Representation of Spatial Heterogeneity of the Baker Creek Watershed in the Land-Surface Hydrology Model, MESH

Haley Brauner

Project Partner- Bruce Davison from Environment and Climate Change Canada [(ECCC)](https://www.canada.ca/en/environment-climate-change.html)

Faculty Advisor- [Dr. Andrew Ireson](https://sens.usask.ca/people/faculty/core-faculty/ireson-andrew.php#AcademicCredentials) School of Environment and Sustainability, University of Saskatchewan

May-August, 2019

A project report submitted in partial fulfilment for the Masters in Water Security degree

Table of Contents

[Acknowledgements 4](#_Toc20298144)

[Executive Summary 5](#_Toc20298145)

[1. Introduction 7](#_Toc20298146)

[Recent Research using the MESH Model 8](#_Toc20298147)

[Represenatation of Spatial Variability 9](#_Toc20298148)

[Objectives 9](#_Toc20298149)

[Summary of the Mkandla Project 9](#_Toc20298150)

[2. Site Description 12](#_Toc20298151)

[Climate 12](#_Toc20298152)

[Land Use 12](#_Toc20298153)

[Vegetation 12](#_Toc20298154)

[Geology and Soils 14](#_Toc20298155)

[Hydrology 15](#_Toc20298156)

[3. Data Analysis 17](#_Toc20298157)

[Objectives 17](#_Toc20298158)

[Data Summary 17](#_Toc20298159)

[Quality Assurance and Quality Control 19](#_Toc20298160)

[Methods 19](#_Toc20298161)

[Results 20](#_Toc20298162)

[4. Model 21](#_Toc20298163)

[Objective 21](#_Toc20298164)

[Model Description 21](#_Toc20298165)

[Methods 22](#_Toc20298166)

[Model Configuration Scenarios 23](#_Toc20298167)

[Parameter selection 24](#_Toc20298168)

[Initial conditions 25](#_Toc20298169)

[Running the models 25](#_Toc20298170)

[Results 26](#_Toc20298171)

[Discussion and Conclusion 34](#_Toc20298172)

[5. Summary of findings 37](#_Toc20298173)

[6. Towards a solution 38](#_Toc20298174)

[Reference List 39](#_Toc20298175)

[Appendices 43](#_Toc20298176)

# Acknowledgements

I wish to thank the following people for their generous support in the success of this project:

Bruce Davison, Environment and Climate Change Canada, for putting forward this project and for his guidance throughout.

Dan Princz, Environment and Climate Change Canada, for taking time out of his busy schedule to provide training, advice, and support for the modelling portion of the project. Without his assistance, this project would not have been possible.

Chris Spence, Environment and Climate Change Canada, for sharing his first-hand knowledge of the Baker Creek Watershed, and for the vast body of scientific knowledge he has published in his years of work in the Baker Creek watershed.

My family, for their support during this Master's program. To my husband, for taking on a larger share of our the household responsibilities, for giving me the time and space to study, and for providing encouragement and soundboarding. Also to my children, Ben and Lucy, for the joy they exude which helped me especially during the difficult times during this process. Also to my parents and in-laws for providing childcare and moral support.

# Executive Summary

The purpose of this project was to compare performance of various representations of spatial heterogeneity in the Baker Creek Watershed (NWT) using the MESH model, building on the work of former Master's of Water Security student, H. Mkandla (2017, University of Saskatchewan). The heterogeneous and variable nature of the earth continues to be a challenge to represent in hydrological models (Haghnegahdar, Tolson, Craig, & Paya, 2015). Therefore, it is important to choose a model configuration(s) that will give the best results within the limititations of observations, time, and computational efficiency.

The project was completed in 3 phases:

1. Replicate the methodology used in the White Gull Creek basin in order to compare the results (Scenario 1 and 2);
2. Increase the complexity of the model by increasing the number of grid cells in the basin (Scenario 3); and
3. Repeat the scenarios 1 and 2 using the PDMROF runoff algorithm (Scenario 1-P and 2-P).

The Baker Creek Watershed is located in the Northwest Territories of Canada near the capital city of Yellowknife in a subarctic climate and zone of discontinuous permafrost. The basin is approximately 155 km2 and receives 289 mm of precipitation annually, 41% of which as snowfall (Environment and Climate Change Canada, 2019a). The landcover in the basin is primarily bedrock, lakes, and coniferous forest hillslopes (Spence & Hedstrom, 2018). The hydrology of the site is dominated by large lakes connected by short streams with a complex runoff regime due to a highly variable contributing area that depends on a number of factors.

Driving data for the model was sourced from three main sources and combined into one continuous set for the modelling period. Streamflow observations from the Water Survey of Canada station 07SB013 - Baker Creek at the Outlet of Lower Martin Lake - were used to calibrate and validate the model (Water Survey of Canada, 2019).

The modelling period was 2006-2016, and each scenario was calibrated for 3 years during a wetter-than-normal period and 2 years during a drier-than-normal period with a 1 year spin-up period; the remaining years were used for model validation. A calibration program called OSTRICH (Matott, 2017) was used to run 1000 iterations per trial for 100 trials using the Nash-Sutcliffe Efficiency (NSE) as the objective parameter (Nash & Sutcliffe, 1970)

All the NSE results from the Baker Creek calibration were much lower than those for White Gull Creek and barely better than the no-model case. Regardless, the trends observed between the calibration and validation periods, and between scenarios, provided useful information about the behavior of the MESH model in the watershed and clues for future improvements to the model setup.

The results of the modelling showed that performance (NSE) improved with increasing complexity, but so did the average run time per trial, and performance during calibration was consistently better than validation with the exception of Scenario 1-P. Using the PDMROF runoff algorithm showed improved performance versus the WATROF algorithm. The model consistently underestimated large spring peakflows and overestimated low spring peakflows; it also either highly underestimated or completely missed fall peakflows.

Going forward, it is recommended that a sensitivity analysis be conducted and key parameters adjusted to improve the model performance. It is also recommended to test a configuration where the Grouped Response Units (GRUs) are defined by hydrological function and location in the watershed rather than by landcover type.

# 1. Introduction

Scientific research in the field of hydrology is continually seeking to better understand, predict, and model the movement of water throughout the earth. This research is important in a plethora of applications with great importance to society. For instance: better prediction of floods and droughts can lead to better management of watersheds with respect to impacts to water supplies, agricultural land, critical infrastructure, etc. A better understanding of the fundamentals of the various aspects of the water cycle can help to predict the potential impacts of a variety of changes - namely direct human impacts and a changing climate.

The heterogeneous and variable nature of the earth continues to be a challenge to represent in hydrological models (Haghnegahdar, Tolson, Craig, & Paya, 2015). There is ongoing work in the field of hydrology related to the best method of representing a watershed in a model and the required level of complexity. With limited watershed characteristic information, validation data, meteorological measurements, computational resources, and time, it is critical that discretization methods provide sufficient results for the modelling purpose in an efficient manner (Haghnegahdar, Tolson, Craig, & Paya, 2015).

The hydrological model being used in this project is the MESH model (Modelisation Environmentale Communautaire (MEC) - Surface and Hydrology). MESH is a land-surface-hydrology model which couples a land-surface scheme (LSS), which represents the vertical movement of water and energy between the atmosphere and earth's surface and subsurface, and a hydrologic model, which characterizes the movement of water horizontally over the land as well as through the soil (Changing Cold Regions Network, 2019a). MESH was developed by Environment and Climate Change Canada (ECCC) for the purposes of streamflow forecasting, building on the model WATCLASS, which coupled the CLASS and WATFLOOD models (Pietroniro et al., 2007). MESH has since been expanded and now allows the modeller to select from a number of LSSs as well as hydrological routing methods and alternative or additional process representations. The MESH model is continually under development by ECCC and by researchers at the Global Institute for Water Security (GIWS) at the University of Saskatchewan, with ongoing improvements to the code and hydrological process representations (University of Saskatchewan, 2019a).

This project builds on the work of H. Mkandla, a former MWS student (Mkandla 2017) which included a comprehensive background discussion on topics pertaining to the project. He included a general discussion of both the benefits and drawbacks of hydrological models, the difference between conceptual vs. physically-based models, and between lumped and distributed models. There are number of sources of error in the hydrological modelling process, which were also discussed in that paper, and include data uncertainty, measurement error, data quality, model parameter uncertainty, lack of data, and the model structure (Mkandla, 2017). A summary of the most common techniques used to deal with spatial variability in hydrological models was also discussed, as well as issues with scale variability between process understanding and representation, measurements, and modelling. Mkandla (2017) also introduced the recent improvements in data availability and computing power, which has subsequently lead to an increase in the feasible level of complexity in the models.

To avoid redundancy, this literature review presents an overview of recent publications using the MESH model, as well as work since 2017 related to the representation of spatial heterogeneity in hydrological models.

## Recent Research using the MESH Model

A number of key studies have been conducted in recent years which evaluate and seek to improve the representation of various hydrological processes and factors in MESH. Of note is a comparison of different routing modules and baseflow algorthms by Abdelhamed et al (2018), incorporation of controlled reservoirs by Yassin et al. (2019), alternative runoff algorithms, PDMROF and LATFLOW, proposed by Mekonnen et al. (2014) and Hossain (2017), respectively, and the parameterization and initialization of organic matter and permafrost conditions by Elshamy et al. (2019)

Abdelhamed et al. (2018) tested two different routing modules in MESH (WF\_R and RTE) and two baseflow algorithms (WATFLOOD and Luo et al., 2012) in the Upper Liard sub-basin in the Yukon. They found that the WATFLOOD algorithm out-performed the other baseflow algorithm, and that using RTE routing showed slighly better performance than WF\_R. They also found that performance metrics improved when using a combined objective function compared to the Nash-Sutcliffe alone (Abdelhamed et al., 2018).

Yassin et al. (2019) developed a method for incorporating a reservoir operation model into the MESH model, called the dynamically zoned target release (DZTR) model. They found that this model improved streamflow prediction versus a no-reservoir case and as compared to other reservoir models, especially for reservoirs where the ratio of storage capacity to annual inflow volume is greater than 0.5 (Yassin et al., 2019). Anis, Razavi, & Wheater (2017) also tested the coupling of MESH with MODSIM-DSS, a tool for modelling water management, and found it improved the representation of water withdrawls from the Bow River, Canada, particularly for irrigation purposes.

Two alternatives to the CLASS and WATROF representations of runoff generation have been developed for the MESH model. PDMROF was developed by Mekonnen et al. (2014) to better represent the fill-and-spill process for prairie wetlands. The algorithm represents the spatial variation in pothole storage and resulting dynamic relationship between surface storage capacity and contributing area for direct runoff due using a probability density function. One drawback of the PDMROF algorithm is that it does not include interflow (Hossain et al., 2016). PDMROF was tested in a prairie basin in Saskatchewan, Canada by Mengistu and Spence (2016) and found to adequately represent the contributing area compared to contributing area mapping. LATFLOW, developed by Hosain (2017), incorporates principles from both WATROF and PDMROF by using the probability density function to represent variable surface storage capacity and contributing area, and the Richards' equation to incorporate interflow (as in WATROF; Kouwen, 2014). The presentation by Hossain (2016) presents a clear conceptual representation of the difference between the CLASS, WATROF, PDMROF, and LATFLOW runoff algorithms.

Elshamy et al. (2019) outlined an approach for parameterizing and initializing areas where permafrost is present yet data is sparse. They found that a soil column depth of 50 m with the soil layer thickness increasing with depth as defined by a scaled power law was most suitable for stabilization of initial soil temperature and moisture conditions during spin-up (Elshamy et al., 2019). Spin-up considered a number of permafrost metrics for one hydrologic year for several iterations, and found that a quasi steady state was reached after 50-100 cycles. They found that the depth and composition of organic soil types as well as the depth to bedrock (SDEP) were important parameters that affected the thermal profile of the subsurface (Elshamy et al., 2019).

## Represenatation of Spatial Variability

Since 2017, there has been little advancement in published literature regarding methods of representing spatial heterogeneity. However, in Haghnegahdar, Tolson, Craig, & Paya (2015), a quantitative methodology was proposed to compare the relative performance of different watershed discretization schemes. The general methodology was to vary the sub-grid complexity based on the number of grouped response units (GRUs) where the GRUs are selected by landcover-only or landcover and soil type, calibrate the configurations for a set period constrained by a select number of time budgets, and validate in neighbouring ungauged sub-basins. They found that good performance in the calibration basins is not an indication of performance in ungauged basins, increased model complexity does not always improve calibration results especially when the calibration runtime is limited due to computational budget, and calibration to a sub-period (i.e 1 year instead of 3 years) may provide comparible results at a fraction of the computational time (Haghnegahdar, Tolson, Craig, & Paya, 2015).

## Objectives

In 2017, a Master's of Water Security student at the University of Saskatchewan completed a project which explored the effect of different representations of spatial heterogeneity in the White Gull Creek watershed in Saskatchewan using the MESH model (Mkandla, 2017).

