Representing Spatial Heterogeneity for the White Gull Creek Watershed using the MESH model

Herbert Mkandla
Global Institute for Water Security
School of Environment and Sustainability
University of Saskatchewan

A project submitted to the College of Graduate and
Postdoctoral Studies for the degree of
Master of Water Security
August 28, 2017

Table of contents

Acknowledgements

Thanks to all the people that supported me especially my classmates in the Water Security program as we broke new ground. I would also like to thank Daniel Princz for his unreserved support and hours spent on helping me understand the many facets of hydrological modeling even when all chips were down he kept the spirits up. Thanks to my project partner, Dr. Bruce Davison of Environment and Climate Change Canada for providing me with all the data for this project and for the guidance as I embarked on this amazing journey. Thank you to my supervisor and program coordinator, Assistant Professor Andrew Ireson for his support and guidance in reviewing, commenting, and providing me with valuable feedback throughout the course of my studies.

I would also like to thank my wife and family for encouraging me to go back to school away from home. I would not have done this without their support and encouragement.

Dedication

I dedicate this research project to my family and my late parents who worked hard to raise me and my siblings to be the people we are today.

I also dedicated this research project to my wife and unborn children. Thank you for encouraging me to go back to school and supporting me all the way. I love you so much.

Abstract

The representation of spatial sub-grid variability in a model has important effects on its performance and efficiency. However, there have been relatively few studies that consider the impact of different ways in which spatial heterogeneity can be represented. In this work, a systematic approach is used to test several ways of representing spatial sub-grid variability in a model. The methodology uses the White Gull Creek watershed located in the southern edge of the Canadian boreal forest, a significant biome for the earth's carbon cycle, to assess the effects of different model setup configurations in the MESH (Modélisation Environnementale Communautaire – Surface and Hydrology) model (Pietroniro et al., 2007). 1999 – 2014 data, obtained from two observation sites located within the basin were utilized to drive the model. An optimization and calibration software; OSTRICH (Matott, 2005), employing the dynamically dimensioned search (DDS) algorithm (Tolson & Shoemaker, 2007), was utilized to calibrate the model parameters. The basin was subdivided into ecodistricts based on available information (Agriculture and Agri-Food Canada, 2017) and/or landscape units based on vegetation cover types (Natural Resources Canada, 2017). The effect of averaged forcing and model parameter data was compared with simulations that used distributed forcing and parameter data. Results indicate that model validation performance increased with complexity albeit with higher computational costs in calibration. Also, simpler configurations show great variability in calibration, while the more complex configuration had good and consistent performance in calibration and better overall validation results.

1. Introduction

Variability of hydrological processes occurs over a wide range of spatial and temporal scales. These processes vary locally, regionally, and globally, from rain drops to hydrological changes over centuries due to variations in climate and geomorphic processes (Skøien, Blöschl, & Western, 2003). Over large geographic regions, spatial variability is influenced by climate, geomorphology, evapotranspiration and ground-water presence. Variability with time governs seasonality and other cycles such as the carbon cycle. Generally, spatial heterogeneity of a watershed is reflected by topography, snowpack, soil, and vegetation which play an important role in the structure of modeling frameworks

In hydrological modeling, spatial variations in landscape impact the analysis and representation of hydrological processes in model simulations. Thus, proper analysis and representation ensures that prediction of hydrological systems' response to natural and anthropogenic changes produces consistent understanding. Also, our ability to assess future conditions while capturing the physics of the watershed adequately depends on how well spatial sub-grid variability can be represented in a model. Since spatial variability is based on point-scale measurements such as rain gauges (Dingman, 2015, p. 23), it is a challenge to collect measurements over large areas to capture all aspects of landscape variability to accurately represent in a model. Fortunately, streamflow measurements can provide integral information on the response of hydrological processes occurring in the watershed (Sciuto & Diekkrüger, 2010). Hence, the hydrologic modeling community tends to target these measurements to test their models.

Spatial heterogeneity results in spatial and temporal variability of hydrological processes which can have local, regional, or global scale influence. Incorporating such variability in models

requires careful discretization of the watershed into appropriate computational units. These computational units are characterized by distinct hydrological responses based on either vegetation type or other physical properties of the watershed such as slope or soil moisture. Therefore, complex model structures are needed to reflect the changing hydrological response of the surface to different climatic conditions, stage of land cover development, and soil moisture availability, whereas simpler structures are required for faster calibration (Franks et al., 1997).

Recognizing that hydrological processes and their physical properties change spatially over the watershed is central for analyzing the effect of spatial heterogeneity in hydrological models. Environment and Climate Change Canada's MESH (Modélisation Environnementale Communautaire – Surface and Hydrology) model (Pietroniro et al., 2007) is applied to the White Gull Creek watershed located on the southern edge of the boreal forest in central Saskatchewan, Canada. The methodology uses a stepwise modification of the MESH model configuration structure to represent spatial variability of a watershed at the sub-grid level. Although, vertical variation of sensible and latent heat fluxes plays a significant role in the hydrological characteristics of a watershed, it is not considered in detail in this study. The degree of spatial heterogeneity and corresponding model complexity, along with how they affect model simulation of streamflow, is the primary focus of this work. The model is calibrated to the observed streamflow data to find a good fit for the model parameters of each configuration. Furthermore, parameter identifiability is assessed for each calibration run.

Representation of hydrological processes at the sub-grid level is assessed by comparing several ways of including spatial sub-grid heterogeneity in a model. Using available data, various model arrangements of complex hydrological properties of the watershed are discretized and their behavior is analyzed. Through optimization and calibration, model variables are adjusted against observed streamflow data to reflect the hydrological response of the system (watershed). This work examines a realistic approach to representing land cover heterogeneity in the model. The Canadian Land Surface Scheme (CLASS) is used within the hydrological model – MESH in this study.

2. Objective of the Study

The overall objective of this study is to perform a systematic investigation of how spatial sub-grid variability can be optimally represented in a case study basin, White Gull Creek, with a case study model, MESH. The idea is to balance physical realism with computational expense. The level of complexity of the MESH model is varied from a simple single watershed scale GRU, through to a landcover dependent discretization with up to ten GRUs.

3. Literature review

Models are important for a variety of reasons in different scientific disciplines, including prediction of future climate change and water resources planning and management. However, due to increasing complexity of environmental change and water-related problems, spatially distributed descriptions of water fluxes and state variables are necessary (Sciuto & Diekkrüger, 2010) for successful hydrologic modeling. These problems are influenced by land-use changes due

to natural and human activities. Hydrologic models have been around for decades and are widely used to analyze these water-related problems (Loucks, Beek, Stedinger, Dijkman, & Villars, 2005).

Hydrological models must be able to deal with hydrological processes occurring across the watershed. These processes are active at different locations (spatial scale) and at different times (temporal scale) (Bierkens et al., 2014). Hydrologic models then, ought to be flexible if different scales are to be considered. The use of hydrological models in assessing the impacts of climate and environmental changes can be limited due to resource limitations, narrow range of measurement systems, and inadequacies of existing data both spatially and temporally (Pechlivanidis, Jackson, Mcintyre, & Wheater, 2011). Despite such drawbacks, hydrological models provide a way to organize complex processes which influence the hydrologic cycle at different scales. Hydrological modeling creates a platform to improve our basic understanding of hydrological processes, validate scientific hypothesis, and increase our abilities to predict and forecast likely impact of future changes (Sciuto & Diekkrüger, 2010), albeit with significant uncertainties.

Traditionally, hydrological models were applied at a watershed scale, adopting an approach somewhere on a spectrum from conceptually based to physically based (Melsen et al., 2016). Their applications vary depending on the problem being considered. In practice, analytical models are used for limited situations, whereas numerical models are widely used for real situations (Sciuto & Diekkrüger, 2010). Models can be classified based on model structure, distribution, and application. In conceptual models, the structure is predefined and not all model parameters are independently measurable, some must be estimated through calibration (Vrugt, Gupta, Bastidas, & Bouten, 2003). Conceptual models can be run as lumped or distributed. Lumped models assume homogeneity of the watershed which is expressed by differential and empirical equations and does not consider spatial variability of the processes (Pechlivanidis et al., 2011). Distributed models on the other hand, incorporate heterogeneity of the physical characteristics of the watershed by partitioning it into homogeneous subunits and solving equations associated with properties of each element (Arabi, Govindaraju, Sophocleous, & Koelliker, 2005). Conceptual models are considered to be more robust and require less data than fully physically-based models (Wada, Graaf, & Beek, 2016).

In physically-based models, all components of hydrological processes use physical governing equations for motion based on continuity, and energy and mass balance. These equations can be solved numerically or analytically (Wheater et al., 1993). The model structure is generally based on small-scale field studies, hence, can be affected by the nature of the experiments themselves. Physically-based models are better at representing landscape and subsurface heterogeneity, plus some parameters can be measured and used without calibration (Beven, 2001; Pechlivanidis et al., 2011). In practice, this cannot be achievable at the scale of interest because such measurements are essentially observed at a point (Dingman, 2015, pp 23; Wheater, 2002). Hybrid models combine elements of other types of models. They are developed to exploit strengths of conceptually-based and physically-based models to deal with problems associated with parameter identifiability among others. They also improve representation of physics-based processes where physical measurements are difficult to obtain (Pechlivanidis et al., 2011).

In hydrological applications, there is a wide range of uncertainty associated with a model's ability to represent hydrological processes and to replicate observed data. A model's capability to facilitate up or downscaling of watershed processes and how realistic model results are, determines the level of confidence we have in them. Sources of uncertainty include: natural uncertainty arising from random effects of physical processes; data uncertainty due to inadequate quality of data caused by systematic errors in measurement systems and techniques; model parameter uncertainty due to use of limited and uncertain observed data (Pechlivanidis et al., 2011). Data and model parameter uncertainties are due to difficulties in obtaining quality historical data to enable meaningful comparison between present and future climate. This is further exacerbated by the sparsity of measurements, both spatially and temporally (Remesan & Holman, 2015). Similarly, magnification of small errors over time and space leads to another cause of uncertainty in modeling (Oreskes & Belitz, 2001). Lastly, model structure uncertainty is influenced by our understanding of basic hydrological processes, which is determined by the data available (Pechlivanidis et al., 2011).

Modeling of hydrological processes has gradually improved and included the incorporation of interactions between landscape and the lower layers of the atmosphere through Land Surface Schemes (LSSs). LSSs have progressed from the conceptual representation of the bucket model of Manabe (1969), to sophisticated schemes that incorporate surface energy and water fluxes interactions with hydrological processes such as the Canadian Land Surface Scheme (CLASS; Verseghy, 1991; Verseghy, McFarlane, & Lazare, 1993); the Variable Infiltration Capacity model (VIC; Liang, Lettenmaier, Wood, & Burges, 1994); the Interactions between Soils, Biosphere, and Atmosphere (ISBA; Noilhan & Mahfouf, 1996; Noilhan & Planton, 1989); or the new Soil, Vegetation, and Snow (SVS; Alavi et al., 2016; Husain et al., 2016) developed at Environment Climate Change Canada for the Global Environmental Multiscale (GEM) atmospheric model.

Land Surface Schemes generally provide large-scale boundary conditions between the land surface and atmosphere to general circulation models (GCMs) or numerical weather prediction (NWP) models. Their emphasis is usually on surface conditions and vertical fluxes, and providing hydrological inputs for flood forecasting. Local processes and surface heterogeneity are generally not assessed depending on the objectives of the investigation being done or are simplified and lumped (Dornes, Pomeroy, Pietroniro, & Verseghy, 2008). Different techniques are utilized to incorporate spatial variability in land surface schemes such as using dominant land cover types to address spatial varying processes over local scales, or by sub-dividing the grid into appropriate units. Responses from these units are then calculated and applied as grid outputs reflecting the land surface variability (Dornes et al., 2008; Verseghy et al., 1993).

Models should be able to describe physical processes in a credible way at a sub-basin scale. Locally varying characteristics of these processes need to be reasonably replicable by the model (Kreye & Meon, 2016; Melsen et al., 2016) to capture any meaningful information. Because of the challenges of measuring nonphysical attributes of the watershed, conceptual parameters are always used where there are no direct measurable properties. However, these are not appropriate to represent spatial sub-grid variability in the real world due to variations in the landscape and the non-linearity nature of the processes involved (Avissar, 1991; Melsen et al., 2016). Also, hydrological processes occur at different time scales and compatibility of these time scales is

problematic to represent (Kampf & Burges, 2007) and have consistent representation of state variables in models.

To improve representation of spatial heterogeneity in hydrological modeling, several approaches that consider watershed lateral variability and discretization were suggested such as; the Hydrological Response Units (HRU; Leavesley, Lichty, Troutman, & Saindon, 1983). This approach considers slope, aspect, land use, vegetation, precipitation distribution, etc., to partition a watershed into homogeneous subunits. The Representative Elementary Areas (REA; Wood, Sivapalan, Beven, & Band, 1988), showed the effect of incorporating spatial scales for catchment responses for runoff generation. The Representative Elementary Watersheds (REWs; Reggiani, Sivapalan, & Hassanizadeh, 1998) approach subdivides the watershed into smaller units where similar physical governing equations operate. Another approach that deal with sub-catchment variability is the 'hydro-landscape' units (Winter, 2001) which takes into account the geophysical features (i.e. topography, land use, soil, etc.) of watershed with a nested discretization of the catchment (Dehotin & Braud, 2008). In the same spirit, the MESH model (Pietroniro et al., 2007) implemented the Grouped Response Unit (GRU; (N. Kouwen, Soulis, Pietroniro, Donald, & Harrington, 1993), to subdivide the watershed into areas with similar hydrological characteristics based on land cover types (see Model Configuration section).

To produce physically realistic model simulations, parameterization of physical processes must be done accurately (Avissar, 1991), however, it has proved a challenge to eliminate uncertainties in the process. This is essential in selecting parameter values, setting boundary and initial conditions in a model (Refsgaard, 1997). When parameterization is done correctly, parameters should be transferable to other watersheds with similar physical and hydrological attributes. Parameter transferability is a necessary exercise that can be used as a measure of the degree to which spatial and temporal variability are represented in the model (Melsen et al., 2016). This can benefit the development of hydrological modeling in areas that currently do not have observational records or have inadequate records. In good models, parameters should be largely transferable over time, because longer time steps are simply extended shorter time steps. In contrast, parameters should not be transferable spatially because land cover and soil characteristics differ from place to place (Melsen et al., 2016).

Traditionally, scale variability makes the systematic analysis of hydrological processes problematic, whether its model structure, measurement techniques, or process understanding. This presents major challenges in reducing the inconsistences between model and data scales in hydrological modeling (Blöschl & Sivapalan, 1995; Kreye & Meon, 2016; Skøien et al., 2003). Unfortunately, models often represent processes at different spatial scales than the observed scales. A process can be conceptual at one scale and physically based at another scale (Dehotin & Braud, 2008). This affects how hydrological processes can be viewed and understood. Often short time-scale is used for observation and modeling but longer time-scale is required for prediction and forecasting of the behavior of the infrastructure design. Similarly, theories and concepts are developed in small-scale lab experiments and are expected to work at large-scale of real world (Blöschl & Sivapalan, 1995; Melsen et al., 2016). To address such conflicts, appropriate model structures are required.

