e) Revise the GRU definition for MESH in the Great Lakes Basin. Currently, the GRU types used to discretize MESH in the Great Lakes Basin are solely defined based on land cover data. Adding more available spatial data, similar to what was done in Chapter 5 of this thesis, can improve MESH predictions. Given that the soil maps may not be available for the entire basin, and since bedrock geology seemed to have an impact on the results, this revision could be as simple as subdividing the current GRU types based on the presence of the Canadian Shield in the region.

f) Perform a global sensitivity analysis (e.g., the Fourier Amplitude Sensitivity Test (FAST, McRae et al., 1982) of MESH to quantitatively rank the importance of all MESH parameters

with respect to streamflow prediction impacts. This would formalize the suite of individual sensitivity analysis experiments conducted to learn more about how the MESH model works.

g) Ideally, the MESH source code would be rewritten to run in parallel and thus utilize parallel computing resources. Then, substantial time savings would be gained during calibration, particularly for large scale studies such as the Great Lakes Basin.

With regards to the comparison framework for evaluating various discretization schemes, the following recommendations are provided for future studies:

a) Evaluate higher resolution discretization schemes. The current study did not investigate the effect of the grid cell resolution on model performance. Therefore, the comparison framework proposed here for assessing various discretization schemes can be extended to also evaluate the same discretization schemes at a higher resolution.

b) Improved GRU definition: the assessment methodology proposed in this study can be extended to include evaluation of discretization schemes that use additional spatial information such as topography, aspect and depth to bedrock to delineate GRU types in a given basin.

c) Apply the comparison framework proposed in this study to more case studies.

d) Apply the comparison framework to synthetic calibration problems. The application of the framework using synthetic experiments was not as successful as expected in this study. Therefore, further modifications can be made in designing the synthetic framework to improve findings. This would include adding random errors to corrupt the synthetically generated "true" flows.

Appendix

A – MESH input files for the best parameter set from the calibration experiment of Chapter 4

