Utility of Stable Isotope Tracer Data in Large Scale Hydrologic Modeling

By

Tegan Linnea Holmes

Thesis submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

Department of Civil Engineering

University of Manitoba

Winnipeg, Manitoba, Canada

Abstract

Hydrologic assessment depends on models in data-sparse regions; however, it is unclear if such models are reliably accurate, or if internal process simulations are reasonable representations of watershed function. This can be problematic for long-term, climate-driven impact assessments that rely on accurate simulation of land-based water storage and loss. Standard model evaluation and calibration approaches focus on the accurate simulation of streamflow, disregarding internal process (storage and flow path) simulation fidelity. Stable isotope tracers are supplementary data sources that can provide additional information on water sources and hydrological processes in remote or large-scale watersheds. Hydrologic models capable of simulating both flows and isotope concentrations can be evaluated against measured isotope data as well as flow data, adding new information to the model calibration and potentially leading to improvements in hydrologic process and flow path representation. A more rigorous evaluation than the previous research case-studies and small-scale basin modeling efforts is needed to provide guidance methods of integration of isotope tracer data into operation-scale hydrologic models.

The aim of this research is to establish guidance for the incorporation of isotope tracer data into hydrologic model calibration in order to maximize the benefits of isotope-enabled simulations for large scale watershed models. The specific objectives of this thesis are to identify parameter sensitivities and parameter value identifiability for isotope tracer simulations, evaluate differences in simulated process fluxes, internal storage and streamflow between models calibrated with and without isotope tracer data, and develop specific recommendations for tracer-aided calibration objectives and isotope simulation performance metrics to maximize flux simulation benefits.

Isotope tracer simulations are more sensitive to parameters relating to soil water fluxes and storages than streamflow simulations, but calibrating isotope simulations does not significantly improve individual parameter identifiability. However, isotope-aided calibration improved process and streamflow component identifiability, with some modest benefits to streamflow simulation. Multi-objective optimization using an isotope simulation performance metric which includes timing error as a secondary calibration objective is recommended in order to maximize the benefits of isotope-aided hydrologic model calibration for large-scale watershed models.

Acknowledgements

First and foremost, I would like to thank my advisor, Dr. Trish Stadnyk, for her support and guidance throughout my time at the University of Manitoba and University of Calgary.

I gratefully acknowledge the assistance of my committee in refining and improving this research: Dr John Gibson, Dr Genevieve Ali, Dr Karen Dow, and Dr Masoud Asadzadeh, who became my co-advisor along the way.

This research was funded by the Natural Sciences and Engineering Research Council of Canada, Global Water Futures and the University of Manitoba, and would not have been possible without data shared by InnoTech Alberta, Manitoba Hydro and the Water Survey of Canada. Thank you to everyone who went forth and collected water samples, allowing me work on computer models unbothered by wildlife.

Thank you to my fellow students and researchers in Water Resources, past and current, both at the University of Manitoba and University of Calgary, you made this journey immeasurably better. For their assistance with this research, a few deserve special acknowledgment: Carly Delavau, for her work on isotopic precipitation inputs; Su Jin Kim, for helping with the initial calibration testing project and providing future climate input files; Andrew Tefs, for information on soils in the Athabasca; and Scott Pokorny, for advice on statistical evaluations, particularly of model ensembles.

I would also like to thank the Odei River (Treaty 5) and Athabasca River (Treaties 6 and 8) who acted as my research sites, and the Nelson, Winnipeg and Churchill Rivers who powered the research.

Finally, a heartfelt thank you to my parents, for their unstinting love and support.

Table of Contents

List of Tables

List of Figures

Contributions of Authors

The manuscripts listed below, which are part of this thesis, were aided by the contributions of multiple co-authors. The results, analysis and discussion are chiefly my own work.

Chapter 3

Dr Tricia Stadnyk provided some guidance with interpretation of the modelling results for this manuscript and helped shape the presentation and discussion of the results. Su Jin Kim created some required forcing data files for isoWATFLOOD modeling and assisted in extracting and processing OSTRICH and isoWATFLOOD output data. Dr Masoud Asadzadeh provided guidance and expertise on analysis and discussion of modelling results, particularly on modeling and parameter uncertainty. All authors contributed to editing the manuscript prior to publication.

Chapter 4

Dr Tricia Stadnyk provided direction on research scope, presentation of results and expertise on the hydrologic modelling results for this manuscript. Dr Masoud Asadzadeh helped shape the discussion of the results and aided in the interpretation of the sensitivity analysis results. Dr John Gibson supplied access to isotope sampling data for the Athabasca basin and provided guidance on framing the study. All authors contributed to editing the manuscript prior to publication.

Chapter 5

Dr Tricia Stadnyk contributed to the framing, research scope, review and editing of the manuscript text. Dr Masoud Asadzadeh aided in the interpretation of model results and was involved in providing feedback for manuscript drafts. Dr John Gibson supplied access to isotope sampling data for the Athabasca basin.

Use of Copyrighted Material

Holmes, T., Stadnyk, T.A., Kim, S.J., Asadzadeh, M., 2020. Regional Calibration With Isotope Tracers Using a Spatially Distributed Model: A Comparison of Methods. Water Resources Research 56. https://doi.org/10.1029/2020WR027447

Holmes, T.L., Stadnyk, T.A., Asadzadeh, M., Gibson, J.J., 2022. Variability in flow and tracer-based performance metric sensitivities reveal regional differences in dominant hydrological processes across the Athabasca River basin. Journal of Hydrology: Regional Studies 41, 101088. https://doi.org/10.1016/j.ejrh.2022.101088

Gibson, J.J., Holmes, T., Stadnyk, T.A., Birks, S.J., Eby, P., Pietroniro, A., 2020. 18O and 2H in streamflow across Canada. Journal of Hydrology: Regional Studies. https://doi.org/10.1016/j.ejrh.2020.100754

1. Introduction

1.1. Research motivation

Models are broadly applied tools in the hydrologic sciences; these mathematical representations of real-world hydrologic systems are intended to reproduce or predict streamflow and are used for a multitude of purposes, from reservoir operations, to researching watershed functioning, to assessing future flood risks. Large-scale hydrologic models can be used to simulate flows based on process representations for vast areas and extensive time-periods, in order to estimate future hydroelectric potential or generate inputs for ecosystem and oceanographic studies (MacDonald et al., 2018; Stadnyk et al., 2020). However, the accuracy of these large-scale models is generally only evaluated against streamflow data, and even these data are rare, with limited record lengths, particularly in northern Canada (Coulibaly et al., 2013; Mishra & Coulibaly, 2009). Data limitations increase the uncertainty in, but also the need for, hydrologic modeling in such high latitude regions. In order to reliably simulate flows in ungauged locations or predict flows under different climatic conditions than the present or recent past, hydrologic models must accurately represent the physical processes generating streamflow (Fatichi et al., 2016). Process or flow path accuracy cannot be determined from streamflow performance data alone, however, as many combinations of process contributions result in the same total streamflow (Beven, 2006). Additional measured data, such as snowpack volumes, groundwater heads, water quality measurements and tracer data, can bring new information into model calibration and the model performance in comparison to these observations may be important in determining if a model is adequate for its intended use (Kirchner, 2006). Unfortunately, information on individual hydrologic processes is even less common than weather or hydrometric data; remoteness and limited accessibility limits the expansion of such data networks.

Stable isotope tracers are supplementary data sources that are particularly well suited to improving our understanding of hydrology in remote or large-scale watersheds. Stable isotopes in water are naturally occurring, non-reactive, and can provide additional information on water sources and hydrological processes due to their variable occurrence in precipitation and evaporating water bodies (Bowen et al., 2019). Hydrologic models capable of simulating both flows and isotope concentrations can be evaluated against measured isotope data as well as flow data, adding new information to the model calibration and potentially leading to improvements in

hydrologic process and flow path representation. A few models have already combined isotope and flow simulations, such as the isoWATLOOD model (Stadnyk et al., 2013; Stadnyk & Holmes, 2020), or the STARR model (van Huijgevoort et al., 2016). To date, research has been focused on a limited number of research catchments where isotope sampling campaigns have been conducted; these basins are typically small-scale and intensively monitored.

Isotope data availability in Canada has recently expanded due to a national water sampling program for stable isotopes at hydrometric gauges funded by the Water Survey of Canada, which was initiated in 2013 (Gibson et al., 2020). Water samples have been collected at 331 gauges greater than 500 km^2 in size, and the network includes waters from nearly 10,000,000 km^2, over half of Canada's area; the dataset from 2013 to 2019 is now publicly available (further information on the data set may be found in Appendix A) (Gibson, Eby, et al., 2021; Gibson et al., 2020). These data have already been used to define isotope framework lines for Canadian waters, both regionally and nationally in Gibson et al. (2020), and for evapotranspiration partitioning in Gibson, Holmes, et al. (2021).

With the increased availability of stable isotope data for Canada's rivers, research on the utility of isotope tracers to hydrologic modeling is timely. A more rigorous evaluation than the previous research case-studies and small-scale basin modeling efforts is needed to provide guidance on the benefits and best methods of integration of isotope tracer data into operation-scale hydrologic models.

1.2. Research objectives and scope

The overall goal of this research is to establish guidance for incorporating isotope tracer data into hydrologic model calibration, in order to maximize the benefits of simulating isotopes in conjunction with flows in large-scale watershed models. This goal will be achieved through the following objectives:

1. Understand the sensitivity and identifiability of parameters for isotope tracer simulations and compare them to those for streamflow simulations;
2. Examine differences in simulated process fluxes, internal storage and streamflow between models calibrated with and without isotope tracer data, and;

3. Develop specific recommendations for tracer-aided calibration objectives and isotope simulation performance metrics to maximize flux simulation benefits for hydrologic models.

The scope of this research is limited to meso- to large-scale isotope tracer-aided model calibration, a hitherto under-explored research topic. The research scope is therefore constrained as follows:

- Modeling and model calibrations will be limited to hydrologic models with associated isotope tracer models which are capable of, and have been previously used for, modeling large watersheds (>100,000 km^2). Furthermore, the hydrologic model used must be process-based (i.e., neither conceptual or lumped), so that changes in process simulation can be evaluated, and both typically measured isotope tracers (δ^2H, $\delta^{18}O$) should be included in the isotope simulation such that the differences, or lack thereof, in model calibration results can be observed. The only model software currently meeting these requirements is the isoWATFLOOD^{TM} model, which will be the only hydrologic modeling software applied in this research.
- The recommendations resulting from this research are intended to be of use in the practice of large-scale modeling, therefore isotope tracer data is limited to the resolution expected for large scale models. No additional sampling was performed to produce high resolution isotope time-series, rather, pre-existing data sets with intermittent sampling similar to the Canada-wide data set were used.

1.3. Thesis outline

This thesis is comprised of six chapter, consisting of an introduction, literature review, research complied in three manuscripts, and conclusions. Chapter 1 provides a brief background of the motivation and scope for the work completed in thesis, with research objectives. Chapter 2 reviews the relevant background literature, summarizing the theory and research relating to isotope tracer-aided hydrologic modeling and model calibration.

Chapter 3 compares calibration objective formulations for tracer-aided model calibration and assesses changes in parameter values and modeled fluxes for different methods of including isotope tracer data in calibration. This work provides a recommendation for tracer-aided

calibration objectives, achieving the first half of Objective 3, and initiates work on Objectives 1 and 2; it has been peer-reviewed and published in *Water Resources Research*:

> *Regional calibration with isotope tracers using a spatially distributed model: a comparison of methods* (2020). Holmes, T., Stadnyk, T., Kim, S.J., and Asadzadeh, M., Water Resources Research. 10.1029/2020WR027447

Chapter 4 examines parameter sensitivities for flow and isotope tracer simulation performance metrics, and their temporal and spatial variations. The differences between isotope and flow performance sensitivities are evaluated, along with links between process simulations and metrics. This work achieves the first half of Objective 1, and it has been peer-reviewed and published in *Journal of Hydrology: Regional Studies*:

> *Variability in flow and tracer-based performance metric sensitivities reveal regional differences in dominant hydrological processes across the Athabasca River Basin* (2022). Holmes, T., Stadnyk, T., Asadzadeh, M., and Gibson, J.J., Journal of Hydrology Regional Studies. 10.1016/j.ejrh.2022.101088

Chapter 5 compares isotope simulation performance metrics, focusing on evaluating differences in parameter identifiability and model calibration results in order to recommend metrics for tracer-aided model calibration. This work is the final step in meeting the thesis objectives, and the manuscript is being prepared for submission in *Journal of Hydrology*.

Chapter 6 summarizes the research completed within this thesis and the major conclusions, and provides recommendations for future work.

1.4.References

Beven, K. (2006). A manifesto for the equifinality thesis. In *Journal of Hydrology* (Vol. 320, pp. 18–36). https://doi.org/10.1016/j.jhydrol.2005.07.007

Bowen, G. J., Cai, Z., Fiorella, R. P., & Putman, A. L. (2019). Isotopes in the Water Cycle: Regional-to Global-Scale Patterns and Applications. *Annual Review of Earth and Planetary Sciences*. https://doi.org/10.1146/annurev-earth-053018

Coulibaly, P., Samuel, J., Pietroniro, A., & Harvey, D. (2013). Evaluation of Canadian national hydrometric network density based on WMO 2008 standards. *Canadian Water Resources Journal*, *38*(2), 159–167. https://doi.org/10.1080/07011784.2013.787181

Fatichi, S., Vivoni, E. R., Ogden, F. L., Ivanov, V. Y., Mirus, B., Gochis, D., et al. (2016). An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. *Journal of Hydrology*, *537*, 45–60. https://doi.org/10.1016/j.jhydrol.2016.03.026

Gibson, J. J., Holmes, T., Stadnyk, T. A., Birks, S. J., Eby, P., & Pietroniro, A. (2020). 18O and 2H in streamflow across Canada. *Journal of Hydrology: Regional Studies*. Elsevier B.V. https://doi.org/10.1016/j.ejrh.2020.100754

Gibson, J. J., Eby, P., Stadnyk, T. A., Holmes, T., Birks, S. J., & Pietroniro, A. (2021). Dataset of 18O and 2H in streamflow across Canada: A national resource for tracing water sources, water balance and predictive modelling. *Data in Brief*, *34*, 106723. https://doi.org/10.1016/j.dib.2021.106723

Gibson, J. J., Holmes, T., Stadnyk, T. A., Birks, S. J., Eby, P., & Pietroniro, A. (2021). Isotopic constraints on water balance and evapotranspiration partitioning in gauged watersheds across Canada. *Journal of Hydrology: Regional Studies*, *37*. https://doi.org/10.1016/j.ejrh.2021.100878

van Huijgevoort, M. H. J., Tetzlaff, D., Sutanudjaja, E. H., & Soulsby, C. (2016). Using high resolution tracer data to constrain water storage, flux and age estimates in a spatially distributed rainfall-runoff model. *Hydrological Processes*, *30*(25), 4761–4778. https://doi.org/10.1002/hyp.10902

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). https://doi.org/10.1029/2005WR004362

MacDonald, M. K., Stadnyk, T. A., Déry, S. J., Braun, M., Gustafsson, D., Isberg, K., & Arheimer, B. (2018). Impacts of 1.5 and 2.0 °C Warming on Pan-Arctic River Discharge Into the Hudson Bay Complex Through 2070. *Geophysical Research Letters*, *45*(15), 7561–7570. https://doi.org/10.1029/2018GL079147

Mishra, A. K., & Coulibaly, P. (2009, June). Developments in hydrometric network design: A review. *Reviews of Geophysics*. https://doi.org/10.1029/2007RG000243

Stadnyk, T. A., Delavau, C., Kouwen, N., & Edwards, T. W. D. (2013). Towards hydrological model calibration and validation: simulation of stable water isotopes using the isoWATFLOOD model. *Hydrological Processes*, *27*(25), 3791–3810. https://doi.org/10.1002/hyp.9695

Stadnyk, T. A., & Holmes, T. L. (2020). On the value of isotope-enabled hydrological model calibration. *Hydrological Sciences Journal*, *65*(9), 1525–1538. https://doi.org/10.1080/02626667.2020.1751847

Stadnyk, T. A, MacDonald, M. K., Tefs, A., Déry, S. J., Koenig, K., Gustafsson, D., et al. (2020). Hydrological modeling of freshwater discharge into Hudson Bay using HYPE. *Elementa: Science of the Anthropocene*, *8*.

2. Theory and background literature

2.1. Stable isotope tracers

2.1.1. Stable isotope hydrology

Water molecules are composed of a number of stable isotopes of hydrogen and oxygen, the most common combinations of which are $^{1}H_2{}^{16}O$, $^{1}H_2{}^{18}O$ and $^{2}H^{1}H\,^{16}O$. The most abundant type of water is $^{1}H_2{}^{16}O$, while $^{1}H_2{}^{18}O$, generally called either oxygen-18 or ^{18}O, represents 0.205% of all water molecules, and $^{2}H^{1}H\,^{16}O$, called deuterium, ^{2}H or D, represents about 0.031% (Gat, 1996). While there are other stable isotopes in water, such as ^{17}O, they are not yet typically measured as they are even less common than deuterium or ^{18}O. The quantity of rare stable isotopes in a water sample is typically expressed as a relative abundance, sometimes as an atom ratio, or R, of D/H or $^{18}O/^{16}O$, but more typically as a δ value, relative to a standard average ocean water isotope abundance (such as Vienna Standard Mean Ocean Water, VSMOW2, or previously VSMOW) (Gat, 1996).

Stable isotopes in water molecules are of interest to hydrologists because small differences in molecular mass result in meaningful variations of isotopic abundance in natural waters. The various stable isotope combinations have different equilibrium vapor pressures, and differing rates of molecular diffusion in the gas phase; these variations have the largest effect during phase change in the water cycle, primarily in the change from liquid to vapor, but also from vapor to liquid and solid to liquid (Gibson and Edwards, 2002). The greater mass of the rare stable isotopes cause evaporation to occur at a slower rate, and condensation at a higher rate, which leads to predictable variations in the molecular composition of water. Evaporated water vapor, δ_E, has a lower (depleted) δ than its source (leading to negative δ values for precipitation), and the δ of water body which is losing water to evaporation will gradually increase (enrich) as lighter isotopes are preferentially removed.

Precipitation is formed from condensed atmospheric vapor, the isotopic composition of which can be described using the Rayleigh distillation equation, with a temperature dependent fractionation factor (Dansgaard, 1964; Gat, 1996). The fractionation factors for ^{18}O and ^{2}H are not identical, but both are determined by the vapor temperature, resulting in a strong correlation between the concentrations of the two isotopes in precipitation. Stable isotopes in precipitation

vary seasonally, particularly for inland regions, due to differences in air temperature and from variations in atmospheric circulation patterns importing water vapor from various sources (Aggarwal et al., 2016; Birks and Edwards, 2009; Brenčič et al., 2015; Vachon et al., 2010).

In Canada, measurement of isotopes in precipitation is sporadic, with precipitation samples being analyzed for individual research projects, and formerly at select locations as part of the Canadian Network for Isotopes in Precipitation (CNIP; 107 stations), part of the Global Network for Isotopes in Precipitation (GNIP) in North America (Birks and Gibson, 2009; Bowen et al., 2019). Due to the scarcity of measured isotope compositions for precipitation, methods have been developed to estimate compositions that can capture seasonal variability in Canadian precipitation, and as part of regional or global climate models (Delavau et al., 2015; Yoshimura, 2015; Yoshimura et al., 2008). Local meteoric water lines (LMWL) in Canada are on average close to the global meteoric water line, but precipitation compositions are depleted in heavy isotopes, particularly in snowfall and at high latitudes, as shown by the average precipitation composition for the Athabasca River basin, δ_I, in Figure 2-1 (Gibson et al., 2020).

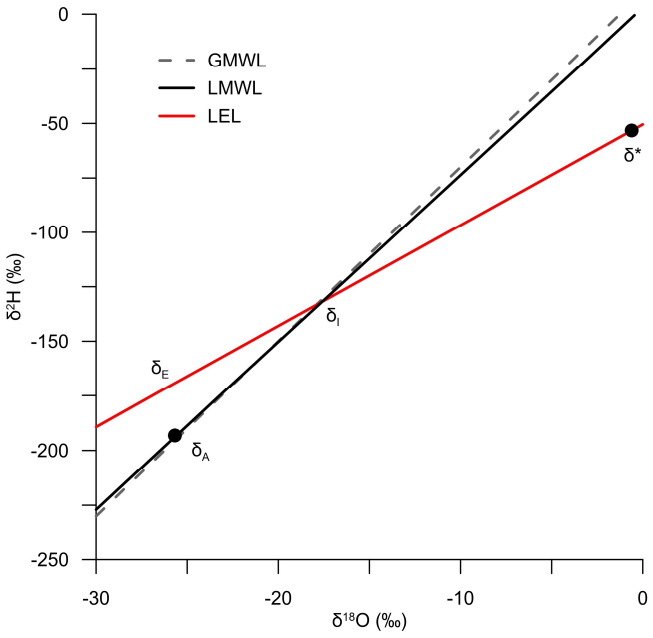

Figure 2-1: Average isotope framework for the Athabasca River basin, Alberta, Canada.

The isotopic composition of a water body undergoing evaporation changes over time: the lighter isotopes are preferentially removed, and the remaining water enriches in the rarer, heavy

isotopes, with different enrichment rates for the different stable isotopes (Yoshimura, 2015). The changes in isotopic composition due to evaporation is apparent when ^{2}H is plotted against ^{18}O in a δ-δ plot; evaporating waters separate from meteoric waters on what is termed a local evaporation line, the slope of which is determined by local conditions, an example of which for the Athabasca River basin is shown in Figure 2-1 (Gibson, et al., 2008). There are two significant components to evaporative fractionation, namely, equilibrium and kinetic separation (Gibson, et al., 2008). Equilibrium isotopic fractionation is temperature dependent, while kinetic separation of isotope species is dependent on both surface properties and relative humidity (Gibson, et al., 2008). Enrichment is limited by local atmospheric conditions, such that a desiccating water body has a limiting isotopic composition, δ^{*}, reached as the remaining water volume approaches zero (Gibson, et al., 2016).

Evaporative fractionation alters the isotopic composition of water in a basin, for all types of storage to some degree, either directly or through mixing with evaporated water. Surface water, such as stream channels, wetlands or lakes, enriches measurably in the open water season; soil water also enriches, seasonally and to a lesser degree, as only the evaporation component of evapotranspiration is fractionation (Gibson and Reid, 2010; Oshun et al., 2016; Sprenger et al., 2016). The stable isotope composition of plant stem water is complex and dependent on plant species but is closely related to soil water compositions over the longer term (Allen et al., 2019; Tetzlaff et al., 2021). Longer residence times in a catchment, from low relief, generally result in more enrichment in riparian zones or wetlands (Klaus et al., 2015b). Near surface soil water in a boreal catchment has been observed to gradually enrich, with slow mixing of enriched water with deeper soil water (Peralta-Tapia et al., 2015). The evaporative enrichment of water in stream channels depends on the balance between the channel flow rate and the evaporative flux; in small headwater streams the isotope composition can even vary over the course of a day with the evaporation rate during low flow periods (Birkel et al., 2012b). The evaporative enrichment in larger channels and lakes is likewise determined by the balance between the throughflow rate and evaporation: high throughflow reduces enrichment of the water body (Gibson et al., 2016, 1996). Enrichment in wetlands is more complex, as wetlands possess properties of both soil storage and channels, with high spatial and temporal variability. Vegetation can reduce enrichment by decreasing the contribution of evaporation in evapotranspiration for wetlands unconnected with the channel, while connected wetlands alternate between the stream isotope

composition and the soil water isotope composition based on the water level in the stream channel (Brooks et al., 2018; Hayashi et al., 2004; Klaus et al., 2015b; Zhang et al., 2018).

Soil water and groundwater both have isotope compositions varying over depth and time; deeper soil water or ground water has a more consistent composition than shallow soil water from mixing, while shallower soil zones change rapidly when new water is added (Chesnaux and Stumpp, 2018; Oshun et al., 2016; Peralta-Tapia et al., 2015; Stumpp et al., 2018). The degree of evaporative enrichment in soil water is also dependent on the vegetation type (Sprenger et al., 2017). Riparian wetlands stabilize the isotope composition of the water entering a stream, as wetlands accumulate water from all soil layers as well as direct runoff and permit more complete mixing than soil storages (Tetzlaff et al., 2014).

Basins with seasonal snowpacks have a distinct decrease in the isotopic composition of streamflow during and immediately following the snow melt period (Beria et al., 2018). The spring freshet in boreal streams consists of a mix of depleted snow meltwater and stored soil or surface water, with snowmelt adding to wetland and lake storages (Hayashi et al., 2004; St Amour et al., 2005). The relative contribution of snowmelt to the spring freshet peak is dependent on the land cover and topography in the basin: connected wetlands increase water mixing and reduce the peak snowmelt contribution (St Amour et al., 2005). Frozen soils have a reduced infiltration capacity, but snow melt does lead to depleted soil water near the surface (Peralta-Tapia et al., 2015). Partitioning of precipitation between evapotranspiration and streamflow is highly seasonal in cold regions; snowpack water is less likely than rainfall to be evapotranspired (Kirchner and Allen, 2020).

2.1.2. Modeling stable isotopes in water

Modeling stable isotopes in terrestrial waters originated in two separate scientific developments: the study of the variation of stable isotopes in global meteoric and oceanic waters, and the development of hydrograph separation techniques using tracers (Craig, 1961; Craig and Gordon, 1965; Dansgaard, 1964; Eriksson, 1963; Pinder and Jones, 1969). Once the isotopic compositions of precipitation and streamflow were known, flow separations could be performed using oxygen-18 as a tracer to identify groundwater or snowmelt contributions to channel flows (Dincer et al., 1970; Sklash et al., 1976). Groundwater was identified as a main contributor to event peak flows, rather than direct rainfall or surface runoff (Sklash et al., 1976). The

development of more complex numerical models for analysis of tracer responses (Niemi, 1977), allowed the determination of not only the contribution of groundwater to streamflow, but also mean transit times of sub-surface flows and groundwater storage volumes using isotope data (Kennedy et al., 1986; Małoszewski et al., 1983; Małoszewski and Zuber, 1982). Stable isotope tracers continue to be useful in hydrograph separations, particluarly when a three-component approach is used (Kirchner, 2019; Klaus and McDonnell, 2013).

Simulating stable isotopes in the hydrologic cycle requires modeling decisions for the mathematical conceptualization of tracer movement through the watershed. The most commonly used strategy is to add state variables to quantify either tracer concentrations within hydrological storages or tracer volumes corresponding to water volumes, and then have tracer movement between storages or into streamflow controlled by flux equations similar to those used to control the flow of water. When the tracer simulation is associated with a preexisting hydrologic model, the number of storages and the connections between them will be based on the structure of the hydrologic model. The ^{18}O simulation added to the HBV model (Bergström et al., 2002) included the same storage layers in each lumped sub-basin as the hydrologic model, with tracer concentrations for each storage used to define tracer fluxes out of each compartment. The monthly water balance model developed by the IAEA, IWBMIso, also uses tracer concentrations for all the original sub-basin and hydrologic response unit storages and the fluxes between them (Belachew et al., 2016). Total tracer volumes for each storage unit can also be used in a similar lumped iso-hydrologic model with multiple ground storages (Dunn et al., 2008, 2007). Birkel et al. (2010) modeled the tracer flux in per mille, and the deuterium response in the stream weighted averages of outputs from the upper and lower storages in the model, as does the STARR model based on this earlier work (van Huijgevoort et al., 2016). Storages may also be sub-divided based on land cover in a basin, in addition to soil depth and saturation (Smith et al., 2016). In distributed models, every water storage compartment has an associated isotope volume, with isotope fluxes determined by the isotope volumetric concentration of the originating compartment (He et al., 2019; Nan et al., 2021b; Stadnyk et al., 2013; Stadnyk and Holmes, 2020).

Tracer simulations generally assume that storage units are completely mixed over the simulation time interval, resulting in uniform tracer concentrations throughout the simulated compartment.

The time interval used in this assumption varies between models, with Birkel et al. (2010), van Huijgevoort et al. (2016) and Smith et al. (2016) using a daily time step in lumped models, while the distributed isoWATFLOOD model uses hourly time steps (Stadnyk et al., 2013), as does the isotope-enabled WASA model (He et al., 2019). Partial mixing models for tracers have been shown to produce different and less identifiable residence times for water than models assuming complete mixing (Fenicia et al., 2010). While partial mixing is believed to be more physically representative, using partial mixing in place of complete mixing adds to model complexity. If additional, separate storages are used for tracer mixing, either as passive storage or for partial mixing, additional model parameters will be required, adding uncertainty to the model (Birkel et al., 2010; Fenicia et al., 2010).

An additional modeling choice that must be made for the simulation of stable isotope tracers specifically is whether or not to include fractionation in the model. The enrichment of water due to evaporative fractionation has been well described theoretically and can be simulated using equations derived experimentation. However, simpler conceptual models may not include fractionation, despite the inclusion of evapotranspiration as a water flux (Birkel et al., 2010). For process-based models, fractionation due to evaporation may be added to simulated evapotranspiration losses from wetlands and soil water (Smith et al., 2016; Stadnyk et al., 2013). Simulating fractionation is necessary to reproduce the isotope mass balance in areas with a high prevalence of wetlands and lakes, and in basins with long retention times (Birkel et al., 2011a; Klaus et al., 2015a). Fractionation during snowmelt can be significant to cold regions models with a high temporal resolution; melting fractionation has been used in both large- and small-scale models, although it may necessitate the introduction of more model parameters (Ala-aho et al., 2017; Yoshimura et al., 2006).

There have been alternative approaches than state variable simulations developed for the modeling of tracers, such as used Markov switching autoregressive models, Bayesian mixing models, or mechanistic transfer functions (Birkel et al., 2012a; Evaristo et al., 2017; Parnell et al., 2013; Weiler et al., 2003). Simulated isotope concentrations and water ages have also been based random particle tracking in models (Davies et al., 2011; Klaus et al., 2015a). McMillan et al. (2012) modeled tracer concentration in streamflow from the convolution of the outflow age distribution with the precipitation concentrations. These methods have provided useful

information on isotope behavior, but they are not currently suitable for integration into conceptual or semi-physically based hydrological models as they are either too computationally intensive or simulate isotopes independently of the flow.

The most significant issue for modeling stable isotopes in water is the limited amount of observational data available (Birkel and Soulsby, 2015). Isotope sampling is often infrequent and sporadic, due to field season timing, and limited staff availability (Gibson et al., 2020). Insufficient data collection frequency limits the utility of stable isotope tracers for studying short duration processes, particularly snowmelt, but also daily evaporation rates (Birkel et al., 2011b; Ohlanders et al., 2013; Wang et al., 2019). However, high-resolution sampling, such as sub-daily sampling for small study basins, may be of limited benefit to hydrologic models, and weekly sampling or sampling in relation to flow conditions can be adequate for tracer -aided models (Stevenson et al., 2021; Tunaley et al., 2017). Data scarcity is also an issue for isotope model inputs such as precipitation; occasional precipitation samples cannot be expected to capture the true variability of precipitation composition, such that researchers turn to modeled data products (Delavau et al., 2017; Nan et al., 2021a; Pangle et al., 2013). The limited number of catchments with isotope data, and sampling sites within those catchments is another issue for stable isotope modeling. Isotope simulations for catchments cannot be validated without measured data and the lack of widespread sampling also prevents the verification of the simulated mixing of stream contributions (Soulsby et al., 2015a; Tetzlaff et al., 2014).

2.2. Tracer-aided hydrologic modeling

2.2.1. Hydrologic models

Models in general use mathematical equations to imitate some real-world phenomenon; hydrologic models are used to simulate streamflow and may also simulate the movement and distribution of water in a watershed. The purpose or uses of a hydrologic model can vary, from predicting flow rates in the short or long term, to the investigation of flow generation processes or the overall basin water balance. The scale, structure and complexity of hydrologic models are also variable, ranging from hillslope to global simulations.

The internal structure and complexity of hydrologic models likewise ranges along a continuum, from fully physics-based models to conceptual models. Physics-based models simulate streamflow and hydrologic processes within the watershed using equations based on the physical

properties of the area; both water and energy balances are included. Verification of internal processes, such as evaporation or groundwater flow, is both possible and desirable for physics-based models, as these processes are both simulated and intended to match real-world fluxes. The greater the physical basis of a model, the greater the required input data for setting up and running the model, and the greater the computational demand for running the model. Physics-based models are best applied in small and intensively studied research basins, where hydrologic storages and fluxes well defined (Singh and Frevert, 2010). In contrast, conceptual models do not use physics to simulate flows, and parameters have no direct or measurable real-world meaning; the accuracy of the model depends entirely on parameter calibration using observed data for the outputs of interest. Unlike physics-based models, conceptual models can be extremely simple (although not all are in practice), and they generally require less input data. However, because the parameter values in conceptual models depend on prior observations, conceptual models can be unreliable when conditions differ significantly from the original development period, or period of calibration, such as when land use changes in the basin or under differing climatic conditions (Clark et al., 2015; Kirchner, 2006). Conceptual models can be applied to any watershed, but the quality of the model will depend on the data record, particularly for streamflow, in the modeled basin. Lying between the extremes of physics-based and purely conceptual models are process-based models, which simulate internal processes in order to simulate total streamflow, but also include conceptual elements and simplifications to reduce forcing and computational demands (Fatichi et al., 2016). Some watersheds also require the simulation of human activities, such as irrigation diversions or reservoir releases, in addition to natural hydrologic processes.

Hydrologic models can also be classified along a continuum between lumped and distributed modeling systems, based on the degree to which the modeled watershed response is aggregated. A fully lumped hydrologic model aggregates the entire simulated watershed into a single response unit with averaged parameters, while a distributed model calculates a hydrologic response from each piece of a finely divided basin area to find the total response of the watershed (Singh and Frevert, 2010). The majority of hydrologic models lie between these two extremes and are either semi-lumped or semi-distributed, with the watershed area divided into sub-basins based on topography or into hydrologic or grouped response units based on expected hydrologic properties such as land cover or soil type. Disaggregating the modeled watershed area

increases the complexity of the model, and the data and computational power required to run the model but also increases the degrees of freedom in the model which can improve the model performance. Distributed models can also allow parameter values to be transferred between watersheds because they are related to a specific land cover or soil type; this transferability is desirable for ungauged watersheds or modeling in remote regions where input and observed data are scarce (Fatichi et al., 2016).

The use of models is an expected part of the study of hydrology, but their general acceptance does not mean that hydrologic models are issue-free, or that efforts to improve them are at an end. Research is focused on improving the reliability of models, either through improving the accuracy of simulations or quantifying the uncertainty of the output (Refsgaard et al., 2022). Some of the methods for improving the simulation performance include using new types or sources of data, such as tracers or remotely sensed data products, experimenting with internal model structure, and applying new model calibration methods to find better parameter values or value ranges (Chlumsky et al., 2021; Fatichi et al., 2016; Guse et al., 2021).

2.2.2. Stable isotope tracers in hydrologic modeling

Stable isotope data has already been put to diverse uses in hydrologic modeling, from traditional isotope-based hydrograph separations (IHS) to model structure and conceptualization analysis, to evaluation of water mixing, storage, and streamflow sources. Modeling the isotopic composition of streamflow and groundwater can provide an estimate of water storage, water age and storage variability within a catchment, based on the modeled water volume required to reproduce the damped isotope response in stream water (Birkel et al., 2020, 2015, 2011a). Conceptualizations of flow pathways from storage compartments to the stream can also be improved using stable isotope data in conjunction with hydrologic simulation. For example, (Bergström et al., 2002) found that modeling by-pass flow into the upper saturated zone improved isotope tracer simulation. Modeling isotopes together with flows indicates that saturated riparian zones or wetlands connected to the main channel should be modeled as mixing storages for upper soil and lower groundwater outflows (Birkel et al., 2011c; Birkel and Soulsby, 2016; Tunaley et al., 2017). Similarly, stable isotopes can be used to identify flow sources within a catchment, for example, relative contributions of soil water, ground water and wetland storage to streamflow in a multi-year simulation, glacier melt contributions, or hillslope and peatland storage

contributions to stormflow peak flows (Dehaspe et al., 2018; Nan et al., 2021b; Piovano et al., 2019; Smith et al., 2016; Soulsby et al., 2015b). Verifying water balance components has also been done at the large scale using the IWBMIso monthly water balance model (Belachew et al., 2016). Isotope data can likewise be used to verify the simulated water ages, or the degree of connectivity (Neill et al., 2019; Piovano et al., 2018). The collection of stable isotope composition data for water other than streamflow, such as groundwater, soil water or wetlands, permits the verification of simulated internal storages. Birkel et al. (2014) successfully used soil water isotope data to improve parameter identifiability for storage parameters, which lead to more accurate streamflow and isotope simulations, while He et al. (2019) found a similar improvement in parameter identifiability using a different hydrologic model.

Streamflow isotope data contains useful information about hydrologic function in a watershed, related to, but independent of, total streamflow rates; there have therefore been multiple efforts to incorporate isotope data into hydrologic model calibration. Isotope tracer data can be used to estimate hydrologic function independently of the model, and the resulting process contribution estimates can be used to constrain model calibrations (Jafari et al., 2021). Alternatively, isotope tracer concentrations can be simulated based on the hydrologic model fluxes and storages. Multiple metrics have been used to quantify the isotope simulation performance, typically with a residual error metric such as the root mean squared error (RMSE) (He et al., 2019; Iorgulescu et al., 2007), normalized RMSE (Smith et al., 2016; Stadnyk and Holmes, 2020) or mean absolute error (Nan et al., 2021b; Piovano et al., 2020, 2018). With higher frequency isotope composition sampling efficiency metrics may be used instead, previously with the Nash-Sutcliffe efficiency (Bergström et al., 2002), but more recently with the Kling-Gupta efficiency (Delavau et al., 2017; He et al., 2019; Tunaley et al., 2017). Isotope performance metrics can be introduced into a traditional automatic calibration by creating a single objective function combining both flow and isotope errors (Bergström et al., 2002; Stadnyk & Holmes, 2020). Alternatively, the model can be calibrated using a Monte Carlo approach, which allows for multiple criteria to be applied in selecting parameter sets, including both flow and isotope errors independently (Delavau et al., 2017; Neill et al., 2019; Piovano et al., 2020, 2019; Smith et al., 2016). There have also been recent studies which used the multi-objective calibration algorithms to calibrate both flow and isotopes simultaneously, with some success (Birkel and Soulsby, 2016; He et al., 2019; Nan et al., 2021b; Tunaley et al., 2017).

2.2.3. The isoWATFLOOD model

The isoWATFLOOD model is the first large-scale isotope-enabled hydrologic model capable of simulating tracer concentrations at a daily or sub-daily temporal resolution, and the only isotope-enabled distributed model functioning in near-continental scale watersheds such as the Mackenzie River basin at the time this research was initiated. The isoWATFLOOD/CHARM model is an open source, distributed model with a mixture of conceptual and physically based process representation (Kouwen, 2018). This model divides the watershed area into grid cells, and then sub-divides the cells into grouped response units (GRU) based on the land cover or soil type. Soil-based GRU are sub-divided vertically into three layers, namely the surface, upper zone and lower zone. The model is best suited to continuous, multi-year simulations, but run times are relatively short, as it was designed for mesoscale watershed simulations. The isotope model in isoWATFLOOD simulates the isotope volume and concentration for all the storages and fluxes used in the hydrologic model, as shown in Figure 2-2 (Holmes, 2016a).

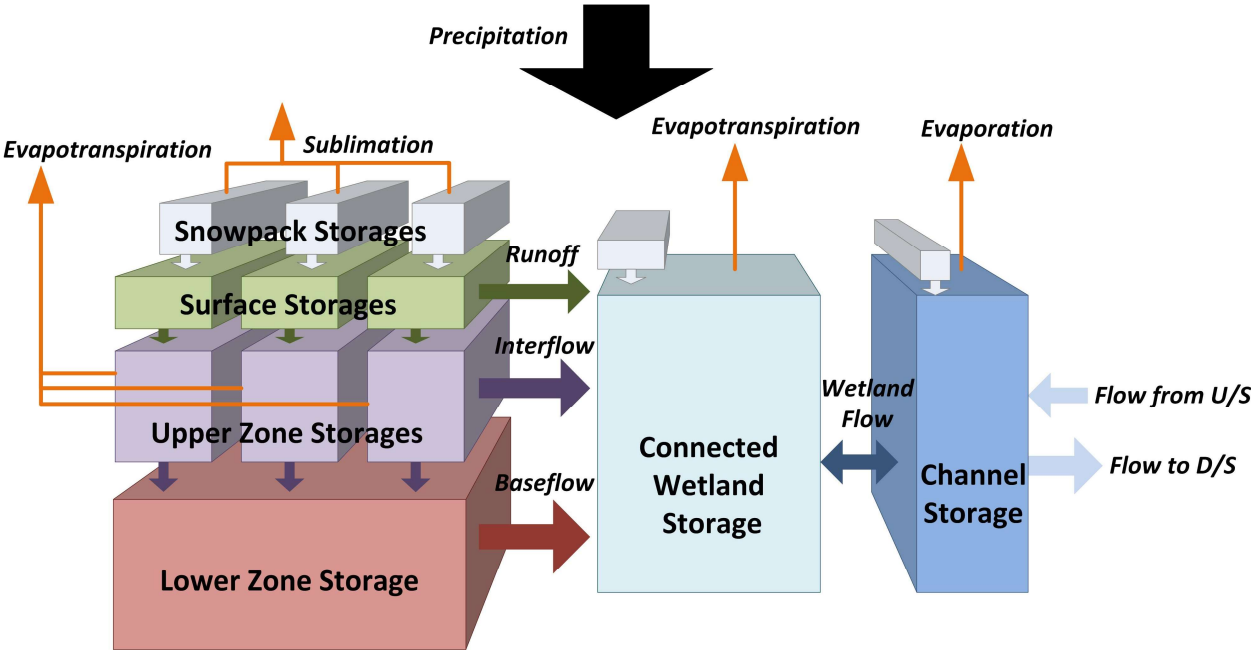

Figure 2-2: Storages and fluxes for one cell in the isoWATFLOOD model (Holmes, 2016a).

Each individual storage unit is modeled separately using a mass balance model, where the change in isotope tracer volume is equal to the isotope tracer inflow less the isotope tracer outflow. All individual storages are assumed to be completely mixed through depth. Non-fractionating fluxes have the same isotope tracer concentration as the source storage; evaporation

fluxes and the evaporation component of evapotranspiration fluxes have a lower concentration of heavy isotopes. The evaporative flux concentration is modeled based on a simplified version of the Craig-Gordon equation and depends on both the concentration of isotopes in the evaporating storage unit and atmospheric conditions (Stadnyk et al., 2013). The oxygen-18 and deuterium volumes are modeled independently; the current model code allows for either the simulation of oxygen-18 alone, or both oxygen-18 and deuterium (Holmes, 2016b; Stadnyk and Holmes, 2020). Beyond the forcings required by the hydrologic model, the isotope simulation requires the isotope concentration of precipitation, and relative humidity data is recommended for evaporative fractionation estimation (Holmes, 2016a).

The isoWATFLOOD model outputs isotope tracer compositions (in δ format) for any user-specified individual simulated storage unit defined in the distributed model space, which can include streams, snowpacks, upper and lower soil zones, and wetlands; basic simulation error statistics are also calculated internally for any sites with measured isotope data series from streamflow. Standard outputs are daily, but the model can be adapted to output at the hourly or sub-hourly simulation time-step. When both isotopes are simulated, isotope framework points (δ_A, δ*, etc.) are calculated from the continuous simulated data.

2.3. Model parameter values

Hydrologic models depend on parameters, and the values assigned to these parameters are key determinants of simulated streamflow, and for process-based models, the contribution of each individual process flux to the streamflow or basin storage. For isotope-enabled models, the parameter values exert a similar influence over simulated isotope concentrations in modeled storages and fluxes.

2.3.1. Identifiability and uncertainty

Parameter uncertainty is only one part of the total uncertainty in hydrologic model output, but it is the component most closely related to model calibration (Renard et al., 2010). Model parameter uncertainty is distinct from uncertainty in parameter values; the first is uncertainty in the modeled output, such as simulated streamflow, while the second is often termed parameter identifiability. There are many possible combinations of parameter values which can produce equally acceptable model results (i.e., similar aggregated errors). However, equally acceptable does not mean equal; different parameter sets will produce different simulated output, and it is

uncertain which is the most accurate, because all of the simulations are plausible (Beven and Freer, 2001). A parameter is well identified only if a narrow range of values for that parameter produces acceptable simulation results. Improvements in parameter value identification do not necessarily result in reductions of uncertainty in the model output; the propagation of uncertainty through the modeling process can be complex and the interactions between parameters within the model structure make parameter identifiability an unreliable proxy for model uncertainty (Guse et al., 2020; Moges et al., 2020; Pechlivanidis et al., 2011). Parameter uncertainty and identifiability are also issues for modelers for different reasons: parameter uncertainty makes flow predictions less precise, while poor parameter identifiability reduces model utility for understanding watershed function (Ebel and Loague, 2006; Guse et al., 2020).

There are two basic ways to estimate parameter uncertainty in hydrologic models: using Bayesian statistics and distributions of parameter values to estimate uncertainty in the output, or by identifying an ensemble of possible parameter sets and determining the output distribution (Li et al., 2010; Matott et al., 2009). Using Bayesian statistics is a rigorous and relatively efficient method, but it is dependent on the assumptions for parameter value distributions and error distribution (Li et al., 2010). Using ensembles of possible simulations requires fewer assumptions, but it can be computationally demanding and is dependent on a number of subjective choices (Li et al., 2010; Xiong et al., 2009). For example, a generalized likelihood uncertainty estimation (GLUE) method will depend on random Monte Carlo sampling to generate parameter sets, which is likely to result in very low fraction of acceptable, or behavioral, solutions (Beven and Binley, 2014, 1992). To improve efficiency over using random parameter values, calibration can be used to identify many behavioral parameter sets, using tools such as DDS-AU (DDS approximation of uncertainty), which can be used to approximate uncertainty bounds on the simulation but not confidence intervals (Tolson and Shoemaker, 2008). A few studies have published estimates parameter identifiability using isotope tracer data, such as He et al. (2019), based on multi-objective calibration results or Delavau et al. (2017), from Monte Carlo parameter sampling.

2.3.2. Parameter sensitivity

In hydrologic modeling, parameter sensitivity is the reaction of the model output to changes in the input parameter value. A sensitivity analysis for model parameters quantifies the sensitivity

of the model output to the parameter value, either locally, around a particular parameter value or globally, across a wide range of possible parameter values. Sensitivity analyses are distinct from calibration, as they do not aim to identify optimal parameter values; however, a sensitivity analysis is still a useful exercise when setting up a hydrologic model. A global sensitivity analysis (GSA) can illuminate the relative importance of different hydrologic processes in a basin model, and direct further research toward the most significant unknowns (Mai et al., 2022; Razavi and Gupta, 2015; Song et al., 2015). The sensitivity of parameters can also be used to inform a future calibration, directing more of the computational budget in calibration to sensitive parameters or address possible model deficiencies (Haghnegahdar et al., 2017). Parameter sensitivity analyses for hydrologic models are generally performed based on the response of the primary output, simulated streamflow, but parameter sensitivity can also be estimated based on associated responses, such as simulated isotope compositions.

