

UNIVERSITY OF CALGARY

Groundwater processes and properties of surficial deposits in the Canadian Rockies

by

Jesse He

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES  
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE  
DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN GEOLOGY AND GEOPHYSICS

CALGARY, ALBERTA

JUNE, 2021

© Jesse He 2021

## Abstract

Alpine surficial deposits such as moraine and talus are important conduits and reservoirs of groundwater and are capable of contributing large fractions of stream runoff. This work describes a combined fieldwork and modeling investigation, aimed toward characterizing the properties and hydrogeological function of various surficial deposits present within a headwater watershed in the Canadian Rockies. Hydrogeological properties such as hydraulic conductivity are highly heterogeneous between and even within surficial deposits. However, a moraine deposit located at the watershed outlet has a dominant influence on groundwater storage and flow. Restriction of discharge by this moraine “gatekeeper” has the effect of bolstering storage and baseflows over the low flow period. The function of this gatekeeper may also allow watershed-scale behavior to be well represented by simple relationships and data-frugal models which do not require detailed characterization of a site. Optimal model design must carefully select a level of complexity to avoid swinging between being over-parameterized and overly simplistic.

## Acknowledgements

I thank my supervisor, Dr. Masaki Hayashi, who has provided opportunities to learn so much, visit research sites across Alberta (and Switzerland too!) and simply do cool research. Thank you for being my steadfast mentor and for providing your knowledge and feedback over the many years.

Many thanks to the field technicians: Brandon Hill, Evan Sieben, Quinn Decent, whom I consider to be the universe's gift to graduate students. Sincere thanks to David Garcia, Tom Wilson, Vishal Singh and the many others who have helped make the fieldwork possible.

This experience has been immensely improved by the many graduate students who've shared in the journey. I specifically acknowledge Ben Roesky, who has been my partner-in-crime through many wearying trips across the mountains. I also acknowledge Laura Beamish and Polina Abdrakhimova for putting up with the brunt of my questions and requests for help.

Access to the field site was provided by the Fortress Mountain Resort. Thanks to Mel, the groundskeeper, for all the tips out there in the field, and to Chevy for accommodating us.

Thanks are also due to Dr. John Pomeroy and the Coldwater Research Lab in Canmore. I am grateful to Logan Fang, Greg Galloway and Eric Courtin for sharing data and expertise.

This research was made possible by funding and scholarships from Global Water Futures (sponsored by the Canada First Research Excellence Fund), Natural Sciences and Engineering Research Council of Canada, Alberta Government and the Canadian Water Resources Association.

I also thank the members of my defense committee: Dr. Larry Bentley and Dr. Martyn Clark for their feedback.

Finally, I thank my friends and family for keeping me sane.

## Table of Contents

Abstract .....	ii
Acknowledgements.....	iii
Table of Contents .....	iv
List of Tables .....	vi
List of Figures .....	vii
1. Introduction.....	1
1.1 Context.....	1
1.2 Thesis organization and objectives .....	2
2. A field characterization of groundwater flow and storage.....	3
2.1 Introduction.....	3
2.2 Study site.....	4
2.3 Methods.....	8
2.4 Results.....	14
2.5 Discussion .....	21
2.6 Figures and Tables .....	25
3. Simulation of the Hathataga groundwater system .....	43
3.1 Introduction.....	43
3.2 Model codes .....	45
3.3 Model construction and implementation .....	47

3.4 Model inter-comparison.....	53
3.5 Results.....	56
3.6 Discussion.....	61
3.7 Figures and Tables .....	70
4. Synthesis and Conclusions.....	88
References.....	91

## List of Tables

Table 2.1 Piezometer specifications. Datum: UTM NAD83, Zone 11N.....	25
Table 2.2 Temperature and electrical conductivity of talus springs measured in summer 2018..	26
Table 3.1 Initial parameters of each model setup. Parameter ranges were determined from measurements (see Section 2.3.6) and educated opinion.....	70
Table 3.2 Estimated values and 95% confidence interval bounds of parameters for all model setups.....	71
Table 3.3 Calculated model NSE and log-NSE for discharge and water levels (when applicable) during calibration and validation. ....	72

## List of Figures

Figure 2.1 (a) Inset map showing location of study site. (b) Close-up view of the outlet spring complex showing three spring branches. (c) Map of Hathataga Watershed showing topography, delineated surface cover, hydrological features and instrumentation. Extent of (b) indicated by red box. Contour interval 50 m. Delineated area of potential ground ice is an approximate estimate. ....	27
Figure 2.2 Surficial landforms within the watershed. (a) (July 13, 2018) Moraine ridges and Hathataga Lake with talus and headwall in background. Image captured from the northern end of the watershed looking south. (b) (July 10, 2019) Meadow and stream, looking east. (c) (July 10, 2019) Talus, looking west. ....	28
Figure 2.3 Map of the Hathataga, Confluence and Galatea watersheds. Red square indicates extent of the inset map. Contour interval 100 m. ....	29
Figure 2.4 Hypsometric curves of Hathataga, Confluence and Galatea watersheds. Normalized area is presented as the fraction of watershed area at or above a given elevation. ....	30
Figure 2.5 (a) 2018 time series of discharge. Discharge was normalized using a catchment area of 0.92 square kilometers; some uncertainty is associated with this normalization due to the western boundary. (b) 2018 time series of hourly average air temperature and daily precipitation. (c) 2019 time series of discharge. (d) 2019 time series of air temperature and precipitation. ....	31
Figure 2.6 (a) Volumetric and (b) area normalized discharge from the Hathataga, Confluence and Galatea watersheds in 2018. Precipitation at AWS1 is also shown. ....	32
Figure 2.7 Piezometer and lake levels in (a) 2018 and (b) 2019. Gaps signify missing data. Dotted lines indicate periods when the piezometer was dry. ....	33
Figure 2.8 Plots of measured deuterium vs oxygen-18 abundance. Local meteoric water line shown as dashed line. (a) Precipitation and snow samples, red outline indicates extent of (b), (c) and (d). (b) Talus springs and waterfall. (c) Groundwater in the moraine and meadow and Hathataga Lake. (d) Discharge from the Hathataga, Galatea and Confluence watersheds. ....	34
Figure 2.9 Major ions measured in samples collected from Hathataga over summer 2018 presented in Stiff diagrams. ....	35
Figure 2.10 Major ions measured in samples of outflow from the Hathataga, Confluence and Galatea watersheds over summer 2018 presented in Stiff diagrams. ....	36
Figure 2.11 Measured lake temperature. (a) 2018. (b) 2019. ....	37
Figure 2.12 Time series of water temperature and EC of Hathataga Outlet discharge measured from a logger deployed at the gauging station (GS1). Manual measurements (shown as points) are used to supplement gaps in logger data. ....	38
Figure 2.13 Time series of (a) EC and (b) water temperatures of discharge from the Hathataga, Galatea and Confluence watersheds in 2018. ....	39
Figure 2.14 (a) Initial peak tracer breakthrough recorded for the moraine and talus tracer tests. (b) Delayed breakthrough at Hathataga Outlet after injection of the talus tracer test. ....	40
Figure 2.15 Hydraulic conductivities measured in different surficial landforms and using different methods. ....	41

Figure 2.16 (a) Time series of calculated moraine hydraulic conductance over 2018 and 2019. (b) Plot of hydraulic conductance against water level in the Lake piezometer. ....	42
Figure 3.1 Graphic illustrating structure of the simple dynamical model reservoir. ....	73
Figure 3.2 Example SFASH model domain. The watershed is represented by a structured 20 m grid. Grid cells are coloured according to discretization of land cover or boundary conditions. White cells are inactive. ....	74
Figure 3.3 Interpolated elevation of bedrock below surficial deposits in Hathataga. Contour interval is 20 m. ....	75
Figure 3.4 Graphic illustrating the differences between model setups. Distributed models (setups 2 to 5) are shown in plan and cross section view. Relative process and parameter complexity is expressed by position on the continuum. ....	76
Figure 3.5 Temporal distribution of model recharge for all recharge scenarios. ....	77
Figure 3.6 Calibration performance of all model setups. Model output is plotted in red and observations in black. ....	78
Figure 3.7 Validation performance of all model setups. Model output is plotted in blue and observations in black. ....	79
Figure 3.8 (a) Estimated discharge sensitivity function of setup 1 over the range of modeled discharge (0.2 mm/d to 15 mm/d). (b) Storage – discharge curve based on the discharge sensitivity function. Values of storage are relative because absolute reservoir storage cannot be determined. ....	80
Figure 3.9 Mean simulated total February-March baseflow for all model setups and recharge scenarios. Error bars span two standard deviations above and below the mean. ....	81
Figure 3.10 Mean difference in total February-March baseflow between recharge scenario 1 (the reference) and recharge scenarios 2 to 4. Differences are expressed as a percentage of total baseflow from scenario 1. Error bars span two standard deviations above and below the mean. ....	82
Figure 3.11 Storage-discharge and discharge sensitivity curves of all model setups. Curves from calibration, validation and pure drainage simulations are shown. Storage-discharge curves are shown in terms of relative and absolute model storage. Relative storage was plotted by setting a zero value equal to the minimum storage of the drainage simulation and shifting all curves accordingly. ....	83
Figure 3.12 Discharge sensitivity and storage-discharge curves of the SDM using an alternative $g(Q)$ function and SFASH setup 5. ....	84
Figure 3.13 (a) Calibration and (b) validation performance of the SDM using an alternative discharge sensitivity function. ....	85
Figure 3.14 Discharge sensitivity and storage-discharge curves of two SDM upper and lower reservoirs connected in series. ....	86
Figure 3.15 (a) Calibration and (b) validation performance of two SDM reservoirs connected in series. ....	87

# 1. Introduction

## 1.1 Context

A safe and stable water supply is essential for both human communities and the natural environment. For many, this supply comes as river flows from mountain watersheds. In Calgary for example, nearly all water needs are provided by the Bow and Elbow Rivers which have headwaters in the Canadian Rocky Mountains. Globally, mountain regions are a critical water supply for ~40% of the world's population (Viviroli et al., 2007).

Future water security in mountain regions is uncertain and will be affected by changes in supply and demand. In terms of supply, warming climate is expected to affect the timing and amount of snowmelt (Bales et al., 2006; Barnett et al., 2005; Beniston and Stoffel, 2014; Jefferson et al., 2008; Mankin et al., 2015; Mote et al., 2005). Change in recharge may shift peak flows earlier and reduce the flows available for later summer and autumn months (Barnett et al., 2005; Stewart et al., 2005; Tague and Grant, 2009). Demand for water is also expected to rise, both in the mountains where development and water extraction is increasing and in lowland regions (Viviroli et al., 2020).

Coarse surficial deposits (e.g., moraine, talus) in mountain ranges are becoming recognized as important reservoirs and contributors to stream flow (Arnoux et al., 2020; Clow et al., 2003; Baraer et al., 2015; Hayashi, 2020; Lauber et al., 2014). Such deposits can supply a majority of discharge in alpine streams especially in the baseflow period (Cochand et al., 2019; Floriancic et al., 2018; Hood and Hayashi, 2015) and may act as a potential buffer to adverse effects from climate change (Tague et al., 2008). Storage and release of water from deposits is influenced by various hydrological attributes/properties. These may include: underlying bedrock topography/geometry (Cochand et al., 2019; Muir et al., 2011; Langston et al., 2011, 2013; McClymont et al., 2010; Harrington et al., 2018); buried ice (Roy and Hayashi, 2009); variations in hydraulic conductivity and other structures (Caballero et al., 2002; Winkler et al., 2016; Harrington et al., 2018).

The distribution and properties of surficial deposits is subject to great spatial variability, which, in turn, varies the hydrological behavior of different watersheds. The extent and effect of this variability is under-characterized in mountain regions due to the logistical difficulty of working in remote and rough terrain. Consequently, there is lacking ability to generalize the results of small watershed or deposit specific work to larger regions. Data scarcity also handicaps efforts to build and calibrate models of alpine regions. These are shortcomings that must be

addressed before characterization and prediction at spatially meaningful scales can take place. Moving forward, there is need for case studies in new study sites which will expand the foundation of alpine hydrological knowledge (Hayashi, 2020). At the same time, conceptualization and modeling of hydrologic behavior at larger scales should be re-examined (Kirchner, 2006). Modeling methodology should be tuned to be as data frugal as allowable, meaning that models only include complexity needed to represent relevant hydrological processes (Hill, 2006).

### 1.2 Thesis organization and objectives

This work is a case study of a new research site in the Canadian Rocky Mountains and is divided into two major parts. In the first part (Chapter 2), data collected from a multi-year field campaign are presented and discussed. These data include observations of hydrological fluxes and properties. Analysis and discussion focuses on characterizing spatially varying properties across the watershed and exploring how these properties may integrate and influence hydrologic behavior at the watershed scale.

The second part (Chapter 3) focuses on the setup and inter-comparison of an ensemble of numerical groundwater flow models for the site. Models in the ensemble span different modeling methodologies (e.g., lumped versus distributed) and levels of parameterization complexity (e.g., homogenous versus heterogeneous). Models are used to explore the effect of different hydrological properties and heterogeneity on overall watershed discharge. Sensitivity of watershed behavior to simple recharge perturbations is also tested. These exercises are used to investigate the benefits and disadvantages of added model complexity and determine more optimized model representations.

The overall objective of this work is to quantify hydrogeological properties of surficial deposits across an alpine watershed and determine how they influence groundwater storage and discharge at the scale of the watershed. Through subsequent modeling work, collected knowledge and understanding is transferred to quantitative numerical models. These models are used with the further aim of evaluating the efficiency and performance of different model designs and determining avenues for improvement.

## 2. A field characterization of groundwater flow and storage

### 2.1 Introduction

Human communities and ecosystems alike depend on flows from mountain regions. In the mountains, coarse surficial deposits (e.g., moraine, talus) are becoming increasingly recognized as important aquifer units (Hayashi, 2020). Groundwater storage within these deposits is filled during the spring and summer by melting snow and precipitation. Over time, stored groundwater is released slowly to streams. This contribution can be important. For example: in the Colorado Rocky Mountains Liu et al. (2004) and Clow et al. (2003) report talus contributions to streamflow of greater than 60% and 75% respectively; In a sub-watershed of the Lake O’Hara hydrological observatory in the Canadian Rockies, a proglacial moraine contributes 70 – 80% of total outflow (Paznekas, 2016).

Many case studies have indicated the important connection between various geological properties and hydrogeological processes (Woodruff and Hewlett, 1970; Weekes et al., 2015). In alpine aquifers, different attributes have been found to influence how groundwater is stored and transmitted. For example, grain size distribution and internal structure have an important impact on aquifer behavior. Fine sediments have low hydraulic conductivity and act as “long-term” storage for groundwater (Caballero et al., 2002; Clow et al., 2003). Layering of coarse and fine materials can cause rapid and delayed flow paths to develop in a given landform. This heterogeneity influences aquifer response to recharge (e.g., Kurylyk and Hayashi, 2017) and can result in complex discharge behavior, which reacts sharply to recharge but also persists for long periods without (e.g., Harrington et al., 2018; Winkler et al., 2016; Kobierska et al., 2015). Aquifer geometry and bedrock topography can also be important controls. In Lake O’Hara, bedrock depressions were found to bolster storage of water in talus and moraine as well as influence flow paths (Muir et al., 2011; Langston et al., 2011). Bedrock topography was also found to regulate discharge and water residence time in a nearby talus-meadow complex (McClymont et al., 2010).

The distribution of surficial deposits and their properties is subject to great spatial variability across mountain regions owing to differences in lithology, sediment deposition and weathering. This variability gives rise to complex, heterogeneous behavior in different locations and at different scales. For example, across a watershed in the Swiss Alps, Floriancic et al. (2018) found large spatial variability in storage potential and baseflow contribution. Observed variability in hydrologic behavior was linked to differences in distribution and properties of

sediment cover. At larger scales, Paznekas and Hayashi (2016) observed a wide range in winter specific discharge (discharge normalized to watershed area) magnitude between watersheds in the Canadian Rockies; Tague and Grant (2004) also found great variability in discharge and recession behavior of watersheds in the Cascade Mountains of Oregon. Both studies indicate a link between discharge behavior and geologic variability.

The properties and heterogeneity of small aquifer units can shape the hydrological behavior of watersheds which, in turn, affect river systems. This link between small-scale details and larger-scale behavior may be captured by simpler, integrative relationships such as watershed storage-discharge curves (e.g., Kirchner, 2009). Storage-discharge curves and other generalizations facilitate study of larger scale regions which generate meaningful quantities of streamflow. Unfortunately, it is unclear how these generalized relationships are quantified and linked to physical properties of the watershed. More field-based research is required to characterize aquifers, their variability, and how they interact to define integrated behavior. To this end, a field campaign was established in and around a small, headwater watershed in the Front Ranges of the Canadian Rocky Mountains. Following previous geophysical work (Christensen et al., 2020), this work attempts to further characterize the properties of landforms across the watershed and investigate their influence on overall watershed hydrological behavior. This analysis focuses on answering the following research questions:

1. What heterogeneity exists within and between surficial units of the site? How does it affect flow paths/response to forcing within the site?
2. How is watershed-scale discharge influenced by interaction and variability of different surficial deposits and their properties? In other words, how does fine-scale heterogeneity scale up in terms of discharge?
3. What proportion of baseflow does the site contribute compared to nearby regions? What factors influence this?

## 2.2 Study site

### *2.2.1 Location and climate*

The study was conducted at the Fortress Ski Area in the Front Ranges of the Canadian Rocky Mountains (50°49'14" N, 115°12'50" W), located 80 km west of Calgary, Alberta (Figure

2.1a). Focus is on a 0.92 km<sup>2</sup>, north-facing, headwater cirque (Figure 2.1c), drained by an outlet spring at its northern tip (Figure 2.1b). This watershed is unofficially labeled “Hathataga” based on naming by the local Stoney people (Crawler et al., 1987). Topography here is rugged, with elevations spanning from 2082 meters above sea level (masl) at the northern outlet to 2850 masl at the headwall in the southern end. Mean annual precipitation at an automatic weather station (AWS1) (Figure 2.1b) was 1129 mm over the years 2016 – 2018. Precipitation occurs mostly as snow from the late fall to early spring period. Average January and July temperatures were -8.8 °C and 11.0 °C, respectively. From the watershed, little evapotranspiration is expected due to limited vegetation cover and shading by the steep headwalls to the south (Hrach, 2020). Higher rates of evapotranspiration will occur in more heavily vegetated/forested areas, but these areas (delineated in Figure 2.1c) make up a relatively small (< 20%) fraction of the watershed.

### 2.2.2 *Bedrock geology*

Bedrock at the southern end of the watershed is predominantly carbonate of the Devonian Palliser Formation (Pana and Elgr, 2013). This thickly bedded and folded unit forms steep headwalls, which feed debris and water to the talus at their foot. The Palliser Fm. is capped by rock of the Exshaw and Banff Formations (Pana and Elgr, 2013). These are more recessive black shales and wackestone which form mountain peaks. Lower-lying regions to the north are underlain by fissile, highly recessive shales and sandstones of the Jurassic Fernie Formation. The Sulfur Mountain Thrust Fault passes through the watershed, separating the Palliser Fm. to the south and the Fernie Fm. to the north (Pana and Elgr, 2013).

Geophysical imaging by Christensen et al. (2020) show generally flat bedrock topography in the northern end of the watershed, underlying moraine. However, some depressions imaged near the outlet spring may indicate the presence of bedrock channels. South of the moraine, bedrock elevations drop, forming a depression which spans 200 – 400 m in length with depths 30 – 40 m below ground surface. As suggested by Christensen et al. (2020), this is likely a glacial overdeepening, similar to those observed in other mountain ranges across the world (Haeberli et al., 2016). At places further south, bedrock underlying the talus appears to have “stepped” topography. Outcrop and geophysical observations also indicate that bedrock underlying much of the watershed may be intensely weathered and fractured at depth.

### 2.2.3 Surficial geology

Several surficial deposits are present within the watershed (Figures 2.1 and 2.2). Steep talus cones and slopes (slope angle generally  $>30^\circ$ ) occur along the southern bedrock headwall. This talus exhibits great lateral heterogeneity at the surface with texture that ranges from “coarse with large, open void spaces” to “clast-supported with all voids filled” (Selby, 1982). This suggests a complex history of deposition and reworking by rockfall, fluvial and avalanche processes (White, 1981). These processes have been observed first-hand during field work. During summer, small waterfalls and rock debris fall onto talus from couloirs above. During the winter, avalanches and rockfall continue to deposit snow and debris onto the talus tops with apparent regularity. Regions of talus with higher fraction of fines are easily visible due to higher vegetation cover (Figure 2.2c). Talus was observed up to 60 m thick in the mid-slope, but decreased to 20 – 30 m downslope (Christensen et al., 2020). Previous geophysical work indicates the presence of ground ice in the talus closer to the eastern edge of the watershed (Figure 2.1c).

A small alpine meadow occurs north of the talus (Figure 2.2b). This grassy region is generally flat, with an area of approximately 5000 m<sup>2</sup>. Geophysical measurements and auger-holes identified an up to ~11 m thick layer of fines below the meadow, which are likely glaciolacustrine in origin (Christensen, 2017). The meadow is bordered by coniferous forest to the east and west. Sediment below forest mainly consists of thin soil and poorly sorted till, however, a fine layer of reddish ash, likely Mazama ash (Bacon, 1983), was found occurring at many places between the soil and till. Excavations along the hillslope just north of Hathataga reveal a similar structure: thin soil covering several meters of till underlain by bedrock.

A recessional moraine is located at the northern end of the watershed. This moraine is composed of poorly sorted, angular, predominantly carbonate clasts. This unit forms high ridges which encircle the lake area. Moraine ridges can reach thicknesses of tens of meters. In places, these ridges are cut-through by channels, possibly from past moraine-dam outbursts. The moraine contains a large fraction of fines. Near the edges, it is overlain by thin soils and vegetation. In the vicinity of the lake, a thin (<2 m) layer of fines covers the moraine. In late-summer, after the lake dries, “drain holes” are visible through this cover. Drain holes are conduits for water to the coarser sediment below. A small moraine is also observed with the talus at the eastern end of the watershed.

#### *2.2.4 Hydrological characteristics*

The Hathataga outlet is a perennial spring complex which drains the watershed from its northern tip. Maximum elevation of the spring complex is highest during the snowmelt period and drops 1 – 2 m with time as higher elevation springs dry up. Spring discharge peaks during snowmelt and begins to consistently recede when snow depletes (around early July in 2018 and 2019). Flows from the spring complex emerge as three distinct stream branches which originate from the western, center and eastern ends of the complex (Figure 2.1b). These branches merge to produce a single outlet stream that flows toward a confluence with Galatea Creek about ~5 km to the north (Figure 2.3).

Other springs are present within the watershed. Several outcrop at the mid-slope and base of talus deposits. Talus base springs at the western and eastern ends of the meadow feed small, losing streams (Figure 2.1 and 2.2b) which join to form a small inlet to Hathataga Lake. These springs/streams flow throughout the spring and summer, but cease during the fall and winter. Discharge of these features is highly sensitive to precipitation events. After drying out in late summer, talus springs and the inlet were observed to “reactivate” after precipitation events. “Groundwater discharge points” (GWD) are also present at the south shore of Hathataga Lake. They release consistent discharge through the summer, becoming exposed as lake levels drop.

At the center of the moraine, Hathataga Lake fills rapidly from snowmelt in the spring and drains over the summer and fall months. At its peak in spring the lake is approximately 2 m deep, but drains dry by late summer. The lake receives water primarily from groundwater (Christensen et al., 2020), as evidenced by the GWD on the southern shore. Additional input comes from the small talus-spring fed inlet stream during the summer and fall. The majority of water leaving the lake is also via groundwater, likely via rapid flow paths through the drain holes observed at the lake bottom. Some water is also lost through evaporation.

#### *2.2.5 Confluence and Galatea*

For comparison, two nearby watersheds are also included in the study (Figure 2.3). The Confluence watershed spans the entire valley and contains Hathataga at its southern end. This 6.5 km<sup>2</sup> watershed is drained by the stream which begins at the Hathataga outlet spring and ends at its confluence with Galatea Creek. Galatea is a 27.8 km<sup>2</sup> watershed, northwest-adjacent to Confluence. It is drained by Galatea Creek which collects flows from multiple sub-watersheds to

the confluence. Given their proximity, Confluence and Galatea have similar climate and geology to each other and Hathataga. However, both are multiple times larger than Hathataga and encompass several small headwater cirques as well as large downstream regions. Compared to Hathataga, Confluence and Galatea consist of a greater proportion of lower elevation (Figure 2.4) and more heavily forested area. Some coarse deposits like talus are present at higher elevations. Little infrastructure exists within Galatea, however, an active pumping station is located on the stream draining Confluence (Figure 2.3). This station was mainly used as a personal water supply for a site caretaker and therefore is not expected to have greatly affected flows in the stream over the study period.

## 2.3 Methods

### *2.3.1 Meteorological measurements*

An automated weather station (AWS1) was installed by the University of Saskatchewan in a clearing just north of Hathataga watershed (Figure 2.1). It is equipped with: an air temperature and relative humidity sensor (Rotronic, HC2-S3); a vane anemometer for wind speed and direction (R. M. Young, 05103); weighing cumulative precipitation gauge (Geonor, T200B with Alter shield); an ultrasonic snow depth sensor (Campbell Science, SR50); and a four component radiometer (Kipp & Zonen, CNR4). Hourly averaged data were manually corrected for noise/errors. Precipitation data were corrected for wind undercatch using the equation of Kochendorfer et al. (2017, Eq. 3).

### *2.3.2 Discharge*

Discharge from Hathataga was measured at a gauging station just north of the spring complex (GS1). Discharge was also measured at several locations downstream, including points on the Hathataga outlet creek (GS5) and Galatea Creek, just upstream of the confluence (GS6) (Figure 2.3). Each gauging station consists of a pressure transducer (Solinst, Levelogger, non-vented; In-situ, Minitroll, vented) in a stilling well installed at the stream bank. Stream stage was recorded at 15-minute intervals throughout the year. To verify transducer data, manual water levels were measured periodically from staff gauges, or from stilling well tops. Measurements from non-vented transducers were compensated with barometric pressure from nearby loggers (Solinst, Barologger) or AWS1 (Campbell Scientific, CS106). Noise from maintenance/freeze-up were manually removed from the data.

To establish a discharge-stage rating curve, stream discharge at each gauging station was measured regularly. Average stream depth and flow rates were recorded with a horizontal-axis flow meter (Global Water, FP101) in segments spanning the width of the stream. Total discharge was calculated from these measurements using the area-velocity method (Dingman, 2002, p.609). The rating curve was then determined by fitting a power function to a plot of discharge vs stream stage. Using this function, a near-continuous record of stream discharge was calculated.

### *2.3.3 Groundwater and lake levels*

Several piezometers were installed at various depths and locations in the moraine and meadow (Figure 2.1; Table 2.1). These piezometers are made of coupled lengths of 1.9-cm (inside diameter) black iron pipe coupled to a screened, 15 or 30 cm drive-point (Solinst, 615N) wrapped in geotextile fabric. Piezometers were installed by hand-auguring a 5 cm diameter hole, inserting the piezometer, then driving them deeper with a slide hammer. Piezometer P05, installed by Christensen (2017), was also used. This piezometer was installed by inserting a PVC pipe into a hand-augered hole in the meadow. Groundwater level was measured and recorded by vented pressure transducers (In-situ, Minitroll; In-situ, Leveltroll) placed at the bottom of each piezometer. Water levels of Hathataga Lake were recorded by a pressure transducer (Solinst, Levellogger) hung in a stilling well. Transducers recorded water levels at 15-minute intervals. These readings were verified periodically by manual water level measurements.

### *2.3.4 Water chemistry*

Water samples were collected routinely from Hathataga Lake and gauging stations. During collection, samples were filtered through 0.45  $\mu\text{m}$  membrane filters into pre-rinsed polyethylene bottles. Regular samples were taken every 1-2 weeks in the summer, and monthly during the winter. Over the spring and summer of 2018, an autosampler (Teledyne Isco, 6712) was placed beside GS1 (Figure 2.1) and collected water samples at daily to weekly intervals. Precipitation was collected every 1-2 weeks in the spring and summer from rainfall collectors deployed within the site (Figure 2.1b). In the winter, snow samples were taken at 20 cm depth intervals from snow pits near AWS1. These samples were combined and melted into one depth-integrated snow sample before analysis.

Temperature and electrical conductivity (EC) of lake, spring and stream waters were measured with hand-held meters during routine sample collection (VWR, 2052-B; Omega Engineering, HH-25TC). All reported EC values were standardized to 25 °C (Hayashi, 2004). Temperature and EC were also logged at a 15-minute interval at the Hathataga outlet by a sensor and data logger (Campbell Scientific, CS547a; Campbell Scientific, CR1000). After collection, samples were stored at 4 °C until analysis. Deuterium and oxygen-18 abundances were measured in samples using an off-axis integrated-cavity spectrometer (Los Gatos Research, DLT-100). Stable isotopes are reported as permil ratios with respect to Vienna Standard Mean Ocean Water (Gonfiantini, 1981). Samples were analyzed for alkalinity using the Gran function plot method (Stumm and Morgan, 1981).

In the summer of 2018, additional samples were collected from several piezometers and smaller springs/streams (Figure 2.1). Piezometers were sampled using a peristaltic pump. When possible, piezometers were pumped out three times and left to refill before sampling. A sample was also taken from a waterfall (Figure 2.1), recharging the top of a talus cone from the bedrock above. These additional samples, along with concurrent “regular” samples, were analyzed for temperature, EC, stable isotopes, alkalinity, and major ions using ion chromatography (Metrohm, 930 Compact IC Flex).

### *2.3.5 Tracer tests*

Two tracer tests were done in the Hathataga watershed (Figure 2.1c). Uranine, a yellow/green fluorescent dye, was used for both tests. This tracer was chosen because it is fairly conservative and has small environmental impact (Taylor and Greene, 2005). The Efficient Hydrologic Tracer-Test Design (EHTD) model (Field, 2008) was used to inform the quantity of tracer to use and estimate timing and magnitude of tracer breakthrough.

The first test took place on September 26, 2018 on the Hathataga lakebed. The tracer was prepared by dissolving 500 g of powdered uranine in approximately 40 L of water from nearby ponds/trickles on the lake bed. The tracer was poured directly into an exposed drain hole at the western end of the lakebed (MTS in Figure 2.1c) and flushed with another ~100 L of water. Injection and flushing of tracer began at ~10:30 and finished at around 10:45. A fluorometer (Turner Designs, Cyclops-7f) connected to a datalogger (Campbell Scientific, CR1000) was deployed at GS1 to monitor tracer breakthrough from the spring.

The second test took place on July 10, 2019 on a talus cone (TTS in Figure 2.1c). 500 g of uranine was dissolved in ~40 L of water from a nearby mid-slope spring (SP3). Tracer was poured onto a spring fed stream which outcropped for a few meters at the surface of the talus before re-infiltrating. Injection began at 13:40 and was flushed with ~ 100 L of spring water until 14:00. To monitor tracer breakthrough, a fluorometer was placed downslope, on a meadow stream fed by springs at the base of the same talus cone. The Hathataga outlet fluorometer was present during the second test as well.

Beforehand, fluorometers were calibrated in a laboratory with eight concentration standards (0, 0.5, 4, 100, 400, 700, 1000, 2000 ppb) prepared from uranine powder and water collected from the Hathataga outlet. Calibration curves relating concentration and output voltage followed a linear trend from the minimum detection limit of 0.01 ppb to an upper limit of about 500 ppb. Beyond the upper limit, curves became non-linear, flattening at the maximum voltage output of 5 V. All concentrations calculated with this curve are with respect to the uranine powder used and not the uranine active ingredient.

### 2.3.6 Hydraulic conductivity

#### Slug tests

Slug tests were conducted in piezometers in the moraine and meadow. Before testing, piezometers were developed with a peristaltic pump or manually using tubing with a foot valve. In lieu of a physical slug, water was poured into piezometers to induce a rise in head. Although not instantaneous, rise in head was still rapid in the narrow, open space within piezometers. Afterward, water levels were allowed to reset. Tests were repeated 2 – 5 times. Testing of piezometer P01d was handled differently: it was pumped out and recovery was monitored over three weeks to determine hydraulic conductivity.

Hydraulic conductivity was calculated from recorded water levels and piezometer dimensions with the Hvorslev (1951) method:

$$K = \frac{r^2 \ln\left(\frac{L}{R}\right)}{2L(t_2 - t_1)} \ln\left(\frac{H_1}{H_2}\right) \quad [2.1]$$

where:  $K$  is hydraulic conductivity ( $\text{m s}^{-1}$ );  $r$  is standpipe inner radius (m);  $L$  is screen length (m);  $R$  is screen outer radius (m);  $H$  and  $t$  are values of two points picked from the early-time straight-line section of a semi-log graph of head (m) and time (s).