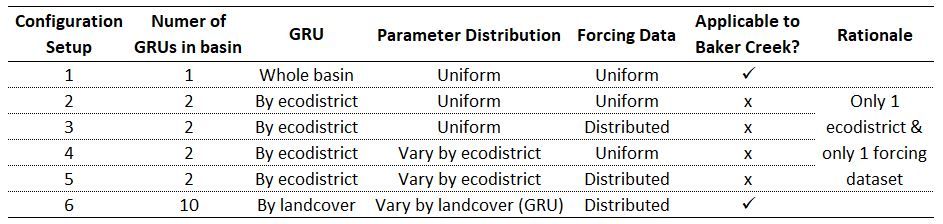
The objectives of the current project are:

1. In the Baker Creek Watershed (NWT), replicate the methodology used by Mkandla (2017) in the Whitegull Creek Watershed (SK) to evaluate the effect of complexity in the representation of spatial heterogeneity in the MESH model on model performance; and
2. Take the work further and explore additional model configuration options to explore their effects on model performance.

## Summary of the Mkandla Project

The purpose of the Mkandla project was to compare the model performance under a number of different representations of sub-grid spatial variability in the MESH model. The model configurations ranged from relatively simple - considering areal-averaged forcing data and model parameters - to more complex setups which considered distributed forcing data and parameters.

The MESH model was used, and configurations were arranged as shown in Table 1.1 below - with the complexity of the model increasing from configuration 1 through configuration 6. These configurations differed by utilizing a uniform forcing dataset (average of two datasets) vs distributing two different forcing datasets by the 2 ecodistrics in the basin, by implementing a uniform parameter set over 1 or 2 GRUs (i.e. whole watershed or by ecodistric), varying the parameters within each GRU by ecodistric, or by distributing the parameters based on landcover type in each ecodistrict. Table 1.1 below, adapted from Mkandla, 2017, shows the differences between the configurations and whether the configuration methodology was used for the modelling of the Baker Creek watershed.

***Table 1.1 - Summary of model configurations for the Mkandla experiment (Mkandla, 2017) and comparison with the Baker Creek modelling scenarios*** 

The model was calibrated using OSTRICH software (Matott, 2017) using the Dynamically-Dimensioned Search (DDS) algorithm, General-purpose Constrained Optimization Platform (GCOP), and the Nash-Sutcliffe Efficiency (NSE, Nash & Sutcliffe, 1970) as the response variable. One hundred calibration trials were run for each configuration with 1000 iterations for each trial (Mkandla, 2017).

The findings from the MESH modelling of the White Gull Creek watershed were:

* As the model complexity increased from configuration 1 through 4, so did the model performance during calibration, with a slight decrease in performance for configuration 5 and 6.
* The model performance during validation increased steadily from configuration 1 through 6.
* In all configurations, the NSE of the validation period was lower than that of the calibration period.
  + Therefore, model performance during calibration was not necessarily an indication of performance during validation.
* The range of NSE values of the calibration runs generally decreased as the model complexity increased, with the exception of configuration 4 and 6, which saw a larger spread between the 25th and 75th percentiles of NSE values compared to scenarios 3 and 5, respectively.
* The NSE values during both the calibration and validation periods for the best calibration run for configuration 1 and 2 were identical, highlighting the issue of equifinality (meaning different parameter sets can provide equally good solutions).
* As the number of GRUs in the model increased, so did the model run time.
* Configurations with uniform forcing data resulted in better performance during calibration, while configurations with distributed forcing data resulted in better performance during calibration (all other factors held constant) with minimal to no increase in run time.
* Parameter identifiability (defined as those below a threshold of 0.3) improved between configurations 1-3 and configurations 4-5, but degraded for configuration 6.
* The parameters below the identifiable threshold of 0.3 included saturated surface soil conductivity for configurations 1-3, minimum leaf-area index (LAI) for broadleaf trees in configuration 3, and minimum LAI, stream channel properties, percent clay in some soil layers, and saturated surface soil conductivity for configurations 4 and 5. Additionally, there were about 5 identifiable parameters in configuration 6.

This project will seek to compare the results of the White Gull Creek modelling with thoses obtained by modelling the Baker Creek Watershed.

# 2. Site Description

The Baker Creek watershed is located in the Northwest Territories (NWT) of Canada, with the outlet defined by the Water Survey of Canada (WSC) hydrometric gauging station 07SB013 - Baker Creek at Outlet of Lower Martin Lake - located approximately 7 km north of the capital city of Yellowknife, NWT (see Figure 2.1a below). The watershed covers an area of approximately 155 km2, with an elevation ranging from approximately 200 meters above sea level (masl) to 266 masl (Spence and Hedstrom, 2018). The watershed is located within the Taiga Shield Ecozone, the Tazin Lake Upland Ecoregion, and the Beaulieu River Ecodistrict (number 260) (Agriculture and Agrifood Canada (AAFC), 2019).

## Climate

The climate in the area can be classified as "subarctic" (Kokelj, 2003) and is characterized by short, cool summers and long, cold winters with daily average temperatures below 0oC between October and mid-April (Environment and Climate Change Canada, 2019a). At the nearby ECCC meteorological station, Yellowknife A, the 30-year daily average temperatures ranged from -25.6oC in January and 17.0 oC in July (Environment and Climate Change Canada, 2019a). The area received an average of 289 mm of precipitation annually, with approximately 41% as snowfall (Environment and Climate Change Canada, 2019a).

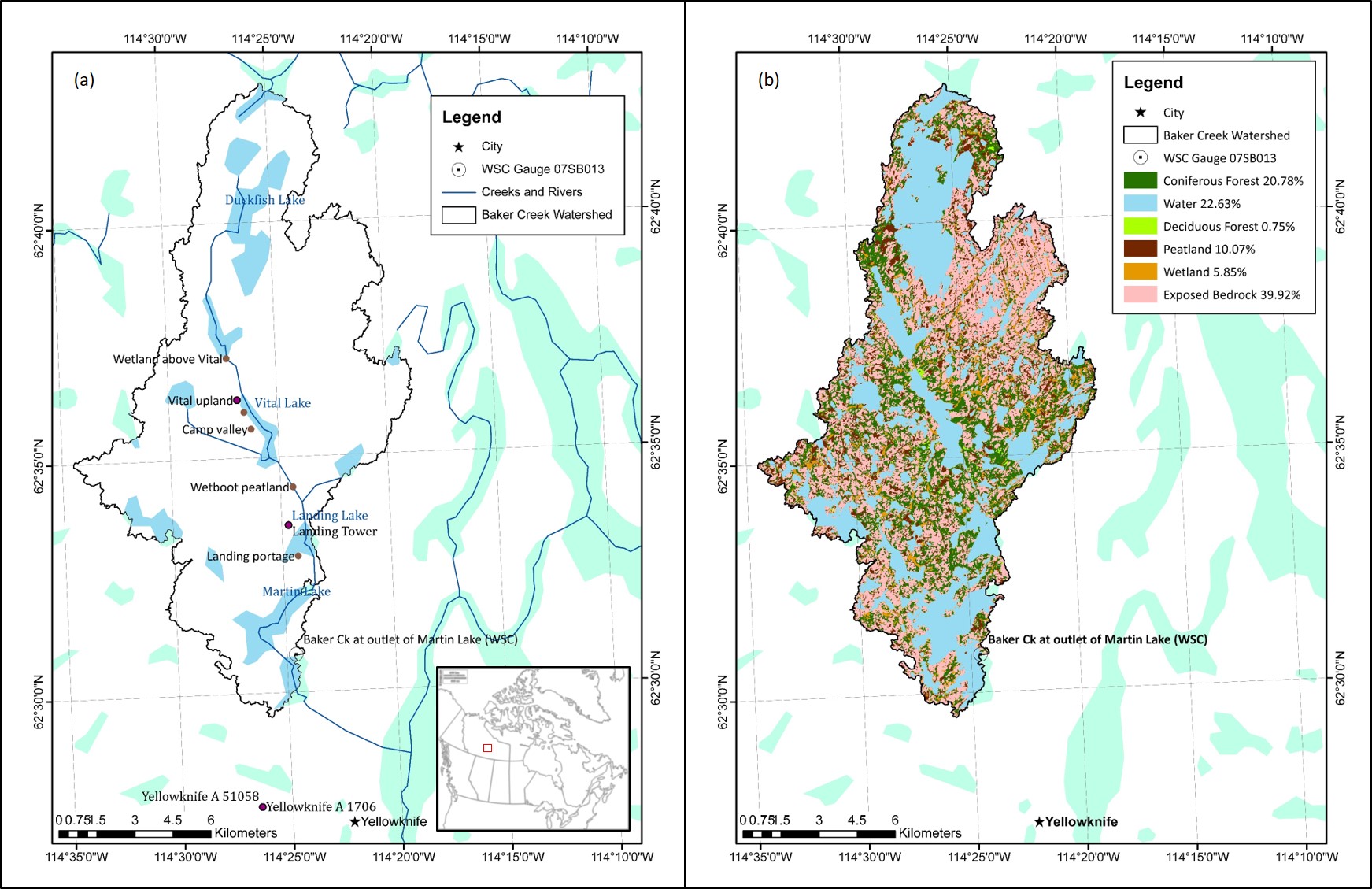
## Land Use

The Baker Creek watershed is largely undeveloped, and there are a number of research stations in the basin (Spence and Hedstrom, 2018). Research by the University of Saskatchewan and Carleton University has been occuring in the basin since 2004, with some water quality data available from 1995 (Changing Cold Regions Network (CCRN), 2019b; Spence and Hedstrom, 2018).

Water bodies and exposed bedrock make up the largest majority of the basin at 62.5% (22.6% and 39.9%, respectively), with forested hillslopes of primarily coniferous trees and some deciduous trees comprising just over 21%, and peatlands and wetlands the remaining 16% (Spence and Hedstrom, 2018); the Baker Creek watershed landcover map is shown in Figure 2.1(b). Photograph 2.1 below shows an example of bedrock upland, forested hillslopes, and a Vital Lake beyond, Photograph 2.2 is a closer view of the bedrock and treed areas at the shore of Vital Lake, and Photograph 2.3 depicts a peatland area near the Tower Peatland site.

## Vegetation

In general, the vegetation in the area can be described as open, short forest (Kokelj, 2003). Vegetation in peatland areas is typically a mixture of nonvascular plants, such as lichens, and vascular plants, such as evergreen shrubs, black spruce, and jack pine with an average vegetation height of about 0.3m (Guan, Westbrook, & Spence, 2010). At a forested valley site, Guan, Westbrook, and Spence (2010) described the vegetation as black spruce-dominated with some willow, birch, prickly rose, dwarf bilberry, and lichen; the area had a mean vegetation height of 7m. At a wetland site between Lake 690 and Vital Lake, vegetation was primarily deciduous shrubs with some black spruce, birch, and tamarack and a mean vegetation height of 1.4 m (Guan, Westbrook, & Spence, 2010).



***Figure 2.1 - Baker Creek Watershed location and major water bodies (a), and landcover (b); based on Spence and Hedstrom, 2018.***



***Photograph 2.1 - Baker Creek watershed from the Tower Peatland instrumentation site (Photo Credit: Cook, 2019)***



***Photograph 2.2 - Overlooking Vital Lake in the Baker Creek watershed (Photo Credit: Cook, 2019)***



***Photograph 2.3 - Peatland near the Tower Peatland instrumentation site (Photo Credit: Cook, 2019)***

## Geology and Soils

The Baker Creek watershed is located in the Canadian Precambrian Shield and contains pockets of lacustrine deposits from the historical inundation glacial activity during the Wisconsin glaciation (Guan, Westbrook, & Spence, 2010). The bedrock in the basin is moderately to highly fractured, and mineral soils (silty and sandy texture) are present in the fissures and valleys from bedrock weathering and erosion (Phillips, Spence, & Pomeroy, 2011). The thickness of overburden above the bedrock ranges from less than 1 m to more than 10 m (Spence & Hedstrom, 2018). The peatland site studied by Guan, Westbrook, & Spence (2010) was comprised of 1.2 m of peat soil with bedrock below, the valley site was characterized by about 0.2m of organic matter underlain by loose gravel, silty clay, and sandly clay, and the wetland site contained 0.2-0.6 m of peat overlying impervious lacustrine clay (Guan, Westbrook, & Spence, 2010).

## Hydrology

The Baker Creek watershed contains a number of large lakes drained by short streams which exit the basin through the gauge at the outlet of Lower Martin Lake, which has a gross drainage area of approximately 155 km2 (Spence et al., 2010). Baker Creek then continues for another 6 km and flows into Great Slave Lake. The flow regime in the basin is highly variable due to the variability of storage capacity in the basin (Spence, 2006; Phillips, Spence, & Pomeroy, 2011). There are 349 lakes in the basin - 97% of which are smaller than 0.5 km2 and eight are larger than 1 km2 (Spence, 2006). The largest and most hydrologically-significant of these lakes include Duckfish Lake (6.2 km2), Vital Lake (1.5 km2), Landing Lake (1.1 km2), and Martin Lake (3 km2) (Spence, 2006). The dominant lakes and general streamflow through the basin is shown in Figure 2.1(a) above.