There are multiple accepted approaches GSA, based on different methods of measuring the response of the output to changes in parameter values: derivative-based, variance-based and variogram-based. Derivative-based GSA estimate the local sensitivity of the output to a parameter by calculating the derivation of the response, or estimating the derivative using a finite difference method, and combine multiple local sensitivities to find the global sensitivity (Song et al., 2015). For example, the best-known derivative-based method, the Morris approach, evaluates the partial derivative of the response with respect to a parameter at multiple locations in the parameter space numerically, and then calculates the average and variance to estimate global sensitivity (Song et al., 2015). The Morris approach is relatively easy to understand and implement, but it is highly dependent on the sampling interval (Song et al., 2015). A more sophisticated alternative is variance-based GSA, such as the Sobol approach, which finds the total response or output variance, and then decomposes it into the variance due to individual parameters or interactions of parameters (Song et al., 2015). Variance-based GSA are truly global, unlike the Morris method, but are also computationally expensive. A more recently developed approach, variogram-based GSA, attempts to both improve the characterization of sensitivity and computational efficiency using variograms, which measure variance of differences in the response surface over the parameter space (Razavi and Gupta, 2015). Global parameter sensitivity can be quantified by integrating the variogram across a range of different scales, as done in the "Variogram Analysis of Response Surfaces" (VARS) framework (Razavi

and Gupta, 2016). The VARS framework is the basis of the VARS-TOOL software, which has recently added features for performing time-varying sensitivity analyses and grouping factors for high-dimensional cases (Razavi et al., 2019).

2.3.3. Calibration and performance metrics

All hydrologic models require parameters, from soil storage volumes to channel roughness factors. In some cases, these parameters can be estimated using field measurements or remotely sensed data, but most models will have parameters for which good or even reasonable values are not known (Singh and Frevert, 2010). Hydrologic models are therefore calibrated; unknown parameter values are optimized by running the model with many different parameter values and selecting those values which result in simulated results which resemble the observed system response. Model calibration is therefore a type of optimization problem, where the calibrated parameter values are decision variables, and the optimal solution provides the best fit between the simulated variable(s) of interest and observed values.

Hydrologic models are complex, so the optimal solution cannot be solved for directly, rather, an approximation of the optimal solution is found by trial and error: creating new combinations of parameter values, running the model with them and evaluating the output to decide if the parameter set is better than the previous best solution. This process can be extremely onerous for complex hydrologic models, as the number of calibrated parameters generally increases with model complexity. Some form of automatic calibration is therefore typically used; this process automates the generation of parameter sets and the evaluation of the output with a calibration algorithm. There are many areas of research relating to automatic calibration; those of particular interest to hydrologists rather than systems analysts can be related to one issue: how to define a "good" model result. Deciding if a model is good or even adequate for some particular purpose is challenging, first in deciding if the simulation is "right" or at least right enough, and secondly, in finding out if the model is right for the "right" reasons (Kirchner, 2006).

In automated calibration, the quality of the model performance needs to be evaluated quantitatively, using some type of error metric to compare the simulation to observed data. These error metrics, or model performance metrics, can be classified into three groups: measures of residual error, comparisons of data sets, and measures of the model efficiency (Bennett et al., 2013). Residual error metrics, such as the root mean squared error, are measurements of the

difference between pairs of simulated and observed data points (Bennett et al., 2013). Residual error metrics are frequently used as they are simple to calculate understand, can measure the model's ability to match the observed timing of events, and can applied even with large gaps in the observed data series, or sampling biases (Bennett et al., 2013; Oliver and Alfonzo, 2018). However, residual error metrics do not measure either the bias or variation of a simulation well, and when used with discontinuous data, these metrics are unaffected by the simulation quality between observations (Bennett et al., 2013).

Direct whole data set comparison metrics are the opposite of residual error metrics; these metrics calculate some property of the observed data population and compare it to the same property of the simulated data population (Bennett et al., 2013). There are many possible properties which can be compared, including, but not limited to, the mean, the range, the variance, or the values of various percentiles; in hydrology these metrics which describe statistical properties of data series are often termed 'flow signatures' (McMillan, 2021; Shafii and Tolson, 2015). Flow signatures can be used to identify dominant hydrologic processes and calibrate models (McMillan, 2021; McMillan et al., 2022). Using a data set comparison metric is the best method for measuring the simulation's match to the general behavior of the observed system (Bennett et al., 2013; Yilmaz et al., 2008). Depending on the exact metric used, this type of comparison can be used in calibration to ensure the simulation matches the observed mean and variability (Shafii and Tolson, 2015). However, direct comparisons are only meaningful if observed data set is representative of the real system behavior, and they provide no information on the match in the time series alignment between the simulation and observations (Oliver and Alfonzo, 2018; Yilmaz et al., 2008).

Model efficiency metrics are intended to be measures of the ability of a model to reproduce observed data, to distinguishing between models or parameter sets in order to select better models or parameters (Bennett et al., 2013; Knoben et al., 2019). There are two efficiency metrics commonly referred to in the literature, the Nash-Sutcliffe efficiency and the Kling-Gupta efficiency. The Nash-Sutcliffe efficiency (NSE) is a modified residual error metric, where the squared error is normalized with the squared error of a model which is only the mean of observations (Nash and Sutcliffe, 1970). The NSE can be a more useful measure of model performance than the squared error metric alone because it is normalized and unitless, and it has

generally accepted values of the limits of acceptability (Moriasi et al., 2015, 2007). The Kling-Gupta efficiency (KGE) metric is a more recent attempt to summarize the overall model match with observations in a single metric with three constituent components: the correlation, the relative variability and the bias (Gupta et al., 2009). The KGE is a useful model performance metric as it explicitly evaluates model bias and relative variability, unlike the NSE, although it does not have the same established reference values (Fowler et al., 2018; Gupta et al., 2009). Traditional performance metrics quantify how well two data series match, particularly when used in combination. However, they are generally inadequate for determining if the simulation is matching the observed data for the right reasons.

Quantifying model performance is not necessarily the only choice to be made in evaluating a model in an automatic calibration. When more than one data series (such as multiple flow gauges) is used in the evaluation or more than one metric is used for one data series, metrics need to be combined to function as a calibration objective. The most common choice for combining metrics is to add or average all the metrics in a single objective function for maximization or minimization in calibration, particularly when using the same type of metric at multiple gauges (Efstratiadis and Koutsoyiannis, 2010). The components of this type of objective function may have different weights, depending on the observed data type or quality, or based on the importance of the data site in the model. Alternatively, metrics can be used in the objective function conditionally, such as penalties for unacceptable simulation errors (Efstratiadis and Koutsoyiannis, 2010). A more computationally intensive, but informative and flexible option is using multi-objective calibration, where more than one objective function is considered simultaneously, and multiple non-dominated solutions are retained rather than a single solution (Asadzadeh and Tolson, 2013; Efstratiadis and Koutsoyiannis, 2010; Sahraei et al., 2020).

Model parameters can be calibrated to produce an optimal match between modeled and observed flows, based on a particular objective, but due to the uncertainty in both the simulated and observed values, it cannot be determined if this optimal parameter set will produce the most accurate flow simulation (Beven, 2016, 2008). Other parameter sets result in simulated flows that agree nearly as well with the observed flow, and these parameter sets may be superior in simulating flows based on a different objective or metric, or during a validation period, or in matching the true, rather than the observed, uncertain flow (Beven, 2016, 2006). The flows

produced by the final results of these different models can therefore be considered equivalent, given there is insufficient information available to determine which is "more right"; this unidentifiability of a best parameter set is termed equifinality (Beven, 2006). Additional observed data, such as snowpack volumes, groundwater heads, water quality measurements and tracer data, can be used to add new information into model calibration, which may assist in determining if a parameter set is truly equivalent to another, or in determining if a model is providing the right answer for the right reasons (Kirchner, 2006).

2.4.Summary of gaps in existing literature

There are research gaps relating to guidance on methods of incorporating isotope tracer data into hydrologic modeling methodologies. This is a particular issue for large scale watershed modeling, as operational scale basins do not have the same data density advantages as small research basins. Large scale basins have lower spatial collection site density, but more importantly, data collection is also sparser temporally, and sampling also tends to be irregular and seasonally biased, due to transportation and other operational limitations. More research is needed to determine how isotope model performance should be evaluated, when the observation data is discontinuous, irregularly spaced and potentially seasonally biased.

Previous research in isotope-enabled hydrologic modeling has been either at a limited spatial scale or a coarse temporal resolution. The majority of studies have used study basins smaller than 100 km^2, and a few have modeled basins of a few hundred to a few thousand square kilometers, such as Delavau, et al. (2017), He, et al. (2019) and Nan, et al. (2021). The IWBMIso model has been used to model large basins (>100,000 km^2) by Belachew, et al. (2016), but at a monthly time interval. The isoWATFLOOD model has also been applied at a large scale with hourly isotope simulation output in Holmes (2016b), but the results are limited to test case of model functionality. There is a significant gap in the research for large or operational scale isotope-enabled hydrologic modeling, for basins with significant infrastructure investment and potential climate change impacts. It remains unknown whether or not the results seen in small research basins, such as improvements in process, water apportionment, or parameter value identification, are seen at the large scale with the isotope data sets currently available.

2.5.References

Aggarwal, P.K., Romatschke, U., Araguas-Araguas, L., Belachew, D., Longstaffe, F.J., Berg, P., Schumacher, C., Funk, A., 2016. Proportions of convective and stratiform precipitation revealed in water isotope ratios. Nature Geoscience 9, 624–629. https://doi.org/10.1038/ngeo2739

Ala-aho, P., Tetzlaff, D., McNamara, J.P., Laudon, H., Kormos, P., Soulsby, C., 2017. Modeling the isotopic evolution of snowpack and snowmelt: Testing a spatially distributed parsimonious approach. Water Resources Research 53, 5813–5830. https://doi.org/10.1002/2017WR020650

Allen, S.T., Kirchner, J.W., Braun, S., Siegwolf, R.T.W., Goldsmith, G.R., 2019. Seasonal origins of soil water used by trees. Hydrology and Earth System Sciences 23, 1199–1210. https://doi.org/10.5194/hess-23-1199-2019

Asadzadeh, M., Tolson, B., 2013. Pareto archived dynamically dimensioned search with hypervolume-based selection for multi-objective optimization. Engineering Optimization 45, 1489–1509.

Belachew, D.L., Leavesley, G., David, O., Patterson, D., Aggarwal, P., Araguas, L., Terzer, S., Carlson, J., 2016. IAEA Isotope-enabled coupled catchment–lake water balance model, IWBMIso: description and validation†. Isotopes in Environmental and Health Studies 52, 427–442. https://doi.org/10.1080/10256016.2015.1113959

Bennett, N.D., Croke, B.F.W., Guariso, G., Guillaume, J.H.A., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, L.T.H., Norton, J.P., Perrin, C., Pierce, S.A., Robson, B., Seppelt, R., Voinov, A.A., Fath, B.D., Andreassian, V., 2013. Characterising performance of environmental models. Environmental Modelling and Software 40, 1–20. https://doi.org/10.1016/j.envsoft.2012.09.011

Bergström, S., Lindström, G., Pettersson, A., 2002. Multi-variable parameter estimation to increase confidence in hydrological modelling. Hydrological Processes 16, 413–421. https://doi.org/10.1002/hyp.332

Beria, H., Larsen, J.R., Ceperley, N.C., Michelon, A., Vennemann, T., Schaefli, B., 2018. Understanding snow hydrological processes through the lens of stable water isotopes. WIREs Water 5. https://doi.org/10.1002/wat2.1311

Beven, K., 2016. Facets of uncertainty: Epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication. Hydrological Sciences Journal 61, 1652–1665. https://doi.org/10.1080/02626667.2015.1031761

Beven, K., 2008. On doing better hydrological science. Hydrological Processes. https://doi.org/10.1002/hyp.7108

Beven, K., 2006. A manifesto for the equifinality thesis, in: Journal of Hydrology. pp. 18–36. https://doi.org/10.1016/j.jhydrol.2005.07.007

Beven, K., Binley, A., 2014. GLUE: 20 years on. Hydrol Process 28, 5897–5918.

Beven, K., Binley, A., 1992. The future of distributed models: Model calibration and uncertainty prediction. Hydrological Processes 6, 279–298. https://doi.org/10.1002/hyp.3360060305

Beven, K., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. J Hydrol (Amst) 249, 11–29.

Birkel, C., Dunn, S.M., Tetzlaff, D., Soulsby, C., 2010. Assessing the value of high-resolution isotope tracer data in the stepwise development of a lumped conceptual rainfall-runoff model. Hydrological Processes 24, 2335–2348. https://doi.org/10.1002/hyp.7763

Birkel, C., Duvert, C., Correa, A., Munksgaard, N.C., Maher, D.T., Hutley, L.B., 2020. Tracer-Aided Modeling in the Low-Relief, Wet-Dry Tropics Suggests Water Ages and DOC Export Are Driven by Seasonal Wetlands and Deep Groundwater. Water Resources Research 56. https://doi.org/10.1029/2019WR026175

Birkel, C., Paroli, R., Spezia, L., Dunn, S.M., Tetzlaff, D., Soulsby, C., 2012a. A new approach to simulating stream isotope dynamics using Markov switching autoregressive models. Advances in Water Resources 46, 20–30. https://doi.org/10.1016/j.advwatres.2012.05.013

Birkel, C., Soulsby, C., 2016. Linking tracers, water age and conceptual models to identify dominant runoff processes in a sparsely monitored humid tropical catchment. Hydrological Processes 30, 4477–4493. https://doi.org/10.1002/hyp.10941

Birkel, C., Soulsby, C., 2015. Advancing tracer-aided rainfall-runoff modelling: A review of progress, problems and unrealised potential. Hydrological Processes 29, 5227–5240. https://doi.org/10.1002/hyp.10594

Birkel, C., Soulsby, C., Tetzlaff, D., 2015. Conceptual modelling to assess how the interplay of hydrological connectivity, catchment storage and tracer dynamics controls nonstationary water age estimates. Hydrological Processes 29, 2956–2969. https://doi.org/10.1002/hyp.10414

Birkel, C., Soulsby, C., Tetzlaff, D., 2014. Developing a consistent process-based conceptualization of catchment functioning using measurements of internal state variables. Water Resources Research 50, 3481–3501. https://doi.org/10.1002/2013WR014925

Birkel, C., Soulsby, C., Tetzlaff, D., 2011a. Modelling catchment-scale water storage dynamics: Reconciling dynamic storage with tracer-inferred passive storage. Hydrological Processes 25, 3924–3936. https://doi.org/10.1002/hyp.8201

Birkel, C., Soulsby, C., Tetzlaff, D., Dunn, S., Spezia, L., 2012b. High-frequency storm event isotope sampling reveals time-variant transit time distributions and influence of diurnal cycles. Hydrological Processes 26, 308–316. https://doi.org/10.1002/hyp.8210

Birkel, C., Tetzlaff, D., Dunn, S.M., Soulsby, C., 2011b. Using lumped conceptual rainfall-runoff models to simulate daily isotope variability with fractionation in a nested mesoscale

catchment. Advances in Water Resources 34, 383–394. https://doi.org/10.1016/j.advwatres.2010.12.006

Birkel, C., Tetzlaff, D., Dunn, S.M., Soulsby, C., 2011c. Using time domain and geographic source tracers to conceptualize streamflow generation processes in lumped rainfall-runoff models. Water Resources Research 47. https://doi.org/10.1029/2010WR009547

Birks, S.J., Edwards, T.W.D., 2009. Atmospheric circulation controls on precipitation isotope-climate relations in western Canada. Tellus, Series B: Chemical and Physical Meteorology 61, 566–576. https://doi.org/10.1111/j.1600-0889.2009.00423.x

Birks, S.J., Gibson, J.J., 2009. Isotope hydrology research in Canada, 2003-2007. Canadian Water Resources Journal 34, 163–176.

Bowen, G.J., Cai, Z., Fiorella, R.P., Putman, A.L., 2019. Isotopes in the Water Cycle: Regional- to Global-Scale Patterns and Applications. Annual Review of Earth and Planetary Sciences 47, 453–479. https://doi.org/10.1146/annurev-earth-053018-060220

Brenčič, M., Kononova, N.K., Vreča, P., 2015. Relation between isotopic composition of precipitation and atmospheric circulation patterns. Journal of Hydrology 529, 1422–1432. https://doi.org/10.1016/j.jhydrol.2015.08.040

Brooks, J.R., Mushet, D.M., Vanderhoof, M.K., Leibowitz, S.G., Christensen, J.R., Neff, B.P., Rosenberry, D.O., Rugh, W.D., Alexander, L.C., 2018. Estimating Wetland Connectivity to Streams in the Prairie Pothole Region: An Isotopic and Remote Sensing Approach. Water Resources Research 54, 955–977. https://doi.org/10.1002/2017WR021016

Chesnaux, R., Stumpp, C., 2018. Advantages and challenges of using soil water isotopes to assess groundwater recharge dominated by snowmelt at a field study located in Canada. Hydrological Sciences Journal 63, 679–695. https://doi.org/10.1080/02626667.2018.1442577

Chlumsky, R., Mai, J., Craig, J.R., Tolson, B.A., 2021. Simultaneous Calibration of Hydrologic Model Structure and Parameters Using a Blended Model. Water Resources Research 57. https://doi.org/10.1029/2020WR029229

Clark, M.P., Fan, Y., Lawrence, D.M., Adam, J.C., Bolster, D., Gochis, D.J., Hooper, R.P., Kumar, M., Leung, L.R., Mackay, D.S., Maxwell, R.M., Shen, C., Swenson, S.C., Zeng, X., 2015. Improving the representation of hydrologic processes in Earth System Models. Water Resources Research 51, 5929–5956. https://doi.org/10.1002/2015WR017096

Craig, H., 1961. Isotopic variations in meteoric waters. Science (1979) 133, 1702–1703.

Craig, H., Gordon, L.I., 1965. Deuterium and oxygen 18 variations in the ocean and the marine atmosphere.

Dansgaard, W., 1964. Stable isotopes in precipitation. Tellus 16, 436–468. https://doi.org/10.1111/j.2153-3490.1964.tb00181.x

Davies, J., Beven, K., Nyberg, L., Rodhe, A., 2011. A discrete particle representation of hillslope hydrology: hypothesis testing in reproducing a tracer experiment at Gårdsjön, Sweden. Hydrological Processes 25, 3602–3612. https://doi.org/10.1002/hyp.8085

Dehaspe, J., Birkel, C., Tetzlaff, D., Sánchez-Murillo, R., Durán-Quesada, A.M., Soulsby, C., 2018. Spatially distributed tracer-aided modelling to explore water and isotope transport, storage and mixing in a pristine, humid tropical catchment. Hydrological Processes 32, 3206–3224. https://doi.org/10.1002/hyp.13258

Delavau, C., Chun, K.P., Stadnyk, T., Birks, S.J., Welker, J.M., 2015. North American precipitation isotope (δ18O) zones revealed in time series modeling across Canada and northern United States. Water Resources Research 51, 1284–1299. https://doi.org/10.1002/2014WR015687

Delavau, C., Stadnyk, T., Holmes, T., 2017. Examining the impacts of precipitation isotope input (δ18Oppt) on distributed, tracer-aided hydrological modelling. Hydrology and Earth System Sciences 21, 2595–2614. https://doi.org/10.5194/hess-21-2595-2017

Dincer, T., Payne, B.R., Florkowski, T., Martinec, J., Tongiorgi, E., 1970. Snowmelt runoff from measurements of tritium and oxygen-18. Water Resources Research 6, 110–124.

Dunn, S.M., Bacon, J.R., Soulsby, C., Tetzlaff, D., Stutter, M.I., Waldron, S., Malcolm, I.A., 2008. Interpretation of homogeneity in δ18O signatures of stream water in a nested sub-catchment system in north-east Scotland. Hydrological Processes 22, 4767–4782. https://doi.org/10.1002/hyp.7088

Dunn, S.M., McDonnell, J.J., Vaché, K.B., 2007. Factors influencing the residence time of catchment waters: A virtual experiment approach. Water Resources Research 43. https://doi.org/10.1029/2006WR005393

Ebel, B.A., Loague, K., 2006. Physics-based hydrologic-response simulation: Seeing through the fog of equifinality. Hydrological Processes 20, 2887–2900. https://doi.org/10.1002/hyp.6388

Efstratiadis, A., Koutsoyiannis, D., 2010. One decade of multi-objective calibration approaches in hydrological modelling: a review. Hydrological Sciences Journal–Journal Des Sciences Hydrologiques 55, 58–78.

Eriksson, E., 1963. Atmospheric tritium as a tool for the study of certain hydrologic aspects of river basins. Tellus 15, 303–308. https://doi.org/10.1111/j.2153-3490.1963.tb01391.x

Evaristo, J., McDonnell, J.J., Clemens, J., 2017. Plant source water apportionment using stable isotopes: A comparison of simple linear, two-compartment mixing model approaches. Hydrological Processes 31, 3750–3758. https://doi.org/10.1002/hyp.11233

Fatichi, S., Vivoni, E.R., Ogden, F.L., Ivanov, V.Y., Mirus, B., Gochis, D., Downer, C.W., Camporese, M., Davison, J.H., Ebel, B., Jones, N., Kim, J., Mascaro, G., Niswonger, R., Restrepo, P., Rigon, R., Shen, C., Sulis, M., Tarboton, D., 2016. An overview of current

applications, challenges, and future trends in distributed process-based models in hydrology. Journal of Hydrology 537, 45–60. https://doi.org/10.1016/j.jhydrol.2016.03.026

Fenicia, F., Wrede, S., Kavetski, D., Pfister, L., Hoffmann, L., Savenije, H.H.G., McDonnell, J.J., 2010. Assessing the impact of mixing assumptions on the estimation of streamwater mean residence time. Hydrological Processes 24, 1730–1741. https://doi.org/10.1002/hyp.7595

Fowler, K., Peel, M., Western, A., Zhang, L., 2018. Improved rainfall-runoff calibration for drying climate: Choice of objective function. Water Resources Research 54, 3392–3408.

Gat, J.R., 1996. Oxygen and hydrogen isotopes in the hydrologic cycle. Annual Review of Earth and Planetary Sciences 24, 225–262.

Gibson, J.J., Birks, S.J., Edwards, T.W.D., 2008. Global prediction of δA and δ2H- δ18O evaporation slopes for lakes and soil water accounting for seasonality. Global Biogeochemical Cycles 22. https://doi.org/10.1029/2007GB002997

Gibson, J.J., Birks, S.J., Yi, Y., 2016. Stable isotope mass balance of lakes: A contemporary perspective. Quaternary Science Reviews 131, 316–328. https://doi.org/10.1016/j.quascirev.2015.04.013

Gibson, J.J., Edwards, T.W.D., 2002. Regional water balance trends and evaporation-transpiration partitioning from a stable isotope survey of lakes in northern Canada. Global Biogeochemical Cycles 16, 10-1-10–14. https://doi.org/10.1029/2001gb001839

Gibson, J.J., Edwards, T.W.D., Prowse, T.D., 1996. Development and validation of an isotopic method for estimating lake evaporation. Hydrol Process 10, 1369–1382.

Gibson, J.J., Holmes, T., Stadnyk, T.A., Birks, S.J., Eby, P., Pietroniro, A., 2020. 18O and 2H in streamflow across Canada. Journal of Hydrology: Regional Studies. https://doi.org/10.1016/j.ejrh.2020.100754

Gibson, J.J., Reid, R., 2010. Stable isotope fingerprint of open-water evaporation losses and effective drainage area fluctuations in a subarctic shield watershed. Journal of Hydrology 381, 142–150. https://doi.org/10.1016/j.jhydrol.2009.11.036

Gupta, H. v., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of Hydrology 377, 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Guse, B., Fatichi, S., Gharari, S., Melsen, L.A., 2021. Advancing Process Representation in Hydrological Models: Integrating New Concepts, Knowledge, and Data. Water Resources Research 57. https://doi.org/10.1029/2021WR030661

Guse, B., Kiesel, J., Pfannerstill, M., Fohrer, N., 2020. Assessing parameter identifiability for multiple performance criteria to constrain model parameters. Hydrological Sciences Journal 65, 1158–1172. https://doi.org/10.1080/02626667.2020.1734204

Haghnegahdar, A., Razavi, S., Yassin, F., Wheater, H., 2017. Multicriteria sensitivity analysis as a diagnostic tool for understanding model behaviour and characterizing model uncertainty. Hydrological Processes 31, 4462–4476. https://doi.org/10.1002/hyp.11358

Hayashi, M., Quinton, W.L., Pietroniro, A., Gibson, J.J., 2004. Hydrologic functions of wetlands in a discontinuous permafrost basin indicated by isotopic and chemical signatures. Journal of Hydrology 296, 81–97. https://doi.org/10.1016/j.jhydrol.2004.03.020

He, Z., Unger-Shayesteh, K., Vorogushyn, S., Weise, S.M., Kalashnikova, O., Gafurov, A., Duethmann, D., Barandun, M., Merz, B., 2019. Constraining hydrological model parameters using water isotopic compositions in a glacierized basin, Central Asia. Journal of Hydrology 571, 332–348. https://doi.org/10.1016/j.jhydrol.2019.01.048

Holmes, T., 2016a. isoWATFLOOD Stable water isotope simulation in the WATFLOOD hydrologic model.

Holmes, T., 2016b. Assessing the value of stable water isotopes in hydrologic modeling: a dual isotope approach (MSc). University of Manitoba, Winnipeg.

Iorgulescu, I., Beven, K.J., Musy, A., 2007. Flow, mixing, and displacement in using a data-based hydrochemical model to predict conservative tracer data. Water Resources Research 43. https://doi.org/10.1029/2005WR004019

Jafari, T., Kiem, A.S., Javadi, S., Nakamura, T., Nishida, K., 2021. Using insights from water isotopes to improve simulation of surface water-groundwater interactions. Science of The Total Environment 798, 149253. https://doi.org/10.1016/j.scitotenv.2021.149253

Kennedy, V.C., Kendall, C., Zellweger, G.W., Wyerman, T.A., Avanzino, R.J., 1986. Determination of the components of stormflow using water chemistry and environmental isotopes, Mattole River basin, California. Journal of Hydrology 84, 107–140.

Kirchner, J.W., 2019. Quantifying new water fractions and transit time distributions using ensemble hydrograph separation: theory and benchmark tests. Hydrology and Earth System Sciences 23, 303–349. https://doi.org/10.5194/hess-23-303-2019

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. https://doi.org/10.1029/2005WR004362

Kirchner, J.W., Allen, S.T., 2020. Seasonal partitioning of precipitation between streamflow and evapotranspiration, inferred from end-member splitting analysis. Hydrology and Earth System Sciences 24, 17–39. https://doi.org/10.5194/hess-24-17-2020

Klaus, J., Chun, K.P., McGuire, K.J., McDonnell, J.J., 2015a. Temporal dynamics of catchment transit times from stable isotope data. Water Resources Research 51, 4208–4223. https://doi.org/10.1002/2014WR016247

Klaus, J., McDonnell, J.J., 2013. Hydrograph separation using stable isotopes: Review and evaluation. Journal of Hydrology 505, 47–64. https://doi.org/10.1016/j.jhydrol.2013.09.006

Klaus, J., McDonnell, J.J., Jackson, C.R., Du, E., Griffiths, N.A., 2015b. Where does streamwater come from in low-relief forested watersheds? A dual-isotope approach. Hydrology and Earth System Sciences 19, 125–135. https://doi.org/10.5194/hess-19-125-2015

Knoben, W.J.M., Freer, J.E., Woods, R.A., 2019. Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores. Hydrology and Earth System Sciences 23, 4323–4331. https://doi.org/10.5194/hess-23-4323-2019

Kouwen, N., 2018. WATFLOOD/CHARM Canadian Hydrological And Routing Model . Waterloo.

Li, L., Xia, J., Xu, C.-Y., Singh, V.P., 2010. Evaluation of the subjective factors of the GLUE method and comparison with the formal Bayesian method in uncertainty assessment of hydrological models. J Hydrol (Amst) 390, 210–221.

Mai, J., Craig, J.R., Tolson, B.A., Arsenault, R., 2022. The sensitivity of simulated streamflow to individual hydrologic processes across North America. Nature Communications 13, 455. https://doi.org/10.1038/s41467-022-28010-7

Małoszewski, P., Rauert, W., Stichler, W., Herrmann, A., 1983. Application of flow models in an alpine catchment area using tritium and deuterium data. Journal of Hydrology 66, 319–330.

Małoszewski, P., Zuber, A., 1982. Determining the turnover time of groundwater systems with the aid of environmental tracers: 1. Models and their applicability. J Hydrol (Amst) 57, 207–231.

Matott, L.S., Babendreier, J.E., Purucker, S.T., 2009. Evaluating uncertainty in integrated environmental models: A review of concepts and tools. Water Resources Research 45.

McMillan, H., Tetzlaff, D., Clark, M., Soulsby, C., 2012. Do time-variable tracers aid the evaluation of hydrological model structure? A multimodel approach. Water Resources Research 48. https://doi.org/10.1029/2011WR011688

McMillan, H.K., 2021. A review of hydrologic signatures and their applications. WIREs Water 8. https://doi.org/10.1002/wat2.1499

McMillan, H.K., Gnann, S.J., Araki, R., 2022. Large Scale Evaluation of Relationships Between Hydrologic Signatures and Processes. Water Resources Research 58. https://doi.org/10.1029/2021WR031751

Moges, E., Demissie, Y., Larsen, L., Yassin, F., 2020. Review: Sources of Hydrological Model Uncertainties and Advances in Their Analysis. Water (Basel) 13, 28. https://doi.org/10.3390/w13010028

Moriasi, D.N., Arnold, J.G., van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans ASABE 50, 885–900.

Moriasi, D.N., Gitau, M.W., Pai, N., Daggupati, P., 2015. Hydrologic and water quality models: Performance measures and evaluation criteria. Trans ASABE 58, 1763–1785. https://doi.org/10.13031/trans.58.10715

Nan, Y., He, Z., Tian, F., Wei, Z., Tian, L., 2021a. Can we use precipitation isotope outputs of isotopic general circulation models to improve hydrological modeling in large mountainous catchments on the Tibetan Plateau? Hydrology and Earth System Sciences 25, 6151–6172. https://doi.org/10.5194/hess-25-6151-2021

Nan, Y., Tian, L., He, Z., Tian, F., Shao, L., 2021b. The value of water isotope data on improving process understanding in a glacierized catchment on the Tibetan Plateau. Hydrology and Earth System Sciences 25, 3653–3673. https://doi.org/10.5194/hess-25-3653-2021

Nash, J.E., Sutcliffe, J. v, 1970. River flow forecasting through conceptual models part I—A discussion of principles. Journal of Hydrology 10, 282–290.

Neill, A.J., Tetzlaff, D., Strachan, N.J.C., Soulsby, C., 2019. To what extent does hydrological connectivity control dynamics of faecal indicator organisms in streams? Initial hypothesis testing using a tracer-aided model. Journal of Hydrology 570, 423–435. https://doi.org/10.1016/j.jhydrol.2018.12.066

Niemi, A.J., 1977. Residence time distributions of variable flow processes. The International Journal of Applied Radiation and Isotopes 28, 855–860.

Ohlanders, N., Rodriguez, M., McPhee, J., 2013. Stable water isotope variation in a Central Andean watershed dominated by glacier and snowmelt. Hydrology and Earth System Sciences 17, 1035–1050. https://doi.org/10.5194/hess-17-1035-2013

Oliver, D.S., Alfonzo, M., 2018. Calibration of imperfect models to biased observations. Computational Geosciences 22, 145–161.

Oshun, J., Dietrich, W.E., Dawson, T.E., Fung, I., 2016. Dynamic, structured heterogeneity of water isotopes inside hillslopes. Water Resources Research 52, 164–189. https://doi.org/10.1002/2015WR017485

Pangle, L.A., Klaus, J., Berman, E.S.F., Gupta, M., McDonnell, J.J., 2013. A new multisource and high-frequency approach to measuring δ2H and δ18O in hydrological field studies. Water Resources Research 49, 7797–7803. https://doi.org/10.1002/2013WR013743

Parnell, A.C., Phillips, D.L., Bearhop, S., Semmens, B.X., Ward, E.J., Moore, J.W., Jackson, A.L., Grey, J., Kelly, D.J., Inger, R., 2013. Bayesian stable isotope mixing models. Environmetrics n/a-n/a. https://doi.org/10.1002/env.2221

Pechlivanidis, I.G., Jackson, B.M., Mcintyre, N.R., Wheater, H.S., 2011. Catchment scale hydrological modelling: A review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications. Global NEST journal 13, 193–214.

Peralta-Tapia, A., Sponseller, R.A., Tetzlaff, D., Soulsby, C., Laudon, H., 2015. Connecting precipitation inputs and soil flow pathways to stream water in contrasting boreal catchments. Hydrological Processes 29, 3546–3555. https://doi.org/10.1002/hyp.10300

Pinder, G.F., Jones, J.F., 1969. Determination of the ground-water component of peak discharge from the chemistry of total runoff. Water Resources Research 5, 438–445.

Piovano, T.I., Tetzlaff, D., Ala-aho, P., Buttle, J., Mitchell, C.P.J., Soulsby, C., 2018. Testing a spatially distributed tracer-aided runoff model in a snow-influenced catchment: Effects of multicriteria calibration on streamwater ages. Hydrological Processes 32, 3089–3107. https://doi.org/10.1002/hyp.13238

Piovano, T.I., Tetzlaff, D., Carey, S.K., Shatilla, N.J., Smith, A., Soulsby, C., 2019. Spatially distributed tracer-aided runoff modelling and dynamics of storage and water ages in a permafrost-influenced catchment. Hydrology and Earth System Sciences 23, 2507–2523. https://doi.org/10.5194/hess-23-2507-2019

Piovano, T.I., Tetzlaff, D., Maneta, M., Buttle, J.M., Carey, S.K., Laudon, H., McNamara, J., Soulsby, C., 2020. Contrasting storage-flux-age interactions revealed by catchment inter-comparison using a tracer-aided runoff model. Journal of Hydrology 590, 125226. https://doi.org/10.1016/j.jhydrol.2020.125226

Razavi, S., Gupta, H. v., 2016. A new framework for comprehensive, robust, and efficient global sensitivity analysis: 1. Theory. Water Resources Research 52, 423–439. https://doi.org/10.1002/2015WR017558

Razavi, S., Gupta, H. v., 2015. What do we mean by sensitivity analysis? the need for comprehensive characterization of "global" sensitivity in Earth and Environmental systems models. Water Resources Research 51, 3070–3092. https://doi.org/10.1002/2014WR016527

Razavi, S., Sheikholeslami, R., Gupta, H. v, Haghnegahdar, A., 2019. VARS-TOOL: A toolbox for comprehensive, efficient, and robust sensitivity and uncertainty analysis. Environmental Modelling & Software 112, 95–107. https://doi.org/https://doi.org/10.1016/j.envsoft.2018.10.005

Refsgaard, J.C., Stisen, S., Koch, J., 2022. Hydrological process knowledge in catchment modelling – Lessons and perspectives from 60 years development. Hydrological Processes 36. https://doi.org/10.1002/hyp.14463

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., Franks, S.W., 2010. Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. Water Resources Research 46.

Sahraei, S., Asadzadeh, M., Unduche, F., 2020. Signature-based multi-modelling and multi-objective calibration of hydrologic models: Application in flood forecasting for Canadian Prairies. Journal of Hydrology 588, 125095. https://doi.org/10.1016/j.jhydrol.2020.125095

Shafii, M., Tolson, B.A., 2015. Optimizing hydrological consistency by incorporating hydrological signatures into model calibration objectives. Water Resources Research 51, 3796–3814. https://doi.org/10.1002/2014WR016520

Singh, V.P., Frevert, D.K., 2010. Watershed models. CRC press.

Sklash, M.G., Farvolden, R.N., Fritz, P., 1976. A conceptual model of watershed response to rainfall, developed through the use of oxygen-18 as a natural tracer. Canadian Journal of Earth Sciences 13, 271–283.

Smith, A., Welch, C., Stadnyk, T., 2016. Assessment of a lumped coupled flow-isotope model in data scarce Boreal catchments. Hydrological Processes 30, 3871–3884. https://doi.org/10.1002/hyp.10835

Song, X., Zhang, J., Zhan, C., Xuan, Y., Ye, M., Xu, C., 2015. Global sensitivity analysis in hydrological modeling: Review of concepts, methods, theoretical framework, and applications. Journal of Hydrology. https://doi.org/10.1016/j.jhydrol.2015.02.013

Soulsby, C., Birkel, C., Geris, J., Dick, J., Tunaley, C., Tetzlaff, D., 2015a. Stream water age distributions controlled by storage dynamics and nonlinear hydrologic connectivity: Modeling with high-resolution isotope data. Water Resources Research 51, 7759–7776. https://doi.org/10.1002/2015WR017888

Soulsby, C., Birkel, C., Geris, J., Tetzlaff, D., 2015b. Spatial aggregation of time-variant stream water ages in urbanizing catchments. Hydrological Processes 29, 3038–3050. https://doi.org/10.1002/hyp.10500

Sprenger, M., Leistert, H., Gimbel, K., Weiler, M., 2016. Illuminating hydrological processes at the soil-vegetation-atmosphere interface with water stable isotopes. Reviews of Geophysics 54, 674–704. https://doi.org/10.1002/2015RG000515

Sprenger, M., Tetzlaff, D., Soulsby, C., 2017. Soil water stable isotopes reveal evaporation dynamics at the soil-plant-atmosphere interface of the critical zone. Hydrology and Earth System Sciences 21, 3839–3856. https://doi.org/10.5194/hess-21-3839-2017

St Amour, N.A., Gibson, J.J., Edwards, T.W.D., Prowse, T.D., Pietroniro, A., 2005. Isotopic time-series partitioning of streamflow components in wetland-dominated catchments, lower Liard river basin, Northwest Territories, Canada. Hydrological Processes 19, 3357–3381. https://doi.org/10.1002/hyp.5975

Stadnyk, T.A., Delavau, C., Kouwen, N., Edwards, T.W.D., 2013. Towards hydrological model calibration and validation: Simulation of stable water isotopes using the isoWATFLOOD model. Hydrological Processes 27, 3791–3810. https://doi.org/10.1002/hyp.9695

Stadnyk, T.A., Holmes, T.L., 2020. On the value of isotope-enabled hydrological model calibration. Hydrological Sciences Journal 65, 1525–1538. https://doi.org/10.1080/02626667.2020.1751847

Stevenson, J.L., Birkel, C., Neill, A.J., Tetzlaff, D., Soulsby, C., 2021. Effects of streamflow isotope sampling strategies on the calibration of a tracer-aided rainfall-runoff model. Hydrological Processes 35. https://doi.org/10.1002/hyp.14223

Stumpp, C., Brüggemann, N., Wingate, L., 2018. Stable Isotope Approaches in Vadose Zone Research. Vadose Zone Journal 17, 180096. https://doi.org/10.2136/vzj2018.05.0096

Tetzlaff, D., Birkel, C., Dick, J., Geris, J., Soulsby, C., 2014. Storage dynamics in hydropedological units control hillslope connectivity, runoff generation, and the evolution of catchment transit time distributions. Water Resources Research 50, 969–985. https://doi.org/10.1002/2013WR014147

Tetzlaff, D., Buttle, J., Carey, S.K., Kohn, M.J., Laudon, H., McNamara, J.P., Smith, A., Sprenger, M., Soulsby, C., 2021. Stable isotopes of water reveal differences in plant – soil water relationships across northern environments. Hydrological Processes 35. https://doi.org/10.1002/hyp.14023

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

Tunaley, C., Tetzlaff, D., Birkel, C., Soulsby, C., 2017. Using high-resolution isotope data and alternative calibration strategies for a tracer-aided runoff model in a nested catchment. Hydrological Processes 31, 3962–3978. https://doi.org/10.1002/hyp.11313

Vachon, R.W., Welker, J.M., White, J.W.C., Vaughn, B.H., 2010. Moisture source temperatures and precipitation δ18O- temperature relationships across the United States. Water Resources Research 46. https://doi.org/10.1029/2009WR008558

van Huijgevoort, M.H.J., Tetzlaff, D., Sutanudjaja, E.H., Soulsby, C., 2016. Using high resolution tracer data to constrain water storage, flux and age estimates in a spatially distributed rainfall-runoff model. Hydrological Processes 30, 4761–4778. https://doi.org/10.1002/hyp.10902

Wang, L., von Freyberg, J., van Meerveld, I., Seibert, J., Kirchner, J.W., 2019. What is the best time to take stream isotope samples for event-based model calibration? Journal of Hydrology 577, 123950. https://doi.org/10.1016/j.jhydrol.2019.123950

Weiler, M., McGlynn, B.L., McGuire, K.J., McDonnell, J.J., 2003. How does rainfall become runoff? A combined tracer and runoff transfer function approach. Water Resources Research 39. https://doi.org/10.1029/2003WR002331

Xiong, L., Wan, M., Wei, X., O'Connor, K.M., 2009. Indices for assessing the prediction bounds of hydrological models and application by generalised likelihood uncertainty estimation. Hydrological Sciences Journal 54, 852–871. https://doi.org/10.1623/hysj.54.5.852

Yilmaz, K.K., Gupta, H. v., Wagener, T., 2008. A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. Water Resources Research 44. https://doi.org/10.1029/2007WR006716

Yoshimura, K., 2015. Stable water isotopes in climatology, meteorology, and hydrology: A review. Journal of the Meteorological Society of Japan. https://doi.org/10.2151/jmsj.2015-036

Yoshimura, K., Kanamitsu, M., Noone, D., Oki, T., 2008. Historical isotope simulation using Reanalysis atmospheric data. Journal of Geophysical Research 113, D19108. https://doi.org/10.1029/2008JD010074

Yoshimura, K., Miyazaki, S., Kanae, S., Oki, T., 2006. Iso-MATSIRO, a land surface model that incorporates stable water isotopes. Global and Planetary Change 51, 90–107. https://doi.org/10.1016/j.gloplacha.2005.12.007

Zhang, J., Zhang, S., Zhang, W., Liu, B., Gong, C., Jiang, M., Lv, X., Sheng, L., 2018. Partitioning daily evapotranspiration from a marsh wetland using stable isotopes in a semiarid region. Hydrology Research 49, 1005–1015. https://doi.org/10.2166/nh.2017.005

3. Regional calibration with isotope tracers using a spatially distributed model: a comparison of methods

Tegan L. Holmes[1], Tricia A. Stadnyk[1, 2], Su Jin Kim[1], and Masoud Asadzadeh[1]
[1]Civil Engineering, University of Manitoba, Winnipeg, Manitoba, R3T 5V6, Canada
[2]Geography, University of Calgary, Calgary, Alberta, T2N 1N4, Canada

Published in 2020 in Water Resources Research (doi: 10.1029/2020WR027447).
Submitted March 2020; accepted August 2020.

This chapter presents a case study for calibrating with both isotope tracer and flow data and compares methods of combining tracer and flow simulation performance in optimization objectives. The parameter values and modeled fluxes produced by different objective formulations are compared, and the results of this chapter lead to a recommendation to use multi-objective optimizations for tracer-aided calibrations.

3.1. Abstract

Accurate representation of flow sources in process-based hydrologic models remains challenging for remote, data-scarce regions. This study applies stable isotope tracers (18O and 2H) in water as auxiliary data for the calibration of the isoWATFLOOD™ model. The most efficient method of those evaluated for introducing isotope data into model calibration was the PA-DDS multi-objective search algorithm. The compromise solutions incorporating isotope data performed slightly inferior in terms of streamflow simulation compared to the calibrated solution using streamflow data only. However, the former solution outperformed the latter one in terms of isotope simulation. Approximation of the model parameter uncertainty into internal flow path partitioning was explored. Inclusion of isotope error facilitated a broader examination of the total parameter space, resulting in significant differences in internal storage and flow paths, most significantly for soil storage and evapotranspiration loss. Isotope-optimized calibration reduced evaporation rates and increased soil moisture content within the model, impacting soil water velocity but not streamflow celerity. Flow-only calibration resulted in artificially narrow model prediction bounds, significantly under-estimating the propagation of parameter uncertainty,

while isotope-informed calibrations yielded more reliable and robust bound on model predictions. Our findings demonstrate that the accuracy of a complex, spatially distributed, and process-based model cannot be judged from one summative flow-based model performance evaluation metric alone.

Plain Language Summary

To predict how water moves through watersheds and adequately anticipate hydrologic response to a changing climate, we depend on numerical models to estimate transport and storage of water on and below Earth's surface. These models are only as good as the observation data we have to test them against, which can be insufficient for many remote parts of the world. Our study introduces a new source of data (i.e., isotope tracers, a fingerprint of water's origin) into a model to improve parameter estimation, and overall reliability. The largest benefit of adding these data is that it nudges the model into a more realistic distribution of water storage and flow paths to the rivers, particularly soil storage and evaporation loss. This is important for improving the reliability of model projections under changing climates and to better understand how the distribution of water on Earth is changing.

3.2. Introduction

Hydrologic models are essential to the study of many water resources systems, where they are used to simulate data, such as ungauged or future flows that cannot be measured. These models require various (often many) parameter values to simulate the highly non-linear hydrologic response of a watershed. It is generally not possible to measure all parameter values, therefore models are calibrated (Beven, 2012). In calibration, model parameter values are chosen to best reproduce observation data in the simulation, typically observed streamflow. The match between the observations and simulation, or model performance, is evaluated to assess the models 'fit for purpose' (Kirchner, 2006), and assumes error does not significantly propagate outside of the calibration period.

Evaluating the simulated flow error, however, may be an inadequate measure of the model's success at reproducing flow paths, water partitioning, and transit time within the real system, particularly over long simulation periods. It may be more desirable to have a hydrologic model simulate flow path partitioning and total streamflow correctly, particularly for applications

assessing long-term water balance under changing climates (Kirchner, 2006). Additional observed data, such as snowpack volume, groundwater head, water quality and tracer data, can inform model calibration and test model process-based performance against alternative observations, as an additional metric of a models 'fit for purpose' (Arnold et al. 2015; Guzman et al. 2015).

One emerging supplementary data source are stable isotopes in water: measurements of the ratio of water molecules containing ^{18}O or ^{2}H to standard water. Isotope tracers are naturally occurring, non-reactive, and can provide additional information on water sources and hydrological processes due to their variable occurrence in precipitation and evaporating water bodies (Birks & Gibson, 2009). A hydrologic model capable of continuously simulating both flow and isotopic composition on time scales relevant for flow path partitioning (i.e., sub-hourly) can be evaluated against both observed isotope and flow data, potentially adding new information to the model calibration. Isotope-enabled model calibration has been tested in a few hydrologic models, established first in catchment-specific, lumped models (e.g., Lindström and Rodhe 1986, de Grosbois et al. 1988, Vache and McDonnell 2004, Son and Sivapalan 2006, Dunn & Bacon 2008). More recently, studies have emerged using both lumped and distributed models applied across a variety of temporal and spatial scales (e.g., Birkel et al., 2010; Hu et al. 2019; Seibert & McDonnell, 2013; Smith et al., 2016; Stadnyk et al., 2013). In these studies, isotope performance metrics have been introduced into traditional automated calibration by creating a single objective (SO) function combining both flow and isotope error (e.g., Bergstrom et al., 2002; Holmes, 2016). One particular challenge has been the disparity in amount and time correspondence of flow (continuous) and isotope tracer (sporadic) data, and the model time scale (most commonly daily or monthly).