### Permeameter measurements

Field-saturated hydraulic conductivity ( $K_{fs}$ ) was measured at four separate locations in the forest northeast of Hathataga using a field-portable constant head permeameter (SoilMoisture, Guelph Permeameter). Values of  $K_{fs}$  are taken as analogous to other measurements of hydraulic conductivity (equation 2.1 and 2.3). During measurement, the permeameter was set up and lowered into 6 cm diameter hand-augered holes. Water was released into the hole at a fixed hydraulic gradient. Once steady state was reached, discharge rates were recorded. This process was repeated at depths ranging from 10 cm to 80 cm below surface in different layers at each measurement point. Field saturated hydraulic conductivity was calculated by:

$$K_{fs} = \frac{Q_s}{\left(\frac{2\pi H^2}{C_p} + \pi a^2 + \frac{2\pi H}{\alpha \times C_p}\right)} \quad [2.2]$$

where  $K_{fs}$  is field-saturated hydraulic conductivity ( $\text{m s}^{-1}$ );  $Q_s$  is steady state discharge ( $\text{m}^3 \text{s}^{-1}$ );  $H$  is depth of water in the hole (m);  $\alpha = 12 \text{ m}^{-1}$  is a coefficient representing the soil capillary length, the value  $12 \text{ m}^{-1}$  is applicable to a wide range of structured soils (Elrick et al., 1989);  $C_p$  is a dimensionless shape factor dependent on piezometer dimensions;  $a$  is the radius of the hole (m).

### Tracer tests

Hydraulic conductivity at larger scales ( $10^1 - 10^2 \text{ m}$ ) was estimated in the moraine and talus using average groundwater flow rates from artificial tracer tests. Average flow rates were calculated as the time between injection and peak concentration in the breakthrough curve divided by distance between the injection and recovery points. Hydraulic conductivity was calculated using Darcy's Law:

$$K = \frac{v \times n_e \times \Delta x}{\Delta h} \quad [2.3]$$

where  $K$  is hydraulic conductivity ( $\text{m s}^{-1}$ );  $v$  is average linear velocity ( $\text{m s}^{-1}$ );  $n_e$  is effective porosity;  $\Delta h$  is head difference between the tracer injection and recovery points;  $\Delta x$  is the lateral distance between the tracer injection and recovery points.

Effective porosity was assumed  $0.3 \pm 0.1$  for coarse sediment (Freeze and Cherry, 1979, p37). Head differences were taken from piezometer measurements (on the moraine) or from surface spring elevation difference (on the talus). Lateral distances were taken as the length of a straight line drawn from the injection and recovery points.

### Hydraulic conductance

Relative changes in hydraulic conductivity of the moraine were estimated using piezometer water levels and outlet discharge (assuming outlet discharge is predominantly sourced from the moraine – see section 2.5.2 for justification). Hydraulic conductance between the moraine and outlet spring is defined as:

$$C = \frac{Q}{\Delta h} \quad [2.4]$$

where  $C$  is hydraulic conductance ( $\text{m}^2\text{s}^{-1}$ );  $Q$  is volumetric discharge at the outlet ( $\text{m}^3\text{s}^{-1}$ );  $\Delta h$  is the difference in head between the moraine and outlet spring piezometers.

Hydraulic conductance is typically used with Cauchy (type III) boundary conditions in numerical groundwater models (Mehl and Hill, 2010) and can be related back to conductivity with the following equation:

$$C = \frac{K_{av}A}{\Delta x} \quad [2.5]$$

where  $K_{av}$  is average hydraulic conductivity ( $\text{m s}^{-1}$ ) over the saturated cross-sectional area ( $A$ ,  $\text{m}^2$ ) perpendicular to flow;  $\Delta x$  is the flow path length between the moraine and outlet spring piezometers.

Both  $A$  and  $\Delta x$  of equation 2.5 are unknown and are assumed to be roughly constant. In reality,  $A$  will vary as water levels rise and fall, but to a much smaller extent than hydraulic conductivity, which often varies across orders of magnitudes. The head gradient between piezometers is also assumed approximately linear. A time series of conductance was calculated from discharge and water levels time series. Variation in conductance can be used to infer how flow paths in the moraine change over time which can indicate how conductivity varies with depth.

#### *2.3.7 Geospatial data*

Depositional landforms present in the Hathataga watershed were identified by field observations, interpretation of aerial photographs (Alberta Environment and Parks, 2014) and previous work by Christensen et al. (2020). Landforms were delineated in ArcMap (ESRI, 2014) with the following categories: exposed bedrock, talus, alpine meadow/forest, moraine.

Topographic data were captured in a 2 m resolution digital elevation model (DEM) provided by the University of Saskatchewan and supplemented by the 20 m Canada DEM (Natural Resources Canada, 2016). These DEMs were used to determine the extents of the Hathataga and several nearby watersheds using ArcMap (ESRI, 2014). Features such as springs and field equipment were surveyed using a differential global positioning system (Hemisphere, S321 Smart Antenna). All coordinates and elevations are referenced to a nearby Alberta Survey Control Marker (no. 486340).

## 2.4 Results

### *2.4.1 Discharge of Hathataga and surrounding regions*

During the spring and early summer, discharge was controlled by snowmelt and fluctuated following periods of high and low temperature (Figure 2.5). Peak discharge rates were 30.6 mm d<sup>-1</sup> (0.326 m<sup>3</sup> s<sup>-1</sup>) and 23.5 mm d<sup>-1</sup> (0.250 m<sup>3</sup> s<sup>-1</sup>) in 2018 and 2019 respectively. No strong diurnal fluctuations in discharge were observed during snowmelt.

Discharge recession began as snow cover depleted, typically by mid-July, and was largely uninterrupted until snowmelt of the following year. Average September discharge rates were 2.0 mm d<sup>-1</sup> (0.021 m<sup>3</sup> s<sup>-1</sup>) and 4.0 mm d<sup>-1</sup> (0.043 m<sup>3</sup> s<sup>-1</sup>) in 2018 and 2019 respectively. Higher baseflows in 2019 may have resulted from delayed snowmelt and/or greater amount of Jul-Aug precipitation (245 mm in 2019 vs 127 in 2018). Winter flow rates (January 1 to May 1) ranged between 0.27 mm d<sup>-1</sup> (0.0029 m<sup>3</sup> s<sup>-1</sup>) and 0.54 mm d<sup>-1</sup> (0.0057 m<sup>3</sup> s<sup>-1</sup>) in 2018 and 2019.

During the recession period, Hathataga discharge appeared insensitive to storms. Only one rainfall event, where 36 mm of precipitation fell between August 31 and September 2, 2019 caused an appreciable rise in discharge (Figure 2.5c). This may suggest a threshold response to recharge, however, the exact threshold is difficult to determine – other large jumps in discharge after precipitation events (e.g., June 21, 2019, where 68 mm of precipitation fell in a single day) were likely also influenced by snowmelt.

Discharge from Confluence and Galatea was also driven by snowmelt in the spring and summer (Figure 2.6). Rise in discharge in the larger watersheds began earlier than Hathataga, likely from contribution of snowmelt in lower elevation regions. In 2018 and 2019, discharge from Confluence and Galatea showed large fluctuations in response to snowmelt as well as clear diurnal fluctuations. Unlike Hathataga, their hydrographs showed sharp responses to several storms between July and October. Over much of the observation period, specific discharge from

both watersheds was lower than Hathataga. By August 16, 2018, specific discharge in both watersheds dropped to approximately  $0.5 \text{ mm d}^{-1}$ . In terms of volumetric discharge, flow at Confluence approximately equaled flow at the Hathataga outlet during this time. This indicates that regions below Hathataga (representing over 80% of the watershed area) contributed negligible flows at the time. This observation has important implications regarding how baseflow is stored and maintained in alpine regions and is further discussed in section 2.5.3.

#### 2.4.2 Groundwater and lake levels

Water levels in all piezometer are plotted in Figure 2.7. Overall, water levels show considerable spatial and temporal variability indicating the dynamic and distributed nature of watershed storage. At PSP6, by the outlet spring, water levels briefly passed the minimum and maximum observable range during low flow and snowmelt respectively, but otherwise remained close to 2082 masl. During peak snowmelt, water was observed discharging  $\sim 1 \text{ m}$  above the piezometer ground surface at 2082.4 masl. PSP6 water levels show diurnal fluctuations during May, following snowmelt in the area. Longer-scale fluctuations during Jun-Jul correlate with rise and fall in outlet discharge. Water levels receded slowly after snow cover depletion in mid-July and PSP6 continued to register water until Feb-Mar, when levels fell below its screen. PSP6 did not noticeably respond to storm events.

In the moraine, water levels of the Lake Piezometer (PLAK) correlated with water levels and discharge at the outlet spring (Figures 2.5a, 2.5c, 2.6a, 2.6b). During snowmelt in May 2018 and 2019, level in the piezometer quickly rose more than 4 m while the Lake filled to a depth of  $\sim 2 \text{ m}$ . Both lake level and piezometric head fluctuated with snowmelt events in spring and early summer, similar to Hathataga outflow and levels at PSP6, but showed no diurnal fluctuations. Before drying up, the lake water levels mirrored water levels in PLAK (Figure 2.7), indicating that the lake is effectively a surface expression of the moraine water table. However, lake levels were consistently 2-3 cm higher, indicating a downward hydraulic gradient at the piezometer. The lake dried up at approximately August 10 and August 23 in 2018 and 2019 respectively. Water levels in PLAK continued to recede slowly until October before falling below the screen (2086.4 m). During the observation period, the lake water level lake did not respond obviously to storm events.

Meadow piezometers P05 and P02d were installed at relatively shallow depths (Table 2.1) and only registered water in the spring and early summer, when the watershed was more

saturated. Both of these piezometers responded quickly to recharge events, showing diurnal fluctuations with snowmelt and sharp increases with storm events (Figure 2.7). Both also receded rapidly, and quickly ran dry in the absence of recharge. After the snowmelt period, water levels in these piezometers would sometimes spike back up from occasional storm events (e.g., September 23, 2018 and July 20, 2019).

In contrast, meadow piezometer P01d was installed deeper but laterally close to P05 and P02d. Head in P01d rose much slower over the snowmelt period than the shallow piezometers and PLAK. Unlike outlet discharge and other piezometers, P01d exhibited no strong fluctuations in response to various snowmelt events, and instead rose and fell smoothly with few stationary points. During recession, P01d level dropped slowly and went dry in November. P01d was largely insensitive to storm events. In July, of 2018, P01d was pumped dry during groundwater sampling and did not recover for several weeks. This behavior suggests that water level in P01d may not be representative of head in the formation and is instead a dampened and lagged reflection.

### 2.4.3 Chemistry & tracer tests

#### Major ions and stable isotopes

A local meteoric water line (LMWL) for the site was fitted to stable isotope abundances measured in precipitation samples. (Figure 2.8a):

$$\delta D = 7.51 \times \delta^{18}O - 1.33 \quad [2.6]$$

Where  $\delta D$  is abundance of Deuterium;  $\delta^{18}O$  is abundance of oxygen-18.

Oxygen-18 abundance in snow samples ranged between -23.7 ‰ and -22.4 ‰. Average  $^{18}O$  abundance in samples collected from rainfall collectors ranged between -22.6 ‰ and -8.0 ‰ over the spring and summer.

Stable isotope abundance and major ion chemistry of groundwater and surface water samples from May 30 to August 23, 2018 are plotted in Figure 2.8 and Figure 2.9, respectively. SP1, SP2, SP3, SP4 and the waterfall are associated with talus cones on the western side of the watershed (Figure 2.1) and are characterized by relatively high total dissolved solids (TDS) and sulfate (Figure 2.9). Stable isotopes of these samples plot on the LMWL, with oxygen-18 abundances between -20.8 ‰ and -18.4 ‰ over the spring and summer. In contrast, flow from SP5 was lower in TDS and proportionally higher in calcium and bicarbonate. Stable isotopes

from SP5 showed greater temporal variability and were generally lighter, with  $\delta^{18}\text{O}$  changing from -21.4 ‰ on July 17, 2018 to -19.5 ‰ August 23, 2018 (Figure 2.8b).

Water sampled from piezometers, the lake and Hathataga outlet appear to have similar TDS and major ion composition. Stable isotope composition of the Hathataga outlet and Lake plot within a close range (Figure 2.8c and 2.8d), with oxygen-18 generally ranging between -21 ‰ and -20 ‰. Groundwater samples from piezometers P01d, P02d and Lake GWDs show a similar spread in stable isotopes to lake and outlet samples.

Galatea and Confluence outflow over spring and summer 2018 were similar to discharge from the Hathataga outlet in both major ion concentrations and stable isotopes (Figures 2.8d and 2.10). Major ions samples of Hathataga outflow did not show great temporal variability (Figure 2.10).

#### Temperature & electrical conductivity (EC)

EC and temperature of springs SP1, SP2, SP3 and SP4 on the western talus ranged from  $\sim 300 \mu\text{S cm}^{-1}$  to  $\sim 1000 \mu\text{S cm}^{-1}$  and  $\sim 3 \text{ }^\circ\text{C}$  to  $\sim 8 \text{ }^\circ\text{C}$  during the summer of 2018 (Table 2.2). In comparison, SP5 which is further east, had EC  $\sim 160 \mu\text{S cm}^{-1}$  and a temperature range of  $0.7 \text{ }^\circ\text{C}$  to  $1.6 \text{ }^\circ\text{C}$ . Hathataga Lake maintained a fairly stable EC of  $\sim 300 \mu\text{S cm}^{-1}$ , however, temperature of the lake changed considerably, heating from  $2 \text{ }^\circ\text{C}$  during snowmelt in early June to over  $11 \text{ }^\circ\text{C}$  around August (Figure 2.11a).

At Hathataga Outlet, water temperature and EC were stable at around  $1.5 \text{ }^\circ\text{C}$  and  $350 \mu\text{S cm}^{-1}$  during the 2018 winter low flow period (Figure 2.12). These values are reflective of the groundwater which sustains flows at that time. With the start of snowmelt, temperature and EC dropped with the influx of snowmelt to the outlet. EC fluctuations in May and June are likely in response to snowmelt. EC remained depressed during snowmelt, but water temperature began to rise up to a maximum of around  $6 \text{ }^\circ\text{C}$ . The three branches of the outlet spring complex (Figure 2.1b) had similar EC values, but decreasing temperatures moving east to west. For example, by late June 2019, temperatures were  $\sim 4 \text{ }^\circ\text{C}$  at the eastern stream branch,  $\sim 3 \text{ }^\circ\text{C}$  at the center branch and  $\sim 1.5 \text{ }^\circ\text{C}$  at the western branch. By late summer, temperatures began to drop and EC rose, approaching its previous winter maximum. Both temperature and EC of outlet discharge were responsive to storm events, with both dropping initially after precipitation, then recovering over the next several days.

In comparison to the Hathataga outlet, manual measurements at Confluence and Galatea follow similar patterns in temperature and EC (Figure 2.13). The larger watersheds, however, saw higher maximum water temperatures in the summer and lower minimum in the winter due to greater surface exposure of their outlet streams.

#### Tracer tests

During the moraine tracer test on September 26, 2018, the first arrival of tracer was detected at GS1 at 11:30 before peaking at 562 ppb at 14:45, 232 minutes after the start of tracer injection (Figure 2.14). Using peak arrival times, average flow velocity over the 150 m straight-line travel distance (Figure 2.1) was  $1.1 \times 10^{-2} \text{ m s}^{-1}$ . Concentrations dropped rapidly after the peak, but remained at lower levels for the next three months. ~78% of tracer was recovered in the month following injection, and ~50% of that mass passed through the outlet within a day after injection. In the following weeks, tracer was detected in water samples from GS1 which is a mixture of waters from all three branches of the Hathataga Outlet (Figure 2.1b). Separate and distinct breakthrough peaks (Figure 2.14a) may indicate the arrival of tracer from different flow paths within the moraine.

With the talus test (on July 10, 2019), tracer was first detected emerging from SP4 at 15:00 and reached a peak of 166 ppb at 15:30, 110 minutes after the start of tracer injection (Figure 2.14a). Given a straight-line travel distance of 125.6 m (Figure 2.1), average flow velocity was calculated to be  $1.9 \times 10^{-2} \text{ m s}^{-1}$ . Similar to the first test, concentrations rose and fell rapidly during the first few hours. Due to an animal attack which severed the connection between the fluorometer and datalogger, fluorometer data at SP4 only span two days after injection. During this period only ~2% of tracer was recovered at SP4. Tracer concentrations were low by the end of record, but it is possible subsequent breakthroughs were missed. On August 22, 2019, 43 days after injection at the talus, another breakthrough in tracer was detected at the Hathataga outlet (Figure 2.14b). This second breakthrough saw gradual changes in concentration and only reached highs of ~4 ppb. However, low tracer levels were maintained at the outlet for months following. Assuming travel times between 43 and 105 days and a distance of 400 m between the injection point (TTS in Figure 2.1c) and the Hathataga outlet, flow velocity was estimated between  $1.0 \times 10^{-4} \text{ m s}^{-1}$  and  $4.4 \times 10^{-5} \text{ m s}^{-1}$ . This second breakthrough is likely the remaining tracer from the talus test passing through the watershed, assuming that any tracer from the moraine test had likely been flushed from the watershed by that point. From the

start of breakthrough to the end of September, about 130 g of tracer passed through the outlet, representing ~26% of initial injected mass. The long, spread out shape of the breakthrough curve indicate mixing and dispersion of the initial mass.

#### *2.4.4 Hydraulic conductivity measurements*

##### Slug tests

Slug tests were performed in PSP6, the PLAK, P01d and P05 in late summer, 2018. The meadow piezometers: P01d and P05 are close to one another but measurements of hydraulic conductivities in these piezometers were  $8.3 \times 10^{-10} \text{ m s}^{-1}$  and  $9.9 \times 10^{-5} - 1.2 \times 10^{-4} \text{ m s}^{-1}$ , respectively (Figure 2.15). Given that P01d is screened almost 2 m deeper than P05, meadow conductivity seems to decrease sharply with depth. In the moraine, conductivities measured in PLAK and PSP6 are higher than the meadow, ranging between  $2.7 \times 10^{-4}$  and  $1.0 \times 10^{-3} \text{ m s}^{-1}$  (Figure 2.15).

##### Permeameter measurement

Field saturated hydraulic conductivities measured by constant-head permeameter vary from  $9.7 \times 10^{-7}$  to  $1.7 \times 10^{-4} \text{ m s}^{-1}$  across the four forest measurement points (Figure 2.15). At all points, sediments were layered, generally consisting of: dark, organic soil overtop a thinner fine, orange ash all underlain by a coarser, poorly sorted till. Upper soil and ash layers were generally several tens of centimeters thick. The till layer, likely extends meters down to bedrock based on observations of open excavations nearby. Hand augured holes were only able to penetrate 10 – 30 cm into the till, but conductivity measurements of the upper till were generally higher than the overlying soil and ash.

##### Tracer tests

Using time to peak concentrations, average flow velocity in the moraine was  $1.1 \times 10^{-2} \text{ m s}^{-1}$ . Head difference between PLAK and PSP6 was 5.90 m over the course of the test and both piezometers are separated by 150 m. Assuming moraine porosity between 0.2 and 0.4, calculated conductivity ranges between  $5.5 \times 10^{-2} - 1.1 \times 10^{-1} \text{ m s}^{-1}$  (Figure 2.15). These values are representative of a rapid, preferential flow path within the moraine because tracer was injected directly into a drain hole within the lake bed.

Average flow velocity in the talus was  $1.9 \times 10^{-2} \text{ m s}^{-1}$ . Travel distance was 125.6 m and elevation difference between the injection and recovery sites was 41 m. Again, assuming

porosity between 0.2 and 0.4, talus conductivity was calculated between  $1.2 \times 10^{-2} - 2.4 \times 10^{-2} \text{ m s}^{-1}$  using these values (Figure 2.15). Because of rapid flow rates and low tracer recovery, these values likely represent only a shallow, conductive flow path within the upper talus.

The second breakthrough of the talus tracer test was used to make a rough estimate of hydraulic conductivity between the injection site and the Hathamaga Outlet. With average flow velocity between  $1.0 \times 10^{-4} \text{ m s}^{-1}$  and  $4.4 \times 10^{-5} \text{ m s}^{-1}$ , and the same assumptions as above, conductivity was estimated ranging between  $3.9 \times 10^{-4} \text{ m s}^{-1}$  and  $8.3 \times 10^{-5} \text{ m s}^{-1}$ . These values reflect the harmonic mean conductivity of the flow path which must travel through different deposits to reach the outlet.

#### Hydraulic conductance

Hydraulic conductance ( $C$ ) (Eq. 2.4) was calculated from the May-October time series of outlet discharge and water levels in PSP6 and PLAK (Figure 2.16a). In both 2018 and 2019, time series of  $C$  reflect discharge, peaking early during the snowmelt period and dropping as flows recede (Figure 2.16a). When plotted against level in the PLAK,  $C$  follows a curve that is exponential in appearance (Figure 2.16b). Over the spring-fall period,  $C$  undergoes more than an order of magnitude change, indicating strong variation of moraine conductivity with depth (Equation 2.5) as water levels rise and fall. Furthermore, as  $C$  is a vertically averaged value, depth-specific hydraulic conductivities could vary several orders of magnitude across the moraine.

It should be reiterated that calculation of  $C$  suffers from uncertainty and is only meant as a rough gauge of how conductivity within the moraine varies with depth. The assumption of constant flow-through area is violated as moraine saturated thickness changes over time – this will reduce the magnitude of  $C$  change in reality. Hydraulic gradient is assumed linear and computed from the head measurements in PSP6 and the Lake piezometer. During snowmelt head at the outlet spring is high and seepage from the spring occurs at elevations above PSP6, which will cause an overestimation of hydraulic gradient (and an underestimation of  $C$ ) between the moraine and the spring.

## 2.5 Discussion

### *2.5.1 Heterogeneity and flow paths in surficial deposits*

#### Talus

Talus deposits within the Hathataga watershed reflect their heterogeneity at the surface – changes in vegetation cover, grain size distribution and texture occur between and within different cones and slopes. Geochemistry of the talus springs suggests at least two distinct flow paths. One is high in sulphate and TDS, and feeds talus springs SP1, SP2, SP3, SP4 and the waterfall to the west. The second flow path feeds SP5 and is distinguished by low TDS and consistently low temperatures. This flow path is likely influenced by ground ice in the talus (Figure 2.1c).

The majority of flow through the talus, however, appears to follow a deeper flow path, which bypasses the springs. This deeper flow path is likely associated with low hydraulic conductivity layers of fines within the talus (Christensen et al., 2020). Only a small fraction of tracer injected above SP4 was initially recovered from the spring. The rest of tracer was most likely transported by this deeper, slower flow path to emerge at the Hathataga Outlet, weeks later.

#### Meadow

Within the meadow, flow in streams and water levels in shallow piezometers responded rapidly to recharge. Water level in P01d, a deeper piezometer, changed more gradually and appeared largely insensitive to short-term recharge events, however, this may not reflect actual behavior of the formation (see Section 2.4.2). Slug tests and pumping of P01d show a steep drop in hydraulic conductivity with depth. This large difference in shallow and deep hydraulic conductivity likely produces fast and slow flow paths within the meadow. For example, during recharge events, rapid lateral flow occurs at the surface or in permeable, shallow sediment, while a deeper flow path transmits a steadier flow of water which persists after shallow flow paths dry out. In this way, the meadow can be capable of responding quickly to recharge events while also maintaining of water levels/storage over longer periods of time.

#### Moraine

Time series of lake water levels and discharge also indicate vertical variability in the moraine. Analysis using calculated hydraulic conductance (Equations 2.4 and 2.5) show that saturated moraine hydraulic conductivity potentially decreases by an order of magnitude over the

observed range in water level (Figure 2.16). Conductivity of the moraine measured from the tracer test (when water level was ~1 m below the lake surface) was also higher than conductivity from the PLAK slug test (which was screened ~ 2 m below the lake surface); although this comparison is far from concrete, it indicates the variability in flow paths through the moraine and may also point to vertical variability of the moraine. The nature of this variability is unknown, but it may result from layering within the moraine, compaction with depth, influence from fractured bedrock below the moraine, and/or variations in bedrock topography.

Depth variability has a strong influence on the storage and transmission of water through the moraine over time. In spring and summer, when the lake is present and water levels are high, large quantities of groundwater can flow through more conductive, shallow moraine sediment, resulting in high discharge rates at the Hathataga Outlet. However, as water levels fall, moraine transmissivity decreases greatly (as a function of both decreasing flow-through area and hydraulic conductivity), allowing for steady low flows to be sustained.

#### *2.5.2 The “gatekeeper”/Hydrologic response at the watershed scale*

Given its location at the northern tip of the watershed (Figure 2.1), (nearly) all flow must pass through the moraine before reaching Hathataga Outlet. This association is supported by strong correlation between moraine water levels and outlet discharge, as well as tracer test results. Furthermore, during spring and summer, there is a large increase in temperature of outlet discharge (Figure 2.12a). This suggests that a large component of discharge is warm water from Hathataga Lake (Figure 2.11), which exposed at the ground surface and heated by solar radiation (Roesky, 2020) before being transmitted through the moraine. Therefore, as the source of (nearly) all watershed flow, the moraine acts as a “gatekeeper” which regulates storage-discharge behavior (Christensen et al., 2020).

Field data indicate great storage capacity with long residence time within the watershed. Oxygen-18 and deuterium abundances in outlet discharge plot tightly on the LMWL (Figure 2.8d);  $^{18}\text{O}$  abundance does not exceed -20 ‰, indicating that snowmelt remains a large component of watershed discharge and storage (Figure 2.8a). This long-term storage capacity is provided by fine, low conductivity layers of various surficial deposits, but is likely further bolstered from restriction of outflow by the moraine. Rather than quickly running off, much recharge during the snowmelt period is trapped by moraine ridges, raising groundwater levels

and Hathataga Lake. This additional storage persists and allows for higher baseflow rates to be maintained into the fall and winter.

Apart from boosting storage, the moraine also regulates other watershed-scale discharge behavior and response to forcing. Much heterogeneity in properties and hydrological processes exist within Hathataga, however, their “signal” may be masked out of outlet discharge by the moraine. In this way, integrated discharge response at the watershed scale may be more influenced by or reflective of characteristics of the moraine. For example, snowmelt and precipitation is transported rapidly in the upper meadow and talus, but must still be routed through the moraine, dampening their response. As a result, the outlet hydrograph (Figure 2.6) generally lacks the sharp peaks associated with rapid flows from these units.

From a practical standpoint, understanding and reproducing the discharge behavior of the Hathataga watershed may only necessitate the characterization of the “gatekeeper”. Heterogeneity and complexity of the rest of the watershed may be secondary to this objective. Previous work indicate that small-scale complexity attenuates at larger scales (e.g., Wood et al., 1988) and that watershed-scale fluxes do not contain enough information to capture detailed heterogeneity (e.g., Jakeman and Hornberger, 1993; Kirchner, 2006). In this particular case, effect of the moraine may warrant a simpler characterization of the watershed as a whole. This will be explored further in Chapter 3.

### *2.5.3 Relative baseflow contribution of Hathataga and surrounding regions*

Although similar in location, climate and water chemistry, the discharge hydrographs of Hathataga, Confluence and Galatea watersheds show dissimilar behavior. Specific discharge from Confluence and Galatea are consistently several times lower than Hathataga during and after the snowmelt period (Figure 2.6). During snowmelt, greater rates of specific discharge likely reflect greater winter snowfall due to higher overall elevations in Hathataga (Figure 2.4). Later on, just before winter freeze-up, total volumetric discharge at the Confluence outlet (GS5) approaches discharge rate from Hathataga (at GS1), indicating almost no contribution from areas between the two gauging stations. These areas, north of Hathataga, generally consisting of forested hillslopes, have a lower proportion of coarse surficial landforms and have limited water storage and retention ability. Galatea also contains mainly forest and sees lower rates of normalized baseflow, reflecting a similar lack of proportional storage capacity.

Hathataga, though small in area, appears to contribute a large quantity of baseflow to downstream areas. Studies elsewhere have shown that specific regions within the mountains are more important contributors to rivers and streams during low flow, usually due to special geological characteristics such as proportion of coarse sediment or younger, softer bedrock (e.g., Floriatic et al., 2018; Paznekas and Hayashi, 2016; Tague and Grant, 2004). Hathataga itself benefits from having a greater proportion of coarse, surficial deposits and the effects of the moraine “gatekeeper”. As discussed above, it is likely that regulation by this “gatekeeper” has a great influence on watershed discharge and retention of storage.

The importance of the gatekeeper may suggest that the arrangement of geological units within a watershed should be an important consideration for characterization of storage-discharge behavior. Specifically, moraine-regulated systems have been studied in other alpine regions (e.g., Roy and Hayashi, 2009; Thompson et al., 2012; Moore et al., 2011) and are prevalent across mountain regions worldwide (Clague and Evans, 2000). Moraine dams and other natural “gatekeepers” in these settings likely help to retain storage and maintain higher low flow rates, similar to what was observed at Hathataga. If so, these “Gatekeeper” watersheds may represent an important groundwater store and regulator of baseflow in the mountains and merit further study.

## 2.6 Figures and Tables

Table 2.1 Piezometer specifications. Datum: UTM NAD83, Zone 11N.

	Northing	Easting	Surface elevation (masl)	Depth from surface (m)	Internal diameter (m)	Screen length (m)
PSP6	5631405.6	625820.9	2082.4	0.97	0.019	0.11
PLAK	5631258.6	625803.4	2088.1	2.02	0.019	0.24
P05	5631086.2	625766.8	2095.2	4.00	0.019	0.24
P02d	5631077.0	625734.3	2096.1	2.80	0.019	0.24
P01d	5631087.5	625774.7	2095.1	5.73	0.019	0.24

Table 2.2 Temperature and electrical conductivity of talus springs measured in summer 2018.

	Date	Temperature (°C)	EC ( $\mu\text{S cm}^{-1}$ )
SP1	2018-07-27	4.8	640
SP2	2018-07-27	5.2	324
SP3	2018-07-27	5.9	986
	2018-08-23	7.8	1067
SP4	2018-07-27	2.9	842
	2018-08-23	7.3	851
SP5	2018-07-27	0.7	159
	2018-08-23	1.6	178

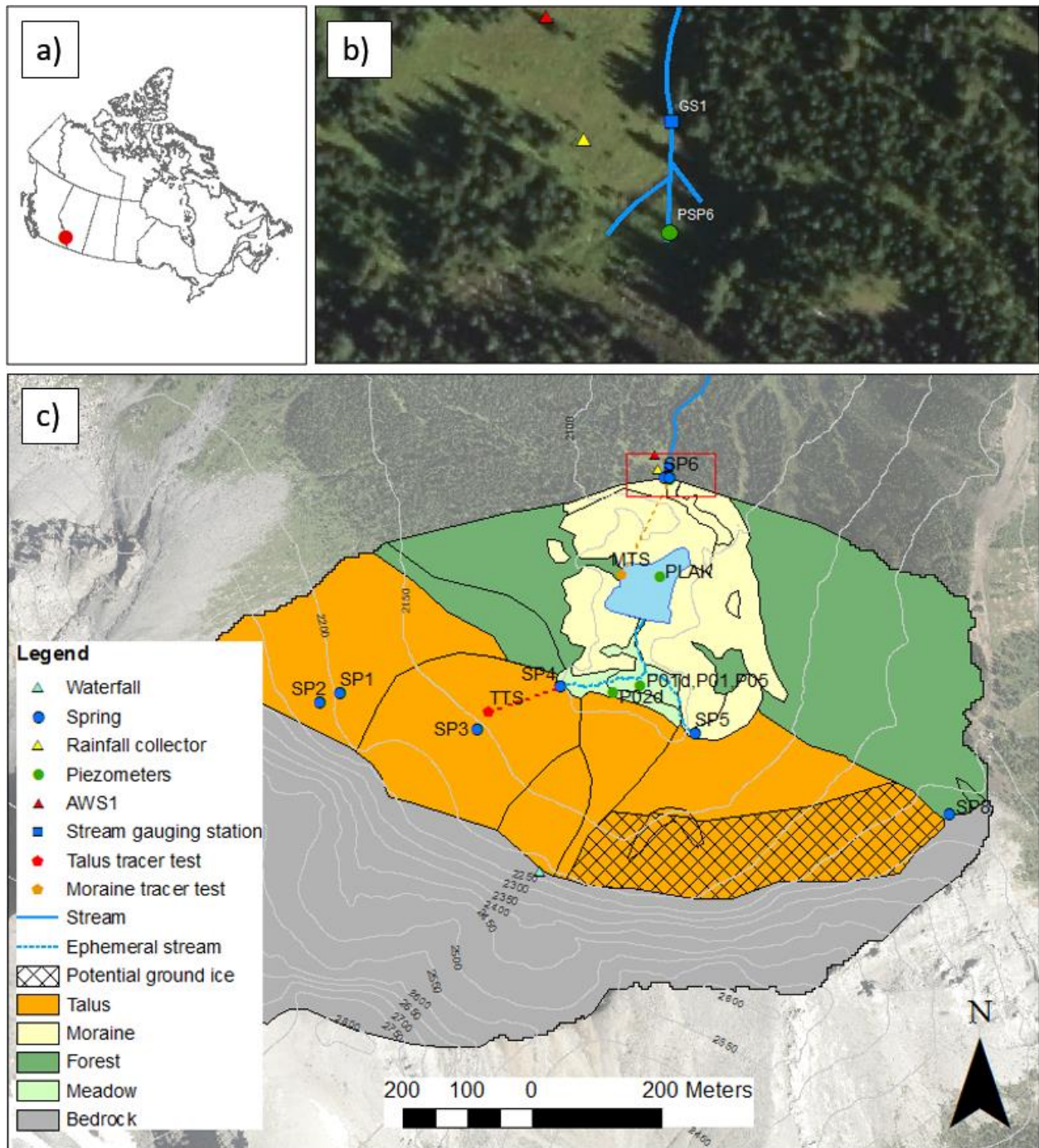


Figure 2.1 (a) Inset map showing location of study site. (b) Close-up view of the outlet spring complex showing three spring branches. (c) Map of Hathataga Watershed showing topography, delineated surface cover, hydrological features and instrumentation. Extent of (b) indicated by red box. Contour interval 50 m. Delineated area of potential ground ice is an approximate estimate.