The average streamflow at the WSC gauge (07SB013) is 0.24 m3 s-1 and annual runoff ratio is 0.17; however, the runoff ratio has been observed to vary by up to three orders of magnitude, as observed by a runoff ratio of 0.005 in 2015 and 0.34 in 2001 (Spence & Hedstrom, 2018; Spence & Woo, 2002). This is due to variable storage and infiltration capacity controlled by depth of soil cover, geometry of bedrock fractures, rainfall intensity, and evaporation (Kokelj, 2003). The basin is in an area of discontinusous permafrost, which also alters the runoff regime depending on the soil conditions (Kokelj, 2003; Guan, Spence, & Westbrook, 2010). The presence of permafrost is influenced by topography, vegetation, snow accumulation, geology, and hydrological conditions, and stark variations in these properties result in highly discontinuous permafrost patterns (Spence & Hedstrom, 2018). Typically, permafrost is found in areas containing organic glaciolacustrine clays and outwash and organic deposits (such as peatland and valley areas) while bedrock and well-drained sandy areas tend to be permafrost-free (Spence & Hedstrom, 2018).

Phillips, Spence, & Pomeroy (2011) stated that there are thresholds in the basin that dominate the hydrological response of the basin, and that the rates and energy driving key hydrological processes are the most influencial factors governing hydraulic connectivity in the basin (Phillips, Spence, & Pomeroy, 2011).

The hydrological response at the outlet of Lower Martin Lake is controled by a number of "gatekeepers", which are elements in the basin that collect, store, and discharge water (Phillips, Spence, & Pomeroy, 2011; Spence & Woo, 2002). In upland areas, the gatekeepers are primarily peatlands and wetlands which control the connectivity of bedrock runoff, and gatekeepers in the downstream areas are large lakes which collect, store, and discharge water from a number of sub-catchments (Phillips, Spence, & Pomeroy, 2011). This nested-nature of storage and release in the basin make the streamflow response of the basin highly variable. Elements that can have the most influence on the connectivity of the basin are lakes with a large lake area compared to contributing area, and lakes closer to the outlet (Phillips, Spence, & Pomeroy, 2011).

The hydrograph at the basin outlet features peak streamflows in the spring due to snowmelt runoff (Water Survey of Canada, 2019; see Figure 4.2), and fall flows can also be observed due to late-summer rains that reduce available storage, intense fall rain events, and reduced evapotranspiration capacity as the trees become dormant (such as the fall peak in 2008; Spence et al., 2010). Backwater conditions have also been observed in the basin, which can delay streamflow propagation to the outlet (Spence, 2006).

# 3. Data Analysis

## Objectives

The objective of the data analysis portion of this project was to prepare the driving data and streamflow data for the model. The meteorological parameters required by MESH are: temperature, humidity, barometric pressure, wind speed, incoming shortwave radiation, incoming longwave radiation, and precipitation. The timestep of the driving data files must be the same frequency of the model timestep (typically 30 or 60 minutes) and be complete for the entire model duration (University of Saskatchewan, 2019a). Preparation of the driving data involved combining driving data from a number of sources into one continuous set.

## Data Summary

Driving data for the model was obtained from 4 sources:

1. Environment and Climate Change Canada's historical weather records for the Yellowknife A station (Environment and Climate Change Canada, 2019).
2. Hydrometeorological and hydrological data collected within the Baker Creek Watershed during research activitys betwen 2003 and 2016 (Spence & Hedstrom, 2018); the Vital Tower is shown in Figure 3.1 below.
3. Environment and Climate Change Canada's Regional Deterministic Prediction System (RDPS), which is based on the GEM-NWP model (Mailhot et al., 2006).
4. Precipitaiton data from the Canadian Precipitation Analysis (CaPA) dataset (Fortin et al., 2018).



***Figure 3.1 - Vital Tower located within the Baker Creek watershed (Photo Credit: A. Cook, 2019)***

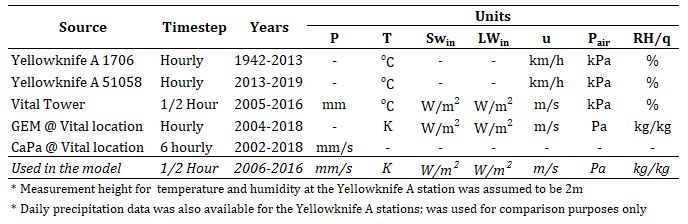
Historical weather records for the ECCC Yellowknife A station are available since 1942 and includes air temperature, wind speed, barometric pressure, and relative humidity at a hourly timestep, and precipitation at a daily timestep (Environment and Climate Change Canada, 2019a).

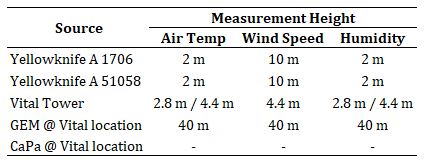
The CaPA datasets are produced by ECCC by combining precipitation data from a number of different sources such as precipitation gauges operated by the Meteorological Service of Canada and other partners, radar information from the Canadian Weather Radar Network, and short-term forecasts provided by the Regional Deterministic Prediction System (RDPS) using the Regional Deterministic Precipitation Analysis (RDPA) (Environment and Climate Change Canada, 2014). The CaPA data is available for much of North America, and consists of a gridded quantitative precipitation estimate (QPE) every 6 hours or 24 hours. For this project, CaPA data was obtained for the grid containing the Vital Tower in the Baker Creek Watershed for the years between 2002 and 2018 at a 6 hour timestep.

Meteorological measurements at an hourly timestep were obtained from ECCC's Global Environmental Multiscale (GEM) model (Mailhot et al., 2006) at a 40 m height from 2004 onward, and from a 2 m height from 2011 onward. GEM is used in the Regional Deterministic Prediction System (RDPS) which produces short-term forecasts (Environment and Climate Change Canada, 2014).

Spence & Hedstrom (2018) includes hydrometeorological data, ground temperature, soil moisture, streamflow, and spatial data for the Baker Creek watershed between 2005 and 2016; a number of these datasets were utilized for the project.

The meta data for the hydrometeorological data sources are shown in Table 3.1 and 3.2 below.

***Table 3.1 - Measurement interval, time period, and units for the hyrometeorological data sources used in the model and units requred by MESH*** 

***Table 3.2 - Measurement heights for the hyrometeorological data sources used in the model*** 

Streamflow measurements were obtained from the Water Survey of Canada for station 07SB013 "Baker Creek at the Outlet of Lower Martin Lake"(Water Survey of Canada, 2019). The end of the modelling period was limited to the end of 2016 since streamflow observations beyond that time were not available at the time of data sourcing.

## Quality Assurance and Quality Control

Quality assurance and quality control checks were performed during data processing by visually comparing plots of meteorological variables from each source. Conservation of the sum of precipitation and of the average of all other variables was ensured by comparing the completed driving dataset with the original source data.

## Methods

All data processing and plotting was done using the R programming language via RStudio (Studio Team, 2018). The general methodology was to characterize each dataset, load into R, plot the data, perform unit conversions, perform timezone shifts (for the GEM and CaPA datasets), interpolate the data, and scale the meteorological observations up to the measurement height required for the model (as necessary). The datasets from each source were prepared separately and converted into a consistent format. Then, the forcing variable from each source was compared, and stitched into one complete set. The priority of source datasets was:

* Vital Tower
* Yellowknife A
* GEM/CaPA for the grid containing Vital Tower

The R scripts used in the driving data preparation and additional driving data preparation methodology are included in Appendix D.

The Yellowknife precipitation data was not used due to the coarse (daily) measurement interval; however, it was used to during analysis to compare with the precipiation data from the other sources. Precipitation data from the Vital Tower was not adjusted for undercatch. Even though GEM data was obtained for both 2 m and 40 m measurement heights from 2004, only the 40 m dataset was used for consistency purposes and to limit the calculations required.

Hourly observations from the Yellowknife and GEM datasets were converted to half hourly intervals using linear interpolation. The CaPA precipitation data was converted from 6 hourly to half hourly intervals by taking the precipitation rate at the time observed and back-filling the 6 hours prior with the same precipitation rate (i.e. the reported precipitation rate is the precipitation rate for the previous 6 hours).

During data inspection, there were some noticeable shifts forward in time or up in magnitude in the Vital data compared to the GEM and/or CaPA data. These shifts were adjusted to match the GEM/CaPA data or sections removed from the final dataset (see Appendix D for more detail).

All data sources were taken as local standard time for Yellowknife (Mountain Standard time, UTC-7) except for the GEM and CaPA data which was in UTC. Therefore, GEM and CaPA data were shifted to MST.

Relative humidity was converted to specific humidity using the equation from the MESH wiki page (University of Saskatchewan, 2019a). Temperature observations were scaled up to 40 m using a dry adiabatic lapse rate of 6.5oC/km. Wind observations were scaled up to 40 m using equations 3.27 and 3.30 a and b from Dingman, 2015 considering an average vegetation height of 2 m.

Streamflow measurements obtained from the Water Survey of Canada at station 07SB013 were loaded into R, and then processed for use in the model. MESH does not read negative flow values and requires a full complete dataset for the model duration; therefore, missing values were replaced with an arbitrary negative number (i.e. -9999), and subsets for the calibration and validation periods were created by converting flows outside the period of interest to negative values.

## Results

The outcome of the data analysis was a continuous dataset for the modelling period for each of the seven hydrometeorological variables requred by MESH, as well as the streamflow data for the model input and post-processing.

# 4. Model

## Objective

The main objective of the modelling exercise was to repeate the experiment by former MWS student, H. Mkandla (2017), in the Baker Creek watershed to determine if the same relationship between model complexity with respect to the representation of spatial heterogeneity in the MESH model would be observed compared to those obtained in the White Gull Creek watershed. The key observations from the White Gull Creek modelling are presented in the *Introduction* section. The main objective is hereafter referred to as "Phase 1" of the project.

The secondary objectives, selected upon reviewing the results of the replication phase, were to a) Calibrate a third modelling scenario considering more than one grid in the basin in order to include sub-basin routing (Phase 2), and b) Assess the model performance using the PDMROF overland flow algorithm (Phase 3).

## Model Description

The modelling was completed using the MESH model (Modelisation Environmentale Communautaire (MEC) - Surface and Hydrology). MESH is a land-surface-hydrology model which couples a land-surface scheme (LSS), which represents the vertical movement of water and energy between the atmosphere and earth's surface and subsurface, and a hydrologic model, which characterizes the movement of water horizontally over the land as well as through the soil (Changing Cold Regions Network, 2019). For this project, the Canadian Land Surface Scheme (CLASS) version 3.6 (Verseghy, 2012) and WATROF (Kouwen, 2014) options were utilized within MESH for Phase 1 and 2, and PDMROF (Mekonnen et al., 2014) was utilized for Phase 3. Phases 1 and 2 utilized MESH version r1024, which was the same version used by Mkandla (2017), and verion r1552 was used for Phase 3.

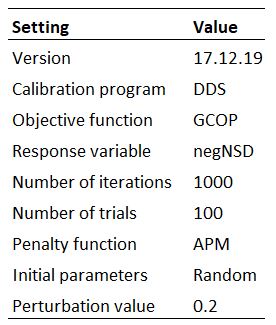
The MESH model uses a Grouped Response Unit (GRU) methodology, which assumes that each GRU will behave hydrologically similarly based on certain properties, such as vegetation, land use, soil type, slope/aspect, etc (Kouwen, 2014). Each GRU is parameterized with one or more of five landcover types - needleleaf trees, broadleaf trees, crops, grass, and urban/barren land/impervious areas (University of Saskatchewan, 2019a), as well as hydrological and soil characteristics. GRUs are not necessarily contiguous throughout the watershed. The MESH setup requires that the watershed be divided into one or more grids, and the portion of each GRU within a grid cell are called tiles (Elshamy et al., 2019). At each timestep, MESH utilizes the LSS (CLASS in this case) to calculate the vertical water and energy balances for each tile.

The OSTRICH software (Matott, 2017) was used for calibrating the model. OSTRICH can be used with any text-base modelling software, and can be configured to utilize a number of different calibration and optimization algorithms (Matott, 2017). The Dynamically-Dimensioned Search (DDS) algorithm was used to calibrate MESH; this algorithm was created for calibrating hydrologic models with a large number of parameters (Tolson and Shoemaker, 2007). The DDS algorithm is for global optimization, it adjusts the search space to find a good solution and avoid poor local optimum values, and has been shown to be computationally efficient compared to other methods (Tolson and Shoemaker, 2007). A General-purpose Constrained Optimization Platform (GCOP) was also used within OSTRICH, which uses a response variable to calculate the system cost and penalty function; the goal of the algorithm is to minimize the cost function (Matott, 2017). For this project, the Nash-Sutcliffe Efficiency (NSE, Nash & Sutcliffe, 1970) was used as the response variable, but since OSTRICH seeks to find the minimum, the negative of NSE was fed to OSTRICH. A script was also written to return an infeasible value of negative-NSE in the case of a model crash. The equation for calculating NSE is:

where modeled streamflow at time *t*, observed (measured) streamflow at time *t*, and mean of observed streamflow.