MESH_parameters_CLASS

```
  GLakes (all)                                                                         01comment line 1                                  | 2X,6A4
  Amin H.                                                                              02comment line 2                                  | 2X,6A4
  March 2014                                                                           03comment line 3                                  | 2X,6A4
     44.75    277.08     40.00     40.00     50.00   -1.0    1 4026   07 04DEGLAT/DEGLON/ZRFHGRD/ZRFMGRD/ZBLDGRD  | 5F10.2,F7.1,3I5
   0.000   0.000   1.000   0.000   0.000   0.000   0.000   4.000   0.000 05Land class type/fcanrow/lamxrow      | 9F8.3
   0.000   0.000  -1.966   0.000   0.000   0.000   0.000   0.000   0.000 06lnz0row/lamnrow                      | 9F8.3
   0.000   0.000   0.060   0.000   0.000   0.000   0.000   3.000   0.000 07alvcrow/cmasrow                      | 9F8.3
   0.000   0.000   0.270   0.000   0.000   0.000   0.000   0.374   0.000 08alicrow/rootrow                      | 9F8.3
   0.000   0.000  99.017   0.000           0.000   0.000  30.000   0.000 09rsmnrow/qa50row                      | 4F8.3,8X,4F8.3
   0.000   0.000   0.754   0.000           0.000   0.000   1.000   0.000 10vpdarow/vpdprow                      | 4F8.3,8X,4F8.3
   0.000   0.000 100.000   0.000           0.000   0.000   5.000   0.000 11psgarow/psgbrow                      | 4F8.3,8X,4F8.3
   1.000   1.504   1.000   79.41                                         12drnrow/sdeprow/farerow/ddenrow       | 3F8.3,F8.2
3.9E-002   1.000   0.1009.6E-004       1                                 13xslprow/grkfrow/manrow/WFCIROW/midrow | 4E8.1,I8
      83.6      83.6      83.6                                           14sand                                 | 3F10.1
      12.0      12.0      12.0                                           15clay                                 | 3F10.1
      4.30      4.30      4.30                                           16org                                  | 3F10.1
     20.00     20.00     20.00     20.00      0.00     20.00             17temperature-soil/can/sno/pnd         | 6F10.2
     0.200     0.200     0.200     0.000     0.000     0.000     0.000   18soil moisture-soil/ice/pnd           | 7F10.3
    0.0000    0.0000      0.00     0.000   000.000     1.000             19rcan/scan/sno/albs/rho/gro           | 2F10.4,F10.2,F10.3,F10.4,F10.3
   0.000   0.000   0.000   1.000   0.000   0.000   0.000   0.000   2.625 05Land class type/fcanrow/lamxrow      | 9F8.3
   0.000   0.000   0.000  -2.996   0.000   0.000   0.000   0.000   2.625 06lnz0row/lamnrow                      | 9F8.3
   0.000   0.000   0.000   0.040   0.000   0.000   0.000   0.000   1.875 07alvcrow/cmasrow                      | 9F8.3
   0.000   0.000   0.000   0.273   0.000   0.000   0.000   0.000   0.669 08alirow/rootrow                       | 9F8.3
   0.000   0.000   0.000 109.498           0.000   0.000   0.000  30.000 09rsmnrow/qa50row                      | 4F8.3,8X,4F8.3
   0.000   0.000   0.000   0.630           0.000   0.000   0.000   1.000 10vpdarow/vpdprow                      | 4F8.3,8X,4F8.3
   0.000   0.000   0.000 100.000           0.000   0.000   0.000   5.000 11psgarow/psgbrow                      | 4F8.3,8X,4F8.3
   1.000   1.470   1.000   72.88                                         12drnrow/sdeprow/farerow/ddenrow       | 3F8.3,F8.2
4.6E-002   1.000   0.1009.2E-005       2                                 13xslprow/grkfrow/manrow/WFCIROW/midrow | 4E8.1,I8
       6.5       6.5       6.5                                           14sand                                 | 3F10.1
       2.6       2.6       2.6                                           15clay                                 | 3F10.1
      5.00      5.00      5.00                                           16org                                  | 3F10.1
     20.00     20.00     20.00     20.00      0.00     20.00             17temperature-soil/can/sno/pnd         | 6F10.2
     0.200     0.200     0.200     0.000     0.000     0.000     0.000   18soil moisture-soil/ice/pnd           | 7F10.3
    0.0000    0.0000      0.00     0.000   000.000     1.000             19rcan/scan/sno/albs/rho/gro           | 2F10.4,F10.2,F10.3,F10.4,F10.3
   0.000   1.000   0.000   0.000   0.000   0.000   5.750   0.000   0.000 05Land class type/fcanrow/lamxrow      | 9F8.3
   0.000   0.372   0.000   0.000   0.000   0.000   3.500   0.000   0.000 06lnz0row/lamnrow                      | 9F8.3
   0.000   0.050   0.000   0.000   0.000   0.000  21.500   0.000   0.000 07alvcrow/cmasrow                      | 9F8.3
   0.000   0.283   0.000   0.000   0.000   0.000   2.316   0.000   0.000 08alirow/rootrow                       | 9F8.3
   0.000 106.689   0.000   0.000           0.000  40.000   0.000   0.000 09rsmnrow/qa50row                      | 4F8.3,8X,4F8.3
   0.000   0.662   0.000   0.000           0.000   0.600   0.000   0.000 10vpdarow/vpdprow                      | 4F8.3,8X,4F8.3
   0.000 100.000   0.000   0.000           0.000   5.000   0.000   0.000 11psgarow/psgbrow                      | 4F8.3,8X,4F8.3
   1.000   0.586   1.000   68.15                                         12drnrow/sdeprow/farerow/ddenrow       | 3F8.3,F8.2
1.6E-002   1.000   0.1004.5E-004       3                                 13xslprow/grkfrow/manrow/WFCIROW/midrow | 4E8.1,I8
      48.1      48.1      48.1                                           14sand                                 | 3F10.1
      48.2      48.2      48.2                                           15clay                                 | 3F10.1
      3.70      3.70      3.70                                           16org                                  | 3F10.1
     20.00     20.00     20.00     20.00      0.00     20.00             17temperature-soil/can/sno/pnd         | 6F10.2
     0.200     0.200     0.200     0.000     0.000     0.000     0.000   18soil moisture-soil/ice/pnd           | 7F10.3
    0.0000    0.0000      0.00     0.000   000.000     1.000             19rcan/scan/sno/albs/rho/gro           | 2F10.4,F10.2,F10.3,F10.4,F10.3
   1.000   0.000   0.000   0.000   0.000   2.000   0.000   0.000   0.000 05Land class type/fcanrow/lamxrow      | 9F8.3
   0.223   0.000   0.000   0.000   0.000   1.050   0.000   0.000   0.000 06lnz0row/lamnrow                      | 9F8.3
   0.030   0.000   0.000   0.000   0.000  20.000   0.000   0.000   0.000 07alvcrow/cmasrow                      | 9F8.3
   0.190   0.000   0.000   0.000   0.000   2.730   0.000   0.000   0.000 08alirow/rootrow                       | 9F8.3
 185.967   0.000   0.000   0.000          30.000   0.000   0.000   0.000 09rsmnrow/qa50row                      | 4F8.3,8X,4F8.3
   0.944   0.000   0.000   0.000           1.050   0.000   0.000   0.000 10vpdarow/vpdprow                      | 4F8.3,8X,4F8.3
 100.000   0.000   0.000   0.000           5.000   0.000   0.000   0.000 11psgarow/psgbrow                      | 4F8.3,8X,4F8.3
   1.000   4.095   1.000   78.77                                         12drnrow/sdeprow/farerow/ddenrow       | 3F8.3,F8.2
1.0E-002   1.000   0.1006.8E-004       4                                 13xslprow/grkfrow/manrow/WFCIROW/midrow | 4E8.1,I8
      57.8      57.8      57.8                                           14sand                                 | 3F10.1
       0.2       0.2       0.2                                           15clay                                 | 3F10.1
      5.00      5.00      5.00                                           16org                                  | 3F10.1
     20.00     20.00     20.00     20.00      0.00     20.00             17temperature-soil/can/sno/pnd         | 6F10.2
     0.200     0.200     0.200     0.000     0.000     0.000     0.000   18soil moisture-soil/ice/pnd           | 7F10.3
    0.0000    0.0000      0.00     0.000   000.000     1.000             19rcan/scan/sno/albs/rho/gro           | 2F10.4,F10.2,F10.3,F10.4,F10.3
   0.000   1.000   0.000   0.000   0.000   0.000   4.500   0.000   0.000 05Land class type/fcanrow/lamxrow      | 9F8.3
   0.000   0.324   0.000   0.000   0.000   0.000   2.683   0.000   0.000 06lnz0row/lamnrow                      | 9F8.3
   0.000   0.042   0.000   0.000   0.000   0.000  21.000   0.000   0.000 07alvcrow/cmasrow                      | 9F8.3
   0.000   0.252   0.000   0.000   0.000   0.000   2.955   0.000   0.000 08alirow/rootrow                       | 9F8.3
   0.000 148.055   0.000   0.000           0.000  35.000   0.000   0.000 09rsmnrow/qa50row                      | 4F8.3,8X,4F8.3
   0.000   0.783   0.000   0.000           0.000   0.825   0.000   0.000 10vpdarow/vpdprow                      | 4F8.3,8X,4F8.3
   0.000 100.000   0.000   0.000           0.000   5.000   0.000   0.000 11psgarow/psgbrow                      | 4F8.3,8X,4F8.3
   1.000   0.440   1.000   27.76                                         12drnrow/sdeprow/farerow/ddenrow       | 3F8.3,F8.2
8.4E-002   1.000   0.1002.1E-004       5                                 13xslprow/grkfrow/manrow/WFCIROW/midrow | 4E8.1,I8
      20.0      20.0      20.0                                           14sand                                 | 3F10.1
      75.8      75.8      75.8                                           15clay                                 | 3F10.1
      4.20      4.20      4.20                                           16org                                  | 3F10.1
     20.00     20.00     20.00     20.00      0.00     20.00             17temperature-soil/can/sno/pnd         | 6F10.2
     0.200     0.200     0.200     0.000     0.000     0.000     0.000   18soil moisture-soil/ice/pnd           | 7F10.3
    0.0000    0.0000      0.00     0.000   000.000     1.000             19rcan/scan/sno/albs/rho/gro           | 2F10.4,F10.2,F10.3,F10.4,F10.3
   0.000   0.000   0.000   0.000   0.000   0.000   0.000   4.500   0.000 05Land class type/fcanrow/lamxrow      | 9F8.3
   0.000   0.000   0.000  -2.300   0.000   0.000   0.000   0.000   0.000 06lnz0row/lamnrow                      | 9F8.3
   0.000   0.000   0.000   0.060   0.000   0.000   0.000   2.550   0.000 07alvcrow/cmasrow                      | 9F8.3
   0.000   0.000   0.000   0.335   0.000   0.000   0.000   1.090   0.000 08alirow/rootrow                       | 9F8.3
   0.000   0.000   0.000 120.000           0.000   0.000  30.000   0.000 09rsmnrow/qa50row                      | 4F8.3,8X,4F8.3
   0.000   0.000   0.000   0.500           0.000   0.000   1.000   0.000 10vpdarow/vpdprow                      | 4F8.3,8X,4F8.3
   0.000   0.000   0.000  33.900           0.000   0.000   7.490   0.000 11psgarow/psgbrow                      | 4F8.3,8X,4F8.3
   1.000   4.100   1.000  2.5000                                         12drnrow/sdeprow/farerow/ddenrow       | 3F8.3,F8.2
  0.0600   1.000  3.0E-2  0.3E-1       6                                 13xslprow/grkfrow/manrow/WFCIROW/midrow | 4E8.1,I8
     65.00     65.00     65.00                                           14sand                                 | 3F10.1
     30.00     30.00     30.00                                           15clay                                 | 3F10.1
      5.00      5.00      5.00                                           16org                                  | 3F10.1
     20.00     20.00     20.00     20.00      0.00     20.00             17temperature-soil/can/sno/pnd         | 6F10.2
     0.200     0.200     0.200     0.000     0.000     0.000     0.000   18soil moisture-soil/ice/pnd           | 7F10.3
    0.0000    0.0000      0.00     0.000   000.000     1.000             19rcan/scan/sno/albs/rho/gro           | 2F10.4,F10.2,F10.3,F10.4,F10.3
   0.000   0.000   0.000   0.000   1.000   0.000   0.000   0.000   0.000 05Land class type/fcanrow/lamxrow      | 9F8.3
   0.000   0.000   0.000   0.000   0.300   0.000   0.000   0.000   0.000 06lnz0row/lamnrow                      | 9F8.3
   0.000   0.000   0.000   0.000   0.090   0.000   0.000   0.000   0.000 07alvcrow/cmasrow                      | 9F8.3
   0.000   0.000   0.000   0.000   0.150   0.000   0.000   0.000   0.000 08alirow/rootrow                       | 9F8.3
   0.000   0.000   0.000   0.000           0.000   0.000   0.000   0.000 09rsmnrow/qa50row                      | 4F8.3,8X,4F8.3
   0.000   0.000   0.000   0.000           0.000   0.000   0.000   0.000 10vpdarow/vpdprow                      | 4F8.3,8X,4F8.3
   0.000   0.000   0.000   0.000           0.000   0.000   0.000   0.000 11psgarow/psgbrow                      | 4F8.3,8X,4F8.3
   1.000   4.100   1.000  2.5000                                         12drnrow/sdeprow/farerow/ddenrow       | 3F8.3,F8.2
  0.0600   1.000  3.0E-2  0.3E-1       7                                 13xslprow/grkfrow/manrow/WFCIROW/midrow | 4E8.1,I8
     65.00     65.00     65.00                                           14sand                                 | 3F10.1
     30.00     30.00     30.00                                           15clay                                 | 3F10.1
      5.00      5.00      5.00                                           16org                                  | 3F10.1
     20.00     20.00     20.00     20.00      0.00     20.00             17temperature-soil/can/sno/pnd         | 6F10.2
     0.200     0.200     0.200     0.000     0.000     0.000     0.000   18soil moisture-soil/ice/pnd           | 7F10.3
    0.0000    0.0000      0.00     0.000   000.000     1.000             19rcan/scan/sno/albs/rho/gro           | 2F10.4,F10.2,F10.3,F10.4,F10.3
       135       318       135       318         0         0             20                                     | *
      1007      2107      1007      2107         0         0             21                                     | *
         1         0       153      2004                                 22IHOUR/IMIN/IDAY/IYEAR                | 4I10
123456789*123456789*123456789*123456789*123456789*123456789*123456789*123  ver 30d
```