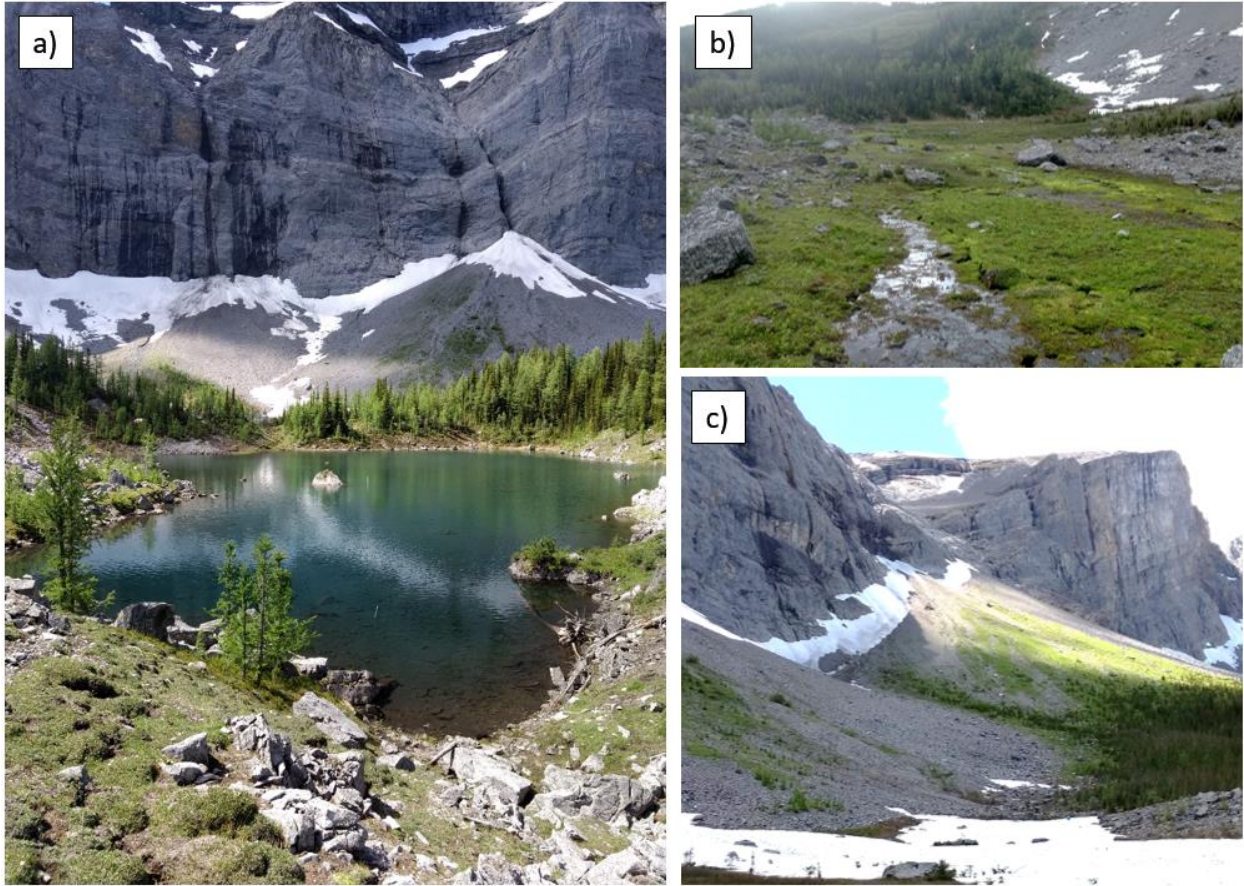


Figure 2.2 Surficial landforms within the watershed. (a) (July 13, 2018) Moraine ridges and Hathataga Lake with talus and headwall in background. Image captured from the northern end of the watershed looking south. (b) (July 10, 2019) Meadow and stream, looking east. (c) (July 10, 2019) Talus, looking west.

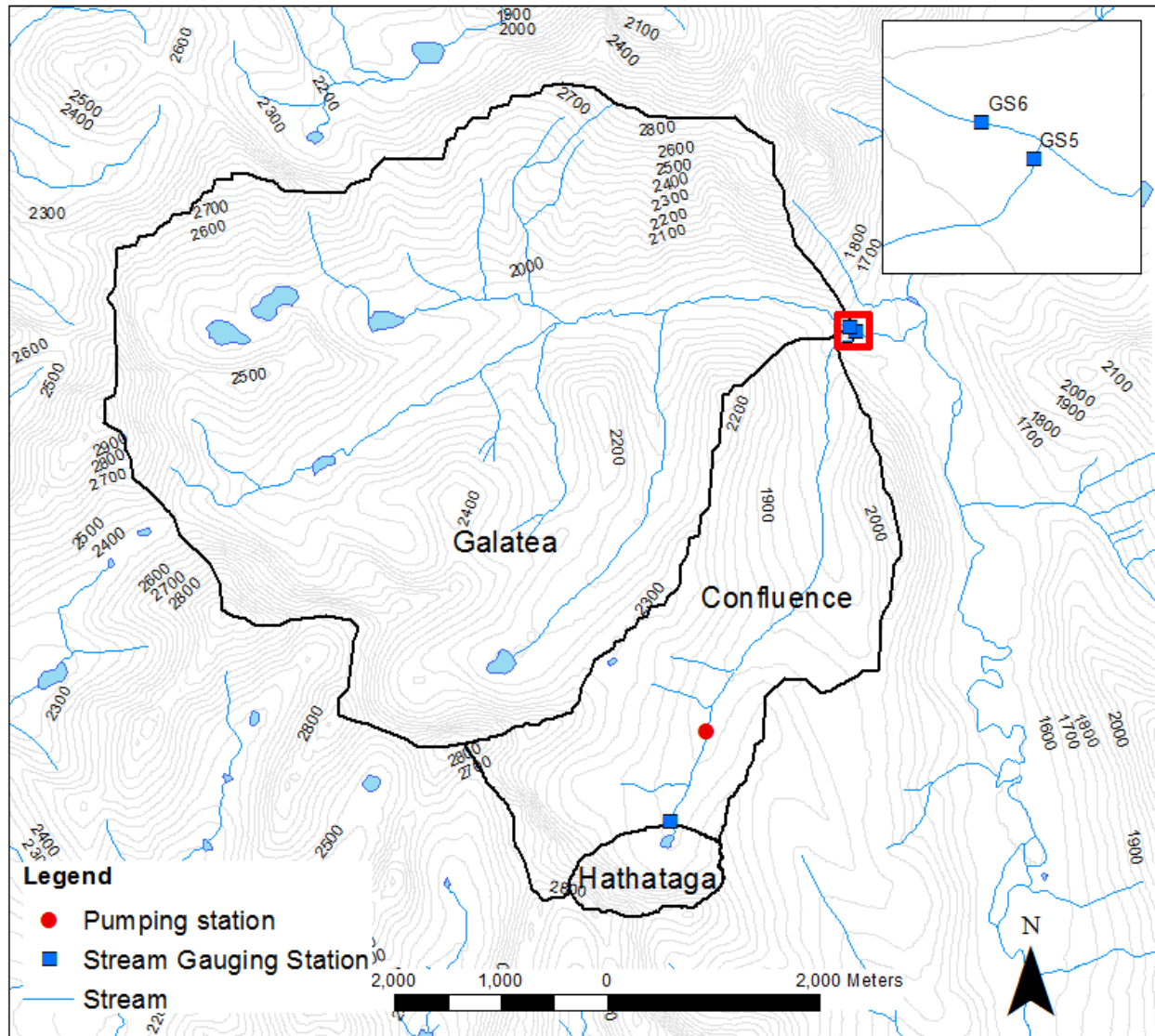


Figure 2.3 Map of the Hathataga, Confluence and Galatea watersheds. Red square indicates extent of the inset map. Contour interval 100 m.

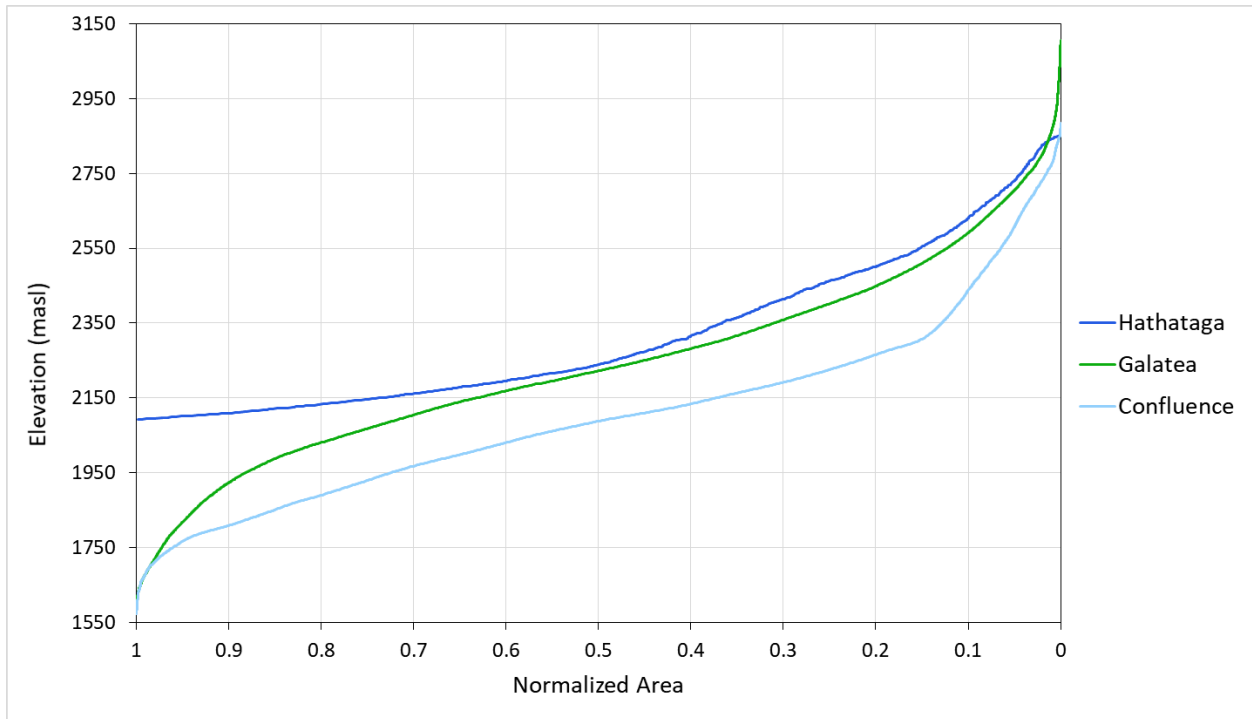


Figure 2.4 Hypsometric curves of Hathataga, Confluence and Galatea watersheds. Normalized area is presented as the fraction of watershed area at or above a given elevation.

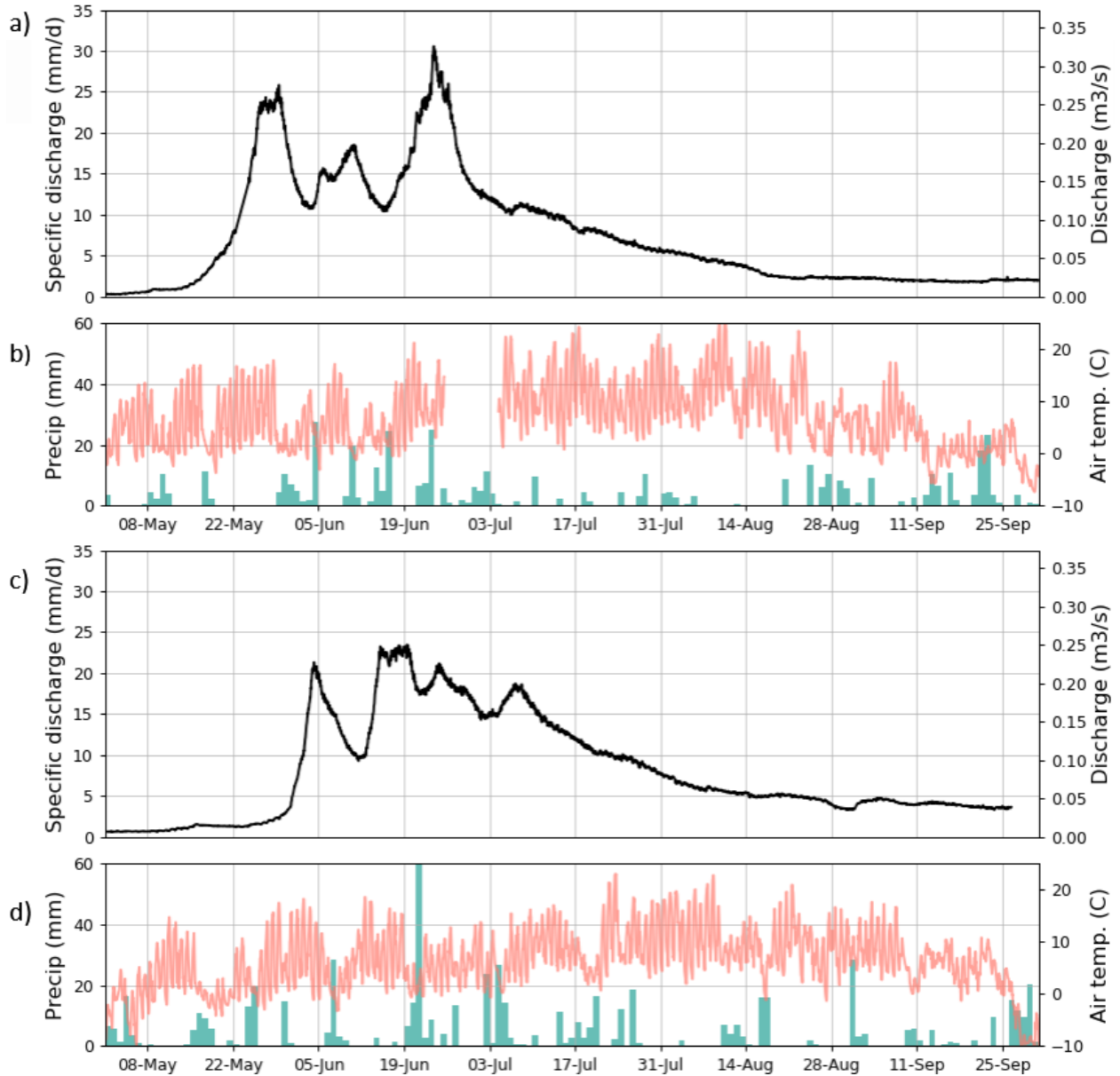


Figure 2.5 (a) 2018 time series of discharge. Discharge was normalized using a catchment area of 0.92 square kilometers; some uncertainty is associated with this normalization due to the western boundary. (b) 2018 time series of hourly average air temperature and daily precipitation. (c) 2019 time series of discharge. (d) 2019 time series of air temperature and precipitation.

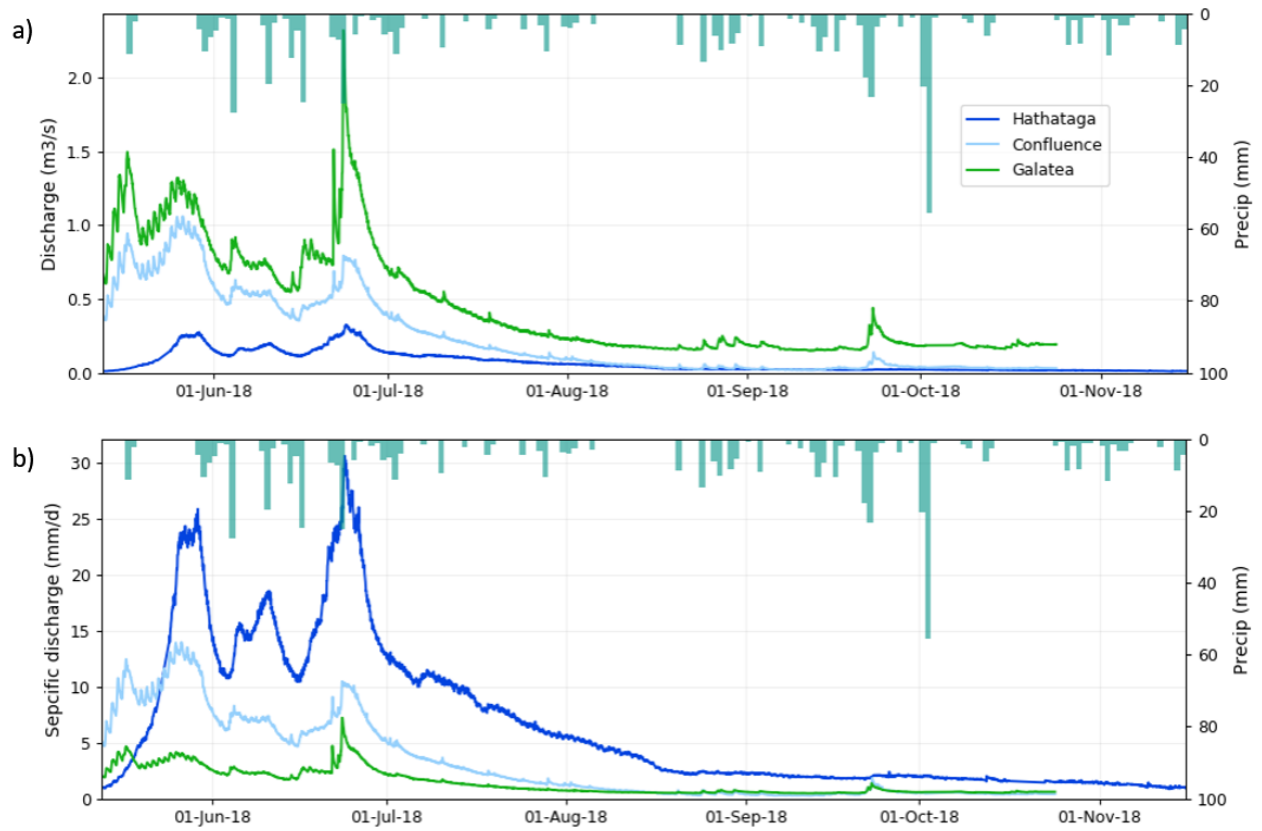


Figure 2.6 (a) Volumetric and (b) area normalized discharge from the Hathataga, Confluence and Galatea watersheds in 2018. Precipitation at AWS1 is also shown.

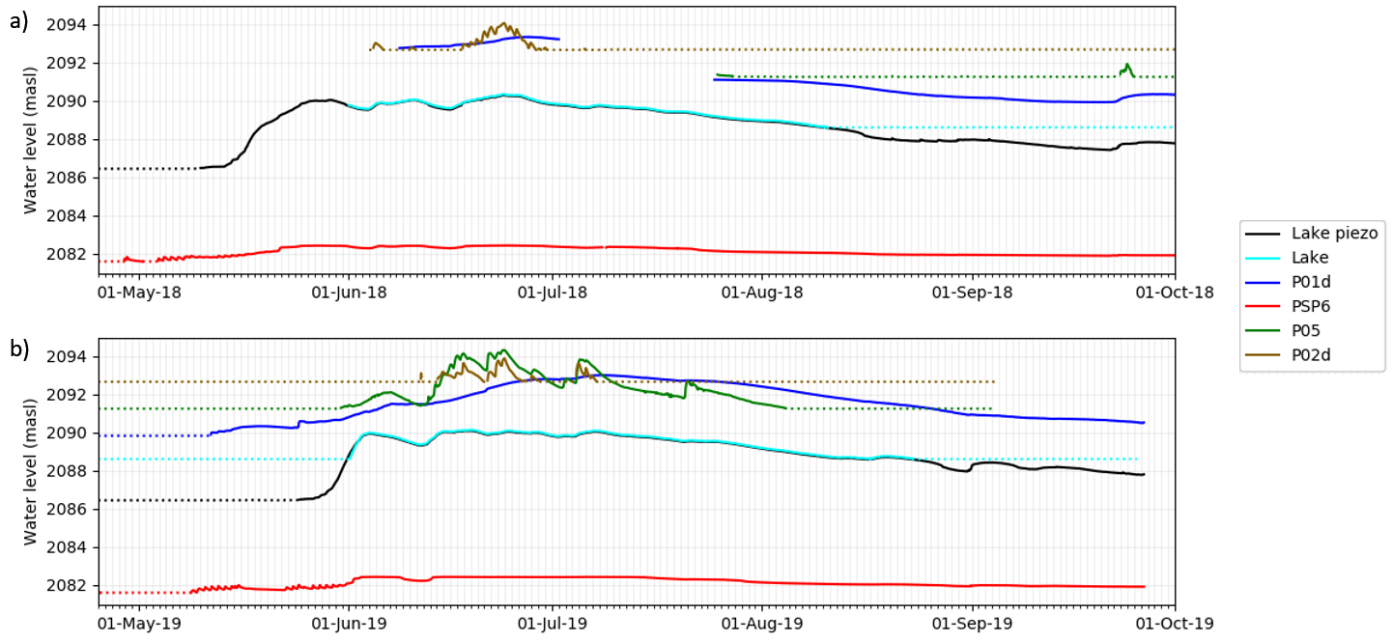


Figure 2.7 Piezometer and lake levels in (a) 2018 and (b) 2019. Gaps signify missing data. Dotted lines indicate periods when the piezometer was dry.

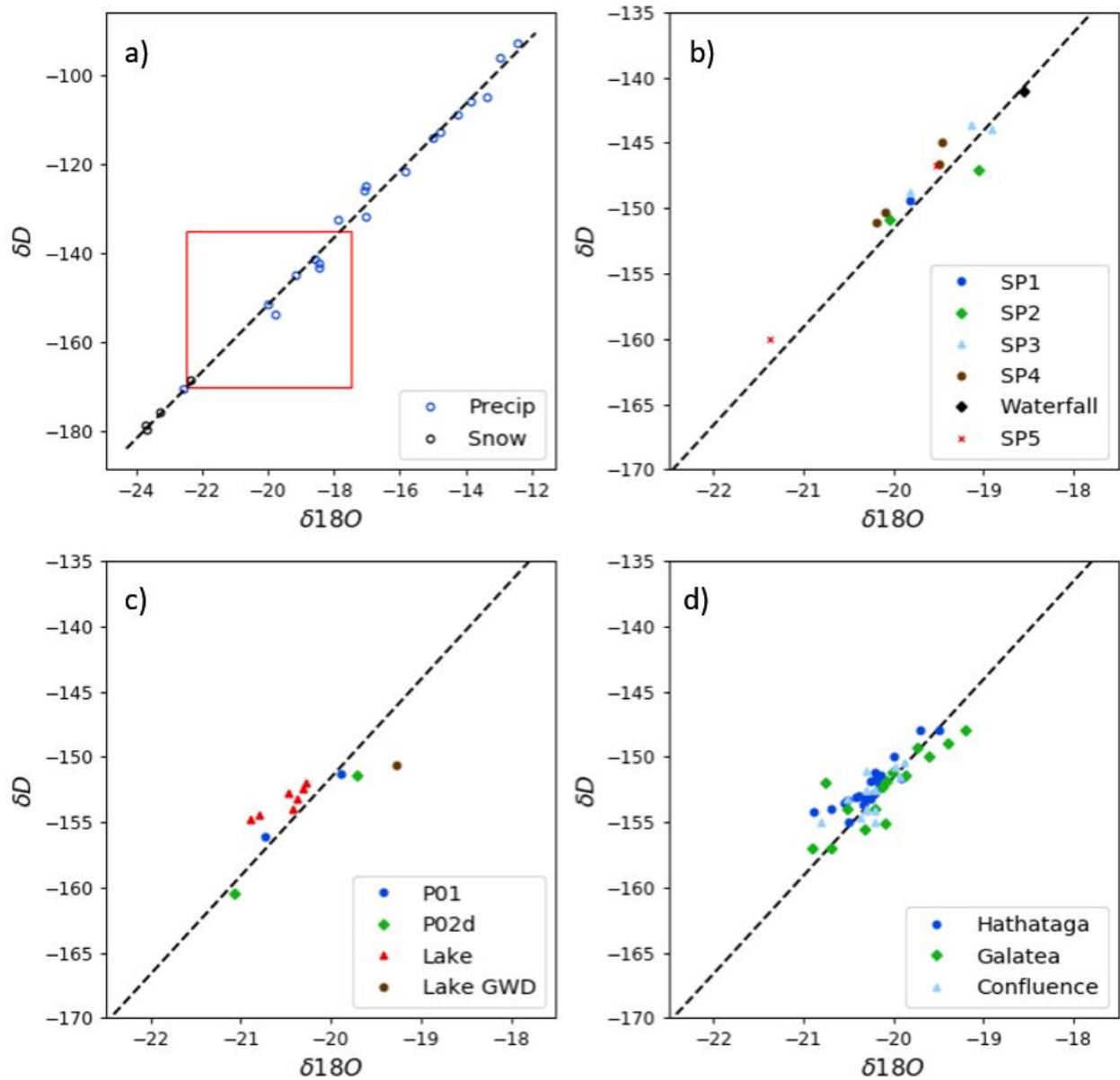


Figure 2.8 Plots of measured deuterium vs oxygen-18 abundance. Local meteoric water line shown as dashed line. (a) Precipitation and snow samples, red outline indicates extent of (b), (c) and (d). (b) Talus springs and waterfall. (c) Groundwater in the moraine and meadow and Hathataga Lake. (d) Discharge from the Hathataga, Galatea and Confluence watersheds.

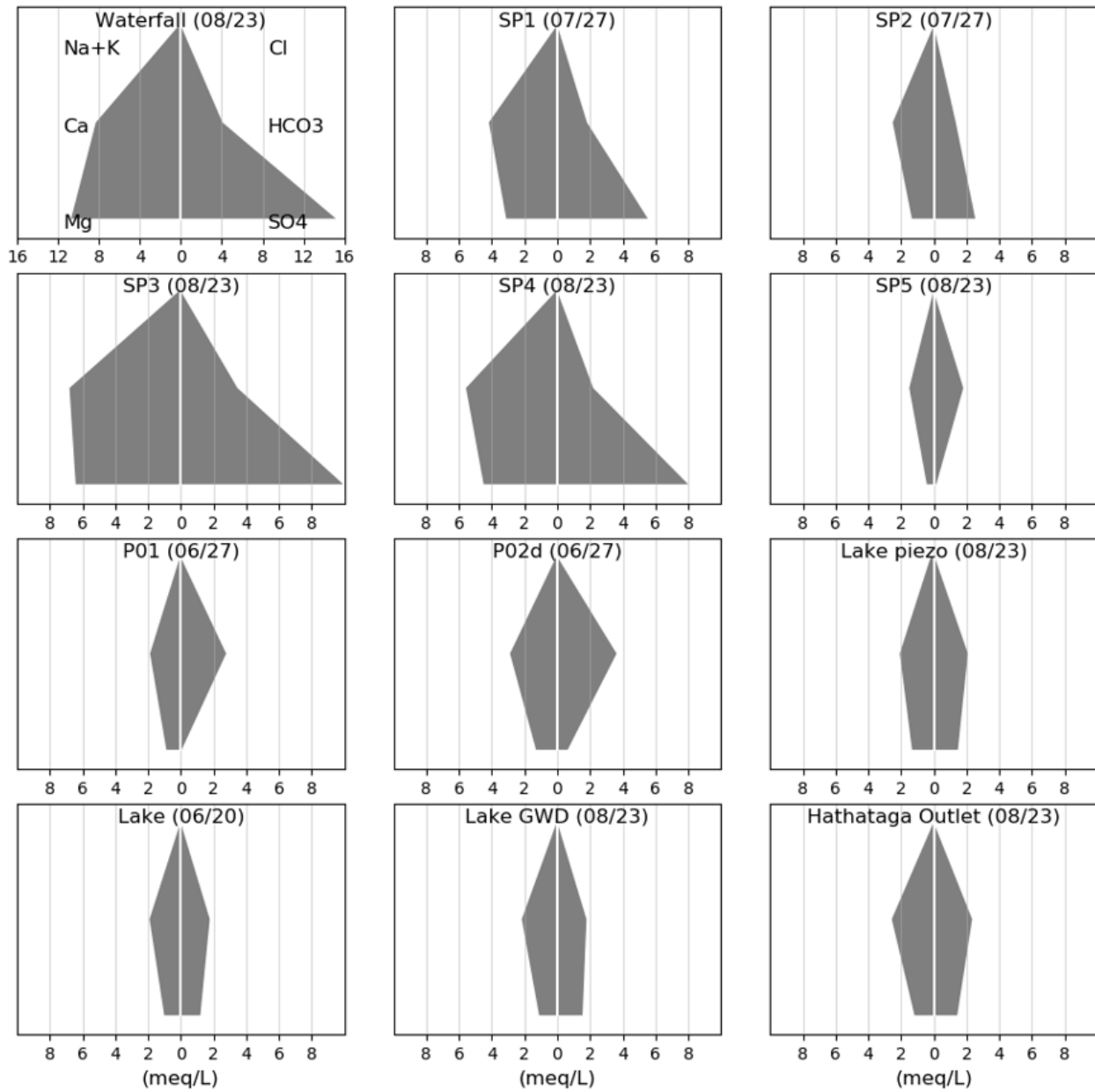


Figure 2.9 Major ions measured in samples collected from Hathataga over summer 2018 presented in Stiff diagrams.

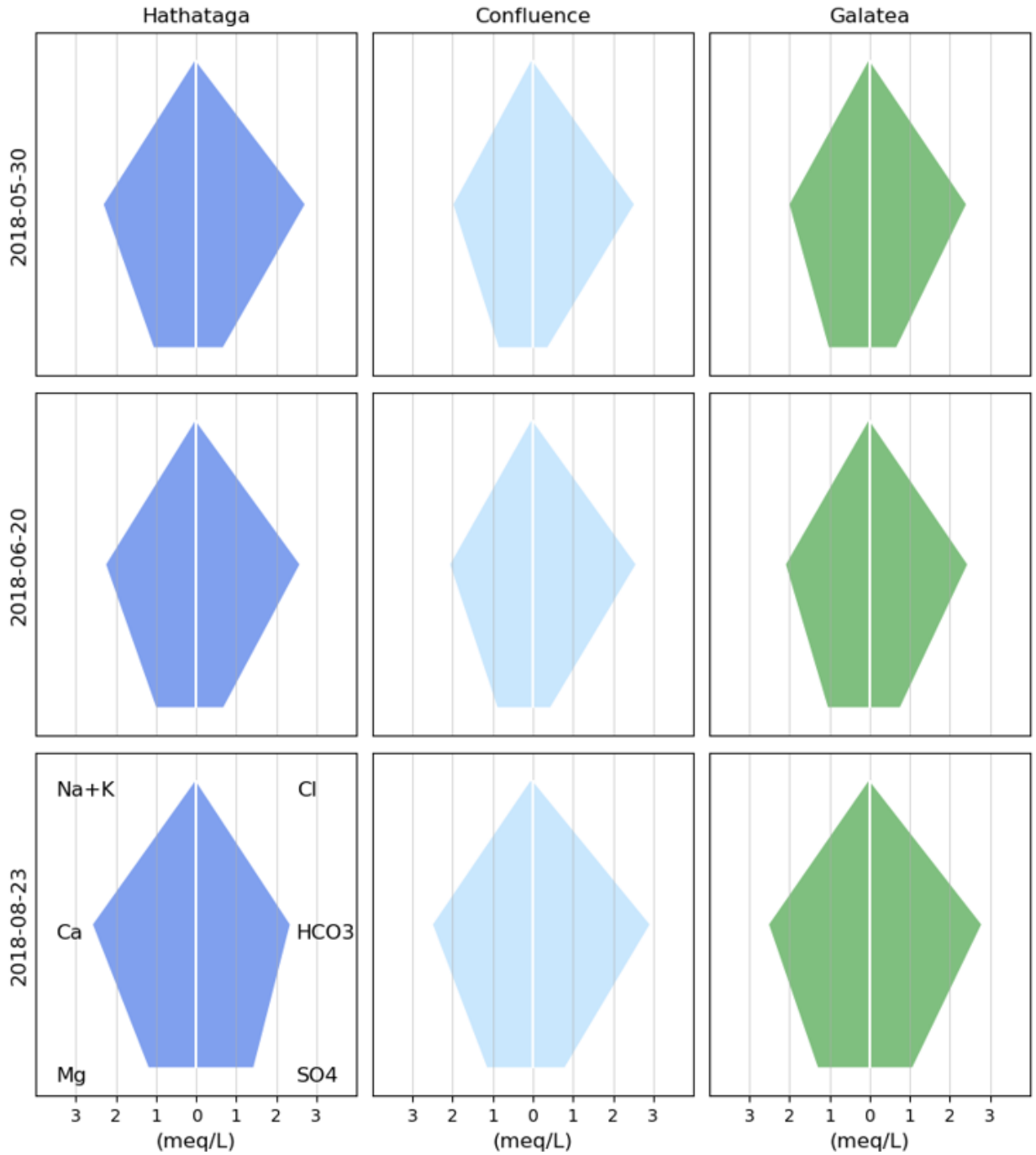


Figure 2.10 Major ions measured in samples of outflow from the Hathataga, Confluence and Galatea watersheds over summer 2018 presented in Stiff diagrams.