Alternatively, models can be calibrated via Monte Carlo approaches, allowing multiple criteria to be applied to parameter selection, including both flow and isotope error (e.g., Delavau et al., 2017; Smith et al., 2016; Piovano, et al., 2019; Neill, et al., 2019; Dehaspe, et al., 2018). Recent studies have used the multi-objective (MO) calibration algorithm NGSAII by Deb et al. (2002) to calibrate both flow and isotopes simultaneously, with some success (e.g., Tunaley et al., 2017; Birkel & Soulsby, 2015; He, et al., 2019). Examining tradeoff between flow and isotope error in spatially distributed, process-based models offers the opportunity to examine complex

interactions among parameters and internal flow path partitioning, and the net propagation of parameter uncertainty into streamflow. Research to determine how best to incorporate isotope data into calibration across large spatial scales using complex, spatially distributed hydrologic models has been lacking, particularly comparisons of different methods and their implications for internal process and flowpath partitioning. Similarly, to ensure hydrologic process variability is captured, it is important the model time step be reflective of at least diurnal variability.

Recent literature demonstrates the benefit of increasing information content and additional parameter constraints when complementary data are used in multi-criteria calibration, although with mixed effects on streamflow performance (Nelson et al. 2018; He et al. 2019). Best practices have yet to be established, however, and some key questions remain surrounding the best calibration methods and metrics to evaluate model performance. Datasets are typically temporally incongruent, which impacts the choice of optimization method (i.e., multi-criteria or multi-objective) and affects error tradeoff during calibration and the resulting flow performance and flow path partitioning within a model.

In this study, we compare model parameterizations resulting from different methods of incorporating stable isotope data into a spatially distributed, continuous (1-hour time step) hydrologic model calibration to assess the impact of parameter uncertainty propagation onto the partitioning of flow paths. Multiple SO functions are used to optimize parameters for an isotope-enabled hydrologic model, and a MO calibration is conducted for comparison. The key research questions we explore in this study are:

(1) What are the tradeoffs between, and implications of, different calibration methodologies;
(2) Does the addition of isotope simulation error to a model calibration change the identified parameter values, and if so,
(3) How do these changes alter the simulated streamflow and flow path partitioning produced by the model?

We explore these questions through the lens of streamflow performance metrics and net change in parameter and flow path contributions to streamflow across a regional scale conducive for long-term water supply assessment.

3.3.Study site and model

3.3.1. Regional background

The Odei River watershed is located in northern Manitoba; with a latitude of 56.2° at the basin center and an area of 6110 km^2, it is a cold region basin with significant seasonality. The Odei River is tributary of the Burntwood River, which flows into the Nelson River as shown in Figure 3-1.

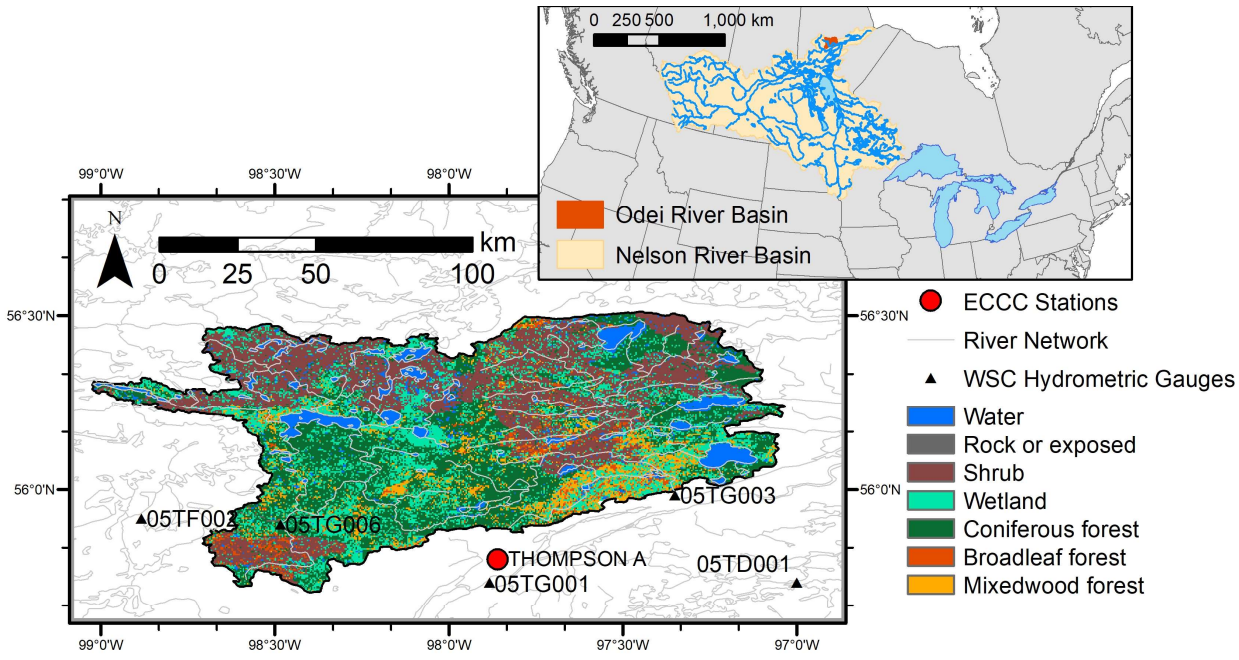

Figure 3-1: Map of the Odei River basin with land cover, and location within the Nelson and lower Nelson basins (inset)

Smith et al. (2016) describe the basin as having subdued topography (average basin slope of 1.4%, ranging from maximum elevation of 334m a.s.l. to 167m a.s.l. at the outlet). Soils are shallow with some exposed bedrock and sporadic permafrost. The basin resides in the Canadian Shield region: soil water storage is limited, and deep groundwater inflows are insignificant; however, extensive wetland complexes provide consistent baseflow, particularly during ice-on periods (Smith et al. 2018). The climate is sub-arctic and sub-humid, with moderate evapotranspiration, low temperature, and low precipitation (Smith et al., 2016). Most precipitation evapotranspires (ET), with an average precipitation and ET of 550 mm and 360 mm per year, respectively. Approximately 1/3 of annual precipitation is snowfall; snowpacks can be

present from October to May. Coniferous forest, shrubs and wetlands are the dominant land cover, with many small lakes and some mixed and broad leaved forest.

3.3.2. Data

Four types of time series forcing data were used to run the isoWATFLOOD^{TM} model: daily total precipitation, hourly air temperature, hourly relative humidity, and monthly average isotopic compositions of precipitation. Precipitation, temperature and humidity were estimated from the following Environment and Climate Change Canada observation gauges: Thompson A, Flin Flon A, The Pas A, Norway House A, South Indian Lake, Gillam A, Kelsey GS CS and Cross Lake A (https://climate.weather.gc.ca/historical_data/search_historic_data_e.html). Forcing data for each grid cell at each time step was derived by inverse distance squared weighting, using all gauges with data for that time interval (Kouwen, 2018). The isotopic composition of precipitation was estimated from an empirical model developed by Delavau, et al. (2015), which uses geographic and climatic indicators in a geospatial interpolation using multiple linear regression. A comparison between simulated isotopic composition of precipitation and observations for the basin was conducted by Stadnyk and Holmes (2020). The watershed model setup for WATFLOOD is described in Holmes (2016).

Historical hydrometric data from the Water Survey of Canada (https://wateroffice.ec.gc.ca) was used for calibrating and validating the model. The Odei River (05TG003) was used for calibration and validation, while the Sapochi River tributary (05TG006) was used only for supplementary validation as the flow record is less complete. Stable isotope data were also used for calibration; $\delta^{18}O$ compositions were measured between 2009 and 2015 and are reported in a previous study (Smith et al. 2016). Samples were collected, on average, once per month in the Odei basin as part of the regional hydrometric monitoring program. The uncertainty in the streamflow data are ±10% on average, with higher uncertainty during peak flow and ice-on periods, while stable isotope data have a precision of ±0.2‰ ($\delta^{18}O$) or ±1.0‰ ($\delta^{2}H$) (Smith, et al., 2016).

3.3.3. Model

We apply an isoWATFLOOD^{TM} model (Stadnyk et al., 2013) of the Odei River basin: an isotope-enabled hydrologic model uniquely capable of regional watershed analysis at an hourly (or less) timestep across large regions in a spatially distributed framework. This affords the

opportunity to examine flow path partitioning outcomes on time and spatial scales that are relevant to source water generation processes.

The WATFLOOD setup divides the basin area into 10 km by 10 km grid cells and utilized land cover data to further subdivide grids into grouped response units (GRU). For the Odei model, six GRU were defined to represent hydrologically distinct land covers, reducing the number of model parameters compared to the previous implementation (Holmes 2016). Combined land classes were chosen based on shared hydrologic properties: coniferous (tall vegetation, constant leaf cover), mixed wood (tall vegetation, seasonally varied leaf cover), shrub (short vegetation), wetland (surface water for at least part of the year, vegetated), water (surface water, no vegetation) and impervious. Grouped land covers for the basin (in percent) were coniferous (30.3), mixed wood (9.9), shrub (29.1), wetland (22), water (8.5) and impervious (0.0).

Forest, shrub, disconnected wetland, and impervious classes are general land GRU having three connected soil storages. Wetland and water classes are treated differently. Connected wetland GRU were defined as 25% of the wetland area, and accumulate outflow from the land GRU. The water class is defined as part of the stream network, with bidirectional flow being possible between the connected wetland and water GRU.

The isoWATFLOOD model was applied to simulate the water and isotope mass balance for all hydrologic storages and fluxes at an hourly, or sub-hourly routing, time step. The monthly average isotopic composition of precipitation is applied as the precipitation signature for every timestep in the model within that month to calculate the incoming isotopic flux. Isotopes are completely mixed through storage depths, and fluxes (other than evaporation) are assigned the same concentration as the source storage. Evaporative fluxes (from soil, wetland or water storages) cause fractionation. A schematic of the model can be found in the supporting information (Figure B1) and process descriptions are provided in Stadnyk and Holmes (2020).

3.4. Methods

3.4.1. Search algorithm

We leverage advances in adaptive search algorithms and isotope hydrology to apply DDS for SO optimization, and PA-DDS for MO optimization (Asadzadeh & Tolson, 2013; Tolson & Shoemaker, 2007) of the isoWATFLOOD model. These algorithms were chosen based on their

efficiency in calibrating highly parametric models, which was prioritized based on the high computational requirements of the distributed hydrologic model. Both are available in a publicly available in the systems analysis toolkit, OSTRICH v17.12.19 (2018).

In SO model calibration, one optimal solution is expected out of each optimization trial. DDS is a stochastic global search single-objective optimization algorithm designed to find a 'good quality' solution within user-specified number of model evaluations (Tolson & Shoemaker, 2008). DDS is effective in calibrating highly parametrized hydrologic models, and the algorithm does not require any algorithm parameter tuning (e.g. population size) prior to calibration (Arsenault et al., 2014; Tolson & Shoemaker, 2008).

In MO optimization with conflicting objectives, a single solution cannot optimize all objectives simultaneously. Instead a Pareto (tradeoff) front exists between the objectives, whereby an improvement in one objective comes with degradation in the other. A solution that is better than another in all objectives is called a dominating solution. The goal of solving a MO optimization problem is to identify a set of non-dominated or suitable compromise solutions. PA-DDS is a stochastic global MO optimization algorithm that uses the DDS search algorithm and archives all non-dominated solutions (Asadzadeh & Tolson, 2013). MO optimization has, to date, not been applied in the context of isotope and flow optimized model calibration using a spatially distributed model applied across a regional scale with significant hydrological complexity or diversity. Applying the PA-DDS algorithm for MO optimization in this study provides the opportunity to assess the amount and degree of error trade off in flow performance required to achieve an acceptable isotope simulation, and the shape of that trade off front. This is valuable information for modelers to determine if the cost in flow performance (to achieve acceptable isotope error) is worthwhile, and what the total cost of this tradeoff will be. The PA-DDS algorithm performs as well as other MO algorithms, stores more non-dominated solutions than the previously applied NGSA-II algorithm (Fenicia et al., 2008), and, like the DDS algorithm, requires no algorithm tuning (Asadzadeh & Tolson, 2013; Yang et al., 2000).

3.4.2. Objective functions

Traditional hydrologic model calibration aims to minimize the residual between measured and simulated values, often for a single type of model output, e.g. streamflow versus hydrometric data from the real system (Moriasi et al., 2015). Quantitative metrics to compute model error in a

single value are most useful for calibration as they can be used in objective functions for automated algorithms. Two common efficiency metrics are used to distinguish between parameter sets: Nash-Sutcliffe (NSE) and the Kling-Gupta (KGE) efficiencies (Bennett et al., 2013). KGE combines three performance metrics into a single error measure to assess the simulation mean, variability and timing with respect to the observations; making it suitable for evaluating both streamflow and isotope performance. The KGE, with equal weighting for all error components, is calculated as (Gupta et al., 2009):

$$KGE = 1 - \sqrt{\left(\frac{\sum_{i=1}^{n}(x_{o,i} - \overline{x}_o)(x_{s,i} - \overline{x}_s)}{\sqrt{\sum_{i=1}^{n}(x_{o,i} - \overline{x}_o)^2}\sqrt{\sum_{i=1}^{n}(x_{s,i} - \overline{x}_s)^2}} - 1\right)^2 + \left(\frac{\sigma_s}{\sigma_o} - 1\right)^2 + \left(\frac{\overline{x}_s}{\overline{x}_o} - 1\right)^2}$$

Where $x_{o,i}$ represents observation i and $x_{s,i}$ is the simulated value at the time observation i was taken. The KGE ranges from 1 to -∞, having a value of 1 for a model having an ideal value for all three components. An ideal value (of 1) is not expected for a simulation, but the value of the KGE will never exceed the error in any one of the components. A good or acceptable value of the KGE metric can therefore be determined based on the maximum allowable error in the bias, relative variability, or correlation of a simulation.

Different model calibration problem formulations conducted in this study are compared in Table 3-1. As a baseline, a single-objective flow-only calibration (SO_F) is performed by DDS to maximize KGE_Q for streamflow across the calibration period. In all other problem formulations, isotope observations are used as auxiliary datasets. Isotope observations in the study area are discontinuous and biased toward the post-freshet, open water season. Simulated data for isotope evaluation, KGE_{18O} and KGE_{2H}, are therefore limited to days having isotope observations. Using only a subset of the simulated isotope time series in the evaluation metrics removes periods without isotope observations from the calibration. The process-based structure of isoWATFLOOD™, however, can constrain the isotope simulation for days or months without observed isotope composition, as demonstrated in Holmes (2016). Isotope KGE (KGE_{18O}, KGE_{2H}) were combined with KGE_Q in weighted average, SO functions: SO_F+O uses equal weights for the KGE_Q and the KGE_{18O}, and SO_F+O+H uses equal weights for KGE_Q and the average of the KGE_{18O} and the KGE_{2H}. Both isotopes were tested because, while the isotopes are

correlated, seasonal evaporative fractionation from lakes and wetlands leads to some additional information content from the second isotope (Figure B2). Isotope model performance was also incorporated into SO calibration as a constraint; the simplest method was to add a penalty for constraint violations, which here was an isotope KGE below an acceptable limit (Mallipeddi & Suganthan, 2010). Acceptable limits were set based on Moriasi et al. (2015), penalizing solutions with an absolute bias > 25% (general satisfactory limit), using a KGE minimum of 0.75. Two penalized objectives were used: a penalty of $2(0.75\text{-}KGE_{18O})$ for $KGE_{18O} < 0.75$ in SO_F-O, and penalties of $2(0.75\text{-}KGE_{18O})$ and $2(0.75\text{-}\ KGE_{2H})$ for $KGE_{18O} < 0.75$ and $KGE_{2H} < 0.75$ in SO_F-O-H. Finally, a MO calibration (MO_FOH) was performed to maximize three objective functions: KGE_Q, KGE_{18O} and KGE_{2H}.

Differences between SO and MO optimizations, or single and dual isotope simulations were assessed using a two-tailed student t-test ($p<0.05$ significant). If the differences in the methods of including isotope data into the model calibration are insignificant, they will be grouped into two categories: isotope optimized (FIO), and flow calibration that was isotope-bounded (FO-IB).

Table 3-1: Summary of MO optimization functions for this study.

Reference name	Isotopes	Optimization function	Penalty
SO_F	-	KGE_Q	-
SO_F+O	$\delta^{18}O$	$0.5KGE_Q+0.5KGE_{18O}$	-
SO_F+O+H	$\delta^{18}O$, $\delta^{2}H$	$0.5KGE_Q+0.25KGE_{18O}+0.25KGE_{2H}$	-
SO_F-O	$\delta^{18}O$	KGE_Q	KGE_{18O} <0.75
SO_F-O-H	$\delta^{18}O$, $\delta^{2}H$	KGE_Q	KGE_{18O} & KGE_{2H} <0.75
MO_FOH	$\delta^{18}O$, $\delta^{2}H$	KGE_Q, KGE_{18O}, KGE_{2H}	-

3.4.3. Model Calibration setup

The Odei River model was calibrated using the OSTRICH optimization toolkit, which has options for both the DDS and PA-DDS algorithms, as well as weighted average and penalized objective functions (Mattot, 2018). OSTRICH generates and compares the performance of different parameter sets; a custom error output was added to WATFLOOD to write a new output

file that includes KGE error metrics for both flow and isotope simulations. Five calibration trials having different random seeds were performed for each of the six optimization functions at 10,000 model evaluations per trial. WATFLOOD is a highly parameterized model having individual parameter values for each GRU class. Sensitive parameters for all modeled storages and fluxes without pre-defined (recommended) parameter values were included in the calibration (Kouwen, 2018). Upper and lower limits for parameter values were set based on previously published limits within the same eco-zone (Delavau et al., 2017). Initial values were randomly generated within the specified limits. The 42 calibrated parameters, their associated modeled storage, and their upper and lower bounds are provided in the supplemental material (Table B1).

The model was calibrated to maximize KGE values for the Odei River streamflow gauge from 2009 to 2015. Daily flow data for the entire calibration period were used to calculate the KGE_Q, while isotope data were limited to the 2011 to 2015 period. Model validation from 2002-2008 used only streamflow data, as isotope simulation in this study was meant only to supplement calibration. Average annual evaporation, transpiration, total evapotranspiration, and soil moisture were calculated for the Odei basin from simulated totals across all cells and GRU for pseudo-verification against the GLEAM v3.3a reanalysis dataset at the point closest to the centroid of the Odei basin. GLEAM data estimate evapotranspiration and its components from a combination of reanalysis climate data and satellite data (Martens et al., 2017; Miralles et al., 2011). No point-based observations of these data were available for further verification.

3.4.4. Model Prediction Bound Approximation

Following Tolson and Shoemaker (2008), model parameter sets selected from the model calibration results deemed to be behavioral and independently generated; therefore, they were were used to estimate model prediction bounds. These prediction bounds give an estimate of the model parameterization uncertainty, assuming that all other sources of uncertainty (i.e., model structural arrangement, inputs and outputs) remain unchanged. Each optimization trial for the five SO model calibration problems produced a single parameter set, resulting in a total of 25 independently generated parameter sets that were deemed to be ‘satisfactory’ (Moriasi et al., 2015). A total of 20 parameter sets were produced when isotope tracers were used in performance evaluation, and five solutions were produced by the model calibration problem that exclusively considered streamflow performance.

Each of the five trials of the MO model calibration produced a set of non-dominated solutions (i.e., solutions providing a suitable compromise between the three objective functions). Non-dominated solutions from the five trials were combined. A subset of non-dominated solutions selected for further analysis based on one objective criteria for isotope-enabled calibration: the five non-dominated solutions that best met each objective were selected. The mean and range (minimum/maximum) of each parameter value were compared among selected solutions with and without isotope tracers to elucidate the impact of isotope tracers on model (parameter) uncertainty. Moreover, differences in internal processes (i.e., flow paths) using these two different sets of solutions were analyzed to examine the information content added by the isotope tracers for the model calibration purposes.

3.5. Results

3.5.1. Calibration results

The SO optimizations and the five MO PA-DDS optimization trials all produced parameter sets which perform well in the calibration period; the number of model evaluation per trial was therefore either sufficient or more than sufficient. There were tradeoffs that occurred between (a) ^{18}O and streamflow KGE values, and (b) ^{18}O and ^{2}H KGE values (Figure 3-2). Additional seed values would likely result in some variation in the final parameter values, as would be expected for any highly parameterized model subject to equifinality (Beven 2011), but this would be unlikely to impact model convergence based on the tight clustering of calibration KGE observed.

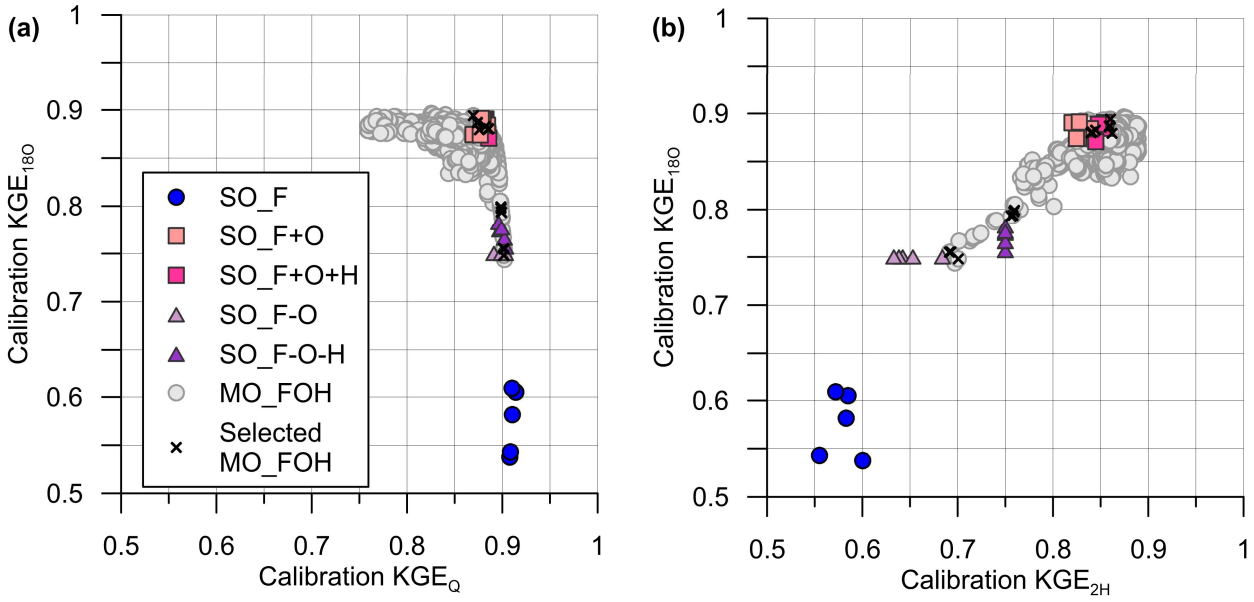

Figure 3-2: Calibration objective tradeoffs for (a) streamflow and $\delta^{18}O$ KGE, and (b) $\delta^{2}H$ and $\delta^{18}O$ KGE.

SO_F, which considered only KGE_Q, not surprisingly resulted in the highest streamflow performance and worst isotope simulation performance (bottom right corner of Figure 3-2a, and bottom left Figure 3-2b). The maximum achieved KGE_Q exceeded 0.9, which indicates a very good performance in flow simulation; however, all isotope KGE values were below 0.65. Achieving optimal streamflow performance from a single (flow-based) objective calibration was expected because the model is only required to satisfy one constraint: no tradeoff is imposed for the model to meet other performance criteria (Nelson et al., 2018; He et al., 2019).

Penalized objective functions (SO_F-O and SO_F-O-H) produced isotope simulations at the acceptable lower limit of 0.75 (Figure 3-2b), while weighted average functions are concentrated to the upper right of the tradeoff front (i.e., where isotope and flow simulation performances are both good). The isotope-penalized solutions exhibited slightly better flow performance, at the expense of isotope performance.

MO_FOH non-dominated solutions represent SO functions including isotope data, and a range of performance outcomes between, or having better, isotope performance (Figure 3-2a). The sub-set of solutions selected from PA-DDS had similar KGE values (for both streamflow and isotopes) to the SO solutions. Tradeoffs between the two isotope simulations were more limited (than flow

and isotopes) as isotope simulation performance was more closely correlated between the two isotope errors than to the streamflow simulation (Figure 3-2b).

Calibrations including both isotopes (SO_F+O+H and SO_F-O-H) had somewhat better isotope performance than calibrations that used only one isotope. All optimization functions that included an isotope error metric were able to reproduce the observed isotope data significantly better than SO_F simulations. Significant differences in flow performance resulted from the choice of objective function; however, differences between SO and MO optimizations, or single and dual isotope simulations were not statistically significant (p-values ranged from 0.25 to 0.99), in either calibration or validation. The time-series flow and isotope simulation results for the calibration period are shown in Figures A3 and 3-5abc.

3.5.2. Final parameters

While there was little variation in streamflow KGE in model calibration, many of the final calibrated parameter values varied not only between each optimization method, but also among each calibration trial (Table 3-2), illustrating the degree of equifinality in the model and parameter uncertainty.

Table 3-2: Summary of mean parameter values and percent usage of the specified possible parameter range; parameters with varying values for different land classes have been summarized using a single, area-weighted value and are indicated with an asterisk.

			Flow-only Optimization (FO)		**Flow Optimized & Isotope-Bounded (FO-IB)**		**Isotope Optimized (IO)**	
Storage	**Parameter**	**WATFLOOD**	**Mean**	***Range (%)***	**Mean**	***Range (%)***	**Mean**	***Range (%)***
Lower zone	Baseflow power	pwr	1.58	*6.7*	1.72	*23.4*	2.24	*48.5*
	Baseflow coefficient	coeff	3.19E-03	*40.7*	1.28E-03	*72.7*	3.20E-04	*21.5*
Channel	Channel roughness	r2n	5.85E-03	*18.9*	5.12E-03	*24.3*	2.71E-03	*21.1*
Connected wetland	Porosity	theta	0.19	*12.9*	0.16	*6.8*	0.19	*14.4*
	Conductivity	kcond	0.248	*2.7*	0.250	*2.2*	0.248	*5.3*
Surface	Infiltration*	ak	23.0	*17.6*	27.1	*55.2*	26.1	*32.9*
	Infiltration, frozen*	akfs	28.8	*34.2*	40.6	*46.9*	34.9	*57.6*
Upper zone	Retention capacity*	retn	43.6	*3.3*	53.7	*16.3*	81.3	*15.6*
	Conductivity (H)*	rec	0.81	*12.9*	0.58	*20.2*	1.38	*55.5*
	Conductivity (V)*	ak2	5.24E-02	*12.0*	5.62E-02	*42.3*	1.46E-01	*51.7*
	Conductivity, frozen	ak2fs	4.90E-03	*0.6*	7.06E-02	*91.2*	1.27E-01	*99.1*
Snowpack	Melt factor*	fm	7.62E-02	*21.1*	6.57E-02	*38.9*	8.67E-02	*48.3*
	Base temperature*	Base	0.77	*34.7*	0.52	*40.6*	0.84	*29.0*
	Sublimation rate*	sublim_rate	2.07E-01	*8.6*	2.37E-01	*15.2*	1.98E-01	*19.2*
Open water	PET to AET factor	fpet	0.73	*85.1*	0.45	*67.4*	0.42	*11.3*

Flow-only (FO) calibration produced the narrowest range of parameter values for most parameter types, although the difference in parameter range between objective functions was often small. Connected wetland porosity (*theta*) and conductivity (*kcond*) parameters, and channel roughness (*r2n*) were unaffected by the inclusion of isotope data and were generally similar among

optimization functions. Parameters related to soil water storage or flux and evaporation, however, exhibited consistent differences among the different model calibration problem formulation. Weighted average functions had a wider calibrated range and higher mean value for soil water parameters (*retn, rec, ak2*), while the isotope-bounded (FO-IB) calibrations resulted in narrower ranges, with FO calibrations having the most limited range and lowest mean parameter values.

The reverse was true for the evaporation factor used to adjust the Hargreaves estimated PET to AET (*fpet*); the most identifiable parameters (i.e., narrowest range) resulted from the isotope optimized (FIO) calibrations. Changes in parameter mean and range as a function of the calibration methodology are visualized on Figure 3-1Figure 3-3abc. Subsequent changes in internal storage (hydrologic partitioning) and flux magnitude associated with parameter values are visualized on Figure 3-3def. Increased values for soil-related parameters in isotope-calibrated models resulted in greater soil moisture retention and faster recharge from upper to lower soil zones (Figure 3-3ef).

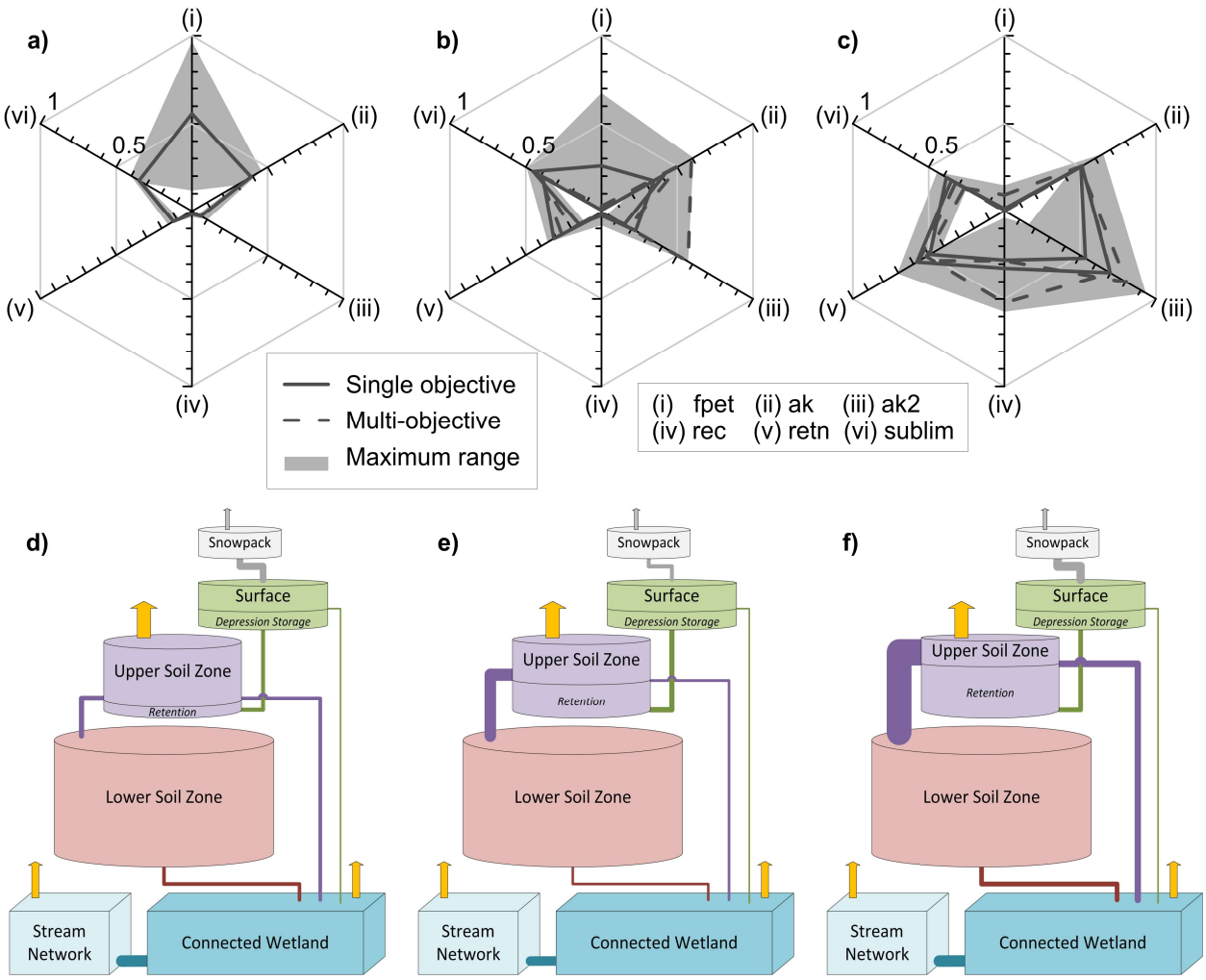

Figure 3-3: Mean parameter values and ranges, and scaled model connectivity (based on area-weighted mean parameter values and relative flux contributions) for flow-only calibrations (a) and (d), isotope-bounded calibrations (b) and (e), and isotope and streamflow optimized calibrations (c) and (f).

Including isotopes in the calibration had a significant impact on water partitioning within the model. When isotope performance was ignored (Figure 3-3a and d), optimal performance resulted in low soil retention capacity and soil conductivity, limited transpiration, and a high mean (and variable) evaporation adjustment factor (i.e., high evaporation loss). When isotope performance was included in calibration as a penalty (Figure 3-3b and e), the mean and range of soil water parameters increased, and the mean and range of the evaporation factor decreased. When isotope performance was optimized along with flow (Figure 3-3c and f), a pronounced and distinct shift between the evaporation and transpiration parameters was observed. Adding new

information into the calibration process resulted in improved identifiability of the evaporation parameter (i.e., a more precise value), and a decrease in the precision of the transpiration parameters. The range of the soil moisture retention (*retn*) parameter, the primary control on the transpiration rate, increased and no longer overlapped with the range identified from the flow-only calibration. All calibrations were sensitive to *retn*, however, each had a different preferred value. Sublimation parameter values were generally not impacted by the addition of isotope data to the model calibration.

3.5.3. Streamflow and isotope performance

The model was validated using flow data from 2002-2008 with model flow performance assessed using three error metrics: Nash-Sutcliffe efficiency (NSE), Kling-Gupta efficiency (KGE), and percent volume error (PVE) (Figure 3-4Figure 3-4). Validation including isotope data would be desirable, but there were insufficient data collected for statistically significant calibration and validation periods; therefore, the choice was made to perform flow-only validation.

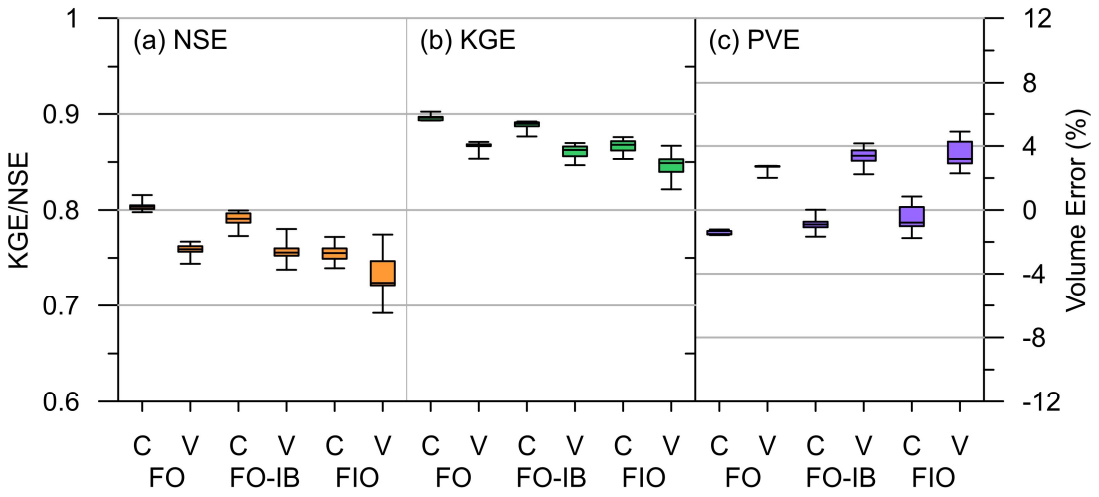

Figure 3-4: Boxplot comparison of (a) NSE, (b) KGE and (c) PVE for the calibration (C) and validation (V) periods, for the flow-only optimizations (FO), flow optimization with isotope boundaries (FO-IB) and flow and isotope optimizations (FIO) at the Odei River gauge, with whiskers extending to maximum and minimum values.

All calibrated parameter sets performed well during both calibration and validation at the basin outlet (05TG003), producing NSE values > 0.65, and KGE values > 0.8. Validation performance

was not as strong as calibration (as anticipated), but differences were typically small (Figure 3-4ab). Streamflow performance at the uncalibrated, upstream gauge (Sapochi River, 05TG006) was also acceptable, but lower than at the calibrated gauge (Figure B4).

Calibrations including isotope data did not perform as well as flow-only calibrations during the calibration period, but their performance difference between calibration and validation periods was significantly smaller (i.e. mean NSE decrease of 0.05 for FO calibrations, versus 0.03 and 0.02 for FO-IB and FIO, respectively, with p-values 0.04 and 0.001). There was no statistically significant difference in NSE and KGE metrics between the FO-IB and FO calibrations during the validation period (p-values 0.58 and 0.25).

Percent volume error for both simulation periods was small relative to the estimated uncertainty in the observed streamflow record (±10%). Volume errors were primarily dependent on the simulation period rather than the calibration methodology, with small underestimation of volume during the calibration period, and slight overestimation during validation for all parameter sets. There was no significant difference in volume between the FIO and FO-IB calibrations in either the calibration or validation periods.

Differences in the streamflow hydrographs were minor for both the calibration (Figure 3-5c) and validation (Figure 3-5d) periods. The range of uncertainty about the mean of each optimized flow series changed significantly among the calibration methodologies, with isotope calibrations generally having wider ranges of simulated flow, particularly around peak flow events. Inclusion of isotopes had the most significant impact on hydrograph uncertainty during the freshet or peak flow period (early April), and the least impact during flow recession (late summer and fall) and ice-on low-flows (winter).

Isotope performance was assessed from 2011-2015 during the period of calibration when data were collected. The isotope time series for $\delta^{18}O$ (Figure 3-5a) and $\delta^{2}H$ (Figure 3-5b) also the most uncertain during the annual spring freshet, however, including isotopes in the calibration generally decreased the uncertainty in isotope simulation. The FO calibration consistently produced the most uncertain and most enriched isotope simulations.

The coarser resolution isotope in precipitation input used in this study (by necessity based on the large, remote domain) contributes some error to the simulated isotope performance and reduce

simulated variability of the isotope in streamflow output. Travel times through the watershed are sufficiently long, however, that there is natural attenuation of the input-output transfer function between the isotope in precipitation and streamflow signal. A monthly average isotope in precipitation input resolution can appropriately describe the precipitation variability in streamflow at meso-scales (e.g., Stadnyk et al. 2013, Delavau et al. 2017). Despite the monthly averaged inputs, the model satisfactorily captures the variability in seasonal processes (e.g., evaporative enrichment, snowmelt depletion; Fig 5a,b, S3).

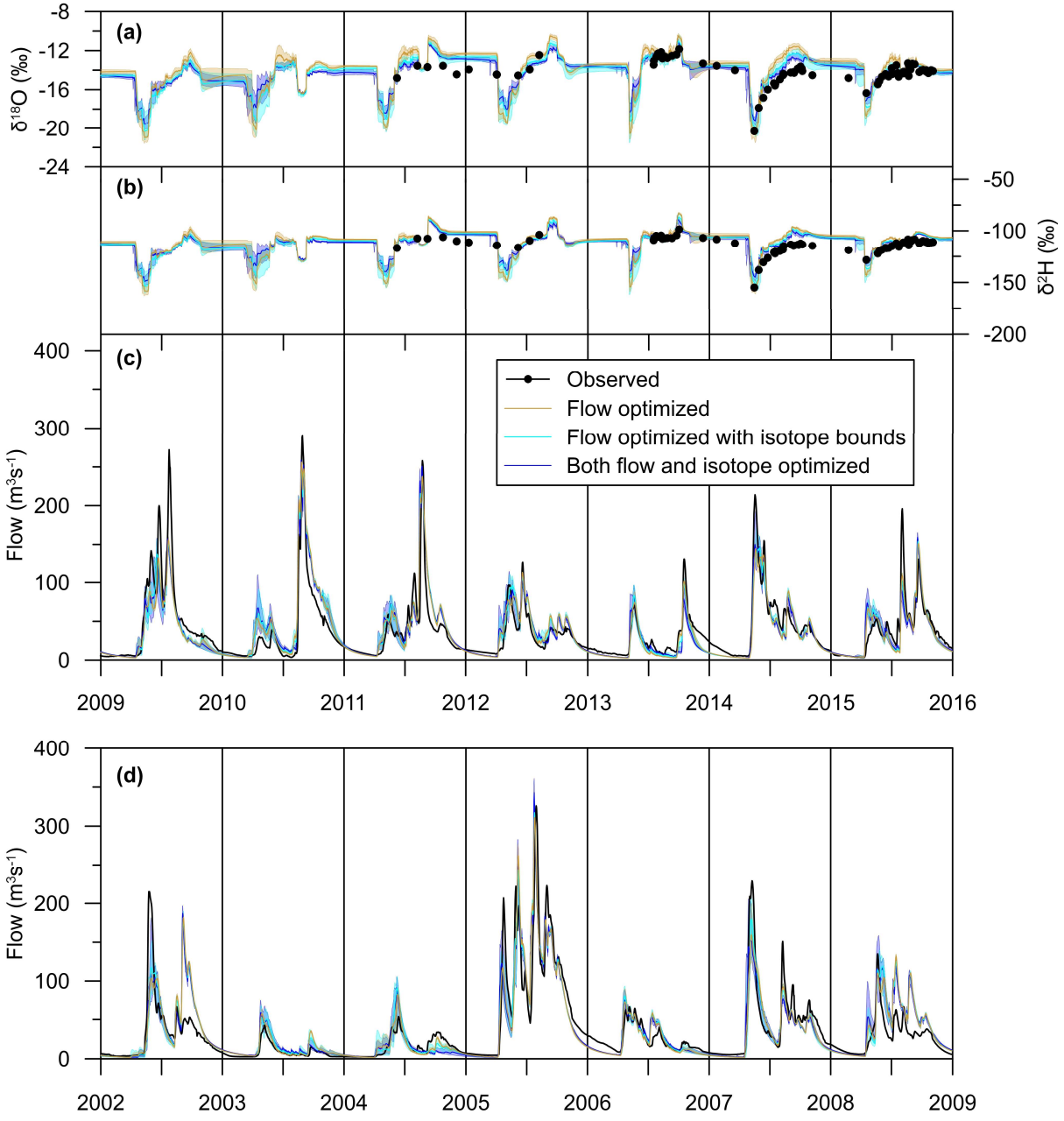

Figure 3-5: Time series plots of (a) $\delta^{18}O$, (b) $\delta^{2}H$ and (c) streamflow for the calibration period (2009-2015) and (d) streamflow for the validation period (2002-2008) at the Odei River gauge. Bounds are maximum and minimum simulated output

3.5.4. Hydrologic Partitioning

Given all calibrated parameter sets simulated total water volume accurately (Figure 3-6a), changes in parameter values represent a change in the internal distribution of water storage (and flux), or hydrologic partitioning within the model. Overland flow from surface storage

contributing to total streamflow was low and largely consistent among calibration methods (Figure 3-6b), with occasional contributions associated with large precipitation events evident during ice-off periods. Streamflow contributions from interflow (Figure 3-6c) and baseflow (Figure 3-6d) were strongly affected by the calibration methodology, but not in any consistent direction. When output from each calibration trial was averaged, the inclusion of isotope data produced lower interflow contributions from lateral flow out of the upper soil zone, and higher contributions from lower soil zone outflows. There was also an increase in the variability of baseflow contributions to streamflow (relative to interflow) associated with inclusion of isotope error in the calibration. However, the primary difference produced by the inclusion of isotope data is an increase in the variability of baseflow versus interflow contributions between calibration trials. The range for both interflow and baseflow contribution was four times greater during the freshet period for the isotope-optimized (FIO) calibrations, relative to FO calibrations.

The average annual evaporation flux was nearly identical between the isotope-informed calibrations (Figure 3-6e), while evaporation from the FO calibrations was 21% higher, on average. The absolute difference in the simulated transpiration flux (Figure 3-6f) was like that for evaporation, however, but smaller for the FO calibrations produced less flux. Since transpiration dominated simulated evapotranspiration in this study basin, the relative difference in transpiration between calibration methods was small.

Recharge from the upper to the lower soil zone varied between calibration methods, not by cumulative values amounts (which share similar annual ranges), but by daily flux rates (Figure 3-6g). As expected from the model connectivity (Figure 3-5def), FO-calibrated models had steady recharge rates during wetter periods, while the isotope calibrated models had much higher recharge rates during (and following) rain or snowmelt events and longer, more frequent periods with minimal recharge. Average annual soil moisture in the models with both isotopes and flow optimized (FIO) were twice that of the FO calibrated models (Figure 3-6h).

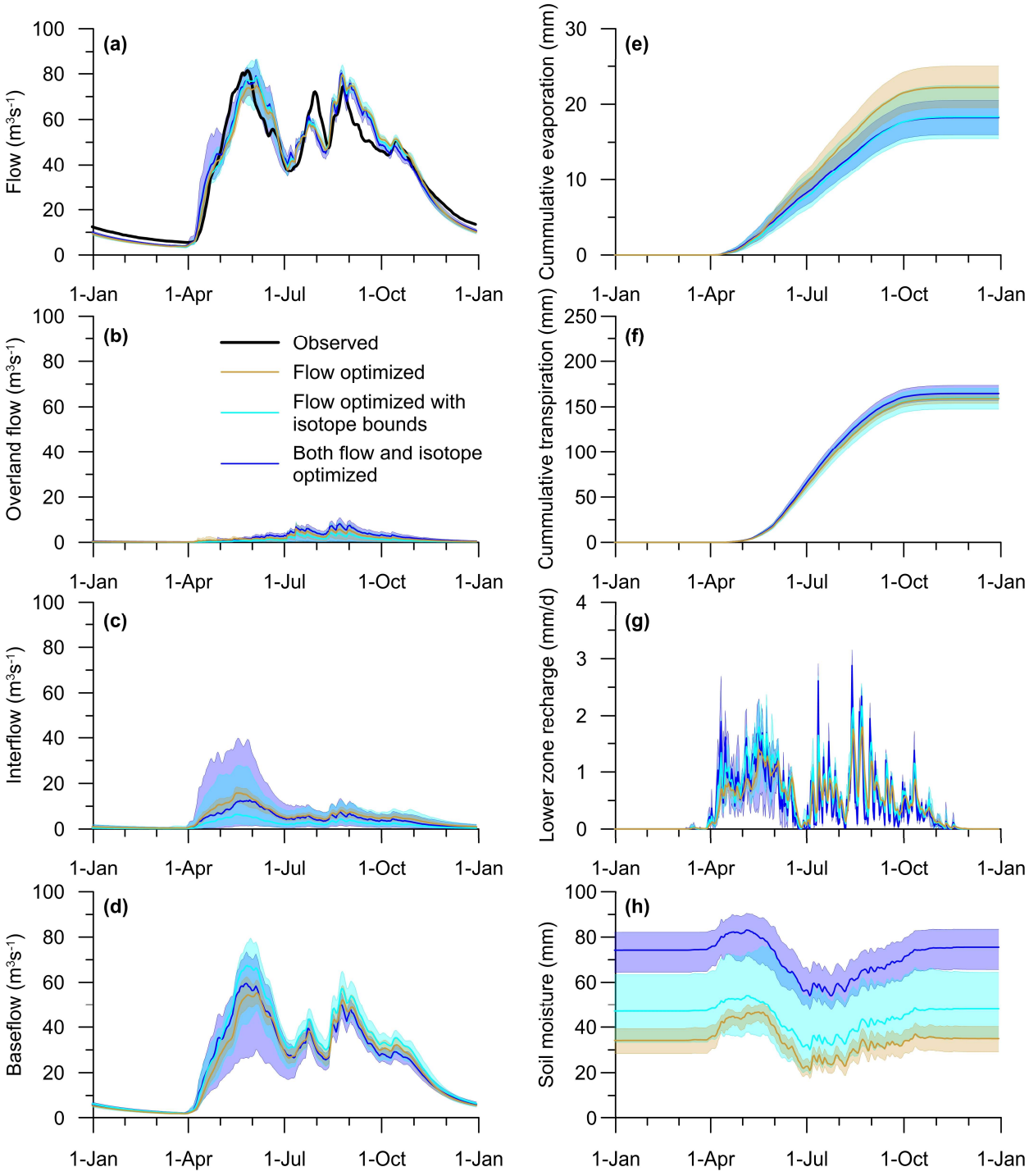

Figure 3-6: Average annual plots for the simulation period (2002-2015) of (a) streamflow, (b) flow from surface runoff, (c) flow from upper zone storages, and (d) flow from lower zone storages at the Odei River gauge, and (e) cumulative evaporation, (f) cumulative transpiration, (g) recharge from the upper to lower soil zones, and (h) average soil moisture over the Odei basin. Bounds are maximum and minimum simulated output.