MESH_parameters_hydrology

```
1.1.a04:MESH Hydrology Parameters input file                                    01version number and comment line 1
##### Option Flags #####                                                       02comment line 2
----#                                                                          03comment line 3
    2                                                                          04Number of option flags
    0                                                                          05[reserved]
    0                                                                          06[reserved]
##### River roughness factor (WF_R2) (5 classes maximum) ##                   07comment line 7
-----#-----#-----#-----#-----#                                                 08comment line 8
 1.862 1.696 0.127 0.629 0.360                                                 09River roughness factor
##### GRU class independent hydrologic parameters #####                       10comment line 10
-------#                                                                       11comment line 11
       5                                                                       12Number of GRU independent hydrologic parameter
   0.800                                                                       13Maximum soil porosity
   4.100                                                                       14Depth from surface to bottom of rooting zone [
   1.000                                                                       15Surface saturations [0.75 - 1.0]
 -10.000                                                                       16Overnight minimum to cause ice lens after majc
 500.000                                                                       17Calculated opportunity time for 1st year [hour
     ##### GRU class dependent hydrologic parameters #####                     18comment line 18
-------#                                                                       19comment line 19
       7                                                                       20Number of GRUs (must match number in mesh_para
       9                                                                       21Number of GRU dependent hydrologic parameters
---------#---------#---------#---------#---------#---------#---------#---------# 22comment line 22
      0.70      0.70      0.97      0.73      0.52      0.10      0.10        23ZSNLROW - limiting snow depth below which cove
      0.08      0.14      0.14      0.07      0.02      0.10      0.10        24ZPLSROW - maximum water ponding depth for snow
      0.12      0.13      0.13      0.04      0.10      0.10      0.10        25ZPLGROW - maximum water ponding depth for snow
      2.10      2.10      2.10      2.10      2.10      2.10      2.10        26FZRCROW - coefficient for the frozen soil infi
      4.00      4.00      4.00      4.00      4.00      4.00      4.00        27BCROW   - Shape factor parameter (PDMROF)
      0.00      0.00      0.00      0.00      0.00      0.00      0.00        28CMINROW - Minimum storage capacity [m] paramet
      1.10      1.10      1.10      1.10      1.10      1.10      1.10        29CMAXROW - Maximum storage [m] parameter (PDMRC
      0.00      0.00      0.00      0.00      0.00      0.00      0.00        30K1ROW - Time constant for the first linear res
      0.00      0.00      0.00      0.00      0.00      0.00      0.00        31K2ROW - Time constant for the second linear re
```