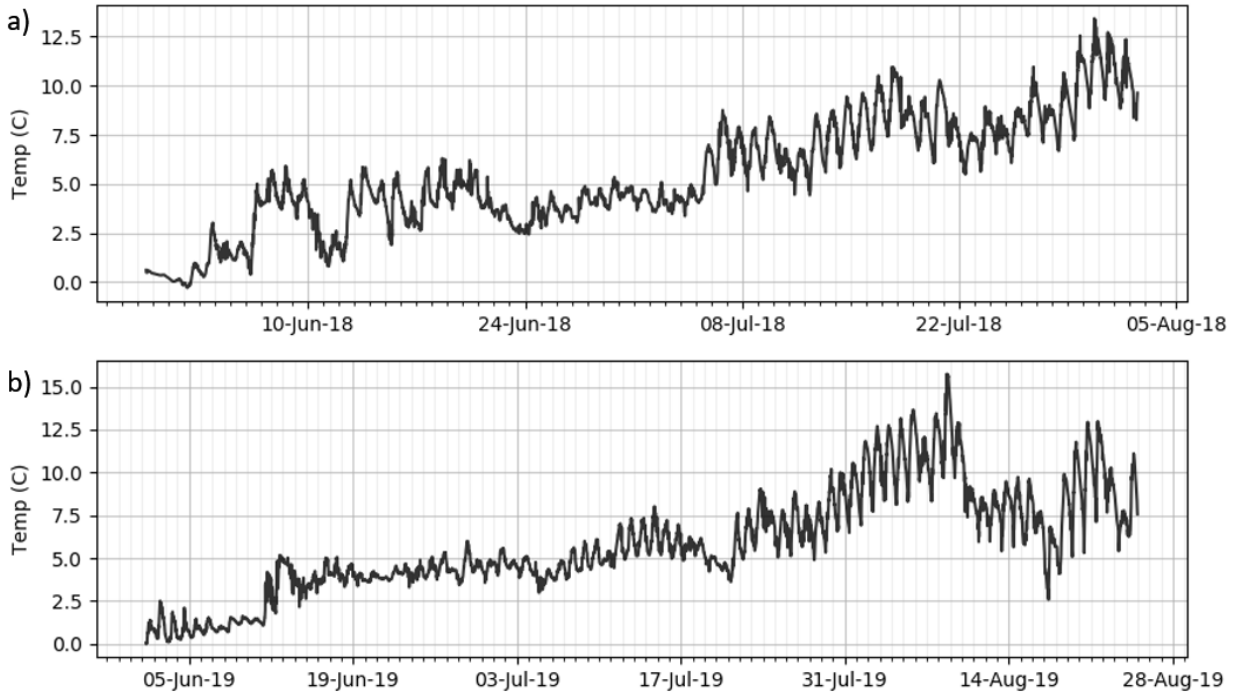


Figure 2.11 Measured lake temperature. (a) 2018. (b) 2019.

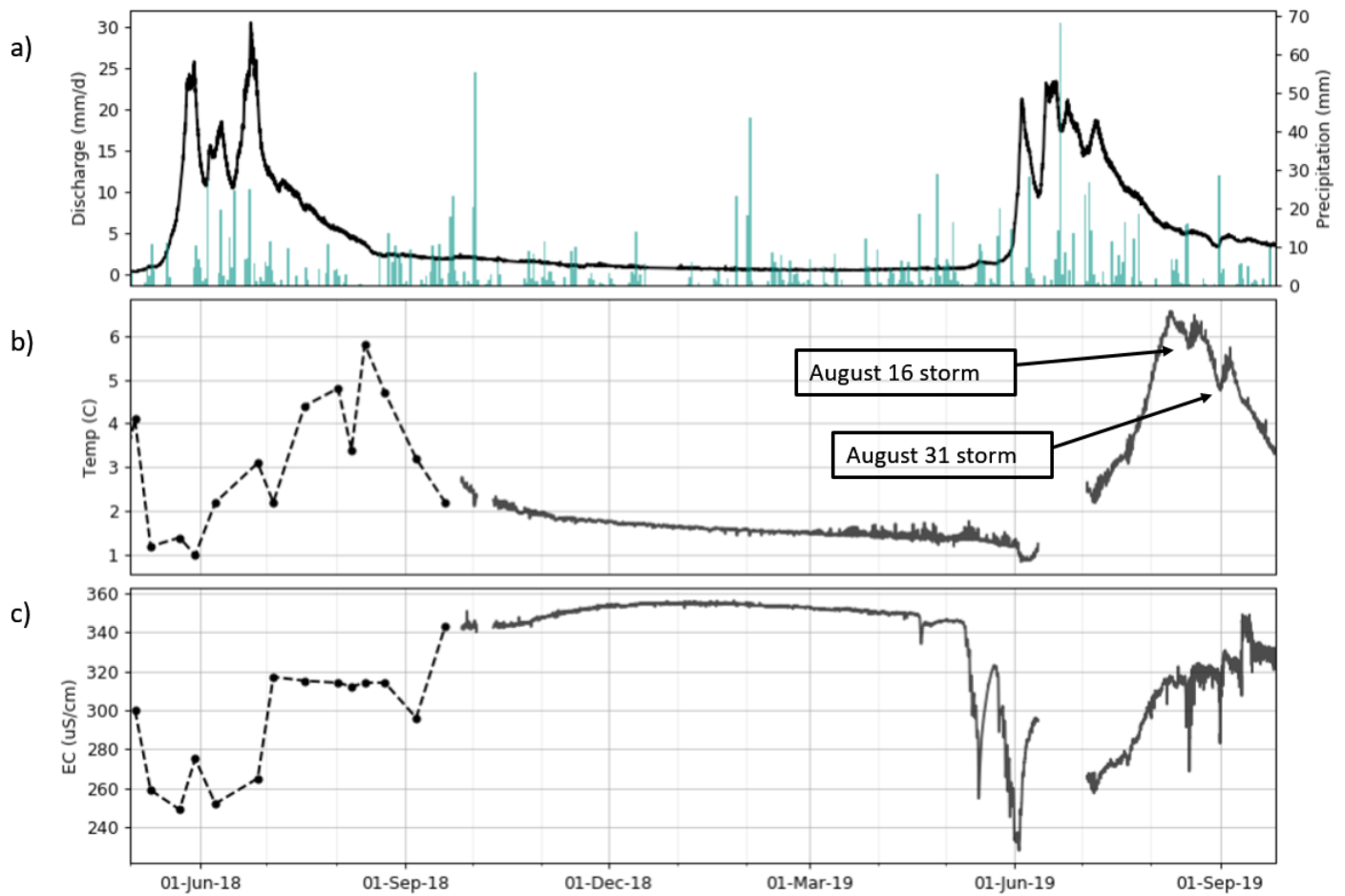


Figure 2.12 Time series of water temperature and EC of Hathataga Outlet discharge measured from a logger deployed at the gauging station (GS1). Manual measurements (shown as points) are used to supplement gaps in logger data.

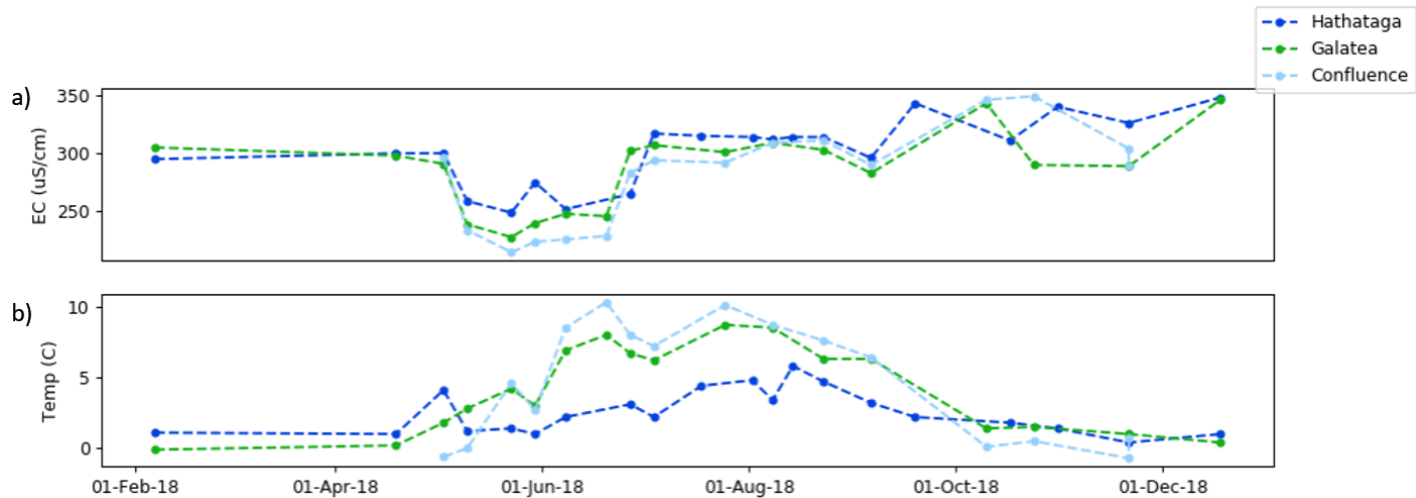


Figure 2.13 Time series of (a) EC and (b) water temperatures of discharge from the Hathataga, Galatea and Confluence watersheds in 2018.

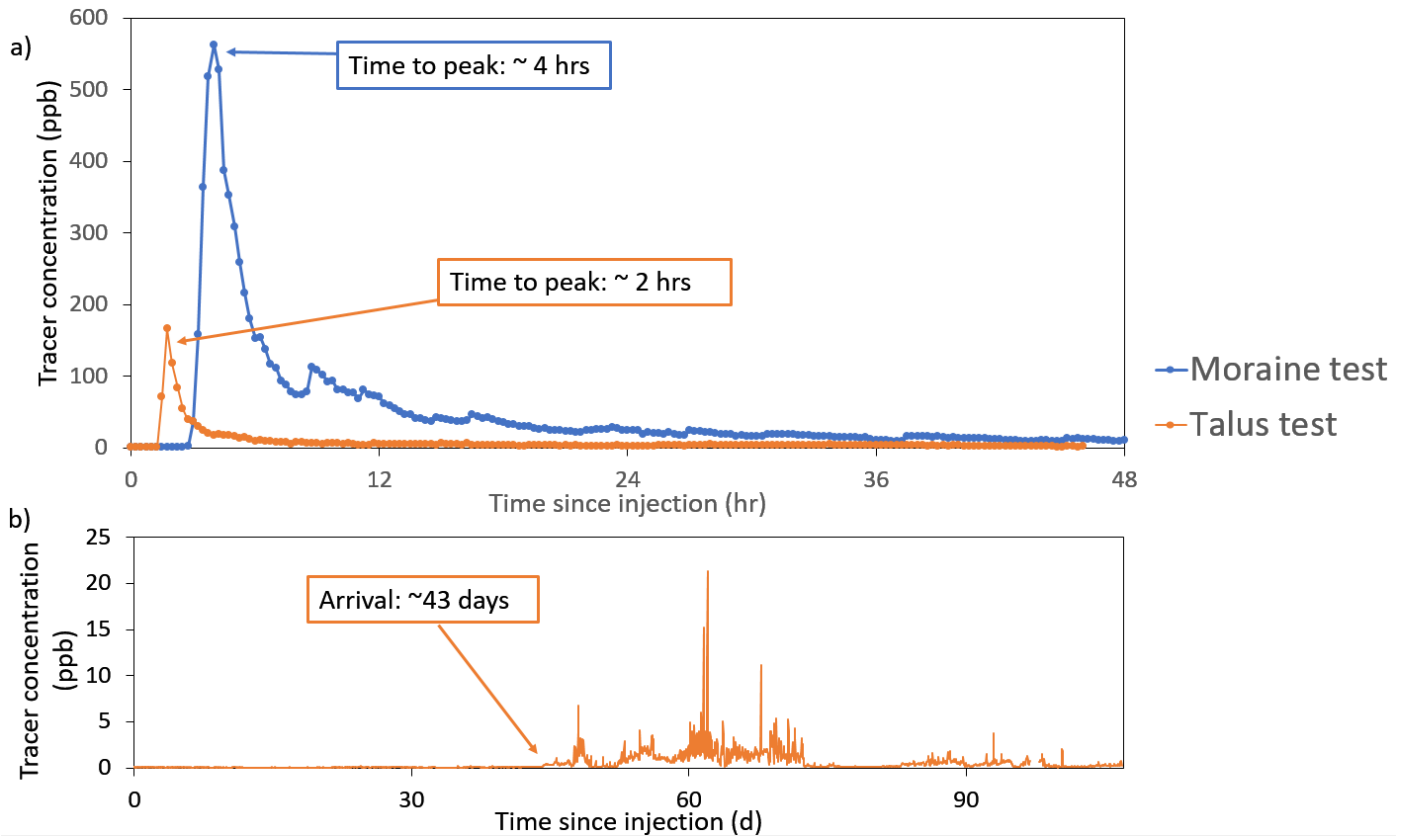


Figure 2.14 (a) Initial peak tracer breakthrough recorded for the moraine and talus tracer tests. (b) Delayed breakthrough at Hathataga Outlet after injection of the talus tracer test.

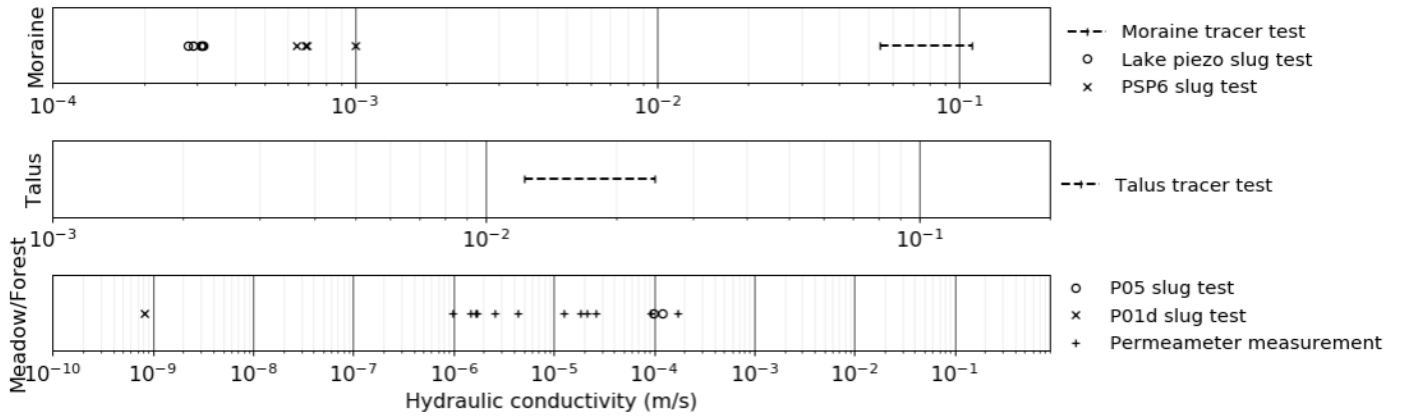


Figure 2.15 Hydraulic conductivities measured in different surficial landforms and using different methods.

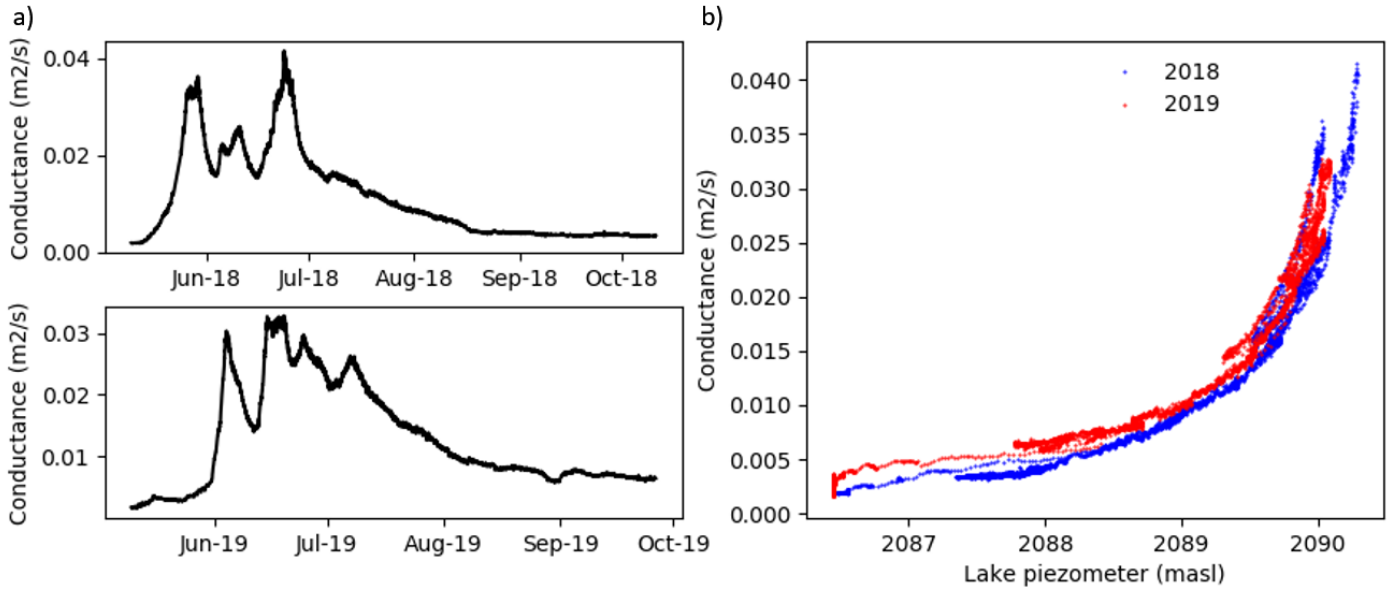


Figure 2.16 (a) Time series of calculated moraine hydraulic conductance over 2018 and 2019. (b) Plot of hydraulic conductance against water level in the Lake piezometer.

### 3. Simulation of the Hathataga groundwater system

#### 3.1 Introduction

Numerical groundwater models are a receptacle for collected data and conceptual understanding. Utilized appropriately, they can be powerful investigative tools for developing understanding of hydrogeological processes and characteristics of a study site. Previous studies have used models to infer hydrogeological properties; aquifer internal structure (e.g., Kurylyk and Hayashi, 2017; Pauritsch et al., 2017); and to study important processes (e.g., Wright et al., 2009; Clark et al., 2011; Wagner et al., 2016). Predictions from models can also be used to assess the impacts of changing climate and water needs. In the face of future water uncertainty, models will play a crucial role as predictive and decision-making tools (DeBeer et al., 2015).

Simulation of hydrogeological processes can be done following a number of different approaches. Models such as Hydrologiska Byråns Vattenbalansavdelning (HBV; Lindstrom, 1997); Precipitation-Runoff-Evapotranspiration HRU Model (PREVAH; Viviroli et al., 2009); and the Cold Regions Hydrological Model (CRHM; Pomeroy et al., 2007) follow a ‘lumped’ approach, where groundwater systems are treated like one or a few (non)linear reservoirs or “buckets”. These lumped models reproduce discharge behavior based on a storage-discharge relationship but do not represent small-scale processes or heterogeneity explicitly. Alternatively, hydrogeological systems can be represented using a distributed numerical model such as MODFLOW (McDonald and Harbaugh, 1988), HYDRUS (Simunek et al., 2006), HydroGeoSphere (Brunner and Simmons, 2012) or FEFLOW (Diersch, 2014). Rather than a bucket, the model domain consists of a one, two- or three-dimensional grid/mesh. This approach allows groundwater storage and discharge to be represented varying across both space and time, and allows for heterogeneity and geometries of the model domain to be explicitly defined. These models can also include complex processes like unsaturated flow or gravity drainage explicitly in the simulation.

Groundwater models, both lumped and distributed, can take varying levels of process and parameter complexity. For example, lumped models can range from a simple single linear reservoir to several non-linear reservoirs connected in series or parallel. Similarly, parameterization of a distributed model domain can be as simple as complete homogeneity (where each grid/mesh node is assigned the same hydraulic property) or can reflect more complex heterogeneity (e.g., parameterization via pilot points or zones of piecewise constancy). The debate for more versus less complexity in hydro(geo)logical models is ongoing and has been

discussed extensively in the literature (e.g., Hill, 2006; Clark et al., 2011; Hrachowitz et al., 2014; Jakeman and Hornberger, 1993; Watson et al., 2013).

In the context of alpine zones, lumped methods may be preferred because of simpler parameterization and ease of use. They are a “data-frugal” option for alpine regions which are usually data scarce. As the study of alpine hydrogeology turns toward characterization and prediction for greater scales, simpler models may be an adequate and efficient method of representing watershed or larger scale behavior. On the other hand, more distributed and highly parameterized models may be better able to simulate groundwater flow and storage in rugged alpine terrain, which depends on complex processes and fine scale spatial heterogeneity. Some studies have indicated that models with more complexity have potential to outperform simpler counterparts in prediction and validation (Clark et al., 2011; Hrachowitz et al., 2014), although over-parameterization is always a risk.

In this chapter, observations and conceptual understanding of the Hathataga watershed are compiled into several groundwater models which are used as investigative tools to test and explore the conclusions from the field-study. Both a bucket type model and distributed models of varying complexity are tested. Following the principle of parsimony (Hill, 2006; Hill and Tiedeman, 2007), the contributions of added model complexity are assessed via inter-model comparison (e.g., Hrachowitz et al., 2014; Clark et al., 2011; Carpenter and Georgakakos, 2006) and used to improve understanding of the watershed. Models are also used to predict response to a simple climate perturbation. Results from these exercises will be used to address the following questions:

1. Which parameters/processes are most important for reproducing/controlling discharge and water levels across the watershed? Posed differently: how much information content is available in common measurements of discharge and water levels?
2. How sensitive is watershed discharge to perturbations in recharge timing (e.g., earlier snowmelt). How well do predictions from different models agree?
3. When simulating and predicting watershed discharge, what is an appropriate level of complexity? How can future modeling be improved?

## 3.2 Model codes

### 3.2.1 Simple dynamical model

Following a parsimonious modeling strategy, the simple **d**ynamical **m**odel (SDM) (Kirchner, 2009) avoids explicitly accounting for specific and fine-scale subsurface processes and instead aims to capture overall or integrated watershed behavior. Watersheds are treated as dynamical systems represented by a reservoir whose discharge is a single valued, increasing function of storage. This storage discharge relationship is represented by a discharge sensitivity function,  $g(Q)$  ( $\text{d}^{-1}$ ). It is the derivative of  $Q$  with respect to watershed storage ( $S$ ) and represents the sensitivity of discharge to changes in storage. This function can be determined from “recession slope analysis” style plots (e.g., Brutsaert and Nieber, 1977), model calibration, or storage measurements (e.g., Brauer et al., 2013). In this work, the functional form of  $g(Q)$  was selected without any theoretical basis – it is represented as a polynomial (cubic) function of log-transformed discharge (similar to Equation 9 in Kirchner, 2009):

$$\ln(g(Q)) = \ln\left(\frac{dQ}{dS}\right) = c_4(\ln(Q))^3 + c_3(\ln(Q))^2 + c_2\ln(Q) + c_1 \quad [3.1]$$

where  $Q$  is watershed discharge ( $\text{mm d}^{-1}$ );  $S$  is watershed storage in units of depth (e.g., mm);  $c_1$ ,  $c_2$ ,  $c_3$ ,  $c_4$  are dimensionless coefficients.

Equation 3.1 is then combined with a watershed mass balance equation:

$$\frac{dS}{dt} = P - ET - Q \quad [3.2]$$

where  $P$  is precipitation ( $\text{mm d}^{-1}$ );  $ET$  is evapotranspiration ( $\text{mm d}^{-1}$ ) (assumed negligible, see Section 2.2.1).

The combination of equations 3.1 and 3.2 results in an initial value problem which can be solved for discharge over time:

$$\frac{dQ}{dt} = \frac{dQ}{dS} \frac{dS}{dt} = g(Q)(P - Q) \quad [3.3]$$

To avoid potential issues with the numerical solution of equation 3.3 (Clark and Kavetski, 2010; Kavetski and Clark, 2011), an implicit multi-step backward differentiation method is used to improve accuracy of the numerical approximation. Also, equation 3.3 is

formulated in terms of log-transformed discharge, which results in a function that is easier to integrate (Kirchner, 2009):

$$\frac{d\ln(Q)}{dt} = \frac{1}{Q} \frac{dQ}{dt} = g(Q) \left( \frac{P}{Q} - 1 \right) \quad [3.4]$$

### 3.2.2 Simple Fill and Spill Hydrology Model

The **Simple Fill and Spill Hydrology** (SFASH) model (Wright et al., 2009) was originally developed to simulate groundwater flow in the active layer overlying permafrost but has since been adopted for work in alpine regions (e.g., Kurylyk and Hayashi, 2017; Donnelly, 2012). SFASH reduces the three-dimensional groundwater flow problem into two dimensions by assuming (i) no vertical flow and (ii) hydraulic gradient equal to the water table slope (i.e. the Dupuit-Forschheimer assumptions). Unsaturated flow is also neglected. With these assumptions, flow in an unconfined aquifer is simulated using the two-dimensional Boussinesq equation:

$$S_y \frac{\partial h}{\partial t} = \frac{\partial}{\partial x} \left( T(b) \frac{\partial h}{\partial x} \right) + \frac{\partial}{\partial y} \left( T(b) \frac{\partial h}{\partial y} \right) + R \quad [3.5]$$

where  $S_y$  is specific yield;  $b$  is aquifer saturated thickness (m);  $h$  is hydraulic head (m);  $t$  is time (s);  $T$  is transmissivity ( $m^2s^{-1}$ );  $x$  and  $y$  are distances in the two ordinate directions (m);  $R$  is recharge rate ( $ms^{-1}$ ).

This non-linear partial differential equation is solved numerically for change in spatially distributed  $b$  and  $h$  with time. The model domain is represented as a structured grid, with nodes that can be individually assigned values of hydraulic conductivity, specific yield, thickness and elevation. No-flow or constant-head boundaries can be specified in the grid. Flows between adjacent nodes are approximated with finite differences. Transmissivity between adjacent nodes is determined by upstream weighting – saturated thickness of the node with the higher head is used to calculate transmissivity. Time stepping is done via the implicit Euler method coupled with Picard iteration – a matrix equation representing mass balance for every node is solved iteratively until the solution converges sufficiently close to the true solution.

Transmissivity feedback can be represented in SFASH by several functions  $K(z)$ , which describe change in hydraulic conductivity with depth. Some of the functions possible are: two-layer step, linear change, exponential change. Only the two-layer step function is used in this

work because it is numerically efficient and there is a lack of field data to inform a more appropriate function:

$$\begin{aligned} K(z) &= K_t, & \text{when } (0 < z < z_t) \\ K(z) &= K_b, & \text{when } (z_t < z < z_b) \end{aligned} \quad [3.6]$$

where:  $K_t$  is conductivity of the upper layer ( $\text{m s}^{-1}$ );  $K_b$  is conductivity of the lower layer ( $\text{m s}^{-1}$ );  $z$  is depth below the ground surface;  $z_t$  is transition depth (m);  $z_b$  is the depth to impermeable bedrock (m).

In Equation 3.6, transition depth,  $z_t$ , is calculated for each grid node as a function of thickness. A parameter,  $f_u$ , ranging from 0 to 1 is used to represent the fractional thickness of the upper layer (e.g., a value of 0.2 means that upper layer thickness is 2 m if total thickness is 10 m at a node).

### 3.3 Model construction and implementation

#### *3.3.1 Model domain geometry and boundaries*

##### SDM

The Hathataga watershed was modeled as a single reservoir with no explicit representation of spatial information (Figure 3.1). This representation can be considered 0-dimensional (Rupp et al., 2009) and takes no information regarding domain geometry or boundaries. Discharge from the reservoir was calculated using a non-linear discharge sensitivity function (Equation 3.1).

##### SFASH

In SFASH, the watershed area was represented as a 45 by 73 grid of nodes with 20 m spacing (Figure 3.2). The extent of the Hathataga watershed was determined using surface elevation (see section 2.3.7) and the Watershed tool in ArcMap (ESRI, 2014). Grid nodes outside of the watershed boundary were set as inactive. As a result, most of the active model domain is surrounded by no-flow boundaries. A specified-head boundary at the northern tip of the domain was used to represent the outlet spring. Head at this boundary was set to match water levels in piezometer PSP6 (Figure 3.3).

Elevation of the top of grid nodes was averaged from a 2 m DEM. Elevation of the bottom of grid nodes was set equal to bedrock elevation. The bedrock surface was estimated with geophysical data collected by Christensen et al. (2020) along several transects (Figure 3.3). Bedrock elevation measured along survey lines across the watershed were extrapolated by

manually contouring between survey lines. Bedrock elevation was then interpolated between contours using the Topo-to-Raster tool in ArcMap (ESRI, 2014) (Figure 3.3). This generated a 20 m DEM of bedrock elevation. Some manual corrections were made for several nodes near the top of the talus, where interpolation resulted in bedrock elevations that exceeded surface elevations.

### *3.3.2 Initial conditions*

#### *SDM*

SDM requires an initial discharge value,  $Q_0$ , to start the simulation.  $Q_0$  was set equal to measured discharge of Hathataga Outlet at the start of the simulation period.

#### *SFASH*

Initializing a SFASH run is more difficult – initial heads must be set for all nodes in the active model domain. These initial values were determined by “spinning up” the model as follows. A base model was run with initially zero saturated thickness. To roughly mimic the yearly recharge cycle at Hathataga,  $\sim 11 \text{ mm d}^{-1}$  of recharge was applied evenly across the model for four months of simulation time (total recharge  $\sim 1300 \text{ mm}$ ); recharge was then halted and the model was allowed to drain over the next eight months. The cycle was repeated for several years until storage and discharge of the model began to stabilize between years. Head in all nodes were saved at the end of the final recharge period. These saved heads were used to initialize subsequent SFASH simulations.

### *3.3.3 Precipitation and snowmelt forcing*

Recharge to models during rainfall events was determined from precipitation observations at AWS1. During rainfall events, model recharge was set equal to observed precipitation rate. This simple representation of recharge assumes negligible ET (see Section 2.2.1) and infiltration from surface runoff (only a few ephemeral streams and the lake exist within the site). However, during the calibration period, several snowfall and preceding melt events occurred. An attempt was made to account for delayed recharge during these brief events as follows:

Precipitation from August to October 2018 was manually flagged as snow or rain using air temperature data. When available, snow events were confirmed from photos or visual

observation. A simple temperature index snowmelt method (e.g., Moore, 1993; Donnelly, 2012) was used to approximate recharge from the melting of fallen snow:

$$M = C(T - T_H)\Delta t \quad [3.7]$$

where  $M$  is melt rate (mm);  $C$  is a melt factor ( $\text{mm } ^\circ\text{C}^{-1} \text{d}^{-1}$ );  $T$  is air temperature ( $^\circ\text{C}$ );  $T_H$  is threshold air temperature (above which snowmelt will occur;  $^\circ\text{C}$ );  $\Delta t$  time step (d).

The melt factor,  $C$  was calculated as:

$$C = C_{min} + 0.5(C_{max} - C_{min}) \left( 1 - \cos \frac{2\pi t}{365} \right) \quad [3.8]$$

where  $C_{min}$  is the minimum melt factor (on the winter solstice;  $\text{mm } ^\circ\text{C}^{-1} \text{d}^{-1}$ );  $C_{max}$  is the maximum melt factor (on the summer solstice;  $\text{mm } ^\circ\text{C}^{-1} \text{d}^{-1}$ );  $t$  is time to the winter solstice (d).

A temperature threshold of  $0.0 \text{ } ^\circ\text{C}$  was used. Values of  $1.0 \text{ mm } ^\circ\text{C}^{-1} \text{d}^{-1}$  and  $7.0 \text{ mm } ^\circ\text{C}^{-1} \text{d}^{-1}$  for  $C_{min}$  and  $C_{max}$ , respectively, were found to produce reasonable snowmelt rates and snowpack depletion times that were consistent with time-lapse photos. Calculated daily snowmelt was distributed across periods where  $T > T_H$  following the snowfall event. Snowmelt was set to zero if it exceeded the amount of accumulated snow from preceding snowfall events.