3.6. Discussion

3.6.1. Tradeoffs when calibrating with isotope data

There is a distinct tradeoff between flow and isotope objectives in MO calibration: improvements to isotope simulation performance degrade flow simulation performance to some extent. These results are consistent with previous studies incorporating isotope error into model calibration (Birkel et al. 2010; 2015). In this study, Pareto optimal solutions include parameter sets that result in a minimal decrease in streamflow performance in the calibration period and a substantial increase in the isotope simulation performance (Figure 3-2a). Using an additional isotope (here, δ^2H) in calibration does not affect streamflow simulation performance, but inclusion did improve the isotope simulation performance. This may not always be the case, however, and will depend on a model's ability to accommodate additional constraints imposed during calibration. For example, Fenicia et al. (2008) saw a more significant degradation in streamflow performance when isotopes were added to an NSGA-II MO model calibration. The extent to which introducing isotopes into calibration degrades model performance likely depends on a model's ability to alter internal flow path contributions and still preserve total streamflow contribution. A predictive measure for this outcome may be the characteristic structure of the model (i.e., lumped versus distributed), which is a key difference between the Fenicia et al (2008) study and our work using a spatially distributed, process-based model.

The two isotopes ($\delta^{18}O$ and δ^2H) are strongly correlated in this basin (Figure B2), as is the model performance in simulating the two isotopes (Figure 3-2b) and therefore, not surprisingly, our results support the use of a single isotope in future automated calibrations. The distinct seasonal evaporative enrichment in the basin was simulated even with only one isotope in the calibration objective. Dual-isotope simulation can still be applied for model validation or process verification using a simulated isotope framework approach (Stadnyk & Holmes, 2020). It should be noted that we did not test the impact of simulating deuterium-only error.

The application of a Pareto-archiving calibration algorithm is novel, and of value when calibrating with competing isotope and flow objectives. The MO calibration trials found parameter sets of similar performance quality to the single-objective calibration trials that include isotope data, as well as parameter sets with intermediate or superior isotope simulation performance. The MO model calibration may not improve the final model performance relative

to a SO calibration, but it does allow for an informed choice between the tradeoff between two types of simulation, or multiple objectives. This capability is particularly useful when calibrating to observational data of differing quality or temporal resolution (e.g., flow and isotopes). The lower confidence in isotope simulation performance (due to sparse observed data records) can be accounted for with increased flexibility by selecting an optimal solution on a case-by-case basis if the modeler is capable of, and interested in, making these judgements.

During validation the statistical performance of all flow simulations is comparable, regardless of the calibration methodology, which negates the apparent advantage of FO calibration that arose during the calibration period (Figure 3-4). Outside the calibration period there are no particular advantages or disadvantages of including isotope data in the model calibration based on streamflow performance metrics. The consistency of good streamflow simulation performance is likely due to the inclusion of a flow performance metric in all calibration methods. It is noted, however, that the validation period was very similar to the calibration period (in both hydrometeorological condition and duration). A longer validation period, or one with more dissimilar conditions, may result in more differentiation between the calibration methodologies (Stadnyk & Holmes, 2020). The results of our study are also dependent on the chosen objective functions; using additional or alternate performance error metrics would likely affect the selection of the final parameter sets (Lilhare et al. 2020).

Improvements in isotope simulation performance come at the expense of the calibration efficiency, as simulating isotope mass concentrations in addition to streamflow results in longer model run times, with the model taking 90 to 130% longer to run with one or two isotope simulations, respectively, than without. The DDS algorithm used in calibration is effective at finding good local optima, as evidenced by the final streamflow performance statistics. The 10,000 model evaluations were more than enough to identify parameter sets producing satisfactory KGE scores (Birkel et al. 2010), consistent with the findings of Asadzadeh et al. (2014). The advantage of using MO calibration when including isotope data is also apparent from a model efficiency point of view: PA-DDS is nearly as fast as DDS. The Pareto-archiving MO calibrations provide more information to the modeler at a lower relative computational cost. Based on this study, MO calibration using PA-DDS or a similar algorithm is the most efficient choice for integrating multiple data types into model calibration. MO calibration identifies the

tradeoff between different objectives, produces a variety of good parameter sets that are equivalent to those found from SO calibration, and does not require a predetermined relationship among the objectives. So long as the tradeoff between calibration objectives is of interest, the slight increase in computational time (i.e., less than 1% of the total calibration time) is more than compensated for by the increased information content available to the modeler.

3.6.2. Impact on simulation uncertainty and parameter identifiability

There is no statistically significant difference in streamflow performance during validation due to the choice of objective function; all calibrated parameter sets produce good simulation results in the validation period. The parameter sets and flow simulations resulting from flow-only calibrations are all similar: both total and partitioned flows are consistent among all calibration trials, and the majority of calibrated parameter values are well identified (i.e., have similar final values). In contrast, isotope-informed calibrations result in substantial variation in the partitioned flow components among the calibration trials, and a wider range of (i.e., more uncertain) total streamflow. Adding isotope simulation performance to the model calibration reduced the identifiability of many calibrated parameters, which is a common side effect of introducing tracer metrics into calibration (Birkel & Soulsby, 2015). Here the impact was most notable on parameters connected to soil storage. Only the open water PET to AET adjustment factor became more identifiable when isotope error was added to the calibration, due to the relatively high sensitivity of the isotope simulation to fractionating evaporative processes.

It is unlikely that the increased simulation variability from isotope-informed calibration is coincidental. Increasing the weight given to isotope performance during the model calibration necessitates a broader exploration of the parameter space to satisfy multiple objectives. When the model is only required to satisfy one objective (e.g., flow-only), there is no (mathematical) reason to substantially deviate from a locally optimal solution. As more diverse constraints are imposed on calibration, more of the parameter space is explored to find acceptable solutions. We see this in the progressive increase in uncertainty resulting from increased calibration complexity from the flow-only to isotope-bounded, and then finally the isotope-optimized calibrations.

Results suggest that including both flow and isotope data in model calibration can mitigate over-fitting of the model. The FO trials all identify a limited area of the parameter space that result in very high KGE (error metric) values for the optimization function. Outside of the calibration

period, however, simulation performance is significantly worse, suggesting the model is not robust to minor variations in hydroclimatic conditions (Beven 2011). The isotope-informed calibrations demonstrate less difference between calibration and validation periods in terms of model performance, an indication that the model is more robust. A different flow simulation error metric would likely result in the identification of an alternate, but equally limited, region of the parameter space. A MO calibration using only streamflow data might also avoid over-fitting. However, in the same way that little value was added by a second isotope in this study (δ^2H), additional flow metrics could be highly correlated and reduce the value of the multi-objective calibration. We find the real value of MO calibration here to be in the distinct and differing nature of the objectives the model must satisfy (i.e., a tradeoff between overall streamflow and model internal processes partitioning), which agrees with the findings of Nijzink et al. (2019).

Given all the streamflow simulations performed equivalently in validation, the FO calibration methodology clearly results in a misleading impression of the model prediction uncertainty. The differences in uncertainty here stem directly from the propagation of parameter uncertainty into the streamflow, which is typically reported as being 'small' relative to other sources (Chen et al. 2011). Modelers tend to place more faith in a model when it is precise, assuming that consistency is a surrogate for accuracy. The FO calibrations perform well and consistently identify a narrow range of soil parameters and partitioned flow components. The isotope-informed models, however, produced a wide range of partitioned flows and soil parameters, yet the isotope-informed calibrations also perform very well in validation.

Furthermore, a well-identified parameter is not necessarily a correctly identified parameter; the reliability of a model (or parameter) can be assessed using sharpness (i.e., how precisely a model replicates a result), where increasing the sharpness can negatively influence the accuracy, or reliability, of a model or parameter value (Zhou et al. 2016). Our results show that the FO calibrations consistently identify a narrow range of soil water retention values across all trials and all land classes. The isotope-informed calibrations, however, disagree with the mean or range of these well-identified values. Calibrations optimized for both flow and isotope performance result in values for soil water retention that are substantially higher and do not overlap with the values resulting from the (well identified) flow-only calibration. These contrasting results suggest one, or both, of the models is incorrectly (but precisely, or unreliably:

Shafi et al., 2015) representing the subsurface distribution of water and resulting transpiration loss. This can have significant implications for long-term water supply projection and water security assessment under changing climates. Calibrating models with regard for process-based implications is advocated by Clark et al. (2016) to preserve confidence in model operational predictions.

3.6.3. Impact on flow path partitioning

Adding isotope data into the calibration process has little effect on the statistical performance of the flow model, but a substantial impact on the simulated internal hydrologic process and flow partitioning. This result is not unexpected based on previous literature on the importance of mixing volumes and internal flow path transit times (Iorgulescu et al. 2007; Fenicia et al. 2008). Our study, however, highlights that calibration methodological choices have a distinct and statistically significant impact on mixing volumes and therefore transit times. Processes affected by simulated fractionation (resulting from evaporation) and mixing volumes are consistently simulated differently when isotopes are included in model calibration. Isotope-informed calibrations result in lower evaporation rates, which reduces evaporative enrichment at the basin outlet (and across the model domain), but also result in higher transpiration rates, thus maintaining similar net evapotranspiration flux (and total water loss) to FO calibrations. The consistent net ET among all calibrations is not unexpected: all optimization functions included the streamflow bias (as a component of KGE_Q).

Parameter sets from the isotope-informed calibrations have twice as much soil moisture storage relative to FO calibrated parameter sets. The increase in basin-wide soil storage increases the mixing volume for the isotope simulations, thus decreasing the average simulated velocity of water through the soil layer. Outflow from the upper soil zone subsequently adjusts to the increased soil storage for the isotope-informed calibrations to preserve streamflow celerity in response to a precipitation event (i.e., small variation in total flow among all calibration trials). There are, however, multiple possible flow paths in the model structure that can alter local velocities while preserving the overall celerity (McDonnell & Beven, 2014). One strength of using a distributed, process-based hydrologic model is that local velocities can vary substantially, representing different transit times, which can be calibrated indirectly by adding isotopes to the model objective function. In this study and many others, there is insufficient observational data

to accurately determine the correct individual flow paths and velocities. Relative contributions of the upper and lower zones to streamflow varied widely among calibration trials, thus the correct separation remains indeterminate, even with the addition of isotopes. Simulated flow partitioning may be verifiable with higher temporal resolutions of isotope data (to distinguish local flow velocities), or by sampling isotope concentrations in soil water and/or xylem water (Knightly et al., 2020). It is notable that, even with a coarse resolution and irregularly sampled isotope data set, the isotope-informed calibrations produce results that are distinct from the flow-only calibrations, and which agree with the findings from previous modeling studies in much smaller, more intensively monitored, basins (e.g., Dehaspe et al. 2018; Tunaley et al 2017).

It is not possible to conclusively determine the process representation that is most correct since detailed field measurements of soil water content and/or evaporation and transpiration rates were not taken (or available) as part of the Lower Nelson stable water isotope monitoring network program. Recent studies, however, suggest that transpiration in the Boreal region should account for approximately 60% of total evapotranspiration water loss, which agrees with all calibration results presented in this study (Wei et al., 2017; Jasechko et al., 2013). Simulated estimates of evaporation and transpiration from isotope-optimized simulations agree more closely with GLEAM model results in the Odei basin than the FO optimized simulations (Table B2) (Martens et al., 2017; Miralles et al., 2011). GLEAM results and isotope-informed calibrations are also consistent with the in-situ measurements of soil moisture in the Canadian boreal forest region (0.1-0.2 m^3/m^3) published by Djamai et al. (2014), unlike the FO calibrations. Overall, we find the isotope-informed calibrations are in better agreement with other process data available for the watershed.

3.7.Conclusion

This study examines the impact different methods of incorporating stable water isotope data into hydrologic model calibration have on model parameterization and subsequent flow path partitioning, and streamflow simulation. We find that inclusion of isotopes in the calibration process significantly alters model parameter values, resulting in substantially different flow path partitions within the model - but with negligible change in total simulated streamflow. Isotope-informed calibrations have significantly different soil storage and evapotranspiration related parameters, which alters the relative contributions of evaporation and transpiration to the

simulated water balance but not overall evapotranspiration loss. Flow path partitioning has a discernable impact on local velocities and transit times through the soil zone, but negligible impact on streamflow celerity. The importance of transpiration indicated by the isotope-informed model results changes our understanding of the basin hydrology and can alter simulated watershed response to changing landcover or climatic conditions. These flow path differences are likely important for coupled hydrologic and water quality modeling, and for land use impact studies. In this study, there was no direct means of quantifying which internal flow path partitions are accurate, though comparison with alternative sources (i.e., GLEAM) of data for the basin indicates isotope-informed calibrations were likely more accurate.

The most efficient method, of those evaluated, to introduce isotope data into the model calibration was the PA-DDS algorithm, as it provides significantly more information content with a minimal increase in computational cost. The MO calibration identifies parameter sets capable of performing as well as those from SO calibrations that considered isotope data. The identification of error tradeoff between multiple objectives, here streamflow and isotope performance, facilitates more informed choices of optimal solutions. Different calibration objective function choices may lead to different optimal parameter selections, an uncertainty that has yet to be fully explored within this study.

This study demonstrates first-hand that precise but unreliable hydrologic partitioning (i.e., internal storage and flux) can in fact result in models despite strong streamflow performance, well above the traditionally used definition of ‘satisfactory’ performance reported in the literature. Strong model performance can be achieved from a highly varying distribution of hydrologic partitions; some of which are likely to be entirely infeasible. The traditional means of flow-only calibration is found to be insufficient for complex, highly parameterized hydrologic models, and we find that additional data are required to support accurate water balance partitioning within the calibration process for distributed, complex hydrologic models. Adding isotope tracers to the objective function produces equivalent model flow performance, yet very different (often opposing) internal water distribution. Though one is likely more correct than another, the real cautionary tale here is that the calibration of a spatially distributed, process-based model should not rely on one flow-based performance metric, and when it does, the reliability of the model is likely being overstated.

Acknowledgments, Samples, and Data

The authors declare no real or perceived financial or other conflicts of interest with respect to the results and publication of this paper. Funding for this study was provided by the Natural Sciences Engineering Research Council (NSERC) collaborative research and development grants program. We thank Manitoba Hydro for their financial and in-kind contributions to this research that facilitated collection and analysis of stable water isotope data. We acknowledge Environment and Climate Change Canada for the collection and dissemination of hydrometric and meteorological data, and physiographic data used for model setup. Data used in this study are publicly available and archived, as referenced. The isoWATFLOOD model code is freeware available on GitHub (https://github.com/h2obabyts/isoWATFLOOD).

3.8.References

Arnold, J. G., Youssef, M. A., Yen, H., White, M. J., Sheshukov, A. Y., Sadeghi, A. M., ... & Haney, E. B. (2015). Hydrological processes and model representation: Impact of soft data on calibration. *Transactions of the ASABE*, *58*(6), 1637-1660.

Arsenault, R., Poulin, A., Côté, P., & Brissette, F. (2014). Comparison of stochastic optimization algorithms in hydrological model calibration. Journal of Hydrologic Engineering, 19(7), 1374-1384. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000938

Asadzadeh, M., & Tolson, B. (2013). Pareto archived dynamically dimensioned search with hypervolume-based selection for multi-objective optimization. Engineering Optimization, 45(12), 1489-1509. https://doi.org/10.1080/0305215X.2012.748046

Asadzadeh, M., Tolson, B. A., & Burn, D. H. (2014). A new selection metric for multiobjective hydrologic model calibration. *Water Resources Research*, *50*(9), 7082-7099.

Bennett, N., Croke, B., Guariso, G., Guillaume, J., Hamilton, S., Jakeman, A., et al. (2013). Characterising performance of environmental models. Environmental Modelling & Software, 40, 1-20. https://doi.org/10.1016/j.envsoft.2012.09.011

Bergstrom, S., Lindstrom, G., & Pettersson, A. (2002). Multi-variable parameter estimation to increase confidence in hydrological modelling. Hydrological Processes, 16(2), 413-421. https://doi.org/10.1002/hyp.332

Beven, K. J. (2006), A manifesto for the equifinality thesis, J. Hydrol., 320, 18–36.

Beven, K. (2011). Rainfall-Runoff Modelling: The Primer (2nd ed.). Chichester, West Sussex, UK: John Wiley & Sons.

Beven, K. (2012). I believe in climate change but how precautionary do we need to be in planning for the future?. *Hydrological Processes*, *25*(9), 1517-1520.

Birkel, C., Dunn, S. M., Tetzlaff, D., & Soulsby, C. (2010). Assessing the value of high-resolution isotope tracer data in the stepwise development of a lumped conceptual rainfall–runoff model. Hydrological Processes, 24(16), 2335-2348. https://doi.org/10.1002/hyp.7763

Birkel, Christian, & Soulsby, C. (2015). Advancing tracer-aided rainfall-runoff modelling: A review of progress, problems and unrealised potential. *Hydrological Processes*, *29*(25), 5227–5240. https://doi.org/10.1002/hyp.10594

Birks, S., & Gibson, J. (2009). Isotope hydrology research in Canada, 2003-2007. Canadian Water Resources Journal, 34(2), 163-176. https://doi.org/10.4296/cwrj3402163

Chen, J, Brissette, FP, Poulin, A, and Leconte, R. 2011. Overall uncertainty study of the hydrological impacts of climate change for a Canadian watershed. *Water Resources Research* 47(12): Article W12509. DOI: https://doi.org/10.1029/2011WR010602

Clark, M.P., R.L. Wilby, E.D., Gutmann, J.A. Vano, S. Gangopadhyay, A.W. Wood, H.J. Fowler, C. Prudhomme, J.R. Arnold, L.D. Brekke (2016). Characterizing uncertainty of the hydrologic impacts of climate change. *Curr. Clim. Change Rep*. 2:55-64.

de Grosbois, E., Hooper, R. P., & Christophersen, N. (1988). A multisignal automatic calibration methodology for hydrochemical models: a case study of the Birkenes model. Water Resources Research, 24(8), 1299-1307.

Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation, 6(2), 182-197. https://doi.org/10.1109/4235.996017

Dehaspe, J., Birkel, C., Tetzlaff, D., Sánchez-Murillo, R., Durán-Quesada, A., & Soulsby, C. (2018). Spatially distributed tracer-aided modelling to explore water and isotope transport, storage and mixing in a pristine, humid tropical catchment. Hydrological Processes, 32(21), 3206-3224. https://doi.org/10.1002/hyp.13258

Delavau, C., Chun, K., Stadnyk, T., Birks, S., & Welker, J. (2015). North American precipitation isotope (delta O-18) zones revealed in time series modeling across Canada and northern United States. Water Resources Research, 51(2), 1284-1299. https://doi.org/10.1002/2014WR015687

Delavau, C., Stadnyk, T., & Holmes, T. (2017). Examining the impacts of precipitation isotope input on distributed, tracer-aided hydrological modelling. Hydrology and Earth System Science, 21(5), 2595-2614. https://doi.org/10.5194/hess-21-2595-2017

Djamai, N., Magagi, R., Goïta, K., Hosseini, M., Cosh, M., Berg, A., & Toth, B. (2015). Evaluation of SMOS soil moisture products over the CanEx-SM10 area. Journal of Hydrology, 520, 254-267. https://doi.org/10.1016/j.jhydrol.2014.11.026

Dunn, S. M., & Bacon, J. R. (2008). Assessing the value of Cl− and δ 18O data in modelling the hydrological behaviour of a small upland catchment in northeast Scotland. Hydrology Research, 39(5-6), 337-358.

Fenicia, F., McDonnell, J. J., & Savenije, H. H. (2008). Learning from model improvement: On the contribution of complementary data to process understanding. Water Resources Research, 44(6).

Gupta, H., Kling, H., Yilmaz, K., & Martinez, G. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of Hydrology, 377(1-2), 80-91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Guzman, J. A., Shirmohammadi, A., Sadeghi, A. M., Wang, X., Chu, M. L., Jha, M. K., ... & Hernandez, J. E. (2015). Uncertainty considerations in calibration and validation of hydrologic and water quality models. *Transactions of the ASABE*, *58*(6), 1745-1762.

He, Z., Unger-Shayesteh, K., Vorogushyn, S., Weise, S., Kalashnikova, O., Gafurov, A., et al. (2019). Constraining hydrological model parameters using water isotopic compositions in a glacierized basin, Central Asia. Journal of Hydrology, 571, 332-348. https://doi.org/10.1016/j.jhydrol.2019.01.048

Holmes, T. (2016). Assessing the value of stable water isotopes in hydrologic modeling: a dual isotope approach. Retrieved from MSpace. (http://hdl.handle.net/1993/31724). Winnipeg, MB: University of Manitoba.

Iorgulescu, I., Beven, K. J., & Musy, A. (2007). Flow, mixing, and displacement in using a data-based hydrochemical model to predict conservative tracer data. Water resources research, 43(3).

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(1), W03S04. https://doi.org/10.1029/2005WR004362

Knightly, J., S. Kuppel, A. Smith, C. Soulsby, M. Sprenger, D. Tetzlaff (2020). Using isotopes to incorporate tree water storage and mixing dynamics into a distributed ecohydrologic modelling framework. Ecohydrol. Accepted 15 Feb 2020. doi: 10.1002/eco2201

Kouwen, N. (2018). *WATFLOOD/CHARM Canadian hydrological and routing model.* Waterloo, ON: University of Waterloo.

Lilhare, R., S. Pokorny, S.J. Dery, T.A. Stadnyk, K. Koenig (2020). Sensitivity analysis and uncertainty assessment in water budgets simulated by the variable infiltration capacity model. *Hydrol. Process.* https://doi.org/10.1002/hyp.13711

Lindström, G., & Rodhe, A. (1986). Modelling Water Exchange and Transit Times in Till Basins Using Oxygen-18. Hydrology Research, 17(4-5), 325-334

Mallipeddi, R., & Suganthan, P. (2010). Ensemble of constraint handling techniques. IEEE Transactions on Evolutionary Computation, 14(4), 561-579. https://doi.org/10.1109/TEVC.2009.2033582

Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A., Fernández-Prieto, D., et al. (2017). GLEAM v3: satellite-based land evaporation on root-zone soil moisture. Geoscientific Model Development, 10(5), 1903-1925. https://doi.org/10.5194/gmd-10-1903-2017

Matott, L. (2018). *OSTRICH - an optimization software toolkit for research involving camputational heuristics.* Buffalo, NY: State University of New York.

McDonnell, J. J., and K. Beven (2014), Debates—The future of hydrological sciences: A (common) path forward? A call to action aimed at understanding velocities, celerities, and residence time distributions of the headwater hydrograph, Water Resour. Res., 50, 5342–5350, doi:10.1002/2013WR015141.

Miralles, D. G., Holmes, T. R., de Jeu, R. A., Gash, J. H., Meesters, A., & Dolman, A. J. (2011). Global land-surface evaporation estimated from satellite-based observations. Hydrology and Earth System Sciences, 453-469. https://doi.org/10.5194/hess-15-453-2011

Moriasi, D. N., Gitau, M. W., & Daggupati, P. (2015). Hydrologic and water quantity models: performance measures and evaluation criteria. Transactions of the ASABE, 58(6), 1763-1785. https://doi.org/10.13031/trans.58.10715

Neill, A., Tetzlaff, D., Strachan, N., & Soulsby, C. (2019). To what extent does hydrological connectivity control dynamics of faecal indicator organisms in streams? Initial hypothesis testing using a tracer-aided model. Journal of Hydrology, 570, 423-435. https://doi.org/10.1016/j.jhydrol.2018.12.066

Nelson, A. M., Moriasi, D. N., Talebizadeh, M., Steiner, J. L., Gowda, P. H., Starks, P. J., & Tadesse, H. K. (2018). Use of soft data for multicriteria calibration and validation of Agricultural Policy Environmental eXtender: Impact on model simulations. Journal of Soil and Water Conservation, 73(6), 623-636. https://doi.org/10.2489/jswc.73.6.623

Nijzink, R. C., Almeida, S., Pechlivanidis, I. G., Capell, R., Gustafssons, D., Arheimer, B., et al. (2018). Constraining conceptual hydrological models with multiple information sources. *Water Resources Research*, *54*, 8332–8362. https://doi.org/10.1029/2017WR021895

Piovano, T., Tetzlaff, D., Carey, S., Shatilla, N., Smith, A., & Soulsby, C. (2019). Spatially distributed tracer-aided runoff modelling and dynamics of storage and water ages in a permafrost-influenced catchment. Hydrology and Earth System Sciences, 23(6), 2507-2523. https://doi.org/10.5194/hess-23-2507-2019

Seibert, J., & McDonnell, J. J. (2013). Gauging the Ungauged Basin: Relative Value of Soft and Hard Data. *Journal of Hydrologic Engineering*, *20*(1), A4014004. https://doi.org/10.1061/(asce)he.1943-5584.0000861

Shafii, M, and Tolson, B. (2015). Optimizing hydrological consistency by incorporating hydrological signatures into model calibration objectives. *Water Resources Research* 51(5): 3796-3814. DOI: https://doi.org/10.1002/2014WR016520

Smith, A., Delavau, C., & Stadnyk, T. (2015). Identification of geographical influences and flow regime characteristics using regional water isotope surveys in the lower Nelson River, Canada. *Canadian Water Resources Journal*, *40*(1). https://doi.org/10.1080/07011784.2014.985512

Smith, A., Welch, C., & Stadnyk, T. (2016). Assessment of a lumped coupled flow-isotope model in Boreal catchments. Hydrological Processes, 30(21), 3871-3884. https://doi.org/10.1002/hyp.10835

Smith, A., Welch, C., & Stadnyk, T. (2018). Assessing the seasonality and uncertainty in evapotranspiration partitioning using a tracer-aided model. *J. Hydrol*. 560: 595-613.

Son, K., & Sivapalan, M. (2006), Improving model structure and reducing parameter uncertainty in conceptual water balance models through the use of auxiliary data, Water Resour. Res., 43, W01415, https://doi.org/0.1029/2006WR005032.

Stadnyk, T. A., Delavau, C., Kouwen, N., & Edwards, T. W. (2013). Towards hydrological model calibration and validation: simulation of stable water isotopes using the isoWATFLOOD model. Hydrological Processes, 27(25), 3791-3810. https://doi.org/10.1002/hyp.9695

Stadnyk, T.A. and Holmes, T. (2020). On the value of isotope-enabled hydrologic model calibration. *Hydrol. Sciences J.*. https://doi.org/10.1080/02626667.2020.1751847

Tolson, B., & Shoemaker, C. (2007). Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. Water Resources Research, 43(1), W01413. https://doi.org/10.1029/2005WR004723

Tolson, B., & Shoemaker, C., (2008). Efficient prediction uncertainty approximation in the calibration of environmental simulation models. *Water Resources Research*, 44(4), 1–19. https://doi.org/10.1029/2007WR005869

Tunaley, C., Tetzlaff, D., Birkel, C., & Soulsby, C. (2017). Using high-resolution isotope data and alternative calibration strategies for a tracer-aided runoff model in a nested catchment. Hydrological Processes, 31(22), 3962-3978. https://doi.org/10.1002/hyp.11313

Vaché, K. B., McDonnell, J. J., & Bolte, J. (2004). On the use of multiple criteria for a posteriori model rejection: Soft data to characterize model performance. Geophysical research letters, 31(21).

Wei, Z., Yoshimura, K., Wang, L., Miralles, D. G., Jasecho, S., & Lee, X. (2017). Revisiting the contribution of transpiration to global terrestrial evapotranspiration. Geophys. Res. Lett., 44(6), 2792-2801. https://doi.org/10.1002/2016GL072235

Yang, D., Herath, S., & Musiake, K. (2000). Comparison of different distributed hydrological models for characterization of catchment spatial variability. *Hydrological Processes*, *14*(3), 403–416. https://doi.org/10.1002/(SICI)1099-1085(20000228)14:3<403::AID-HYP945>3.0.CO;2-3

Zhou, R, Li, Y, Lu, D, Liu, H, and Zhou, H. 2016. An optimization based sampling approach for multiple metrics uncertainty analysis using generalized likelihood uncertainty estimation. *Journal of Hydrology*, 540: 274-286. DOI: https://doi.org/10.1016/j.jhydrol.2016.06.030

3.9.Extension to future flows and floods

The various calibrated models presented here in Chapter 3 were run with an ensemble of climate scenarios, in order to assess the impact of including isotope tracer data in model calibration on the simulation of hydrologic process under possible future climates and over longer-term periods. Two separate periods were simulated:

- Historical simulations for 1981 to 2010
- Future simulations for 2051 to 2070

The historical period was run using observed weather data, from the same source as in Section 3.3.2, while the future period was run with 19 bias corrected CMIP5 climate simulations (14 GCMs and 2 RCPs: 2.5 and 8.5), spanning a wide range of possible future climatic conditions [1,2]. These future inputs cover 90% of the regional climate variability in all CMIP5 climate simulations across 1/3 of the continental interior of Canada [3]. Annual precipitation and temperature are summarized for both the historical and future periods in Figure 3-7.

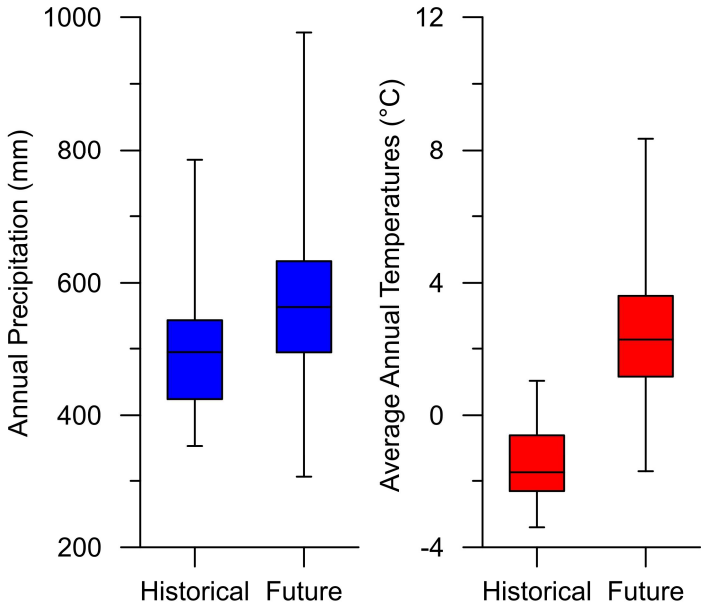

Figure 3-7: Annual precipitation totals and average annual temperatures inputs for the Odei River basin model in the historical (1981-2010) and future (2051-2070) simulation periods.

The average annual hydrographs for both periods and all model calibration types are shown in Figure 3-8, with the maximum and minimum simulation limits to represent the total range of variability.

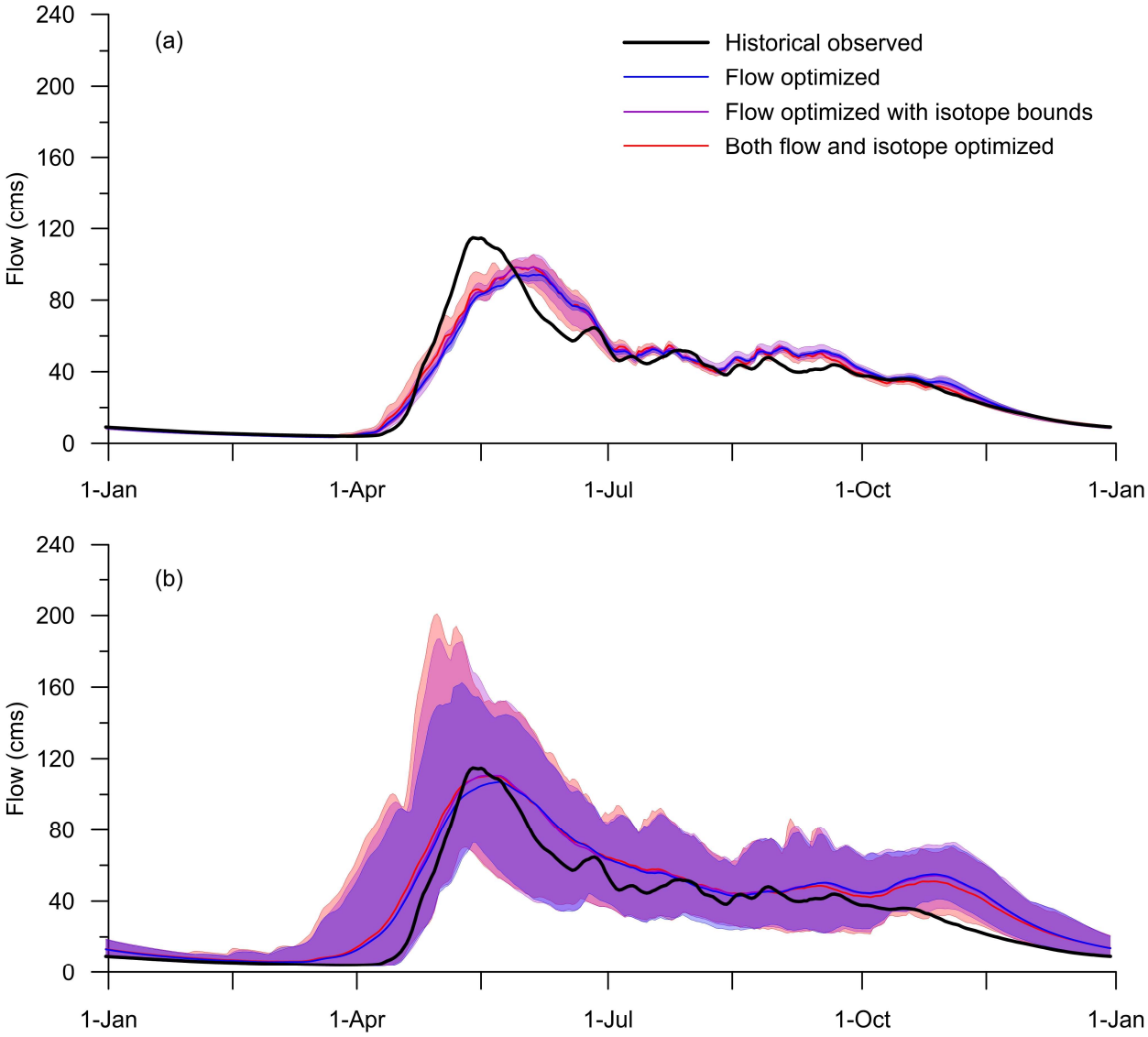

Figure 3-8: Average annual hydrographs for (a) the historical simulation period (1981-2010) and (b) the future simulation period (2051-2070), with the shaded area representing the maximum and minimum simulation limits.

All simulations were on average similar in the historical period, with a good fit to observations (Figure 3-8a). Future streamflow has a wider uncertainty envelope, which is expected given the uncertainty inherited from climate projections, although an earlier melt and later freeze are consistent all scenarios (Figure 3-8b). Parameter sets from isotope-enabled calibrations have higher freshet peak flows, and lower autumn recession flows. Differences in the calibration methodology produced larger changes in the simulated fluxes and storages than in overall

streamflow (Figure 3-9), pointing to some hydrologic processes being more sensitive to isotope tracers.

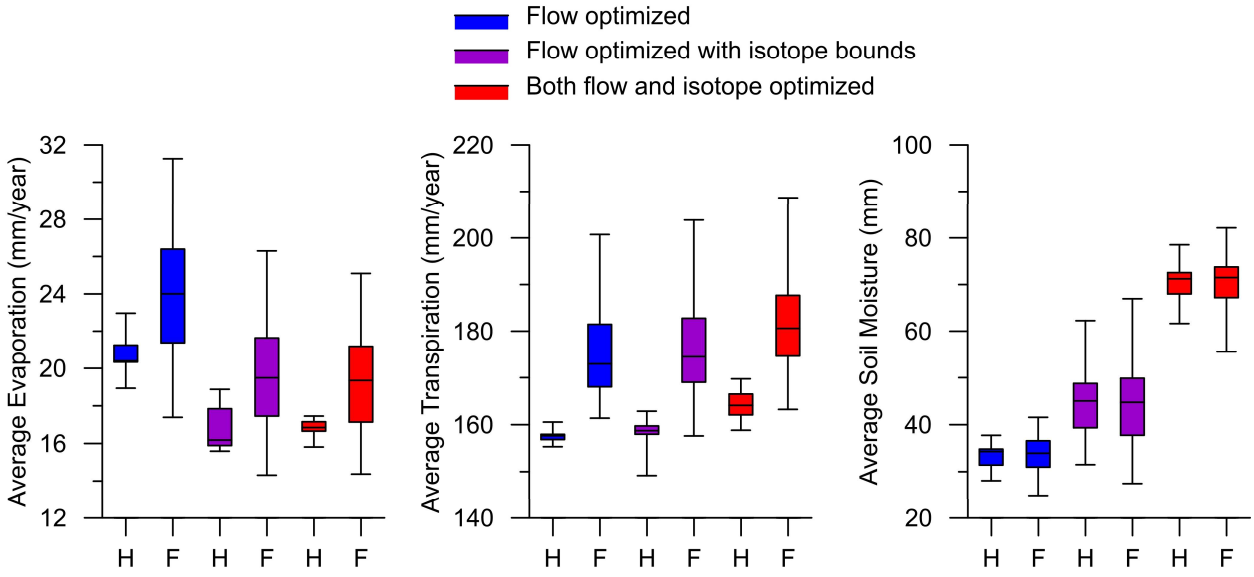

Figure 3-9: Simulated basin-wide fluxes for the historical (H) and future (F) periods.

Fluxes vary between calibration method in both the historical and future periods, with basin-wide evaporation decreasing and transpiration increasing as more weight is given to isotope simulation performance in the calibration. Future evaporation and transpiration depend more on the climate input than parameter values, while soil moisture is more dependent on model parameters.

Differences between the calibration methods are even more obvious for extreme events when streamflow is presented as an annual peak flow timeseries in a flood frequency diagram (Figure 3-10).

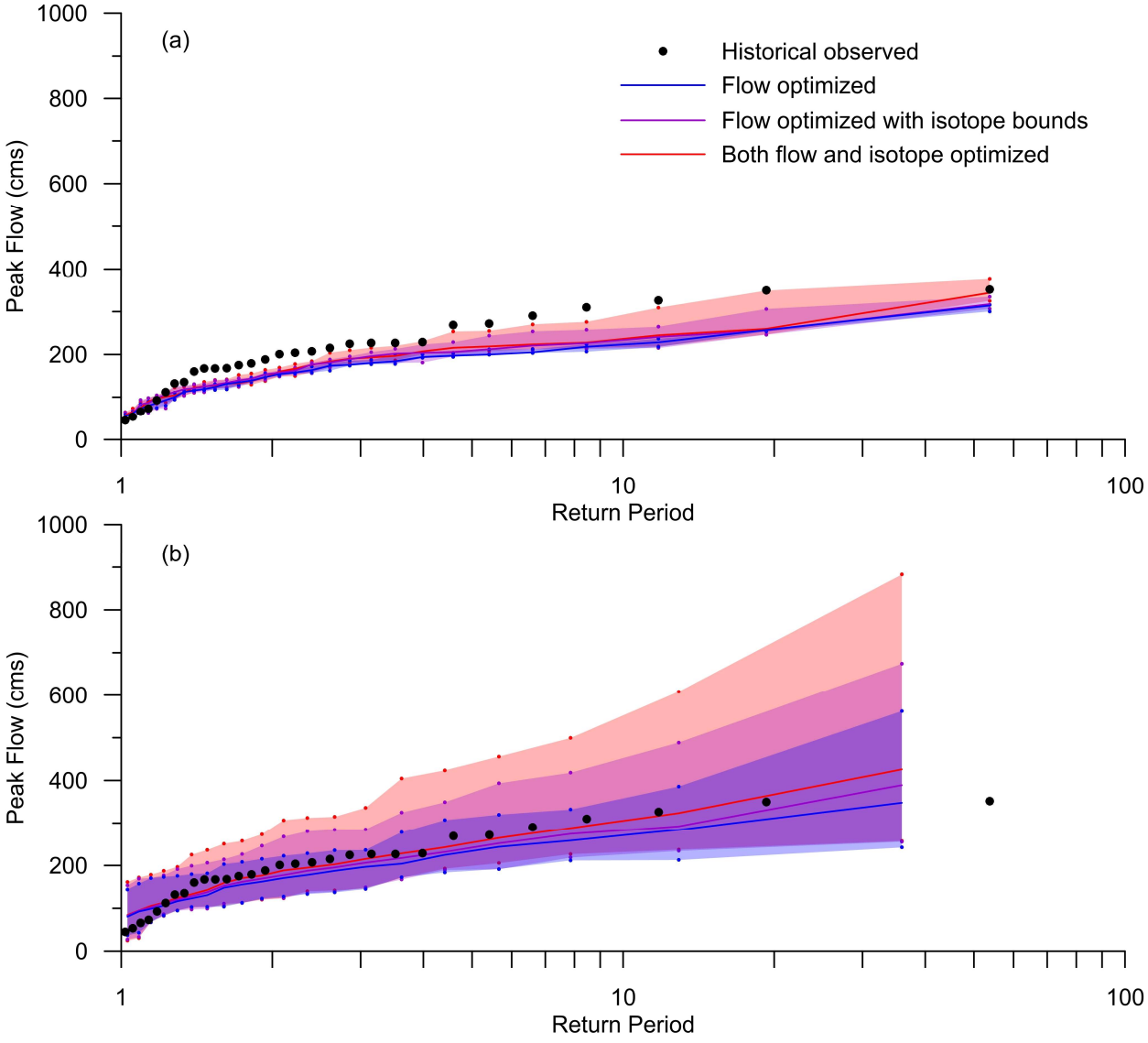

Figure 3-10: Flood frequency for (a) the historical simulation period (1981-2010) and (b) the future simulation period (2051-2070), with the shaded area representing the maximum and minimum simulation limits.

In the 1981-2010 period all simulations have fairly similar estimated flood magnitudes, considering the range of model uncertainty. The flow and isotope optimized model, however, does a better job at capturing the highest flow quantiles than the other calibrations. When run with climate change scenario forcings, the behavior of the different calibrated parameter sets diverges: the median 20-year flood estimated using an isotope-enabled calibration is 60 cms higher, and the maximum 275 cms higher, than from the flow calibrations (20 to 60% higher).

The accuracy of these future flood estimates cannot be ascertained, but the current maximum recorded flow for the Odei River, in 2017 and therefore not included in the historical simulations above, can be used as a test case for extreme conditions (Figure 3-11).

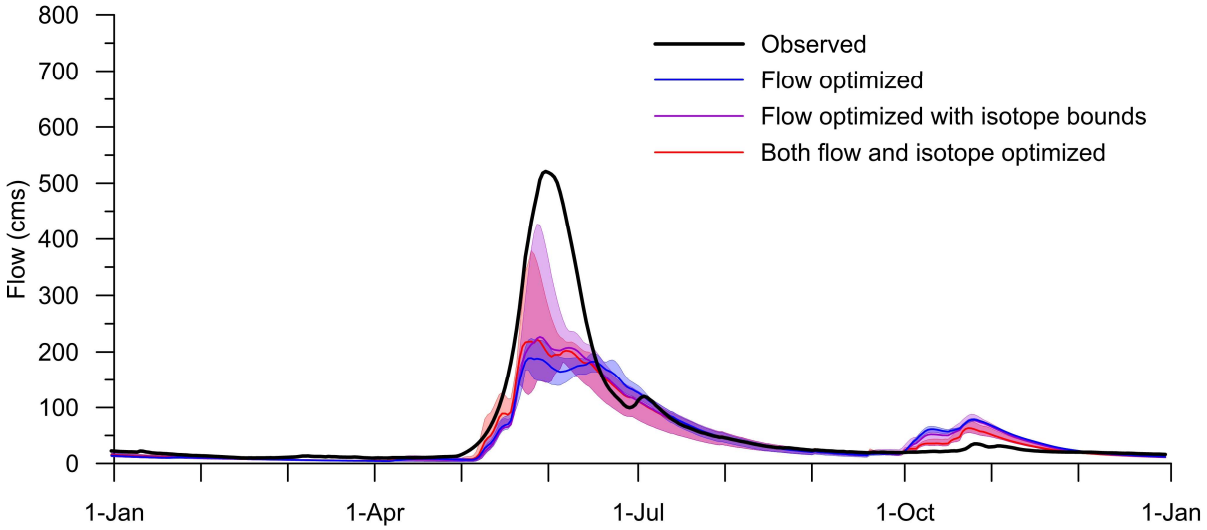

Figure 3-11: Simulated and observed hydrographs for 2017, the year with highest recorded peak flows to date, with the shaded area representing the maximum and minimum simulation limits.

The recorded flows for the flood peak are uncertain, due to the lack of reliability of the rating curve outside normal conditions, but the flow-only optimized models peak at less than half the recorded peak. The mean values for models calibrated using isotope data are also substantially lower than the observations, but the upper limits from isotope calibrated simulations are with the plausible uncertainty range of observed flows (±20%). The flood peaks estimated by isotope-enabled calibrations are decidedly more plausible than those from calibrations using only flow data. This is particularly important when establishing models for flood or peak flow scenario prediction under future climates.