MESH_input_soil_levels

```
      0.10    0.10
      0.25    0.35
      3.75    4.10
```

MESH_input_run_options

```
MESH input run options file                    01comment line 1                | *
##### Control Flags #####                      02comment line 2                | *
----#                                          03comment line 3                | *
   17                                          04Number of control flags       | I5
HOURLYFLAG            60                       05hourly flag                   | A20, I4
BASINRAINFLAG          1                       06basin rain flag               | A20, I4
BASINLONGWAVEFLAG      1                       07basin longwave flag           | A20, I4
BASINSHORTWAVEFLAG     1                       08basin shortwave flag          | A20, I4
BASINTEMPERATUREFLAG   1                       09basin temperature flag        | A20, I4
BASINWINDFLAG          1                       10basin wind flag               | A20, I4
BASINPRESFLAG          1                       11basin pressure flag           | A20, I4
BASINHUMIDITYFLAG      1                       12basin humidity flag           | A20, I4
SHDFILEFLAG            1                       13basin shed file flag          | A20, I4
SAVERESUMEFLAG         0
R2COUTPUTFLAG          0
SUBBASINFLAG           1
RESUMEFLAG             1
AUTOCALIBRATIONFLAG    1
PREEMPTIONFLAG         1
OBJFNFLAG              3
SOILINIFLAG            4
##### Output Grid selection #####          15comment line 15                | *
----#                                      16comment line 16                | *
    0    #Maximum 5 points                 17Number of output grid points  | I5
---------#---------#---------#---------#---------# 18comment line 18         | *
       019       037       020       037       957 19Grid number            | 5I10
         1         1         3         3         4 20Land class             | 5I10
CLASSOUT1                                  21Output directory               | 5A10
##### Output Directory #####               22comment line 22                | *
---------#                                  23comment line 23                | *
BASINAVG1                                  24Output Directory for total-basin files | A10
##### Simulation Run Times #####           25comment line 25                | *
---#---#---#---#                           26comment line 26                | *
2004 153  01   0                           27Start year, day, hour, minute  | 4I4
2005 273  24   0                           28Stop year, day, hour, minute   | 4I4
```

B – Supplemental Results for Chapter 5

In order to further investigate the findings of the comparison framework presented in Chapter 5, for assessing watershed discretization schemes, the same framework was applied using synthetic calibration experiments and the results are briefly discussed here. The motivation behind this synthetic analysis was to perform a cleaner comparison between various schemes where the model and the data uncertainties are eliminated. This fact allows for a better evaluation of the effect of various discretization schemes and calibration budgets on model performance.

Details of the synthetic experiments are all identical to the real case study, except that a synthetic "true" flow was generated based on the most complex scheme (16-GRU) and replaced the measured flow for calibrations. In other words, all discretization schemes, including the 16-GRU scheme itself, were calibrated against the "true' flow.

Results for these synthetic experiments are shown in Figure A1 for the overall calibration and validation model performance (Figure A1a and A1b), and the individual validation model performance at each individual sub-basin (Figure A1c and A1d). Overall, consistent results can be seen between this synthetic experiment and the real experiment presented in Chapter 5. As seen in sub-basin V2 (Figure A1d), validation model performance does not necessarily improve with increased discretization complexity and the calibration budget is an important factor. This behavior is even more pronounced here, for the 16-GRU scheme, compared with the real experiment. Surprisingly, the median model performance of the 7-GRU scheme constantly decreases with a longer calibration budget. This behavior could be due to the stochastic nature of DDS algorithm used in calibration. More experiments are needed to further investigate these observations.