The Nash-Sutcliffe efficiency is a the difference between initial and residual variance divided by the initial variance; when the residual variance is equal to the intial variance, NSE=0 and the simulated values are the same as the no-model values (Nash & Sutcliffe, 1970). The NSE value can range from to +1, with a value of 0 meaning that the average of the streamflow is just as close of a guess as the model output (Nash & Sutcliffe, 1970) and a value of 1 representing a perfect match between observed and simulated values.

The OSTRICH configuration settings are presented in Table 4.1 below.

***Table 4.1 - OSTRICH configuration details***  


For all scenarios, there were four soil layers defined with thicknesses of 0.15 m (shallowest), 0.25 m, 0.70 m, and 3.00 m (deepest) for a total depth of 4.1 m.

## Methods

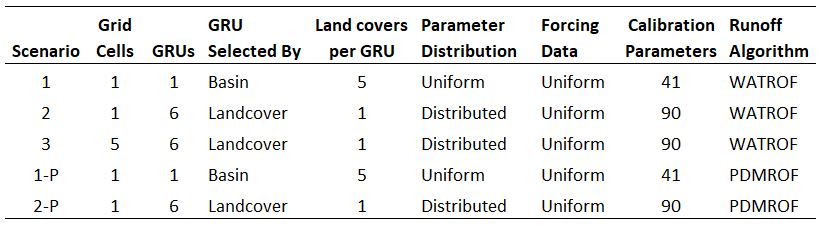
This section presents the methods used in designing the model configuration for each scenario, selecting parameters and initial conditions, and running the model.

### Model Configuration Scenarios

To the extent possible, the methodology used in the Mkandla project (Mkandla, 2017) were followed for Phase 1 of this project while also adjusting the setup where appropriate to incorporate the characteristics of the Baker Creek watershed. All of Mkandla's MESH modelling files were available for this project, so the same setup and procedure was closely followed.

In the Mkandla experiment (2017), six different model configurations were used. These configurations differed by utilizing a uniform, averaged forcing dataset vs distributing two different forcing datasets betwetten the 2 ecodistrics, by implementing a uniform parameter set over 1 or 2 GRUs (i.e. whole watershed or by ecodistric), varying the parameters within each GRU by ecodistric, or by distributing the parameters based on landcover type in each ecodistrict (see Table 1.1, in the *Introduction* section). Since the watershed area was relatively small and the goal of the project was to explore the effect of the representation of sub-grid spatial heterogeneity within MESH, all scenarios considered one grid cell (Mkandla, 2017).

The Baker Creek watershed is approximately 155 km2, which is approximately one-quarter of the size of the White Gull Creek watershed (629 km2), and is located within one ecodistrict (see *Site Description*. Additonally, there was only one set of driving data for the basin (see *Data Analysis*). Therefore, the replication phase (Phase 1) of this project considered only two different model configurations; the details are presented in Table 4.2 below.

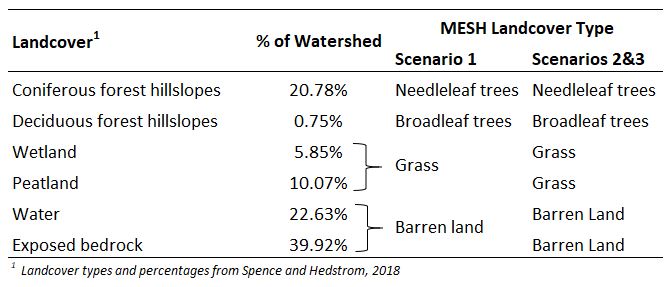
***Table 4.2 - Model configuration details*** 

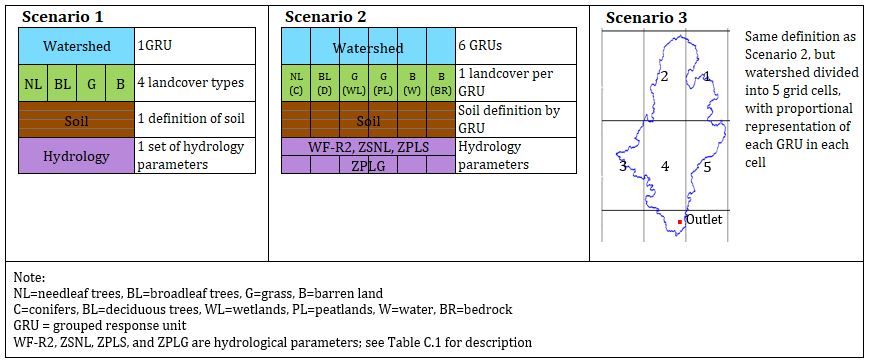
Scenario 1 considered the basin to contain one GRU - meaning that the entire watershed is assumed to behave hydrologically similar. This GRU contained five landcover types based on the CLASS parameterization (Verseghy, 2012). The land cover types defined in Spence and Hedstrom (2018) were used, with the CLASS landcover type "grass" being used to define wetlands and peatlands, and the "barren land" type being used to define open water and exposed bedrock areas (see Table 4.3 and Figure 4.1 below).

In Scenario 2, the basin was divided into 6 GRUs defined by landcover type, with the vegetation, surface, hydrology, and soils parameterized individually for each GRU (see Figure 4.1 below).

In Scenario 3, the parameterization was the same as for Scenario 2 but the watershed was divided into five grid cells (see Figure 4.1 below) with tiles in each cell representing the a portion of each GRU. This configuration was chosen as it was the next logical step in increasing the complexity of the model configuration.

Scenarios 1-P and 2-P were parameterized the same as Scenarios 1 and 2, respectively, except that parameters for WATROF were replaced with those for PDMROF (Mekonnen et al., 2014; University of Saskatchewan, 2019a). It was hypothesized that due to the large number of lakes, dominance of large lakes, and fill-and-spill nature of the basin, PDMROF may improve the predictive power of the model.

***Table 4.3 - Landcover parameterization in the MESH model*** 



***Figure 4.1 - Conceptual figure of MESH scenario configurations***

### Parameter selection

The general methodology for parameter selection was to choose values (for non-calibrated parameters) and ranges (for calibrated parameters) that would best represent the physical conditions in the Baker Creek basin based on literature and personal experience of research personnel (Spence, 2019). Table C.1 in Appendix C presents the parameter values, rationale, and referenced sources for non-calibrated parameters, and Table C.2 presents the parameter ranges, rationale, and referenced sources for calibrated parameters. Soil characteristics were first selected for Scenario 2, and an areal-weighted average by landcover type (Spence and Hedstrom, 2018) was used to select the values and ranges to represent the "effective" soil characteristics for Scenario 1. Since bedrock and peat are parameterized in MESH using integer flags (University of Saskatchewan, 2019a), bedrock layers were appoximated as 100% clay, and peat layers were approximated as 100% organic matter.

### Initial conditions

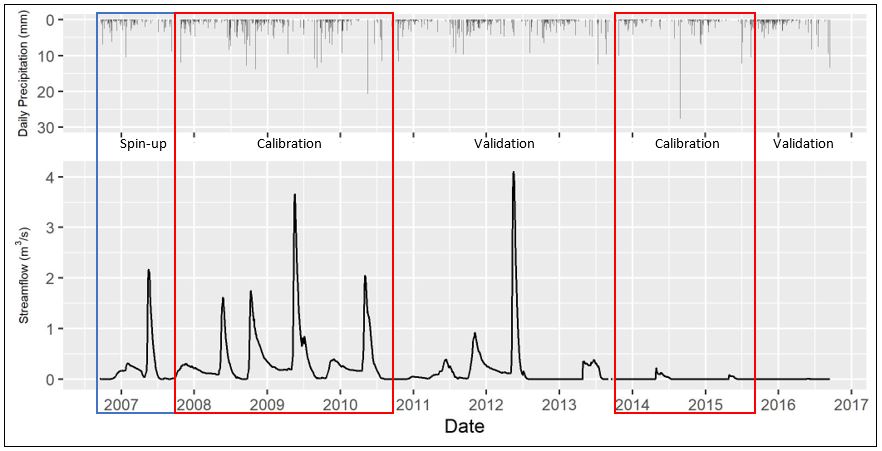
The model start time was selected to be in the fall at a time when the vegetation was entering the dormant stage, when there was no snow in the watershed, and there had been little to no rain in the few days prior. These conditions allow the initial condition of snow depth, snow density, and water ponding depth in the canopy and on the ground to be set to zero, which avoids introducing bias into the model by estimating these values. Normally, a hydrological model start date is selected as October 1st - the "typical" start of the hydrologic year, but due to the northern latitude and precipitation conditions in 2006 when the model was started, the start date was chosen as September 15th.

Initial soil temperature and moisture for the surface layers were calculated from the average of the observed measurements on September 15 from the data reported in Spence and Hedstrom (2018) and the temperature and moisture profiles for the deeper layers were estimated based on the trends reported in Morse, Wolfe, Kokelj, & Gaanderse (2016). For Scenarios 2, 3, and 2-P, variations based on landcover type were considered in selecting the initial conditions.

### Running the models

As discussed above, the model start date was selected as September 15, 2006 and the first year of the model run was used as a spin-up period prior to calibration. During the years of data availability (2005-2016), there was a general period of wetter-than-normal conditions in the watershed from 2008 to 2011, and a drier-than-normal period from 2012-2016 (Spence and Hedstrom, 2016). The model calibration period was divided into two sections - one in the wet period and one in the dry period - in order to condition the model to respond to both states. The model was calibrated from September 15, 2007 through September 15, 2010, and from September 15 2013 through September 14, 2015 (5 years total); the validation periods were September 15, 2010 through September 14, 2013, and September 15, 2015 through September 13, 2016 (4 years total). Figure 4.2 below shows the hydrograph for WSC station 07SB013 at the outlet of the watershed for the years between 2005-2016 along with the selected spin-up, calibration, and validation periods.

MESH was run using a half hour timestep and a reference height of 40 m for all measurements.



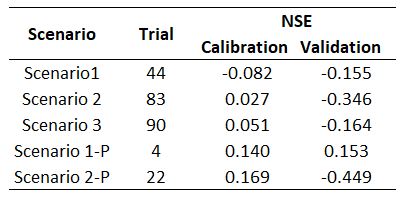
***Figure 4.2 - Streamflow at WSC station 07SB013 Baker Creek at the Outlet of Lower Martin Lake and the daily precipitation forcing data, showing model spin-up, calibration, and validation periods***

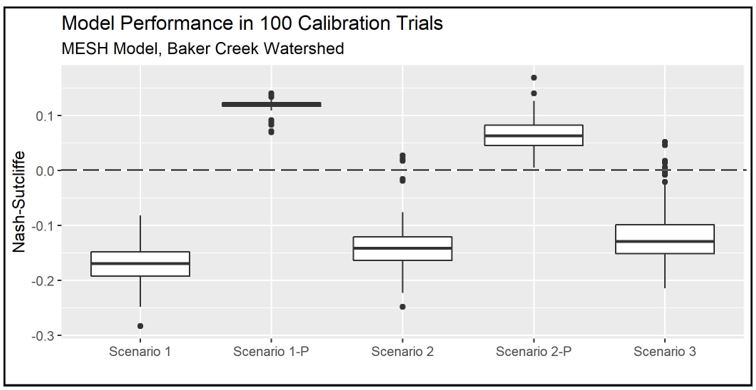
The model scenarios were calibrated using OSTRICH for 100 trials - each using a different seed number - with 1000 iterations of MESH for each trial. MESH and OSTRICH were compiled and the calibrations were run on the University of Sasktchewan's High-Performance Computing cluster, PLATO (University of Saskatchewan, 2019b). In order to reduce total model run time, the scenarios were divided into batches and run concurrently; Scenarios 1, 2, 1-P, and 2-P were divided into 10 batches, and Scenario 3 was divided into 15 batches.

## Results

Results were processed and visualizations created using RStudio (Studio Team, 2018); the post-processing code is included in Appendix D. Water balance plots for each scenario are included in Appendix C.

The trial number and NSE value for the best calibration run and corresponding validation NSE are presented in Table 4.4 below, and a figure showing box plots of the NSE of calibration results for the 100 trials of each model scenario are presented in Figure 4.3. The ranked NSE values for the calibration of all scenarios is presented in Figure C.1 (Appendix C). It is observed that the best calibration result for Scenario 1 is below zero, meaning that the average observed streamflow provides a better estimate of watershed response than the model result (Nash & Sutcliffe, 1970). The best calibration results for Scenarios 2 and 3 are only slightly above zero while the majority of the results from the 100 trials were below zero. With increasing model complexity from Scenario 1 to 2 to 3, the model performance in terms of NSE also improved. In both cases, the configuration utilizing PDMROF performed better than the one using WATROF. While Scenario 2 results were better than Scenario 1, Scenario 1-P results were better than Scenario 2-P.