The appropriate level of complexity required to represent spatial sub-grid variability in hydrological models is an unresolved issue that is critical to modeling. It is essential to compare and validate alternative modeling systems against simpler available models (Wood, Lettenmaier, & Zartarian, 1992) to determine their computational feasibility in representing spatial heterogeneity. Although, efforts to represent spatial sub-grid variability and incorporate appropriate levels of model complexity have been slow, due in part to the intensive computational resource requirements. The last few decades have seen advances in computational power that has led to improvements in hydrological processes modeling. Similarly, improvements in information technology, data availability and data storage capabilities has fueled rapid development of more complex models which incorporate spatially varying parameters such as; land cover, soil moisture, soil types, and variable driving data. This improvement has created opportunities in the development of hyper-resolution global hydrological models (Melsen et al., 2016).

Hyper-resolution could improve representation of spatial sub-grid heterogeneity in models and offer better initial boundary conditions for regional models. Modeling and prediction of human impact on watersheds require fine spatial scale representation of water-pathways, irrigation, inter-basin water diversion, and reservoir management. Flood and drought management also need high resolution, however, spatial scales used in current models are insufficient to provide adequate locally relevant hydrological information globally (Bierkens et al., 2014; Wada et al., 2016). Recent improvements in computational capabilities and data storage means hyper-resolution modeling is feasible for larger domains and longer simulations (Bierkens et al., 2014). This can be constrained by the availability of data at fine spatial scales, which are not readily available (Sciuto & Diekkrüger, 2010). Availability of remote sensing data (satellite, drone, planes, etc.) is expected to close the data gap between field measurements resolution and modeling resolution (Dehotin & Braud, 2008). Hyper-resolution influences the choice of the level of discretization to handle spatial landscape heterogeneity (Dehotin & Braud, 2008). It affects how dominant hydrological processes are represented in a model. This will require further development in spatial resolution of hydrological and land surface models (Wood et al., 2011). (See Wood et al (2011) for detailed discussion).

This study, contributes to the discussion about the appropriate approach to representing spatial sub-grid heterogeneity. Representing spatial heterogeneity in a model provides an indication on how well spatial discretization of the watershed could impact the simulation of water fluxes and state variables at different scales. Advances in computational resources and availability of data, provides a platform to evaluate the influence of spatial landscape discretization on model results, and compare with effects of lumped and semi-distributed vegetation and soil parameters as well as meteorological forcing data.

4. Study Site/Area

a) Basin description

The White Gull Creek watershed (Figure 1) is in central Saskatchewan, 60Km to the North East of Prince Albert on the southern edge of the Canadian boreal forest, between 53.99° and 54.13°N latitudes and 104.62° and 105.08°W longitudes. The watershed comprises a drainage area of 629 km^2 with approximately 26 km^2 of that covered by the White Gull Lake, which was not actively

contributing to the main stream flow during the entire study period due to low levels of water in the lake (Barr et al., 2012; Davison et al., 2016), and is not considered in this study.

The boreal forest is in an important biome for the earth's carbon cycle, which is tightly linked with the local hydrological cycle. The White Gull Creek basin falls into two distinct ecodistricts; the Whiteswan Uplands ecodistrict and White Gull Plains ecodistrict. An ecodistrict is defined as the smallest unit of land characterized based on an ecological assemblage of regional landforms, relief, surficial geologic material, soil development, water bodies, vegetation, land use classes, range of annual precipitation, and mean annual temperature similarities, according to Agriculture and Agri-food Canada (Smith & Marshall, 1996). The basin is characterized by a continental climate, with a wide range of annual temperature, long severe winters and mild to warm summers, with approximately 30% of its annual precipitation falling as snow (Barr et al., 2012; Ireson et al., 2015).

Land cover type is diverse, and has wide reaching influence on the basin's hydrological processes due to topography, drainage, soil structure, along with varied regeneration history after natural and anthropological disturbances, which is influenced by site conditions and establishment method (Amiro et al., 2004). Accordingly, the effect of surface geology along with soil drainage and soil moisture availability is evident in the spatial distribution of vegetation. Vegetation cover type is relatively uniform forests of Black and White Spruce, Tamarack, Jack Pine, Old Aspen, interspersed with wetlands and lakes which reflect the heterogeneity of the landscape. Peatlands dominate the poorly drained deeper depressions with shallow groundwater table. Black Spruce forests and wetlands are found mostly in the low laying areas of the Whiteswan Uplands on the western side of the watershed. Jack Pine forests occupy the rapid draining sandy plains on the east with a considerably deeper groundwater table, while mixed wood and aspen forests are situated on the well-drained sandy loamy uplands (Amiro et al., 2004; Barr et al., 2012; Bartlett et al., 2006; Davison et al., 2016). The topography and the heterogeneous land-cover type presents unique hydrological characteristics for hydrological modeling and other studies.

White Gull Creek basin is located within one of the sites which was part of a large-scale intensive research experiment that began in the early 1990s as the Boreal Ecosystem–Atmosphere Study (BOREAS, Sellers et al., 1997). The influence of the Boreal forest on weather and climate conditions and its role as a carbon sink was believed to be significant. Therefore, to improve the understanding of the connections between biophysical processes and the conditions that govern energy, water, heat and atmospheric gases and their relationship with ecosystems, two research sites were chosen in Canada; North BOREAS near Thompson Manitoba and South BOREAS located North East of Prince Albert. Additionally, BOREAS was used to develop and improve process modeling over large areas by implementing latest technologies of the time (ORNL DAAC, 2017; Sellers et al., 1997). Meteorological, flux, soil moisture, and temperature data were collected from monitoring sites established within and around the White Gull Creek basin. Two of the monitoring sites are located within the basin, one in each ecodistrict. The data collected since, has been used to advances studies in hydrology and hydrological modeling. The work of BOREAS has continued to improve hydrological causes as part of the Boreal Ecosystem Research and Monitoring Sites (BERMS) and Fluxnet-Canada network since 1997, bringing together government and university scientists studying the effect of climate change on the Canadian forest ecosystems (ORNL DAAC, 2017).

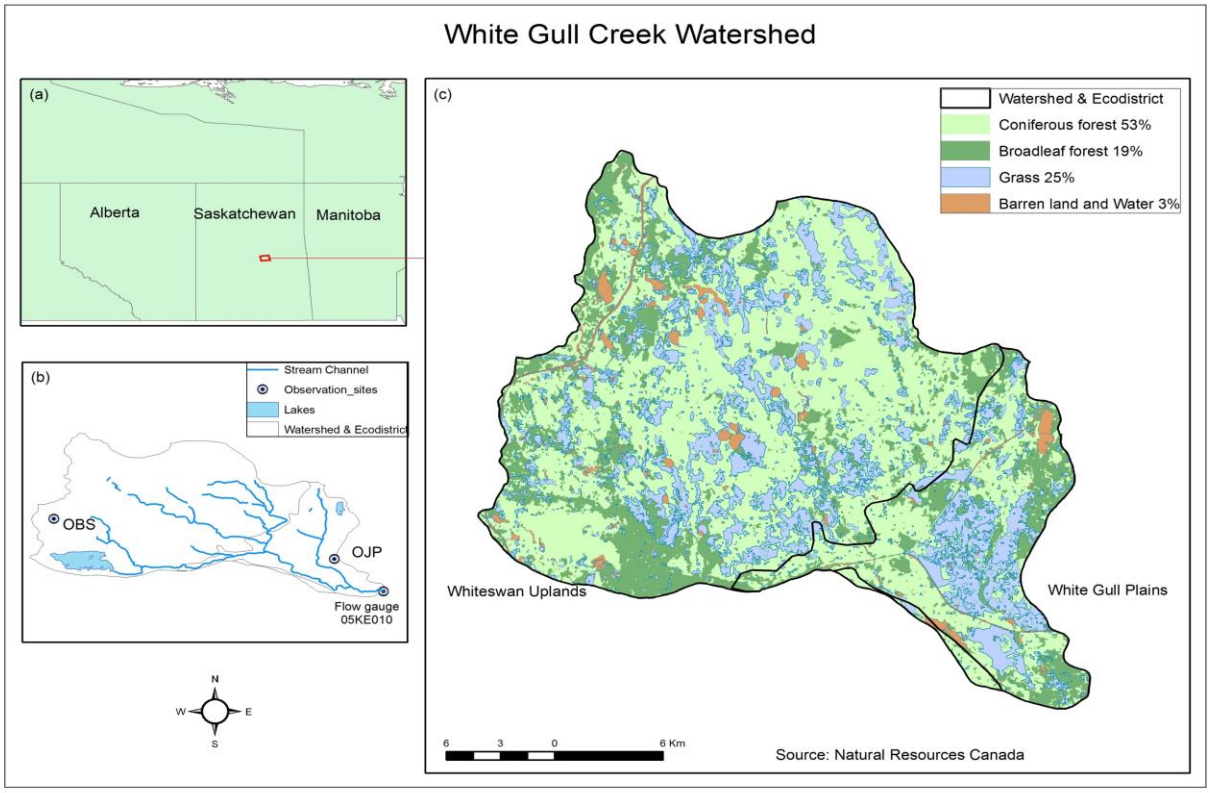

Figure 1. (a): Location of White Gull Creek watershed in Saskatchewan, Canada; (b): observational sites, stream channel network, and stream flow gauge; (c): ecodistricts and land cover types.

b) Basin data

Land cover and ecological data was required to set up the MESH model configurations, streamflow data to fit the simulated streamflow to, while meteorological data was needed to drive the model. The driving data utilized are total precipitation (rain and snow), wind speed, specific humidity, atmospheric pressure, temperature, and incoming longwave and shortwave solar radiation, all at 30-minute time intervals. This data was measured since 1994 at the two observational sites located within the boundaries of the White Gull Creek basin; the Old Black Spruce (OBS) and the Old Jack Pine (OJP), whereas the stream flow measurements used were observed since 1993 at the streamflow gauge number 05KE010 (Water Office Canada, 2017) at the basin outlet (Figure 1(b)).

The forcing data was used either as an average of the measurements from the two sites to provide uniform forcing data over the entire basin, or separated according to location of the observation site within the basin depending on the model setup configuration. For example, the Whiteswan Uplands on the west utilized data from the OBS observation site, while the OJP site measurements were employed on the White Gull Plains to the east of the basin.

Figure 1(c), shows the land cover map as reclassified from the Land Cover, circa 2000-Vector (LCC2000-V) database available at the Natural Resources Canada (2017) website, which was used

to set up different model configurations (see Model Configurations section for further discussion). Land cover types with similar features were regrouped into one of the four categories recognized by the CLASS model (described below). Four land cover types were utilized for most of the configurations whereas a fifth type; mixed forest, was added to represent areas with considerable composition of mixed wood and aspen forest.

5. Method

a) Model description

WATFLOOD (Kouwen et al., 1993), is a deterministic, semi-distributed flow model which utilizes the grouped response unit, GRU, approach (Kouwen et al., 1993; Soulis et al., 2005; Comeau et al., 2009). A GRU may be defined as an area, or collection of areas, of land within some domain for which the properties (i.e. parameters) are considered uniform. GRUs may be used to represent sub-grid scale heterogeneity, in which case as well as having uniform parameters, the GRU also has uniform states and fluxes. The area represented by a single sub-grid scale GRU does not have to be contiguous – the GRU represents a set portion of the grid cell, whether that land area is physically located as a single area, or fragmented over the grid cell. A GRU may also be used to distribute the parameters over multiple grid cells, to parsimoniously distribute the parameters over a watershed or larger regional domain. In this case, the states and fluxes of GRUs will vary in different grid cells, while the parameters will be treated as uniform over the entire domain, or else some subdomain, such as an ecodistrict. In WATFLOOD the outputs from each GRU within a single grid cell are summed together, since there is no horizontal exchange between sub-grid GRUs, and produce the lumped output of the grid cell. This output is then is channeled through the stream network, as the model does not allow horizontal transfer of water between grid cells other than through a stream channel (Kouwen et al., 1993; Kouwen, 2009; Haghnegahdar et al., 2014). This approach was implemented in the MESH model (Pietroniro et al., 2007).

The Canadian Land Surface Scheme (CLASS, Verseghy, 1991; Verseghy et al., 1993), is a land surface scheme widely used to simulate the vertical exchange of energy and water between the land and atmosphere through the soil and vegetation. CLASS solves the energy balance of the land cover and soil layers, which is coupled with the water balance (Dornes et al., 2008). This is implemented within various Canadian General Circulation Models (Haghnegahdar et al., 2014; Parviainen & Pomeroy, 2000). CLASS is used to provide lateral flow processes such as overland flow, interflow, groundwater flow, blowing snow redistribution, and so on, to hydrologic-land surface models including MESH. Being Canadian, there is a focus on representing snow and frozen land. Vertical exchanges are solved for each sub-grid scale GRU. The vertical movement of water through the soil layers is handled by the Richard's equation. Interflow out of a GRU is generated as a function of bulk saturation of each soil layer over a given time (Soulis et al., 2000), and baseflow comes from a unit gradient, or free drainage, boundary condition at the bottom of the soil column within each GRU (Haghnegahdar et al., 2014; Soulis et al., 2000). The version of CLASS used in this study recognizes five land cover classes: needleleaf forest, broadleaf forest, crop, grass, and barren land. Since there are no agricultural crops in this basin, the crop class was not considered. A single GRU includes all five land cover classes, each with a fractional coverage that sums to 100%, which over lie the soil profile. In this way, a GRU can be set to a single land cover class, by setting that land cover class to 100% coverage, and all other classes to zero. The MESH

model (Pietroniro et al., 2007) evolved from WATCLASS (Soulis et al., 2000), a hybrid that brought together the strengths of CLASS and the routing power of WATFLOOD to be coupled with any GCM. This linking of models eliminates the need to treat land surface parameters and routing schemes independently, resulting in inconsistent watershed state variables.

b) Model Configuration

Watershed discretization is an important aspect of model configuration. Decisions about how to discretize a watershed will depend upon the model structure, knowledge of the watershed (which includes published literature, technical documents and expert opinion), and data availability. In recognizing that in physical reality there is infinite variability and complexity that is impossible to deal with, the choice of model structure becomes vital in representing spatial sub-grid variability. Either a simple model structure, where everything is uniform (lumped model) – i.e. driving data and model parameters (soil and vegetation), or a complex model structure with a fine spatial resolution – where everything is distributed, can be used depending on the objectives of the investigation. As such, care should be exercised in balancing the physical realism and computational expense, see Table 1 for model structure comparison. Configurations developed here represent the major hydrological processes that affect stream flow as simple as considered possible, while maintaining a physically realistic conceptual basis.

Table 1.Pros and cons of simple vs complex model structures

Model structure	**Pros**	**Cons**
Simple	- Quick to setup and run - Cheap	- Not representative of physical reality - Simplify assumptions - Course resolution
Complex	- Fine resolution - Should be more realistic (physically) - Provides more detailed information	- Needs more data - Needs more computational resources (expensive) - Challenging to setup and run

In this study, we explore different ways to represent spatial heterogeneity in a single grid cell model – i.e. a model where there is no internal lateral redistribution of water, but with a heterogeneous vertical water balance using a sub-grid GRU approach. The basin was covered by two ecodistricts (described above in Section 4a). Driving data is either applied uniformly over the entire watershed (based on an area weighted average of the observed driving data), or is applied uniformly over each ecodistrict (based on the observations from that ecodistrict). Various GRU approaches are considered: the simplest case is to have a single GRU, with one set of land cover classes over the entire watershed; the next simplest case is to have two GRUs, one for each ecodistrict, but with the same parameters for the soils and landcover in both GRUs, i.e. uniform over the watershed; the next case is to have two GRUs where the parameters are different between the two ecodistrict; and the final case is to have one GRU for each land cover class (five), for each ecodistrict (two), resulting in ten GRUs. The model configurations that result from this are summarized in Table 2 below.