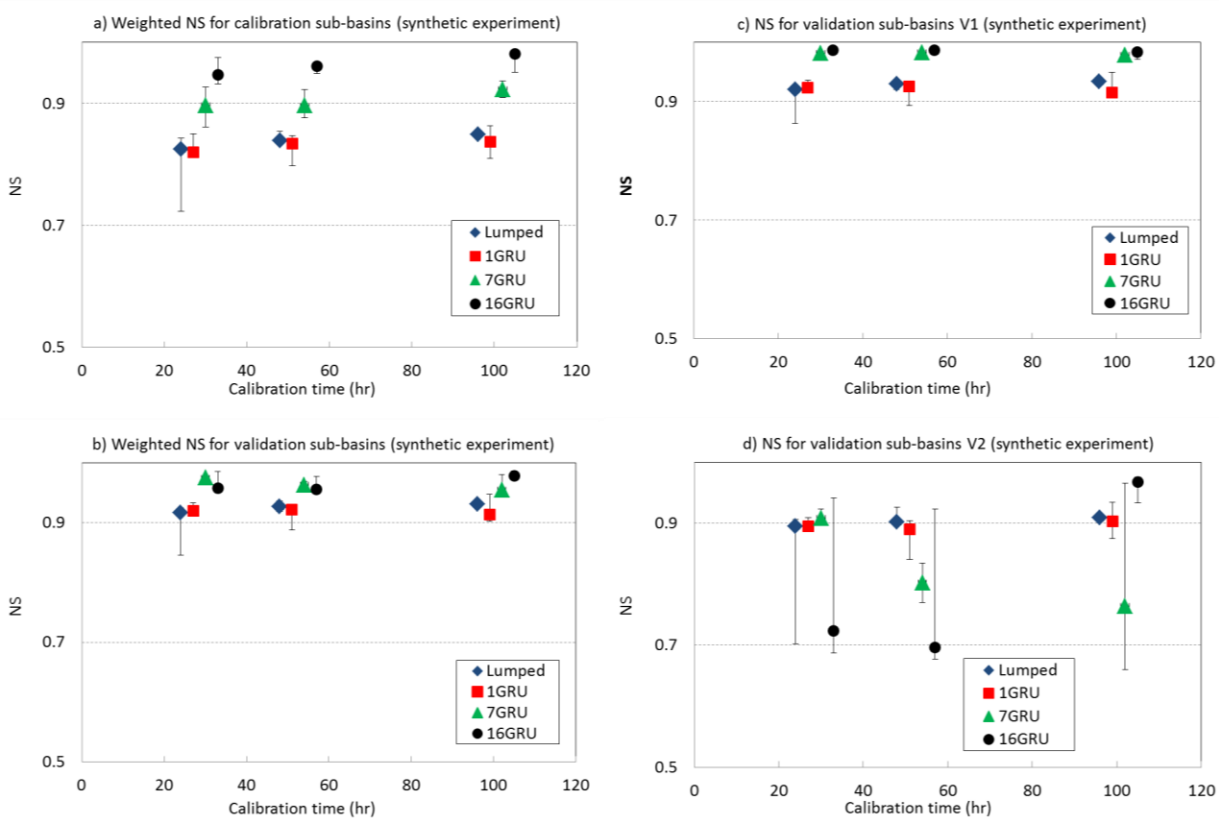

Figure 28. Synthetic experiment results for the overall a) calibration and b) validation model performance, and the individual validation model performance at sub-basin a) V1, and b) V2, after calibration over the main period (October 2004 to September 2007), for three calibration budgets of 24, 48 and 96 hours. Symbols indicate the median performance for the three calibration replicates. Minimum and maximum values are displayed with bars. For better display purposes, lateral positions are shifted by 3 hours and minimum y-axis was fixed at 0.5.

References

Abbaspour, K. C., Yang, J., Maximov, I., Siber, R., Bogner, K., Mieleitner, J., et al. (2007). Modelling hydrology and water quality in the pre-alpine/alpine thur watershed using SWAT. Journal of Hydrology, 333(2-4), 413-430.

Abbaspour, K., Johnson, C., & van Genuchten, M. (2004). Estimating uncertain flow and transport parameters using a sequential uncertainty fitting procedure. Vadose Zone Journal, 3(4), 1340-1352.

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

Ajami, N., Gupta, H., Wagener, T., & Sorooshian, S. (2004). Calibration of a semi-distributed hydrologic model for streamflow estimation along a river system. Journal of Hydrology, 298(1-4), 112-135.

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

Arnold, J. G., Srinivasan, R., Muttiah, R. S., & Williams, J. R. (1998). Large area hydrologic modeling and assessment part i: Model development. JAWRA Journal of the American Water Resources Association, 34(1), 73-89.

Bandaragoda, C., Tarboton, D. G., & Woods, R. (2004). Application of TOPNET in the distributed model intercomparison project. Journal of Hydrology, 298(1-4), 178-201.

Benoit, R., Pellerin, P., Kouwen, N., Ritchie, H., Donaldson, N., Joe, P., et al. (2000). Toward the use of coupled atmospheric and hydrologic models at regional scale. Monthly Weather Review, 128(6), 1681-1706.

Beven, K., & Binley, A. (1992). The future of distributed models - model calibration and uncertainty prediction rid F-8707-2011.Hydrological Processes, 6(3), 279-298.

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

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

Boyle, D., Gupta, H., Sorooshian, S., Koren, V., Zhang, Z., & Smith, M. (2001). Toward improved streamflow forecasts: Value of semidistributed modeling. Water Resources Research, 37(11), 2749-2759.

Branger, F., Braud, I., Debionne, S., Viallet, P., Dehotin, J., Henine, H., et al. (2010). Towards multi-scale integrated hydrological models using the LIQUID® framework. overview of the concepts and first application examples. Environmental Modelling & Software,25(12), 1672-1681.

Brazier, R., Beven, K., Freer, J., & Rowan, J. (2000). Equifinality and uncertainty in physically based soil erosion models: Application of the glue methodology to WEPP-the water erosion prediction project-for sites in the UK and USA RID F-8707-2011 RID C-7335-2009 RID F-6539-2011. Earth Surface Processes and Landforms, 25(8), 825-845.

Burnash, R. J. C., Ferral, R. L., & McGuire, R. A. (1973). A generalized streamflow simulation system: Conceptual modeling for digital computers No. U.S. Dep. of Commerce, Natl. Weather Serv. and State of Calif., Dep. of Water Resour., Sacramento, Calif.)