### 3.3.4 Simulation parameters/Convergence criteria

Both the SDM and SFASH numerically solve differential equations. To reduce the impact of numerical artifacts (e.g., Kavetski and Clark, 2011), both models use an implicit time stepping scheme (see Section 3.2). Conservative settings for solver convergence (maximum allowable change between nonlinear iterations) were set for all simulations with SFASH – a convergence criteria of  $5 \times 10^{-6} \text{ m}$  was set as the minimum allowable difference in head between Picard iterations before time was allowed to advance to the next step. Both models implement adaptive time stepping schemes, but were limited to a maximum step of 1 hour. SFASH simulations were also evaluated via mass balance: storage change, inflows and outflows were summed over the simulation period and checked for balance within 1% error.

### 3.3.5 Calibration and validation

SDM and SFASH model parameters (e.g., coefficients of the discharge sensitivity function, hydraulic conductivity, specific yield) were calibrated so they: (i) provided acceptable

match between model state variables/fluxes and observations; (ii) were realistic in terms of expert intuition or “hydrosense” (Anderson et al., 2015).

The **Parameter Estimation** (PEST) family of inverse codes was used to estimate parameters of both models. Specifically, PEST++ (Welter et al., 2015) and PESTHP (Doherty, 2020) were utilized. Briefly, PEST++ contains several quality-of-life improvements over vanilla PEST (Doherty, 2018) and has some additional functionality (e.g., more methods for optimization and uncertainty analysis). PESTHP is optimized to run models in parallel (e.g., using a computing cluster).

To estimate model parameters which could potentially minimize fit between observations and model output, PEST solves an indirect inverse problem (Yeh, 1986), in which a least-squares objective function is minimized:

$$\varphi = \sum w_i(r_i)^2 \quad [3.9]$$

Where  $\varphi$  is the value of the objective function;  $w_i$  is the weight of the  $i^{\text{th}}$  observation (data point in a discharge or water level time series);  $r_i$  is the value of the  $i^{\text{th}}$  residual between an observed and simulated quantity.

Equation 3.9 represents the fit between model outputs and corresponding field observations. Observations used for model calibration were outlet discharge and lake piezometer water levels over the period of July 17, 2018 to February 28, 2019. Observation weights were assigned subjectively, but were reflective of confidence in the observation (e.g., error and relevance of the measurement). Zero weights were assigned to periods when observations were unavailable (e.g., data gaps). Prior to calibration, the PyEMU library (White et al., 2016) was used to adjust observation weighting to avoid either discharge or water level observations from dominating the other in the objective function.

Model calibration began by running a model with a set of initial parameters that represent “best guesses”. Afterwards the sensitivity of model outputs to different parameters was determined during a sequence of model runs where a single parameter was perturbed at a time (the influence coefficient method described in Yeh, 1986). These calculated sensitivities were used to estimate an updated set of parameters, which produce a lower objective function (a better model fit). The cycle of sensitivity calculation and parameter updating was repeated until a minimum in the objective function was reached. The updated parameter set that produces this

objective minima is considered to be optimized. This general process is referred to as a gradient or regression method (Hill and Tiedeman, 2007).

Following calibration, models were also validated. During validation, models with calibrated parameters were run with forcing from the summer of 2019 (July 16 to September 26) and compared against observed discharge and water levels. Good performance in the validation phase indicates that calibrated models can produce realistic behavior in times/conditions outside of the calibration period and are not “over-fit”.

### 3.3.6 Parameter uncertainty

Estimated model parameters are those which provide the minimized objective function (Equation 3.9) after calibration. However, it is difficult to find a “true” set of parameter values due to significant model simplification and limited/noisy measurements. Estimated parameters are therefore uncertain, and should be expressed probabilistically. A useful statistical framework for this in the context of model parameters and calibration are Bayesian methods. A simple form of Bayes’ rule is:

$$p(\mathbf{k}|\mathbf{h}) \propto p(\mathbf{k})p(\mathbf{h}|\mathbf{k}) \quad [3.10]$$

where  $p(\mathbf{k}|\mathbf{h})$  is the posterior (after calibration) probability density function (PDF) of parameters  $\mathbf{k}$  conditioned by calibration data  $\mathbf{h}$ ;  $p(\mathbf{k})$  is the prior parameter PDF;  $p(\mathbf{h}|\mathbf{k})$  is the likelihood function which reflects model performance – the better the fit between model output and calibration data, the higher the value of this function.

Equation 3.10 is generally difficult to solve analytically. In this work, a simple and fast linear analysis method (Hill and Tiedeman, 2007, chapter 7) is used to approximate  $p(\mathbf{k}|\mathbf{h})$ . Uncertainty in estimated parameters is calculated using a form of equation 3.10 which assumes linear model behavior (i.e. consistent sensitivity of model output to parameters across parameter space) and that parameters and observation noise lie within multi-gaussian PDFs. This equation is used to calculate the posterior parameter covariance matrix (a symmetric matrix that holds parameter variances along the diagonal and covariances on the off-diagonals):

$$\mathbf{C}'(\mathbf{k}) = \mathbf{C}(\mathbf{k}) - \mathbf{C}(\mathbf{k})\mathbf{J}^T[\mathbf{J}\mathbf{C}(\mathbf{k})\mathbf{J}^T + \mathbf{C}(\boldsymbol{\varepsilon})]^{-1}\mathbf{J}\mathbf{C}(\mathbf{k}) \quad [3.11]$$

Where  $\mathbf{C}'(\mathbf{k})$  is the posterior parameter covariance matrix;  $\mathbf{C}(\mathbf{k})$  is the prior parameter covariance matrix;  $\mathbf{J}$  is the jacobian matrix of partial derivatives of observations with respect to parameters;  $\mathbf{C}(\boldsymbol{\varepsilon})$  is a matrix holding variances of noise associated with measurements, generally these

variances are assumed to be independent and based on precision of the method or sensor used to collect the measurement.

PEST++ and PyEMU were used to solve Equation 3.11. The two user inputs required are the prior covariance matrix,  $C(\mathbf{k})$ , and the observation noise covariance matrix,  $C(\epsilon)$ . A diagonal  $C(\mathbf{k})$  was input using prior parameter variances. Prior variance was determined from field measurements when possible (section 2.4.4) or based on values reported by studies of similar alpine deposits (e.g., Table 1 of Hood and Hayashi, 2015). A diagonal  $C(\epsilon)$  was determined from observation weights based on measurement error and reliability (see section 3.3.5). The jacobian used was based on estimated parameter values at the end of calibration. Calculated parameter variances were then used to calculate 95% confidence intervals. Because Equation 3.11 assumes that parameters are normally distributed, confidence intervals are calculated as:

$$\mathbf{k}_b = \mathbf{k} \pm 2\sigma \quad [3.12]$$

Where  $k_b$  is the upper or lower bound of the confidence interval;  $k$  is the parameter value; and  $\sigma$  is the parameter standard deviation (square root of variance).

To further facilitate comparisons of estimation uncertainty between parameters, the parameter coefficient of variation (Hill and Tiedeman, 2007, Equation 7.4) was calculated as:

$$\textit{Parameter coefficient of variation} = \frac{\sigma}{k} \quad [3.13]$$

The parameter coefficient of variation is a dimensionless statistic that can be used to evaluate the relative uncertainty of parameters with different units or ranges in magnitude. Parameters to which observations are more sensitive will be estimated with more precision, reflected by a lower coefficient of variation. This statistic was not calculated for parameters which have negative or zero values.

The sensitivity and uncertainty of parameter values estimated by calibration were evaluated comparing prior and posterior variances, calculated confidence intervals, and parameter coefficients of variation. These values and other linear, jacobian-based statistics (e.g., Hill and Tiedeman, 2007; Doherty, 2015) may suffer, however, in cases where models are highly non-linear.

### 3.4 Model inter-comparison

Several SDM and SFASH models were calibrated and validated using field observations. Their performance was compared to investigate potentially important watershed processes/properties. Later calibrated models were run under different, simple recharge scenarios to explore how baseflow may be sensitive between different model representations.

#### *3.4.1 Description of model setups*

Five model representations of the Hathataga watershed were used. They are numbered 1 to 5 in order of increasing process and parameter complexity (Figure 3.4). Each model setup introduces a new process or property (as described below) and their comparison was used to gauge the importance of each refinement. See Table 3.1 for a summary of all model parameters.

##### Setup 1: Simple dynamical model

This is the simplest model, representing the watershed as a single non-linear reservoir (Figure 3.1). A third-order polynomial was used for  $g(Q)$  (Equation 3.1).

##### Setup 2: SFASH, homogenous, no transmissivity feedback

This setup is a departure from the bucket representation of setup 1. The watershed is represented by a distributed model domain, which explicitly models groundwater flow and storage across the watershed. Parameterization, however, is kept simple by neglecting any heterogeneity – only a single value of storage and hydraulic conductivity were assigned to the entire model domain.

##### Setup 3: SFASH, lateral homogeneity, two-layer step change in vertical conductivity

Building upon the distributed model domain in setup 2, this model introduces vertical hydraulic conductivity variability to the laterally homogenous model. This is represented by Equation 3.6, as a high conductivity layer ( $K_t$ ) overlaying a low conductivity layer ( $K_b$ ).

##### Setup 4: SFASH, lateral zones of piecewise constancy, no transmissivity feedback

Setup 4 is also derived from setup 2. Lateral heterogeneity was introduced by separating the model domain into zones of piecewise constancy. Zones were delineated based on land cover and deposit types identified from air photos and field observation (Figure 3.2). Each zone was assigned conductivity and storage parameters after the surficial deposit type they represent. Surficial deposits were partitioned into the following zones: moraine; forest + meadow; talus.

Setup 5: SFASH, lateral zones of piecewise constancy, two-layer step change in vertical conductivity

This is the most highly parameterized model setup. The watershed is represented by a distributed model domain. Vertical and lateral variability in hydraulic conductivity are represented in the same manner as setups 3 and 4, respectively.

*3.4.2 Model performance evaluation*

Performance of models during both calibration and validation was judged visually and with the use of a “goodness of fit” measure: the Nash Sutcliffe Efficiency (NSE) (Nash and Sutcliffe, 1970):

$$NSE = 1 - \frac{\sum_{i=1}^N (o_i - m_i)^2}{\sum_{i=1}^N (o_i - \bar{o})^2} \quad [3.14]$$

where  $o_i$  is the  $i^{\text{th}}$  observed value;  $m_i$  is the  $i^{\text{th}}$  simulated value;  $\bar{o}$  is the average of all observations;  $N$  is the total number of observation/simulated values.

For discharge data, log-NSE was also calculated. The calculation of log-NSE is the same as NSE, but observations and simulated values are first log transformed. Because fit to higher flows or water levels may dominate NSE, log-NSE was included to evaluate model performance with greater “weighting” to the baseflow period.

Both goodness of fit measures are simple, intuitive and consistent for all models (unlike Equation 3.9 which can use different weighting between models). Values of NSE and log-NSE range from  $<0$  to 1. A value of 1 indicates a perfect fit between model results and observations whereas, negative values indicate that the average value of observations produces a better fit than the model.

*3.4.3 Recharge scenarios*

To test the sensitivity of simulated baseflow magnitude to changes in recharge, calibrated models were used under several simple recharge scenarios (Figure 3.5) as explained below. All recharge scenarios are highly simplistic, but can give insight into how model discharge is affected by change in the temporal distribution of recharge via relative comparison. The four scenarios described below focus on comparing two aspects of recharge timing: (i) earlier shift in snowmelt period and (ii) importance of subsequent summer/fall precipitation. Comparison of output across model setups may give insight on important processes and properties which control

the sensitivity of baseflow. Results from this comparison may also provide a useful reference for future work where models are used to make discharge predictions.

For each model setup, 100 members of an ensemble were run for each scenario. These ensembles were constructed by sampling parameters from their estimated posterior confidence intervals (Equation 3.12). Average modeled baseflow magnitudes (calculated here as the sum of February-March discharge) were compared between setups.

All recharge scenarios span a year, starting on April 1, and cover a range in recharge timing and intensity. Scenarios are as follows:

*Scenario 1: snowmelt*

This is the “base” scenario. Recharge begins on May 1, linearly increasing over a 10-day period to a maximum rate of 11 mm d<sup>-1</sup>. The 10-day rise represents a quick ramp up in snowmelt as temperatures rise in spring, and the maximum flow rate is a rough estimate of an average snowmelt rate. This maximum rate is maintained until August 4 for a total of 1000 mm cumulative recharge. In this scenario all recharge comes as a large and consistent snowmelt period without any storm events.

*Scenario 2: post-snowmelt storms*

In this scenario, recharge is partitioned into a snowmelt period and subsequent storm period. The snowmelt period lasts from May 1 to August 4, similar to scenario 1, but the maximum recharge rate is reduced to 8.8 mm d<sup>-1</sup> in order to add the storm period while maintaining the same total recharge amount. As a result cumulative recharge of the snowmelt period is 800 mm. In the August to October period following snowmelt, several storms occur, each lasting 32 hours with a maximum precipitation rate of 20 mm d<sup>-1</sup>. These values are based roughly on observed storms which typically release ~20 to 30 mm over one to several days. Cumulative recharge of these storm events is 200 mm for a total of 1000 mm.

*Scenario 3: early snowmelt*

This scenario is nearly identical to the first. The snowmelt period is shifted one month earlier, beginning on April 1 and ending on July 5. This scenario is meant to represent a shift toward warmer climate, where snowmelt is expected to begin earlier.

#### Scenario 4: early snowmelt, post-snowmelt storms

In this scenario, the snowmelt period is shifted one month earlier, beginning on April 1 and ending on July 5. The storm period is identical to scenario 2, lasting from August to October.

### 3.5 Results

#### 3.5.1 Calibration and validation

Model parameters were estimated through calibration (Equation 3.10). Post-calibration (posterior) PDFs are reported in terms of their estimated value and 95% confidence intervals in Table 3.2. Estimated parameter values are those which result in the minimum objective function value found by PEST. Confidence intervals around estimates were calculated using Equations 3.11 and 3.12. When these intervals resulted in unrealistic values (e.g., a negative lower boundary of storage), realistic boundaries were enforced (Table 3.1).

Estimated parameter values were used in calibrated models whose performance and behavior are described in this section. Discharge and lake piezometer water levels of calibrated models are plotted against observations for the calibration period (July 17, 2018 to February 28, 2019) in Figure 3.6 and for the validation period (July 16 to September 26, 2019) in Figure 3.7.

#### Setup 1: SDM

The SDM discharge sensitivity function (Equation 3.1) with estimated coefficients is plotted in Figure 3.8a. This calibrated sensitivity function takes a low value for the low discharge range ( $0.2 \text{ mm d}^{-1}$  to  $1 \text{ mm d}^{-1}$ ). This results in slow discharge recession at low flow, shown by the slope of the discharge – storage relationship in Figure 3.8b. With higher flows, the sensitivity function increases more rapidly by orders of magnitude.

During calibration, discharge of the SDM matched observations well at early time (Figure 3.6), reproducing recession from  $8 \text{ mm d}^{-1}$  to less than  $3 \text{ mm d}^{-1}$ . Toward mid-August, when observed recession slowed down, modeled discharge reacted strongly to recharge events, rising above the observed hydrograph. This resulted in sharp and unrealistic hydrograph behavior and a large over estimation of discharge. Modeled discharge receded quickly when recharge stopped (around October 1) and was better at matching observed flows later in the winter months. NSE and log-NSE of discharge in the calibration period were similar (Table 3.3), indicating that the model was equally good at matching higher and lower observed flows.

Fit between modeled and observed discharge was worse during validation, reflected by lower values of NSE and log-NSE (Table 3.3). Discharge was overestimated for the majority of

the period (Figure 3.7). Modeled discharge again had sharp and unrealistically large responses to precipitation events.

Setup 2: SFASH, homogenous, no transmissivity feedback

Estimated watershed hydraulic conductivity and specific yield were  $6.5 \times 10^{-4} \text{ m s}^{-1}$  and 0.14 respectively; these properties (especially conductivity) match well with measured and inferred properties of the moraine (e.g., moraine slug test conductivity between  $\sim 10^{-4}$  to  $\sim 10^{-3} \text{ m s}^{-1}$ ).

In calibration, performance of setup 2 is mixed (Figure 3.6). Discharge recession from higher flows at early time (mid-July to mid-August) and lower flows later on (December onward) were well matched. However, modeled discharge during late summer and fall was larger than the observed hydrograph, resulting in an overestimation. Discharge of setup 2 appears to respond to recharge events more than observations do (peaks in September and October), but does not have the same, sharp responses of setup 1. Simulated water levels at the lake piezometer were overestimated at early time and underestimated at late time. The degree of mismatch was significant, exceeding 2 m difference at the start and end of the observation time series. During the middle period, fit was better, but modeled water levels still showed a stronger response to recharge events than observed.

Compared against the validation data set, fit to observed discharge appeared to improve. Model discharge accurately follows general trends in discharge over the simulation period without any huge over- or under-estimation. Shorter term discharge fluctuations (e.g., a rise in discharge at the start of September) were not captured by the model. Water levels, on the other hand, are overestimated throughout the entire period. This overestimation is worst at early time, exceeding 4 m difference. This gap closes with time, as modeled water levels drop faster than observed. NSE of water levels during validation was negative (Table 3.3), reflecting poor model performance.

Setup 3: SFASH, lateral homogeneity, two-layer step change in vertical conductivity

Estimated upper and lower conductivity of the watershed were  $1.8 \times 10^{-3} \text{ m s}^{-1}$  and  $1.6 \times 10^{-4} \text{ m s}^{-1}$  respectively. These values, again, match closely with measurements in the moraine. The estimated proportion of the upper layer (see Section 3.2.2) was 0.71, meaning that the upper

layer represented just over 70% of the thickness of a given model grid node. Specific storage was estimated as 0.22.

Inclusion of vertical variability in setup 3 resulted in somewhat better performance during calibration over setup 2 (Figure 3.6). The modeled discharge hydrograph sees a slight improvement in NSE and log-NSE (Table 3.3) and appears smoother and less sensitive to recharge events. Overestimation of discharge was again apparent starting from mid-August. Recession of model discharge in the early period (up to mid-August) was also too steep, causing a small underestimation of discharge. Compared to setup 2, modeled water levels fit much more closely to calibration data, never deviating by more than 1 m. However, small fluctuations in water levels were not perfectly reproduced. There also was a slight over and underestimation of water levels in July and late November, respectively.

During validation, fit to discharge was good (Table 3.3) but still failed to match shorter term fluctuations (Figure 3.7). Fit of modeled water levels was worse than calibration. There was a ~2 m overestimation of water level at early time. As the simulation progressed, modeled water levels approached observations to closely match them during September. Although clearly better than setup 2, NSE of water levels for setup 3 during validation was still negative (Table 3.3), indicating poor model performance.

#### Setup 4: SFASH, lateral zones of piecewise constancy, no transmissivity feedback

Estimated parameters of moraine hydraulic conductivity and specific yield were  $6.9 \times 10^{-4} \text{ m s}^{-1}$  and 0.2 respectively. This moraine conductivity is similar to what was estimated in setup 2 and to field measurements. Conductivity and specific yield of the forest + meadow were  $2.2 \times 10^{-4} \text{ m s}^{-1}$  and 0.1 respectively. Estimated conductivity of the talus was  $1.1 \times 10^{-4} \text{ m s}^{-1}$ , lower than the forest + meadow and much lower than the values obtained from field tracer tests. Estimated specific yield of the talus was 0.1.

The addition of lateral heterogeneity in setup 4 does not appear to produce a large increase in performance compared to setup 2. In calibration, modeled discharge overestimated observations in the middle period, but fit early and late times well (Figure 3.6), similar to setup 3. Fit of modeled water levels was a little better than setup 2, but worse than setup 3. Similar to setup 2, water levels were over and underestimated during early and late times, respectively.

Modeled discharge and water levels during validation appear similar to setup 2 (Figure 3.7). Modeled discharge matched observations, but appear too smooth and insensitive to small

fluctuations. Water levels were highly overestimated to a degree similar to setup 2, with highly negative NSE (Table 3.3).

Setup 5: SFASH, lateral zones of piecewise constancy, two-layer step change in vertical conductivity

All estimated parameters of setup 5 are listed in Table 3.2. The estimated upper and lower conductivity of the moraine were  $1.2 \times 10^{-2} \text{ m s}^{-1}$  and  $1.1 \times 10^{-4} \text{ m s}^{-1}$  respectively. Both layers were estimated to have roughly equal thickness. Moraine specific yield was 0.12. In the talus, estimated upper and lower conductivity were much lower than field measurements, ranging from around  $10^{-5}$  to  $10^{-4} \text{ ms}^{-1}$ . The lower layer was estimated to be dominant, occupying (nearly) all thickness of a grid node. In the meadow, conductivities were estimated ranging from around  $10^{-7}$  to  $10^{-4} \text{ ms}^{-1}$  which is in agreement with field measurements. Upper and lower layer thickness in this zone was also estimated to be roughly equal. Specific yield of the talus and forest + meadow were estimated as 0.35 and 0.40 respectively. These values are 3 to 4 times larger than estimated for these zones in setup 4. For the forest + meadow, which are composed of fine soils, these values are similar to porosity of soils in other alpine sites (Hood, 2013).

The added complexity in setup 5 appears to benefit model performance. During calibration model discharge fit observations well throughout the simulation and lacked the period of overestimation seen in every other setup (Figure 3.6). Calculated NSE and log-NSE were the highest of all setups (Table 3.3). Similar to setup 1, modeled discharge reacted sharply to recharge events, however, it did not increase as strongly. Water levels also showed the closest fit out of all setups and appeared to respond sharply to recharge. From October onward, water levels leveled out and did not decrease by much more. This behavior is not seen in other setups – other models show water levels decreasing by several meters between October and March.

Performance of setup 5 suffers in the validation period. Modeled discharge largely underestimated observations, especially during the second half of the period, during lower flows. Sharp responses to recharge were more pronounced and contrast observed behavior. NSE and log-NSE of discharge in the validation period were the lowest out of all setups (Table 3.3). A large difference between NSE and log-NSE reflects worse model performance for lower flows. In contrast, water levels did better, and showed the closest fit out of all model setups in validation.

### 3.5.2 *Parameter uncertainty*

Estimated parameter coefficients of variation, and prior and posterior confidence intervals are listed in Tables 3.1 and 3.2. Prior to calibration, model parameter ranges were informed solely by educated opinion, field measurements and literature values. Through calibration, parameter values were further constrained by information in the calibration dataset (Equation 3.10). Parameters to which model output (corresponding to calibration observations) was more sensitive were estimated more precisely, reflected by their narrower confidence intervals and smaller coefficients of variation. These parameters are implied to be more important to the calibration because they have a greater influence on model output corresponding to calibration data. In contrast, parameters whose prior and posterior variance are similar did not greatly benefit from the calibration process because they did not control model output as much.

Parameters in setups 2 to 3 were estimated with high precision (Table 3.2). These models have few adjustable parameters (2 to 4), therefore it is sensible that none appear to be redundant. Results from setup 1 are hard to interpret intuitively, given the abstract nature of its parameters. Setups 2 and 3 show (obviously) that watershed hydraulic conductivity and specific yield are important in affecting modeled discharge and water levels.

In the laterally distributed models (setups 4 and 5), uncertainty of parameters belonging to different delineated surficial deposits show an interesting contrast. Generally moraine parameters were estimated with more precision and had much smaller values of parameter coefficient of variation (Table 3.2). Properties of the talus and vegetation, while not unimportant, were not estimated as confidently. In setup 5, uncertainty of upper talus hydraulic conductivity was not reduced by calibration at all. This apparent redundancy was a consequence of non-linear model behavior: because the estimated proportional thickness of the upper talus was zero (Table 3.2), hydraulic conductivity of the upper talus layer had no effect on model output and was therefore not constrained by calibration in that particular region of parameter space. If this proportional thickness were allowed to increase, model sensitivity to talus upper layer conductivity would presumably increase as well.

### 3.5.3 *Recharge scenarios*

Ensembles (100 members) for each model setup (1 to 5) were run for each recharge scenario. Average baseflow, calculated as the sum of February-March discharge, are plotted in Figure 3.9 for each setup and recharge scenario. The magnitude of average baseflow varied

greatly between different model setups. For example, baseflow of setup 5 was several times larger than the other setups for all scenarios. Sensitivity of baseflow to different recharge scenarios was also variable. Average baseflows of setups 1 and 5 were similar in all recharge scenarios, indicating insensitivity to recharge timing, whereas setups 2, 3 and 4 show much greater baseflow sensitivity (Figures 3.9 and 3.10). Variability in baseflow between ensemble members is shown by error bars which have a length of two standard deviations in Figure 3.9. Error bars for setups 2 and 3 are much narrower than other setups and generally do not overlap. This is reflective of relatively narrow parameter confidence intervals from which ensemble members of these setups were sampled (Table 3.2).

Percentage difference in baseflow between recharge scenario 1 (the base case) and the other scenarios was calculated for each ensemble member and plotted in Figure 3.10. Although magnitude of baseflow may be highly variable for a given ensemble (error bars for setups 1, 3, 5 in Figure 3.9), the differences in baseflow between recharge scenarios were consistent. Recharge scenario 2, where precipitation occurs in summer and fall caused increases in modeled February-March baseflow for all models, suggesting importance of these later-time precipitation events. The importance of recharge timing is also shown from scenario 3, where earlier snowmelt caused consistent decreases in modeled baseflow. However, scenario 4 results indicate that post-snowmelt storms can compensate for an early shift in snowmelt, showing moderate to large increases in baseflow in setups 1 to 4. In setup 5, averaged baseflow in scenario 4 still decreases, but less so compared to scenario 3. Overall, results from these scenarios indicate that the distribution of previous recharge influences model baseflow, but to differing degrees for different models.

## 3.6 Discussion

### *3.6.1 Comparison of model performance and parameter values*

The performance of various model setups are compared to explore and determine which processes and parameters may be a good representation of watershed behavior. Similar multi-model comparisons have permitted useful insights in other work (e.g., Buytaert and Beven, 2011; Clark et al., 2011; Hrachowitz et al., 2014). Results from calibration show that no model setup performed perfectly, however fit to observations generally improved with complexity. This is unsurprising as additional parameters and process representation usually grant models greater flexibility to match the calibration data set.

Setup 1 (single reservoir SDM model) performs adequately during parts of the calibration and validation period, but has a relatively poor fit compared to the distributed models. More importantly, parameters of this model are highly abstracted and hard to explicitly relate to real processes or properties. Because of these issues, this model is excluded from the following discussion.

### *The moraine gatekeeper*

In Chapter 2, the Hathataga moraine was identified as a “gatekeeper”, which regulates and controls groundwater flow from the watershed. Because this moraine dominates watershed discharge behavior, it was hypothesized that modeling of Hathataga may only necessitate characterization of the moraine. Setups 2 and 3 can be seen as test-cases for this hypothesis, where only a single “unit” is used to characterize the entire watershed. In these models, it is encouraging to see that overall estimated watershed conductivity (Table 3.2) matches well with the range of measured moraine values (Figure 2.14). This may indicate that the “effective” conductivity of the watershed is indeed reflective of the moraine rather than other units. These models (particularly setup 3) also provide a mostly adequate (Table 3.3) fit with observations during calibration and validation.

Models which include different units within the watershed also demonstrate the importance of the moraine gatekeeper. In setups 4 and 5, lateral heterogeneity is expressed by zones corresponding to different surficial deposits at the land surface. In these models, estimated parameters pertaining to the moraine have a greater influence on model output and are estimated with smaller coefficients of variation compared to parameters of the talus and forest + meadow (Table 3.2). This reflects greater sensitivity of discharge and water levels to the moraine, suggesting its importance for controlling watershed hydrologic behavior. This supports the conclusions of the field study, which describe the moraine as an influential “gatekeeper”. Parameters pertaining to other units, however, still have an influence on the output of these models, which may justify their inclusion in the structure of models; this is discussed further in the following section.

### *Lateral and vertical heterogeneity*

Setup 2 is a homogenous, distributed SFASH model and can serve as a “base model” to compare the refinements in setups 3 to 5. Simple representations of vertical and lateral

heterogeneity in hydraulic conductivity are introduced to the model in setups 3 and 4, respectively. The overall discharge hydrographs produced by both these models appear visually similar to one another (Figure 3.6 and 3.7) and only offer a slightly improved fit to observations. It is therefore difficult to distinguish these models setups based on only discharge data. This finding is similar to that of Pauritsch et al. (2017), who found that discharge data alone was insufficient to resolve watershed internal structures through inverse modeling.

However, when water levels are considered, the difference in outputs between setups 3 and 4 is much more distinct. During calibration, setup 4 grossly over- and under-estimates observed water levels at different times (Figure 3.6), whereas in setup 3, water levels are generally well fit. This difference demonstrates the importance of representing vertical variability in models. When hydraulic conductivity is uniform, like in setups 2 and 4, model transmissivity only changes with saturated thickness, and lacks versatility to reproduce the observed range of water levels. However, when transmissivity varies with saturated thickness and (just) a simple conductivity step function, water levels can be closely reproduced. The importance of vertical variability is also reflected in narrowly estimated confidence intervals of parameters which control transition depth (Table 3.2).

The refinements in setup 3 and 4 fail to provide satisfying fit to all observations, indicating that both models lack certain elements important for determining watershed hydrological functioning. However, when the lateral and vertical heterogeneity of setups 3 and 4 are jointly represented in setup 5, performance during calibration is considerably improved (Figure 3.6 and Table 3.3). This observation suggests that lateral variability also has an important influence on watershed storage and discharge, and must be represented for a model to perform accurately. The talus and forest + meadow units may have some control on watershed storage and total discharge that is not completely masked away by the moraine. The importance of these other units may vary with time and conditions of the watershed. For example, during the low flow period, low conductivity regions of forest + meadow could contribute a larger fraction of watershed discharge compared to the other units which drain and deplete more quickly. During this time, the forest + meadow would have a larger influence on the magnitude and recession of watershed discharge.

Overall, inter-model comparison indicates that neither chosen representation of simple lateral and vertical heterogeneity in hydraulic conductivity is sufficient to achieve perfect model

performance on their own. In terms of discharge data, setups 3 and 4 appear to perform equally well and only slightly better than setup 2. This indicates that discharge data contains insufficient information to make inferences on model structure. Addition of water level data allows for better differentiation between the laterally and vertically varying parameterization schemes and indicates that vertical variability provides a greater benefit than lateral variability. When both types of heterogeneity are included within the model, better calibration performance can be achieved. However, it is important to note that despite representing both lateral and vertical variability, setup 5 is clearly not perfect. Unrealistic behavior such as small but sharp spikes in discharge during storms, and poor performance during validation (Figure 3.7 and Table 3.3) indicate that setup 5 may already be over-parameterized.

#### Model storage-discharge

Storage-discharge (S-Q) relationships were constructed for all model setups (top row of Figure 3.11). S-Q curves for calibration and validation periods are plotted alongside a purely drainage simulation (where models were allowed to drain from relatively saturated conditions without any recharge). For the most part, S-Q curves of setups 2-5 appear visually similar, showing a steady increase in discharge with storage. In SFASH models, sharp departures from the drainage case occur in response to recharge events during the calibration and validation simulations, resulting in non-unique curves. These departures can be large, especially for setup 5, indicating the potential influence of recharge history and initial conditions on watershed behavior. In comparison, the single-reservoir SDM model is incapable of producing this behavior and must follow a single-valued S-Q curve.

From S-Q curves, discharge sensitivity relationships ( $g(Q)$ ) were approximated numerically for all SFASH models (bottom row of Figure 3.11). These approximations are heavily contaminated by numerical noise, however, visual inspection of the drainage  $g(Q)$  and corresponding S-Q curves allow for some comparisons to be made. Generally, SFASH models appear to behave similarly, with discharge steadily increasing alongside storage. Setups 3 and 5, which include vertical variability, have slightly lower values of  $g(Q)$  (and less steep S-Q curves) during low flow than setups 2 and 4. In contrast, characteristics of setup 1 are vastly different, with estimated values of  $g(Q)$  that are generally orders of magnitude higher than the other setups over the simulated range in discharge. These high values of discharge sensitivity result in sharp changes in discharge, compared to the other models (e.g., Figures 3.6 and 3.7). S-Q of setup 1 is

also unlike any of the SFASH curves, showing a slow and gradual increase in discharge with storage before a sudden and sharp increase (after reaching  $Q \sim 2 \text{ mm d}^{-1}$ ). Poor performance and mismatch in S-Q of setup 1 compared to the SFASH models may suggest that the selected functional form of  $g(Q)$  (Equation 3.1) is ill-suited for representing the Hathataga watershed.