Table 2. Configuration set up used in MESH.

Configuration Setup	Forcing Data	Number of GRUs	Parameter Distribution	Number of Parameters
1	Uniform	1	Uniform	33
2	Uniform	2	Uniform	33
3	Distributed	2	Uniform	33
4	Uniform	2	Vary by ecodistrict	65
5	Distributed	2	Vary by ecodistrict	65
6	Distributed	10	Vary by Land Cover	92

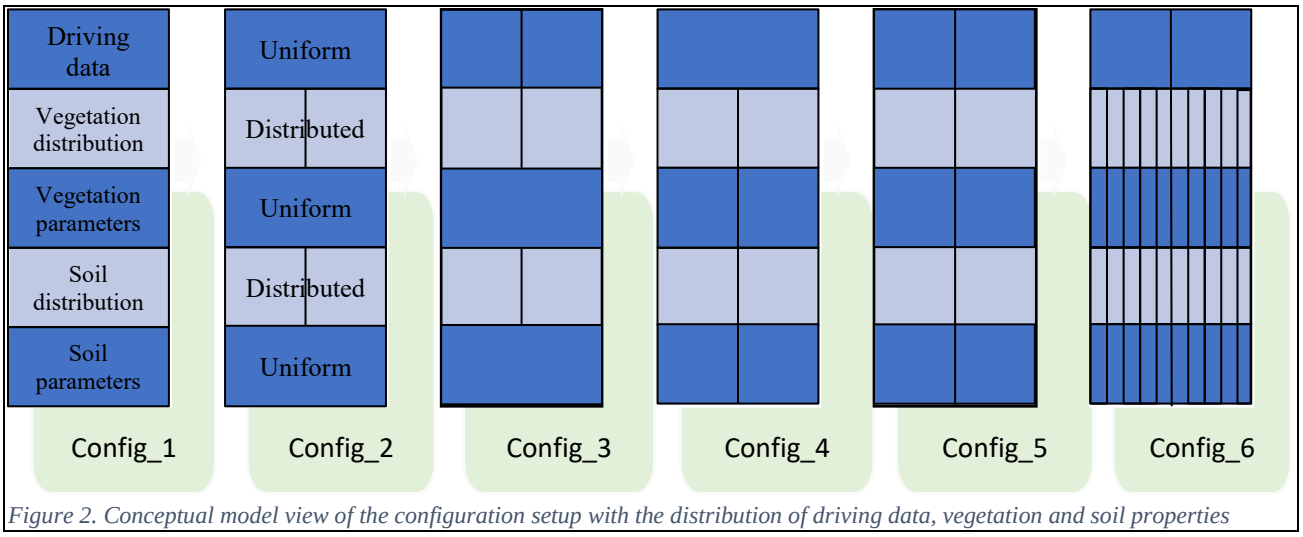

Figure 2. Conceptual model view of the configuration setup with the distribution of driving data, vegetation and soil properties

A conceptual layout of each configuration in the model is shown in Figure 2. Each layer (blue) represents the input (driving data) or model parameters (i.e. vegetation distribution refers to the physical location of the vegetation in the basin; vegetation parameters – e.g. leaf area index (LAI); soil distribution – e.g. location in the basin; soil parameters – e.g. soil texture). Each configuration will now be described in detail as follows.

Detailed description of each model configuration

Configuration 1 (Figure 3) – the entire watershed represents the model grid domain and is used to determine model parameter distribution. The watershed is kept as one unit and the driving data (i.e. precipitation, incoming longwave and shortwave radiation, specific humidity, atmospheric pressure, wind speed, and temperature) are averaged and applied uniformly over the entire watershed. These variables represent the average measurements observed from the two observation sites (OBS and OJP) located within the watershed and are used without considering the spatial location of landforms in the watershed. The model parameters (33) such as the vegetation information (e.g. albedo, stomata resistance, roughness length, leaf area index, etc.), soil texture, and so on, are distributed equally across the land cover types. This means each land

cover type (coniferous forest, broadleaf forest, grass, and barren lands & water) gets the same model token for a given parameter with its value distributed according to the fractional contribution of the total land area of that cover type to the overall watershed area. The hydrological response from each land cover class is calculated and contributes to the main stream flow.

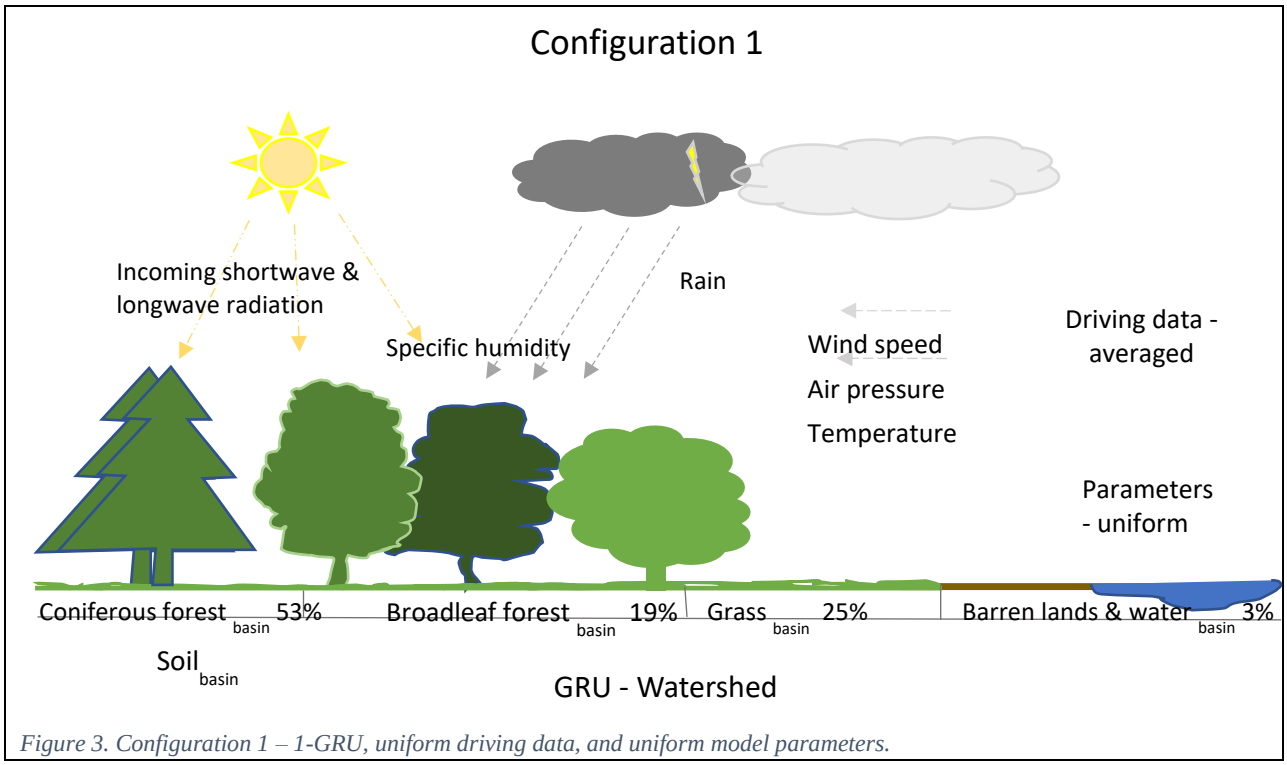

Figure 3. Configuration 1 – 1-GRU, uniform driving data, and uniform model parameters.

Configuration 2 (Figure 4) – the watershed is covered by two ecodistrict – zoning information produced by Agriculture and Agri-Food Canada (available online from The National Ecological Framework website), namely, the Whiteswan Uplands (zone 658) and the White Gull Plains (zone 659). The soils and land cover parameters are the same between the two areas but in different proportions over each ecodistrict. Again, as described in configuration 1 above, the parameter value given to each unit is proportional to the weighted average of that land cover type in the respective ecodistrict. The forcing data is a calculated mean of the measurements obtained from the OBS and OJP observation sites and is applied to both side of the watershed. The hydrological response from each land cover class in the respective ecodistrict contributes to the stream channel network which guides the flow to the outlet of the basin.

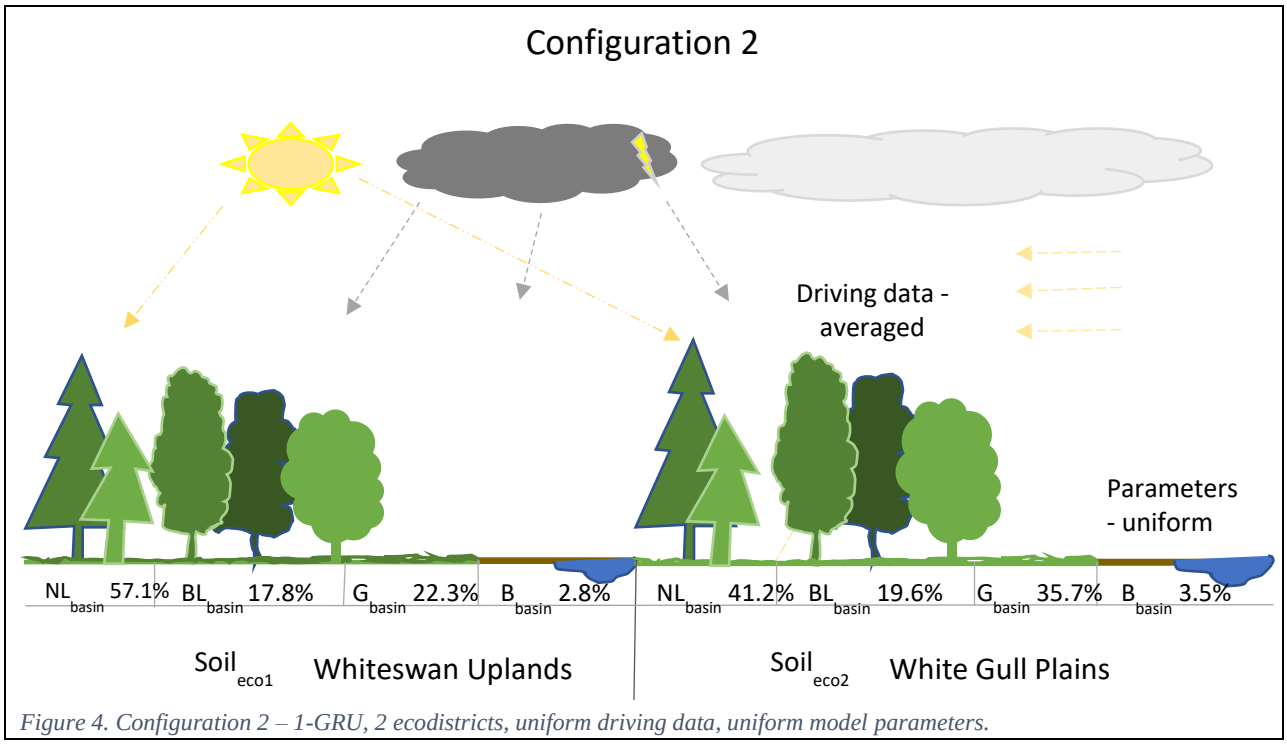

Figure 4. Configuration 2 – 1-GRU, 2 ecodistricts, uniform driving data, uniform model parameters.

Configuration 3 (Figure 5) – the forcing data are distributed over the entire watershed based on the 2 ecodistricts. The values used are the past measurements obtained from the OBS and OJP observational sites. Thus, the measurements from the OBS are used for Whiteswan Uplands to the west, whereas the OJP measurements are applied to White Gull Plains to the east of the watershed. The parameter distribution is maintained evenly over the entire watershed as described in configuration 2 above.

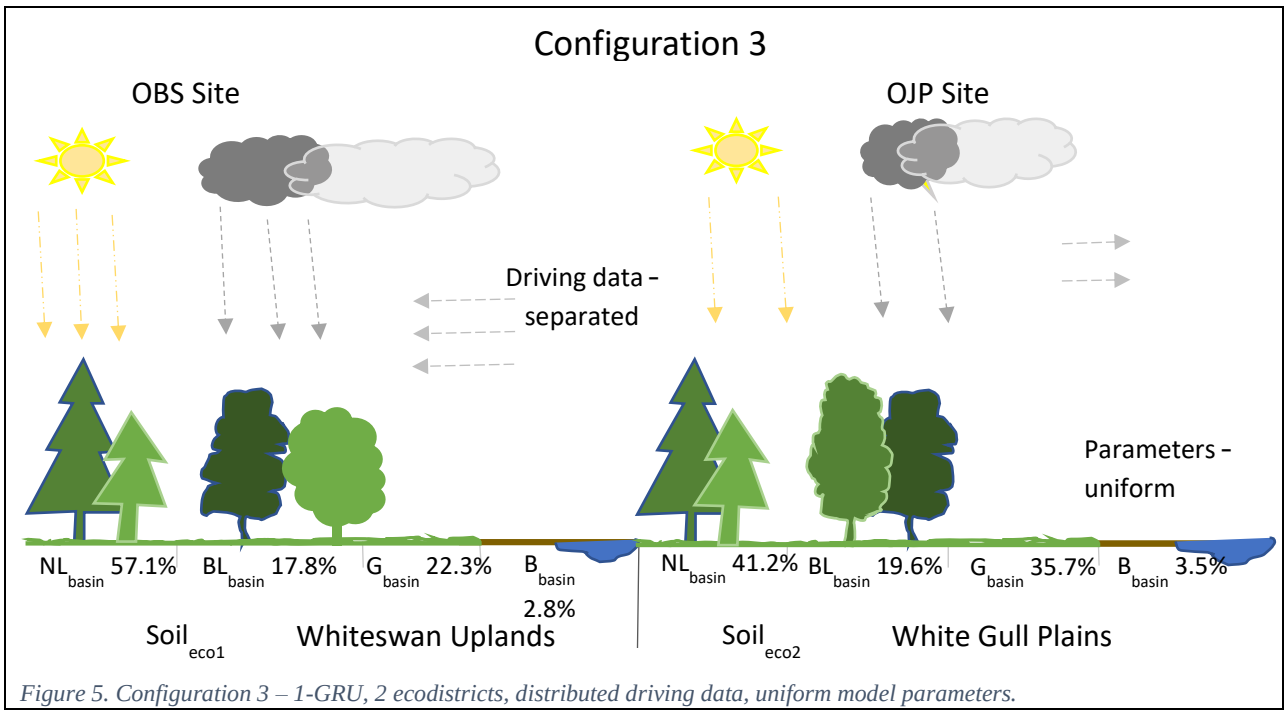

Figure 5. Configuration 3 – 1-GRU, 2 ecodistricts, distributed driving data, uniform model parameters.