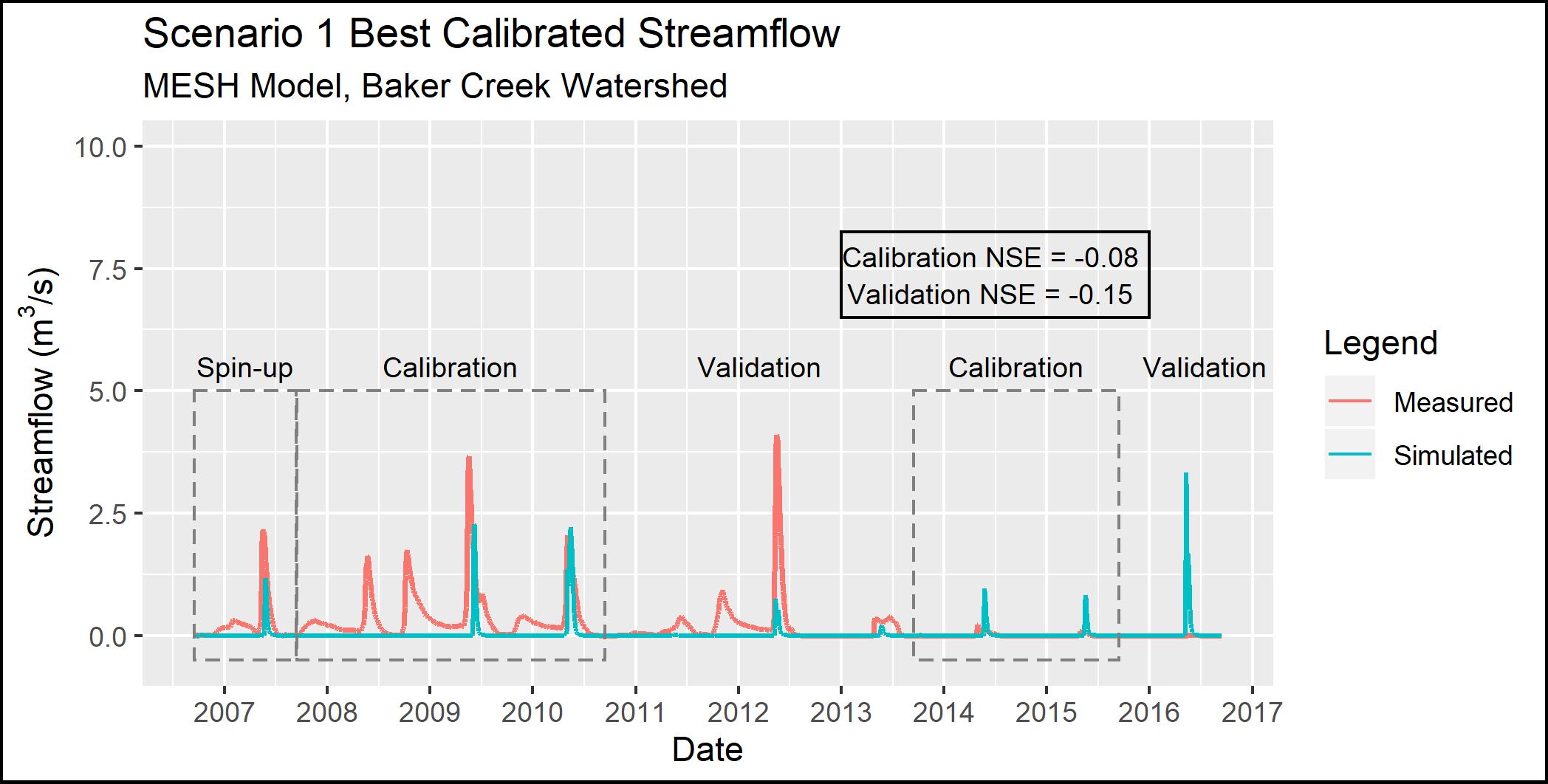
***Table 4.4 - Summary of Nash-Sutcliffe Efficiency (NSE) values for the calibration and validation periods of each model scenario***  




***Figure 4.3 - A comparison of the model performance (NSE) for each configuration scenario. The horizontal line represents the NSE limit below which model results are worse then no-model observed streamflow***

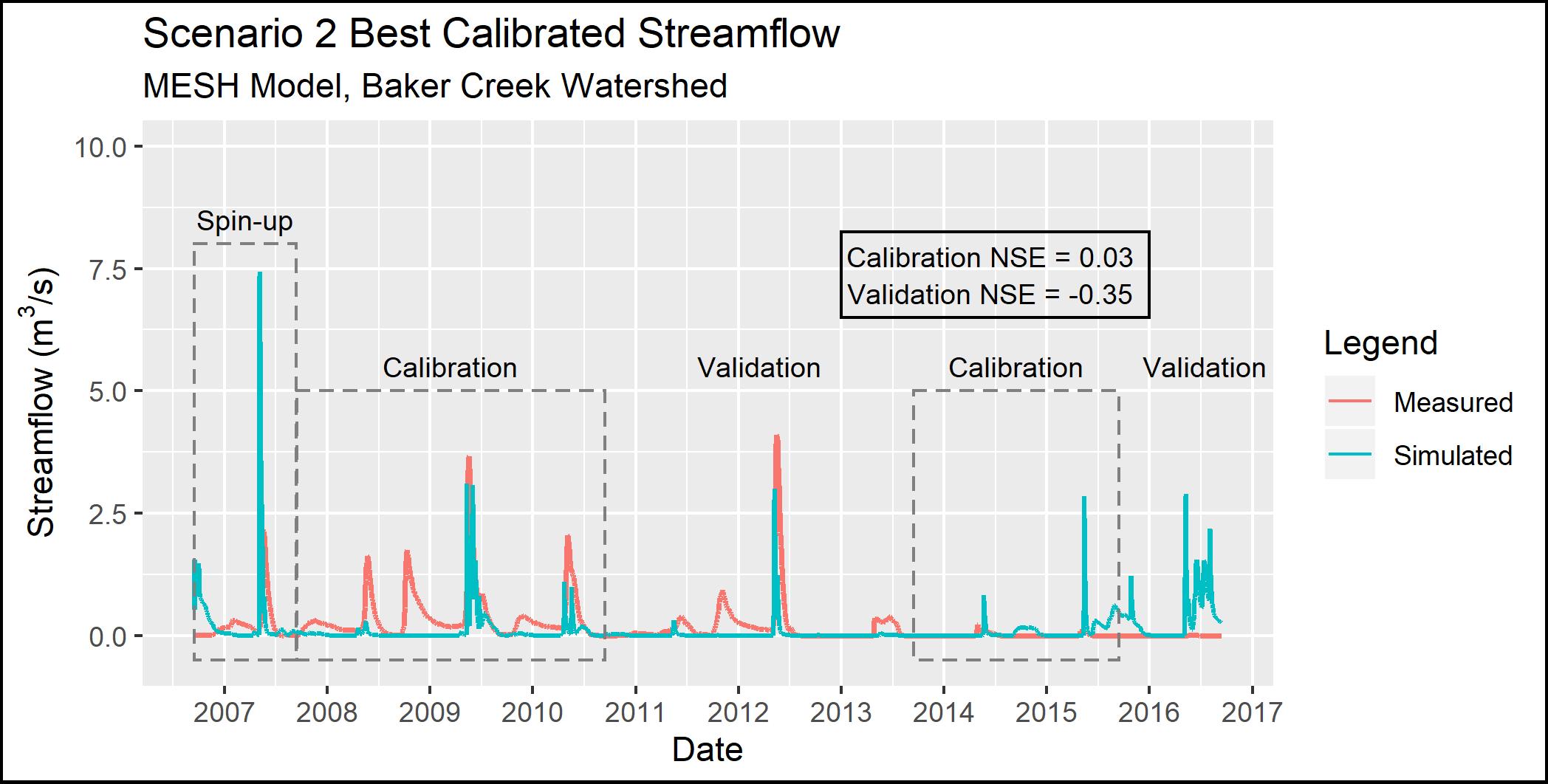
Hydrographs showing measured vs. best calibration streamflow, as well as NSE values for the calibration and validation periods, are presented in Figure 4.4 through 4.8.

As shown in Figure 4.4 below, the streamflow results from Scenario 1 do not match measured streamflow well. Only the magnitude of the spring peak in 2010 is matched, while the rest of the spring peak flows are greatly underestimated (or missed entirely), and the spring peak flows during the dry period are greatly overestimated. Additionally, fall peak flows and winter flows are not represented by the model.



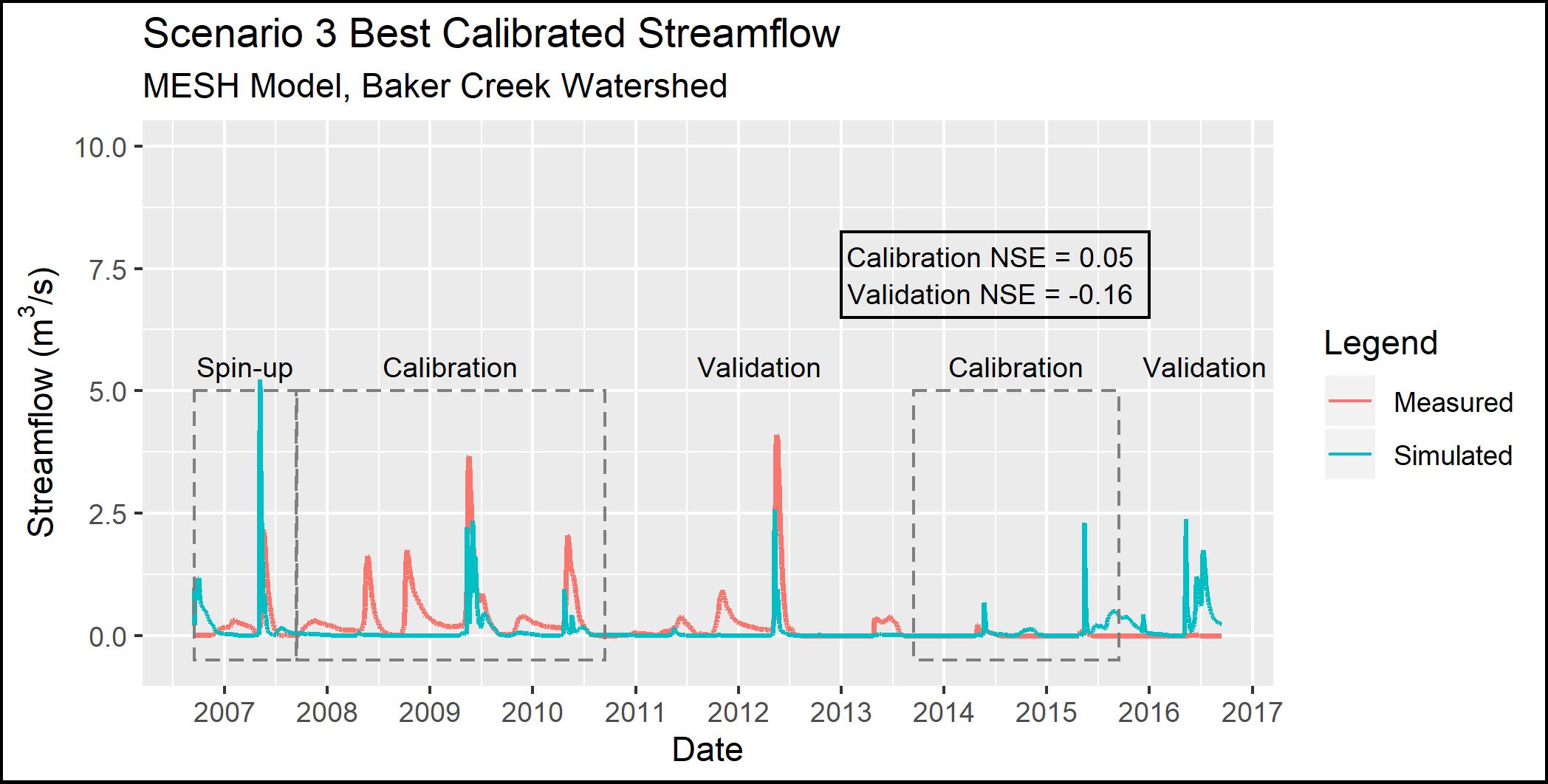
***Figure 4.4 - Observed and simulated streamflow for Scenario 1***

As shown in Figure 4.5 below, the Scenario 2 model configuration simulates the peak flows in the spring of 2009, 2010, and 2012 reasonably well, yet misses the fall/winter flows, especially the fall peak flows in 2008, 2009, and 2011. This scenario estimates the spring peak flows better than Scenario 1 for most years, and there is a small amount of fall flow in 2009. The model also tends to under-estimate peak flows during the wet period (2008-2012) and over-estimate peak flows during the drier period (2013 onward).