[1] Berg, P., Donnelly, C., and Gustafsson, D. (2018). *Near-real-time adjusted reanalysis forcing data for hydrology*. Hydrol. Earth Syst. Sci., 22, 989–1000, doi:10.5194/hess-22-989-2018

[2] Braun, M., Thiombiano, A., Vieira, M., Stadnyk, T.A. 2021. Representing climate evolution in ensembles of GCM simulations for the Hudson Bay System. *Elementa: Sci. of the Anthro.*, 9 (1), 00011, doi:10.1525/elementa.2021.00011

[3] MacDonald, M. et al. (2018) *Impacts of 1.5 and 2.0 °C Warming on Pan-Arctic River Discharge Into the Hudson Bay Complex Through 2070.* Geophys. Res., 45(15), 7561–7570, doi:10.1029/2018GL079147

4. Variability in flow and tracer-based performance metric sensitivities reveal regional differences in dominant hydrological processes across the Athabasca River basin

Tegan L. Holmes[1, 2], Tricia A. Stadnyk[2, 1], Masoud Asadzadeh[1], John J. Gibson[3, 4]
[1] University of Manitoba, Civil Engineering, Winnipeg MB R3T 5V6
[2] University of Calgary, Geography, Calgary AB T2N 1N4 Canada
[3] InnoTech Alberta, 3-4476 Markham Street, Victoria BC V8Z 7X8 Canada
[4] University of Victoria, Geography, Victoria BC V8W 3R4 Canada

Published in 2022 in Journal of Hydrology: Regional Studies (doi: 10.1016/j.ejrh.2022.101088).
Submitted May 2021; accepted April 2022.

This chapter investigates parameter and process sensitivities for a variety of isotope tracer simulation performance metrics and compares them to those of flow simulation performance metrics. Spatial and temporal variations in sensitivity are also assessed. Isotope tracers are found to be more sensitive to soil processes which control water age and mixing volume.

4.1. Abstract

Study region: Athabasca River basin, Alberta, Canada (156,000 km2).

Study focus: Hydrology often relies upon hydrologic models in data-sparse regions; however, it is unclear if such models are reliably accurate, or if internal process simulations are reasonable representations of watershed function. Standard model evaluation and calibration approaches often prioritize accurate reproduction of recorded streamflow, ignoring process simulation fidelity, regardless of the intended model application. This study evaluates whether combined use of streamflow and isotope tracer performance metrics can improve representation of simulated streamflow-generating processes within a large river basin, the Athabasca watershed, to inform calibration of a process-based, distributed hydrologic model.

New hydrological insights for the region: Flow-based performance metrics were found to be sensitive to processes influencing streamflow volume and timing, but insensitive to internal flow

paths and storage volumes. Although somewhat less reliable than flow metrics, isotope tracer performance metrics are found to be most sensitive to processes influencing mixing and water age, and appreciably responsive to many other processes. We demonstrate that process-based hydrologic models for rivers such as the Athabasca River cannot be optimally calibrated using streamflow metrics alone, as such optimizations cannot tune parameters or process representations to which the objective function is insensitive. Importantly, isotope tracers have demonstrable value for informing process-based hydrologic model optimization by providing a window into the sub-surface black box within complex regional-scale simulations.

4.2. Introduction

Hydrologic models are broadly used to simulate flow generating processes in watersheds, typically with the goal of producing runoff and streamflow assessments. The flow timing and water volumes from these assessments affect predictions of ecosystem function and resilience, water supply and hydroelectric generation, and the extent of damage resulting from flooding and drought (Carlisle et al., 2011; Wan et al., 2021; Buttle et al., 2016). An accurate prediction of streamflow in both the short-term and long-term (for climate change projections) is therefore of key importance for water resources operations and planning. Standard methods of evaluating models tend to judge simulation quality based on the accurate reproduction of historical streamflow, while the fidelity of process simulation is often ignored, regardless of the intended use of the simulation results (Clark et al., 2011).

Increasing the physical basis of a model to improve process simulation and decrease model dependence on historical data is intuitively attractive, but also necessitates higher resolution input data and increases the computational demand for running the model (Clark et al., 2017; Peters-Lidard et al., 2017). Accurately modeling streamflow is particularly challenging in remote regions as data availability is generally low: streamflow and weather gauges are rare and have limited record lengths (Coulibaly et al., 2013). Data limitations increase the uncertainty, but also the need, for hydrologic modeling in mid- to high latitude regions with limited accessibility. To reliably simulate flows in ungauged basins, or to predict flows under non-stationary climatic conditions, it is critical for hydrologic models to accurately represent the physical processes generating streamflow (Duethmann et al., 2020). Information on individual hydrologic processes

is even more rare than weather or hydrometric data, and limited accessibility in remote regions stalls the expansion of data networks.

Additional observations, such as stable isotope tracer data (i.e. ratios of water molecules containing ^{18}O or ^{2}H to standard water), have been used to add information on hydrologic processes. Stable isotope tracers are naturally occurring, non-reactive tracers of water source and processes resulting from their variable occurrence in precipitation and evaporating water bodies (Birkel and Soulsby, 2015; Bowen et al., 2019). A hydrologic model capable of simulating both flow and isotope tracer composition can therefore be more robustly evaluated against additional observed data. A few models have already combined isotope and flow simulations, such as the isoWATFLOOD model (Stadnyk et al., 2013), the IWBMIso model (Belachew et al., 2016), or the STARR model (van Huijgevoort et al., 2016). Linked hydrologic-tracer simulations can be compared to both flow and tracer observations, further expanding the options for model evaluation, simulation performance metric choice, and the identification of more physically meaningful model parameter values during model calibration (Holmes et al., 2020; Stadnyk and Holmes, 2020; Tunaley et al., 2017; Yamanaka and Ma, 2017).

Model performance metrics are used to evaluate model accuracy, or as objective functions in automated model calibration algorithms. They are broadly categorized as either residual error metrics, based on aggregation of differences between simulated and observed data point pairs; or data set comparison metrics, based on differences between a population property of the observed and the simulated data sets (Bennett et al., 2013). Despite the wealth of literature on performance metrics, there remains no general consensus on the best performance metric(s) to evaluate either flow or isotope tracer simulations, and moreover, these two types of data typically differ in temporal resolution and consistency of sampling, which can significantly impact metric accuracy and therefore selection (Bennett et al., 2013; Mizukami et al., 2019). The Kling-Gupta (KGE) and Nash-Sutcliffe efficiency (NSE) metrics are frequently used in the literature, but flow signature metrics are also shown to improve the evaluation of specific flow simulation characteristics and identify hydrologically consistent parameter sets (Shafii and Tolson, 2015; Knoben et al., 2019; Sahraei et al., 2020). Isotope simulations are most frequently evaluated using some variant of a residual error metric, but other metrics, such as the Kling-Gupta efficiency have also been applied (He et al., 2019; Tunaley et al., 2017). These metrics all

evaluate the model performance at local points (either individually or as an averaged performance) using integrated (or cumulative) data, as both flow and flow tracers are the final summation of a multitude of hydrologic processes across a watershed.

A process-based model can potentially be verified along individual flow paths, as such processes are both simulated and intended to match real-world fluxes. It cannot be assumed that metrics designed to evaluate cumulative model performance will also be capable of evaluating or identifying the individual processes contributing to flow (Oreskes et al., 1994). Previous research on process-based model calibration identifies that streamflow performance only informs a sub-set of simulated processes (Acero Triana et al., 2019; Newman et al., 2017), and that process sensitivity to model performance varies seasonally (Bajracharya et al., 2020; Pfannerstill et al., 2015; Wagener et al., 2003). The literature is, however, short in analyzing the processes tracer performance metric are sensitive to. The actual capacities of metrics (for both streamflow and tracer simulations) to react to changes in the internal simulation of critical hydrologic processes (on both inter- and intra-annual time scales) would be of considerable utility in designing calibration strategies for process-based models (Mizukami et al., 2019). Sensitivity analyses are well-adapted to address this point, and relative sensitivities of model parameters have previously been used to inform model calibration (Razavi and Gupta, 2015; Song et al., 2015; Haghnegahdar et al., 2017). Sensitivity analyses are distinct from model calibration, as they do not identify optimal or even necessarily good parameter values, but rather they identify linkages between parameters and metrics.

This study will evaluate whether performance metrics respond to changes in simulated streamflow-generating processes for the purposes of guiding hydrologic modeling choices. To this end, global sensitivity analyses are utilized to answer the following questions:

1. Which processes are flow simulation performance metrics sensitive to, and is there any temporal or spatial variability in this sensitivity; and
2. Are there processes which isotope tracer metrics are sensitive to that streamflow metrics are insensitive to, and vice-versa.

These results will be used to assess the value of various metrics or datasets in adequately informing the calibration of a process-based hydrologic model, rather than comparing calibration outputs as has been done previously. Our aim is to provide guidance for the selection of

performance metrics in tracer-aided calibration, and an awareness of inherent trade-off between traditional flow-based calibration and tracer-aided calibration. Our study focuses on the Canadian Oil Sands region in Alberta, Canada, where it is critical to assess the reliability of water supply forecasts given this region is undergoing significant future change resulting from anthropogenic development and climate change, including glacial retreat, permafrost thaw and increased forest fires (J. J. Gibson et al., 2019; Nenzén et al., 2020; Stahl et al., 2008). Understanding the hydrology and water supply of this region is a key goal of the Alberta Oil Sands Monitoring strategy (Government of Canada, 2021).

4.3.Methods

4.3.1. Athabasca River Basin

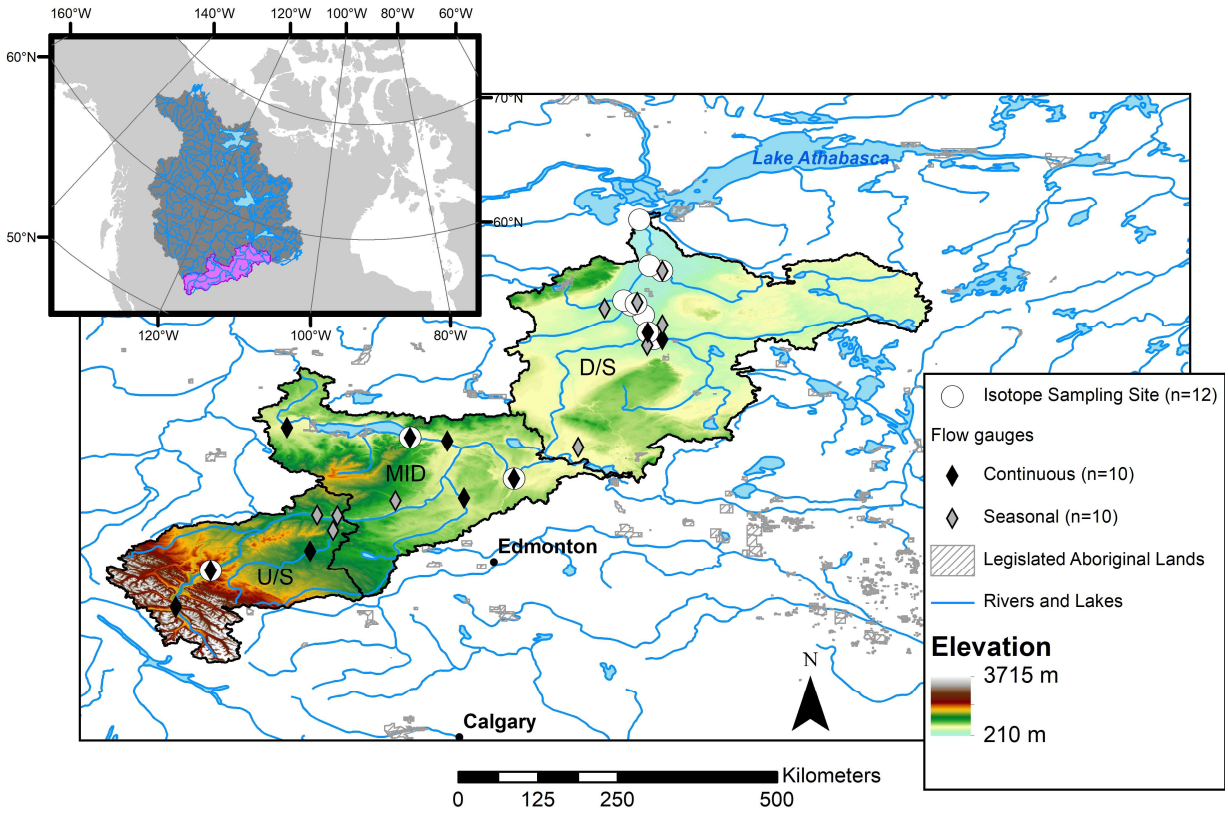

Figure 4-1: The Athabasca watershed with Water Survey of Canada flow gauges and sampling sites for isotope compositions of streamflow. The Mackenzie River basin with the Athabasca River watershed highlighted is shown in the inset.

The Athabasca River runs north-east from the Rocky Mountains to Lake Athabasca and the Peace-Athabasca Delta. It is the most southerly part of the Mackenzie River basin (Figure 4-1). The Athabasca River watershed is located in the north of the Canadian provinces of Alberta and Saskatchewan, on Treaty 6 and 8 territory. The total watershed area is 156,000 km^2; elevations and land use vary widely from upstream to down. The upper reaches of the Athabasca are alpine or foothills regions, with steep slopes and some glaciers, most notably the Athabasca Glacier in the Columbia Icefield (Intsiful and Ambinakudige, 2021). The lower reaches, which coincide with the Athabasca Oil Sands region, have subdued relief and abundant wetlands. The soils in the basin are primarily loam, with higher clay prevalence in the mid-reach, some sandy or coarse soil in the downstream region and some exposed rock or shallow soil in the upstream areas (Shangguan et al., 2014). The rock underlying the Athabasca basin is predominately sedimentary stone from the Cretaceous and Paleogene, with some older rocks exposed in the Rocky Mountains and a small area with Precambrian granite in the north (Alberta Geological Survey, 2013). Substantial agricultural activity occurs between the upper and lower reaches; boreal coniferous forests are prevalent throughout the basin. Minimal quantities of water are diverted for agriculture, while approximately 1% of annual flow is used for activities in the oil sands (Rosa et al., 2017). There is some sporadic permafrost in the region which is actively degrading; deeper bedrock formations contribute small amounts of flow to the Athabasca River and its tributaries (3-5% of annual flow) (Gibson et al., 2016; Vitt et al., 2000). The landscape features of the Athabasca River basin, and its upstream, middle and downstream regions are summarized in Table 4-1.

Table 4-1: Topographic, soil and land cover data summary for the Athabasca River basin and its upstream (U/S), mid-reach (MID), and downstream (D/S) regions as in this study (see Figure 4-1 for region boundaries).

		All	U/S	MID	D/S
Area (km^2)		156,000	32,200	46,000	77,700
Slope (%)	Average	0.35	1.21	0.28	0.15
Elevation (m)	Maximum	3715	3715	1379	866
	Minimum	211	689	494	211
	Mean	659	1375	725	522
Soil (%)	Sand/coarse	9.4	0.0	0.0	19.0
	Loam (low clay)	62.0	85.0	44.9	62.5
	Loam (with clay)	25.1	14.7	49.0	15.2
	Clay/clay mix	3.5	0.2	6.1	3.4
Land cover (%)	Grass	8.1	5.3	23.0	0.4
	Wetland	11.0	4.0	12.3	13.2
	Mixed Wood	15.1	10.8	30.7	7.5
	Coniferous	54.1	65.2	25.0	66.8
	Shrub	6.2	7.1	4.1	7.1
	Impervious	0.0	0.1	0.0	0.0
	Barren	1.4	5.9	0.0	0.4
	Water	3.8	0.8	4.7	4.6
	Glacier	0.2	0.9	0.0	0.0

The climate of the Athabasca watershed is highly seasonal; mean monthly temperatures (averaged across the entire basin) range from -19 °C to +17 °C, with a mean annual temperature of 0°C over the study period (2002-2015). The long-term average annual precipitation is 450 mm; in the downstream reaches, approximately 60% of precipitation falls as rain, but the upstream reaches are colder than the basin average and a larger fraction falls as snow (Environment and Climate Change Canada, 2020).

The Athabasca River basin is, in many respects, an ideal watershed case study for the utility of isotope tracers in large-scale hydrologic modeling. The basin contains a wide range of elevations and land cover, from the glacial headwaters, through mixed-use grasslands and ending in wetland-dominated boreal forest, within a moderately-sized basin. The rivers in the Athabasca

watershed are not regulated by any major reservoirs or hydro-electric developments. The Athabasca River basin is also relatively accessible, compared to many mid- to high-latitude watersheds, and oil sands developments have led to expanded research and longer-term monitoring in the region.

4.3.2. Hydrologic model setup & parameterization

Model

The Athabasca River watershed was modeled using CHARM/WATFLOOD and its associated dual-isotope simulation model, isoWATFLOOD. CHARM is an open source, distributed hydrologic model with a mixture of physically based and conceptual process representation (Kouwen, 2018). The isotope tracer models for CHARM simulate the isotopic concentrations of oxygen-18 and deuterium in all of the storages and fluxes used in the original hydrologic model; individual hydrologic storages are assumed to be completely mixed through depth, and fluxes generally have the same concentration as the source storage, except evaporative fluxes, which are subject to isotopic fractionation (Stadnyk and Holmes, 2020). Both the hydrologic and tracer simulations run on an hourly time-step, with daily simulated model output.

The Athabasca River basin model divides the watershed area into 320 grid cells, with a nominal cell size of 0.4° longitude by 0.2° latitude (actual cell sizes are adjusted based on drainage area); each cell is subdivided into 10 grouped response unit (GRU) types, based on land cover data from the ESA (European Space Agency, 2017). The majority of these GRU types are modeled with soil layers, namely the grass (8.1%), coniferous (54.1%) and mixed forest (15.1%), shrub (6.2%), disconnected wetland (8.8%), and barren (1.4%) classes. The glacier (0.2%) and impervious (0.03%) classes have no modeled soil storages (all rain and snowmelt becomes direct runoff), and glacier GRU also generate glacier melt flows. Open water (3.8%) and wetlands connected to the stream network (2.2%) also have no soil storages, but rain or snowmelt is added directly to the wetland, channel or lake rather than running off.

Process representation

The CHARM/WATFLOOD model has two soil layers, both of which can generate sub-surface flows to the channel network. Rain or snowmelt can either infiltrate to the upper soil zone or runoff. Water in the upper soil zone may recharge the lower soil zone, flow out to the channel

network or connected wetlands, or evapotranspire. Water in the lower soil zone may only flow out to the channel network or connected wetlands. All types of GRU have potential snowpack storages. Snow and glacier melt rates are calculated as a function of air temperature and melting snowpacks cover fractional areas of each GRU. The upper soil zone under a snowpack is considered frozen, and all soil fluxes (infiltration, recharge and interflow) have substantially reduced rates in frozen soils, but permafrost is not included in the model. Connected wetlands, where present, collect outflows from all GRU with soil layers, and have bi-directional flow with the channel network, with direction determined by the relative water levels. More detailed descriptions and full equations can be found in Holmes (2016).

All processes listed above have parameters controlling the simulated flux; parameter values can be consistent across GRU, or separate GRU can have different parameter values specific to that class. As a mixed physically based and conceptual model, there are a very large number of parameters which can be altered in setting up and calibrating a watershed model (over 250 for the Athabasca model). However, the vast majority have minimal impact on the simulation (e.g. overland flow roughness factors), or can be estimated from the literature (surface depression storage caps). This study will focus on potentially significant parameters identified by previous studies, including Holmes et al. (2020), and model developer recommendations, with a minimum of one parameter per simulated process, as listed in Table 4-2.

Table 4-2: List of potential significant parameters included in the sensitivity analysis, including the parameter names, the process affected by the parameter and the affected GRU with a list of the decoupled land classes to which coupled parameters are applied

Parameter description	Parameter name	Internal name	Hydrologic Process	Applicable GRU	Decoupled classes
Surface soil conductivity	k F (surf)	ak	Infiltration	All soil-based	-
Horizontal upper soil zone conductivity	k F (horz)	rec	Interflow	All soil-based	-
PET to AET factor	PET F	fpet	Evaporation	Water, connected wetland	-
Snowmelt rate factor	melt rate	fm	Snowmelt	All	low vegetation (grass+shrub), coniferous, mixed, bare (barren+impervious), disconnected wetland, connected wetland, water
Upper soil zone soil water retention cap	soil ret	retn	Soil storage and ET	All soil-based	grass, coniferous, mixed, barren, shrub, wetland
Vertical upper soil zone conductivity	k F (vert)	ak2	Recharge	All soil-based	-
Baseflow equation constant	C	flz	Baseflow	All soil-based	Upstream, mid-basin, downstream
Baseflow equation power	pwr	pwr	Baseflow	All soil-based	-
Channel roughness factor	n	r2n	Channel velocity	Water	Upstream, mid-basin, downstream
Wetland porosity	θ (wet)	theta	Wetland storage	Connected wetland	-
Wetland conductivity	k (wet)	kcond	Wetland velocity	Connected wetland	-
Glacier melt factor	glac F	gladjust	Glacier melt	Glacier	-

Meteorological Data

The hydrologic and isotope tracer models were run using four meteorological forcings: hourly air temperature and humidity, daily total precipitation and monthly average isotopic compositions of precipitaiton. The precipitation, temperature and humidity forcings were based on observations at Environment and Climate Change Canada (ECCC) weather stations (Environment and Climate Change Canada, 2020). Forcing data for each grid cell at each time step were estimated using inverse distance squared weighting, with a temperature lapse rate of -5 °C/km and a precipitation lapse rate of 0.2 mm/km; 56 weather stations were included in the calculation, provided there was observation data for that time interval (Minder et al., 2010; Kouwen, 2018). The isotopic compositions of precipitation were estimated from the empirical model developed by Delavau et al. (2015), which uses a geospatial interpolation and a multiple linear regression of geographic and climatic indicators. No field measurements of isotopes in precipitation within the Athabasca basin boundaries were used in the development or validation of the geospatial isotope model, but meteoric water samples from both immediately south and north of the watershed were included (Delavau et al., 2011). The climate zone models covering the Athabasca basin had modeled precipitation residual IQR of 3.6 and 4.7‰ (for $\delta^{18}O$) in validation and the model adequately captured the seasonality of isotopes in precipitation (i.e. highly depleted precipitation in winter and annual variation of 15‰ for $\delta^{18}O$) (Delavau et al., 2015).

Flow and Isotope Data

Historical hydrometric data from the Water Survey of Canada were used to calculate model performance metrics (Environment and Climate Change Canada, 2018). A total of 20 continuous or seasonal (i.e. continuous only during the open water season) hydrometric stations with daily data (m^3s^{-1}) between 2002 and 2015 were used in the analysis, listed in Table 4-3 (see Figure 4-1 for spatial distribution, and Appendix C for average annual hydrographs and isotope sample data). Gauged areas ranged between 960 and 133,000 km^2. The uncertainty in the streamflow data are approximately ±10% on average, with higher uncertainty during peak flow and ice-on periods (Kiang et al., 2018; Westerberg et al., 2020).

Table 4-3: Hydrometric gauges and isotope sampling sites in the Athabasca River basin

		Latitude (°)	Longitude (°)	Isotope Samples	Drainage Area (km^2)	Operation Schedule
07AA002	ATHABASCA RIVER NEAR JASPER	52.91	-118.06		3870	Continuous
07AD002	ATHABASCA RIVER AT HINTON	53.42	-117.57	159	9760	Continuous
07AE001	ATHABASCA RIVER NEAR WINDFALL	54.21	-116.06		19600	Seasonal
07AG004	MCLEOD RIVER NEAR WHITECOURT	53.99	-115.84		9110	Seasonal
07AG007	MCLEOD RIVER NEAR ROSEVEAR	53.70	-116.16		7140	Continuous
07AH001	FREEMAN RIVER NEAR FORT ASSINIBOINE	54.41	-114.96		1660	Seasonal
07AH003	SAKWATAMAU RIVER NEAR WHITECOURT	54.20	-115.78		1150	Seasonal
07BC002	PEMBINA RIVER AT JARVIE	54.45	-113.99		13100	Continuous
07BE001	ATHABASCA RIVER AT ATHABASCA	54.72	-113.29	146	74600	Continuous
07BF002	WEST PRAIRIE RIVER NEAR HIGH PRAIRIE	55.45	-116.49		1150	Continuous
07BK001	LESSER SLAVE RIVER AT SLAVE LAKE	55.31	-114.76	17	13600	Continuous
07BK007	DRIFTWOOD RIVER NEAR THE MOUTH	55.26	-114.23		2100	Continuous
07CA006	WANDERING RIVER NEAR WANDERING RIVER	55.17	-112.39		1120	Seasonal
07CD001	CLEARWATER RIVER AT DRAPER	56.68	-111.20	44	30800	Continuous
07CD004	HANGINGSTONE RIVER AT FORT MCMURRAY	56.60	-111.41		960	Seasonal
07DA001	ATHABASCA RIVER BELOW MCMURRAY	56.78	-111.40	126	133000	Continuous
07DA006	STEEPBANK RIVER NEAR FORT MCMURRAY	56.89	-111.20	37	1320	Seasonal
07DA008	MUSKEG RIVER NEAR FORT MACKAY	57.21	-111.55	70	1460	Seasonal
07DB001	MACKAY RIVER NEAR FORT MACKAY	57.12	-112.01	26	5570	Seasonal
07DC001	FIREBAG RIVER NEAR THE MOUTH	57.65	-111.20	44	6500	Seasonal
07DD011	Athabasca R. at Old Fort	58.37	-111.52	120	160000	-
AB07DA0750	Ells River	57.30	-111.68	36	2500	-
AB07DA0980	Athabasca River u/s Firebag	57.72	-111.38	68	154400	-

A monthly water isotope sampling campaign on the Athabasca River and several tributaries was conducted for the Alberta Environmental Monitoring, Evaluation and Reporting Agency's Long-Term River Network monitoring program (Gibson et al., 2016). Sampling at hydrometric gauges in the Athabasca basin began in 2002, and continued through 2014, with variable sampling frequency; some years, sampling occurred approximately monthly, while some gauges have data gaps longer than one year. All water samples were sealed in 30 mL high-density polyethylene bottles and analyzed at either the University of Waterloo Environmental Isotope Laboratory or at Alberta Innovates Technology Futures, Victoria (Gibson et al., 2016). High-density polyethylene bottles have been shown to be effective at preventing isotopic fractionation, and all samples were sealed and analysed within 1 year of sample collection (J. J. Gibson et al., 2019; Spangenberg, 2012). Water samples were analyzed using a Micromass IsoPrime Dual Inlet/Gas Chromatograph pre-2009, and from 2009 on, using a Thermo Scientific Delta V Advangage Dual Inlet/HDevice system, with an estimated analytical uncertainty of ±0.1‰ for oxygen-18 and ±1‰ for deuterium for both periods (Gibson et al., 2016). Isotope results are reported in δ notation in permil (‰), relative to V-SMOW.

4.3.3. Performance metrics

A variety of metrics were selected to quantify simulation performance, based on the most commonly applied metrics from the literature. As noted throughout this section, metrics are sensitive to various characteristics of the distribution of the error residuals, and therefore including multiple metrics in model evaluations is expected to expand the number of sensitive parameters. Only simulated data on days that have flow or isotope observations are considered for calculating these performance metrics. Firstly, the normalized root mean square error was used for both the flow and isotope simulations, calculated as:

$$NRMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_{s,i} - x_{o,i}\right)^2}/\overline{x}_o \qquad (1)$$

Where n is number of observations, $x_{o,i}$ is observation i, $x_{s,i}$ is the corresponding simulated value and $\overline{x}_o$ is the observation mean. The traditional Nash-Sutcliffe efficiency (NSE), another residual error metric, and the log transform of NSE were calculated exclusively for the flow simulation (Nash and Sutcliffe, 1970). While $NRMSE$ and NSE are highly sensitive to large

residuals that often happen due to mis-timed simulations in high-flow periods, $logNSE$ is more sensitive to small residuals that happen during low-flow periods.

$$NSE = 1 - \frac{\sum_{i=1}^{n}(x_{s,i}-x_{o,i})^2}{\sum_{i=1}^{n}(x_{o,i}-\overline{x}_o)^2} \quad (2)$$

$$logNSE = 1 - \frac{\sum_{i=1}^{n}\left(log(x_{s,i})-lo\ \ (x_{o,i})\right)^2}{\sum_{i=1}^{n}(log(x_{o,i})-\overline{lo\ \ (x_o)})^2} \quad (3)$$

The Kling-Gupta efficiency (KGE) metric, and all three of its constituent components (i.e. the correlation r, the relative variability α and the bias β) were used for both the isotope and flow simulations. KGE and NSE share the same components r, α, and β, but KGE gives them the same weight as opposed to NSE that relatively undermines variability (Gupta et al., 2009).

$$r = \frac{\sum_{i=1}^{n}(x_{o,i}-\overline{x}_o)(x_{s,i}-\overline{x}_s)}{\sqrt{\sum_{i=1}^{n}(x_{o,i}-\overline{x}_o)^2}\sqrt{\sum_{i=1}^{n}(x_{s,i}-\overline{x}_s)^2}} \quad (4)$$

$$\alpha = \frac{\sigma_s}{\sigma_o} \quad (5)$$

$$\beta = \frac{\overline{x}_s}{\overline{x}_o} \quad (6)$$

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \quad (7)$$

Where σ_s and σ_o are the standard deviations of the simulated and observed data.
Three popular flow signature metrics were also applied to both the isotope and flow simulations. Equation 8 was used to calculate the slope of the flow duration curve ($SFDC$) (for flows) and the slope of the duration curve (SDC) (for isotopes) (Viglione et al., 2013):

$$SDC = 100\left(\frac{x_{s,30}-x_{s,70}}{40\overline{x}_s} - \frac{x_{o,30}-x_{o,70}}{40\overline{x}_o}\right) \quad (8)$$

Where x_{30} and x_{70} are the data with exceedance probabilities of 30% and 70%. The high- and low-flow signatures, at the 5% and 95% exceedance probabilities were calculated using equations 9 and 10:

$$Q_5 = \frac{x_{o,5}-x_{s,5}}{x_{o,5}} \quad (9)$$

$$Q_{95} = \frac{x_{o,95}-x_{s,95}}{x_{o,95}} \quad (10)$$

Where x_5 and x_{95} are the data with exceedance probabilities of 5% and 95%.

There are multiple other possible permutations of the SDC, and high- and low-flow signatures, which include slopes in log scale, alternate exceedance probabilities, and mean high and low flow comparisons, but these are all fundamentally variants evaluating the same parts of the hydrograph (Bajracharya et al., 2020; Shafii and Tolson, 2015; Yilmaz et al., 2008). The relative variability and bias components of KGE can also be flow signatures based on discharge statistics (Shafii and Tolson, 2015).

Finally, three metrics quantifying the representativeness of the simulated slope of the isotope-derived local mixing line (LML), which uses both isotope simulations in combination, were included in the analysis: the LML slope error, the LML intercept error and the LML fit error (Stadnyk and Holmes, 2020). The LML slope error is the difference between the best fit slope for the simulated river isotope compositions and the best fit slope for the observed river isotope compositions:

$$LML\ mE = \frac{\sum_{i=1}^{n}(O_{s,i}-\overline{O}_s)(D_{s,i}-\overline{D}_s)}{\sum_{i=1}^{n}(O_{s,i}-\overline{O}_s)^2} - \frac{\sum_{i=1}^{n}(O_{o,i}-\overline{O}_o)(D_{o,i}-\overline{D}_o)}{\sum_{i=1}^{n}(O_{o,i}-\overline{O}_o)^2} \quad (11)$$

Where $O_{o,i}$ is oxygen-18 observation i, and $O_{s,i}$ is the simulated oxygen-18 value for the time observation i was taken, and $D_{o,i}$ and $D_{s,i}$ are likewise the observation and simulated value for the deuterium data. Similarly, the LML intercept error is the difference between the best fit line intercept for the simulated river isotope compositions and the best fit line intercept for the observed river isotope compositions:

$$LML\ bE = \overline{D}_s - \frac{\overline{O}_s \sum_{i=1}^{n}(O_{s,i}-\overline{O}_s)(D_{s,i}-\overline{D}_s)}{\sum_{i=1}^{n}(O_{s,i}-\overline{O}_s)^2} - \overline{D}_o + \frac{\overline{O}_o \sum_{i=1}^{n}(O_{o,i}-\overline{O}_o)(D_{o,i}-\overline{D}_o)}{\sum_{i=1}^{n}(O_{o,i}-\overline{O}_o)^2} \quad (12)$$

The LML fit error is simply the difference between the R^2 values for the best fit lines through the simulated and observed river isotope compositions:

$$LML\ RE = R_s^2 - R_o^2 \quad (13)$$

The performance metrics used in this study are summarize in Table 4-4, which includes a qualitative assessment of which error types each metric responds to (as, according to the literature previously referenced here, the sensitivity of metrics to various error types differ). Error types considered are simulation timing, simulation bias, the simulation variability, and

errors in high quantiles (e.g. peak flows, enriched isotope concentrations) and low quantiles (e.g. low flow periods, snowmelt freshet isotope signatures).

Table 4-4: Summary of the 29 performance metrics considered, listing which simulation types each metric was applied to, and a qualitative assessment of the simulation error types the metric responds to (filled circles indicate strong responses and empty circles indicate some response).

	Simulation			Error type				
		Isotope						
	Flow	^{18}O	^{2}H	Timing	Bias	Variability	Upper quantile	Lower quantile
NRMSE	X	X	X	●	●	○	○	○
NSE	X			●	●	○	○	○
logNSE	X			●	○			●
KGE	X	X	X	●	●	●	○	○
β (bias)	X	X	X		●			
α (var)	X	X	X			●	○	○
Correlation	X	X	X	●				
SDC	X	X	X		○	●		
Q5	X	X	X				●	
Q95	X	X	X					●
LML mE		X			○	●	○	○
LML bE		X			●	○	○	○
LML RE		X				●		

4.3.4. Parameter sensitivity and visualization

Parameter sensitivity analyses quantify the variation in a response variable - either a model performance metric or simply a model output variable - to changes in parameter values, either locally (around a particular parameter value) or globally (across a wide range of possible parameter values). A global sensitivity analysis (GSA) can illuminate the relative importance of different hydrologic processes within a particular watershed (Razavi and Gupta, 2015; Song et al., 2015), but the response is often aggregated across broad areas and longer time periods, such that results represent an "average" or aggregate level of sensitivity. This can be misleading for directing further research and in identifying the most significant unknowns (Bajracharya et al., 2020).

This study uses variogram-based GSA, an approach which aims to both improve the characterization of sensitivity and computational efficiency using the variogram (measuring variance of differences in the response surface over the parameter space) and quantifying global parameter sensitivity by integrating the variogram across multiple scales (Razavi and Gupta, 2016a). This method has been implemented in the "Variogram Analysis of Response Surfaces" (VARS) framework, the basis of the VARS-TOOL software (Razavi et al., 2019). This tool was selected for the GSA in this study due to the relative efficiency of the method, which was required to apply GSA to a large-scale process-based hydrologic model. The methodology for the application of VARS to the isoWATFLOOD model is illustrated in Figure 4-2.

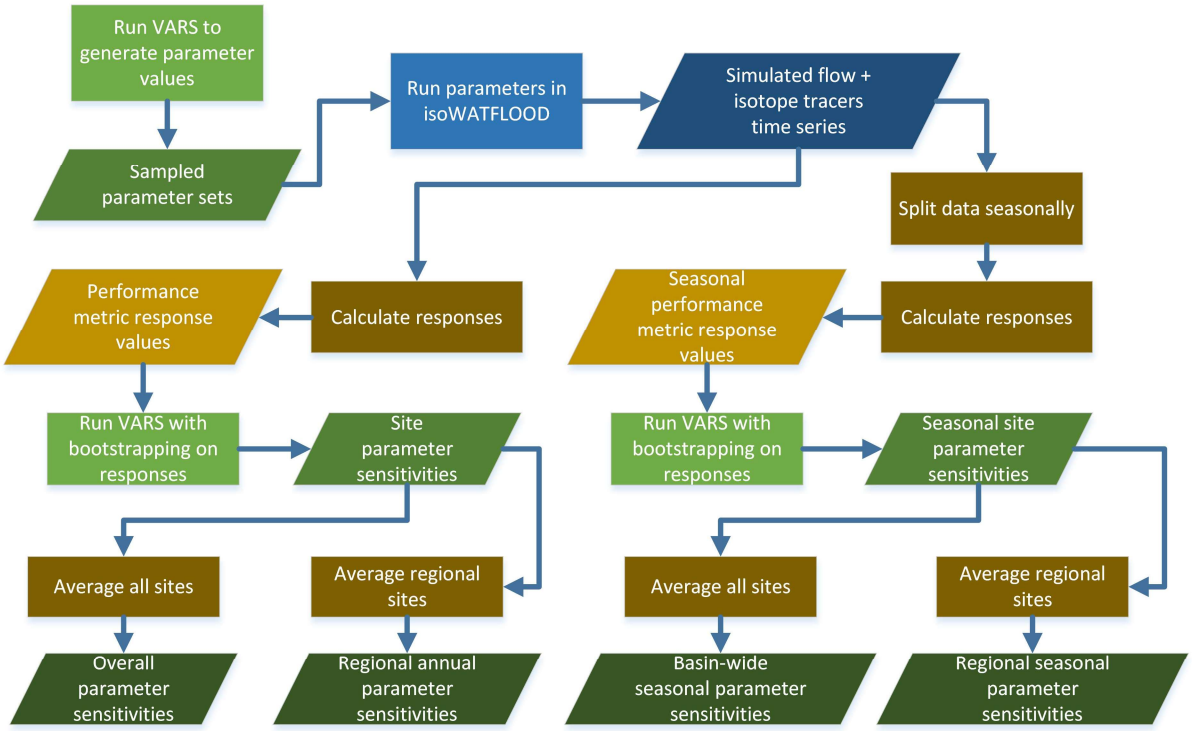

Figure 4-2: Flow chart of the methodology for isoWATFLOOD simulations and generating parameter sensitivities from the VARS analyses. Processes are indicated with rectangles and data with parallelograms; VARS is shaded in green, isoWATFLOOD in blue and external scripts in brown.

The VARS tool was used to generate parameter sets using star sampling (a sampling methodology for computationally efficient coverage of the full parameter space): from 200 star centers, a sampling resolution of 0.1 of the total parameter range used to generate cross sections

of uniformly spaced parameter cross sections, a total of 48,800 parameter sets were generated (specifics of the sampling space are listed in Table C1) (Razavi and Gupta, 2016b). According to Razavi and Gupta (2016b), this sample size was deemed sufficient for VARS to calculate sensitivity reliably for our study. In VARS, the variograms are integrated for multiple perturbation scales, but the $IVARS_{50}$ (integrated variogram across a range of scales, from 0 to 50% of the scale range) index is the most comprehensive index for global, rather than local, sensitivity and is therefore used exclusively in our results (Razavi and Gupta, 2016a). The 90% confidence intervals on the sensitivity results were estimated via the internal VARS bootstrap procedure, using 1000 sampling iterations. All sensitivity results in our study were normalized, with values relative to the total sensitivity of a given metric (i.e. the individual IVARS50 values are normalized using the sum of the IVARS50 for all parameters). Parameter sensitivities were calculated for all gauges and isotope sampling sites, and the corresponding sensitivity indices were averaged, either for the entire watershed, or for all observation points within a defined region. Parameter sensitivities quantify the responsiveness of metrics to changes in parameter values and are not intended to identify goodness of fit or optimal parameter values. Sensitivity analyses were similarly performed on the data that was split seasonally (December-January-February (DJF), March-April-May (MAM), June-July-August (JJA), and September-October-November (SON)) to improve the temporal resolution of the analysis and assess seasonal changes in parameter and process sensitivity.

The individual sensitivities for parameters in a process-based hydrologic model can inform our understanding of the model and basin function, but they can also be grouped to illuminate the sensitivity of the model to specific simulated hydrologic processes. Changing a parameter value changes the simulated process that depends on that parameter value within the model, and hence the contribution of that process to total streamflow. If changing a parameter value, and therefore the magnitude or timing of a process, has no effect on the value of a performance metric, then that metric is considered insensitive to both the parameter and the process. Multiple parameters can influence a process, thus only if the metric is insensitive to all the parameters controlling a process is it truly insensitive to the process. The number of analysed parameters varied between processes, ranging from one to seven individual parameters per process (the parameters associated with each process are provided in Table 4-2). The overall process sensitivity is summarized as the average sensitivity of all individual parameters controlling the process

simulation; assuming that if all parameters were equally sensitive, all processes would have equal shares of sensitivity in the visualizations.

4.4.Results

4.4.1. Parameter Sensitivity

The relative parameter sensitivities for all analysed metrics are presented in Figure 4-3; insensitive parameters are highlighted in blue, and parameters which dominate the metric response are highlighted in orange shades (darker shading is more sensitive/insensitive). The 90% confidence intervals for the relative sensitivity values are indicated by the red bars above the sensitivity values; precise values for the confidence intervals are presented in Table C2 in the supplementary material.

		NRMSE	NSE	logNSE	KGE	β (bias)	α (var)	Correlation	SFDC	Q5	Q95	O NRMSE	O KGE	O β (bias)	O α (var)	O Correlation	O SDC	O Q5	O Q95	H NRMSE	H KGE	H β (bias)	H α (var)	H Correlation	H SDC	H Q5	H Q95	LML mE	LML bE	LML RE
	k F (surf)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	k F (horz)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.11	0.01	0.12	0.01	0.03	0.00	0.08	0.08	0.11	0.03	0.12	0.03	0.02	0.02	0.15	0.08	0.03	0.16	0.04
	PET F	0.00	0.00	0.64	0.11	0.05	0.10	0.01	0.04	0.00	0.07	0.02	0.03	0.01	0.03	0.09	0.04	0.07	0.02	0.00	0.02	0.00	0.02	0.04	0.00	0.00	0.01	0.27	0.07	0.19
Snowpack	melt (gr)	0.07	0.10	0.01	0.05	0.00	0.03	0.09	0.09	0.05	0.00	0.00	0.01	0.00	0.01	0.05	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.07	0.01	0.00	0.00	0.03	0.01	0.03
Snowpack	melt (con)	0.37	0.40	0.00	0.19	0.04	0.10	0.44	0.19	0.17	0.00	0.02	0.01	0.02	0.01	0.05	0.01	0.02	0.04	0.02	0.04	0.01	0.04	0.05	0.06	0.00	0.04	0.06	0.02	0.08
Snowpack	melt (mix)	0.08	0.08	0.00	0.02	0.01	0.02	0.03	0.03	0.05	0.00	0.05	0.01	0.07	0.01	0.03	0.00	0.06	0.04	0.05	0.01	0.07	0.01	0.04	0.00	0.08	0.04	0.02	0.07	0.03
Snowpack	melt (bar)	0.04	0.04	0.00	0.07	0.39	0.03	0.06	0.21	0.08	0.00	0.12	0.47	0.13	0.46	0.14	0.29	0.08	0.11	0.13	0.09	0.13	0.09	0.17	0.16	0.16	0.12	0.03	0.13	0.02
Snowpack	melt (bog)	0.02	0.01	0.01	0.03	0.00	0.02	0.05	0.05	0.02	0.01	0.06	0.01	0.07	0.01	0.02	0.01	0.06	0.06	0.06	0.02	0.07	0.02	0.02	0.05	0.08	0.06	0.03	0.07	0.03
Snowpack	melt (fen)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Snowpack	melt (wat)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.02	0.04	0.00	0.05	0.01	0.00	0.00	0.00	0.00	0.06	0.01	0.00	0.00	0.04	0.01	0.06
Retention	soil (gr)	0.04	0.06	0.00	0.01	0.05	0.00	0.00	0.00	0.06	0.01	0.11	0.01	0.12	0.01	0.03	0.01	0.08	0.08	0.11	0.01	0.12	0.01	0.04	0.04	0.16	0.09	0.03	0.12	0.03
Retention	soil (con)	0.14	0.11	0.01	0.27	0.22	0.41	0.08	0.09	0.33	0.00	0.05	0.01	0.05	0.01	0.05	0.01	0.04	0.04	0.05	0.00	0.05	0.00	0.04	0.02	0.05	0.03	0.04	0.05	0.05
Retention	soil (mix)	0.02	0.02	0.00	0.01	0.05	0.00	0.00	0.01	0.04	0.00	0.04	0.24	0.04	0.25	0.04	0.39	0.03	0.05	0.04	0.29	0.03	0.29	0.04	0.08	0.00	0.05	0.05	0.02	0.05
Retention	soil (bar)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.01	0.01	0.00	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.00	0.01
Retention	soil (shr)	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.05	0.01	0.05	0.01	0.02	0.01	0.05	0.05	0.05	0.03	0.05	0.03	0.02	0.05	0.06	0.05	0.03	0.04	0.03
Retention	soil (wet)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.02	0.00	0.02	0.03	0.01	0.02	0.02	0.00	0.06	0.00	0.06	0.02	0.00	0.00	0.01	0.03	0.01	0.03
	k F (vert)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.01	0.02	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.02	0.01	0.03
Baseflow	C (MID)	0.03	0.02	0.03	0.02	0.00	0.04	0.02	0.01	0.03	0.09	0.10	0.01	0.09	0.01	0.04	0.02	0.05	0.10	0.11	0.11	0.10	0.11	0.04	0.12	0.05	0.10	0.05	0.05	0.05
Baseflow	C (DS)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.02	0.05	0.02	0.02	0.02	0.04	0.06	0.06	0.04	0.05	0.04	0.02	0.08	0.04	0.07	0.02	0.04	0.02
Baseflow	C (US)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.02	0.03	0.01	0.03	0.01	0.08	0.00	0.03	0.01	0.00	0.01
Baseflow	pwr	0.08	0.07	0.03	0.06	0.00	0.07	0.05	0.03	0.04	0.06	0.09	0.05	0.08	0.05	0.05	0.14	0.06	0.09	0.10	0.12	0.09	0.12	0.05	0.14	0.05	0.09	0.05	0.05	0.05
Channel	n (MID)	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.00	0.02	0.07	0.00	0.04	0.03	0.00	0.03	0.00	0.03	0.08	0.01	0.00	0.02	0.04	0.01	0.06
Channel	n (DS)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.01	0.03	0.01	0.02	0.00	0.02	0.04	0.03	0.02	0.02	0.02	0.01	0.01	0.00	0.04	0.03	0.01	0.02
Channel	n (US)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	θ (wet)	0.01	0.01	0.17	0.03	0.00	0.04	0.03	0.04	0.03	0.41	0.04	0.01	0.05	0.01	0.03	0.00	0.05	0.03	0.04	0.00	0.05	0.00	0.04	0.01	0.07	0.03	0.03	0.05	0.03
	k (wet)	0.06	0.04	0.06	0.07	0.00	0.09	0.05	0.11	0.06	0.32	0.01	0.02	0.00	0.02	0.03	0.00	0.02	0.03	0.01	0.04	0.00	0.04	0.03	0.01	0.00	0.03	0.04	0.01	0.04
	glac F	0.02	0.02	0.00	0.05	0.18	0.02	0.06	0.09	0.02	0.00	0.00	0.00	0.00	0.00	0.06	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.02	0.00	0.00	0.01	0.00	0.00

Figure 4-3: Relative parameter sensitivity for assessed metrics; insensitive parameters are highlighted in blue and highly sensitive parameters shaded in orange (darker shading is more sensitive/insensitive). Red bars summarize 90% uncertainty range in sensitivity values (displaying 0 to 0.5 relative sensitivity, where sensitivity values with higher uncertainty have longer bars). Parameter names and descriptions are provided in Table 4-1, and performance metric information in Table 4-4.