Configuration 4 (Figure 6) – In this arrangement, the model parameters increase from 33 to 65 by applying the different parameter sets for the soils and landcover in the two ecodistricts and varying the values across the watershed based on the size of the GRUs. This means that each land cover fraction is estimated with respect to the GRU portion in the watershed, the model parameter values are different between the two GRU based on the land cover type. Uniform distributed forcing data (average of the measured data from the 2 observation sites) were utilized for this setup. The output generated by the different land cover classes contributes to the main stream flow.

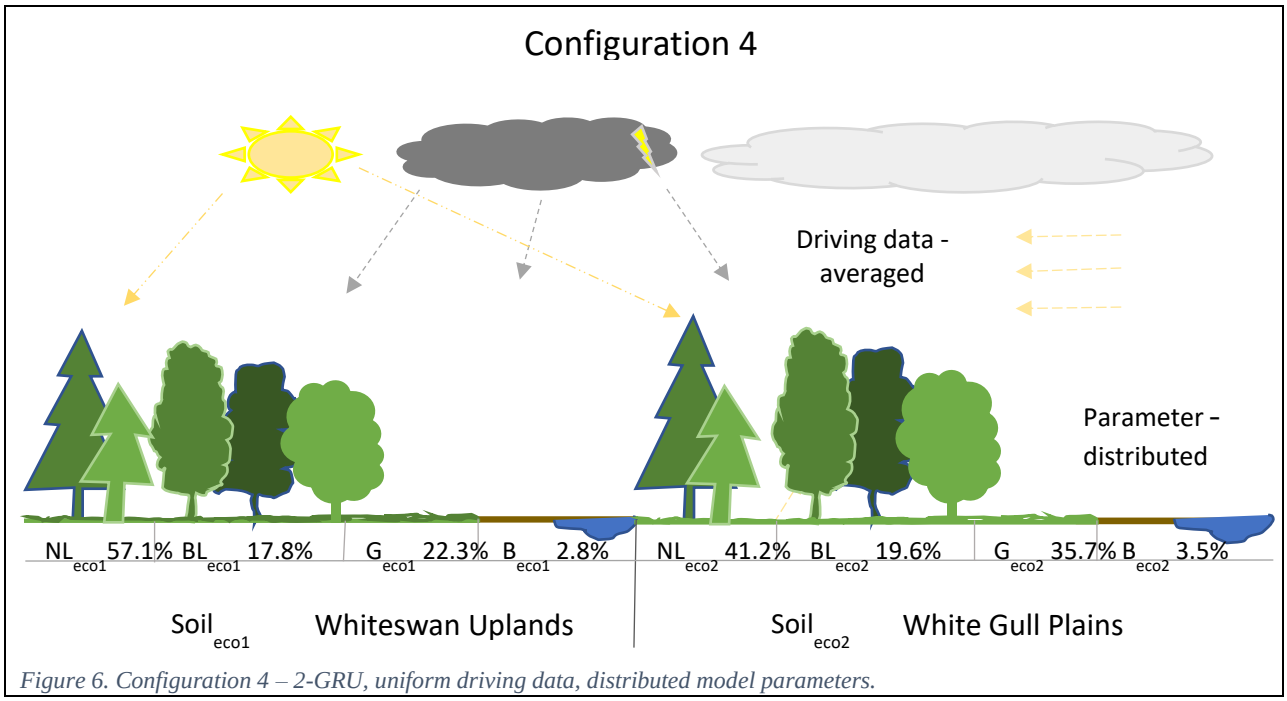

Figure 6. Configuration 4 – 2-GRU, uniform driving data, distributed model parameters.

Configuration 5 (Figure 7) – in this 2-GRU setup, the parameter distribution is the same as in configuration 4 above, but the driving data are distributed such that measurements obtained from the OBS and OJP observational sites are applied in their respective ecodistrict i.e. Whiteswan Uplands is forced with data from the OBS site, whereas White Gull Plains get the information from the OJP site. The response from each GRU is calculated and contributes to the overall streamflow at the basin outlet.

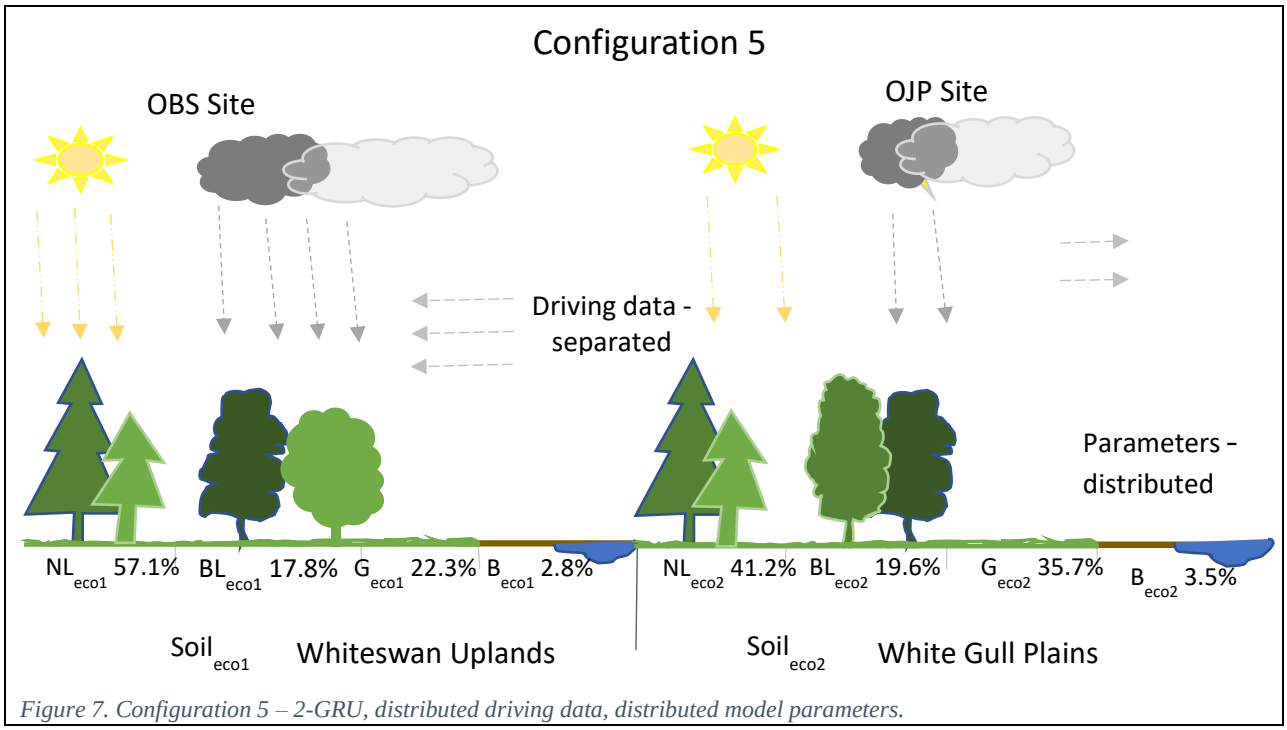

Figure 7. Configuration 5 – 2-GRU, distributed driving data, distributed model parameters.

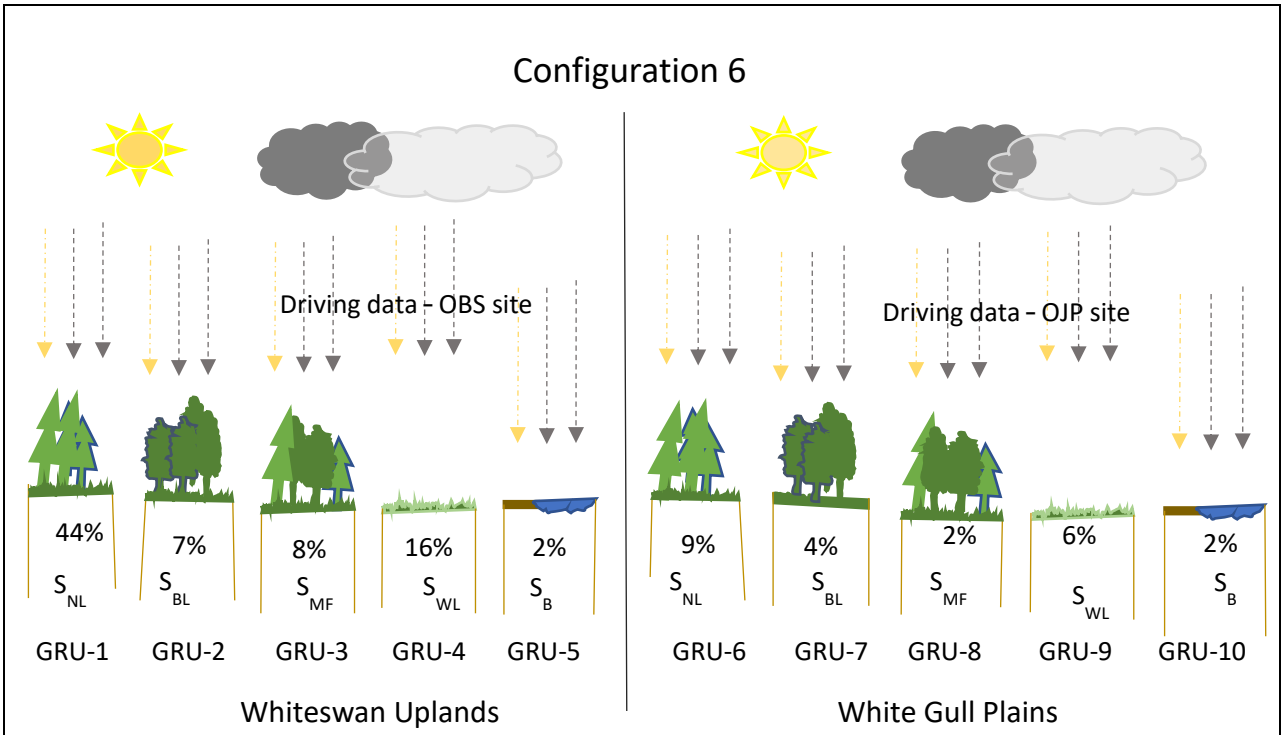

Figure 8. Configuration 6 – 10-GRU, distributed driving data, distributed model parameters.

Configuration 6 (Figure 8) – in this arrangement, available spatial information is utilized in subdividing the watershed into one GRU for each land cover class (five), for each ecodistrict (two), resulting in ten GRUs. The distribution of the driving data follows the two ecodistrict lines, i.e.

each land cover class gets a separate set of driving data based on the location of its GRU in the watershed. 92 different model parameters were used among the ten GRUs. Their distribution is also influenced by the physical location of the land cover types, such that each land cover class gets a unique parameter value representing surface heterogeneity on the physical landscape. The hydrological responses from these GRUs are summed and channeled through the stream channel network to contribute to the stream flow at the watershed outlet.

c) Model Calibration and Validation

Model calibration is a procedure of searching the parameter space for optimal values "by comparing model estimations for a given set of assumed conditions with observed data for the same conditions" (Moriasi et al., 2007). In models such as MESH there are large numbers of parameters (tens to hundreds), with the potential number of parameters increasing as more spatial detail is included. If no calibration is performed, i.e. a-priori parameter values are used, we know that the model performance will be poor (Davison et al., 2016). It is also not practically possible to calibrate every parameter (each additional parameter adds a new dimension to the parameter space). We therefore start the calibration process by identifying, based on experience, those parameters that we expect to be sensitive (i.e. impact on the model output) and uncertain (i.e. take a value that depends on highly variable factors, such as soil properties). These are termed "free parameters", and are subset of all the model parameters that can be allowed to vary in the calibration process. In this study, six model configurations were tested, all of which needed to be calibrated, and most of which have different numbers of free-parameters. Since we wish to compare the results from these different configurations, we sought to minimize any bias that might be introduced from the calibration process itself. For example, a model with fewer parameters would require less model realizations during the calibration procedure than a model with many parameters, and would the less susceptible to false (or local) optima. Therefore, 100 calibration runs with different initial parameter guesses were run for each model configuration, and each calibration run comprised of a fixed number of realizations (1000). This produces 100 sets of optimal parameters for each configuration (which if the calibration was perfect would be identical), and we compare the distribution of these sets. The model was calibrated using the observed daily discharge data from the stream flow gauge located at the basin outlet from 1999 – 2007. The first year was used as a warmup period to initialized the model. The model evaluation efficiency approach of Nash and Sutcliffe (1970) was used as the objective function. The Nash-Sutcliffe (NS) efficiency index, is a normalized statistical indicator that defines the degree of predicted signal variance compared to the observed information (Nash & Sutcliffe, 1970). The statistical and graphical presentation, displays how close the simulated streamflow and the measured data are to each other. NS efficiency is computed as follows:

$$NSE = 1 - \left[\frac{\sum_{i=1}^{n}\left(Y_i^{obs} - Y_i^{sim}\right)^2}{\sum_{i=1}^{n}\left(Y_i^{obs} - Y^{mean}\right)^2}\right]$$

where Y_i^{obs} is the observed streamflow values, Y_i^{sim} is simulated streamflow values, Y^{mean} is the mean of the observed streamflow data, and *n* is the total number of observations evaluated.

The NS efficiency values range between 1.0 and $-\infty$, with 1 being a perfect fit, whereas less than 1 is considered acceptable and below zero is undesirable performance (Moriasi et al., 2007). Generally, NS values that are above the threshold of 0.50 are considered satisfactory predictions,

although can change subject to project goals (Moriasi et al., 2007; Refsgaard, 1997). The research cluster (the PLATO server) of the University of Saskatchewan was used to run the model-independent OSTRICH software (Matott, 2005), which employs the dynamically dimensioned search (DDS) system (Tolson & Shoemaker, 2007) for the automation of the parameter optimization. To minimize any bias that might be introduced from the calibration process itself, every calibration run begun from a different initial starting point determined by a random seed generated by the system. This yielded a total of 600 000 runs over a 14-day period. For efficiency, the one hundred calibration runs for each configuration were sub-divided into groups of 3 and ran concurrently to shorten the time required to complete the calibration process. The more complex structure (configuration 6) required six groups to reduce the length of the experiment.

Model validation involves running the model using the calibrated parameter values over a period other than the one used for calibration. This process tests the model's ability to make reasonably accurate predictions which can lead to the model results being acceptable or rejected depending on the objectives of the project (Moriasi et al., 2007; Refsgaard, 1997). Using observed streamflow data for model validation exposes its performance and weaknesses in its behavior, however, such tests are inadequate (Arora, Chiew, & Grayson, 2001). Notwithstanding such observations, the model was validated against the archived daily stream flow data observed from 2007 – 2014.

6. Results

a) Model output

The distribution of the model performance for 100 calibration runs for each configuration is shown in Figure 9 (see Figure 21 for ranked model performance). Each box illustrations the variability of the model performance. The box represents the 25^{th} and 75^{th} percentiles of the model performance, while the median is represented by the line in the box. The whiskers (error bars) depict the minimum and maximum values outside the 1^{st} and 3^{rd} quartiles. The results show a slight increase in model performance with increasing complexity, but starts to degrade after configuration 4. Though the performance seems to degrade, the narrow spread in configuration 6 imply a reliably consistent model performance with no bad values. All other configurations display a considerable level of susceptibility to bad results.