#### Other model limitations

The inability of models to reproduce certain aspects of observed discharge and water level time series may also be reflective of other model limitations outside of the representation or omission of heterogeneity discussed above. For example, model recharge is assumed equal to precipitation rate measured at AWS1 or calculated by a highly simplified temperature index snowmelt model (Equation 3.7). This fails to account for any spatial variability or elevation effects. Similarly, initial conditions of all SFASH model runs were determined using steady and spatially uniform recharge. This is likely a poor substitute for actual distribution of storage across the watershed, which is a product of a more complex and variable recharge history. These inadequacies may contribute to some of the inconsistent model behavior observed between calibration and validation simulations.

Models also lack explicit representation of detailed processes which may influence certain behaviors. For example, uncharacteristically rapid changes in discharge and water level are observed to occur in late-August and early-September of 2019 (observed values in Figure 3.7). These changes are roughly coincident with depletion of Hathataga Lake and probably reflect the transition from surface water to ground water regimes – storage levels and discharge fall and rise more rapidly in the porous medium of the moraine after loss of the surface water body. This occurs because of a rapid drop in specific yield from open water of the lake to porosity of the moraine sediment. Models however, fail to capture this sudden change in specific yield and do not reproduce these small scale fluctuations.

#### *3.6.2 Baseflow sensitivity to recharge perturbation*

Simulations with models under different recharge scenarios produce wide ranges of February-March total baseflow, both between scenarios and models. Although the motivation for this exercise was not to make accurate predictions of future baseflow for the Hathataga watershed, this has implications for work where similar models are used to predict response of discharge to future change. All models provide at least a passable match to low flows during

calibration and/or validation (see Figure 3.6 and 3.7; log-NSE in Table 3.2), which indicates suitability to simulate and make predictions of baseflow for the watershed. However, baseflow totals simulated by one setup can be several times larger or smaller than another (Figure 3.9). The uncertainty of predicted baseflow using multiple models is therefore high, and larger than indicated by linear uncertainty analysis applied to each individual model (error bars in Figure 3.9). This result indicates that the structure/design of a predictive model should be handled carefully (Watson et al., 2013), and based upon more than just basic goodness-of-fit measures like NSE.

Despite large variability in simulated baseflow magnitude, models are more consistent when reflecting differences between recharge scenarios (Figure 3.10). Generally, baseflow of all models shows some degree of sensitivity to recharge timing. The importance of recharge from storm events after the snowmelt period are indicated by results from scenarios 2 and 4. These storm events can result in increases of modeled baseflow even when snowmelt is shifted earlier. Different models have varying levels of sensitivity to recharge scenarios. Setup 1 and 5 show more stability against changes in recharge. In the case of setup 1, low values of the estimated discharge sensitivity function (Figure 3.8) cause insensitivity during low flows. With setup 5, the cause for insensitivity is more intuitive – the forest + meadow was assigned a small value of lower layer hydraulic conductivity ( $K_{bot,v}$  in Table 3.2), allowing it to act as a “longer term” storage that appears less responsive to changes in recharge. Setups 2 to 4 lacked this kind of long response time/low conductivity storage and were therefore more influenced by short term changes to recharge.

Model results for recharge scenarios can be compared against February-March baseflow measured at Hathamaga in previous years. Despite large differences in conditions and forcing between the model scenarios and reality, this comparison can test if simulations are in the right order of magnitude. Observed February-March total baseflow was ~15 mm in 2017 and ~25 mm in 2018. This magnitude of observed baseflow is well within the range of setups 1 to 4, but much lower than baseflow from setup 5 (Figure 3.9). Similarly, the difference in total baseflow between 2017 and 2018 is large, signifying sensitivity to year-to-year changes in recharge (assuming other factors influencing discharge are stationary). Precipitation measured at AWS1 in the hydrological years leading up to the 2017 and 2018 baseflow period was 1070 mm and 1215 mm, respectively. Variability in observed baseflow therefore may reflect difference in both

recharge timing and magnitude. However, these baseflow observations still point to faults in setup 5, and emphasize that good calibration performance based on NSE may not be enough to determine an optimal predictive model.

### *3.6.3 Optimizing the structure of alpine groundwater models*

Discussion in the previous two sections demonstrate the importance of different aspects of model construction for both performance during calibration/validation and in making predictions. As the modeling of alpine groundwater systems moves toward larger scales and unstudied regions (Hayashi, 2020), key questions relating to model design and implementation must be faced. How complex should a model be? This has been a long discussed topic (e.g., Yeh and Yoon, 1981; Hill, 2006; Hrachowitz et al., 2014), but is particularly important given data scarcity of alpine regions.

Lessons from the inter-model comparison can be used to explore potential routes towards more optimal model structure. For distributed SFASH models (setups 2 to 5), this comparison has shown that simpler models (e.g., setup 2) are unable to achieve good fit to multiple calibration observations. In contrast, more complex models (e.g., setup 5) do see generally improved fit in calibration, but yield inconsistent validation performance and unreasonable predictions (Figures 3.7 and 3.9). This echoes findings of previous work, which indicate that at a certain point, greater model complexity does not necessarily improve a model (e.g., Jakeman and Hornberger, 1993; Kunstmann et al., 2006). More optimal model structures for simulating groundwater storage and discharge in Hathataga likely sit between setups 2 and 5 in terms of complexity (Figure 3.4). The moraine is important for controlling discharge, but lateral and vertical variability should be captured and represented in some form.

Compared to distributed SFASH models, the SDM reservoir model is an even simpler but easier-to-use method of modeling groundwater. The implementation of the SDM in this study (setup 1) fails to achieve good fit in either calibration or validation, indicating that more complexity is required. However, avenues to enhance the performance of the SDM are not intuitive and require good understanding of the processes and properties that influence watershed hydrogeological behavior; how they scale/generalize; and how they are quantified in terms of storage-discharge relationships. Using lessons from the previous modeling exercises, two attempts were made to try to improve performance of the SDM in the following.

### Alternative $g(Q)$ SDM

The functional form of the discharge sensitivity function  $g(Q)$  (Equation 3.1) in the original model design was selected arbitrarily, following Kirchner (2009). However, without any strong theoretical basis, the form of Equation 3.1 is likely not optimal and may contribute to poor performance of the model at Hathataga. Compared to the better performing SFASH models, calibrated  $g(Q)$  and the storage-discharge relationship of setup 1 appear unlike (Figure 3.11). To explore the effects of an alternative form for  $g(Q)$ , it was reformulated as:

$$g(Q) = c_1 Q^{c_2} + c_3 \quad [3.15]$$

where  $Q$  is watershed discharge ( $\text{mm d}^{-1}$ );  $c_1$ ,  $c_2$ ,  $c_3$  are coefficients determined through calibration.

Coefficients  $c_1$ ,  $c_2$ ,  $c_3$  were set manually to produce a  $g(Q)$  and S-Q curve which closely matched that of setup 5 during drainage (Figure 3.12). This was done in an attempt to recreate the behavior of setup 5.

The SDM model with this alternative  $g(Q)$  was run over the calibration and validation periods (Figure 3.13). It is apparent that simply matching  $g(Q)$  with setup 5 does not result in similar discharge behavior during calibration. Like the original SDM model (setup 1), discharge recession with the alternative  $g(Q)$  match well with observations during early and late periods, but sharp response to recharge during the months of August and September result in large overestimation of discharge. This result suggests that simply changing  $g(Q)$  to match a better performing model does not offer a meaningful benefit to the SDM. To better capture response to recharge, it may be important to represent hysteresis in the storage-discharge (and discharge sensitivity) relationship. For example, S-Q curves of setup 5 (Figure 3.11) are non-unique, showing that response of discharge to recharge events is variable and does not necessarily follow the drainage curve. Representation of this hysteresis could be added by redesigning the model with multiple reservoirs (Kirchner, 2009).

### Serial SDM

An upper reservoir was added to the model and linked to a lower reservoir in series (discharge from the upper reservoir feeds the lower reservoir). Both reservoirs were parameterized by separate, quadratic  $g(Q)$  functions (similar to Equation 3.1 except  $c_4 = 0$ ; Figure 3.14) and a “proportionality” ratio which partitions recharge between them. The idea

behind this model structure was to represent both vertical heterogeneity (through  $g(Q)$  functions) and lateral heterogeneity/multiple storages (through two separate reservoirs). Reservoirs were connected in series rather than in parallel to imitate the conceptual structure of the Hathataga watershed, where the moraine (lower reservoir) is fed flow from the surficial deposits and regions further above (upper reservoir). Multiple reservoirs also enable the model to have hysteretic behavior because total watershed discharge (from the lower reservoir) can have non-unique values as a function of total storage (of both reservoirs).

This enhanced “serial SDM” was calibrated using PEST and subjected to validation. A quick look at modeled discharge shows substantially improved performance during calibration (Figure 3.15a) compared to the original SDM. Compared to other iterations of the SDM, sharp responses and discharge over-prediction during calibration are reduced. This is attributed to the additional reservoir, which allows the model to effectively bank additional recharge and limit the increase in discharge in response to storm events. However, issues remain during validation (Figure 3.15b). Visually, output of this model is similar in magnitude and behavior to that of setup 5. This is likely because the structure of this “serial SDM” model roughly emulates that of the more complex SFASH model. Overall, these results are encouraging and support the claim that hydrogeological processes scale up into simpler relationships at the watershed scale and can therefore be represented in simpler models (e.g., Kirchner, 2009; Teuling et al., 2010).

Overall, these exercises and the preceding model inter-comparison show the importance of model structure and design. Groundwater models must carefully balance their process and parameter complexity based on the limitations of working in alpine regions. Adding complexity to simpler models can improve fit to data but may introduce/compensate for structural defects in the model design. Model design and improvement should be rooted in some conceptual background and subjected to reality checks during calibration and use. Future modeling work will rely on more field data collection to provide measurements and understanding to explore improvements on model designs.

### 3.7 Figures and Tables

Table 3.1 Initial parameters of each model setup. Parameter ranges were determined from measurements (see Section 2.3.6) and educated opinion.

	Parameter	Lower bound	Upper bound	Description
Setup 1	$c_1$	-10.0	10.0	$g(Q)$ coefficient
	$c_2$	-5.0	5.0	$g(Q)$ coefficient
	$c_3$	-5.0	5.0	$g(Q)$ coefficient
	$c_4$	-1.0	1.0	$g(Q)$ coefficient
Setup 2	$K$ (m/s)	0.0	0.1	Watershed hydraulic conductivity
	$S_y$	0.1	0.4	Watershed specific yield
Setup 3	$K_{top}$ (m/s)	$1.0 \times 10^{-3}$	$5.0 \times 10^{-2}$	Watershed upper layer hydraulic conductivity
	$K_{bot}$ (m/s)	$1.0 \times 10^{-7}$	$1.0 \times 10^{-3}$	Watershed lower layer hydraulic conductivity
	$S_y$	0.1	0.4	Watershed specific yield
	$z_t$	0.0	1.0	Transition depth
Setup 4	$K_m$ (m/s)	$1.1 \times 10^{-5}$	$1.1 \times 10^{-2}$	Moraine hydraulic conductivity
	$K_t$ (m/s)	$1.1 \times 10^{-4}$	$1.1 \times 10^{-2}$	Talus hydraulic conductivity
	$K_v$ (m/s)	$1.1 \times 10^{-6}$	$1.4 \times 10^{-3}$	Forest + meadow hydraulic conductivity
	$S_{y,m}$	0.1	0.4	Moraine specific yield
	$S_{y,t}$	0.1	0.4	Talus specific yield
	$S_{y,v}$	0.1	0.4	Forest + meadow specific yield
Setup 5	$K_{top,m}$ (m/s)	$4.4 \times 10^{-4}$	$5.0 \times 10^{-2}$	Moraine upper layer hydraulic conductivity
	$K_{bot,m}$ (m/s)	$1.0 \times 10^{-5}$	$4.4 \times 10^{-4}$	Moraine lower layer hydraulic conductivity
	$S_{y,m}$	0.1	0.4	Moraine specific yield
	$z_{t,m}$	0.0	1.0	Moraine transition depth
	$K_{top,t}$ (m/s)	$4.4 \times 10^{-4}$	$1.0 \times 10^{-1}$	Talus upper layer hydraulic conductivity
	$K_{bot,t}$ (m/s)	$1.0 \times 10^{-5}$	$4.4 \times 10^{-4}$	Talus lower layer hydraulic conductivity
	$S_{y,t}$	0.1	0.4	Talus specific yield
	$z_{t,t}$	0.0	1.0	Talus transition depth
	$K_{top,v}$ (m/s)	$1.0 \times 10^{-5}$	$1.0 \times 10^{-3}$	Forest + meadow upper layer hydraulic conductivity
	$K_{bot,v}$ (m/s)	$1.0 \times 10^{-7}$	$1.0 \times 10^{-5}$	Forest + meadow lower layer hydraulic conductivity
	$S_{y,v}$	0.1	0.4	Forest + meadow specific yield
	$z_{t,v}$	0.0	1.0	Forest + meadow transition depth

Table 3.2 Estimated values and 95% confidence interval bounds of parameters for all model setups.

	Parameter	Expected value	Lower bound	Upper bound	Parameter coefficient of variation	Description
Setup 1	$c_1$	-5.45	-7.84	-3.06	-	g(Q) coefficient
	$c_2$	1.52	-0.16	3.20	-	g(Q) coefficient
	$c_3$	1.04	0.55	1.54	-	g(Q) coefficient
	$c_4$	0.23	-0.05	0.52	-	g(Q) coefficient
Setup 2	$K$ (m/s)	$6.53 \times 10^{-4}$	$6.32 \times 10^{-4}$	$6.74 \times 10^{-4}$	$1.63 \times 10^{-2}$	Watershed hydraulic conductivity
	$S_y$	0.14	0.13	0.15	$3.43 \times 10^{-2}$	Watershed specific yield
Setup 3	$K_{top}$ (m/s)	$1.85 \times 10^{-3}$	$1.80 \times 10^{-3}$	$1.91 \times 10^{-3}$	$1.52 \times 10^{-2}$	Watershed upper layer hydraulic conductivity
	$K_{bot}$ (m/s)	$1.64 \times 10^{-4}$	$1.49 \times 10^{-4}$	$1.79 \times 10^{-4}$	$4.55 \times 10^{-2}$	Watershed lower layer hydraulic conductivity
	$S_y$	0.22	0.21	0.23	$1.11 \times 10^{-2}$	Watershed specific yield
	$z_t$	0.72	0.71	0.72	$5.38 \times 10^{-3}$	Transition depth
Setup 4	$K_m$ (m/s)	$6.91 \times 10^{-4}$	$6.19 \times 10^{-4}$	$7.62 \times 10^{-4}$	$5.17 \times 10^{-2}$	Moraine hydraulic conductivity
	$K_t$ (m/s)	$1.11 \times 10^{-4}$	$1.11 \times 10^{-4}$	$2.40 \times 10^{-4}$	0.57	Talus hydraulic conductivity
	$K_v$ (m/s)	$2.2 \times 10^{-4}$	$1.1 \times 10^{-6}$	$6.2 \times 10^{-4}$	0.90	Forest + meadow hydraulic conductivity
	$S_{y,m}$	0.20	0.12	0.28	0.21	Moraine specific yield
	$S_{y,t}$	0.10	0.10	0.20	0.52	Talus specific yield
	$S_{y,v}$	0.10	0.10	0.22	0.60	Forest + meadow specific yield
Setup 5	$K_{top,m}$ (m/s)	$1.17 \times 10^{-2}$	$9.63 \times 10^{-3}$	$1.38 \times 10^{-2}$	$8.79 \times 10^{-2}$	Moraine upper layer hydraulic conductivity
	$K_{bot,m}$ (m/s)	$1.14 \times 10^{-4}$	$7.84 \times 10^{-5}$	$1.49 \times 10^{-4}$	0.15	Moraine lower layer hydraulic conductivity
	$S_{y,m}$	0.12	0.10	0.18	0.23	Moraine specific yield
	$z_{t,m}$	0.54	0.53	0.55	$1.19 \times 10^{-2}$	Moraine transition depth
	$K_{top,t}$ (m/s)	$6.31 \times 10^{-4}$	$4.44 \times 10^{-4}$	$1.0 \times 10^{-1}$	39.47	Talus upper layer hydraulic conductivity
	$K_{bot,t}$ (m/s)	$1.00 \times 10^{-5}$	$1.00 \times 10^{-5}$	$2.06 \times 10^{-5}$	0.38	Talus lower layer hydraulic conductivity
	$S_{y,t}$	0.35	0.25	0.40	0.19	Talus specific yield
	$z_{t,t}$	0.00	0.00	0.41	-	Talus transition depth
	$K_{top,v}$ (m/s)	$9.49 \times 10^{-5}$	$6.09 \times 10^{-5}$	$1.29 \times 10^{-4}$	0.19	Forest + meadow upper layer hydraulic conductivity
	$K_{bot,v}$ (m/s)	$1.00 \times 10^{-7}$	$1.00 \times 10^{-7}$	$1.94 \times 10^{-6}$	2.86	Forest + meadow lower layer hydraulic conductivity
	$S_{y,v}$	0.40	0.28	0.40	0.16	Forest + meadow specific yield
$z_{t,v}$	0.44	0.34	0.54	$7.44 \times 10^{-2}$	Forest + meadow transition depth	

Table 3.3 Calculated model NSE and log-NSE for discharge and water levels (when applicable) during calibration and validation.

	Discharge				Water levels	
	Calibration		Validation		Calibration	Validation
	NSE	Log-NSE	NSE	Log-NSE	NSE	NSE
Setup 1	0.92	0.90	0.84	0.77	-	-
Setup 2	0.89	0.87	0.96	0.93	0.08	-11.87
Setup 3	0.90	0.88	0.96	0.93	0.92	-0.40
Setup 4	0.89	0.87	0.96	0.94	0.48	-8.66
Setup 5	0.98	0.93	0.79	0.25	0.96	0.88

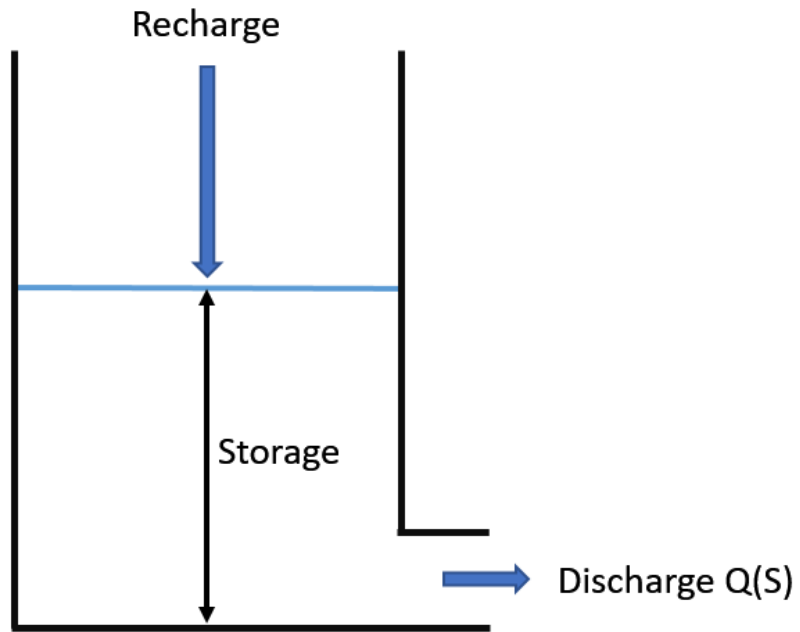


Figure 3.1 Graphic illustrating structure of the simple dynamical model reservoir.

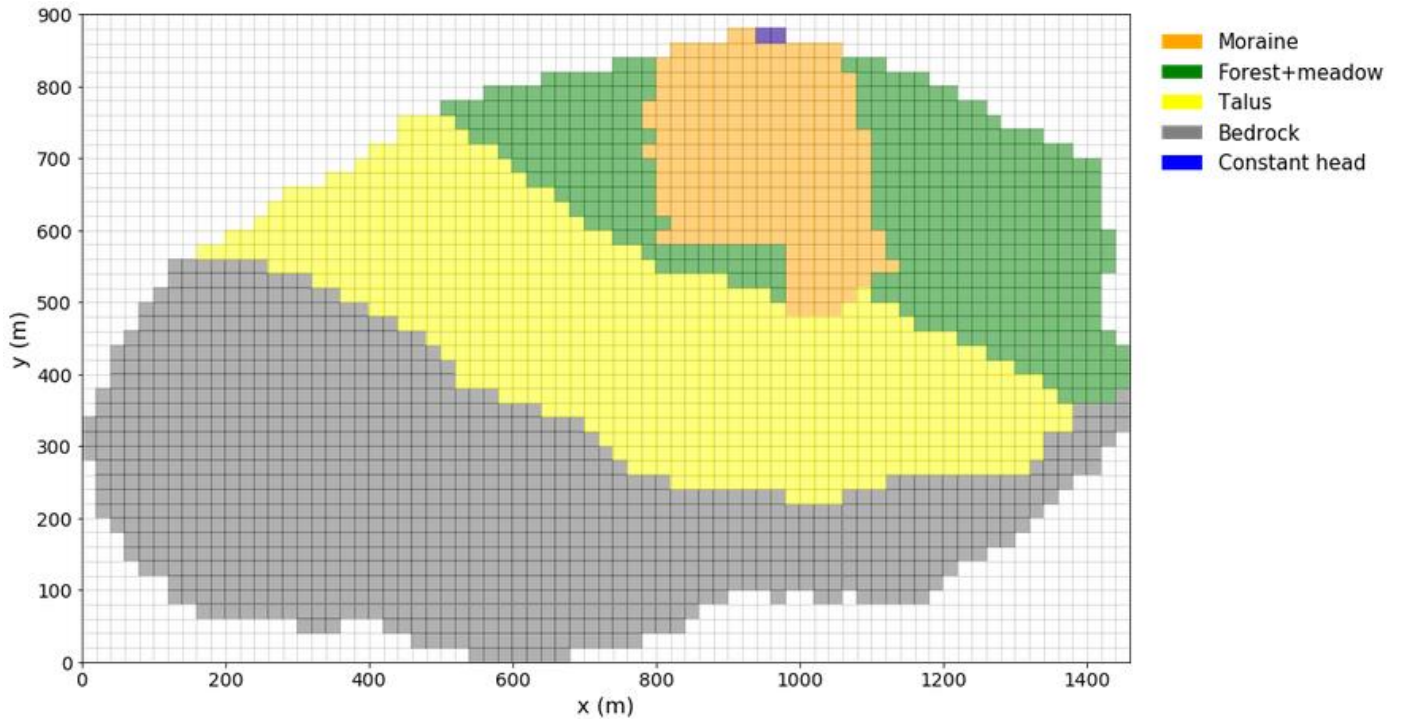


Figure 3.2 Example SFASH model domain. The watershed is represented by a structured 20 m grid. Grid cells are coloured according to discretization of land cover or boundary conditions. White cells are inactive.

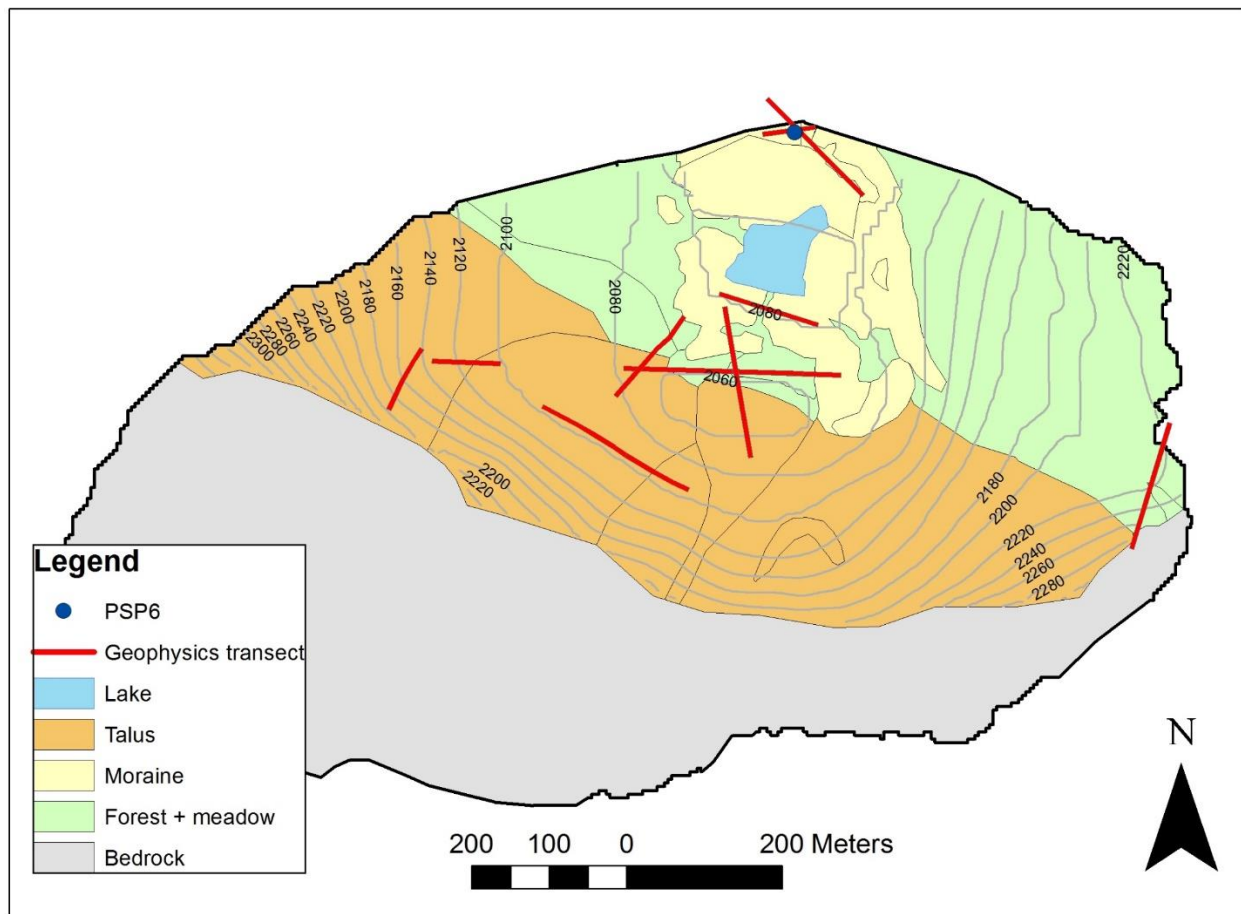


Figure 3.3 Interpolated elevation of bedrock below surficial deposits in Hathataga. Contour interval is 20 m.

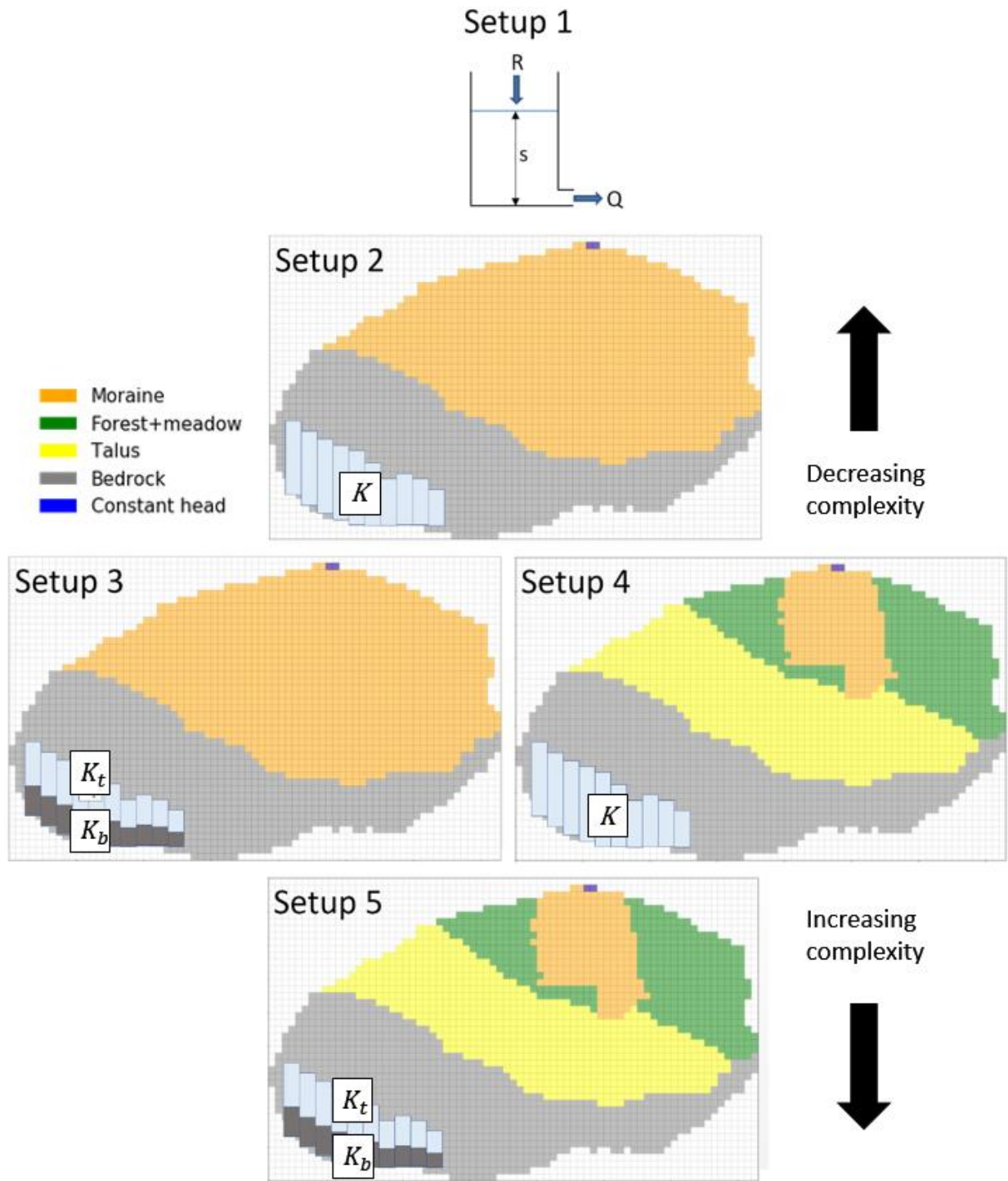


Figure 3.4 Graphic illustrating the differences between model setups. Distributed models (setups 2 to 5) are shown in plan and cross section view. Relative process and parameter complexity is expressed by position on the continuum.

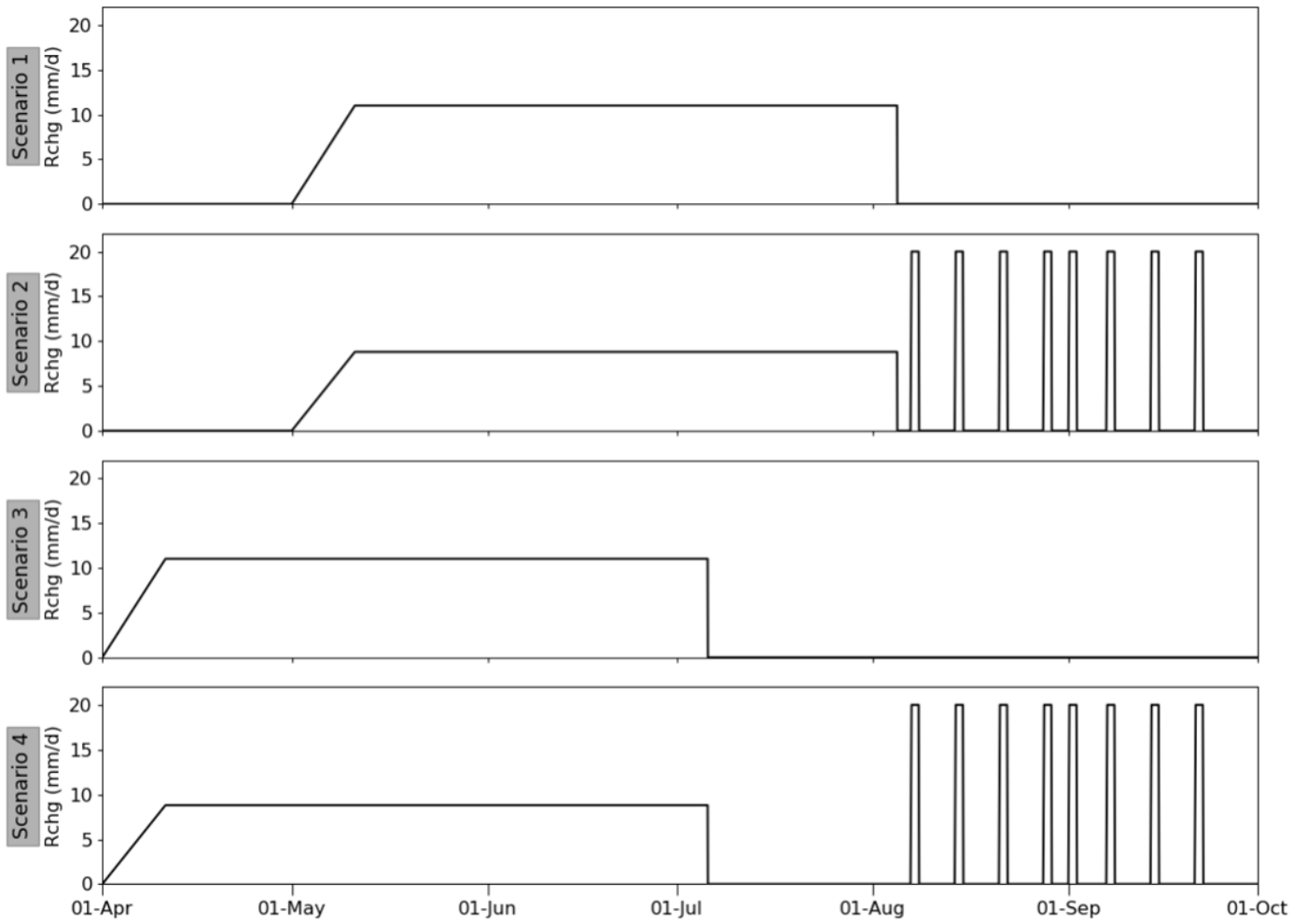


Figure 3.5 Temporal distribution of model recharge for all recharge scenarios.