Cheng, C. T., Wu, X. Y., & Chau, K. W. (2005). Multiple criteria rainfall-runoff model calibration using a parallel genetic algorithm in a cluster of computers. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, 50(6), 1069-1087.

Clarke, R. T. (1973). A review of some mathematical models used in hydrology, with observations on their calibration and use. Journal of Hydrology, 19(1), 1-20.

Craig, J. R., & Snowdon, A. P. (2011). Raven users and Developer's manual v0.9 No. 2011, Draft Technical Report, Dept. of Civil & Environmental Engineering, University of Waterloo.

Davison, B., Pohl, S., Domes, P., Marsh, P., Pietroniro, A., & MacKay, M. (2006). Characterizing snowmelt variability in a land-surface-hydrologic model. Atmosphere-Ocean, 44(3), 271-287.

Deacu, D., Fortin, V., Klyszejko, E., Spence, C., & Blanken, P. D. (2012). Predicting the net basin supply to the great lakes with a hydrometeorological model. Journal of Hydrometeorology, 13(6), 1739-1759.

Dehotin, J., & Braud, I. (2008). Which spatial discretization for distributed hydrological models? proposition of a methodology and illustration for medium to large-scale catchments. Hydrology and Earth System Sciences (HESS), 12(3), 769-796.

Donald, J. R., Soulis, E. D., Kouwen, N., & Pietroniro, A. (1995). A land cover-based snow cover representation for distributed hydrologic models. Water Resources Research, 31(4), 995-1009.

Dorner, S. M., Anderson, W. B., Slawson, R. M., Kouwen, N., & Huck, P. M. (2006). Hydrologic modeling of pathogen fate and transport. Environmental Science & Technology, 40(15), 4746-4753.

Dornes, P. F., Tolson, B. A., Davison, B., Pietroniro, A., Pomeroy, J. W., & Marsh, P. (2008). Regionalisation of land surface hydrological model parameters in subarctic and arctic environments. Physics and Chemistry of the Earth, 33(17-18), 1081-1089.

Duan, Q., Gupta, V., & Sorooshian, S. (1993). Shuffled complex evolution approach for effective and efficient global minimization. Journal of Optimization Theory and Applications, 76(3), 501-521.

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

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

Dupont, F., Chittibabu, P., Fortin, V., Rao, Y. R., & Lu, Y. (2012). Assessment of a NEMO-based hydrodynamic modelling system for the great lakes. Water Quality Research Journal of Canada, 47(3-4), 198-214.

Falcone, J. (2011). GAGES-II: Geospatial attributes of gages for evaluating streamflow. Reston, Virginia: U.S. Geological Survey.

Flügel, W. (1995). Delineating hydrological response units by geographical information system analyses for regional hydrological modelling using PRMS/MMS in the drainage basin of the river bröl, germany. Hydrological Processes, 9(3-4), 423-436.

Fry, L. M., Gronewold, A. D., Fortin, V., Buan, S., Clites, A. H., Luukkonen, C., et al,2014. The Great Lakes Runoff Intercomparison Project Phase 1: Lake Michigan (GRIP-M). Journal of Hydrology ;519, Part D(0):3448-65.

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

Gray, G. A., & Kolda, T. G. (2006). Algorithm 856: APPSPACK 4.0: Asynchronous parallel pattern search for derivative-free optimization. Acm Transactions on Mathematical Software, 32(3), 485-507.

Gronewold, A. D., Clites, A. H., Hunter, T. S., & Stow, C. A. (2011). An appraisal of the great lakes advanced hydrologic prediction system. Journal of Great Lakes Research, 37(3), 577-583.

Gronewold, A. D., & Stow, C. A. (2014). Water loss from the great lakes. Science, 343(6175), 1084-1085.

Gronewold, A. D., & Fortin, V. (2012). Advancing great lakes hydrological science through targeted binational collaborative research. Bulletin of the American Meteorological Society, 93(12), 1921-1925.

Gupta, H., Sorooshian, S., & Yapo, P. (1998). Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. Water Resources Research, 34(4), 751-763.

Haghnegahdar, A., Tolson, B. A., Seglenieks, F. R., Klyszejko, E., & Fortin, V. (2010). Comparing the calibrated performance of the WATFLOOD and MESH distributed hydrological models in the great lakes basin. Water 2010: Hydrology, hydraulics and water resources in an uncertain environment, Quebec city, Quebec, Canada, July 5-7.

Haghnegahdar, A., Tolson, B. A., Davison, B., Seglenieks, F. R., Klyszejko, E., Soulis, E. D., et al. (2014). Calibrating environment Canada's MESH modelling system over the Great Lakes Basin. Atmosphere-Ocean, 1-13.

Haverkamp, S., Srinivasan, R., Frede, H. G., & Santhi, C. (2002). Subwatershed spatial analysis tool: Discretization of a distributed hydrologic model by statistical criteria1. JAWRA Journal of the American Water Resources Association, 38(6), 1723-1733.

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

Islam, M. S., Oki, T., Kanae, S., Hanasaki, N., Agata, Y., & Yoshimura, K. (2007). A grid-based assessment of global water scarcity including virtual water trading. Water Resources Management, 21(1), 19-33.

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.

Kampf, S. K., & Burges, S. J. (2007). A framework for classifying and comparing distributed hillslope and catchment hydrologic models.Water Resources Research, 43(5), W05423.

Khu, S., & Werner, M. (2003). Reduction of Monte-carlo simulation runs for uncertainty estimation in hydrological modelling RID C-8144-2009. Hydrology and Earth System Sciences, 7(5), 680-692.

Kouwen, N., Soulis, E. D., Pietroniro, A., Donald, J., & Harrington, R. A. (1993). Grouped response units for distributed hydrologic modeling. Journal of Water Resources Planning and Management-Asce, 119(3), 289-305.

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

Kuczera, G., & Parent, E. (1998). Monte carlo assessment of parameter uncertainty in conceptual catchment models: The metropolis algorithm. Journal of Hydrology, 211(1-4), 69-85.