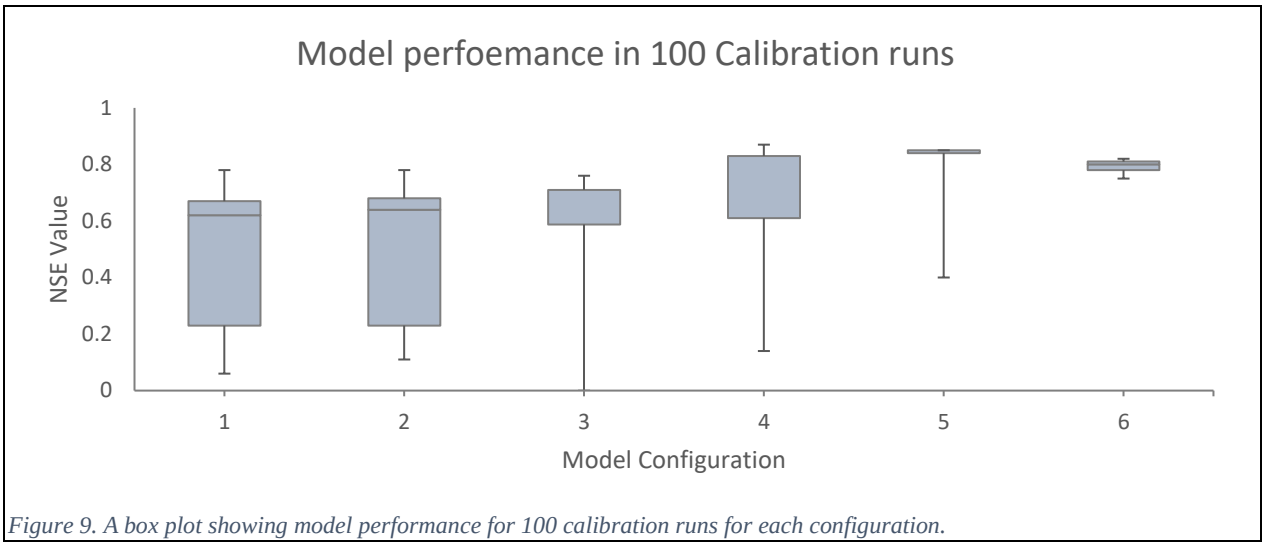

Figure 9. A box plot showing model performance for 100 calibration runs for each configuration.

The following figures show the performance of the model in validating the best calibration run for each configuration. Although, there are marginal differences in the calibration results, model validation produced three identical results. This exposed the inherent problem of equifinality associated with calibration, where many sets of solutions exist for the same problem, or simply put, different parameter sets producing the same solution (Beven, 2002). This is shown in the calibration results of configuration 1 (Figure 10) and configuration 2 (Figure 11). Similarly, the validation of the first three model structures (shown in Figures 10, 11, and 12) reveal the equifinality issue.

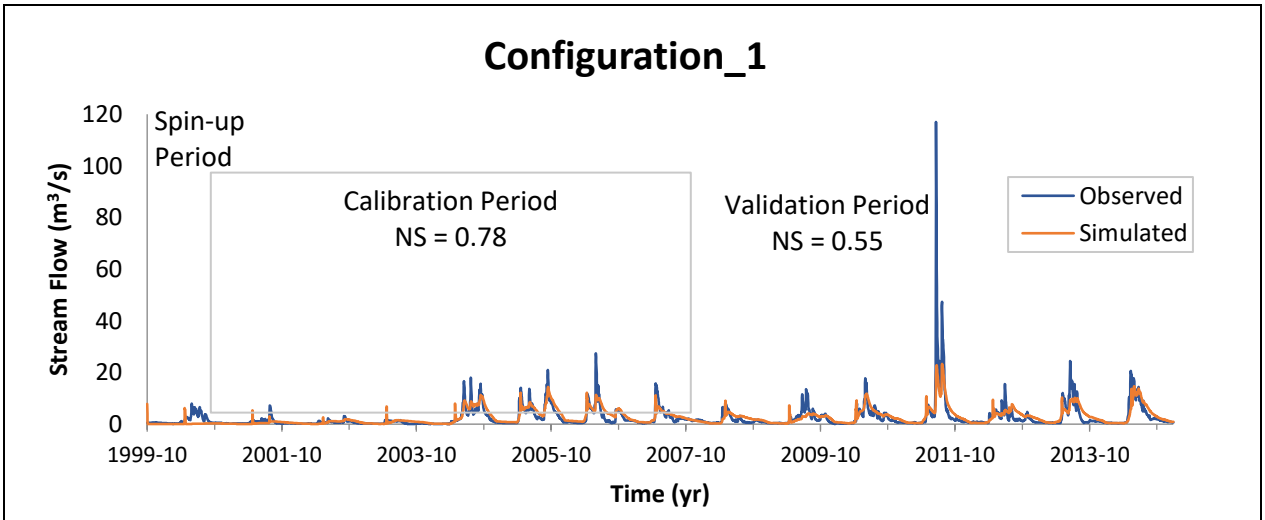

Figure 10. Observed and simulated hydrograph for Configuration 1 showing the spin-up period (1999 - 2000), calibration period (2000 - 2007), and validation period (2007 - 2014).

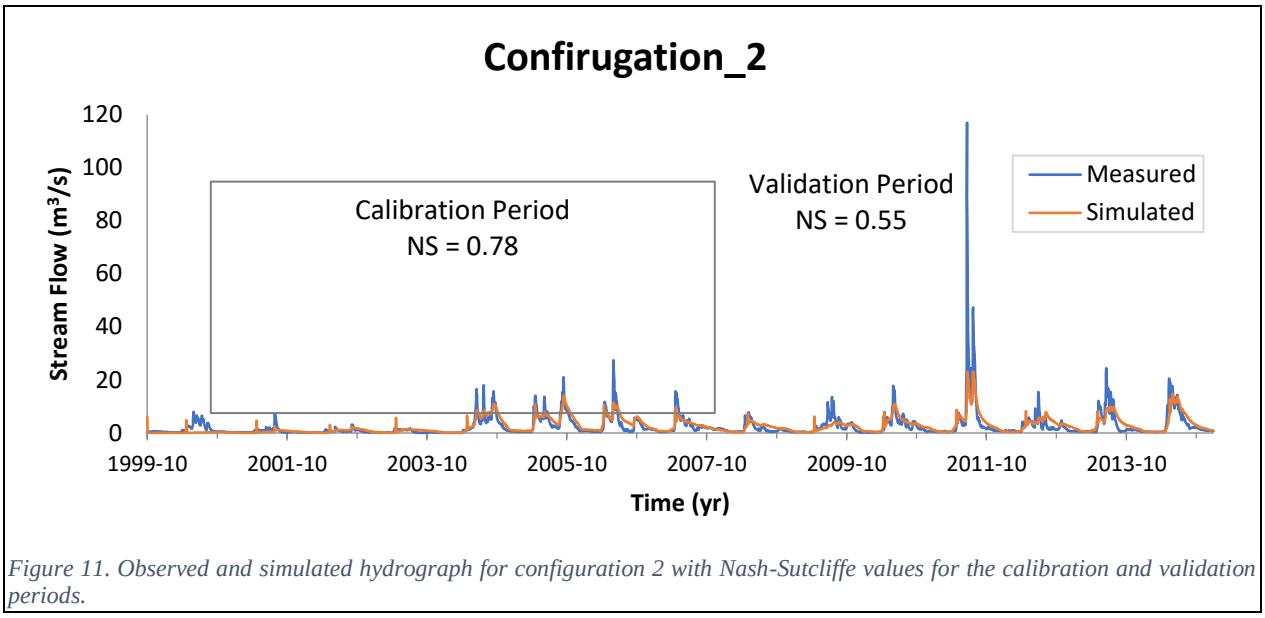

Figure 11. Observed and simulated hydrograph for configuration 2 with Nash-Sutcliffe values for the calibration and validation periods.

Configuration 3 (Figure 12) shows a decrease in performance in calibration, though the only difference with configuration 2 is that forcing data is spatially discretized. The validation results show an example of many different parameter sets producing same solution – equifinality, for all three configurations.

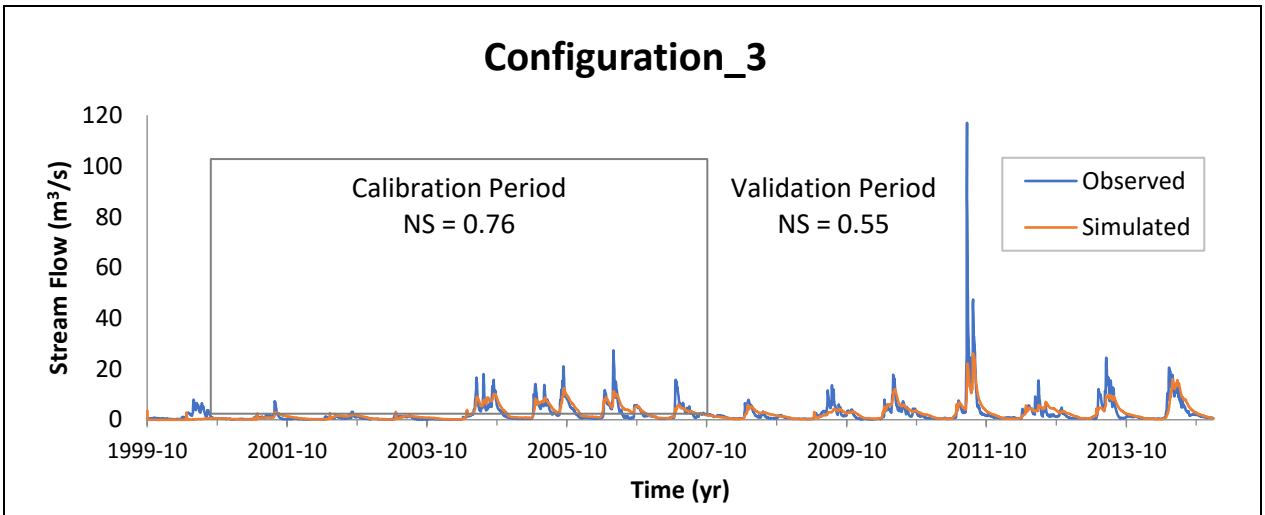

Figure 12.Observed and simulated hydrograph for configuration 3.

Figures 13 and 14 show a considerable improvement in model performance for both periods, with configuration 4 producing the best calibration run of the entire experiment. The increase in performance is consistent with the addition of more model parameters (see Table 1), which describe spatial landscape heterogeneity both explicitly and indirect.

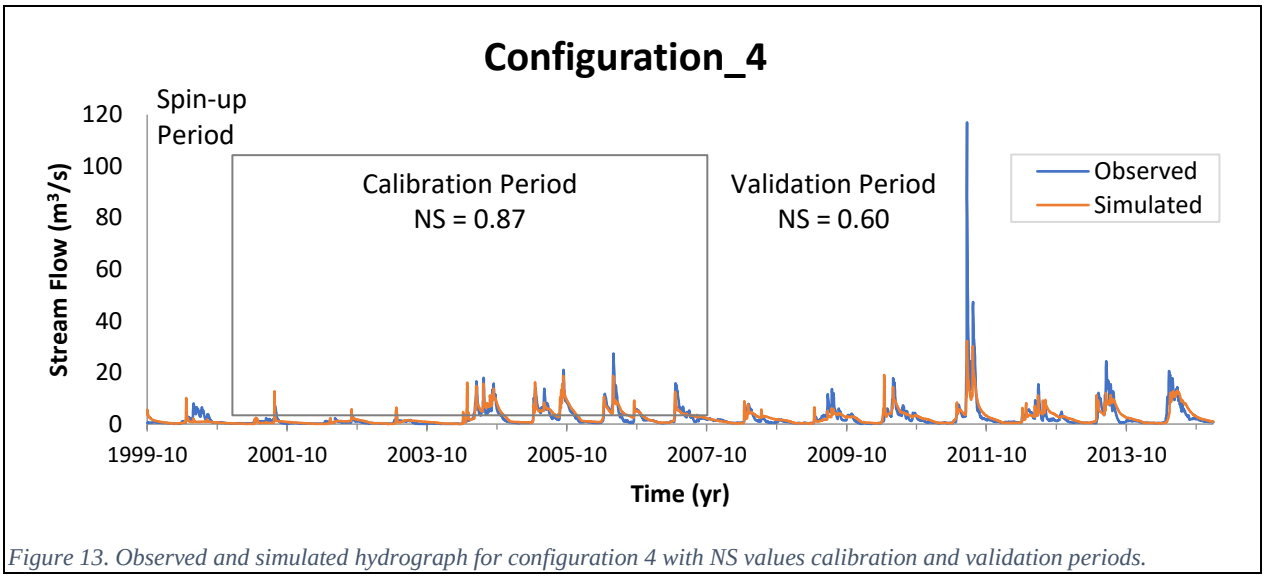

Figure 13. Observed and simulated hydrograph for configuration 4 with NS values calibration and validation periods.

Configuration 5 (Figure 14), shows an improved performance in validation, though there is a slightly lower NS value in calibration than configuration 4.

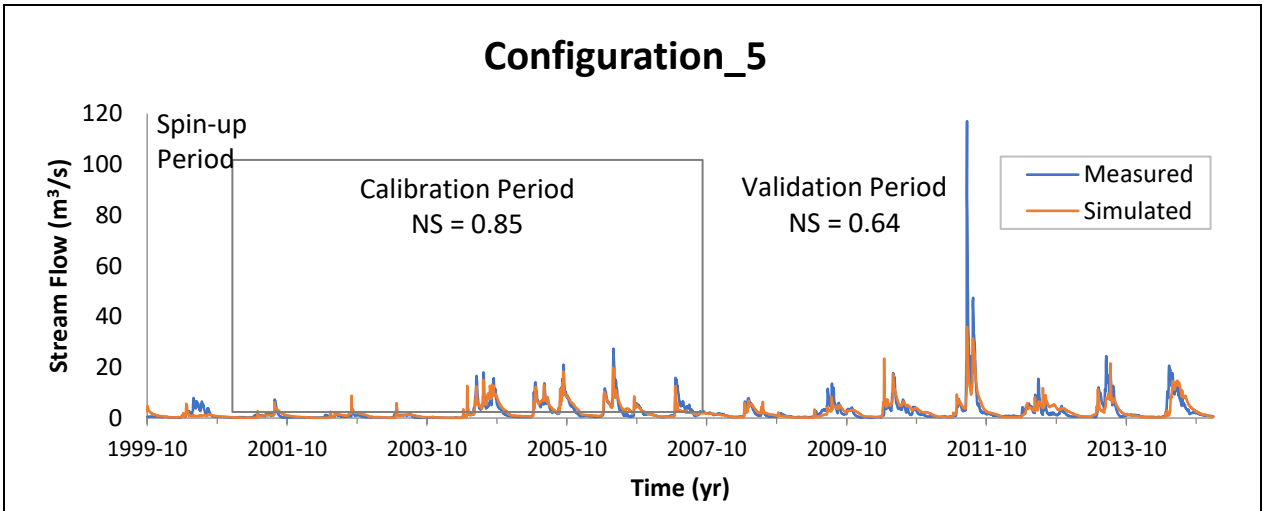

Figure 14.Observed and simulated hydrograph for configuration 5 showing the model performance for the calibration and the validation periods.