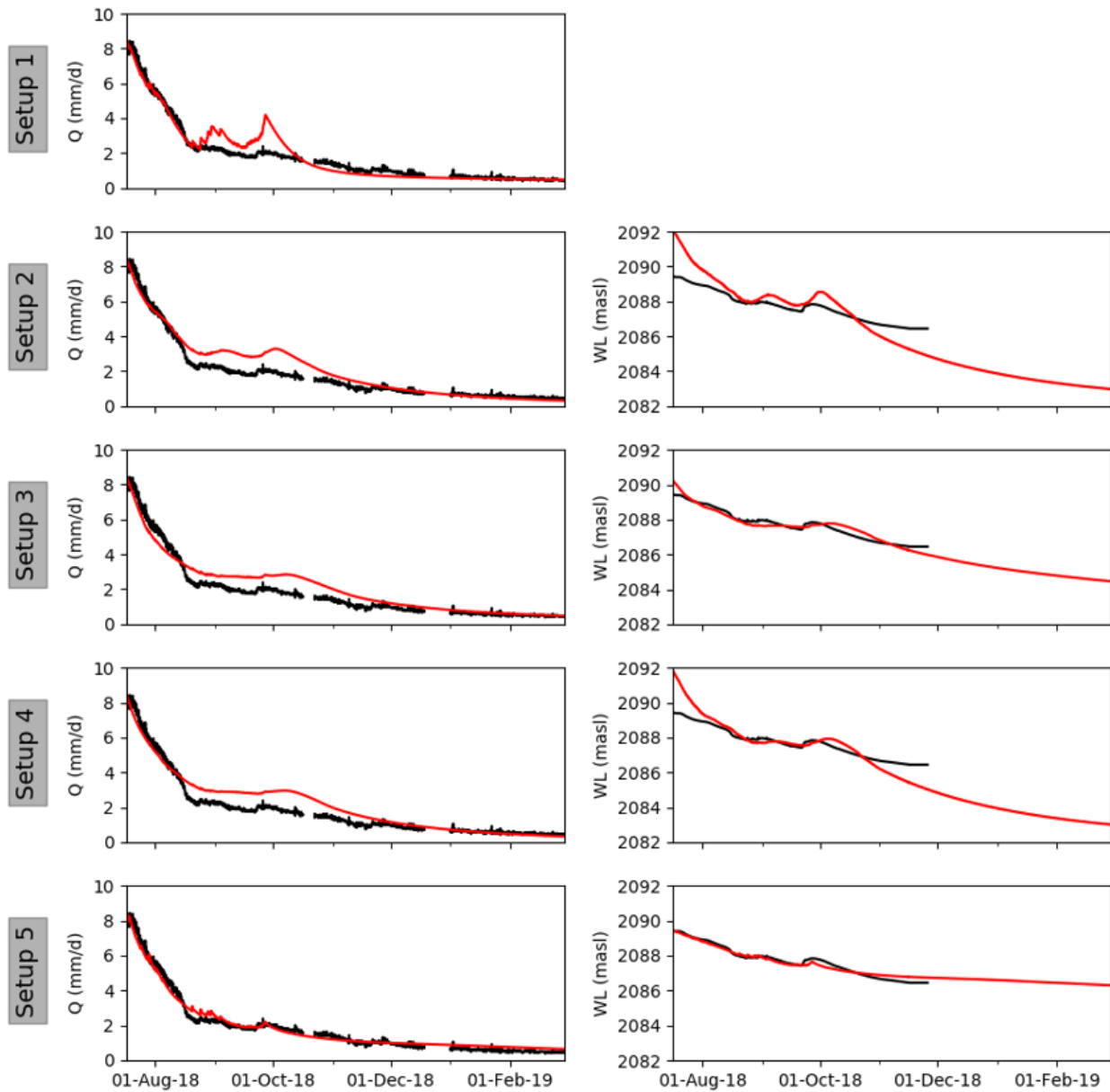


Figure 3.6 Calibration performance of all model setups. Model output is plotted in red and observations in black.

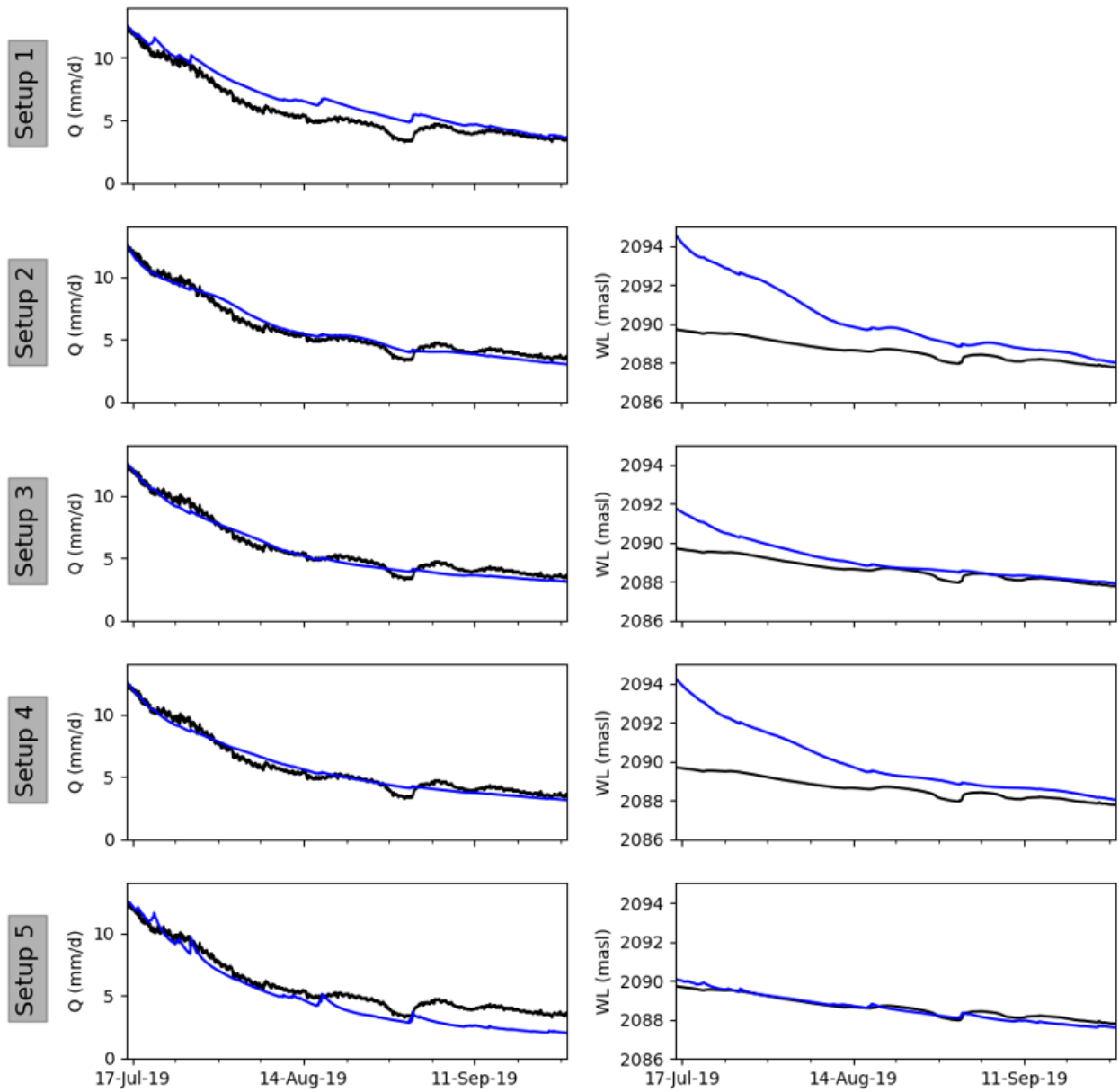


Figure 3.7 Validation performance of all model setups. Model output is plotted in blue and observations in black.

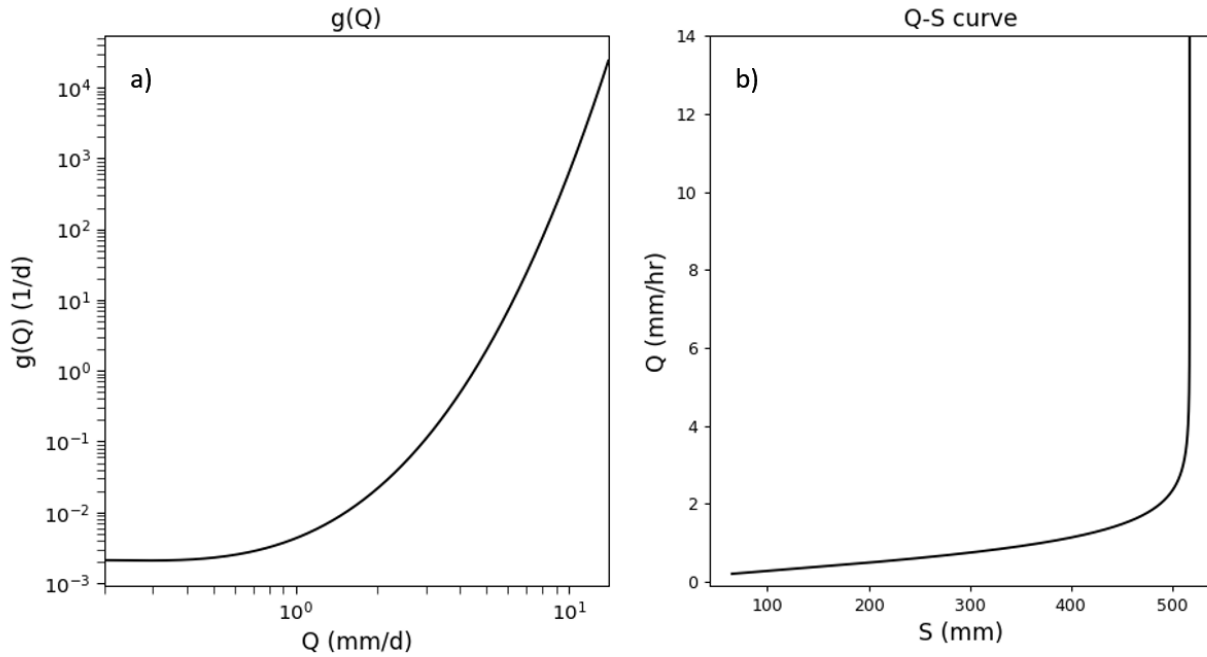


Figure 3.8 (a) Estimated discharge sensitivity function of setup 1 over the range of modeled discharge (0.2 mm/d to 15 mm/d). (b) Storage – discharge curve based on the discharge sensitivity function. Values of storage are relative because absolute reservoir storage cannot be determined.

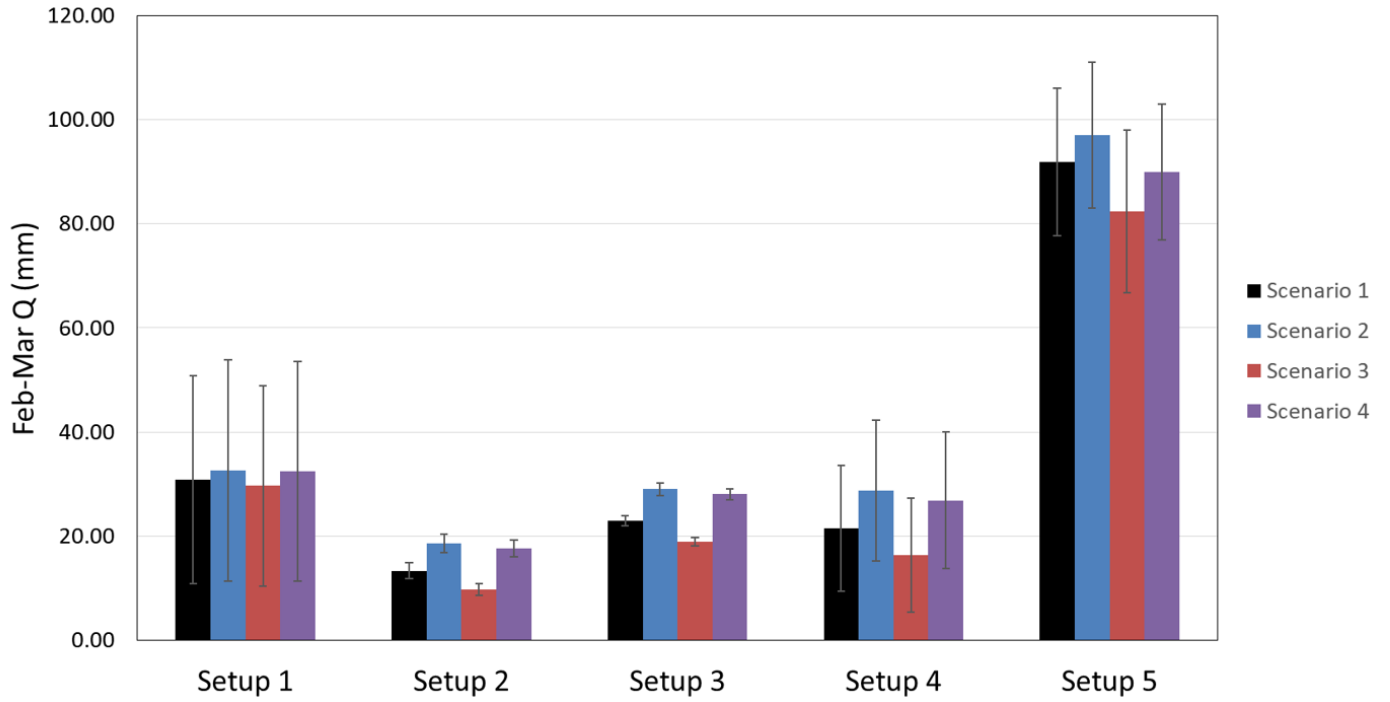


Figure 3.9 Mean simulated total February-March baseflow for all model setups and recharge scenarios. Error bars span two standard deviations above and below the mean.

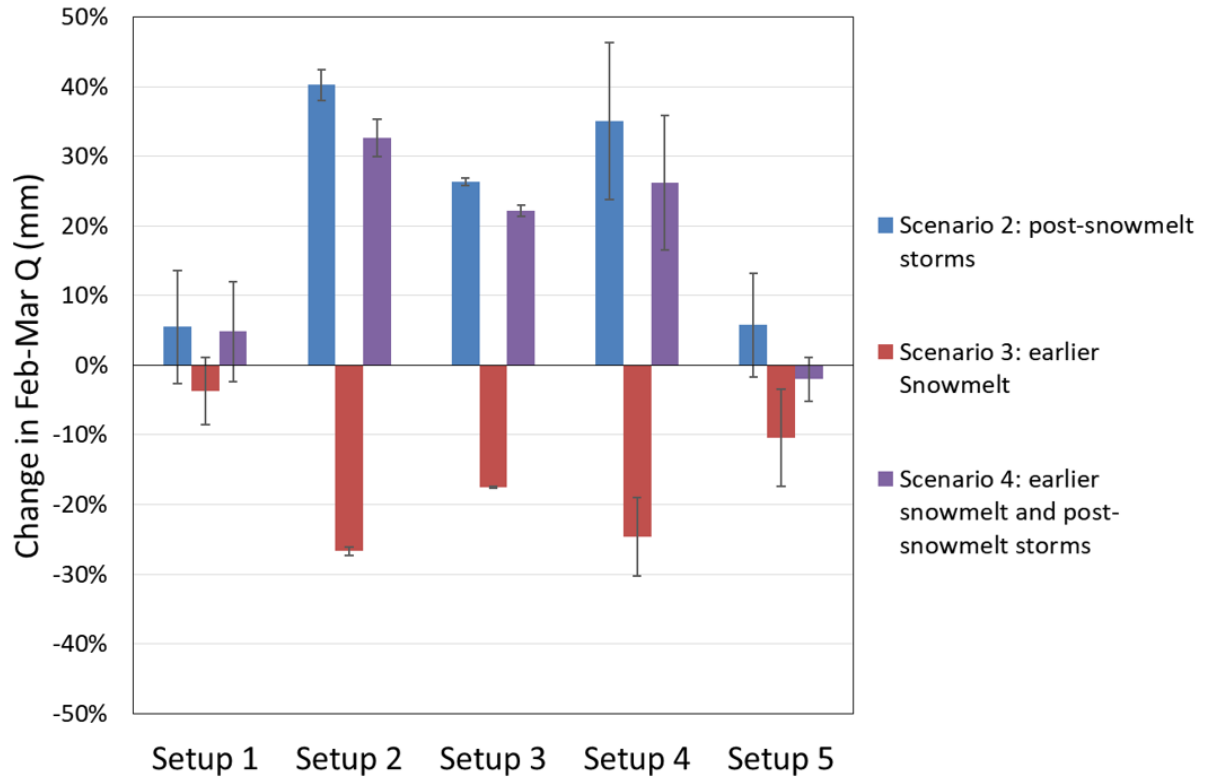


Figure 3.10 Mean difference in total February-March baseflow between recharge scenario 1 (the reference) and recharge scenarios 2 to 4. Differences are expressed as a percentage of total baseflow from scenario 1. Error bars span two standard deviations above and below the mean.

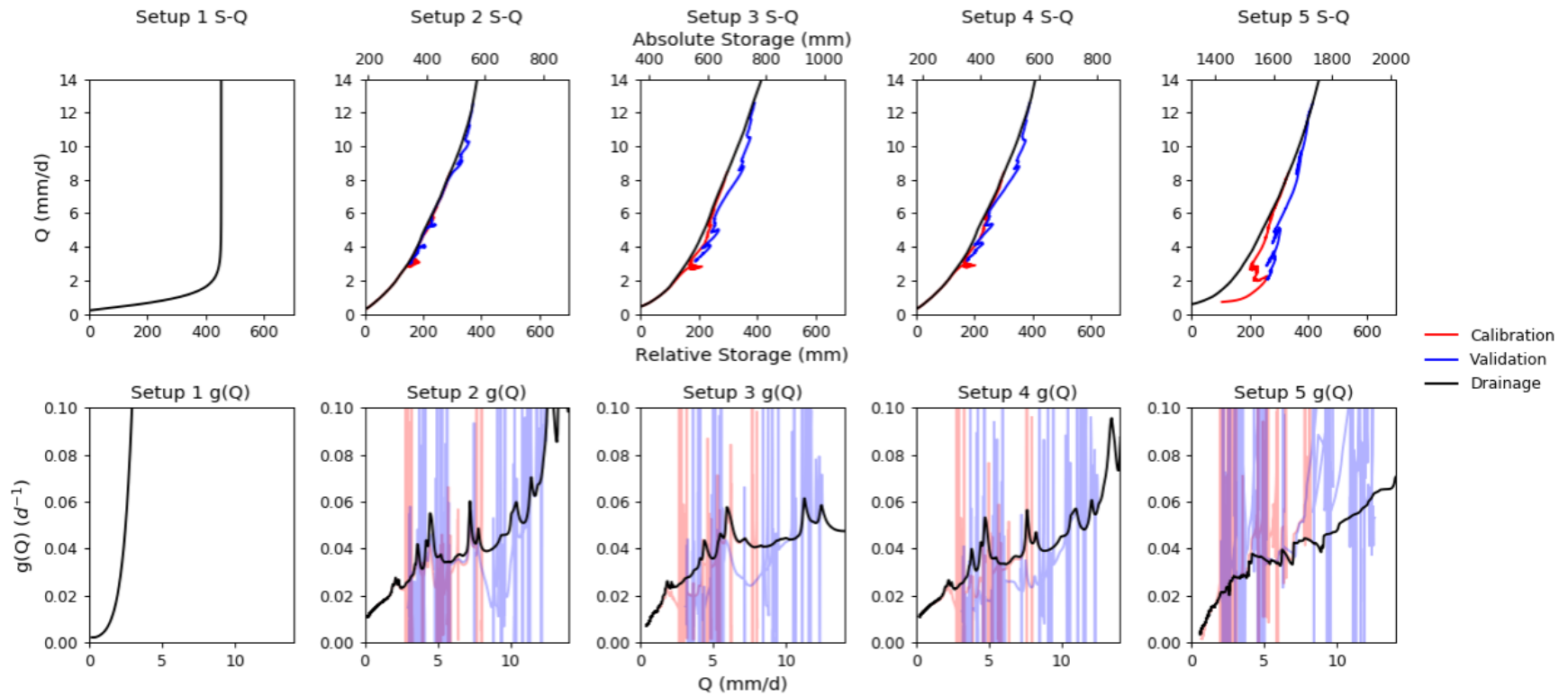


Figure 3.11 Storage-discharge and discharge sensitivity curves of all model setups. Curves from calibration, validation and pure drainage simulations are shown. Storage-discharge curves are shown in terms of relative and absolute model storage. Relative storage was plotted by setting a zero value equal to the minimum storage of the drainage simulation and shifting all curves accordingly.

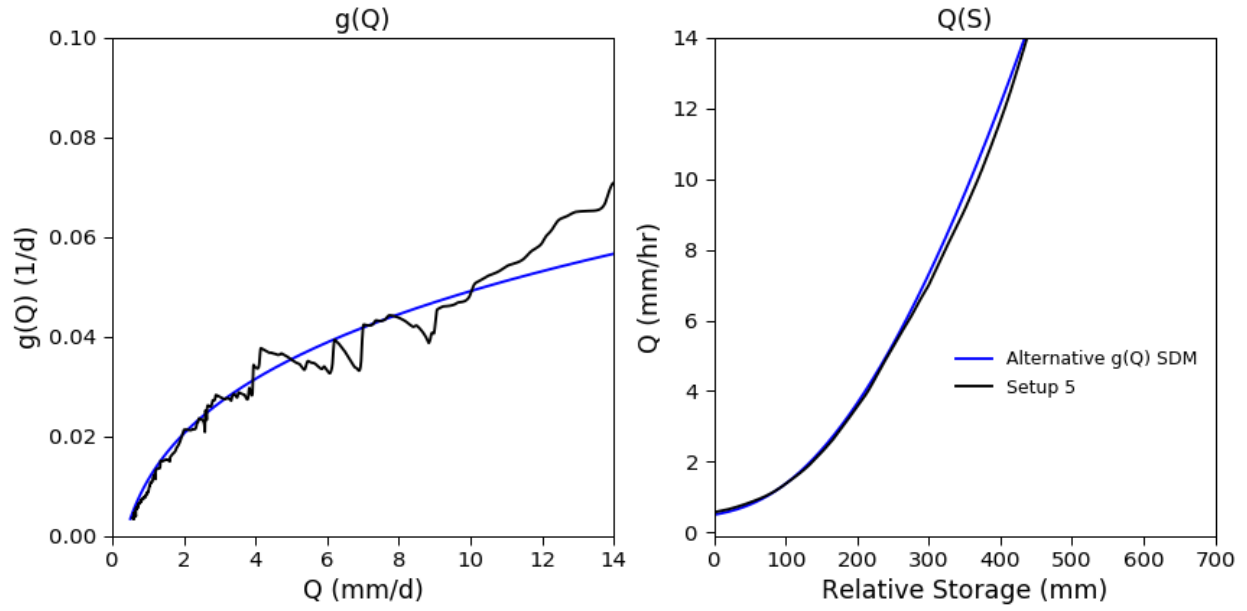


Figure 3.12 Discharge sensitivity and storage-discharge curves of the SDM using an alternative  $g(Q)$  function and SFASH setup 5.

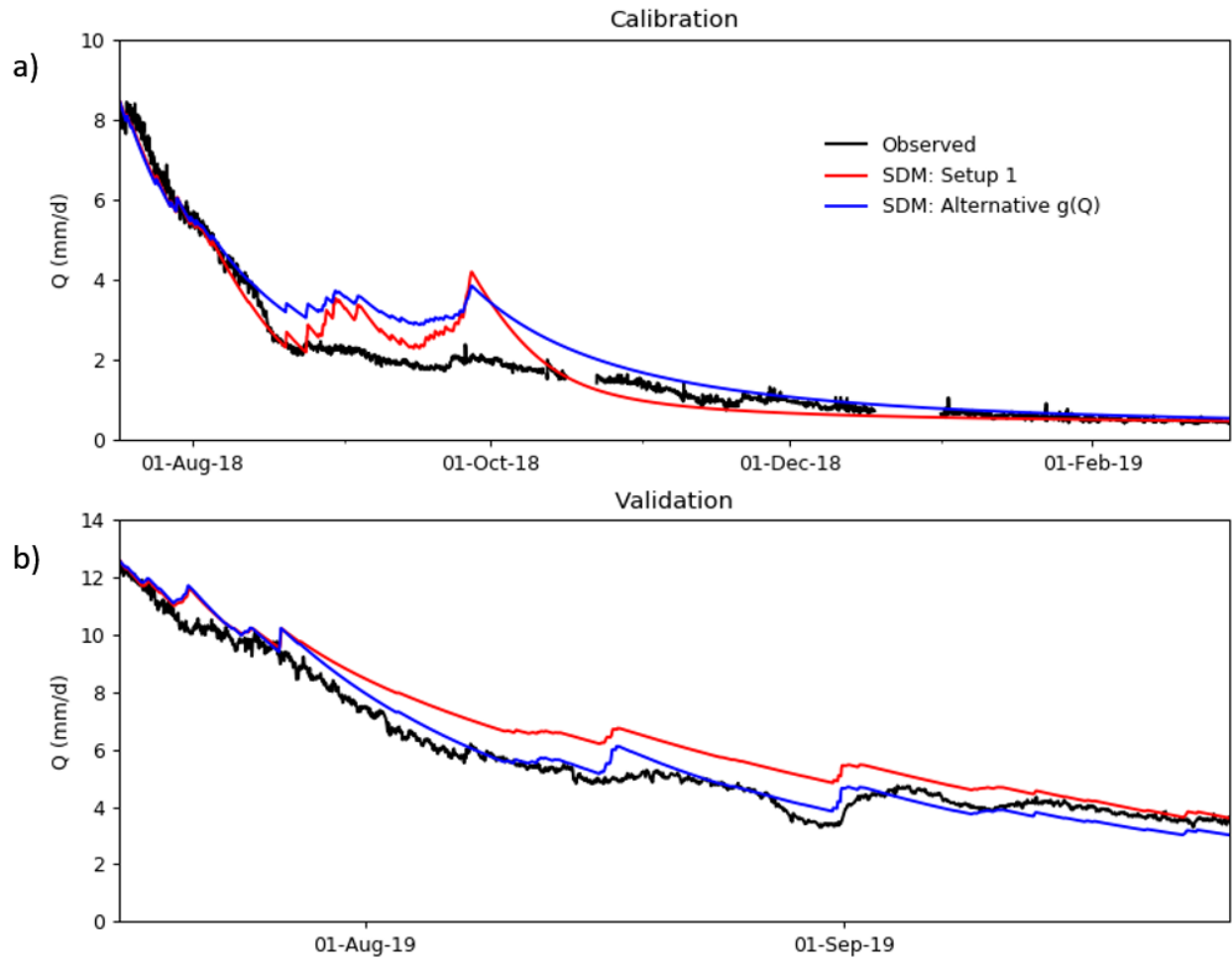


Figure 3.13 (a) Calibration and (b) validation performance of the SDM using an alternative discharge sensitivity function.

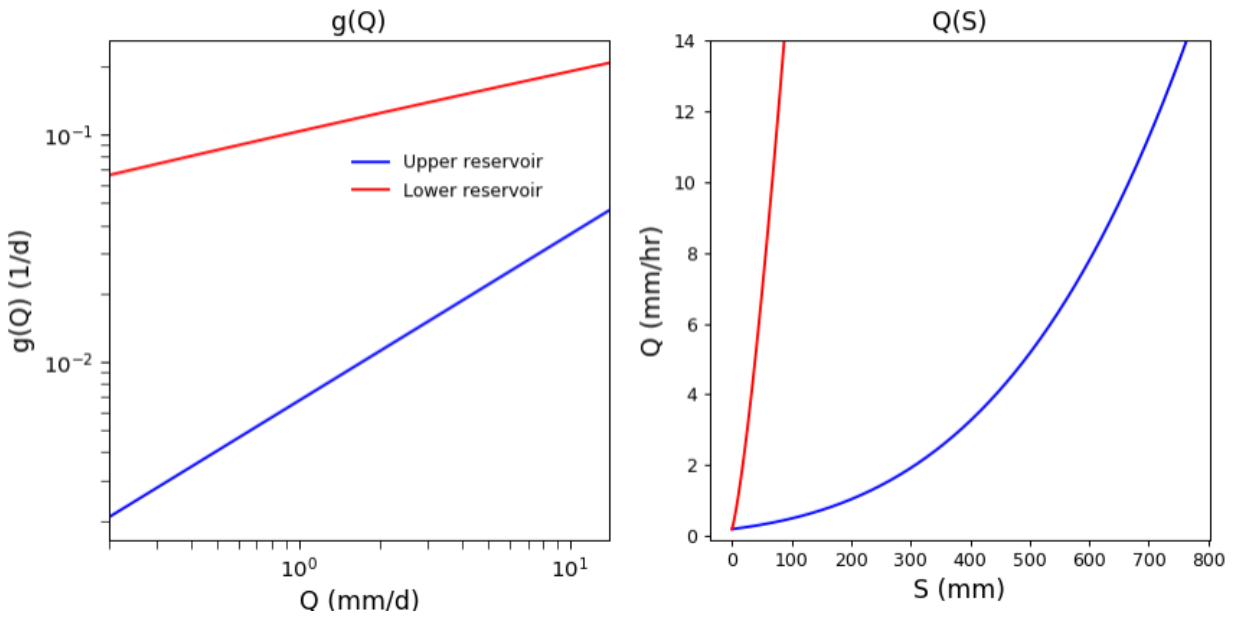


Figure 3.14 Discharge sensitivity and storage-discharge curves of two SDM upper and lower reservoirs connected in series.

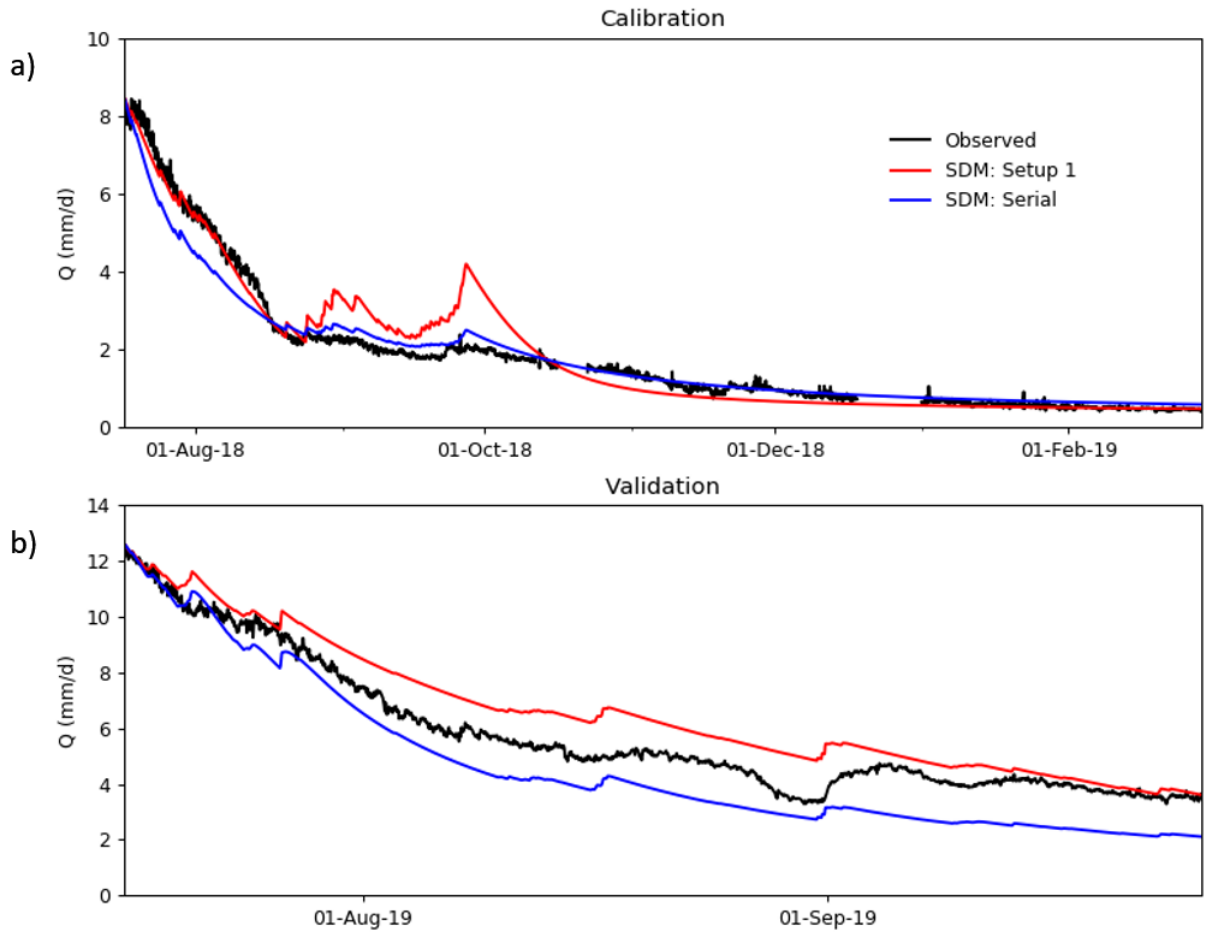


Figure 3.15 (a) Calibration and (b) validation performance of two SDM reservoirs connected in series.