The majority of flow simulation performance metrics (left) are sensitive to a consistent subset of parameters and are likewise insensitive to another subset of parameters. In general, flow metrics are highly sensitive to the snowmelt and soil retention parameters in the dominant land class

(coniferous forest). Flow metrics are also sensitive to snowmelt parameters in other prevalent land classes, wetland, evaporation and glacier parameters, and baseflow parameters in the dominate river class. All of these parameters have significant control over peak streamflow volume and timing. There are a few outliers to these general tendencies: logNSE, bias and Q95. The logNSE and Q95 metrics are largely insensitive to snowmelt and soil retention parameters; however, logNSE is extremely sensitive to the evaporation parameter (PET F), while the Q95 metric is highly sensitive to wetland parameters (k_{wet} and θ). The flow bias metric has a higher relative sensitivity to glacier and barren ground snowpack melt and is insensitive to baseflow and wetland parameters.

Flow performance metrics are uniformly insensitive to all three soil conductivity parameters (k_{surf} F, k_{horz} F, k_{vert} F), channel roughness parameters, snowmelt parameters for connected wetlands and open water, and soil retention and baseflow parameters in non-dominant classes. Combined, these trends in relative sensitivity results mean that most flow metrics have correlated sensitivities (correlation values for the sensitivities of all metrics are provided in Table C3). Only the logNSE and Q95 metrics are not highly correlated with any other flow metrics. Confidence intervals on parameter sensitivities are narrow, meaning that parameter sensitivity rankings for flow metrics are not highly dependent on the sampled parameter space (i.e. are reliable, such that a small subset of simulations with extreme results are not determining the overall sensitivity estimate). These narrow confidence intervals are indicative of a smoother response surface, where gradual changes in parameter values do not result in discontinuous changes in performance metric values.

In contrast to flow performance metrics, isotope simulation performance metrics are at least somewhat sensitive to most parameters (fewer blue cells in the right of Figure 4-3). There was no parameter to which all isotope metrics were highly sensitive, but all isotope metrics were at least moderately sensitive to the barren ground snowmelt parameter and the power parameter determining baseflow. Unlike the flow metrics, the majority of isotope metric sensitivities have wide confidence intervals, meaning that the relative sensitivities are often more uncertain for isotope metrics. The isotope simulations' correlations, and the LML slope and R^2 errors have the most reliable sensitivity estimates, with confidence bounds comparable to those of the flow metrics. As isotope performance metrics are all somewhat sensitive to nearly all parameters, the

relative sensitivities for isotope metrics are generally correlated with each other. The LML slope and R^2 errors are least correlated with other isotope metrics due to their high sensitivity to the PET adjustment factor.

Flow and isotope performance metrics have distinct parameter sensitivities, with only a few parameters having similar relative sensitivities for both metric types. The surface conductivity, snowmelt in connected wetlands, and the alpine channel roughness parameters were uninfluential for all metrics; no parameters were sensitive for all metrics. Isotope metric sensitivities were not generally correlated with flow metric sensitivities; of the flow simulation metrics, the bias sensitivities were most similar to isotope metric sensitivities, due to the relative importance of the barren snowmelt parameter for both the flow bias and isotope metrics.

4.4.2. Process Sensitivity

The parameter sensitivities from Figure 4-3 were grouped based on the process they affect, in order to evaluate what metrics in this study are sensitive to specific simulated processes. Overall process sensitivities for all sampling sites or hydrometric gauges (i.e. what would be used in a conventional model calibration) are shown in Figure 4-4.

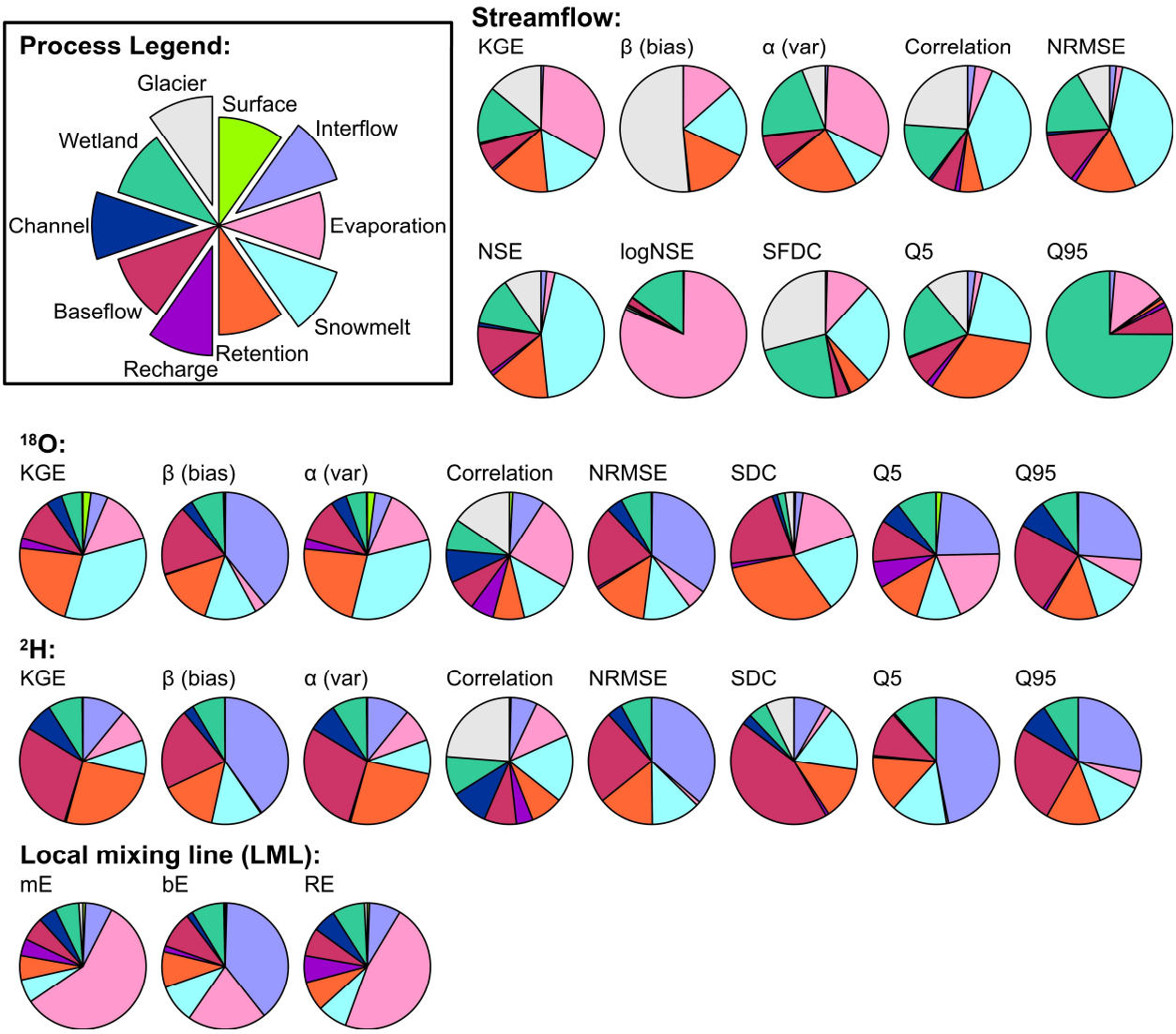

Figure 4-4: Overall relative process sensitivities for all evaluated performance metrics (Table 4-4), averaged for all observation locations over the entire simulation period.

Only six of the ten modeled processes dominate flow performance metric sensitivities, with some variation in the relative shares for each metric. The channel and upper zone soil fluxes have small or negligible shares of the overall flow metric sensitivity. Many of the evaluated metrics' sensitivities are dominated by a single process, particularly evaporation for logNSE, and wetlands for Q95, and (to a lesser degree) glacier for bias, and snowmelt for NRMSE, NSE and correlation. The KGE metric has the most balanced shares of process sensitivity, for the six processes dominating the streamflow response.

There is a substantial variation in the process sensitivities for isotope simulation performance metrics, possibly due in part to the uncertainty in the parameter sensitivity values. Isotope metrics are consistently sensitive to soil retention, snowmelt and baseflow, and generally sensitive to interflow (i.e., the horizontal upper soil zone flux). No single isotope metric is sensitive to all processes, but there at least one isotope metric is slightly sensitive for all ten processes. The second isotope simulation, ^{2}H, was not needed to cover all ten processes, since the ^{18}O simulation alone is sensitive to all processes. Given the uncertainty in many of the sensitivity estimates for isotope tracer metrics, the differing sensitivities of the two tracers is likely insignificant.

In comparison to the flow metrics, isotope metrics cover sub-surface soil processes better and emphasize the channel roughness more. However, only the isotope correlations are highly sensitive to glacier melt, meaning isotope simulations do not generally respond to changes in simulated glacier melt. The isotope metrics have smaller shares of the overall sensitivity dedicated to wetlands, evaporation, and snowmelt than flow metrics, although they are significant processes for both types of metric.

Performance metrics can be calculated on a seasonal basis, and observation locations can likewise be geographically separated for a more detailed analysis of process sensitivity across the Athabasca region (Figures 4-5 and 4-6, for flow and isotope metrics respectively).

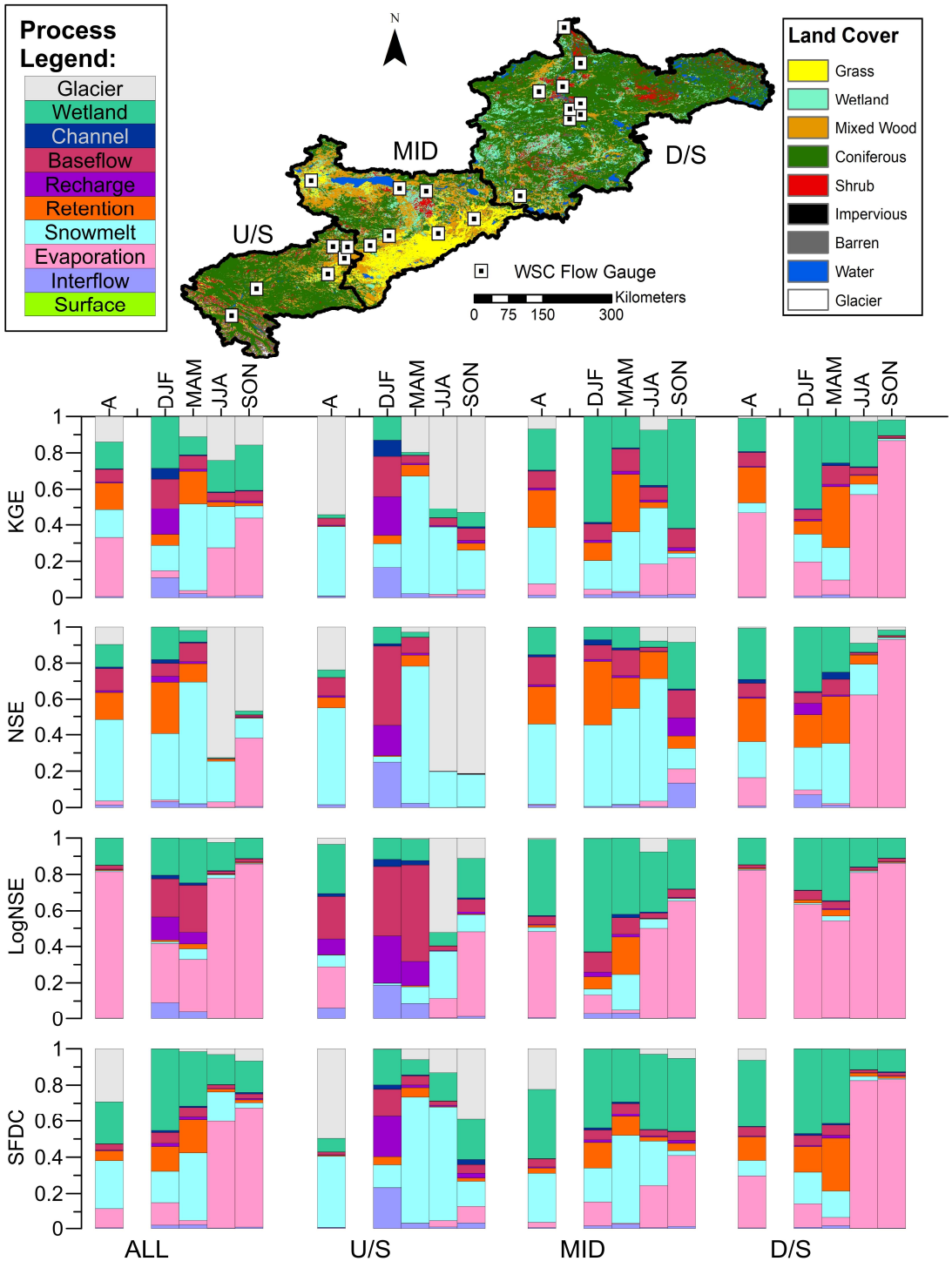

Figure 4-5: Regional flow metric process sensitivity for the upstream (U/S), mid-basin (MID) and downstream (D/S) reaches, and basin-wide (ALL), temporally aggregated by season (DJF-December, January, February; MAM-March, April, May, JJA-June, July, August, SON-September, October, November) and full year (A).

As would be expected from the overall annual results in Figure 4-4, the KGE, NSE and SFDC metrics are sensitive to six of the ten simulated hydrologic processes for either the average of all sites in the basin (ALL), or the basin-wide annual (A) sensitivities, with some variation in the ranking of these processes. NSE is generally more sensitive to snowmelt, KGE is more consistently sensitive to evaporation, and glacier melt and wetlands are most influential over the annual SFDC. It is important to note that cumulative error metrics have 'overall' sensitivities that can be estimated from subsets, with weighting (e.g. overall snowmelt sensitivity for the NSE will be between the maximum and minimum seasonal sensitivity). This is not the case for population-based error metrics: the SFDC calculated from seasonally separated data may have different sensitive parameters than the SFDC calculated from unseparated data (e.g. SFDC more sensitive to glaciers when considering whole years, than seasonally separately data).

Glacier melt is frequently the most influential process in the mountainous upstream, headwater region of the Athabasca basin in the summer and fall but is largely irrelevant to flow metrics in the mid- or downstream basins. Among flow metrics, snowmelt is an influential process across the entire basin, but it is most sensitive in the period covering the freshet (MAM and to some extent, JJA). Similarly, evaporation is most influential in summer and fall, when the bulk of evaporation loss occurs. In contrast to these seasonally varying processes, flow metrics are sensitive to wetland fluxes year-round, more so in the lower slope mid- and downstream regions of the basin where wetlands are more prevalent. Sub-dividing data (from Figure 4-4) geographically and temporally does render some soil flux sensitivities noticeable. All flow metrics in Figure 4-5 are sensitive to interflow and lower zone recharge in the lowest flow data sub-division: winter flows in the upstream basin. Other low flow times and locations are likely to be sensitive to interflow and recharge; these times and places are also likely to have baseflow as a relatively sensitive process. Soil retention is an influential process but has the highest relative sensitivity in spring (MAM), coinciding with most snowmelt, indicating that the primary cause of flow metrics' sensitivity to retention is the soil's ability to absorb snowmelt. No flow metric is sensitive to surface infiltration during any season or region; therefore, these metrics are completely insensitive to whether water on the soil surface infiltrates or runs off directly to wetlands. Flow metrics can be somewhat sensitive to channel parameters, in very low-flow, or high (peak) flow periods. Parameters controlling flow velocity in the channel can influence flow

simulation timing, but wetlands have a much larger influence on the magnitude and timing of streamflow in this watershed.

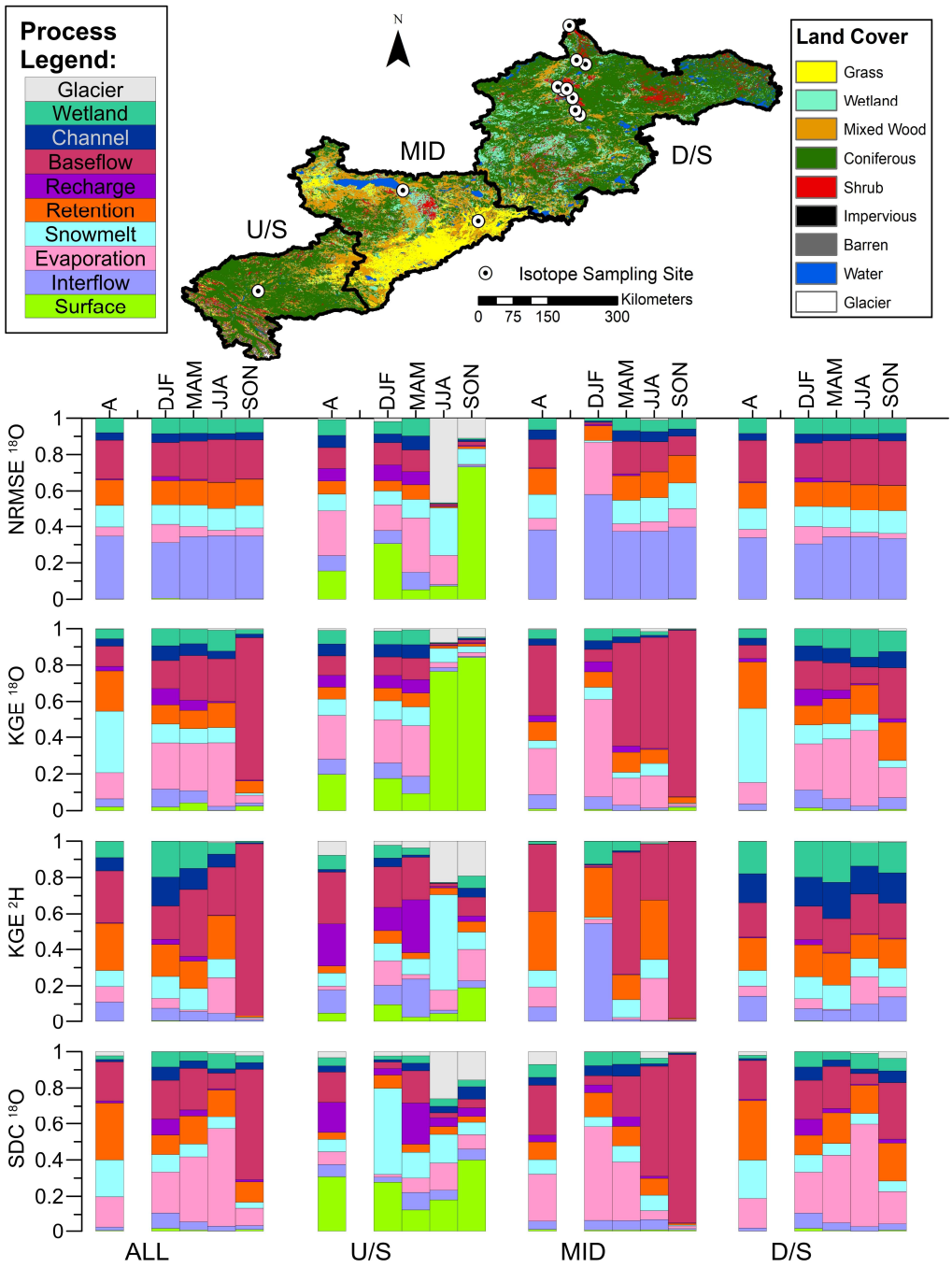

Figure 4-6: Regional isotope tracer metric process sensitivities for the upstream (U/S), mid-basin (MID) and downstream (D/S), and basin-wide (ALL), temporally aggregated by season (DJF-December, January, February; MAM-March, April, May, JJA-June, July, August, SON-September, October, November) and full year (A).

Isotope sampling locations are biased toward the downstream portion of the basin, in the oil sands region of the Athabasca River basin (reference map, Figure 4-6). While the data sampling resolution in the upstream and mid-basin regions are of the same quality as the best locations in the downstream region, the smaller number of sites limits confidence in the generalizability of the sensitivities for the upstream and mid-basin regions. The large confidence intervals on isotope metric sensitivity (Figure 4-3) likewise adds to the uncertainty in regional and seasonal process sensitivity results.

Some general observations from the isotope metric sensitivities may still, however, be drawn. Soil fluxes are the main theme of isotope sensitivity: all isotope performance metrics are sensitive to some combination of infiltration, interflow, recharge, soil retention and baseflow, with these soil processes dominating the sensitivity in most seasons and regions. Baseflow is approximately the only sensitive process in the mid-basin during fall (SON). The isotope simulation at the upstream sampling site is clearly sensitive to surface infiltration, and other downstream sites are not completely insensitive. Isotope performance metrics are more sensitive to recharge in winter (DJF) and spring (MAM), but there is no clear seasonal pattern to soil water retention sensitivity for isotope metrics, unlike flow metrics. Overall, there is considerable variation in the most sensitive processes for the various metrics, locations, and time periods, and every modeled process is significant in at least one relative sensitivity sub-division.

Interestingly, snowmelt and evaporation are not generally the most influential processes for isotope metrics, even though both have distinctive signals in the isotope data; these processes are also sensitive processes outside of their main seasons of occurrence. Isotope metrics are much less sensitive to glacier melt than flow metrics, though it remains an influential process in the headwater basin during the summer and fall. On the other hand, compared to flow metrics, isotope metrics are more sensitive to channel velocity.

The downstream transfer of process sensitivity for flow and isotope tracer metrics is illustrated in Figure 4-7, with a pair-wise comparison of sensitivities for nested watersheds along the Athabasca River mainstem, from upstream to downstream.

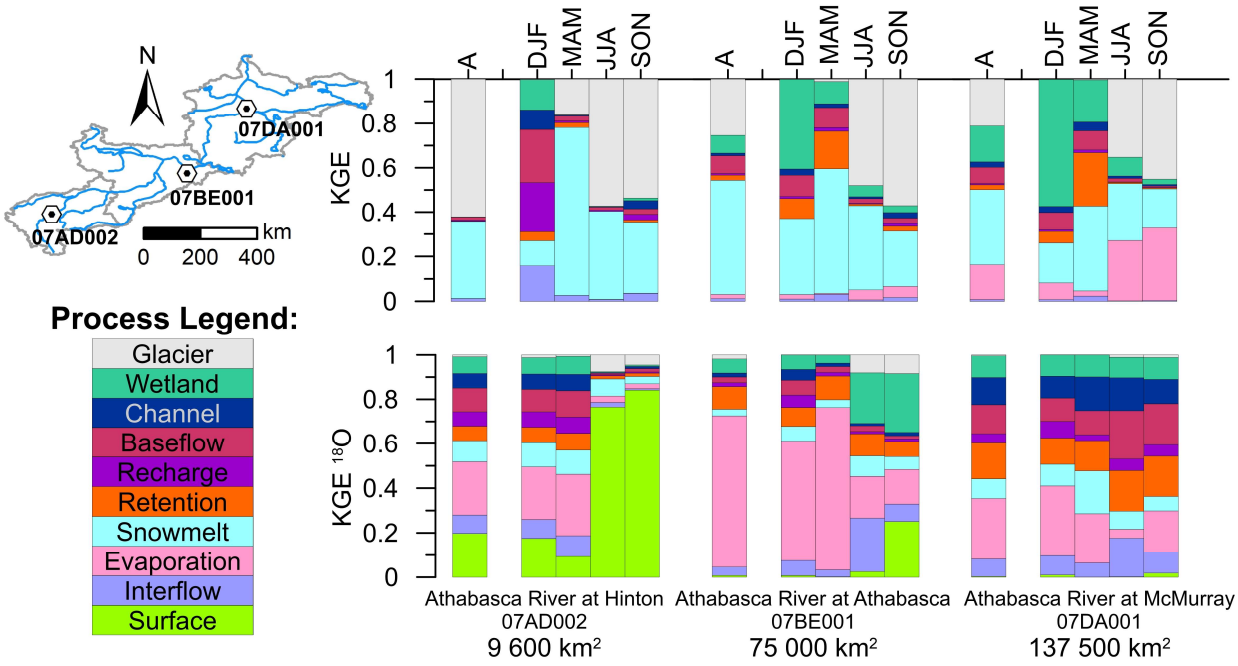

Figure 4-7: Flow and isotope tracer KGE process sensitivities for sites along the Athabasca River mainstem demonstrating the downstream transfer of process sensitivity aggregated seasonally (DJF, MAM, JJA, SON) and full year (A).

The flow metric sensitivity shows significant downstream transfer of process sensitivity along the mainstem of the Athabasca River; glacier melt is the dominant process in the Athabasca headwaters and remains a significant process in the simulation at Fort McMurray, over 1000 km downstream. The flow gauges on the Athabasca River itself are outliers from the regional sensitivities (Figure 4-5) due to influence of upstream areas: glacier and snowmelt have much larger shares of the relative sensitivity. Isotope tracer sensitivities, in contrast, have limited downstream transfer of process sensitivity. At the downstream end of the Athabasca River, isotope tracer sensitivities closely resemble those of local tributaries rather than upstream gauges. Isotope tracer sensitivities for the largest watershed areas also have limited seasonal variation, unlike flow sensitivities for the same observation location.

4.5. Discussion

4.5.1. Evaluating with blinders: what flow-based metrics 'see'

Flow simulation performance metrics respond primarily to processes with substantial influence on either water volume (i.e. evaporation and glacier melt) or peak flow timing (i.e. snowmelt,

wetland and soil retention). The processes with the largest influence on flow metric values vary significantly within the Athabasca River basin. Glacier melt is hugely significant to most flow simulation performance metrics in the mountain headwaters of the river, but averaged flow metrics are less sensitive to the magnitude of glacial melt outside the alpine region. Conversely, wetland retention and evaporation make only a modest contribution to flow metric responses in the headwaters, where there is a limited area of wetlands and open water, but these two processes are among the most important to flow metrics in the downstream oil sands region. This finding is supported by Gibson et al (2019) who found similar relationships between headwater and lowland regions using isotope-derived estimates of water yield. Snowmelt and soil water retention are sensitive processes across the entire Athabasca River basin, but the scope of their influence on flow metric response is somewhat limited temporally: the spring freshet (MAM) is most sensitive to both melt rates and the upper soil zone water retention capacity. From the timing of maximum sensitivity to soil water retention, it is clear that flow metrics respond primarily to the capacity of the upper zone to absorb runoff and damp peak flows, rather than its ability to retain water in the longer term and affect evapotranspiration. Based on the flow simulation alone, it would appear that spatial variation in the most influential processes within the Athabasca basin depends only on the prevalence of glaciers or wetlands.

Although the model was not calibrated in this study, sensitivity results indicate that tuning the simulation of just six of the ten modeled processes in the model would be sufficient to generate a 'good' flow simulation, assuming a 'good' simulation is considered to be one where the simulated hydrographs closely resemble the observed hydrographs. If this simple measure of accuracy is sufficient for the intended application of the hydrologic model, a simulation that is well-calibrated to optimize KGE or some other combination of flow simulation performance metrics can be considered fit for its purpose (e.g., short-term peak flow forecasting). In such situations, however, it would be unclear why a process-based hydrologic model is being used in the first place. The blind spots of flow metrics are of concern for potential model applications where soil fluxes are important, or where the fidelity of process representation matters. For example, when total basin storage, water age or flow paths are relevant outputs (i.e. for water supply assessments or long-term climate change studies), calibrating the model to optimize only flow performance will likely prove to be inadequate (Kirchner, 2006). In fact, this study demonstrates that increasing structural complexity (i.e. more parameters and processes) is likely

to result in a larger decision space of 'acceptable' solutions derived from different hydrologic partitions, or proportional flow path contributions (Figures 4-3 and 4-4). It should be noted that many applications of hydrologic models and scenarios in the Canadian Oil Sands region require the accurate simulation of both water storage and flow paths for contaminant tracing or climate and land use impact assessment.

4.5.2. The added value of isotope-aided metrics

Isotope simulation performance metrics are sensitive to a wider variety of model processes than flow metrics (Figure 4-3). It is well-known from previous research that some processes are only locally or periodically hydrologically significant, and that parameter sensitivities therefore change depending on the time period or location within the watershed (Herman et al., 2013; Höllering et al., 2018). We show here that there are, in fact, *no* processes that flow performance metrics are sensitive to that isotope performance metrics are not (Figure 4-4). In addition to those processes which flow metrics are sensitive to, isotope metrics are additionally sensitive to soil water fluxes (*k F, C,* and *pwr*) and channel roughness (*n*), and are more sensitive to soil water storages (*soil*). In our modeling, isotope metrics were sensitive to infiltration rates in association with exposed or barren ground, lower and upper soil zone flows in association with grassland and mixed deciduous and coniferous forest, and a mix of soil and wetland properties with channel velocities in association with wetlands and coniferous forest (Figure 4-6). Snowmelt and evaporation were less influential than soil processes, in spite of their importance to the water balance, although this may be an artifact of the sampling resolution. Model sensitivity to sub-surface processes is a reflection of the significance of mixing volumes in isotope tracer simulations; the variability of the isotope tracer simulation in other models is also largely determined by the volume of simulated water in storage (the water age) (Birkel et al., 2011; Klaus et al., 2015; Rodriguez and Klaus, 2019). Flow simulations are dependent on the amount and timing of flow, while an isotope tracer simulation, with a concentration output, is dependent on the age of flow (flow path length) and fractionation processes (surface versus subsurface flow paths). Flow metrics are therefore better at detecting flow volume or timing errors, such as glacier melt rate errors (Figure 4-5), and isotope tracer metrics respond most to errors in flow paths, such as soil water flux rates (Figure 4-6).

There is little seasonal variation in process sensitivity for isotope tracers in the downstream region of the Athabasca River basin; the inherent mixing within large upstream areas, sub-surface storages or extensive regional wetlands limits the temporal variation in isotope data (isotope data provided in Appendix C). Isotope tracers are therefore better at providing information on processes in smaller basins than larger ones. The damped isotope signals in larger watersheds provides useful information on long-term process contributions, but a high-resolution isotope dataset from a smaller watershed can clarify individual process contributions for specific events or types of events, as illustrated by Figure 4-7. Regions of low water yield have been reported within the middle and downstream portions of the Athabasca River basin, where it is believed there may be buried channels and a shift toward more vertical flow exchange as opposed to lateral surface runoff (Gibson et al., 2019). In fact, isotope tracer process sensitivities reflect this finding from surface water dominance in the headwater basin toward more soil storage dominated processes in the mid and downstream reaches; this is, however, not reflected in flow-based process sensitivities (Figure 4-7). To specifically diagnose lateral and vertical flow exchange processes, other tracer types or enhanced sampling resolution may be necessary to fully delineate the geographical and temporal significance of these processes. Isotope tracer datasets can be better leveraged when sub-basin scale or type are explicitly considered: a single headwater sampling site, or a site with different land cover or topography, can add far more valuable information than adding more gauges along the mainstem of a river.

Isotope tracer metrics in conjunction with flow metrics provide a more complete picture of the influential processes in the Athabasca River basin than flow metrics alone. When both data types are considered, the alpine headwaters are affected not only glacial and snow melt, but also infiltration and surface runoff (Figure 4-6). In the central portion of the Athabasca watershed, isotope tracer data can highlight the importance of soil storage and fluxes year-round, which flow performance metrics ignore. Both isotope tracer and streamflow simulations agree on the critical importance of wetlands and evaporation rates in the downstream regions of the Athabasca River, but isotope tracer metrics are also responsive to the path through the soil that water takes to reach those wetlands (Figure 4-6), which is intrinsically linked to residence time or water age. There is also an interesting possibility that for processes both flow and isotope performance metrics are sensitive to, the different simulation types may have contradictory optimization outcomes. Evaporation, for example, reduces simulated flow but increases both the

magnitude and annual variability of isotope tracer concentrations (i.e. seasonally enriched streamflow); what may appear to be equifinality when streamflow alone is considered, may not be equivalent when both tracers and streamflow are evaluated (Beven, 2006; Kirchner, 2006). For example, a model optimized with a flow performance metric could therefore have a much higher evaporative loss than the same model optimized with an isotope performance metric (Holmes et al., 2020).

This study utilized one hydrologic model with multiple simulated outputs to produce a suite of flow *and* isotope tracer-based simulations. The exact proportions we report for process sensitivity in relation to various metrics and proportion of simulated flow are specific to the hydrologic model used (here, isoWATFLOOD) as they are a reflection of the model's internal structure and the algorithms that numerically define each process. Our findings, however, are model agnostic in terms of the cautionary tale they tell of over-reliance on flow data for model evaluation, or rather the missing information content when calibration is based on flow data alone. The value of adding isotope tracer data is that water age and flow paths are directly incorporated into model evaluation, which correlates to internal process function and model structure. This outcome would occur generally for physically based models, as it is a reflection of adding metrics and data capable of diagnosing such storage and flux interactions. Some findings relating to snowmelt sensitivity are only transferrable to watersheds under similar climates (i.e., seasonal basins in mid- to high-latitude regions), and outcomes would differ for lower latitude regions experiencing exclusively rainfall and much higher proportions of evaporative loss. The actual value added by the isotope tracers in any particular application ultimately depends on the degree to which the isotopes fractionate throughout the regional hydrologic cycle, and the isotope concentration distinctness of processes (or end members) in the watershed.

4.5.3. On the selection of performance metrics for model calibration

The various isotope and flow performance metrics have different relative advantages in the context of hydrologic model parameter calibration. Flow metrics consistently have reliable parameter sensitivities (Figure 4-3), due to a regular response surface for flow performance. These reliable sensitivities are advantageous in model calibration as they identify consistently insensitive parameters and remove them from the calibration; a smoother performance response

surface also facilitates searching in optimization. KGE is the flow performance metric with the broadest range of process sensitivities, and it is therefore the best choice for a stand-alone flow metric in a process-based optimization (out of those evaluated in this study). The NSE is a possible alternative, although unlike the KGE, it is skewed toward snowmelt (i.e. the primary peak flow generating mechanism, with high magnitude residual error, in the Athabasca basin). Using more specialized metrics, such as logNSE or flow signatures, can highlight particular processes, but these metrics did not respond strongly to processes also not covered by the KGE metric (Figure 4-4). Juggling different metrics or data subsets (e.g. Q_{95} or alpine flow gauges) can highlight particular processes (e.g. wetland fluxes or glacial melt) far better than averaged general response metrics such as KGE, but does little to expose the internal soil processes. The components of KGE may be just as useful as specialized metrics for rebalancing process sensitivities. When only streamflow is evaluated, process-based hydrologic models can behave as something like a black box for simulated flow pathways, since streamflow simply tracks how much water comes out of the landscape (Blöschl et al., 2019). Streamflow performance is therefore unresponsive to changes in water flow paths alone; flow metrics are indifferent to how much water is stored internally within a simulation, or how long precipitation takes to reach the channel network, as long as the correct volume of water reaches the river at the right time.

In contrast to flow metrics, many isotope performance metrics have low reliability for parameter sensitivity (Figure 4-3), in that a small region of the parameter space can have an outsized influence over the relative sensitivity of a parameter. As an example, the isotope concentration simulation can perform extremely poorly if one combination of soil conductivity and soil water retention parameters results in the desiccation of a fractionating storage unit, yet desiccation (and poor simulation performance) can be avoided by slight changes in any one of three parameters. Every isotope performance metric included in the analysis was sensitive to soil fluxes and storage, indicating that responsiveness to internal flow paths is an inherent property of the isotope tracer simulation, which the flow simulation alone does not have. Isotopic sensitivity to subsurface flow paths has been identified previously in the literature (Delavau et al., 2017; Stadnyk and Holmes, 2020) which is broadly why isotopes are considered excellent hydrologic tracers (Klaus and McDonnell, 2013). Therefore, unlike flow signatures, isotope tracer performance metrics can cover processes missed by streamflow KGE. An isotope tracer

simulation can produce a better-informed hydrologic model, but the utility of the added information is dependent on the application.

Isotope performance metrics have a larger number of sensitive parameters than flow performance metrics (Figure 4-3); including isotope metrics in model optimization therefore increases the scope of the optimization: more parameters need to be included in the optimization, but more parameters will actually be optimized. Both the isotope tracer and seasonal flow metric sensitivity results oppose the common practice of removing parameters from calibration based on simple sensitivity analyses: the considerable variation in sensitive processes for the various metrics, locations, and time periods meant every modeled process is significant to the model at some place or time. The most reliable isotope parameter sensitivity metrics were the LML errors and the correlations between simulated and observed isotope data; they are relatively unaffected by desiccation events in the simulation which can lead to substantially different model responses in highly localized parts of the decision space (Sahraei et al., 2020).

The more reliable isotope sensitivity estimates (i.e., correlation and NRMSE) have similar process sensitivities for both tracers, which is anticipated under similar atmospheric forcing. However, the advantage of simulating both isotopes is that it allows the calculation of a simulated LML, and therefore LML errors; of the isotope metrics, LML error metrics were the most sensitive to evaporation. No evidence was found to support using KGE for isotope simulation evaluation in place of the traditional residual error metrics (e.g., NRMSE). KGE sensitivities were no more reliable, and the same processes were influential for both NRMSE and KGE metrics. Furthermore, the KGE sensitivity was dominated by its variability component, however, trying to evaluate the variability error of a simulation based on sporadic observations is highly dubious. Just as using the variability in observations to normalize the squared error (i.e. using the NSE) is not recommended for discontinuous data because the data sample may not be representative of the true population variability, the variability error of the KGE metric is not recommended for calibrating tracer simulations when only sparse observations are available. The correlation or bias components of the KGE are better supplements to a residual error metric in isotope simulation evaluation (e.g. in a multi-objective calibration problem formulation) with discontinuous or sparse observed datasets.

Simulating both isotope tracers does not increase the number of sensitive processes, as all processes are sensitive to some degree to either of the two tracers. The differing sensitivities of the two tracers for some metrics cannot be attributed to the properties of the tracers due to the uncertainties in the tracer sensitivity results. It must also be noted that this analysis has not been extended to include either uncertainty in observed data values, or from sampling (analytic uncertainties for isotope data are relatively low, however observations are sparse both spatially and temporally). Multi-objective optimization methods are highly suitable for calibrating hydrologic models with both tracer and flow data, as they allow a transparent choice in the trade-off between simulation qualities; the importance and uncertainty of an accurate tracer simulation can be balanced by the modeller.

4.6. Conclusions

This study highlights the important regional hydrologic differences between the upper, middle, and lower basins of the Athabasca River. The Oil Sands Monitoring program is concerned with cumulative effects assessment, which requires knowledge of the impacts to more than just streamflow (or total volume), and accurate projections of future water supply depend on the accurate partitioning of processes controlling the overall water balance. A 'black box' model calibrated without specific consideration of these process can – and likely will – result in inaccurate partitioning of water in soils, which directly influences projection of evapotranspiration (air-land), and infiltration or baseflow (land-subsurface) flow paths, skewing future projections of streamflow.

The scope of this study was limited to sensitivity analyses in the Athabasca River basin, but there are some conclusions applicable to model calibration or evaluation more generally:

- Flow simulation performance metrics alone provide an incomplete picture of hydrologic process regional variation and significance.
- KGE is the best stand-alone flow performance metric for process-based optimization as it exhibited the broadest range of process sensitivity.
- Flow signature metrics can highlight specific processes already covered by the generalized KGE but do not add new ones.
- Including an isotope tracer simulation expands the number of processes which can be evaluated.

- Residual error metrics, or bias and correlation are all reasonable measures of simulation performance for isotope tracers.

In conclusion, we suggest that a process-based hydrologic model cannot be considered fully calibrated if the performance of the model is only evaluated with streamflow metrics, because it is not possible for an optimization to tune parameters or processes to which the calibration objective is insensitive. Either the streamflow-insensitive internal flux simulation should be ignored as unreliable, or the model calibration should be expanded to include relevant datasets. Isotope tracers have demonstrable value for informing process-based hydrologic model calibration, although further research is needed on isotope-enabled calibration methodologies and the effects of metric choice on simulated streamflow-generating processes.

Acknowledgements

The authors acknowledge this study occurred within and about Treaty 8 and 6 regions, lands which are, or have historically been, home to no less than nine Indigenous peoples of Canada: the Dane-zaa, Sekani, Secwepemc (Shuswap), Salish, Ktunaxa, Nakoda/Stoney, Woodland Cree, Chipewyan (Denesoline), and Métis. The origin of the meaning behind 'Athabasca River' is derived from the Woodland Cree word *aeapaskāw* meaning" where there are plants one after another". The authors gratefully acknowledge those who have contributed to data collection required to conduct this study, including the Water Survey of Canada, Environment and Climate Change Canada and Innotech Alberta. This research was supported by the Natural Sciences and Engineering Research Council of Canada [CRD 462584-2013], and Global Water Futures [NSERC CFREF-GWF 418474].

4.7.References

Acero Triana, J.S., Chu, M.L., Guzman, J.A., Moriasi, D.N., Steiner, J.L., 2019. Beyond model metrics: The perils of calibrating hydrologic models. Journal of Hydrology 578. https://doi.org/10.1016/j.jhydrol.2019.124032

Alberta Geological Survey, 2013. Bedrock Geology of Alberta [WWW Document]. URL https://open.canada.ca/data/en/dataset/5155d48c-ce34-4493-b4f6-fb4eb94fb348 (accessed 9.20.21).

Bajracharya, A., Awoye, H., Stadnyk, T., Asadzadeh, M., 2020. Time variant sensitivity analysis of hydrological model parameters in a cold region using flow signatures. Water (Switzerland) 12. https://doi.org/10.3390/W12040961

Belachew, D.L., Leavesley, G., David, O., Patterson, D., Aggarwal, P., Araguas, L., Terzer, S., Carlson, J., 2016. IAEA Isotope-enabled coupled catchment–lake water balance model, IWBMIso: description and validation†. Isotopes in Environmental and Health Studies 52, 427–442. https://doi.org/10.1080/10256016.2015.1113959

Bennett, N.D., Croke, B.F.W., Guariso, G., Guillaume, J.H.A., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, L.T.H., Norton, J.P., Perrin, C., Pierce, S.A., Robson, B., Seppelt, R., Voinov, A.A., Fath, B.D., Andreassian, V., 2013. Characterising performance of environmental models. Environmental Modelling and Software 40, 1–20. https://doi.org/10.1016/j.envsoft.2012.09.011

Beven, K., 2006. A manifesto for the equifinality thesis, in: Journal of Hydrology. pp. 18–36. https://doi.org/10.1016/j.jhydrol.2005.07.007

Birkel, C., Soulsby, C., 2015. Advancing tracer-aided rainfall-runoff modelling: A review of progress, problems and unrealised potential. Hydrological Processes 29, 5227–5240. https://doi.org/10.1002/hyp.10594

Birkel, C., Soulsby, C., Tetzlaff, D., 2011. Modelling catchment-scale water storage dynamics: Reconciling dynamic storage with tracer-inferred passive storage. Hydrological Processes 25, 3924–3936. https://doi.org/10.1002/hyp.8201

Blöschl, G., Bierkens, M.F.P., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., Kirchner, J.W., McDonnell, J.J., Savenije, H.H.G., Sivapalan, M., Stumpp, C., Toth, E., Volpi, E., Carr, G., Lupton, C., Salinas, J., Széles, B., Viglione, A., Aksoy, H., Allen, S.T., Amin, A., Andréassian, V., Arheimer, B., Aryal, S.K., Baker, V., Bardsley, E., Barendrecht, M.H., Bartosova, A., Batelaan, O., Berghuijs, W.R., Beven, K., Blume, T., Bogaard, T., Borges de Amorim, P., Böttcher, M.E., Boulet, G., Breinl, K., Brilly, M., Brocca, L., Buytaert, W., Castellarin, A., Castelletti, A., Chen, X., Chen, Yangbo, Chen, Yuanfang, Chifflard, P., Claps, P., Clark, M.P., Collins, A.L., Croke, B., Dathe, A., David, P.C., de Barros, F.P.J., de Rooij, G., di Baldassarre, G., Driscoll, J.M., Duethmann, D., Dwivedi, R., Eris, E., Farmer, W.H., Feiccabrino, J., Ferguson, G., Ferrari, E., Ferraris, S., Fersch, B., Finger, D., Foglia, L., Fowler, K., Gartsman, B., Gascoin, S., Gaume, E., Gelfan, A., Geris, J., Gharari, S., Gleeson, T., Glendell, M., Gonzalez Bevacqua, A., González-Dugo, M.P., Grimaldi, S., Gupta, A.B., Guse, B., Han, D., Hannah, D., Harpold, A., Haun, S., Heal, K., Helfricht, K., Herrnegger, M., Hipsey, M., Hlaváčiková, H., Hohmann, C., Holko, L., Hopkinson, C., Hrachowitz, M., Illangasekare, T.H., Inam, A., Innocente, C., Istanbulluoglu, E., Jarihani, B., Kalantari, Z., Kalvans, A., Khanal, S., Khatami, S., Kiesel, J., Kirkby, M., Knoben, W., Kochanek, K., Kohnová, S., Kolechkina, A., Krause, S., Kreamer, D., Kreibich, H., Kunstmann, H., Lange, H., Liberato, M.L.R., Lindquist, E., Link, T., Liu, J., Loucks, D.P., Luce, C., Mahé, G., Makarieva, O., Malard, J., Mashtayeva, S., Maskey, S., Mas-Pla, J., Mavrova-Guirguinova, M., Mazzoleni, M., Mernild, S., Misstear, B.D., Montanari, A., Müller-Thomy, H., Nabizadeh, A., Nardi, F., Neale, C., Nesterova, N., Nurtaev, B., Odongo, V.O., Panda, S., Pande, S., Pang, Z., Papacharalampous, G., Perrin, C., Pfister, L., Pimentel, R., Polo, M.J., Post, D., Prieto Sierra, C., Ramos, M.-H., Renner, M., Reynolds,

J.E., Ridolfi, E., Rigon, R., Riva, M., Robertson, D.E., Rosso, R., Roy, T., Sá, J.H.M., Salvadori, G., Sandells, M., Schaefli, B., Schumann, A., Scolobig, A., Seibert, J., Servat, E., Shafiei, M., Sharma, A., Sidibe, M., Sidle, R.C., Skaugen, T., Smith, H., Spiessl, S.M., Stein, L., Steinsland, I., Strasser, U., Su, B., Szolgay, J., Tarboton, D., Tauro, F., Thirel, G., Tian, F., Tong, R., Tussupova, K., Tyralis, H., Uijlenhoet, R., van Beek, R., van der Ent, R.J., van der Ploeg, M., van Loon, A.F., van Meerveld, I., van Nooijen, R., van Oel, P.R., Vidal, J.-P., von Freyberg, J., Vorogushyn, S., Wachniew, P., Wade, A.J., Ward, P., Westerberg, I.K., White, C., Wood, E.F., Woods, R., Xu, Z., Yilmaz, K.K., Zhang, Y., 2019. Twenty-three unsolved problems in hydrology (UPH) – a community perspective. Hydrological Sciences Journal 64, 1141–1158. https://doi.org/10.1080/02626667.2019.1620507

Bowen, G.J., Cai, Z., Fiorella, R.P., Putman, A.L., 2019. Isotopes in the Water Cycle: Regional- to Global-Scale Patterns and Applications. Annual Review of Earth and Planetary Sciences. https://doi.org/10.1146/annurev-earth-053018

Buttle, J.M., Allen, D.M., Caissie, D., Davison, B., Hayashi, M., Peters, D.L., Pomeroy, J.W., Simonovic, S., St-Hilaire, A., Whitfield, P.H., 2016. Flood processes in Canada: Regional and special aspects. Canadian Water Resources Journal 41, 7–30. https://doi.org/10.1080/07011784.2015.1131629

Carlisle, D.M., Wolock, D.M., Meador, M.R., 2011. Alteration of streamflow magnitudes and potential ecological consequences: a multiregional assessment. Frontiers in Ecology and the Environment 9, 264–270. https://doi.org/10.1890/100053

Clark, M.P., Bierkens, M.F.P., Samaniego, L., Woods, R.A., Uijlenhoet, R., Bennett, K.E., Pauwels, V.R.N., Cai, X., Wood, A.W., Peters-Lidard, C.D., 2017. The evolution of process-based hydrologic models: Historical challenges and the collective quest for physical realism. Hydrology and Earth System Sciences 21, 3427–3440. https://doi.org/10.5194/hess-21-3427-2017

Clark, M.P., Kavetski, D., Fenicia, F., 2011. Pursuing the method of multiple working hypotheses for hydrological modeling. Water Resources Research 47. https://doi.org/10.1029/2010WR009827

Clark, M.P., Vrugt, J.A., 2006. Unraveling uncertainties in hydrologic model calibration: Addressing the problem of compensatory parameters. Geophysical Research Letters 33, L06406. https://doi.org/10.1029/2005GL025604

Coulibaly, P., Samuel, J., Pietroniro, A., Harvey, D., 2013. Evaluation of Canadian national hydrometric network density based on WMO 2008 standards. Canadian Water Resources Journal 38, 159–167. https://doi.org/10.1080/07011784.2013.787181

Delavau, C., Chun, K.P., Stadnyk, T., Birks, S.J., Welker, J.M., 2015. North American precipitation isotope (δ18O) zones revealed in time series modeling across Canada and northern United States. Water Resources Research 51, 1284–1299. https://doi.org/10.1002/2014WR015687

Delavau, C., Stadnyk, T., Birks, J., 2011. Model based spatial distribution of oxygen-18 isotopes in precipitation across Canada. Canadian Water Resources Journal 36. https://doi.org/10.4296/cwrj3604875

Delavau, C.J., Stadnyk, T., Holmes, T., 2017. Examining the impacts of precipitation isotope input δ18Oppt) on distributed, tracer-aided hydrological modelling. Hydrology and Earth System Sciences 21, 2595–2614. https://doi.org/10.5194/hess-21-2595-2017

Duethmann, D., Blöschl, G., Parajka, J., 2020. Why does a conceptual hydrological model fail to correctly predict discharge changes in response to climate change? Hydrology and Earth System Sciences 24, 3493–3511. https://doi.org/10.5194/hess-24-3493-2020

Environment and Climate Change Canada, 2020. Historical climate data [WWW Document]. URL https://climate.weather.gc.ca/historical_data/search_historic_data_e.html (accessed 5.26.21).