In configuration 6 (Figure 15), the model performed well in validation period than all other configurations, although it was not the best performer in the calibration period. The model captures the shape of the hydrograph, but still misses peak flows in both periods. The model under-estimate the extreme event of June 2011 by a wide margin. This shows the limitation of the model structure applied.

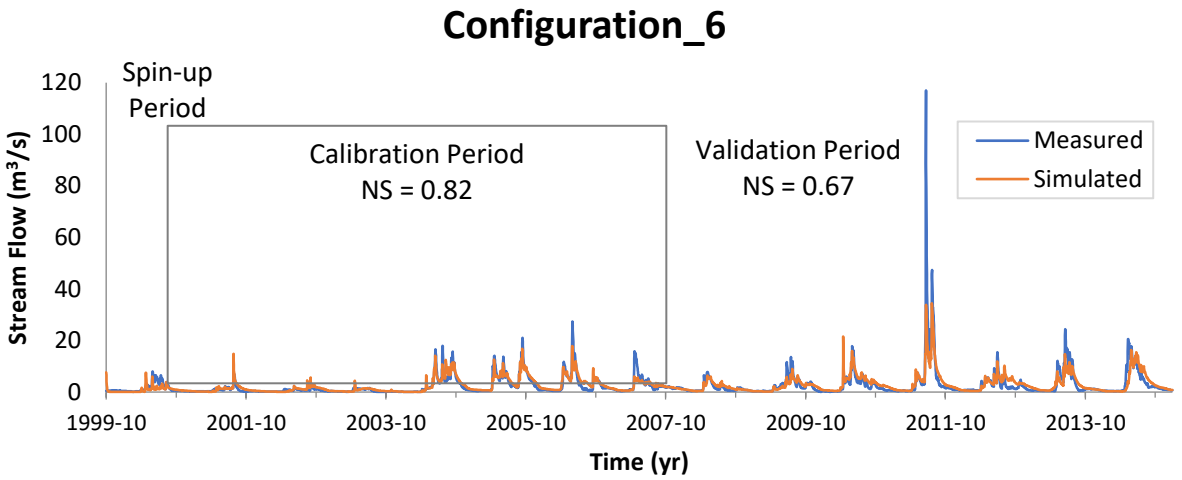

Figure 15.Observed and simulated hydrograph for configuration 6 showing model performance during calibration and validation periods.

The box plot in Figure 16, shows an assessment of the model performance validation of the best 10 calibration runs for each configuration. This represent the top 10% performance of the model (see also Figure 21 in Appendix). The results reveal substantial variability for configurations 1, 2, 3, and 4 in both calibration and validation periods. However, there is a decrease in performance in the calibration process for configurations 5 and 6. There is a steady improvement in model validation from the lumped configuration 1 to complex configuration 6. Would the trend continue if more complex configurations were added beyond the current level?

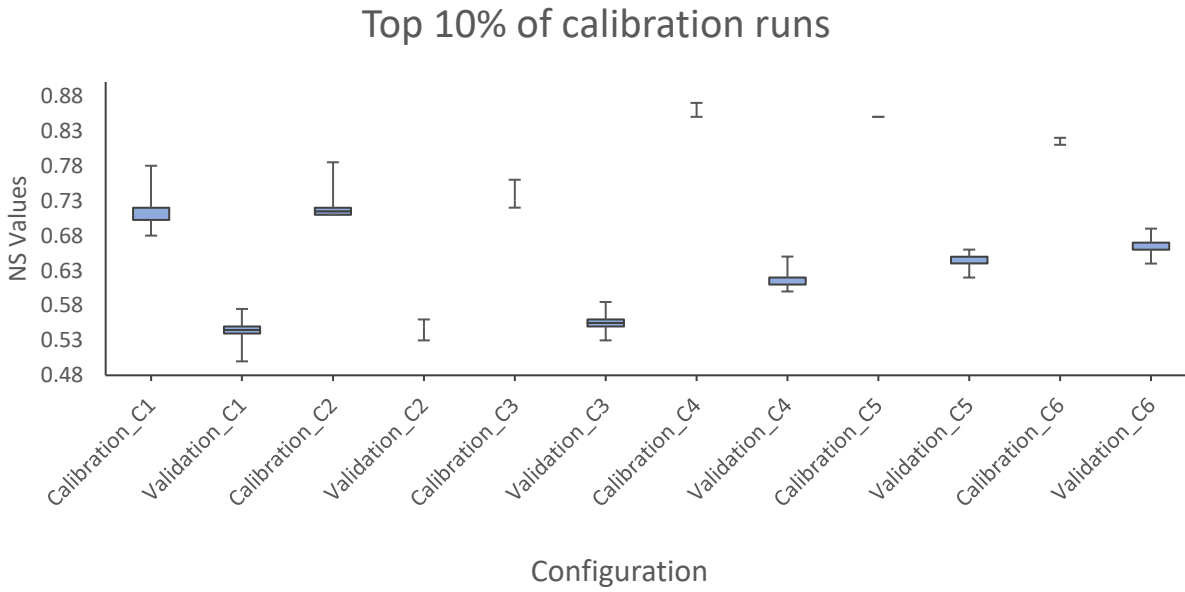

Figure 16. A box plot showing model performance for the validation of the best 10 calibration runs for each configuration.

The average computational run-time required to complete each calibration run, the calibration and validation results of the best calibration performance for each configuration are shown in Figure

17. The calibration run-time increases as additional GRUs are added to the model, increasing the computational expense of the model. Hence, as expected there is a good correlation between level of complexity and the amount of time needed to complete each calibration run. Model performance in calibration, however, does not linearly increase with addition of more GRUs, whereas that is the case in model validation. The results (Figure 17), also reveal that configurations (refer to Table 2 for configuration structure) forced with uniform driving data are associated with higher calibration results, while configurations with distributed forcing performed better in the validation period for same level of complexity (i.e. configurations 2 vs 3 and 4 vs 5). However, there is no significant difference in the corresponding computational time requirements for the same configurations.

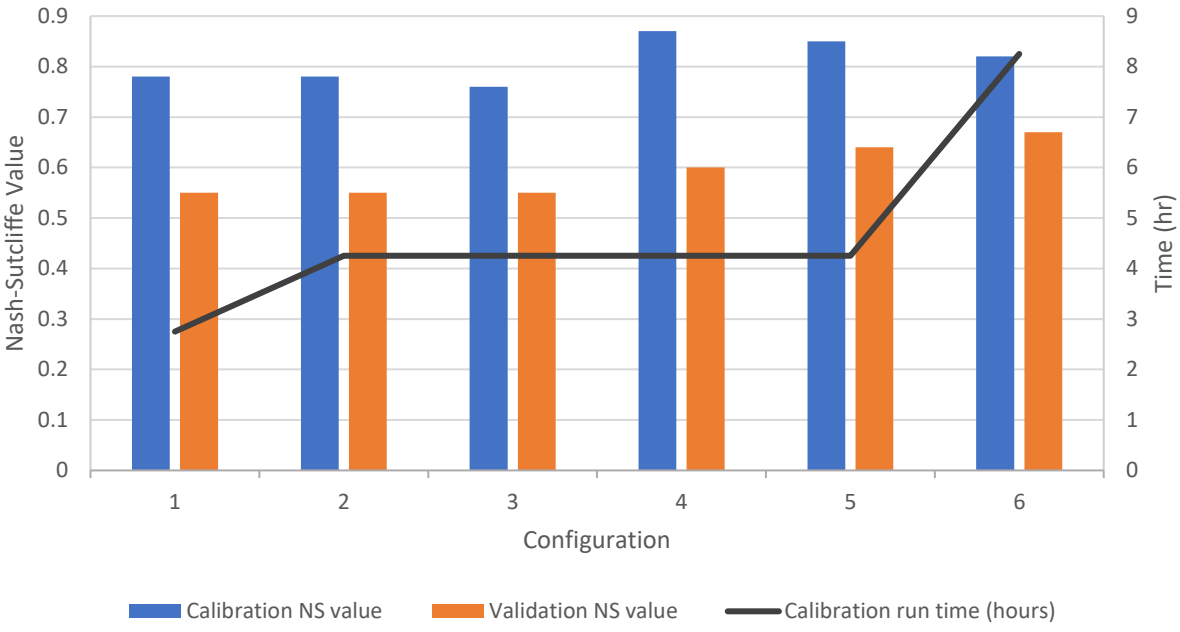

Figure 17. Model performance (for the best calibration run) vs average run times for each calibration.

b) Parameter identifiability

Parameter identifiability is a measure of how uniquely a parameter value is found by the calibration process. In some cases, the same model performance can be achieved with different parameter values – the problem of equifinality – and in this case, we do not know the true value that the parameter should take. Here we performed 100 calibrations for each model configuration, which allows us to explore parameter identifiability. An identifiable parameter will have very little spread in its optimized values, whereas a completely unidentifiable parameter might take any value over the range considered. In this study, we use a simple metric of parameter identifiability, which is the range between the 10^{th} and 90^{th} percentiles of the normalized parameter ranges. Parameter ranges were normalized so that each parameter takes a value from zero to one, and hence the identifiability metric also takes a value from zero (perfectly identifiable) to one (completely unidentifiable). We also consider a threshold, arbitrarily set to 0.3, which is used to categorize a

parameter as identifiable (when it's identifiability metric is less than or equal to 0.3) or unidentifiable (the metric is greater than 0.3).

In Figure 18 – 20, the behavior of the model in finding optimum values for each parameter in the 100 calibration runs for each configuration is shown. The calibration of configurations 1 and 2 (Figure 18) reveals that most parameters are not constrained, the range of their values is too wide to capture any significant information. In configuration 3, the broadleaf minimum leaf area index (LMN-BL) is the most identifiable parameter while other parameters are unidentifiable. We note that only the saturated surface soil conductivity (wcfi1) parameter is more identifiable in all 3 configurations.

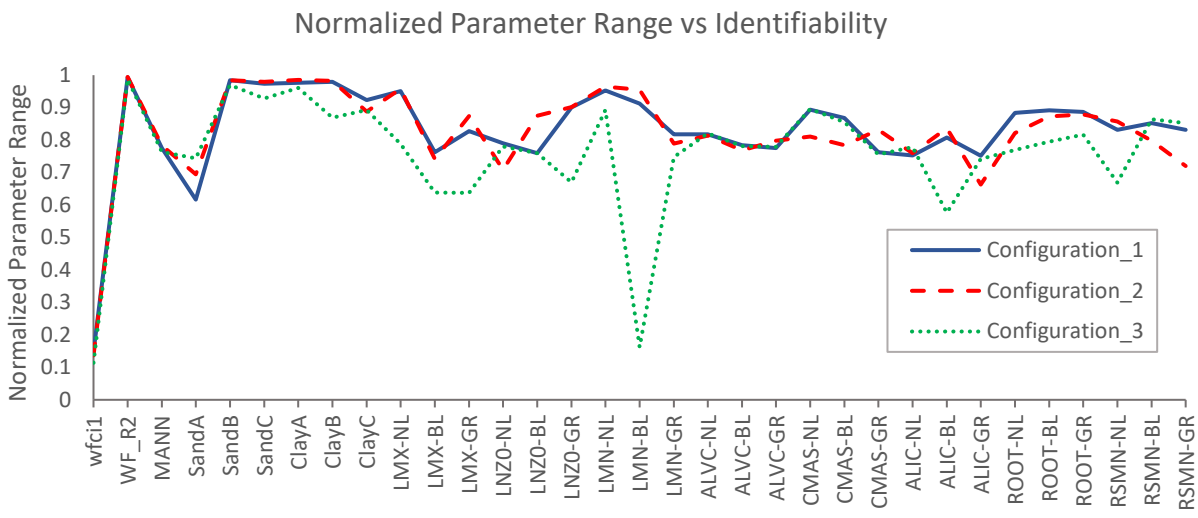

Figure 18. Parameter identifiability during the calibration process (33 Model parameters). The blue solid line represents configuration 1, the red dashed line shows configuration 2, and the green dotted line represents configuration 3. The graph represents the difference between the 10th and 90th percentile parameter ranges.

For configuration 4 and 5, shown in Figure 19, several parameters are uniquely identifiable by the calibration process, including the LMN-BL, WR_R2, ClayA3, ClayB3, and wfci1, representing the minimum leaf area index, stream channel properties, clay soil layers, and saturated surface soil conductivity respectively. These parameters may be sensitive, but might not have a clear influence on the simulation of the stream flow as the range of their values is poorly defined (i.e. not zero – perfectly identifiable). The results show that parameter identifiability improves in configuration 5. While configuration 4 has fewer identifiable parameters than configuration 5, it still gives better NS values in calibration though with more susceptibility to bad runs.

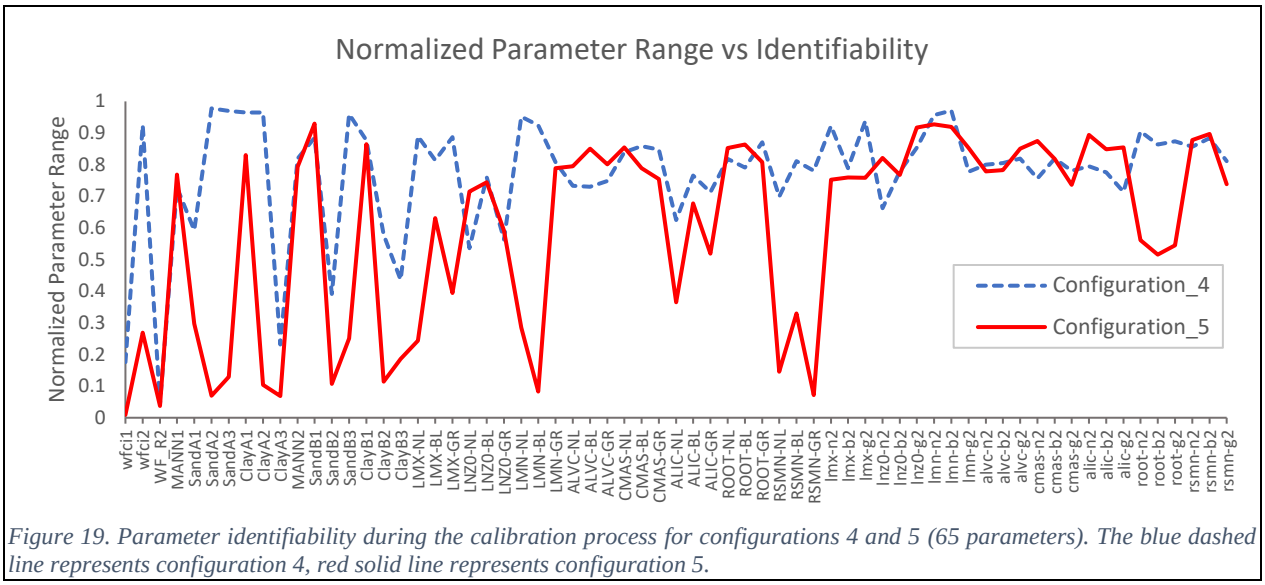

Figure 19. Parameter identifiability during the calibration process for configurations 4 and 5 (65 parameters). The blue dashed line represents configuration 4, red solid line represents configuration 5.