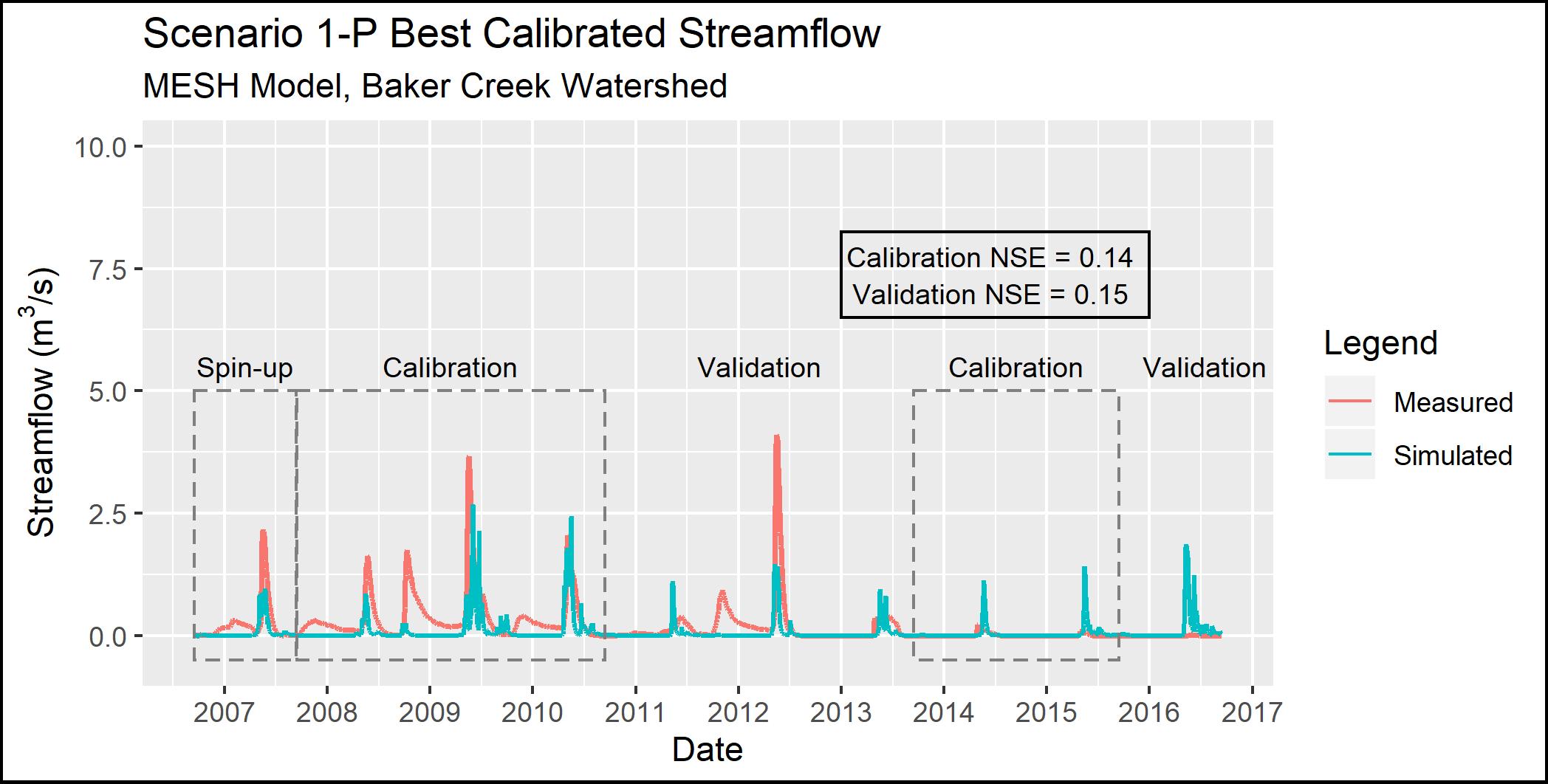
***Figure 4.5 - Observed and simulated streamflow for Scenario 2***

The results for Scenario 3 are comparible to Scenario 2, as shown in Figure 4.6 below (compared to Figure 4.5 above). This similarity indicates that increasing the complexity by adding more grid cells in the basin and including hydraulic routing through the basin resulted in only minimal improvement in model accuracy.



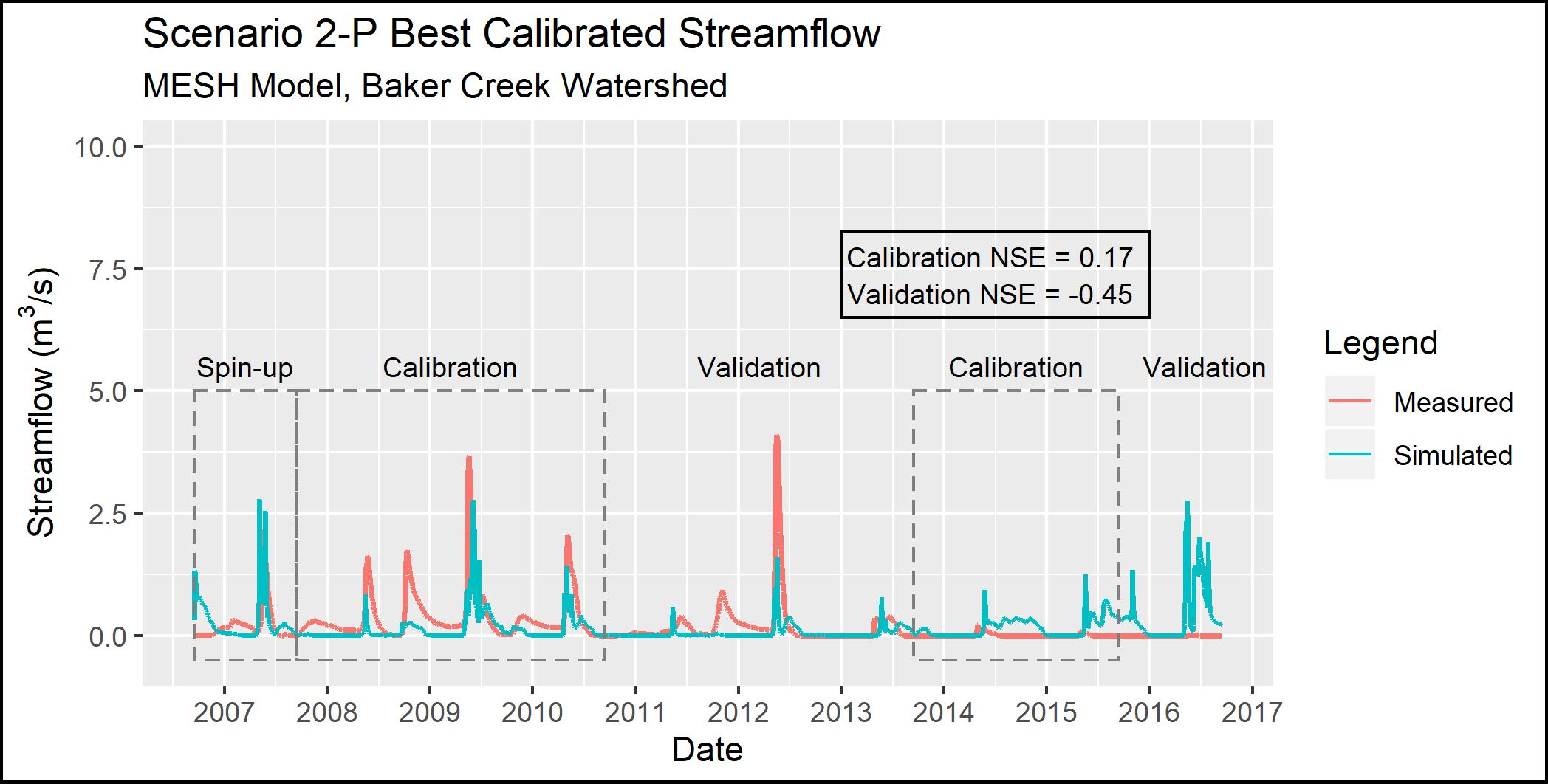
***Figure 4.6 - Observed and simulated streamflow for Scenario 3***

The results for Scenario 1-P (which used the Scenario 1 parameterization but switched PDMROF for WATROF) are shown in Figure 4.7. Compared to Scenario 1, the Scenario 1-P model is a better reflection of the spring peaks, with the most notable improvements in spring 2008 and 2011. Scenario 1-P shows a small amount of flow in the fall of 2009, but all other fall peaks are negligibly represented by the model.



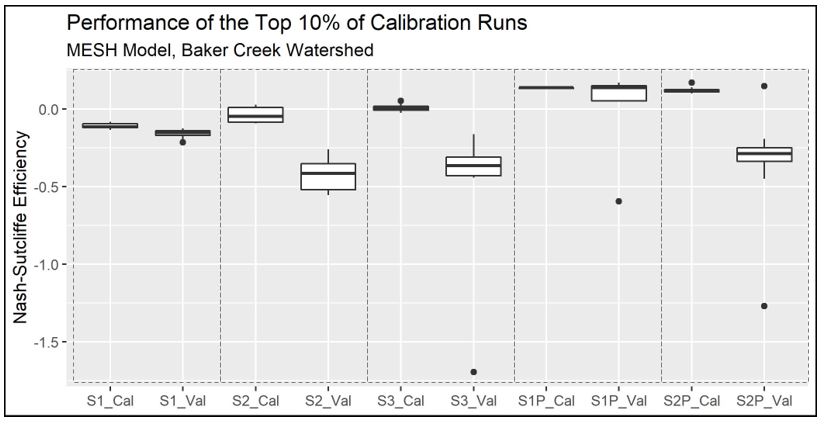
***Figure 4.7 - Observed and simulated streamflow for Scenario 1-P***

The results for Scenario 2-P (which used the Scenario 2 parameterization but switched PDMROF for WATROF) are shown in Figure 4.8. Compared to Scenario 2, the Scenario 2-P spring peak flows are lower in 2009 and 2012, and the fall peaks are more visible in 2008 and 2009 though still much lower than observed.



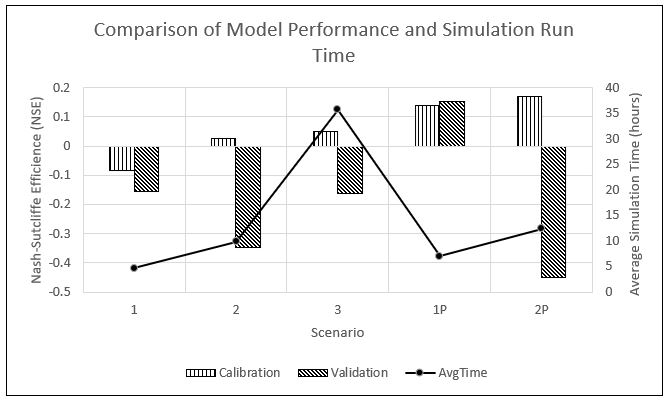
***Figure 4.8 - Observed and simulated streamflow for Scenario 2-P***

Figure 4.9 (below) presents a comparison between the top 10% (10 out of 100) of the calibration trials and the corresponding validation results. In every scenario except Scenario 1-P, the performance during the validation period was worse than during the calibration period.



***Figure 4.9 - Comparison of the performance of the top 10% (10 of 100) of trials during both the calibration (Cal) and validation (Val) periods***

Figure 4.10 below shows a comparison between model performance during calibration and the average run time of the calibration. As model complexity increased for Scenario2 1-3, so did the average time for a single calibration run. The calibration of Scenario 1 took 4.71 hours per trial on average, while Scenarios 2 and 3 took 9.86 and 35.71 hours, respectively. The difference in run time between Scenario 2 and 3 is large with only a marginal (0.024) increase in NSE. Scenario 1-P took 7.04 hours per run - 49% longer to than Scenario 1 - and Scenario 2-P took 12.40 hours per trial, which is 26% longer than Scenario 2. Therefore, there was a tradeoff between model performance, which increased with more GRUs and by changing the runoff algorithm, and run time, which also increased with additional model complexity.



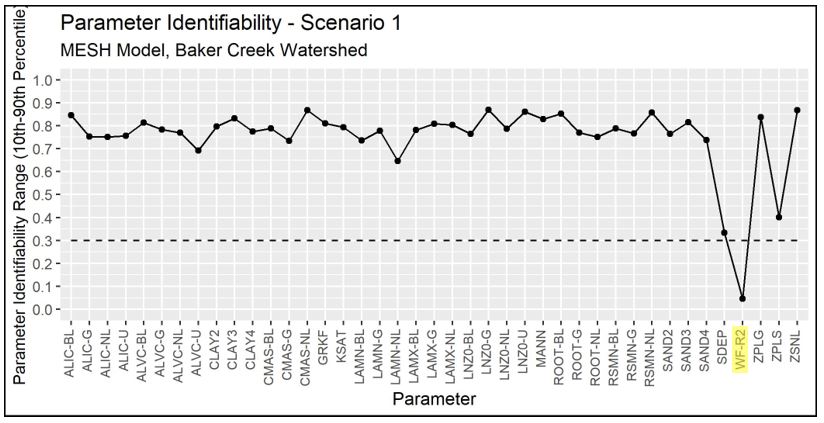
***Figure 4.10 - Comparison of the model performance and calibration run time for each scenario.***

Identifiability is a measure of the degree to which the true value of a parameter can be identified from a model (Gábor, Villaverde, & Banga, 2017). The value of a parameter may not be uniquely determined due to insensitivity of the outputs to the parameter, parameter interdependence, and poor data quality (Gábor, Villaverde, & Banga, 2017).