Kuczera, G., Kavetski, D., Renard, B., & Thyer, M. (2010). A limited-memory acceleration strategy for MCMC sampling in hierarchical Bayesian calibration of hydrological models. Water Resources Research, 46, W07602.

Kuzmin, V., Seo, D., & Koren, V. (2008). Fast and efficient optimization of hydrologic model parameters using a priori estimates and stepwise line search. Journal of Hydrology, 353(1-2), 109-128.

Leavesley, G. H., Lichty, B. M., Troutman, L. G., & Saindon, L. G. (1983). Precipitation-runoff modeling system: User's manual No. USGS Water-Resources Investigations Rep. 83-4238, USGS, Denver.

MacLean, A. J. (2009). Calibration and analysis of the MESH hydrological model applied to cold regions. MSc thesis. University of Waterloo.

Madsen, H. (2003). Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. Advances in Water Resources, 26(2), 205-216.

Mahfouf, J., Brasnett, B., & Gagnon, S. (2007). A Canadian precipitation analysis (CaPA) project: Description and preliminary results. Atmosphere-Ocean, 45(1), 1-17.

Mailhot, J., Bélair, S., Lefaivre, L., Bilodeau, B., Desgagné, M., Girard, C., et al. (2006). The 15- km version of the canadian regional forecast system. Atmosphere-Ocean, 44(2), 133-149.

Matott, L. S. (2005). An optimization software tool: Documentation and user's guide. State Univ. of N. Y. at buffalo, buffalo.

McCabe, G. J., & Wolock, D. M. (2011). Century-scale variability in global annual runoff examined using a water balance model. International Journal of Climatology, 31(12), 1739-1748.

McRae, G. J., Tilden, J. W., & Seinfeld, J. H. (1982). Global sensitivity analysis—a computational implementation of the Fourier amplitude sensitivity test (FAST). Computers & Chemical Engineering, 6(1), 15-25.

Mekonnen, M. A., Wheater, H. S., Ireson, A. M., Spence, C., Davison, B., & Pietroniro, A. (2014). Towards an improved land surface scheme for prairie landscapes. Journal of Hydrology, 511(0), 105-116.

Millerd, F. (2005). The economic impact of climate change on Canadian commercial navigation on the great lakes. Canadian Water Resources Journal, 30(4), 269-280.

Moriasi, D. N., Wilson, B. N., Douglas-Mankin, K. R., Arnold, J. G., & Gowda, P. H. (2012). Hydrologic and water quality models: Use, calibration, and validation. Transactions of the ASABE, 55(4), 1241-1247.

Mugunthan, P., & Shoemaker, C. A. (2006). Assessing the impacts of parameter uncertainty for computationally expensive groundwater models. Water Resources Research, 42(10), W10428.

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, B. P., & Nicholas, J. R. (2005). Uncertainty in the great lakes water balance No. U.S. Geological Survey Scientific Investigations Report 2004-5100). Reston, Va.: U.S. Dept. of the Interior, U.S. Geological Survey.

Neitsch, S. L., Arnold, J. G., Kiniry, J. R., & Williams, J. R. (2001). Soil and water assessment tool user's manual version 2000 No. U.S. Dept. of Agric. Agric. Res. Serv., Temple, Tex.

Neitsch, S. L., Arnold, J. G., Kiniry, J. R., Srinivasan, R., & Wiliams J.R. (2005). Soil and water assessment tool user's manual version 2000. No. GSWRL Report 02-02, BRC Report 02-06, TR-192. College Station, TX: Texas Water Resources Institute.

Ostfeld, A., & Salomons, S. (2005). A hybrid genetic-instance based learning algorithm for CE-QUAL-W2 calibration. Journal of Hydrology, 310(1-4), 122-142.

Petrucci, G., & Bonhomme, C. (2014). The dilemma of spatial representation for urban hydrology semi-distributed modelling: Trade-offs among complexity, calibration and geographical data. Journal of Hydrology, 517, 997-1007.

Pietroniro, A., Fortin, V., Kouwen, N., Neal, C., Turcotte, R., Davison, B., et al. (2007). Development of the MESH modelling system for hydrological ensemble forecasting of the Laurentian Great Lakes at the regional scale. Hydrology and Earth System Sciences, 11(4), 1279-1294.

Pietroniro, A., Prowse, T., Hamlin, L., Kouwen, N., & Soulis, R. (1996). Application of a grouped response unit hydrological model to a northern wetland region. Hydrological Processes, 10(10), 1245-1261.

Pietroniro, A., & Soulis, E. (2003). A hydrology modelling framework for the Mackenzie GEWEX programme. Hydrological Processes, 17(3), 673-676.

Pitman, A. J. (2003). The evolution of, and revolution in, land surface schemes designed for climate models. International Journal of Climatology, 23(5), 479-510.

Pokhrel, P., Gupta, H. V., & Wagener, T. (2008). A spatial regularization approach to parameter estimation for a distributed watershed model. Water Resources Research, 44(12), W12419.

Pokhrel, P., & Gupta, H. V. (2010). On the use of spatial regularization strategies to improve calibration of distributed watershed models. Water Resources Research, 46

Razavi, S., & Tolson, B. A. (2013). An efficient framework for hydrologic model calibration on long data periods. Water Resources Research, 49(12), 8418-8431.

Razavi, S., Tolson, B. A., & Burn, D. H. (2012). Review of surrogate modeling in water resources. Water Resources Research, 48(7), W07401.

Razavi, S., Tolson, B. A., Matott, L. S., Thomson, N. R., MacLean, A., & Seglenieks, F. R. (2010). Reducing the computational cost of automatic calibration through model preemption. Water Resources Research, 46(11), W11523.

Reggiani, P., Sivapalan, M., & Hassanizadeh, S. M. (1998). A unifying framework for watershed thermodynamics: Balance equations for mass, momentum, energy and entropy, and the second law of thermodynamics. Advances in Water Resources, 22(4), 367-398.