Environment and Climate Change Canada, 2018. Water Survey of Canada: Historical hydrometric data [WWW Document]. URL https://wateroffice.ec.gc.ca (accessed 5.26.21).

Gibson, J.J., Birks, S.J., Moncur, M., 2019a. Mapping water yield distribution across the South Athabasca Oil Sands (SAOS) area: Baseline surveys applying isotope mass balance of lakes. Journal of Hydrology: Regional Studies 21. https://doi.org/10.1016/j.ejrh.2018.11.001

Gibson, J.J., Yi, Y., Birks, S.J., 2019b. Isotopic tracing of hydrologic drivers including permafrost thaw status for lakes across Northeastern Alberta, Canada: A 16-year, 50-lake assessment. Journal of Hydrology: Regional Studies 26. https://doi.org/10.1016/j.ejrh.2019.100643

Gibson, J.J., Yi, Y., Birks, S.J., 2016. Isotope-based partitioning of streamflow in the oil sands region, northern Alberta: Towards a monitoring strategy for assessing flow sources and water quality controls. Journal of Hydrology: Regional Studies 5, 131–148. https://doi.org/10.1016/j.ejrh.2015.12.062

Government of Canada, 2021. Canada-Alberta oil sands environmental monitoring [WWW Document]. URL https://www.canada.ca/en/environment-climate-change/services/oil-sands-monitoring.html (accessed 5.30.21).

Gupta, H. v., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of Hydrology 377, 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Haghnegahdar, A., Razavi, S., Yassin, F., Wheater, H., 2017. Multicriteria sensitivity analysis as a diagnostic tool for understanding model behaviour and characterizing model uncertainty. Hydrological Processes 31, 4462–4476. https://doi.org/10.1002/hyp.11358

He, Z., Unger-Shayesteh, K., Vorogushyn, S., Weise, S.M., Kalashnikova, O., Gafurov, A., Duethmann, D., Barandun, M., Merz, B., 2019. Constraining hydrological model

parameters using water isotopic compositions in a glacierized basin, Central Asia. Journal of Hydrology 571, 332–348. https://doi.org/10.1016/j.jhydrol.2019.01.048

Herman, J.D., Kollat, J.B., Reed, P.M., Wagener, T., 2013. From maps to movies: High-resolution time-varying sensitivity analysis for spatially distributed watershed models. Hydrology and Earth System Sciences 17, 5109–5125. https://doi.org/10.5194/hess-17-5109-2013

Höllering, S., Wienhöfer, J., Ihringer, J., Samaniego, L., Zehe, E., 2018. Regional analysis of parameter sensitivity for simulation of streamflow and hydrological fingerprints. Hydrology and Earth System Sciences 22, 203–220. https://doi.org/10.5194/hess-22-203-2018

Holmes, T., 2016. isoWATFLOOD Stable water isotope simulation in the WATFLOOD hydrologic model.

Holmes, T., Stadnyk, T.A., Kim, S.J., Asadzadeh, M., 2020. Regional Calibration With Isotope Tracers Using a Spatially Distributed Model: A Comparison of Methods. Water Resources Research 56. https://doi.org/10.1029/2020WR027447

Intsiful, A., Ambinakudige, S., 2021. Glacier cover change assessment of the Columbia Icefield in the Canadian rocky mountains, Canada (1985-2018). Geosciences (Switzerland) 11, 1–9. https://doi.org/10.3390/geosciences11010019

Kiang, J.E., Gazoorian, C., McMillan, H., Coxon, G., Le Coz, J., Westerberg, I.K., Belleville, A., Sevrez, D., Sikorska, A.E., Petersen-Øverleir, A., Reitan, T., Freer, J., Renard, B., Mansanarez, V., Mason, R., 2018. A Comparison of Methods for Streamflow Uncertainty Estimation. Water Resources Research 54, 7149–7176. https://doi.org/10.1029/2018WR022708

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. https://doi.org/10.1029/2005WR004362

Klaus, J., Chun, K.P., McGuire, K.J., McDonnell, J.J., 2015. Temporal dynamics of catchment transit times from stable isotope data. Water Resources Research 51, 4208–4223. https://doi.org/10.1002/2014WR016247

Klaus, J., McDonnell, J.J., 2013. Hydrograph separation using stable isotopes: Review and evaluation. Journal of Hydrology 505, 47–64. https://doi.org/10.1016/J.JHYDROL.2013.09.006

Knoben, W.J.M., Freer, J.E., Woods, R.A., 2019. Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores. Hydrology and Earth System Sciences 23, 4323–4331. https://doi.org/10.5194/hess-23-4323-2019

Kouwen, N., 2018. WATFLOOD/WATROUTE Hydrological Model Routing & Flood Foresting System [WWW Document]. URL www.watflood.ca

Kumar, R., Samaniego, L., Attinger, S., 2013. Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations. Water Resources Research 49, 360–379. https://doi.org/10.1029/2012WR012195

Lacroix, M.P., Martz, L.W., Kite, G.W., Garbrecht, J., 2002. Using digital terrain analysis modeling techniques for the parameterization of a hydrologic model. Environmental Modelling & Software 17, 125–134. https://doi.org/10.1016/S1364-8152(01)00042-1

Minder, J.R., Mote, P.W., Lundquist, J.D., 2010. Surface temperature lapse rates over complex terrain: Lessons from the Cascade Mountains. Journal of Geophysical Research 115, D14122. https://doi.org/10.1029/2009JD013493

Mizukami, N., Rakovec, O., Newman, A.J., Clark, M.P., Wood, A.W., Gupta, H. v., Kumar, R., 2019. On the choice of calibration metrics for "high-flow" estimation using hydrologic models. Hydrology and Earth System Sciences 23, 2601–2614. https://doi.org/10.5194/hess-23-2601-2019

Najafi, M.R., Moradkhani, H., Jung, I.W., 2011. Assessing the uncertainties of hydrologic model selection in climate change impact studies. Hydrological Processes 25, 2814–2826. https://doi.org/10.1002/hyp.8043

Nash, J.E., Sutcliffe, J. v, 1970. River flow forecasting through conceptual models part I—A discussion of principles. Journal of Hydrology 10, 282–290.

Nenzén, H.K., Price, D.T., Boulanger, Y., Taylor, A.R., Cyr, D., Campbell, E., 2020. Projected climate change effects on Alberta's boreal forests imply future challenges for oil sands reclamation. Restoration Ecology 28, 39–50.

Newman, A.J., Mizukami, N., Clark, M.P., Wood, A.W., Nijssen, B., Nearing, G., 2017. Benchmarking of a Physically Based Hydrologic Model. Journal of Hydrometeorology 18, 2215–2225. https://doi.org/10.1175/JHM-D-16-0284.1

Oreskes, N., Shrader-Frechette, K., Belitz, K., 1994. Verification , Validation , and Confirmation of Numerical Models in the Earth Sciences Naomi Oreskes ; Kristin Shrader-Frechette ; Kenneth Belitz. Science 263, 641–646.

Peters-Lidard, C.D., Clark, M., Samaniego, L., Verhoest, N.E.C., van Emmerik, T., Uijlenhoet, R., Achieng, K., Franz, T.E., Woods, R., 2017. Scaling, similarity, and the fourth paradigm for hydrology. Hydrology and Earth System Sciences 21, 3701–3713. https://doi.org/10.5194/hess-21-3701-2017

Pfannerstill, M., Guse, B., Reusser, D., Fohrer, N., 2015. Process verification of a hydrological model using a temporal parameter sensitivity analysis. Hydrology and Earth System Sciences 19, 4365–4376. https://doi.org/10.5194/hess-19-4365-2015

Razavi, S., Gupta, H. v., 2016a. A new framework for comprehensive, robust, and efficient global sensitivity analysis: 1. Theory. Water Resources Research 52, 423–439. https://doi.org/10.1002/2015WR017558

Razavi, S., Gupta, H. v., 2016b. A new framework for comprehensive, robust, and efficient global sensitivity analysis: 2. Application. Water Resources Research 52, 440–455. https://doi.org/10.1002/2015WR017559

Razavi, S., Gupta, H. v., 2015. What do we mean by sensitivity analysis? the need for comprehensive characterization of "global" sensitivity in Earth and Environmental systems models. Water Resources Research 51, 3070–3092. https://doi.org/10.1002/2014WR016527

Razavi, S., Sheikholeslami, R., Gupta, H. v, Haghnegahdar, A., 2019. VARS-TOOL: A toolbox for comprehensive, efficient, and robust sensitivity and uncertainty analysis. Environmental Modelling & Software 112, 95–107. https://doi.org/https://doi.org/10.1016/j.envsoft.2018.10.005

Rodriguez, N.B., Klaus, J., 2019. Catchment Travel Times From Composite StorAge Selection Functions Representing the Superposition of Streamflow Generation Processes. Water Resources Research 55, 9292–9314. https://doi.org/10.1029/2019WR024973

Rosa, L., Davis, K.F., Rulli, M.C., D'Odorico, P., 2017. Environmental consequences of oil production from oil sands. Earth's Future 5, 158–170. https://doi.org/https://doi.org/10.1002/2016EF000484

Sahraei, S., Asadzadeh, M., Unduche, F., 2020. Signature-based multi-modelling and multi-objective calibration of hydrologic models: Application in flood forecasting for Canadian Prairies. Journal of Hydrology 588, 125095. https://doi.org/10.1016/J.JHYDROL.2020.125095

Schoups, G., van de Giesen, N.C., Savenije, H.H.G., 2008. Model complexity control for hydrologic prediction. Water Resources Research 44. https://doi.org/10.1029/2008WR006836

Shafii, M., Tolson, B.A., 2015. Optimizing hydrological consistency by incorporating hydrological signatures into model calibration objectives. Water Resources Research 51, 3796–3814. https://doi.org/10.1002/2014WR016520

Shangguan, W., Dai, Y., Duan, Q., Liu, B., Yuan, H., 2014. A global soil data set for earth system modeling. Journal of Advances in Modeling Earth Systems 6. https://doi.org/10.1002/2013MS000293

Song, X., Zhang, J., Zhan, C., Xuan, Y., Ye, M., Xu, C., 2015. Global sensitivity analysis in hydrological modeling: Review of concepts, methods, theoretical framework, and applications. Journal of Hydrology. https://doi.org/10.1016/j.jhydrol.2015.02.013

Spangenberg, J.E., 2012. Caution on the storage of waters and aqueous solutions in plastic containers for hydrogen and oxygen stable isotope analysis. Rapid Communications in Mass Spectrometry 26. https://doi.org/10.1002/rcm.6386

Stadnyk, T.A., Delavau, C., Kouwen, N., Edwards, T.W.D., 2013. Towards hydrological model calibration and validation: Simulation of stable water isotopes using the isoWATFLOOD model. Hydrological Processes 27, 3791–3810. https://doi.org/10.1002/hyp.9695

Stadnyk, T.A., Holmes, T.L., 2020. On the value of isotope-enabled hydrological model calibration. Hydrological Sciences Journal 65, 1525–1538. https://doi.org/10.1080/02626667.2020.1751847

Stahl, K., Moore, R.D., Shea, J.M., Hutchinson, D., Cannon, A.J., 2008. Coupled modelling of glacier and streamflow response to future climate scenarios. Water Resources Research 44. https://doi.org/10.1029/2007WR005956

Tunaley, C., Tetzlaff, D., Birkel, C., Soulsby, C., 2017. Using high-resolution isotope data and alternative calibration strategies for a tracer-aided runoff model in a nested catchment. Hydrological Processes 31, 3962–3978. https://doi.org/10.1002/hyp.11313

van Huijgevoort, M.H.J., Tetzlaff, D., Sutanudjaja, E.H., Soulsby, C., 2016. Using high resolution tracer data to constrain water storage, flux and age estimates in a spatially distributed rainfall-runoff model. Hydrological Processes 30, 4761–4778. https://doi.org/10.1002/hyp.10902

Viglione, A., Parajka, J., Rogger, M., Salinas, J.L., Laaha, G., Sivapalan, M., Blöschl, G., 2013. Comparative assessment of predictions in ungauged basins - Part 3: Runoff signatures in Austria. Hydrology and Earth System Sciences 17, 2263–2279. https://doi.org/10.5194/hess-17-2263-2013

Vitt, D.H., Halsey, L.A., Zoltai, S.C., 2000. The changing landscape of Canada's western boreal forest: the current dynamics of permafrost. Canadian Journal of Forest Research 30, 283–287.

Wagener, T., McIntyre, N., Lees, M.J., Wheater, H.S., Gupta, H. V., 2003. Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis. Hydrological Processes 17, 455–476. https://doi.org/10.1002/HYP.1135

Wan, W., Zhao, J., Popat, E., Herbert, C., Döll, P., 2021. Analyzing the Impact of Streamflow Drought on Hydroelectricity Production: A Global-Scale Study. Water Resources Research 57, e2020WR028087. https://doi.org/10.1029/2020WR028087

Westerberg, I.K., Sikorska-Senoner, A.E., Viviroli, D., Vis, M., Seibert, J., 2020. Hydrological model calibration with uncertain discharge data. Hydrological Sciences Journal 1–16. https://doi.org/10.1080/02626667.2020.1735638

Wood, E.F., 1976. An analysis of the effects of parameter uncertainty in deterministic hydrologic models. Water Resources Research 12, 925–932. https://doi.org/10.1029/WR012i005p00925

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, 29–47. https://doi.org/10.1016/0022-1694(88)90090-X

Yamanaka, T., Ma, W., 2017. Runoff prediction in a poorly gauged basin using isotope-calibrated models. Journal of Hydrology 544, 567–574. https://doi.org/https://doi.org/10.1016/j.jhydrol.2016.12.005

Yilmaz, K.K., Gupta, H. v., Wagener, T., 2008. A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. Water Resources Research 44. https://doi.org/10.1029/2007WR006716

5. Guidance on large scale hydrologic model calibration with isotope tracers

Tegan L. Holmes[1, 2], Tricia A. Stadnyk[2, 1], Masoud Asadzadeh[1], John J. Gibson[3, 4]
[1] University of Manitoba, Civil Engineering, Winnipeg MB R3T 5V6
[2] University of Calgary, Geography, Calgary AB T2N 1N4 Canada
[3] InnoTech Alberta, 3-4476 Markham Street, Victoria BC V8Z 7X8 Canada
[4] University of Victoria, Geography, Victoria BC V8W 3R4 Canada

Manuscript in preparation for submission in Journal of Hydrology.

This chapter compares and evaluates calibration and identifiability results for a variety of isotope tracer simulation performance metrics. Process fluxes and streamflow simulations for isotope-aided calibrations are assessed. Isotope-enabled calibrations are found to have better process identifiability and multi-objective optimizations using an isotope simulation performance metric which includes timing error as a secondary calibration objective are recommended for large-scale model calibrations.

5.1. Abstract

Standard hydrologic model evaluation and calibration approaches focus on the accurate simulation of streamflow, disregarding internal process simulations. Stable isotope tracers can provide additional information on water sources, flow paths and processes, which can be used to inform model calibration. This study assesses the added value of isotope data in comparison to current best-practice flow-only calibration and evaluates of the merits and limitations of isotope simulation performance metrics for the purposes of model calibration. An isotope-enabled process-based hydrologic model of a large, mid- to high-latitude watershed with a multi-year isotope sampling program was assessed and calibrated using global sensitivity analyses, Monte Carlo simulations and multi-objective optimizations. Isotope tracer data improves process and streamflow component identifiability and produces some minor improvements in individual parameter value identifiability. Calibrating to optimize both flow and isotope simulation performance produced better flow simulation ensembles, with improved observation capture and

validation performance, than calibrating to optimize flow simulations alone. Using an isotope simulation performance metric which includes timing error as a secondary calibration objective leads to more robust streamflow modeling, even in mesoscale watersheds, and with limited isotope observation datasets.

5.2. Introduction

Hydrologic models are an essential tool for hydrologists, used to predict runoff or streamflow in both the short-term (forecasting) and long-term (climate change projection). The purpose or application of hydrologic models vary, from overall basin water balance estimates, to predicting flood volumes, or the investigation of flow generation processes, such as baseflow estimation. Likewise, hydrologic models vary in structure and complexity, from fully physically-based to conceptual models, and in scale, from hillslope to global simulation of water fluxes. Accurately modeling streamflow with any type of model is challenging, and even more so in remote regions due to low data availability: streamflow and weather gauges are both rare and have limited record lengths (Coulibaly et al., 2013). These data limitations increase both the need for, and the uncertainty in, hydrologic modeling in mid- to high latitude regions that have short or sparse data records. To reliably predict streamflow under climatic conditions different than those in limited recorded data or estimate flows in ungauged locations, hydrologic models must accurately represent the basic physical processes which generate streamflow (Duethmann et al., 2020).

A process-based hydrologic model is intended to simulate river flows by simulating the hydrologic processes in the watershed. The underlying design aspiration is that the combination of sub-models (for evaporation, groundwater and so forth) will be accurate in and of themselves, and their summation will be an accurate streamflow timeseries (with some inescapable uncertainty). This complete watershed model should then be able to accurately simulate streamflow for any weather input the modeler cares to throw at it, and the only non-stationarity concern is changes to physical properties of the basin (such as changes to vegetation) (Fatichi et al., 2016). This aspiration has yet to be actualized, but an approximation thereof might be achievable.

One of the major hurdles facing large scale process-based models is the limited quantity of information available on water movement in watersheds (Fatichi et al., 2016; Kirchner, 2006; Stevenson et al., 2021). Only in the most intensively monitored research sites is the daily

transpiration loss or wetland flux rate measured; streamflow records, with potentially some intermittent water tracer or point process observations, are all that can be hoped for in modeling at the meso-scale (Gibson et al., 2020). This lack of data poses a major challenge for the reliability of process-based models: how can the modeler be assured of the accuracy of the sub-models when the only available observations are the final, cumulative, streamflow? Can any confidence be placed in the model when innumerable combinations of hydrologic process simulations add up to the same total streamflow, and only the total streamflow is verified?

Part of the solution is ensemble-based modeling, or multiple parameter sets within a single model framework used to generate envelopes of simulated streamflow (Matott et al., 2009; Moges et al., 2020; Uusitalo et al., 2015). Another component is increasing the information available to assess model accuracy, both at the internal process level (i.e., hydrologic compartments) and summative (i.e., streamflow) (Kirchner, 2006). Previous research has shown stable isotope tracer data (i.e. ratios of water molecules containing ^{18}O or ^{2}H to standard water) can provide additional information for the evaluation of hydrologic models (He et al., 2019; Holmes et al., 2020; Stadnyk and Holmes, 2020; Tunaley et al., 2017). Stable isotope tracers are particularly useful in remote or inaccessible watersheds as they are non-reactive and naturally occurring, while their variable concentration in precipitation and evaporating water bodies can provide additional information on water sources and hydrological processes (Brooks et al., 2018; Oshun et al., 2016; Peralta-Tapia et al., 2015).

Observational data, whether flow or tracer observations, can be fed into the model development by calibrating parameter values. Hydrologic models depend on parameters, which are used to quantify and adjust the relationships between modeled storages and fluxes. In process-based models, which aim to emulate real-world hydrologic storages and mass fluxes, some parameter values can be estimated using field measurements or remotely sensed data; however, there are generally other parameters (with limited physical basis) for which good or reasonable values are unmeasurable or unknown (Acero Triana et al., 2019; Fatichi et al., 2016). These unknown parameter values necessitate model calibration, where parameter values are selected or adjusted to achieve an acceptable model performance. There are two approaches commonly used in the literature: trying vast numbers of parameter value combinations (generated either at random or using a sampled distribution) and using those with the best model performance or using a search

algorithm to locate optimal parameter value combinations (Beven and Binley, 1992; Efstratiadis and Koutsoyiannis, 2010; Pechlivanidis et al., 2011). There are benefits to both approaches: generating large numbers of independent solutions allows for much more complex statistical analyses and can ensure coverage of the entire parameter space, while search algorithms are much better at identifying good quality solutions, at a lower computational cost (Acero Triana et al., 2019; Pechlivanidis et al., 2011).

Regardless of the approach chosen, the calibration methodology will depend on quantifying the quality of the simulated model output using some sort of performance metric or metrics (Bennett et al., 2013). The quantified performance can be used to identify the best solutions, set an acceptable model performance threshold, or be used by a search algorithm as an objective. Performance metrics are not limited to simulated streamflow; if the hydrologic model is capable of simulating both flow and isotope tracer compositions, then both data types can be used quantify the model performance quality (He et al., 2019; Nan et al., 2021; Stadnyk and Holmes, 2020; Tunaley et al., 2017). To date, there are no universal guidelines on the best performance metrics to calibrate either flow or isotope tracer simulations, and metric selection has largely been ad-hoc or 'best guess' in the tracer-aided calibration literature (Holmes et al., 2020). Tracer and streamflow data typically differ in temporal resolution or consistency of sampling, which can influence metric selection (Bennett et al., 2013; Mizukami et al., 2019). Both the Kling-Gupta and Nash-Sutcliffe efficiency metrics are frequently used as flow simulation performance metrics in the current literature, and flow signature metrics are often recommended to evaluate specific flow simulation characteristics or identify hydrologically consistent parameter sets (Shafii and Tolson, 2015; Knoben et al., 2019; Sahraei et al., 2020) . Due to irregular or infrequent sampling, isotope tracer simulations are often evaluated using some variant of a residual error metric, but other metrics, such as the Kling-Gupta efficiency have also been applied (He et al., 2019; Stadnyk and Holmes, 2020; Tunaley et al., 2017). These different metrics vary in their sensitivity to different properties of the data series (e.g. the Nash-Sutcliffe efficiency is highly sensitive to the timing of peak flows, but is not very responsive to bias errors in low-flow periods). By using a mixture of multiple metrics (either to evaluate different aspects of the flow data, or to evaluate both flow and tracer data), model evaluations or calibrations can be rendered more reliable or comprehensive.

The desired outcome of calibrating a process-based hydrologic model is both a precise and accurate simulation of streamflow, resulting from a precise and accurate representation of the internal hydrologic processes. This perfect calibration result is out of reach for models of real-world watersheds, but some approximation of each of the elements remains desirable. An accurate simulation of streamflow matches all known flow observations, considering both the uncertainty in both the simulated and observed values. A precise simulation of streamflow has low uncertainty and a narrow range of simulated values at any given time; increasing the simulation precision generally decreases the accuracy, as wide uncertainty bounds are more likely to include observed values. For the internal process representation, much of the accuracy, or lack thereof, is determined by the model structure or conceptualization, but it is also partially determined by the parameter values resulting from the calibration process (e.g., the simulated infiltration accuracy depends on the equation used to estimate it, the scale of the area it is applied to, and the value assigned to the conductivity parameter). The precision of the process simulations depends only on the identifiability of the parameter values (provided the model structure and equations are static). If calibration results in a narrow range of values for a parameter, that parameter is well-identified, and if all parameters influencing a modeled process are well-identified, the range of contributions from the process will be small. However, identifiable parameters are not necessarily accurate, or representative of the real-world system the model is intended to emulate; due to issues such as model structure errors or forcing data uncertainty, parameters which consistently produce optimal performance metric results may also be consistently wrong (e.g., over-estimating evaporation to compensate for biased precipitation inputs or not including sublimation in the model). These four competing objectives are not technically mutually exclusive, but improvements in one often come at the expense of another (e.g., improving flow simulation accuracy generally reduces both simulation precision and parameter identifiability) (Xiong et al., 2009).

Adding in tracer data to the calibration process can improve the accuracy of the internal process representation, particularly for soil water storages, and can also improve the identifiability of some parameters (He et al., 2019; Holmes et al., 2020). However, parameter identifiability can also decrease when calibrating with isotope tracer data, when parameter values producing optimal flow simulation results are contradicted by the tracer simulation (Holmes et al., 2020). Effects on flow simulation accuracy and precision are likewise unclear; including isotope data in

the calibration in previous studies has decreased the flow simulation performance in the calibration period, and the flow simulation uncertainty has tended to increase. On the other hand, flow simulation performance in validation may improve when calibrating with tracer data. The value added by isotope tracer data to calibration remains difficult to assess, as previous studies have either used only mixed flow and isotope data, used different calibration methods for flow-only and mixed flow and isotope calibration (limiting the comparability) or used a single isotope tracer performance metric with a fixed weight in calibration (limiting the generalizability).

This study aims to address two remaining gaps in the literature on isotope-aided modeling at the large-scale: the added value of isotope data relative to current best-practice flow-only calibrations, and a comparison of the merits and limitations of isotope simulation performance metrics for the purposes of model calibration. In particular:

- Does isotope tracer data alter parameter identifiability, and if so, is this change an improvement?
- Does isotope tracer data lead to different flow simulation results, and if so, is this change an improvement?
- Can a tracer performance metric be recommended to maximize the benefits for future calibrations?

Our study is conducted in the Oil Sands region of Alberta, within the Athabasca River basin. This region was chosen because there exists one of the longest isotope in river sampling programs in Canada, several hydrometric gauges for calibration of subbasins and mainstem, and because it is high-latitude and representative of Canada's cold regions seasonal hydrology. In addition, there is significant interest in better understanding groundwater or baseflow contributions to streamflow, which requires accurate process-based information from hydrological models.

5.3. Methods

5.3.1. Hydrologic Model and Study Area

5.3.1.1. The CHARM/isoWATFLOOD Model

Assessing parameter identifiability and flow simulation changes from including isotope tracers in calibration requires an isotope-enabled hydrologic; this study uses the CHARM/WATFLOOD

hydrologic model and isoWATFLOOD, its associated dual-isotope tracer simulation model. CHARM and isoWATFLOOD are open source, distributed models which use a mixture of physically based and conceptual process representations for relatively computational efficient modeling of meso-scale watersheds (Kouwen, 2018). The isoWATFLOOD tracer models simulate the isotopic concentrations of oxygen-18 and deuterium in all of the storages and fluxes used in the CHARM hydrologic model. The isotope tracer model assumes that individual hydrologic storages are completely mixed through depth, and fluxes have the same concentration as the source storage, except evaporative fluxes, which are subject to isotopic fractionation (Stadnyk and Holmes, 2020). The linked hydrologic and tracer simulations both run on an hourly time-step and output daily simulation results.

The CHARM/WATFLOOD model structure divides the watershed area into grid cells with defined drainage directions, and then divides each cell into grouped response units or GRU (with the area of each GRU generally determined by landcover data), as shown in Figure 5-1. The default structure of a GRU (used for most types of landcover, such as forests, grass or shrubland) has a vertical soil column into two soil layers, the upper and lower soil zones, both of which contribute flows to the channel network. Rain or snowmelt ponded on the surface can either infiltrate into the upper soil zone or runoff directly to the channel network. Water in the upper soil zone may recharge the lower soil zone, evapotranspire, or be added to the channel network. Water in the lower soil zone may only flow out to the channel network; the model does not include fluxes between lower zone storages in adjacent grid cells, so regional groundwater fluxes are not 128ncludeed in the simulation. All types of GRU have separate snowpack storages which accumulate and melt independently and melting snowpacks cover only fractional areas of the GRU as the melt progresses. Both the snow and glacier melt rates are calculated as a function of air temperature. The upper soil zone under a snowpack is assumed to be frozen, and all three soil fluxes (i.e., recharge, infiltration and interflow) have substantially reduced rates in frozen soils; permafrost is not modeled. There are four special GRU classes without soil layers: glacier, impervious, connected wetland and water. In the glacier and impervious classes all rain and snowmelt becomes direct runoff, and glacier GRU also generate glacier melt flows. Connected wetlands, where present, accumulate all lateral outflows from GRU with soil layers, and have a bi-directional connection with stream channels, the flow direction being determined by the relative water levels in the wetland and stream. The water GRU is the end point for all flows

generated within the grid cell, and it is also the only GRU connected to other cells, with inflows from upstream cells and an outflow downstream. More detailed descriptions of process representation and equations can be found in Holmes (2016).

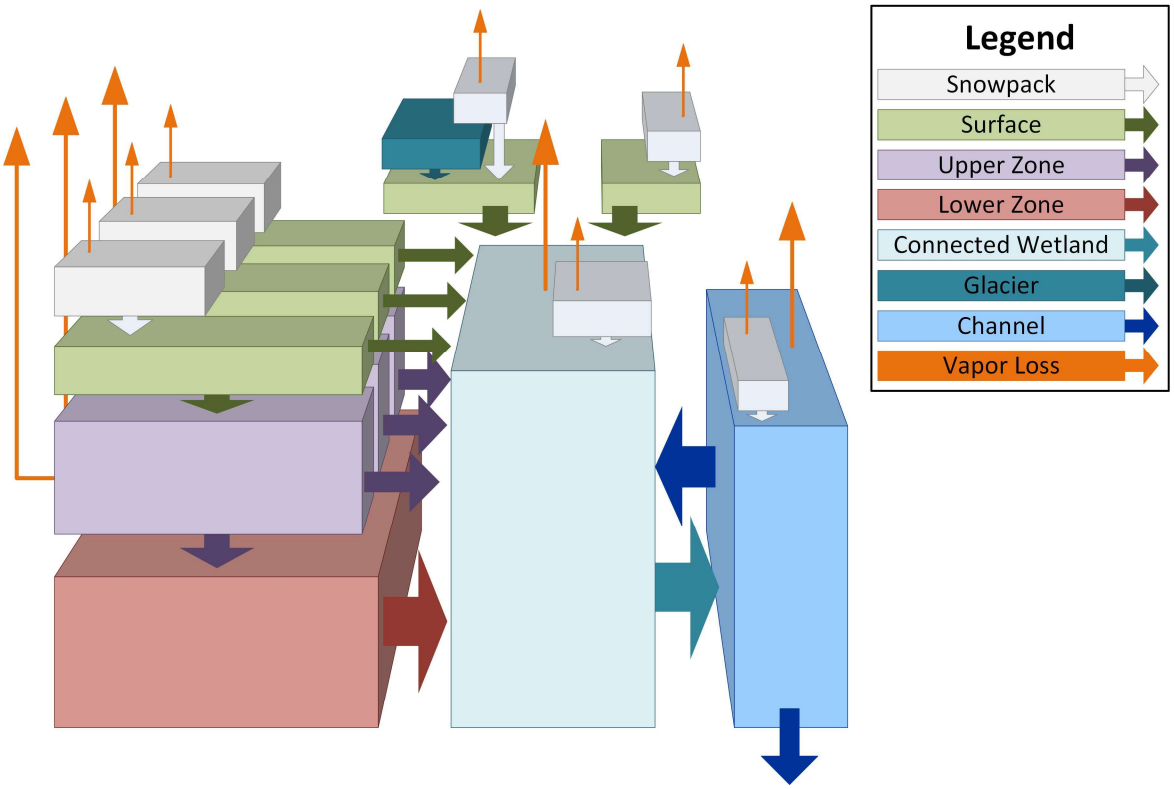

Figure 5-1: Schematic of grouped response units (GRU) representing water storage and fluxes between storages within a single headwater grid cell, with glaciers, connected wetlands and three soil-based GRU.

All processes in the model have parameters controlling the simulated fluxes, and the values of these parameters can either be consistent across all GRU, or separate GRU can have individual parameter values specific to that class. There are a very large number of parameters (over 300 for the Athabasca basin model) which can be altered in set up and calibration of a CHARM/isoWATFLOOD model, but the vast majority have minimal impact on the simulation (e.g., overland flow roughness factors), or can be estimated from the literature (e.g., canopy interception caps). This study will focus on the calibration of the highest priority parameters identified by previous studies, including Holmes et al. (2020), and model developer

recommendations, while still including a minimum of one parameter per simulated process in the calibration, listed in Table 5-1.

Table 5-1: List of potential significant parameters included in the calibration, including the parameter names, the process affected by the parameter and the affected GRU. Parameters with separate values for different GRU classes are italicized.

Parameter		Internal name	Process	Applicable GRU
Surface soil conductivity	k F (surf)	ak	Infiltration	All soil-based
Horizontal upper soil zone conductivity	k F (horz)	rec	Interflow	All soil-based
Open water PET to AET factor	PET F	fpet	Evaporation	Water, connected wetland
Snowmelt rate factor	*melt rate*	*fm*	*Snowmelt*	*All*
Upper soil zone soil water retention cap	*soil ret*	*retn*	*Soil storage and ET*	*All soil-based*
Vertical upper soil zone conductivity	k F (vert)	ak2	Recharge	All soil-based
Baseflow equation constant	*C*	*flz*	*Baseflow*	*All soil-based*
Baseflow equation power	pwr	pwr	Baseflow	All soil-based
Channel roughness factor	*n*	*r2n*	*Channel velocity*	*Water*
Wetland porosity	θ (wet)	theta	Wetland storage	Connected wetland
Wetland conductivity	k (wet)	kcond	Wetland velocity	Connected wetland
Glacier melt factor	glac F	gladjust	Glacier melt	Glacier

5.3.1.2. Athabasca Watershed

The Athabasca River watershed is the most southerly part of the Mackenzie River basin and is located in the north of the Canadian provinces of Alberta and Saskatchewan, on Treaty 6 and 8 land (Figure 5-2). The Athabasca River runs north-east from the Rocky Mountains to the Peace-

Athabasca Delta and Lake Athabasca, with a total watershed area of 156,000 km^2. The more steeply sloped upper reaches of the Athabasca lie in alpine or foothills regions, and are covered predominately by coniferous forests, with some barren ground and glaciers, most notably the Athabasca Glacier in the Columbia Icefield (Intsiful and Ambinakudige, 2021). Further downstream, there is substantial agricultural activity, although both forests and wetlands remain common; the largest tributaries in this region are the Pembina, McLeod and Lesser Slave Rivers. The lower reaches of the Athabasca River basin have abundant wetlands and coniferous forest, and subdued relief; the Clearwater, Firebag and MacKay Rivers are the largest tributaries in the downstream portion of the Athabasca basin, which coincides with the Athabasca Oil Sands region. The Athabasca basin is predominately underlain by Cretaceous and Paleogene sedimentary stone, while soils are largely loam, with higher clay prevalence in the middle of the basin and more some sandy or coarse soil in the downstream region (Shangguan et al., 2014)(Alberta Geological Survey, 2013). Deeper bedrock or regional groundwater flows contribute a small amount to the stream flow of the Athabasca River and its tributaries (3-5% of annual flow); there also remains some sporadic permafrost which is actively degrading (Gibson et al., 2016; Vitt et al., 2000). The flow of the Athabasca River is not regulated, either for flood control or hydroelectric generation, but small amounts of water are diverted for agricultural use, and approximately 1% of the annual flow in the Athabasca near Fort McMurray is used for activities in the oil sands (Rosa et al., 2017).

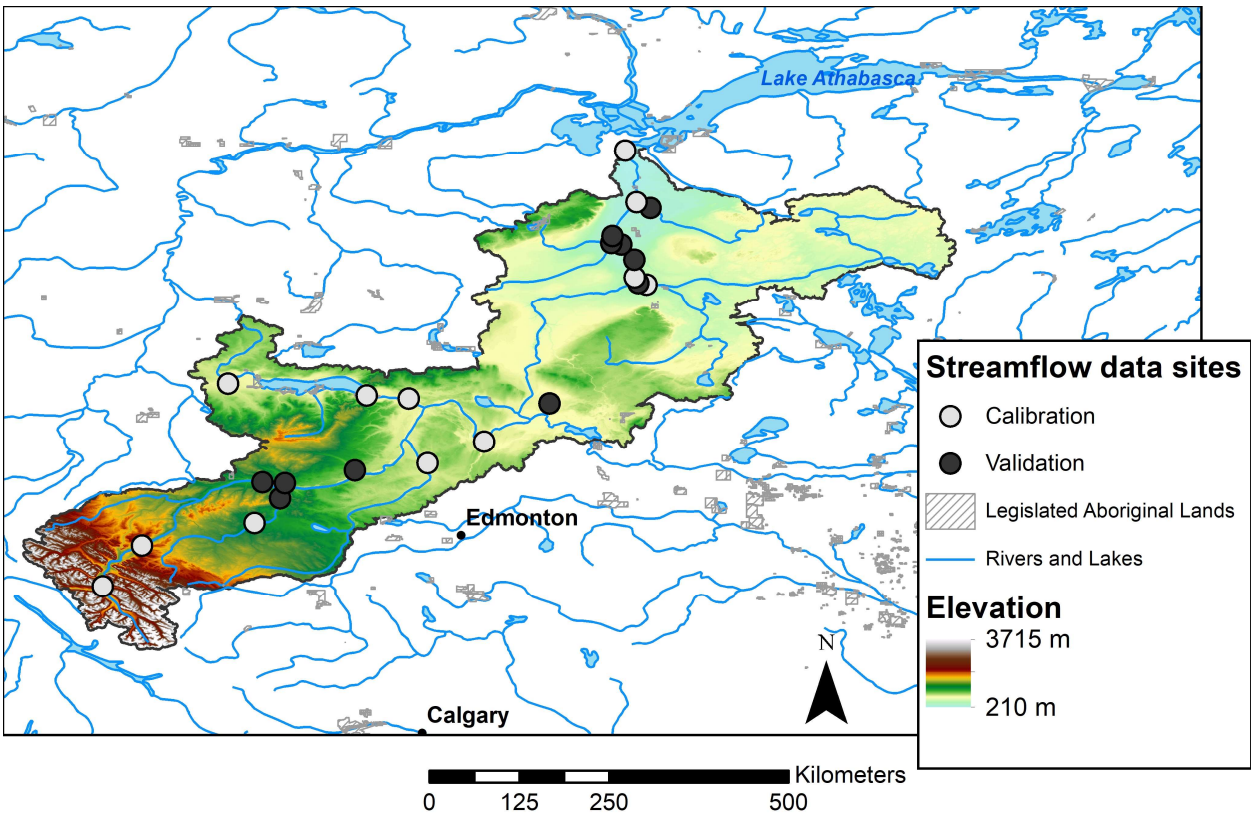

Figure 5-2: Map of the Athabasca River watershed, with calibration and validation sites shown.

The climate of the Athabasca River watershed is highly seasonal, with the highest temperature and precipitation in July, and the lowest in January or February (basin-wide mean monthly temperatures vary from -19 °C to +17 °C). Temperatures are generally below freezing for 5 to 6 months of the year, but approximately 60% of the 450 mm average annual precipitation falls as rain (precipitation rates are significantly higher in the summer months); the high elevation headwaters are colder than the basin average and a larger fraction of precipitation falls as snow (Environment and Climate Change Canada, 2020).

The CHARM model of Athabasca watershed used in this study divides the basin area into 320 grid cells; each cell has a nominal size of 0.4° longitude by 0.2° latitude, but actual cell sizes are adjusted based on delineated drainage areas. In turn, cells are subdivided into 10 different grouped response unit types, listed in Table 5-2, with GRU areas determined based on land cover data from the ESA (European Space Agency, 2017).

Table 5-2: Grouped response units for the Athabasca watershed model with landcover types and prevalence.

GRU Name	GRU type	Prevalence (%)	Landcover types
Grass	soil	8.1	herbaceous, cropland, grassland
Disconnected wetland	soil	8.8	wetlands: all vegetation heights
Connected wetland	connected wetland	2.2	wetlands: all vegetation heights
Mixed forest	soil	15.1	mixed and deciduous forest
Coniferous forest	soil	54.1	coniferous needle leafed forest
Shrub	soil	6.2	shrubland and sparse trees
Impervious	impervious	0.03	urban, consolidated bare ground
Barren	soil	1.4	sparse short vegetation, bare ground
Water	water	3.8	open water
Glacier	glacier	0.2	permanent ice and snow

Most GRU types in CHARM are modeled with two soil layers; in this study, the coniferous and mixed forest, disconnected wetland, grass, shrub, and barren classes are modeled with soil layers. For the glacier and impervious classes all rain and snowmelt becomes direct runoff, and glacier GRU also generate glacier melt flows. Rain or snowmelt in open water or connected wetlands GRU is simply added directly to the wetland, channel or lake modeled network.

5.3.1.3. Forcing and Evaluation Data

Meteorological Data

The coupled isotope-hydrologic models were run using four meteorological forcings: hourly air temperature and humidity, daily total precipitation and monthly average isotopic compositions of precipitaiton. The precipitation, temperature and humidity forcings were based on observations at Environment and Climate Change Canada (ECCC) weather gauges in the watershed and surounding area (Environment and Climate Change Canada, 2020). Forcing data for individual grid cells were estimated at each time step using an inverse distance squared weighting interpolation and a temperature lapse rate of -5 °C/km and a precipitation lapse rate of 0.2 mm/km; overall, 56 weather gauges were included in the calculations, provided there was observation data at the gauge for that time interval (Minder et al., 2010)(Kouwen, 2018). Due to the scarcity of observations of isotopes in precipitation, the forced isotopic compositions of

precipitation were estimated from the empirical model developed by Delavau et al. (2015). This model is a geospatial interpolation which extends historical isotope observations spatially and temporally using a multiple linear regression of geographic and climatic indicators; the model adequately captured the seasonality and annual variability of isotopes in precipitation in the region (Delavau et al., 2015).

Flow and Isotope Data

Simulated model outputs were compared to historical hydrometric data from the Water Survey of Canada (Environment and Climate Change Canada, 2018). From of a total of 20 continuous or seasonal (i.e. continuous only during the open water season) flow gauges with daily data between 2002 and 2015, 10 were used to calibrate the model, and the remaining 10 were used for validation, as listed in Table 3 (see Figure 5-2 for spatial distribution). Gauged areas for the Athabasca and its tributaries range between 1000 and 137,500 km^2. The streamflow data have an uncertainty of approximately ±10%, with higher uncertainty during peak flow and ice-on periods (Kiang et al., 2018; Westerberg et al., 2020).

Table 5-3: Hydrometric gauges and isotope sampling sites in the Athabasca River basin.

		Latitude (°)	Longitude (°)	Isotope Samples	Drainage Area (km^2)	Operation Schedule	Data Application
07AA002	ATHABASCA RIVER NEAR JASPER	52.91	-118.06		3870	Continuous	Calibration
07AD002	ATHABASCA RIVER AT HINTON	53.42	-117.57	159	9760	Continuous	Calibration
07AE001	ATHABASCA RIVER NEAR WINDFALL	54.21	-116.06		19600	Seasonal	Validation
07AG004	MCLEOD RIVER NEAR WHITECOURT	53.99	-115.84		9110	Seasonal	Validation
07AG007	MCLEOD RIVER NEAR ROSEVEAR	53.70	-116.16		7140	Continuous	Calibration
07AH001	FREEMAN RIVER NEAR FORT ASSINIBOINE	54.41	-114.96		1660	Seasonal	Validation
07AH003	SAKWATAMAU RIVER NEAR WHITECOURT	54.20	-115.78		1150	Seasonal	Validation
07BC002	PEMBINA RIVER AT JARVIE	54.45	-113.99		13100	Continuous	Calibration
07BE001	ATHABASCA RIVER AT ATHABASCA	54.72	-113.29	146	74600	Continuous	Calibration
07BF002	WEST PRAIRIE RIVER NEAR HIGH PRAIRIE	55.45	-116.49		1150	Continuous	Calibration
07BK001	LESSER SLAVE RIVER AT SLAVE LAKE	55.31	-114.76	17	13600	Continuous	Calibration
07BK007	DRIFTWOOD RIVER NEAR THE MOUTH	55.26	-114.23		2100	Continuous	Calibration
07CA006	WANDERING RIVER NEAR WANDERING RIVER	55.17	-112.39		1120	Seasonal	Validation
07CD001	CLEARWATER RIVER AT DRAPER	56.68	-111.20	44	30800	Continuous	Calibration
07CD004	HANGINGSTONE RIVER AT FORT MCMURRAY	56.60	-111.41		960	Seasonal	Validation
07DA001	ATHABASCA RIVER BELOW MCMURRAY	56.78	-111.40	126	133000	Continuous	Calibration
07DA006	STEEPBANK RIVER NEAR FORT MCMURRAY	56.89	-111.20	37	1320	Seasonal	Validation
07DA008	MUSKEG RIVER NEAR FORT MACKAY	57.21	-111.55	70	1460	Seasonal	Validation
07DB001	MACKAY RIVER NEAR FORT MACKAY	57.12	-112.01	26	5570	Seasonal	Validation
07DC001	FIREBAG RIVER NEAR THE MOUTH	57.65	-111.20	44	6390	Seasonal	Validation
07DD011	ATHABASCA RIVER AT OLD FORT	58.37	-111.52	120	156000	Sampling Only	Calibration
AB07DA0750	ELLS RIVER	57.30	-111.68	36	2500	Sampling Only	Validation
AB07DA0980	ATHABASCA RIVER UPSTREAM OF FIREBAG	57.72	-111.38	68	154400	Sampling Only	Calibration

In addition to the WSC flow gauge data, a water isotope sampling campaign was conducted for the Alberta Environmental Monitoring, Evaluation and Reporting Agency's Long-Term River Network monitoring program on the Athabasca River and some of its tributaries (Gibson et al., 2016). Sampling at hydrometric gauges in the Athabasca basin began in 2002 and continued through 2014, with an approximately monthly sampling frequency, although not all sampling series began in 2002, and some gauges have data gaps over one year. Water samples for isotope tracer analysis were sealed in 30 mL high-density polyethylene bottles and analyzed at either the University of Waterloo Environmental Isotope Laboratory or at Alberta Innovates Technology Futures, Victoria (Gibson et al., 2016). The high-density polyethylene bottles used have been shown to be effective at preventing isotopic fractionation, and all samples were also sealed and analyzed within 1 year of sample collection (Gibson et al., 2019; Spangenberg, 2012). The isotope tracer compositions of the water samples were analyzed using a Micromass IsoPrime Dual Inlet/Gas Chromatograph pre-2009, and from 2009 on, using a Thermo Scientific Delta V Advangage Dual Inlet/HDevice system; both machines have estimated analytical uncertainties of ±0.1‰ for oxygen-18 and ±1‰ for deuterium, and all results are reported relative to VSMOW (Gibson et al., 2016).