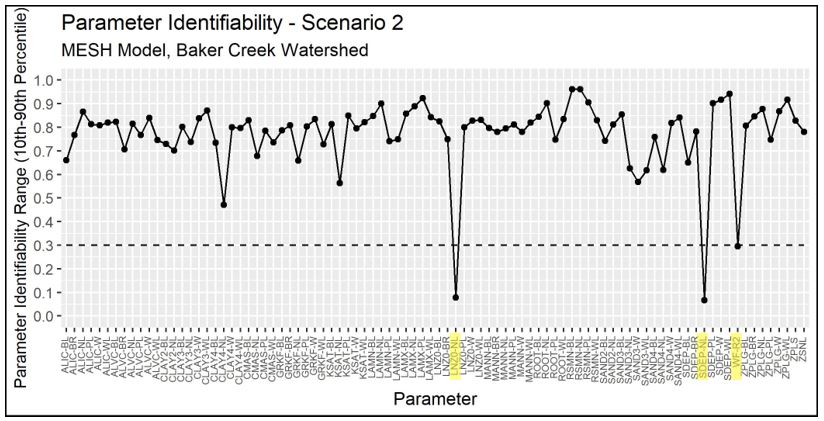
The same methodology as Mkandla 2017 was used, considering a parameter to be identifiable if the differece between the 10th and 90th percentile of the normalized parameter range is less than 0.3 using a normalization range of 0 to 1. A parameter identifiability range of 0 would mean the parameter is perfectly identifiable, and a value of 1 would mean the parameters is completely unidentifiable. The normalized values are calculated by:

where *max* and *min* is the range of the calibrated values for a given parameter, *max'* and *min'* is the normalized range, and *value* is the parameter value being converted.

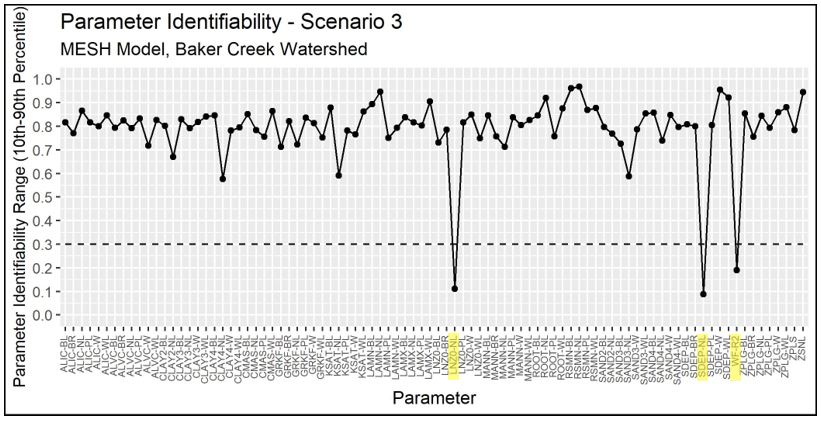
Plots of the parameter identifiability for all scenarios are shown in Figure 4.11 through Figure 4.15. WF-R2, which is a channel roughness factor combining channel shape, width to depth ratio, and Manning's n, are identifiable in Scenarios 1, 2, and 3. In Scenarios 2, 3, and 2-P, natural log of roughness height (LNZO-NL, which corresponds to tree height) and permeable depth soil (SDEP-NL) are identifiable for the needleaf/conifer GRU. In Scenario 1-P, the identifiable parameters are minimum leaf area index of needleaf forest (LAMN-NL), and the limiting snow depth to consider 100% coverage (ZSNL). It is noted that LNZO-NL and LAMN-NL are is just above 0.3 in Scenarios 1-P and 2-P, respectively.



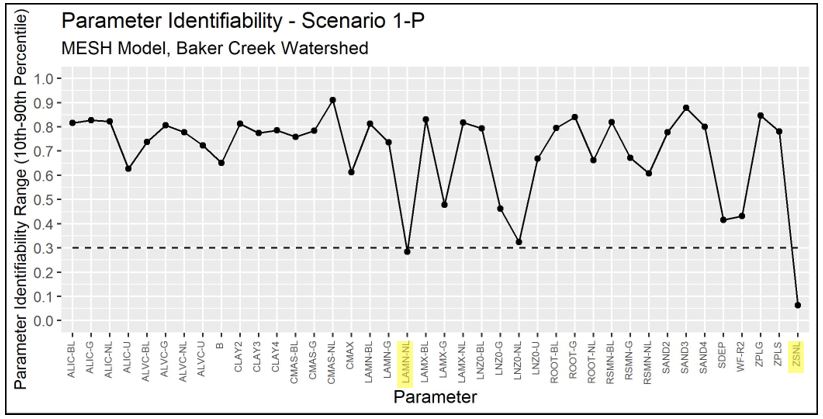
***Figure 4.11 - Identifiability of calibrated parameters for Scenario 1***



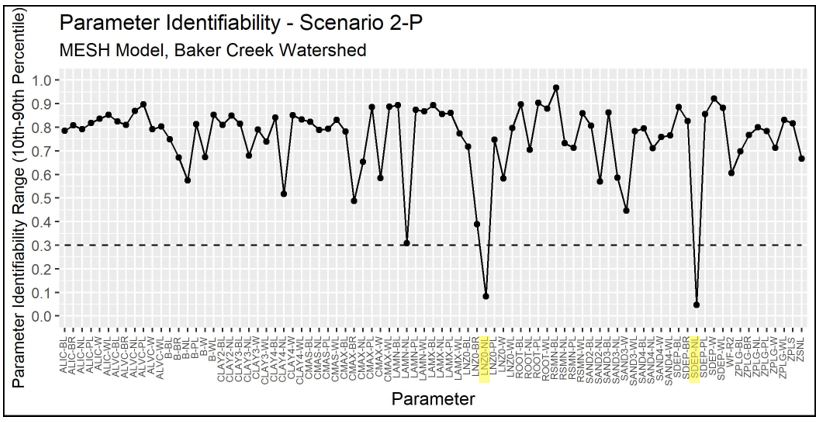
***Figure 4.12 - Identifiability of calibrated parameters for Scenario 2***



***Figure 4.13 - Identifiability of calibrated parameters for Scenario 3***



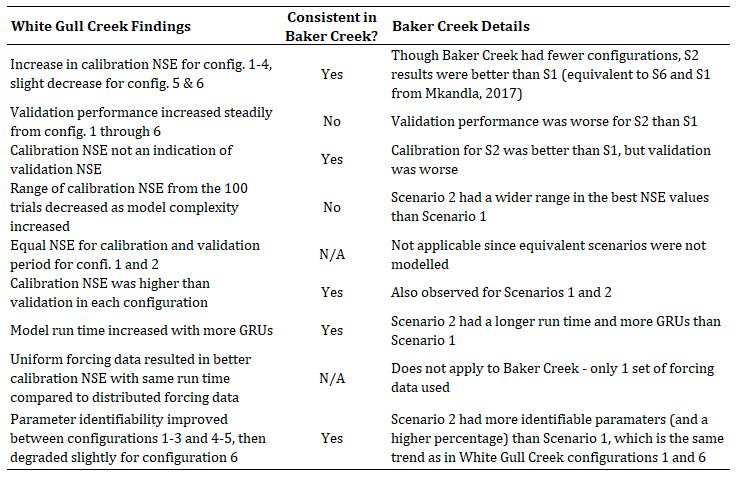
***Figure 4.14 - Identifiability of calibrated parameters for Scenario 1-P***



***Figure 4.15 - Identifiability of calibrated parameters for Scenario 2-P***

## Discussion and Conclusion

The objective of Phase 1 (Scenarios 1 and 2) was to replicate the modelling methodology from the White Gull Creek basin (Mkandla, 2017) and compare the conclusions. A summary of the comparison of the White Gull Creek and Baker Creek MESH modelling is presented in Table 4.5 below.

***Table 4.5 - Comparison of results between White Gull Creek and Baker Creek MESH Modelling***  


It is also noted that all the NSE results from the Baker Creek calibration were much lower than those for White Gull Creek; possible improvements are discussed below and in the *Towards a Solution* section. Regardless, the trends observed between the calibration and validation periods, and between scenarios, can provide useful information about the behavior of the MESH model in the watershed.

In all Scenarios, the model under-estimates most peak flows during the wet period (2008-2012) and over-estimates peak flows during the dry period (2013 onward). This is likely because the calibration was trying to match both high and low streamflow periods but did not do very well at matching either of them. Also, the use of NSE as the performance metric tended to steer the calibration toward matching peak flows rather than low flows.

The very poor performance (Nash-Sutcliffe values near zero) could be due to the calibration period including both a wetter-than-average and drier-than-average flow periods. Ideally, the model should be able to represent the hydrological processes in the basin so that the streamflow could be estimated in both wet and dry periods, but there are still some improvements to be made. Additionally, calibrating to streamflow when streamflow is near or at zero is not ideal, and it would be better to calibrate to another variable such as evapotranspiration or snow depth.

Bedrock is also not well-represented by the MESH model. MESH (i.e. CLASS) considers bedrock to be impermeable and runoff proceeds directly to the stream. However, as discussed in Spence, 2006, bedrock areas often retain water in fractures, soil-filled zones, and local depressions before it is transmitted to the stream or evaporated. The amount of runoff generated by bedrock areas depends on the available storage, evaporation demand, and rainfall amount and intensity (Spence & Woo, 2002). Phillips, Spence, & Pomeroy (2011) also stated that the connectivity of bedrock areas is controlled by peatlands and wetlands in the upland areas.

The MESH model configurations may also be limited in by the represenation of soil thermal properties and frozen soil representation appropriately. The basin is in a zone of discontinuous permafrost and the presence of seasonally frozen soils can vary spatially and temporally depending on soil moisture conditions (Spence, 2006; Elshamy et al., 2019).

The under-representation of fall peak flows indicates that the hydrological process(es) most influencial in the release of water at that time (i.e. storage and release of water from the lakes) are not well-represented in MESH. In general, it seems that the configurations of the model were not able to capture the storage and release behavior of the large lakes in the basin.

For the Scenario 1 and Scenario 2 configurations, the PDMROF runoff algorithm led to better results than the WATROF algorithm.

It is interesting that Scenario 1-P performed better than Scenario 2-P, even though the performance of Scenario 1 was worse than Scenario 2. This could potentially be because the 1-P model was better able to constrain the contributing area of the whole basin with only one value of CMAX and B while the segmented contributing area was segmented by GRU (landcover type) in 2-P. It could also be that due to the interaction between the larger number of calibration parameters in Scenario 2-P, the calibration was not able to converge on a better solution within the 1000 iterations.

Depending on the application, the marginal improvement in performance between Scenarios 2 and 3 may not be worth the 262% increase in calibration time.

# 5. Summary of findings

The project was completed in three phases; Phase 1 consisted of Scenario 1 and 2 which replicated the methodology of the MESH modelling in the White Gull Creek watershed (Mkandla, 2017), Phase 2 consisted of Scenario 3, which added complexity by increasing the number of grids in the basin from 1 to 5, and Phase 3, which repeated the calibration of Scenarios 1 and 2 but used the PDMROF runoff algorithm instead of WATROF. A summary of the main findings from the project are discussed below.

Overall, the model performance was very poor for all scenarios, with a maximum optimal Nash-Sutcliffe Efficiency (NSE) value of 0.169 for Scenario 1-P, and minimum optimal NSE value of -0.082. However, meaningful observations can still be made, and there is hope that with a little more work on the model and a few small, key tweaks, that performance could be substantially improved.

In Phase 1, comparing the Baker Creek project to the White Gull Creek project, the relationship between model complexity and NSE value, as well as between model complexity and run time, was consistent in both projects. In both models, parameterizing the model using 1 GRU for each landcover type resulted in better NSE values but longer run times compared to using only 1 GRU for the whole basin. Additionally, the more complex configurations resulted in improved identifiability, although Mkandla (2017) found that there was some degradation in the most complex configration compared to the intermediary ones (which weren't replicated in Baker Creek). Both projects also found that calibration NSE does not give an indication of how well the model will perform during validation, and found that the NSE values were consistently higher during calibration than validation.

In Phase 2, there was some improvement in model performance of Scenario 3 in terms of NSE compared to Scenarios 1 and 2. However, the increase in NSE value of 0.024 came at a high computational cost of an additional 25.8 hours per trial, or 262% longer than Scenario 2. That said, were the NSE results better, there may be some applications where the cost is worth the improvement in performance.

In Phase 3, the main findings were that PDMROF performed better than WATROF, but took slightly longer to run (2.3 hours and 2.5 hours longer for Scenarios 1-P and 2-P, respectively). Given the hydrological characteristics of the Baker Creek Basin - the large number of lakes and dominance of large lakes along the main drainage path- this outcome is consistent with what was expected.