Roberts, J., Pryse-Phillips, A., & Snelgrove, K. (2012). Modeling the potential impacts of climate change on a small watershed in labrador, canada. Canadian Water Resources Journal / Revue Canadienne Des Ressources Hydriques, 37(3), 231-251.

Rouhani, H., Willems, P., & Feyen, J. (2009). Effect of watershed delineation and areal rainfall distribution on runoff prediction using the SWAT model. Hydrology Research, 40(6), 505-519.

Samaniego, L., Kumar, R., & Attinger, S. (2010). Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. Water Resources Research, 46(5), W05523.

Schutte, J., Reinbolt, J., Fregly, B., Haftka, R., & George, A. (2004). Parallel global optimization with the particle swarm algorithm. International Journal for Numerical Methods in Engineering, 61(13), 2296-2315.

Seglenieks, F. (2009). Creation of a gridded time series of hydrological variables for Canada. PhD dissertation. University of Waterloo.

Sellers, P. J., Dickinson, R. E., Randall, D. A., Betts, A. K., Hall, F. G., Berry, J. A., et al. (1997). Modeling the exchanges of energy, water, and carbon between continents and the atmosphere. Science, 275(5299), 502-509.

Sellers, P. J., Mintz, Y., Sud, Y. C., & Dalcher, A. (1986). A simple biosphere model (SIB) for use within general circulation models. Journal of the Atmospheric Sciences, 43(6), 505-531.

Sheets, R.A., Dumouchelle, D.H., Feinstein, D.T., 2005. Ground-Water Modeling of Pumping Effects near Regional Ground-Water Divides and River/Aquifer Systems— Results and Implications of Numerical Experiments, USGS, Scientific Investigations Report 2005–5141.

Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshmi, V., et al. (2003). IAHS decade on predictions in ungauged basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences. Hydrological Sciences Journal, 48(6), 857-880.

Sorooshian, S., Duan, Q., & Gupta, V. K. (1993). Calibration of rainfall-runoff models: Application of global optimization to the Sacramento soil moisture accounting model. Water Resources Research, 29(4), 1185-1194.

Soulis, E., Snelgrove, K., Kouwen, N., Seglenieks, F., & Verseghy, D. (2000). Towards closing the vertical water balance in Canadian atmospheric models: Coupling of the land surface scheme CLASS with the distributed hydrological model WATFLOOD. Atmosphere-Ocean, 38(1), 251-269.

Srinivasan, M., Gerard-Marchant, P., Veith, T., Gburek, W., & Steenhuis, T. (2005). Watershed scale modeling of critical source areas of runoff generation and phosphorus transport. Journal of the American Water Resources Association, 41(2), 361-375.

Stadnyk, T. A., Delavau, C., Kouwen, N., & Edwards, T. W. D. (2013). Towards hydrological model calibration and validation: Simulation of stable water isotopes using the isoWATFLOOD model. Hydrological Processes, 27(25), 3791-3810.

Tang, Y., Reed, P., & Wagener, T. (2006). How effective and efficient are multiobjective evolutionary algorithms at hydrologic model calibration? Hydrol. Earth Syst.Sci., 10(2), 289-307.

Tolson, B. A., & Shoemaker, C. A. (2007). Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. Water Resources Research, 43(1), W01413 1-16.

Tolson, B. A., & Shoemaker, C. A. (2008). Efficient prediction uncertainty approximation in the calibration of environmental simulation models. Water Resources Research, 44(4), W04411.

Tonkin, M. J., & Doherty, J. (2005). A hybrid regularized inversion methodology for highly parameterized environmental models. Water Resources Research, 41(10), W10412.

Verseghy, D. (2011). CLASS – THE CANADIAN LAND SURFACE SCHEME (VERSION 3.5), technical documentation (version 1) Environment Canada.

Verseghy, D. L. (1991). Class-a Canadian land surface scheme for GCMs .1. soil model. International Journal of Climatology, 11(2), 111-133.

Verseghy, D. L., McFarlane, N. A., & Lazare, M. (1993). Class - a Canadian land-surface scheme for GCMs.2. vegetation model and coupled runs. International Journal of Climatology, 13(4), 347-370.

Vrugt, J., Gupta, H., Bouten, W., & Sorooshian, S. (2003). A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. Water Resources Research, 39(8), 1201.

Vrugt, J. A., Ó Nualláin, B., Robinson, B. A., Bouten, W., Dekker, S. C., & Sloot, P. M. A. (2006). Application of parallel computing to stochastic parameter estimation in environmental models. Computers & Geosciences, 32(8), 1139-1155.

Wang, Q. J. (1991). The genetic algorithm and its application to calibrating conceptual rainfall-runoff models. Water Resources Research,27(9), 2467-2471.

Wilson, C. P. V. (1982). Groundwater hydrology (2nd edn) by D. K. Todd. Wiley, New York, 1980. 552 pp. ISBN 0 471 08641 X. Geological Journal, 17(4), 345-345.

Winter, T. C. (2001). The concept of hydrologic landscapes. JAWRA Journal of the American Water Resources Association, 37(2), 335-349.

Wood, E. F., Sivapalan, M., Beven, K., & Band, L. (1988). Effects of spatial variability and scale with implications to hydrologic modeling. Journal of Hydrology, 102(1-4), 29-47.

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

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

Yatheendradas, S., Wagener, T., Gupta, H., Unkrich, C., Goodrich, D., Schaffner, M., et al. (2008). Understanding uncertainty in distributed flash flood forecasting for semiarid regions. Water Resources Research, 44(5), W05S19.

Zhang, X., Srinivasan, R., & Van Liew, M. (2009). Approximating SWAT model using artificial neural network and support vector machine rid B-7907-2009 rid D-3937-2009. Journal of the American Water Resources Association, 45(2), 460-474.