In configuration 6 (Figure 20), model parameters show moderate identifiability (sensitivity) regardless of the initial conditions. This is consistent with the perception that more complexity brings additional identifiability problems as there are more parameters in the parameter space to be calibrated. Similarly, there may be multiple local/false optimal values for which longer calibration runs may be needed to resolve the results.

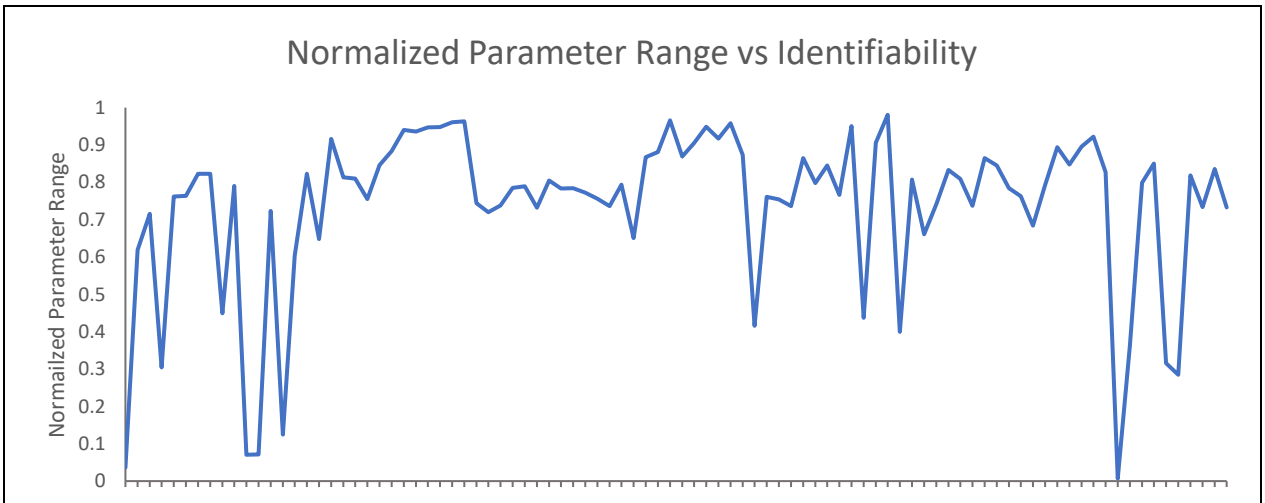

Figure 20. Parameter identifiability during the calibration process for configuration 6 (92 Model parameters). The graph represents the difference between the 10th and 90th percentiles of the normalized parameter ranges.

7. Discussion

In configuration 1 (Figure 3), a simplified structure (lumped model) is used to represent spatial heterogeneity of the physical characteristics of the watershed. Model parameters are discretized based on land cover classes and the spatial information about landforms is ignored. Where

measurements are available, such as meteorological data, weighted average values are used, or else they are estimated through calibration (Vrugt et al., 2003). As shown in the box plot in Figure 9, this model structure (configuration 1) reveals great variability and susceptibility to bad results in the calibration period (see Figure 21 in Appendix for ranked calibration performance for all configurations). The model performance (for the best calibration run shown in Figure 10) is good in capturing the shape of the hydrograph, however, high flows during the wet years were not accurately captured for both calibration and validation periods. In the validation of the top 10 calibration runs (Figure 16), the model displays a very unstable and weak performance. As expected, configuration 1 required the least amount of time to complete each calibration run (Figure 17).

The model structure in configuration 2 (Figure 4) is slightly different from configuration 1, in that information about the physical dimensions of ecodistricts is incorporated but the average values are used regardless. The model displays poor, unreliable performance with great variability and susceptibility to bad results over 100 calibration runs (Figure 9). The result shown in Figure 11 for the best calibration run in configuration 2 is as expected – identical to results in configuration 1. The model fails to capture the peak flows for the entire study period, with the extreme event of June 2011 extremely under estimated by a significant factor. This could partly be due to wetland soils and ponding features of the watershed not being represented explicitly in the model (Davison et al., 2016). The model also, fails to estimate low flows during winter months and drier years. Upon validating the top 10 runs of the calibration process (shown in Figure 16), the results show poor and unstable performance as configuration 1. There is an increase in computational time needed to complete each calibration run for this arrangement (Figure 17), almost double the time for configuration 1.

Configuration 3 (shown in Figure 5), introduces a separation of driving data based on ecodistrict, while model parameter distribution remains unchanged. In Figure 9, the model shows that configuration 3 is prone to producing very bad results despite improved performance in validation. Configuration 3 results shown in Figure 12, reveals a degraded model performance in calibration, before recovering in the validation period to perform better than configurations 1 and 2. Despite a decrease in performance during calibration, increase in complexity improved performance in model validation, with no increase in computational time (Figure 17).

Configuration 4 (Figure 6), introduces more model parameters (Table 2) by separating each parameter based on the two ecodistricts. The model performance in 100 calibration runs (Figure 9) reveals that this structure may the best calibration, but is susceptible to poor results. The results (for best calibration run shown in Figure 13), indicate that configuration 4 does well in estimating high flows in the calibration period. However, the overall performance is weak. The validation period is moderately good, with a slight increase in performance, although the high flows are under estimated. Validation of the top 10 calibration runs (Figure 16) display great unpredictability in model performance.

In configuration 5 (Figure 7), driving data and model parameters are discretely distributed based on ecodistrict. Configuration 5 is consistently good most in most of the 100 calibration runs, as shown in Figure 9, but is susceptible to bad performance. Addition of more complexity (shown in Figure 14), marginally degrades model performance in calibration but improves validation

significantly. The model does well in capturing most of the peak flows, though the extreme event of 2011 is missed.

Configuration 6 (Figure 8), shows the representation of spatial sub-grid heterogeneity of a watershed using a 10 GRU model structure. Soil and vegetation parameters are distributed based on land cover classes and ecodistrict. Configuration 6 shows the most reliability and consistent performance in calibration, with no bad values, as presented in Figure 9. The results (Figure 15) show an improved model performance in validation. The model captures most of the peak flows in the drier years, for the entire study period. High flows in wetter years are under estimated slightly, apart from the extreme event of June 2011, which is greatly under estimated. This is due in part to how wetland soils and ponding features are treated in the model, as described above. The performance in validation is nonetheless significantly better than all other configuration.

Model validation for the best 10 calibration runs for each configuration, presented in Figure 16 show that the model structure with best calibration run did not produce the best validation, i.e. configuration 4 had the best calibration value of 0.87, but configuration 6 performed considerably better in the validation period despite having 0.82 in calibration. Furthermore, the best validation NS values did not come from the best calibration run for all configurations. There is no clear trend followed by the validation process to determine the best validation NS value. It will be interesting to see how the model performs when more than 15 – 20% of the calibration runs are validated rather than the standard top calibration run.

Lumped model structure performed as well as the semi-distributed, but with great variability. As the watershed was discretized from 1 GRU to 2 GRUs and eventually to 10 GRUs, the model performance increased significantly. Spatial location is ignored in the 1-GRU structure, whereas some locations (ecodistrict) are identified in the 2-GRU arrangement. Additional spatial information is used in the distribution of parameters (soil and vegetation), which undoubtedly improves the model performance.

The distribution of driving data over the basin does have an impact on the model output. Averaging total precipitation from the two observation sites performs better in calibration and splitting (keeping the observed data in the respective ecodistricts) does well in validation. As shown (see Table 2 and Figure 17), uniform application of meteorological information is associated with higher performance for the calibration period (for a similar level of complexity), while the opposite is true for the validation period. Varying driving data alone is not enough to capture and represent sub-grid variability in a model. To reduce variability of errors that may be introduced in the model, same driving data is used for all configurations. It is assumed that would not affect the comparison of the setup and complexity of the configurations structure (Haghnegahdar et al., 2015).

In this study, the number of parameters in the model that required calibration was one of the criteria used to define the level of complexity. As the number of parameters gradually increase from 33 to 65 (configuration 4 and 5) to 92 in configuration 6 (see Table 2), more complexity is assumed. The 1-GRU (lumped) arrangement was quick to calibration, taking an average of two hours and 15 minutes per calibration run. 2-GRU configurations (with 33 or 65 parameters) demanded 4 hours and 15 minutes of computational run-time to complete each calibration run. Further increase in the number of model parameters (92 in the 10-GRU configuration 6) required longer run-time

(8 hours and 15 minutes). A summary of computational run-time required to complete each calibration run and how that varies with increase in model complexity is shown in Figure 17. This shows that improvement in model performance could be a result of providing the search algorithm with more spatial information and time to find better solutions.

While a good parameter set is important in obtaining a good fit, there is no guarantee that it will work for another system (Beven, 2001). In some cases, the same model parameters may behave different in another model structure, but it's still possible to achieve the same model performance with different parameter sets. In this experiment, parameter identifiability is poorly defined for the less complex configurations (Figure 18), but slightly improves with added complexity (Figure 19). However, as the number of parameters increases in configuration 6 (most complex configuration shown in Figure 20), parameter identifiability decreases as there are more parameters to identify in the parameter space. This comes with increase in model run-time in calibration. Parameter identifiability problems could be caused by a false or local optimal value that may exist for which the model optimization process converges on in the parameter space. Alternatively, it could be that there are not enough searches (1000 guesses for each calibration run) for effective optimization especially for more complex configurations. Similarly, model parameters may interact with other parameters in the model, such that their identifiability depends on the value of others in the parameter space and may not be detectable by the model.

In summary, results show that adding more spatial complexity to the model structure may not provide the best model performance in general, but could provide slight gains at higher computational cost. For that reason, computational run-time should play a vital role in the selection of a better configuration depending on the objectives of the project. Although, this process was not exhaustive on the level of complexity that could be added, results show considerably improvement in model performance when varying parameters (soil and vegetation) based on ecodistricts. Model performance in calibration may reach a peak as complexity increases. In contrast, model validation seems to improve with added complexity, at the expense of computational run-time in calibration. Results also reveal that model performance during calibration is not an indicator of its performance in validation. Therefore, it is not adequate to validate only the best calibration run. Additionally, increasing model complexity improves the consistency of the calibration results at the expense of parameter identifiability.

The remaining questions that could be examined in future work include:

- Could the inconsistency in 'good' calibration be due to the algorithm (DDS), or perhaps the number of runs performed in the calibration process?
- Would hyper-resolution modelling give similar results? (i.e. semi-distributed vs distributed [hyper-resolution] vs "hyper-resolution semi-distributed" – hyper-resolution + many GRUs per grid.
- Could things be improved by making the model more complex? Do we get similar results with a different land surface scheme? How does changing the model physics impact the results?

The limitations associated with this study include longer computational time required to perform numerous calibration runs for marginal gains. This might be a challenge for large-scale experiments. Also, finding sufficient data at appropriate scale, to adequately discretize large-scale or complex watersheds might limit the applicability of this approach. Arguably, the experiment

has shown the importance of identifying reliable solution schemes to represent the behavior of hydrological processes at sub-grid level while considering model complexity, computational efficiency, and data availability.

8. Conclusion

In conclusion, semi-distributed model structure outperforms lumped model in both calibration and validation. Also, varying model parameters by ecodistrict provides a clear improvement over lumping vegetation parameters across ecodistricts. At this spatial scale, distributing the forcing data doesn't seem to make much difference when calibrating. The most complex configuration is the most consistent and produces the best validation results. Parameter identifiability is poorly defined for all configurations, with a slight improvement in the most complex structure. Finally, based on the analysis of the current study, it's a toss-up as to whether configurations 4, 5 or configuration 6 is the best for representing spatial sub-grid variability in a model. More work is needed to examine the remaining questions.

9. Appendix

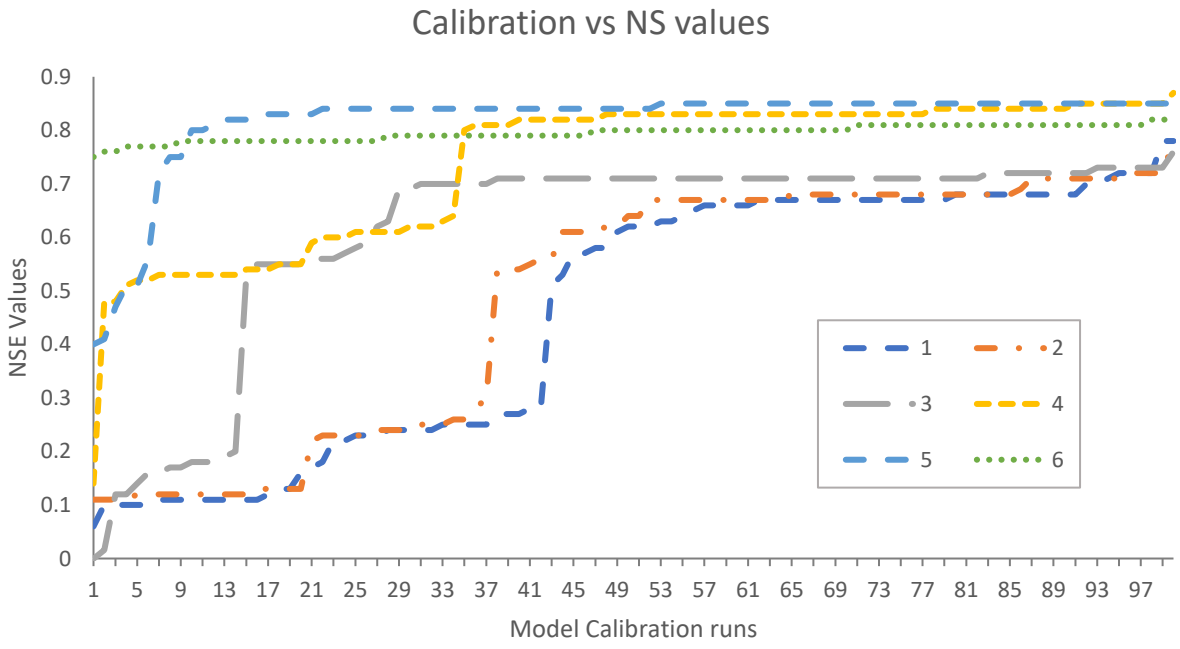

Figure 21. Model performance (ranked) for the 100 calibration runs for each configuration.

Table 3 Parameters ranges used in the MESH model set up and calibration (adapted from Table B4 (Davison et al., 2016)).