In general, calibrating to streamflow and using NSE as the objective parameter was not ideal given the low runoff ratios in the basin (Spence & Hedstrom, 2018). Additonally, calibrating streamflow to both a wetter-than-normal and drier-than-normal period pushed the limits of MESH's capabilities. The effect of the contrasting conditions during calibration and the use of the NSE was apparent in the underestimation of the majority of low-flow periods and overestimations of low flows - particularly during the spring peaks. All model scenarios also struggled to adequately represent the fall peaks / winter streamflow, with Scenario 1 doing the worst and Scenario 2-P doing the best.

# 6. Towards a solution

Overall, the MESH model performance was very poor for all the scenarios. That said, there were valuable trends and comparisons observed. Future recommendations for further research to improve understanding and outcome of the model are as follows:

* Explore the effect of calibrating to logNSE instead of NSE.
  + Given the prevalence of low flows in the basin, this would likely improve the overall NSE value.
* Revisit the parameters and consider modifying some of the values and/or expanding the parameter ranges to see if better results can be obtained. This may include swapping out which parameters were calibrated.
  + Concurrently, it would be a good idea to perform a sensitivity analysis to hone in on the most sensitive parameters. This may result in a reduction of the number of calibrations required, and thus allow for more calibration iterations in the same run time. Examples of calibration tools include VisId (Gábor, Villaverde, & Banga, 2017) and VARS (Razavi & Gupta, 2016).
* Improve the parameterization of soils in Scenario 1 to better represent an "effective" soil type for the entire basin.
* Consider adding a reservoir(s) to represent the behavior of the large lakes in the basin (see Yassin et al., 2019).
* Consider using the LATFLOW as another alternative to represent the fill-spill behavior of the basin (see Hosain, 2017).
* Since streamflow is such a small part of the water balance in this basin - especially during dry years - consider calibrating to a differnt metric, such as evaporation and/or snow depth since both are available for the basin for several years.
* Calibrate just to the spring peaks, just the fall peaks, and/or only the higher flow period. This may result in better performance, but may also provide more information on the key parameters for each of those dominant behaviors.
* Iteratively run the spin-up period to initialize the model (particularly the soil moisture and thermal properties) rather than only 1 year. The initial conditions were chosen based on the average soil moisutre and temperature measurements over a number of years, which may not have been representative of the conditions at the start of the modelling period. This may be particularly important due to the discontinuous permafrost in the area and the lack of deeper soil temperature measurements.
* Define GRUs by hydrological response in the watershed rather than by landcover type, since the same landcover in different areas of the basin can respond differently.
* Improve the parameterization of bedrock in the model, either explicitly or effectively.

# Reference List

Abdelhamed, M., Razavi, S., Elshamy, M., & Wheater, H. (2018). Assessing baseflow simulation using MESH in Upper Liard Sub-Basin. In Global Water Futures Inaugural Science Meeting. https://doi.org/10.13140/RG.2.2.30541.33764

Agriculture and Agrifood Canada (AAFC). (2019, May 27). National Ecological Framework. Retrieved from http://sis.agr.gc.ca/cansis/nsdb/ecostrat/index.html

Anis, M. R., Razavi, S., & Wheater, H. (2017). An integrated modelling framework for regulated river systems in Land Surface Hydrological Models, 19.

Changing Cold Regions Network. (2019a, August 10). Hydrological Models. Retrieved from http://ccrnetwork.ca/science/models/hydrological-models/index.php

Changing Cold Regions Network. (2019b, August 13). 14. Baker Creek, NWT. Retrieved from http://ccrnetwork.ca/science/WECC/sub-arctic-lowlands/baker-creek.php

Cook, A. (2019). Photographs in the Baker Creek Watershed. Northwest Territories, Canada.

Dingman, S. L. (2015). Physical hydrology. Progress in Physical Geography (3rd ed.). Long Grove, IL: Waveland Press Inc.

Elshamy, M. E., Princz, D., Sapriza-Azuri, G., Pietroniro, A., Wheater, H. S., & Razavi, S. (2019). On the Configuration and Initialization of a Large Scale Hydrological Land Surface Model to Represent Permafrost. Hydrology and Earth System Sciences Discussions, (May), 1–49. https://doi.org/10.5194/hess-2019-206

Environment and Climate Change Canada. (2014, October 29). Regional Deterministic Precipitation Analysis (RDPA-CaPA). Retrieved from http://collaboration.cmc.ec.gc.ca/cmc/cmoi/product\_guide/docs/lib/capa\_information\_leaflet\_20141118\_en.pdf

Environment and Climate Change Canada. (2019, May). Historical Data. Retrieved from Government of Canada: http://climate.weather.gc.ca/historical\_data/search\_historic\_data\_e.html

Environment and Climate Change Canada. (2019, August 13). Yellowknife A, Climate ID 2204100. Retrieved from Canadian Climate Normals 1981-2010 Station Data: http://climate.weather.gc.ca/climate\_normals/index\_e.html

Fish Crossing. (2019, June). Manning's n Values. Retrieved from http://www.fsl.orst.edu/geowater/FX3/help/FX3\_Help.html#8\_Hydraulic\_Reference/Flow\_Profiles.htm

Fortin, V., Roy, G., Stadnyk, T., Koenig, K., Gasset, N., & Mahidjiba, A. (2018). Ten Years of Science Based on the Canadian Precipitation Analysis: A CaPA System Overview and Literature Review†. Atmosphere - Ocean, 56(3), 178–196. https://doi.org/10.1080/07055900.2018.1474728

Gábor, A., Villaverde, A. F., & Banga, J. R. (2017). Parameter identifiability analysis and visualization in large-scale kinetic models of biosystems. BMC Systems Biology, 11(1), 1–16. https://doi.org/10.1186/s12918-017-0428-y

Guan, X. J., Spence, C., & Westbrook, C. J. (2010). Shallow soil moisture - Ground thaw interactions and controls - Part 2: Influences of water and energy fluxes. Hydrology and Earth System Sciences, 14(7), 1387–1400. https://doi.org/10.5194/hess-14-1387-2010

Guan, X. J., Westbrook, C. J., & Spence, C. (2010). Shallow soil moisture - Ground thaw interactions and controls - Part 1: Spatiotemporal patterns and correlations over a subarctic landscape. Hydrology and Earth System Sciences, 14(7), 1375–1386. https://doi.org/10.5194/hess-14-1375-2010

Haghnegahdar, A. (2015). An Improved Framework for Watershed Discretization and Model Calibration: Application to the Great Lakes Basin. University of Waterloo.

Haghnegahdar, A., Tolson, B. A., Craig, J. R., & Paya, K. T. (2015). Assessing the performance of a semi-distributed hydrological model under various watershed discretization schemes. Hydrological Processes, 29(18), 4018–4031. https://doi.org/10.1002/hyp.10550

Hossain, K, Elshorbagy, A., Bharath, R., Davison, B., & Wheater, H. S. (2016). A Comparative Study of the Runoff Generation Algorithms in MESH Hydrological Model, (May).

Hossain, Kamrul. (2017). Towards a Systems Modelling Approach for a Large-Scale Canadian Prairie Watershed. Master’s Thesis. University of Saskatchewan.

Kokelj, S. A. (2003). Hydrologic overview of the North and South Slave regions. Yellowknife, NWT, Canada.

Kouwen, N. (2014). WATFLOOD/ WATROUTE Hydrological Model Routing & Flow Forecasting System. Waterloo, Ontario.

Luo, Y., Arnold, J., Allen, P., & Chen, X. (2012). Baseflow simulation using SWAT model in an inland river basin in Tianshan Mountains, Northwest China. Hydrology and Earth System Sciences, 16(4), 1259–1267. https://doi.org/10.5194/hess-16-1259-2012

Mailhot, J., Bélair, S., Lefaivre, L., Bilodeau, B., Desgagné, M., Girard, C., … Qaddouri, A. (2006). The 15-km version of the Canadian regional forecast system. Atmosphere - Ocean, 44(2), 133–149. https://doi.org/10.3137/ao.440202

Matott, L. (2017). OSTRICH: An Optimization Software Tool, Documentation and User's Guide, Version 17.12.19. 79 pages. University of Buffalo Center for Computational Research. Retrieved from www.eng.buffalo.edu/~lsmatott/Ostrich/OstrichMain.html

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, 105–116. https://doi.org/10.1016/j.jhydrol.2014.01.020

Mengistu, S. G., & Spence, C. (2016). Testing the ability of a semidistributed hydrological model to simulate contributing area. Journal of the American Water Resources Association, (52), 4399–4415. https://doi.org/10.1002/ 2016WR018760

Mkandla, H. (2017). Representing Spatial Heterogeneity for the White Gull Creek Watershed using the MESH model. University of Saskatchewan, Global Institute for Water Security. Saskatoon: School of Environment and Sustainability.

Morse, P. D., Wolfe, S. A., Kokelj, S. V., & Gaanderse, A. J. R. (2016). The Occurrence and Thermal Disequilibrium State of Permafrost in Forest Ecotopes of the Great Slave Region, Northwest Territories, Canada. Permafrost and Periglacial Processes, 27(2), 145–162. https://doi.org/10.1002/ppp.1858

Nash, J. E., & Sutcliffe, J. V. (1970). River Flow Forecasting Through Conceptual Models Part I: A Discussion of Principles. Journal of Hydrology, 10, 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Phillips, R. W., Spence, C., & Pomeroy, J. W. (2011). Connectivity and runoff dynamics in heterogeneous basins. Hydrological Processes, 25(19), 3061–3075. https://doi.org/10.1002/hyp.8123

Pietroniro, A., Fortin, V., Kouwen, N., Neal, C., Turcotte, R., Davison, B., … Pellerin, P. (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. https://doi.org/10.5194/hess-11-1279-2007

Razavi, S., & Gupta, H. V. (2016). A new framework for comprehensive, robust, and efficient global sensitivity analysis: 1. Theory. Water Resources Research, 423–439. https://doi.org/10.1002/2015WR017559.

Spence, C. (2006). Hydrological processes and streamflow in a lake dominated watercourse. Hydrological Processes, 20, 3665–3681. https://doi.org/10.1002/hyp

Spence, C. (2019, June 10). Baker Creek Conditions and Parameter Selection Meeting. (H. Brauner, Interviewer)

Spence, C., Guan, X. J., Phillips, R., Hedstrom, N., Granger, R., & Reid, B. (2010). Storage dynamics and streamflow in a catchment with a variable contributing area. Hydrological Processes, 24(16), 2209–2221. https://doi.org/10.1002/hyp.7492

Spence, Christopher, & Hedstrom, N. (2018). Hydrometeorological data from Baker Creek Research Watershed, Northwest Territories, Canada. Earth System Science Data, 10(4), 1753–1767. https://doi.org/10.5194/essd-10-1753-2018

Spence, Christopher, & Woo, M. (2008). Hydrology of the Northwestern Subarctic Canadian Shield. In M. Woo (Ed.), Cold Region Atmospheric and Hydrologic Studies (pp. 235–256). Hamilton, ON: Springer.

Spence, Christopher, & Woo, M. K. (2002). Hydrology of subarctic Canadian shield: Bedrock upland. Journal of Hydrology, 262(1–4), 111–127. https://doi.org/10.1016/S0022-1694(02)00010-0

Studio Team. (2018). RStudio: Integrated Development for R. Boston, MA: RStudio, Inc. Retrieved from http://www.rstudio.com/

Tolson, B. A., & Shoemaker, C. A. (2007). Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. Water Resources Research, 43(1), 1–16. https://doi.org/10.1029/2005WR004723

University of Saskatchewan. (2019a, August 7). About MESH. Retrieved from MESH - A Community Hydrology-Land Surface Model: https://wiki.usask.ca/display/MESH/About+MESH

University of Saskatchewan. (2019b, August 12). Plato. Retrieved from Service Desk (usask account required for access): https://jira.usask.ca/servicedesk/customer/kb/view/1548517815?q=Plato&q\_time=1565636151297

Verseghy, D. (2012). CLASS – The Canadian Land Surface Scheme (version 3.6) - Technical Documentation. Internal Report, Climate Research Division, Science and Technology Branch, Environmental Canada, (February).

Water Survey of Canada. (2019, May 13). Historical Hydrometric Data Search. Retrieved from Government of Canada: https://wateroffice.ec.gc.ca/search/historical\_e.html

Yassin, F., Razavi, S., Elshamy, M., Davison, B., Sapriza-Azuri, G., & Wheater, H. (2019). Representation of Water Management in Hydrological and Land Surface Models. Hydrology and Earth System Sciences Discussions, (January), 1–35. https://doi.org/10.5194/hess-2019-7

# Appendices

[Appendix A - Project Plan](Appendices/Appendix%20A%20-%20MWS%20Capstone%20Project%20Plan.pdf)

[Appendix B - Monthly Progress Reports](Appendices/Appendix%20B%20-%20Monthly%20Project%20Updates.pdf)

[Appendix C - Additional Tables and Figures](Appendices/Appendix%20C.pdf)

[Appendix D - R Code](Appendices/Appendix%20D.pdf)