#### 4. Synthesis and Conclusions

Mountains are a critical source of water for downstream communities and ecosystems. In these regions, surficial deposits (e.g., moraine, talus, alpine meadow) capture, retain and steadily release water to streams. The influence of these deposits on watershed hydrological behavior can be considerable. For example, groundwater released from storage can account for large fractions of streamflow especially during the low flow period. Unfortunately, the harsh conditions and rugged terrain of alpine regions has limited the number and scale of hydrogeological studies. To continue developing knowledge and understanding of alpine regions, there is need for collection of more data across new and diverse sites. Future work must also aim towards exploring more efficient methods to study and represent/model these regions without long and costly site investigations.

##### *Watershed gatekeepers*

Fieldwork in the Hathataga watershed has produced a large dataset of water fluxes, head, water chemistry and hydraulic properties measured and recorded from 2017 to 2019. This work has identified great heterogeneity within and between various surficial deposits, indicating the immense complexity of even just a small first-order alpine watershed. Hydrogeological properties at fine scales influence storage and release of groundwater across the watershed, but are integrated at the watershed scale and may be approximated by simpler relationships. At Hathataga, a potential mechanism for this simplification is believed to be a moraine located adjacent to the watershed outlet. This moraine is fed by other surficial deposits and functions as a “gatekeeper” which regulates flow leaving the watershed. The moraine exerts a strong influence on overall discharge from Hathataga, dampening response to recharge events and bolstering low flows.

The influence of the moraine gatekeeper remains important even at larger scales. Hathataga makes up only a small fraction of the Confluence watershed, yet can be responsible for the majority/entirety of flow from Confluence during fall and winter. During these periods, the moraine may effectively control discharge of a region spanning several square kilometers. Similar gatekeepers likely exist and have similar function in numerous other watersheds, and could collectively influence discharge from sizable portions of alpine regions. Therefore, it would be valuable to identify and study potential gatekeepers elsewhere to investigate their role in the hydrological functioning of mountain regions.

### *Design and optimization of alpine groundwater models*

The existence and function of alpine gatekeepers has implications for how Hathataga and similar sites can be efficiently characterized and represented. Future modeling efforts could save time and effort by focusing on watershed characteristics which play a greater role in controlling hydrological behavior. At Hathataga, inter-model comparison shows that the moraine is the dominant influence on watershed discharge and water levels, supporting its role as a gatekeeper. Within the moraine, vertical variation in hydraulic conductivity and flow paths is important for improving fit to observed discharge and water levels in distributed models. However, representation of other deposits/lateral heterogeneity can also improve model performance.

Simple reservoir models can also be used to simulate groundwater storage and discharge. These models represent the watershed with a storage-discharge relationship. This has the benefit of being relatively simple to parameterize as well as fast and easy to use. However, there are drawbacks which can discourage the use of these models in alpine regions. Storage-discharge relationships may be too simplistic to accurately reproduce watershed behavior. At Hathataga, it is demonstrated that storage-discharge hysteresis is important but cannot be captured by a single storage-discharge relationship. Initial attempts to use the SDM reservoir model at Hathataga were unable to achieve good performance with any single reservoir. SDM performance was only improved with additional reservoirs.

Another drawback of reservoir models may apply in steeply sloping alpine regions, where drainage is driven by gravity rather than pressure. Here, the storage-discharge relationship can begin breaking down because flow will be controlled by topographical factors more so than watershed storage - the hydraulic gradient driving flow will approach bed slope angle on steep slopes, rather than reflect differences in storage. To assess the suitability of reservoir models, the influence of steep slopes on storage and discharge in a given watershed needs to be investigated. In certain cases, the effects of topography on watershed discharge may potentially be masked away by gatekeepers or averaged out by other means.

Further work should continue to test and optimize model design by exploring different parameterizations of heterogeneity for hydraulic conductivity and other parameters. Model limitations (e.g., relating to distribution of recharge forcing and initial conditions) must also be addressed because they affect model behavior and can confound the comparison and interpretation of model results.

### Model predictions

In this work, different model setups yielded highly variable and contradictory predictions of low flow in response to various recharge scenarios. The same models largely provided adequate fits to low flows during the calibration and validation periods. Despite the removal of uncertainty related to forcing (recharge scenarios were fabricated and kept the same for all models) predictions were highly uncertain, beyond even the range indicated from linear uncertainty analysis of individual models. These results show that model predictions can be heavily influenced by decisions made during model set up and calibration. Model preparation and interpretation should be carefully considered. The consistency and precision of model predictions might be increased by calibrating models specific for use-cases. For example, to make predictions of low flows the models in this study could be recalibrated with alternate/additional objective functions which prioritize low flows with greater weight (e.g., log-NSE of discharge).

### Conclusions

The hydrogeological properties of various surficial deposits in Hathataga are highly heterogeneous. Despite this variability, hydrological behavior of the watershed appears to be dominated by a moraine gatekeeper. The gatekeeper restricts and regulates discharge from the watershed outlet, intercepting rapid flows from other areas within the watershed (e.g., talus, steep slopes). This has the effect of bolstering storage to maintain relatively high rates of baseflow later on. Based on the results from this case study, gatekeepers may have an important influence on discharge in mountain regions and merit further study. Additionally, the presence and function of the gatekeeper has important implications for future study and modeling of Hathataga and similar watersheds. By masking out and intercepting the signal from various, heterogeneous watershed aquifers, gatekeepers represent a mechanism through which small scale complexity is averaged up to the scale of watersheds. This has the potential to be well represented by more efficient and simple relationships and models. Model inter-comparison suggests that optimal model designs should include some representation of lateral and vertical heterogeneity and watershed storage-discharge hysteresis.

## References

- Alberta Environment and Parks (2014). Air Photos, RGB Ortho Southern Alberta Post Flood 2013. [online] <https://maps.alberta.ca/genesis/rest/services>.
- Anderson, M., Woessner, W., & Hunt, R. J. (2015). *Applied Groundwater Modeling*. Elsevier. <https://doi.org/10.1016/C2009-0-21563-7>
- Arnoux, M., Halloran, L. J. S., Berdat, E., & Hunkeler, D. (2020). Characterizing seasonal groundwater storage in alpine catchments using time-lapse gravimetry, water stable isotopes and water balance methods. *Hydrological Processes*, 34(22), 4319–4333. <https://doi.org/10.1002/hyp.13884>
- Bacon, C. R. (1983). Eruptive history of Mount Mazama and Crater Lake Caldera, Cascade Range, U.S.A. *Journal of Volcanology and Geothermal Research*, 18(1–4), 57–115. [https://doi.org/10.1016/0377-0273\(83\)90004-5](https://doi.org/10.1016/0377-0273(83)90004-5)
- Bales, R. C., Molotch, N. P., Painter, T. H., Dettinger, M. D., Rice, R., & Dozier, J. (2006). Mountain hydrology of the western United States. *Water Resources Research*, 42(8). <https://doi.org/10.1029/2005WR004387>
- Baraer, M., McKenzie, J., Mark, B. G., Gordon, R., Bury, J., Condom, T., Fortner, S. K. (2015). Contribution of groundwater to the outflow from ungauged glacierized catchments: a multi-site study in the tropical Cordillera Blanca, Peru. *Hydrological Processes*, 29(11), 2561–2581. <https://doi.org/10.1002/hyp.10386>
- Barnett, T. P., Adam, J. C., & Lettenmaier, D. P. (2005). Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature*, 438(7066), 303–309. <https://doi.org/10.1038/nature04141>
- Beniston, M., & Stoffel, M. (2014). Assessing the impacts of climatic change on mountain water resources. *Science of The Total Environment*, 493, 1129–1137. <https://doi.org/10.1016/j.scitotenv.2013.11.122>
- Brauer, C. C., Teuling, A. J., Torfs, P. J. J. F., & Uijlenhoet, R. (2013). Investigating storage-discharge relations in a lowland catchment using hydrograph fitting, recession analysis, and soil moisture data. *Water Resources Research*, 49(7), 4257–4264. <https://doi.org/10.1002/wrcr.20320>
- Brunner, P., & Simmons, C. T. (2012). HydroGeoSphere: a fully integrated, physically based hydrological model. *Ground Water*, 50(2), 170-176.
- Brutsaert, W., & Nieber, J. L. (1977). Regionalized drought flow hydrographs from a mature glaciated plateau. *Water Resources Research*, 13(3), 637–643. <https://doi.org/10.1029/WR013i003p00637>

- Buytaert, W., & Beven, K. (2011). Models as multiple working hypotheses: Hydrological simulation of tropical alpine wetlands. *Hydrological Processes*, 25(11), 1784–1799. <https://doi.org/10.1002/hyp.7936>
- Caballero, Y., Jomelli, V., Chevallier, P., & Ribstein, P. (2002). Hydrological characteristics of slope deposits in high tropical mountains (Cordillera Real, Bolivia). *Catena*, 47(2), 101–116. [https://doi.org/10.1016/S0341-8162\(01\)00179-5](https://doi.org/10.1016/S0341-8162(01)00179-5)
- Carpenter, T. M., & Georgakakos, K. P. (2006). Intercomparison of lumped versus distributed hydrologic model ensemble simulations on operational forecast scales. *Journal of Hydrology*, 329(1–2), 174–185. <https://doi.org/10.1016/j.jhydrol.2006.02.013>
- Christensen, C. W. (2017). A geophysical study of alpine groundwater processes and their geologic controls in the southeastern Canadian Rocky Mountains. MSc thesis, University of Calgary.
- Christensen, C. W., Hayashi, M., & Bentley, L. R. (2020). Hydrogeological characterization of an alpine aquifer system in the Canadian Rocky Mountains. *Hydrogeology Journal*, 28(5), 1871–1890. <https://doi.org/10.1007/s10040-020-02153-7>
- Clague, J., & Evans, S. (2000). A review of catastrophic drainage of moraine-dammed lakes in British Columbia. *Quaternary Science Reviews*, 19(17–18), 1763–1783. [https://doi.org/10.1016/S0277-3791\(00\)00090-1](https://doi.org/10.1016/S0277-3791(00)00090-1)
- Clark, M. P., & Kavetski, D. (2010). Ancient numerical daemons of conceptual hydrological modeling: 1. Fidelity and efficiency of time stepping schemes. *Water Resources Research*, 46(10). <https://doi.org/10.1029/2009WR008894>
- Clark, M. P., McMillan, H. K., Collins, D. B. G., Kavetski, D., & Woods, R. A. (2011). Hydrological field data from a modeller's perspective: Part 2: Process-based evaluation of model hypotheses. *Hydrological Processes*, 25(4), 523–543. <https://doi.org/10.1002/hyp.7902>
- Clark, M. P., B. Nijssen, J. D. Lundquist, D. Kavetski, D. E. Rupp, R. A. Woods, J. E. Freer, E. D. Gutmann, A. W. Wood, D. J. Gochis, R. M. Rasmussen, D. G. Tarboton, V. Mahat, G. N. Flerchinger, D. G. Marks. (2015). A unified approach for process-based hydrologic modeling: 2. Model implementation and case studies. *Water Resources Research*, 51(4), 2515–2542. <https://doi.org/10.1002/2015WR017200>
- Clow, D. W., Schrott, L., Webb, R., Campbell, D. H., Torizzo, A., & Dornblaser, M. (2003). Ground water occurrence and contributions to stream flow in an alpine catchment. Colorado Front Range. *Ground Water*, 41(9), 937–950. <https://doi.org/10.1111/j.1745-6584.2003.tb02436.x>

- Cochand, M., Christe, P., Ornstein, P., & Hunkeler, D. (2019). Groundwater Storage in High Alpine Catchments and Its Contribution to Streamflow. *Water Resources Research*, 55(4), 2613–2630. <https://doi.org/10.1029/2018WR022989>
- Crawler M, Labelle C, Mark D, Mark W, Willie K (1987) Ozade – Mmnotha Wapta Makochi, Stoney Place Names. Chiniki Band Council, Morley, Alberta, 109 pp.
- Debeer, C. M., Wheeler, H. S., Quinton, W. L., Carey, S. K., Stewart, R. E., MacKay, M. D., & Marsh, P. (2015). The Changing Cold Regions Network: Observation, diagnosis and prediction of environmental change in the Saskatchewan and Mackenzie River Basins, Canada. *Science China Earth Sciences*, 58(1), 46–60. <https://doi.org/10.1007/s11430-014-5001-6>
- Diersch, H.-J. G. (2014). FEFLOW. Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-38739-5>
- Dingman, S. L. (2002). *Physical Hydrology*. Prentice Hall Inc. New Jersey.
- Doherty, J. E. (2020). PEST\_HP PEST for Highly Parallelized Computing Environments. Watermark Numerical Computing, Brisbane, Australia.
- Doherty, J. (2018). PEST model-independent parameter estimation – User Manual (7<sup>th</sup> edition with slight additions). Watermark Numerical Computing, Brisbane, Australia, 3338, 3349.
- Doherty, J. (2015). Calibration and uncertainty analysis for complex environmental models. Brisbane, Australia: Watermark Numerical Computing.
- Donnelly, C. (2012). Investigation of groundwater flow pathways in an alpine catchment using a coupled snowmelt-groundwater flow model. MSc thesis, University of Calgary.
- Elrick, D. E., Reynolds, W. D., & Tan, K. A. (1989). Hydraulic Conductivity Measurements in the Unsaturated Zone Using Improved Well Analyses. *Groundwater Monitoring & Remediation*, 9(3), 184–193. <https://doi.org/10.1111/j.1745-6592.1989.tb01162.x>
- ESRI: ArcMap, 10.3 ed., ArcGIS, Esri, 2014.
- Fang, X., Pomeroy, J. W., Ellis, C. R., MacDonald, M. K., Debeer, C. M., & Brown, T. (2013). Multi-variable evaluation of hydrological model predictions for a headwater basin in the Canadian Rocky Mountains. *Hydrology and Earth System Sciences*, 17(4), 1635–1659. <https://doi.org/10.5194/hess-17-1635-2013>
- Field, M. S., (2008). Tracer-Test Planning Using the Efficient Hydrologic Tracer-Test Design (EHTD) Program. *Environmental Protection*, 191. <https://doi.org/14.07.2015>
- Floriatic, M. G., van Meerveld, I., Smoorenburg, M., Margreth, M., Naef, F., Kirchner, J. W., & Molnar, P. (2018). Spatio-temporal variability in contributions to low flows in the high

Alpine Poschiavino catchment. *Hydrological Processes*, 32(26), 3938–3953.  
<https://doi.org/10.1002/hyp.13302>

Freeze, R. A., and J. A. Cherry (1979), *Groundwater*. Prentice-Hall, Englewood Cliffs, NJ.

Gonfiantini R. 1981. The d-notation and the mass spectrometric measurement technique. In *Stable Isotope Hydrology: Deuterium and Oxygen-18 in the Water Cycle*, Gat JR, Gonfiantini R (eds). International Atomic Energy Agency: Vienna, Austria; 35–84, Chapter 4.

Haerberli, W., Linsbauer, A., Cochachin, A., Salazar, C., & Fischer, U. H. (2016). On the morphological characteristics of overdeepenings in high-mountain glacier beds. *Earth Surface Processes and Landforms*, 41(13), 1980–1990. <https://doi.org/10.1002/esp.3966>

Harrington, J. S., Mozil, A., Hayashi, M., & Bentley, L. R. (2018). Groundwater flow and storage processes in an inactive rock glacier. *Hydrological Processes*, 32(20), 3070–3088. <https://doi.org/10.1002/hyp.13248>

Hayashi, M. (2004). Temperature-Electrical Conductivity Relation of Water for Environmental Monitoring and Geophysical Data Inversion. *Environmental Monitoring and Assessment*, 96(1–3), 119–128. <https://doi.org/10.1023/B:EMAS.0000031719.83065.68>

Hayashi, M. (2020). *Alpine Hydrogeology: The Critical Role of Groundwater in Sourcing the Headwaters of the World*. *Groundwater*. <https://doi.org/10.1111/gwat.12965>

Hill, M. C. (2006). The Practical Use of Simplicity in Developing Ground Water Models. *Ground Water*, 44(6), 775–781. <https://doi.org/10.1111/j.1745-6584.2006.00227.x>

Hill, M. C., & Tiedeman, C. R. (2007). *Effective groundwater model calibration: with analysis of data, sensitivities, predictions, and uncertainty*. John Wiley & Sons.

Hood, J. L., (2013). *Quantifying snowmelt inputs in an alpine basin for the purpose of investigating the role of groundwater storage*. PhD thesis, University of Calgary.

Hood, J. L., & Hayashi, M. (2015). Characterization of snowmelt flux and groundwater storage in an alpine headwater basin. *Journal of Hydrology*, 521, 482–497. <https://doi.org/10.1016/j.jhydrol.2014.12.041>

Hrach, D. (2020). *Quantifying the role of shade on microclimate conditions and water use efficiency of a subalpine wetland in the Canadian Rocky Mountains, Kananaskis, Alberta*. MSc thesis, University of Waterloo.

Hrachowitz, M., Fovet, O., Ruiz, L., Euser, T., Gharari, S., Nijzink, R., Freer, J., Savenije, H., Gascuel-Oudou, C. (2014). Process consistency in models: The importance of system signatures, expert knowledge, and process complexity. *Water Resources Research*, 50(9), 7445–7469. <https://doi.org/10.1002/2014WR015484>

- Hvorslev, M. J. (1951). Time lag and soil permeability in ground-water observations (No. 36). Waterways Experiment Station, Corps of Engineers, US Army.
- Jakeman, A. J., & Hornberger, G. M. (1993). How much complexity is warranted in a rainfall-runoff model? *Water Resources Research*, 29(8), 2637–2649. <https://doi.org/10.1029/93WR00877>
- Jefferson, A., Nolin, A., Lewis, S., & Tague, C. (2008). Hydrogeologic controls on streamflow sensitivity to climate variation. *Hydrological Processes*, 22(22), 4371–4385. <https://doi.org/10.1002/hyp.7041>
- Kavetski, D., & Clark, M. P. (2011). Numerical troubles in conceptual hydrology: Approximations, absurdities and impact on hypothesis testing. *Hydrological Processes*, 25(4), 661–670. <https://doi.org/10.1002/hyp.7899>
- Kirchner, J. W. (2006). Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resources Research*, 42(3), 1–5. <https://doi.org/10.1029/2005WR004362>
- Kirchner, J. W. (2009). Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward. *Water Resources Research*, 45(2), 1–34. <https://doi.org/10.1029/2008WR006912>
- Kobierska, F., Jonas, T., Kirchner, J. W., & Bernasconi, S. M. (2015). Linking baseflow separation and groundwater storage dynamics in an alpine basin (Dammagletscher, Switzerland). *Hydrology and Earth System Sciences*, 19(8), 3681–3693. <https://doi.org/10.5194/hess-19-3681-2015>
- Kochendorfer, J., Nitu, R., Wolff, M., Mekis, E., Rasmussen, R., Baker, B., ... Brækkan, R. (2017). Analysis of single-Alter-shielded and unshielded measurements of mixed and solid precipitation from WMO-SPICE, 3525–3542.
- Kunstmann, H., Krause, J., & Mayr, S. (2006). Inverse distributed hydrological modelling of Alpine catchments. *Hydrology and Earth System Sciences*, 10(3), 395–412. <https://doi.org/10.5194/hess-10-395-2006>
- Kurylyk, B. L., & Hayashi, M. (2017). Inferring hydraulic properties of alpine aquifers from the propagation of diurnal snowmelt signals. *Water Resources Research*, 53(5), 4271–4285. <https://doi.org/10.1002/2016WR019651>
- Langston, G., Bentley, L. R., Hayashi, M., McClymont, A., & Pidlisecky, A. (2011). Internal structure and hydrological functions of an alpine proglacial moraine. *Hydrological Processes*, 2982(May), n/a-n/a. <https://doi.org/10.1002/hyp.8144>

- Langston, G., Hayashi, M., & Roy, J. W. (2013). Quantifying groundwater-surface water interactions in a proglacial moraine using heat and solute tracers. *Water Resources Research*, 49(9), 5411–5426. <https://doi.org/10.1002/wrcr.20372>
- Lauber, U., Kotyla, P., Morche, D., & Goldscheider, N. (2014). Hydrogeology of an Alpine rockfall aquifer system and its role in flood attenuation and maintaining baseflow. *Hydrology and Earth System Sciences*, 18(11), 4437–4452. <https://doi.org/10.5194/hess-18-4437-2014>
- Lindstrom, G., Johansson, B., Persson, M., Gardelin, M. and Bergstrom, S. 1997. Development and test of the distributed HBV-96 hydrological model. *Journal of Hydrology* 20: 272-288.
- Liu, F., Williams, M. W., & Caine, N. (2004). Source waters and flow paths in an alpine catchment, Colorado Front Range, United States. *Water Resources Research*, 40(9), 1–16. <https://doi.org/10.1029/2004WR003076>
- Mankin, J. S., Viviroli, D., Singh, D., Hoekstra, A. Y., & Diffenbaugh, N. S. (2015). The potential for snow to supply human water demand in the present and future. *Environmental Research Letters*, 10(11). <https://doi.org/10.1088/1748-9326/10/11/114016>
- McClymont, A. F., Hayashi, M., Bentley, L. R., Muir, D., & Ernst, E. (2010). Groundwater flow and storage within an alpine meadow-talus complex. *Hydrology and Earth System Sciences*, 14(6), 859–872. <https://doi.org/10.5194/hess-14-859-2010>
- McDonald, M. G., & Harbaugh, A. W. (1988). A modular three-dimensional finite-difference ground-water flow model. US Geological Survey.
- Mehl, S., & Hill, M. C. (2010). Grid-size dependence of Cauchy boundary conditions used to simulate stream–aquifer interactions. *Advances in Water Resources*, 33(4), 430–442. <https://doi.org/10.1016/j.advwatres.2010.01.008>
- Melsen, L. A., Teuling, A. J., van Berkum, S. W., Torfs, P. J. J. F., & Uijlenhoet, R. (2014). Catchments as simple dynamical systems: A case study on methods and data requirements for parameter identification. *Water Resources Research*, 50(7), 5577–5596. <https://doi.org/10.1002/2013WR014720>
- Moore, J. R., Boleve, A., Sanders, J. W., & Glaser, S. D. (2011). Self-potential investigation of moraine dam seepage. *Journal of Applied Geophysics*, 74(4), 277–286. <https://doi.org/10.1016/j.jappgeo.2011.06.014>
- Moore, R. D. (1993). Application of a conceptual streamflow model in a glacierized drainage basin. *Journal of Hydrology*, 150(1), 151–168. [https://doi.org/10.1016/0022-1694\(93\)90159-7](https://doi.org/10.1016/0022-1694(93)90159-7)

- Mote, P. W., Hamlet, A. F., Clark, M. P., & Lettenmaier, D. P. (2005). Declining mountain snowpack in western North America. *Bulletin of the American Meteorological Society*, 86(1), 39–50. <https://doi.org/10.1175/BAMS-86-1-39>
- Muir, D. L., Hayashi, M., & Mcclymont, A. F. (2011). Hydrological storage and transmission characteristics of an alpine talus. *Hydrological Processes*, 25(March), 2954–2966. <https://doi.org/10.1002/hyp.8060>
- Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, 10(3), 282–290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)
- Natural Resources Canada, 2016. Canadian Digital Elevation Model. <https://open.canada.ca/en>
- Pana, D.I., Elgr, R. (2013). Geology of the Alberta Rocky Mountains and Foothills. Alberta Geological Survey.
- Pauritsch, M., Wagner, T., Winkler, G., & Birk, S. (2017). Investigating groundwater flow components in an Alpine relict rock glacier (Austria) using a numerical model, 371–383. <https://doi.org/10.1007/s10040-016-1484-x>
- Paznekas, A. (2016). Groundwater Contribution to Winter Streamflows in Alpine Watersheds. MSc thesis, University of Calgary.
- Paznekas, A., & Hayashi, M. (2016). Groundwater contribution to winter streamflow in the Canadian Rockies. *Canadian Water Resources Journal*, 41(4), 484–499. <https://doi.org/10.1080/07011784.2015.1060870>
- Pomeroy, J. W., Gray, D. M., Brown, T., Hedstrom, N. R., Quinton, W. L., Granger, R. J., & Carey, S. K. (2007). The cold regions hydrological model: a platform for basing process representation and model structure on physical evidence. *Hydrological Processes*, 21(19), 2650–2667. <https://doi.org/10.1002/hyp.6787>
- Roesky, B. (2020). The Thermal Regime and Physical Characteristics of a Groundwater Fed Alpine Headwater Stream. MSc thesis, University of Calgary.
- Roy, J. W., & Hayashi, M. (2009). Multiple, distinct groundwater flow systems of a single moraine-talus feature in an alpine watershed. *Journal of Hydrology*, 373(1–2), 139–150. <https://doi.org/10.1016/j.jhydrol.2009.04.018>
- Rupp, D. E., Schmidt, J., Woods, R. A., & Bidwell, V. J. (2009). Analytical assessment and parameter estimation of a low-dimensional groundwater model. *Journal of Hydrology*, 377(1–2), 143–154. <https://doi.org/10.1016/j.jhydrol.2009.08.018>
- Selby, M. (1982). *Hillslope Materials and Processes*. Oxford: Oxford University Press.

- Simunek, J., Van Genuchten, M. T., & Sejna, M. (2006). The HYDRUS software package for simulating the two-and three-dimensional movement of water, heat, and multiple solutes in variably-saturated media. Technical manual, 1.
- Stewart, I. T., Cayan, D. R., & Dettinger, M. D. (2005). Changes toward Earlier Streamflow Timing across Western North America. *Journal of Climate*, 18(8), 1136–1155. <https://doi.org/10.1175/JCLI3321.1>
- Stumm, W., & Morgan, J. (1981). *Aquatic chemistry: An introduction emphasizing chemical equilibria in natural waters*. John Wiley.
- Tague, C., & Grant, G. E. (2004). A geological framework for interpreting the low-flow regimes of Cascade streams, Willamette River Basin, Oregon. *Water Resources Research*, 40(4), 1–9. <https://doi.org/10.1029/2003WR002629>
- Tague, C., & Grant, G. E. (2009). Groundwater dynamics mediate low-flow response to global warming in snow-dominated alpine regions. *Water Resources Research*, 45(7), 1–12. <https://doi.org/10.1029/2008WR007179>
- Tague, C., Grant, G., Farrell, M., Choate, J., & Jefferson, A. (2008). Deep groundwater mediates streamflow response to climate warming in the Oregon Cascades. *Climatic Change*, 86(1–2), 189–210. <https://doi.org/10.1007/s10584-007-9294-8>
- Taylor, C. J., & Greene, E. a. (2005). Hydrogeologic Characterization and Methods Used in the Investigation of Karst Hydrology. *Field Techniques for Estimating Water Fluxes Between Surface Water and Ground Water*, 75–114. <https://doi.org/PNR61>
- Teuling, A. J., Lehner, I., Kirchner, J. W., & Seneviratne, S. I. (2010). Catchments as simple dynamical systems: Experience from a Swiss prealpine catchment. *Water Resources Research*, 46(10), 1–15. <https://doi.org/10.1029/2009WR008777>
- Thompson, S., Kulesa, B., & Luckman, A. (2012). Integrated electrical resistivity tomography (ERT) and self-potential (SP) techniques for assessing hydrological processes within glacial lake moraine dams. *Journal of Glaciology*, 58(211), 849–858. <https://doi.org/10.3189/2012JoG11J235>
- Trefry, M. G., & Muffels, C. (2007). FEFLOW: A Finite-Element Ground Water Flow and Transport Modeling Tool. *Ground Water*, 45(5), 525–528. <https://doi.org/10.1111/j.1745-6584.2007.00358.x>
- Viviroli, D., Zappa, M., Gurtz, J., & Weingartner, R. (2009). An introduction to the hydrological modelling system PREVAH and its pre- and post-processing-tools. *Environmental Modelling and Software*, 24(10), 1209–1222. <https://doi.org/10.1016/j.envsoft.2009.04.001>

- Viviroli, D., Dürr, H. H., Messerli, B., Meybeck, M., & Weingartner, R. (2007). Mountains of the world, water towers for humanity: Typology, mapping, and global significance. *Water Resources Research*, 43(7), 1–13. <https://doi.org/10.1029/2006WR005653>
- Viviroli, D., Kummu, M., Meybeck, M., Kallio, M., & Wada, Y. (2020). Increasing dependence of lowland populations on mountain water resources. *Nature Sustainability*. <https://doi.org/10.1038/s41893-020-0559-9>
- Wagner, T., Pauritsch, M., & Winkler, G. (2016). Impact of relict rock glaciers on spring and stream flow of alpine watersheds: Examples of the Niedere Tauern Range, Eastern Alps (Austria). *Austrian Journal of Earth Sciences*, 109(1). <https://doi.org/10.17738/ajes.2016.0006>
- Watson, T. A., Doherty, J. E., & Christensen, S. (2013). Parameter and predictive outcomes of model simplification. *Water Resources Research*, 49(7), 3952–3977. <https://doi.org/10.1002/wrcr.20145>
- Weekes, A. A., Torgersen, C. E., Montgomery, D. R., Woodward, A., & Bolton, S. M. (2015). Hydrologic response to valley-scale structure in alpine headwaters. *Hydrological Processes*, 29(3), 356–372. <https://doi.org/10.1002/hyp.10141>
- Welter, D. E., White, J. T., Hunt, R. J., & Doherty, J. E. (2015). Approaches in highly parameterized inversion—PEST++ Version 3, a Parameter ESTimation and uncertainty analysis software suite optimized for large environmental models (No. 7-C12). US Geological Survey.
- White, J. T., Fienen, M. N., & Doherty, J. E. (2016). A python framework for environmental model uncertainty analysis. *Environmental Modelling and Software*, 85, 217–228. <https://doi.org/10.1016/j.envsoft.2016.08.017>
- White, S. (1981). Alpine Mass Movement Forms (Noncatastrophic): Classification, Description, and Significance. *Arctic and Alpine Research*, 13(2), 127–137.
- Winkler, G., Wagner, T., Pauritsch, M., Birk, S., Kellerer-Pirklbauer, A., Benischke, R., ... Hergarten, S. (2016). Identification and assessment of groundwater flow and storage components of the relict Schöneben Rock Glacier, Niedere Tauern Range, Eastern Alps (Austria). *Hydrogeology Journal*, 24(4), 937–953. <https://doi.org/10.1007/s10040-015-1348-9>
- Wood, E. F., Sivapalan, M., Beven, K., & Band, L. (1988). Effects of spatial variability and scale with implications to hydrologic modeling. *Journal of Hydrology*, 102(1–4), 29–47. [https://doi.org/10.1016/0022-1694\(88\)90090-X](https://doi.org/10.1016/0022-1694(88)90090-X)
- Woodruff, J. F., & Hewlett, J. D. (1970). Predicting and Mapping the Average Hydrologic Response for the Eastern United States. *Water Resources Research*, 6(5), 1312–1326. <https://doi.org/10.1029/WR006i005p01312>

- Wright, N., Hayashi, M., & Quinton, W. L. (2009). Spatial and temporal variations in active layer thawing and their implication on runoff generation in peat-covered permafrost terrain, 45, 1–13. <https://doi.org/10.1029/2008WR006880>
- Yeh, W. W. (1986). Review of Parameter Identification Procedures in Groundwater Hydrology: The Inverse Problem. *Water Resources Research*, 22(2), 95–108. <https://doi.org/10.1029/WR022i002p00095>
- Yeh, W. W.-G., & Yoon, Y. S. (1981). Aquifer parameter identification with optimum dimension in parameterization. *Water Resources Research*, 17(3), 664–672. <https://doi.org/10.1029/WR017i003p00664>