Name	Initial	Lower limit	Upper limit
wfci1	0.00001	0.00001	0.1
WF_R2	0.3	0.3	3.0

MANN	0.01	0.01	0.16
SandA	50.0	50.0	80.0
SandB	50.0	50.0	80.0
SandC	50.0	50.0	80.0
ClayA	5.0	5.0	20.0
ClayB	5.0	5.0	20.0
ClayC	5.0	5.0	20.0
LMX-NL	1.8	1.8	3.0
LMX-BL	3.0	3.0	10.0
LMX-GR	4.0	4.0	6.0
LNZ0-NL	0.0	0.0	0.69
LNZ0-BL	0.0	0.0	1.1
LNZ0-GR	-3.912	-3.912	-1.966
LMN-NL	0.5	0.5	1.8
LMN-BL	0.1	0.1	3.0
LMN-GR	0.0	0.0	4.0
ALVC-NL	0.02	0.02	0.1
ALVC-BL	0.02	0.02	0.1
ALVC-GR	0.02	0.02	0.1
CMAS-NL	10	10	40
CMAS-BL	5	5	35
CMAS-GR	1	1	5
ALIC-NL	0.1	0.1	0.3
ALIC-BL	0.2	0.2	0.4
ALIC-GR	0.2	0.2	0.4
ROOT-NL	0.05	0.05	3.0
ROOT-BL	0.05	0.05	3.0
ROOT-GR	0.05	0.05	3.0
RSMN-NL	150	150	250
RSMN-BL	75	75	175
RSMN-GR	50	50	150

Reference

Agriculture and Agri-Food Canada. (2017). The National Ecological Framework for Canada. http://sis.agr.gc.ca/cansis/nsdb/ecostrat/gis_data.html

Alavi, N., Bélair, S., Fortin, V., Zhang, S., Husain, S. Z., Carrera, M. L., & Abrahamowicz, M. (2016). Warm Season Evaluation of Soil Moisture Prediction in the Soil, Vegetation and Snow (SVS) Scheme. *Journal of Hydrometeorology*, *17*. https://doi.org/10.1175/JHM-D-15-0189.1

Amiro, B. D., Barr, A. G., Black, T. A., Iwashita, H., Kljun, N., McCaughey, J. H., … Saigusa, N. (2004). Carbon, energy and water fluxes at mature and disturbed forest sites, Saskatchewan, Canada. *Agricultural and Forest Meteorology*, *136*(3–4), 237–251. https://doi.org/10.1016/j.agrformet.2004.11.012

Arabi, M., Govindaraju, R., Sophocleous, M., & Koelliker, J. K. (2005). Use of Distributed Models for Watershed Management. In V. P. Singh & D. K. Frevert (Eds.), *Watershed Models* (pp. 502–526). CRC Press. https://doi.org/doi:10.1201/9781420037432.ch20

Arora, V. K., Chiew, F. H. S., & Grayson, R. B. (2001). Effect of sub-grid-scale variability of

soil moisture and precipitation intensity on surface runoff and streamflow. *Journal of Geophysical Research-Atmospheres*, *106*(D15), 17073–17091. https://doi.org/10.1029/2001jd900037

Avissar, R. (1991). A statistical-dynamical approach to parameterize subgrid-scale land-surface heterogeneity in climate models. *Journal of Survey*, *12*(1–3), 155–178. https://doi.org/10.1007/BF01903417

Barr, A. G., van der Kamp, G., Black, T. A., McCaughey, J. H., & Nesic, Z. (2012). Energy balance closure at the BERMS flux towers in relation to the water balance of the White Gull Creek watershed 1999-2009. *Agricultural and Forest Meteorology*, *153*, 3–13. https://doi.org/10.1016/j.agrformet.2011.05.017

Bartlett, P. A., Mackay, M. D., Verseghy, D. L., Bartlett, P. A., Mackay, M. D., Verseghy, D. L., … Verseghy, D. L. (2006). Modified snow algorithms in the Canadian land surface scheme : Model runs and sensitivity analysis at three boreal forest stands. *Atmosphere-Ocean*, *43*(3), 207–222. https://doi.org/10.3137/ao.440301

Beven, K. (2001). How far can we go in distributed hydrological modelling?, *5*(1), 1–12.

Beven, K. (2002). Towards a Coherent Philosophy for Modelling the Environment Linked references are available on JSTOR for this article : Towards a coherent philosophy for modelling the environment, *458*(2026), 2465–2484. Retrieved from http://www.jstor.org/stable/3067324

Bierkens, M. F. P., Bell, V. A., Burek, P., Chaney, N., Condon, L. E., David, C. H., … Wood, E. F. (2014). Hyper-resolution global hydrological modelling : what is next ? " Everywhere and locally relevant ." *Hydrological Processes*, *320*(October 2014), 310–320. https://doi.org/10.1002/hyp.10391

Blöschl, G., & Sivapalan, M. (1995). Scale Issues In Hydrological Modelling : a Review. *Hydrological Processes*, *9*, 251–290.

Comeau, L. E. L., Pietroniro, A., & Demuth, M. N. (2009). Glacier contribution to the North and South Saskatchewan Rivers. *Hydrological Processes*, *23*, 2640–2653. https://doi.org/10.1002/hyp

Davison, B., Pietroniro, A., Fortin, V., Leconte, R., Mamo, M., Yau, M. K., … Yau, M. K. (2016). What is Missing from the Prescription of Hydrology for Land Surface Schemes? *Journal of Hydrometeorology*, *17*(7), 2013–2039. https://doi.org/10.1175/JHM-D-15-0172.1

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*, *12*, 769–796.

Dingman, S. L. (2015). *Physical Hydrology: Third Edition*. Waveland Press. Retrieved from https://books.google.ca/books?id=rUUaBgAAQBAJ

Dornes, P. F., Pomeroy, J. W., Pietroniro, A., & Verseghy, D. L. (2008). Effects of Spatial Aggregation of Initial Conditions and Forcing Data on Modeling Snowmelt Using a Land Surface Scheme. *Journal of Hydrometeorology*, *9*(4), 789–803. https://doi.org/10.1175/2007JHM958.1

Franks, S. W., Beven, K. J., Quinn, P. F., & Wright, I. R. (1997). On the sensitivity of soil-vegetation-atmosphere transfer (SVAT) schemes : equifinality and the problem of robust calibration. *Agricultural and Forest Meteorology*, *3*, 63–75.

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,

4031(July), 4018–4031. https://doi.org/10.1002/hyp.10550
Haghnegahdar, A., Tolson, B. A., Davison, B., Seglenieks, F. R., Klyszejko, E., Soulis, E. D., … Matott, L. S. (2014). Calibrating Environment Canada's MESH Modelling System over the Great Lakes Basin. *Atmosphere-Ocean*, *52*(4), 281–293. https://doi.org/10.1080/07055900.2014.939131
Husain, S. Z., Alavi, N., Bélair, S., Carrera, M., Zhang, S., Fortin, V., … Gauthier, N. (2016). The Multi-Budget Soil, Vegetation, and Snow (SVS) Scheme for Land Surface Parameterization: Offline Warm Season Evaluation. *Journal of Hydrometeorology*, *17*. https://doi.org/10.1175/JHM-D-15-0228.1
Ireson, A. M., Barr, A. G., Johnstone, J. F., Mamet, S. D., Kamp, G. Van Der, Whitfield, C. J., … Sagin, J. (2015). The changing water cycle: the Boreal Plains ecozone of Western Canada. *Wiley Interdisciplinary Reviews: Water*, *2*(5), 505–521. https://doi.org/10.1002/wat2.1098
Kampf, S. K., & Burges, S. J. (2007). A framework for classifying and comparing distributed hillslope and catchment hydrologic models. *Water Resources Research*, *43*(5). https://doi.org/10.1029/2006WR005370
Kouwen, N. (2009). *WATFLOOD/ WATROUTE Hydrological Model Routing & Flow Forecasting System, Since 1972* (2009th ed.).
Kouwen, N., Soulis, E. D., Pietroniro, A., Donald, J., & Harrington, R. A. (1993). Grouped response units for distributed hydrologic modelling. *Journal of Water Resources Planning and Management*, *119*(3), 289–305.
Kreye, P., & Meon, G. (2016). Subgrid spatial variability of soil hydraulic functions for hydrological modelling. *Hydrology and Earth System Sciences*, *20*(6), 2557–2571. https://doi.org/10.5194/hess-20-2557-2016
Leavesley, G. H., Lichty, R. W., Troutman, B. M., & Saindon, L. G. (1983). *Precipitation-Runoff modeling system: User's manual*. Denver, Colorado.
Liang, X., Lettenmaier, D. P., Wood, E. F., & Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research: Atmospheres*, *99*(D7), 14415–14428. https://doi.org/10.1029/94JD00483
Loucks, D. P., Beek, E. Van, Stedinger, J. R., Dijkman, J. P. M., & Villars, M. T. (2005). *Water Resource Systems Planning and Management: An Introduction to Methods, Models, and Applications*. (D. P. Loucks & E. Van Beek, Eds.). Deltares and UNESCO-IHE.
Matott, L. S. (2005). *OSTRICH: An Optimization Software Tool; Documentation and User' s Guide, version 1.6*. State University of New York, Buffalo: Dep. of Civil, Structural, and Environmental Engineering. University of Buffalo.
Melsen, L., Teuling, A., Torfs, P., Zappa, M., Mizukami, N., Clark, M., & Uijlenhoet, R. (2016). Representation of spatial and temporal variability in large-domain hydrological models: Case study for a mesoscale pre-Alpine basin. *Hydrology and Earth System Sciences*, *20*(6), 2207–2226. https://doi.org/10.5194/hess-20-2207-2016
Moges, M. (2015). *Exploring the Ability of a Distributed Hydrological Land Surface Model in Simulating Hydrological Processes in the Boreal Forest Environment*.
Moriasi, D. N., J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, & T. L. Veith. (2007). Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Transactions of the ASABE*, *50*(3), 885–900. https://doi.org/10.13031/2013.23153

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

Natural Resources Canada. (2017). Land Cover, circa 2000-Vector. Retrieved from http://ftp.geogratis.gc.ca/pub/nrcan_rncan/vector/geobase_lcc_csc/shp_en/

Noilhan, J., & Mahfouf, J.-F. (1996). The ISBA land surface parameterisation scheme. *Global and Planetary Change*, *13*, 145–159.

Noilhan, J., & Planton, S. (1989). A simple parameterization of land surface processes for meteorological models.pdf. *Monthly Weather Review*.

Oreskes, N., & Belitz, K. (2001). Philosophical Issues in Model Assessment. In M. G. Anderson & P. D. Bates (Eds.), *Model Validation; Perspective in Hydrological Science* (pp. 23–41). John Wiley & Sons.

ORNL DAAC. (2017). BOREAS webpage.

Parviainen, J., & Pomeroy, J. W. (2000). Multiple-scale modelling of forest snow sublimation: Initial findings. *Hydrological Processes*, *14*(15), 2669–2681. https://doi.org/10.1002/1099-1085(20001030)14:15<2669::AID-HYP85>3.0.CO;2-Q

Pechlivanidis, I. G., Jackson, B. M., Mcintyre, N. R., & Wheater, H. S. (2011). CATCHMENT SCALE HYDROLOGICAL MODELLING : A REVIEW OF MODEL TYPES , CALIBRATION APPROACHES AND UNCERTAINTY ANALYSIS METHODS IN THE CONTEXT OF RECENT DEVELOPMENTS IN TECHNOLOGY AND. *Global NEST*, *13*(3), 193–214.

Pietroniro, A., Fortin, V., Kouwen, N., Neal, C., Turcotte, R., Davison, B., … Pellerin, P. (2007). Using the MESH modelling system for hydrological ensemble forecasting of the Laurentian Great Lakes at the regional scale. *Hydrology and Earth System Sciences*, *3*(4), 1279–1294. https://doi.org/10.5194/hess-11-1279-2007

Refsgaard, J. C. (1997). Parameterisation, calibration and validation of distributed hydrological models. *Journal of Hydrology*, *198*(1–4), 69–97. https://doi.org/10.1016/S0022-1694(96)03329-X

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, *22*(4), 367–398.

Remesan, R., & Holman, I. P. (2015). Effect of baseline meteorological data selection on hydrological modelling of climate change scenarios. *Journal of Hydrology*, *528*, 631–642. https://doi.org/10.1016/j.jhydrol.2015.06.026

Sciuto, G., & Diekkrüger, B. (2010). Influence of Soil Heterogeneity and Spatial Discretization on Catchment Water Balance Modeling. *VadoseZone*, *9*, 955–969. https://doi.org/10.2136/vzj2009.0166

Sellers, P. J., Hall, F. G., Kelly, R. D., Black, A., Baldocchi, D., Berry, J., … Guertin, F. E. (1997). BOREAS in 1997: Experiment overview, scientific results, and future directions. *Journal of Geophysical Research*, *102*(D24), 28731. https://doi.org/10.1029/97JD03300

Skøien, J. O., Blöschl, G., & Western, A. W. (2003). Characteristic space scales and timescales in hydrology. *Water Resources Research*, *39*(10), n/a-n/a. https://doi.org/10.1029/2002WR001736

Smith, S., & Marshall, I. (1996). *A NATIONAL ECOLOGICAL FRAMEWORK FOR CANADA*.

Soulis, E. D., Kouwen, N., Pietroniro, A., Seglenieks, F. R., Snelgrove, K. R., Pellerin, P., … Martz, L. W. (2005). A framework for hydrological modelling in MAGS. *Predictions in*

Ungauged Basins: Approaches for Canada's Cold Regions, 119–138. Retrieved from http://202.154.59.182/ejournal/files/A framework for hydrological modelling in MAGS.pdf

Soulis, E. D., Snelgrove, K. R., Kouwen, N., Seglenieks, F., & Verseghy, D. L. (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. https://doi.org/10.1080/07055900.2000.9649648

Tolson, B. A., & Shoemaker, C. A. (2007). Dynamically dimensioned search algorithm for computationally efficient watershed model calibration, *43*(August 2006), 1–16. https://doi.org/10.1029/2005WR004723

Verseghy, D. L. (1991). Canadian Land Surface Scheme for GCMS I. Soil model. *International Journal of Climatology*, *11*, 111–133. https://doi.org/10.1002/joc.3370110202

Verseghy, D. L., McFarlane, N. A., & Lazare, M. (1993). CLASS-A CANADIAN LAND SURFACE SCHEME FOR GCMS, 11. VEGETATION MODEL AND COUPLED RUNS. *International Journal of Climatology*, *13*, 347–370.

Vrugt, J. A., Gupta, H. V, Bastidas, L. A., & Bouten, W. (2003). Effective and efficient algorithm for multiobjective optimization of hydrologic models. *Water Resources Research*, *39*(8), 1–19. https://doi.org/10.1029/2002WR001746

Wada, Y., Graaf, I. E. M., & Beek, L. P. H. Van. (2016). Journal of Advances in Modeling Earth Systems. *Journal of Advanves in Modeling Earth Systems*, *8*, 735–763. https://doi.org/10.1002/2015MS000618.Received

Wheater, H. S. (2002). Progress in and Prospects for Fluvial Flood Modelling. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *360*(1796), 1409–1431.

Winter, T. C. (2001). The concept of hydrologic landscapes'. *JAWRA Journal of the American Water Resources Association*, *37*(2), 335–349.

Wood, E. F., Lettenmaier, D. P., & Zartarian, V. G. (1992). A land-surface hydrology parameterization with subgrid variability for general circulation models. *Journal of Geophysical Research*, *97*(D3), 2717–2728. https://doi.org/10.1029/91JD01786

Wood, E. F., Roundy, J. K., Troy, T. J., Beek, L. P. H. Van, Bierkens, M. F. P., Blyth, E., … Whitehead, P. (2011). Hyperresolution global land surface modeling : Meeting a grand challenge for monitoring Earth ' s terrestrial water. *Water Resources Management*, *47*, 1–10. https://doi.org/10.1029/2010WR010090

Wood, E. F., Sivapalan, M., Beven, K. J., & Band, L. (1988). Effects of spatial variability and scale with implications to hydrologic modeling, *102*, 29–47.