5.3.2. Calibration Methods

Assessing the value added by isotope data to calibration, in comparison to current best-practice flow-only calibrations requires the application of a multi-stage calibration process, completed both with and without isotope tracer data. The overall methodology is outlined in the flow chart in Figure 5-3, and details are covered in the following sections.

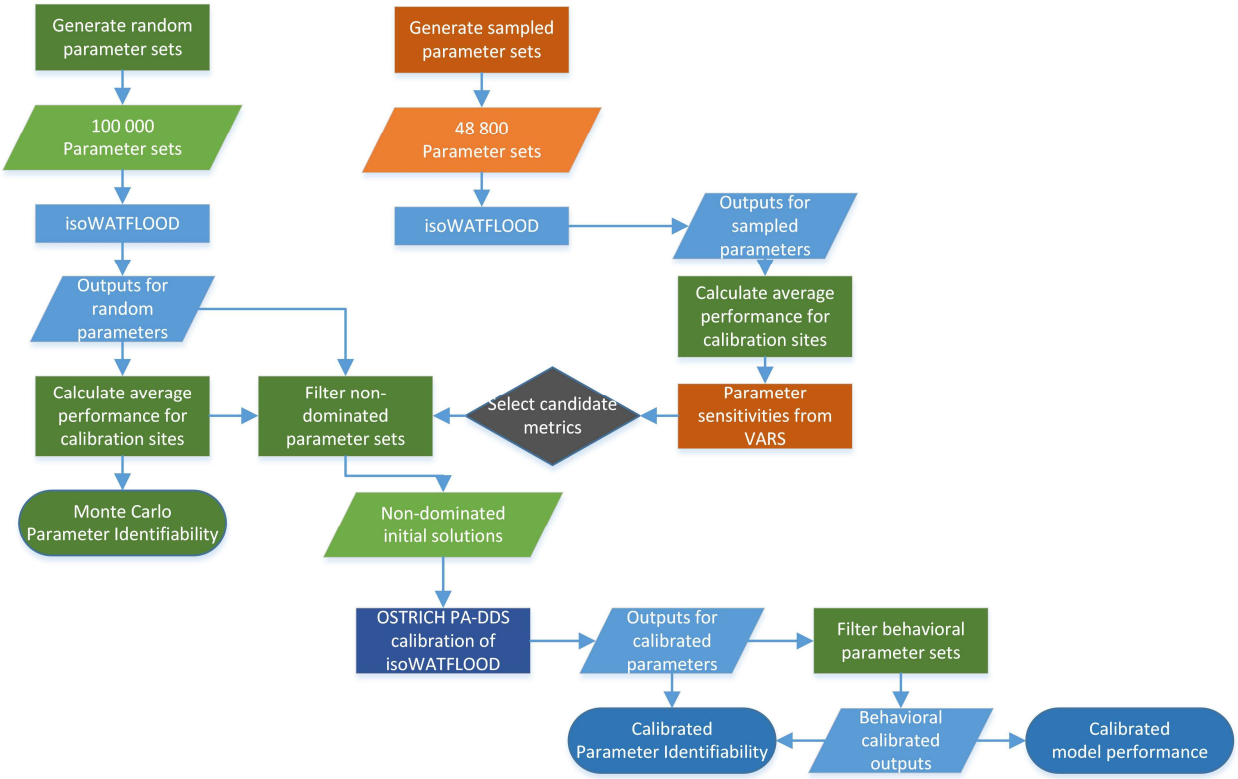

Figure 5-3: Flow chart of the study methodology, with isoWATFLOOD processes and outputs in blue, VARS in orange, and independent scripts in green.

Two separate assessments of parameter identifiability were performed. The first was a Monte Carlo style evaluation, based on a large number of randomly generated parameter sets, which has the advantages of independent sampling and limited user decisions. The second was a set of optimizations, using current recommended methods (assessing parameter sensitivities, selecting useful metrics, running multi-objective optimizations and selecting behavioral parameter sets), that produces higher quality solutions, but also depends on a series of user choices.

5.3.2.1. Model Performance Metrics

A variety of metrics were evaluated as potential calibration objectives, with the selection based on the most commonly applied metrics from the literature. These metrics vary in their responsiveness to different types of simulation error (such as timing or volume bias), and likewise vary in their parameter sensitivities (Holmes et al., 2022). In calculating all performance metrics, only simulated data on days that have flow or isotope observations are considered.

The normalized root mean square error metric, a simple residual error metric, was evaluated as a potential objective for both the flow and isotope simulations, calculated as:

$$NRMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_{s,i} - x_{o,i}\right)^2} / \overline{x}_o \quad (1)$$

Where n is number of observations, $x_{o,i}$ is observation i, $x_{s,i}$ is the corresponding simulated value and $\overline{x}_o$ is the observation mean. The Nash-Sutcliffe efficiency (NSE), another residual error metric traditionally used to evaluate hydrograph fits, and the log transform version of NSE were calculated exclusively for the flow simulation (Nash and Sutcliffe, 1970). Both NRMSE and NSE are highly sensitive to large residuals that often occur with mis-timed simulations in high-flow periods, while logNSE is relatively sensitive to small residuals that happen during low-flow periods.

$$NSE = 1 - \frac{\sum_{i=1}^{n}\left(x_{s,i} - x_{o,i}\right)^2}{\sum_{i=1}^{n}\left(x_{o,i} - \overline{x}_o\right)^2} \quad (2)$$

$$logNSE = 1 - \frac{\sum_{i=1}^{n}\left(log(x_{s,i}) - log(x_{o,i})\right)^2}{\sum_{i=1}^{n}\left(log(x_{o,i}) - \overline{lo\ \ (x_o)}\right)^2} \quad (3)$$

The Kling-Gupta efficiency (KGE) metric, and all three of its constituent components (the correlation r, the relative variability α and the bias β) were evaluated as possible objectives for both the isotope and flow simulations. KGE and NSE share the same components r, α, and β, but KGE gives each component equal weight unlike NSE, which discounts the variability (Gupta et al., 2009).

$$r = \frac{\sum_{i=1}^{n}\left(x_{o,i} - \overline{x}_o\right)\left(x_{s,i} - \overline{x}_s\right)}{\sqrt{\sum_{i=1}^{n}\left(x_{o,i} - \overline{x}_o\right)^2}\sqrt{\sum_{i=1}^{n}\left(x_{s,i} - \overline{x}_s\right)^2}} \quad (4)$$

$$\alpha = \frac{\sigma_s}{\sigma_o} \quad (5)$$

$$\beta = \frac{\overline{x}_s}{\overline{x}_o} \quad (6)$$

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \quad (7)$$

Where σ_s and σ_o are the standard deviations of the simulated and observed data.

Three popular flow signature metrics were also evaluated as potential calibration objectives for both the isotope and flow simulations. Equation 8 was used to calculate the error in the slope of the flow duration curve (SFDC) (for flows) and the slope of the duration curve (SDC) (for isotopes) (Viglione et al., 2013):

$$SDC = 100\left(\frac{x_{S,30} - x_{S,70}}{40\overline{x}_S} - \frac{x_{O,30} - x_{O,70}}{40\overline{\ }_O}\right) \quad (8)$$

Where x_{30} and x_{70} are the data with exceedance probabilities of 30% and 70%. The high- and low-flow signatures, at the 5% and 95% exceedance probabilities were calculated using equations 9 and 10:

$$Q_5 = \frac{x_{O,5} - x_{S,5}}{x_{O,5}} \quad (9)$$

$$Q_{95} = \frac{x_{O,95} - x_{S,95}}{x_{O,95}} \quad (10)$$

Where x_5 and x_{95} are the data with exceedance probabilities of 5% and 95%. The previously described relative variability and bias metrics (which are components of KGE) can also be considered flow signatures based on discharge statistics (Shafii and Tolson, 2015).

Finally, a metric quantifying the representativeness of the simulated slope of the local mixing line (LML) was included in the analysis (Stadnyk and Holmes, 2020). The LML slope error, which uses the simulation of both isotopes in combination, is the difference between the best fit slope for the simulated river isotope compositions and the best fit slope for the observed river isotope compositions:

$$LML\ mE = \frac{\sum_{i=1}^{n}(O_{S,i} - \overline{O}_S)(D_{S,i} - \overline{D}_S)}{\sum_{i=1}^{n}(O_{S,i} - \overline{O}_S)^2} - \frac{\sum_{i=1}^{n}(O_{O,i} - \overline{O}_O)(D_{O,i} - \overline{D}_O)}{\sum_{i=1}^{n}(O_{O,i} - \overline{O}_O)^2} \quad (11)$$

Where $O_{O,i}$ is oxygen-18 observation i, and $O_{S,i}$ is the simulated oxygen-18 value for the time observation i was taken, and $D_{O,i}$ and $D_{S,i}$ are likewise the observation and simulated value for the deuterium data.

In addition to individual simulation performance metrics, which can potentially be used as calibration objectives, there are also ensemble performance metrics, which quantify the performance of groups of simulations (e.g., the outcome of multi-objective calibration). In this study ensemble performance is assessed using two indices: the containment ratio, and the relative

band-width. The containment ratio is simply the ratio of observations which are contained within the ensemble upper and lower bounds to the total number observations (Xiong et al., 2009). The average relative band-width is calculated as (Xiong et al., 2009):

$$RB = \frac{\sum_{i=1}^{n}\left(x_{s,i}^{u} - x_{s,i}^{l}\right)/x_{o,i}}{n} \quad (12)$$

Where n is number of observations, $x_{o,i}$ is observation i, $x_{s,i}^{u}$ is the maximum corresponding simulated value from the ensemble (i.e., the upper simulated bound) and $x_{s,i}^{l}$ is the minimum corresponding simulated value from the ensemble (i.e., the lower simulated bound).

5.3.2.2. Parameter sensitivity and metric selection

Parameter sensitivities for the various performance metrics are evaluated to design informed calibration strategies utilizing isotope tracer data. The best method is a global sensitivity analysis (GSA), which evaluates parameter sensitivities over the entire specified parameter space (Song et al., 2015). This study uses VARS, a GSA approach which is uses variograms (measuring variance of differences in the response surface over the parameter space) and quantifies global parameter sensitivity by integrating the variogram across multiple scales (Razavi and Gupta, 2016). VARS, which is implemented in the VARS-TOOL software package, was selected for the GSA over older approaches for this study due to the relative computational efficiency of the method, which is a key consideration in applying GSA to a large-scale process-based hydrologic model (Razavi et al., 2019). The VARS-TOOL software was used to generate a total of 48,800 sampled parameter sets (specifics of the sampling space are listed in Table D1). Variograms can integrated for different perturbation scales, but the $IVARS_{50}$ (integrated variogram across a range of scales, from 0 to 50% of the scale range) index has been found to be the best index for global sensitivity and is therefore used in this study (Razavi and Gupta, 2016). The 90% confidence intervals on the sensitivity results were estimated with bootstrapping, using 1000 sampling iterations.

The sensitivity results in this study are normalized, such that all values are relative to the total sensitivity of a given metric (i.e., the individual IVARS50 values are normalized using the sum of the IVARS50 for all parameters). Parameter sensitivities were calculated for the averaged performance at all calibration sites (see Table 5-3 for the list of calibration locations), for all 27 simulation performance metrics described in 5.3.2.1. These parameter sensitivities quantify the

responsiveness of metrics to changes in parameter values and do not identify goodness of fit or optimal parameter values, rather, they indicate a metric's ability to distinguish between 'good' or 'bad' parameter values.

Parameter sensitivity values are used to select possible metrics for the multi-objective optimization (Section 5.3.2.4). To select candidate metrics for this study in a more objective and reproducible manner:

1. Metrics (for flow and isotope data separately) are ordered from most types of error covered (as categorized in Holmes et al. (2022)), to least
2. Parameter sensitivities are classified as highly sensitive (over 10% of total sensitivity, bootstrapped sensitivity uncertainty less than 10% of total sensitivity), likely sensitive (over 3% of total sensitivity, sensitivity uncertainty less than 25%), possibly sensitive (over 1% of total sensitivity), and no detected sensitivity (less than 1% of total sensitivity)
3. The metric with the most broadly distributed sensitivity (i.e., highest number of sensitive parameters) is selected as the primary candidate metric for a given data type
4. When searching for second candidate metric for the same data type, the metric increasing the sensitivity class for the most parameters, in combination with the primary candidate, is selected
5. In cases of a tie (i.e., equal numbers of sensitive parameters) the metric covering the most types of error is chosen

Seven different combinations of data types are used to select candidate metric pairs: flow-only, isotope-only, oxygen-18 only, deuterium only, flow and oxygen-18, flow and deuterium, and flow and both isotope tracers.

5.3.2.3. Monte Carlo Analysis

Exploring the relationship between parameter values and model performance requires sampling the full range of all parameters and running the simulation with each parameter set to generate simulation results. For this study, a Monte Carlo analysis is performed, where all 27 calibrated parameters are sampled across their entire calibration range using a uniform random distribution. A total of 100,000 random parameter sets are generated and used in the isoWATFLOOD model of the Athabasca watershed. The simulation performance for the resulting flow and tracer outputs

was quantified using the all the individual simulation performance metrics (Section 2.2.1). Using the average performance for calibration sites, the Monte Carlo results are used to assess the parameter identifiability for each performance metric independently, by selecting the best 0.1% solutions for each metric and finding the distribution (i.e., range, median and interquartile range) of parameter values for those 100 simulations. The key benefit of the Monte Carlo analysis is that the parameter sets, and therefore simulation results, are completely independent of each other, which permits robust statistical treatment of the results; Monte Carlo simulations have also been used to calibrate models with mixed flow and isotope data and should therefore be included in the method comparison (Delavau et al., 2017; Neill et al., 2019; Piovano et al., 2020). However, using randomly generated parameter sets is computationally inefficient, and the best performing simulations will have a very low likelihood of matching the quality of solutions produced by an optimization algorithm (Efstratiadis and Koutsoyiannis, 2010).

5.3.2.4. Multi-Objective Optimization

The current best practice for hydrologic model calibration is the use of optimization algorithms, particularly multi-objective automated search algorithms. An optimization algorithm may start from a random point in the parameter space (or from a specified initial solution), but rather than continuing to generate random solutions, it will the adjacent region of the parameter space for better solutions. The algorithm judges solutions as better or worse based on the value of the objective function, or objective functions for multi-objective algorithms; the selection of an objective is therefore a key decision in model calibration.

This study uses the PA-DDS (Pareto archiving dynamically dimensioned search) algorithm for multi-objective model calibration, as recommended by (Holmes et al., 2020). This algorithm is based on the DDS algorithm (Tolson and Shoemaker, 2007), a computationally efficient search method where the search space is gradually constrained as the algorithm completed iterations, but with the additional ability to retain an archive of solutions which are not Pareto dominated (i.e., all solutions which are not categorically outperformed, considering multiple objectives) (Asadzadeh and Tolson, 2013). Model calibrations were performed using the OSTRICH program, a model-independent calibration tool which implements the PA-DDS algorithm among others (Matott, 2017). For each set of candidate objectives, five separate calibration trials were run, with different random seeds (the method for choosing performance metrics as possible

candidates is covered in 2.2.2). Each trial ran for 1000 iterations, and all non-dominated solutions from each of the five trials were retained in the initial results analyses. The five trials for each candidate performance metric pair shared initial solutions: non-dominated solutions for the candidate metrics were selected from the Monte Carlo results, as better-quality initial solutions can reduce the computational budget required for optimization. While PA-DDS is not limited to two objectives, this study is limited to paired objectives, as increasing the number of objectives generally increases the computational requirements and would add unnecessary complications to comparing data types in calibration. Calibrations using an optimization algorithm generally have better simulation performance than those using random parameter sets in a Monte Carlo analysis and using multi-objective optimization generates substantial numbers of parameter sets producing good flow or tracer simulations. However, the solutions from a particular trial are not fully independent of each other (as the algorithm generates new solutions from progressive modifications of previous ones), which limits or complicates the statistical treatment of the results.

5.4. Results

5.4.1. Monte Carlo Parameter Identifiability

The best 100 solutions from the 100,000 randomly generated parameter sets run in the Monte Carlo analysis are selected for each flow and isotope performance metric independently. Box-whisker plots of all parameters (normalized to the calibration range) and all assessed metrics are shown in Figure 5-4.

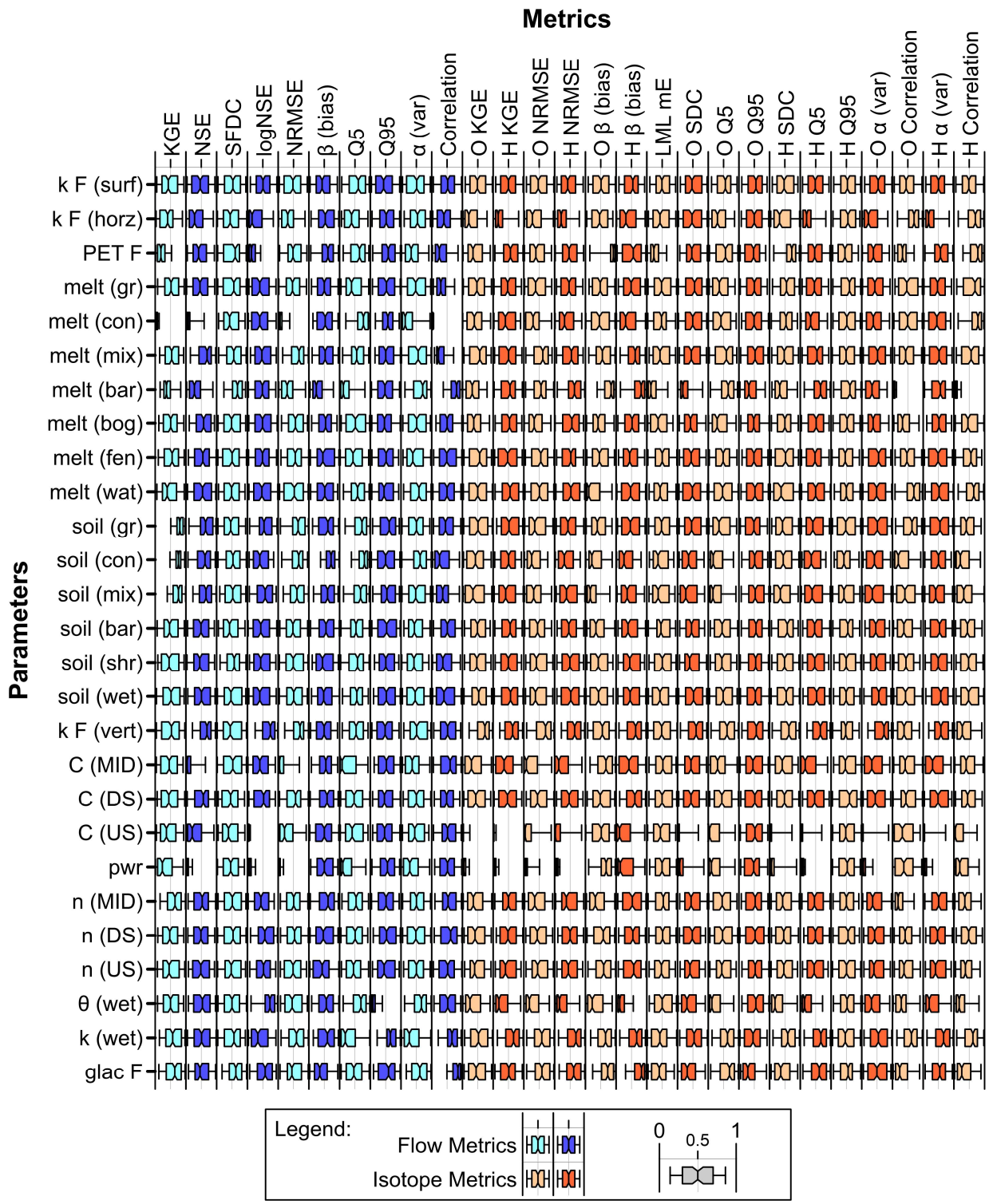

Figure 5-4: Box-whisker plots of normalized parameter values for the best 100 solutions from the Monte Carlo analysis, based on each performance metric independently. Whiskers extend to the 5/95 percentile, and flow and isotope metrics are shown in blue and orange respectively.

Overall, there is limited identifiability for most parameters, with the majority of box-whisker plots exceeding 50% of the calibration range. Some metrics identify specific values for a few parameters, namely melt rates for coniferous forest and barren ground, the baseflow power, and the baseflow coefficient in the mountains. Most parameters have some relationship with at least one performance metric, where the parameter values for the 100 selected solutions are either limited to or skewed towards only part of the total calibration range. For example, the distribution of values for the horizontal conductivity factor is skewed for the majority of performance metrics, and parameter value range is somewhat identified for the wetland porosity parameter with a few metrics. Only the surface soil conductivity parameter and the downstream channel roughness are completely unidentifiable, retaining a uniform distribution of parameter values across the entire possible parameter range. Generally, the flow and isotope performance metrics are aligned in either identifying or not identifying parameter values, but not in all cases; the flow and oxygen-18 performance metrics identify opposite ends of the possible parameter value range for the soil water retention capacities for coniferous and mix forest. While intriguing, the Monte Carlo results do not identify parameter values well, and are inconclusive on the utility of isotope data for improving parameter identifiability. The parameter space, with 27 interacting calibrated parameters, is both large and complex and even the best simulations produced were of mediocre quality (e.g., best flow KGE range between 0.48 and 0.52).

5.4.2. Parameter Sensitivity and Candidate Objective Metric Selection

A VARS global sensitivity analysis was performed for all evaluated performance metrics independently, using the average performance at calibration sites as the response variable. The normalized sensitivities and sensitivity reliabilities are shown in Figure 5-5.

		KGE	NSE	SFDC	logNSE	NRMSE	β (bias)	Q5	Q95	α (var)	Correlation	O KGE	H KGE	O NRMSE	H NRMSE	O β (bias)	H β (bias)	LML mE	O SDC	O Q5	O Q95	H SDC	H Q5	H Q95	O α (var)	O Correlation	H α (var)	H Correlation
	k F (surf)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
	k F (horz)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.08	0.22	0.23	0.25	0.25	0.03	0.02	0.10	0.17	0.04	0.16	0.18	0.04	0.01	0.07	0.01
	PET F	0.01	0.00	0.02	0.44	0.00	0.01	0.00	0.05	0.08	0.00	0.14	0.00	0.01	0.00	0.01	0.00	0.25	0.09	0.08	0.00	0.00	0.00	0.00	0.14	0.10	0.00	0.01
Snowpack	melt (gr)	0.01	0.01	0.11	0.01	0.01	0.00	0.00	0.00	0.06	0.06	0.02	0.00	0.00	0.00	0.00	0.00	0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.02	0.01	0.00	0.02
Snowpack	melt (con)	0.09	0.06	0.14	0.01	0.13	0.00	0.01	0.00	0.23	0.52	0.06	0.14	0.01	0.01	0.00	0.00	0.06	0.01	0.00	0.08	0.01	0.00	0.08	0.06	0.07	0.15	0.04
Snowpack	melt (mix)	0.01	0.01	0.02	0.00	0.02	0.00	0.00	0.00	0.05	0.03	0.02	0.03	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.02	0.00	0.03	0.01
Snowpack	melt (bar)	0.01	0.00	0.30	0.01	0.01	0.04	0.01	0.00	0.05	0.05	0.04	0.10	0.13	0.13	0.11	0.10	0.06	0.20	0.00	0.17	0.46	0.00	0.16	0.04	0.26	0.10	0.36
Snowpack	melt (bog)	0.00	0.00	0.03	0.01	0.00	0.00	0.00	0.01	0.02	0.01	0.02	0.06	0.08	0.08	0.06	0.06	0.04	0.08	0.01	0.10	0.24	0.00	0.10	0.02	0.01	0.06	0.02
Snowpack	melt (fen)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Snowpack	melt (wat)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.04	0.01	0.01	0.00	0.00	0.00	0.00	0.05	0.02	0.00	0.02
Retention	soil (gr)	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.16	0.16	0.22	0.22	0.00	0.02	0.42	0.09	0.00	0.68	0.09	0.01	0.00	0.00	0.02
Retention	soil (con)	0.02	0.02	0.02	0.01	0.03	0.02	0.05	0.01	0.04	0.01	0.02	0.03	0.07	0.07	0.06	0.06	0.03	0.02	0.01	0.07	0.04	0.00	0.07	0.02	0.03	0.03	0.03
Retention	soil (mix)	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.03	0.00	0.00	0.00	0.00	0.00	0.03	0.05	0.04	0.00	0.00	0.00	0.00	0.03	0.01	0.00	0.02
Retention	soil (bar)	0.04	0.04	0.01	0.01	0.03	0.04	0.04	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.01
Retention	soil (shr)	0.02	0.02	0.00	0.01	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00
Retention	soil (wet)	0.02	0.02	0.00	0.01	0.02	0.02	0.02	0.01	0.01	0.01	0.04	0.06	0.00	0.00	0.00	0.00	0.02	0.01	0.02	0.00	0.00	0.01	0.00	0.04	0.02	0.06	0.02
	k F (vert)	0.02	0.02	0.00	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.02	0.02	0.00	0.01
Baseflow	C (MID)	0.03	0.03	0.01	0.03	0.03	0.02	0.03	0.07	0.05	0.02	0.04	0.00	0.00	0.00	0.00	0.00	0.04	0.02	0.01	0.00	0.00	0.00	0.00	0.04	0.03	0.00	0.01
Baseflow	C (DS)	0.02	0.02	0.00	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.06	0.13	0.12	0.13	0.14	0.14	0.03	0.01	0.07	0.10	0.00	0.12	0.10	0.06	0.02	0.13	0.03
Baseflow	C (US)	0.02	0.02	0.01	0.04	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.02	0.03	0.01	0.00	0.01	0.00	0.00	0.01	0.02	0.00	0.02
Baseflow	pwr	0.08	0.08	0.04	0.08	0.08	0.07	0.08	0.07	0.11	0.04	0.07	0.04	0.15	0.16	0.13	0.15	0.07	0.10	0.04	0.14	0.12	0.01	0.14	0.07	0.06	0.04	0.06
Channel	n (MID)	0.04	0.04	0.01	0.02	0.03	0.04	0.04	0.02	0.02	0.02	0.04	0.03	0.00	0.00	0.00	0.00	0.04	0.03	0.02	0.01	0.00	0.00	0.01	0.04	0.05	0.03	0.03
Channel	n (DS)	0.04	0.04	0.01	0.02	0.04	0.04	0.04	0.01	0.01	0.01	0.05	0.05	0.01	0.00	0.00	0.00	0.02	0.03	0.02	0.02	0.00	0.00	0.02	0.05	0.02	0.05	0.01
Channel	n (US)	0.11	0.12	0.01	0.02	0.10	0.12	0.12	0.01	0.01	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.02	0.04	0.02	0.00	0.00	0.00	0.00	0.02	0.03	0.00	0.01
	θ (wet)	0.10	0.10	0.02	0.13	0.09	0.11	0.11	0.34	0.05	0.03	0.02	0.01	0.00	0.00	0.00	0.00	0.03	0.04	0.02	0.00	0.01	0.00	0.00	0.02	0.03	0.01	0.03
	k (wet)	0.11	0.12	0.06	0.07	0.11	0.11	0.12	0.28	0.11	0.03	0.12	0.22	0.02	0.02	0.00	0.00	0.04	0.04	0.04	0.05	0.01	0.01	0.05	0.12	0.04	0.22	0.05
	glac F	0.20	0.21	0.17	0.02	0.18	0.24	0.22	0.02	0.05	0.08	0.03	0.00	0.00	0.00	0.00	0.00	0.03	0.08	0.02	0.00	0.03	0.00	0.00	0.03	0.11	0.00	0.16

Figure 5-5: Relative parameter sensitivity for averaged calibration site performance; insensitive parameters are highlighted in blue and highly sensitive parameters shaded in orange (darker shading is more sensitive/insensitive). Red bars summarize 90% uncertainty range in sensitivity values (displaying 0 to 0.5 relative sensitivity, where sensitivity values with higher uncertainty have longer bars).

A few parameters are consistently insensitive for all evaluated metrics, namely the surface soil conductivity (i.e., kF (surf)) and connected wetland snowmelt rate (i.e., melt(fen)), but generally

parameters are sensitive for some metrics and not others. Both flow and isotope metrics are sensitive to some snowmelt and baseflow parameters. Isotope performance metrics are more sensitive to upper zone soil retention parameters, relative to the flow performance metrics, but these sensitivities are not very reliable. Likewise, flow metrics are more sensitive to wetland and glacier parameters but not always reliably. The LML slope error, and the oxygen-18 KGE and SDC standout for their balanced sensitivities, with nearly equal shares of the total sensitivity for each individual parameter.

The sensitivity results share some similarities with the Monte Carlo identifiability results (Figure 5-4), but with too many anomalies to make either analysis type a good predictor of the other. The melt rate for coniferous forest is both sensitive and identifiable using flow performance metrics, but glacier melt is sensitive and generally poorly identified with the same. Unidentifiable parameters match with consistently insensitive ones, yet isotope metrics are better at identifying baseflow coefficients in the upstream and worse at identifying the coefficient in the downstream than would be predicted based on the parameter sensitivities.

The data presented in Figure 5-5 is the basis of the metric candidate identification process, where performance metric pairs providing the widest parameter sensitivity coverage (with some additional conditions as listed in 5.3.2.2) are selected for use in multi-objective calibration. The categorized parameter sensitivity coverage for the candidate paired performance metrics is shown in Table 5-4.

Table 5-4: Parameter sensitivity coverage for paired performance metrics; the highest sensitivity category for the metric pair is shown as green/solid point for highly sensitive parameters, yellow/hollow point for likely sensitive parameters, and red/small point for possibly sensitive parameters.

Data type	**Q**	**Q O**	**Q O**	**Q O H**	**O**	**Q H**	**O H**	**H**
Objectives	KGE Q	KGE Q	KGE Q	KGE Q	KGE O	KGE Q	KGE O	KGE H
	SFDC	KGE O	NRMSE O	LML mE	NRMSE O	KGE H	H corr	H corr
k F (surf)		◦			◦		◦	
k F (horz)		○	◦	○	○	○	○	○
PET F	◦	●	◦	○	●	◦	●	◦
melt (gr)	●	◦	◦	◦	◦	◦	◦	◦
melt (con)	●	○	○	○	○	○	○	○
melt (mix)	◦	◦	◦	◦	◦		◦	○
melt (bar)	●	○	◦	○	○	◦	●	○
melt (bog)	○	◦	○	○	○	○	◦	○
melt (fen)								
melt (wat)		○		○	○		○	◦
soil (gr)			◦		◦		◦	◦
soil (con)	◦	◦	○	○	○	○	○	○
soil (mix)		◦		○	○		○	◦
soil (bar)	○	○	○	○	◦	○	◦	◦
soil (shr)	◦	◦	◦	◦	◦	◦	◦	
soil (wet)	◦	○	◦	◦	○	○	○	○
k F (vert)	◦	◦	◦	◦	◦	◦	◦	◦
C (MID)	○	○	○	○	○	○	○	◦
C (DS)	◦	○	◦	○	○	◦	○	○
C (US)	◦	◦	◦	◦	◦	◦	◦	◦
pwr	○	○	○	○	○	○	○	○
n (MID)	○	○	○	○	○	○	○	○
n (DS)	○	○	○	○	○	○	○	○
n (US)	○	○	○	○	◦	○	◦	◦
θ (wet)	○	○	○	○	◦	○	○	○
k (wet)	○	○	○	○	◦	○	○	○
glac F	●	○	◦	○	○	◦	●	○

Candidate pairs were selected when considering only flow metrics, flow and isotope metrics and isotope metrics only (see section 5.3.2.2 for the selection rules). One additional candidate pair

was added, as the NRMSE for isotope simulation performance is too commonly used to be overlooked in this analysis. Overall, there is limited variation in the parameter sensitivity coverage between the eight different candidate calibration metric pairs; by maximizing the number of sensitive parameters within the data constraints, the resulting coverage is fairly similar, with most parameters being either slightly or somewhat sensitive to at least one of the paired metrics. The lower zone, wetland and channel parameters have similar sensitivity for all candidate pairs, and there is more variation between candidates for upper zone, melt and evaporation parameter sensitivities.

5.4.3. Multi-Objective Calibration

Each candidate pair of performance metrics were used as objectives in five independent multi-objective calibration trials, using non-dominated solutions (for that candidate pair) from the Monte Carlo analysis as initial solutions. The outcomes of these 40 calibration trials are shown in Figure 5-6, along with the initial solutions.

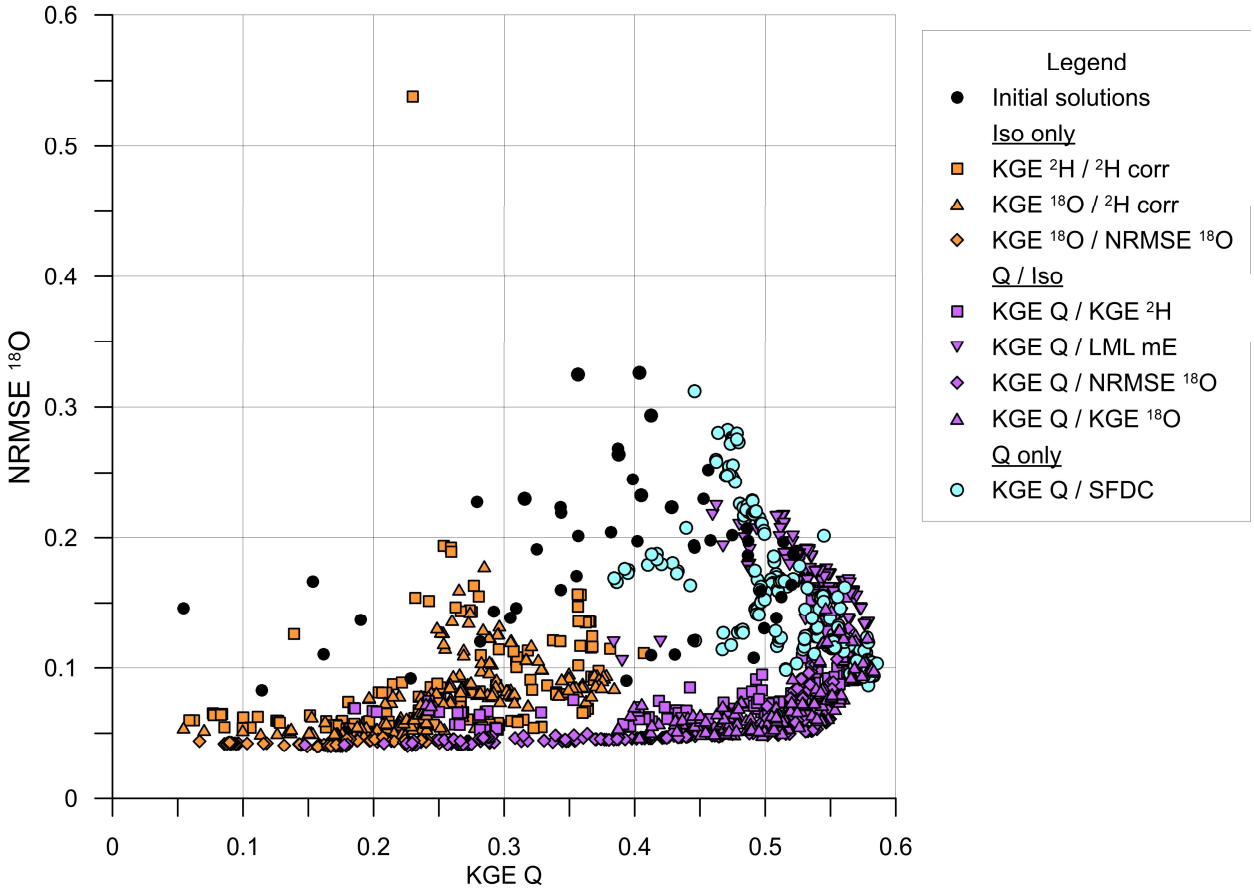

Figure 5-6: Calibration performance for the PA-DDS solutions for all candidate metric combinations.

Simulation performance after calibration is substantially improved compared to the initial random solutions (e.g., most solutions with KGE Q as an objective now have KGE Q exceeding 0.5). Figure 5-6 is not strictly a tradeoff front plot, as neither KGE Q nor NRMSE ^{18}O are objectives for all calibrations, but it still clearly depicts the cost of calibrating with an isotope tracer objective. The best flow simulation performance definitely does not have the best isotope tracer simulation performance. If equivalent weight is given to isotope and flow performance, the flow simulation performance in calibration will be lower than if flow performance is considered alone. However, considering the uncertainty in performance metric values, a decrease in calibration KGE Q of 0.03 is not particularly serious. The two isotope tracers are similar but not identical; when KGE ^{2}H is an objective, the calibration performance does not quite overlap with solutions found with KGE ^{18}O as an objective. On the other hand, using NRMSE ^{18}O or KGE ^{18}O as objectives produces equivalent calibration performances. There is a clear segregation of

objective pair types: the flow-only pair has the best flow performance, the isotope-only pairs have the best isotope simulation performance, and the mix flow-isotope pairs fall between the other two, with some overlap. The difference between the best flow performance for the flow-only and mixed flow-isotope calibrations is negligible. The isotope-only calibrations do not generally have good flow simulations, with a mean and median flow KGE of 0.28, flow simulations are of similar quality to those from random parameter sets in the Monte Carlo analysis (mean and median KGE 0.20 and 0.21). Isotope-only calibrations only outperformed random solutions in avoiding very poor flow simulations, with a minimum flow KGE of 0.05, compared to -0.27 from the Monte Carlo analysis.

The relationships between calibration performance and calibrated parameter values for several key parameters are illustrated in Figure 5-7 (only solutions with NRMSE ^{18}O less than 0.5 are shown).

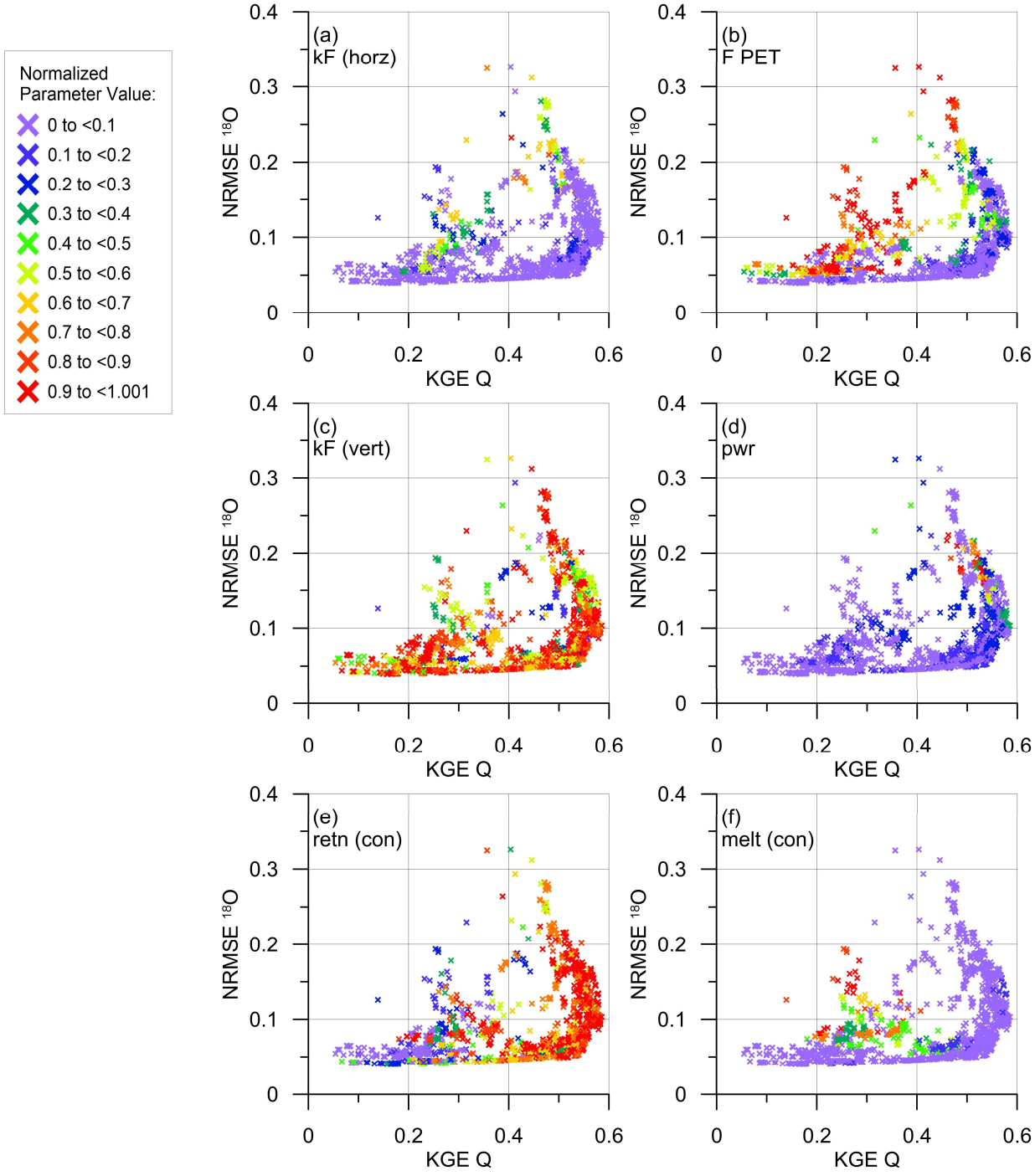

Figure 5-7: Calibration performance for all PA-DDS solutions, with point color indicating normalized parameter values.

These highly influential parameters generally have identified parameter values or correlated simulation performance and parameter values. Some are identified by both simulation types (i.e.,

kF, the horizontal and vertical soil conductivity), but others are better identified by either the flow simulation performance (i.e., coniferous forest soil retention) or the isotope performance (i.e., the evaporation adjustment factor), where parameter values are only consistent for the best performing simulations of one type.

The final step in the calibration process is the selection of acceptable solutions from the total output of the 40 non-dominated solution Pareto-fronts produced by the multi-objective optimization algorithm. As the flow simulation is the primary model output of interest for a hydrologic model, only solutions with calibration KGE Q greater than 0.5 will be considered behavioral solutions. This threshold entirely removes isotope-only calibrations from further analysis, as no behavioral solutions were produced by calibrating to isotope tracer performance alone. The average flow simulation at validation sites for behavioral solutions is shown in Figure 5-8, along with the calibration performance.

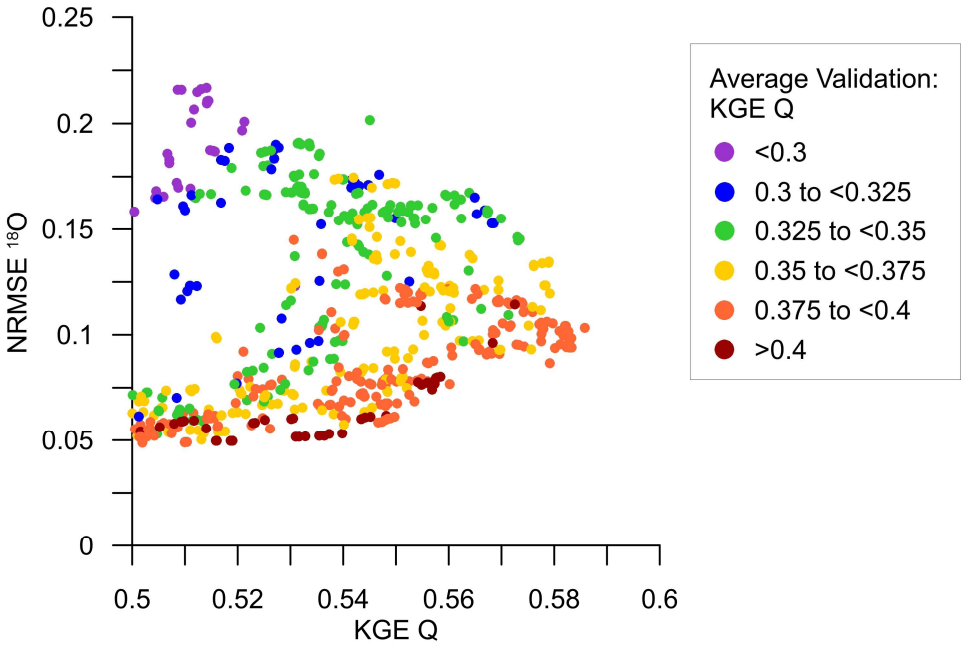

Figure 5-8: Calibration and validation performance for PA-DDS solutions with behavioral flow (KGE Q ≥ 0.5) calibration performance.

As is typical for hydrologic models, the validation performance decreases relative to the calibration performance, with average KGE Q exceeding 0.5 for calibration sites, and ranging

between 0.26 and 0.42 for validation sites. The best validation flow simulation performance for behavioral solutions is best predicted by the isotope simulation performance in calibration, not the flow simulation performance in calibration; the best validation KGE Q are for solutions with low NRMSE ^{18}O, and the worst are for solutions with higher NRMSE ^{18}O.

The behavioral solutions from the calibrated parameter sets have better parameter identifiability than the best solutions from random parameter sets, although without the same statistical robustness: solutions generated from the same PA-DDS multi-objective calibration trial are not truly independent, although the five trials per candidate objective pair are independent. The normalized parameter values from all behavioral solutions, for all parameters, are shown in the histograms in Figure 5-9.

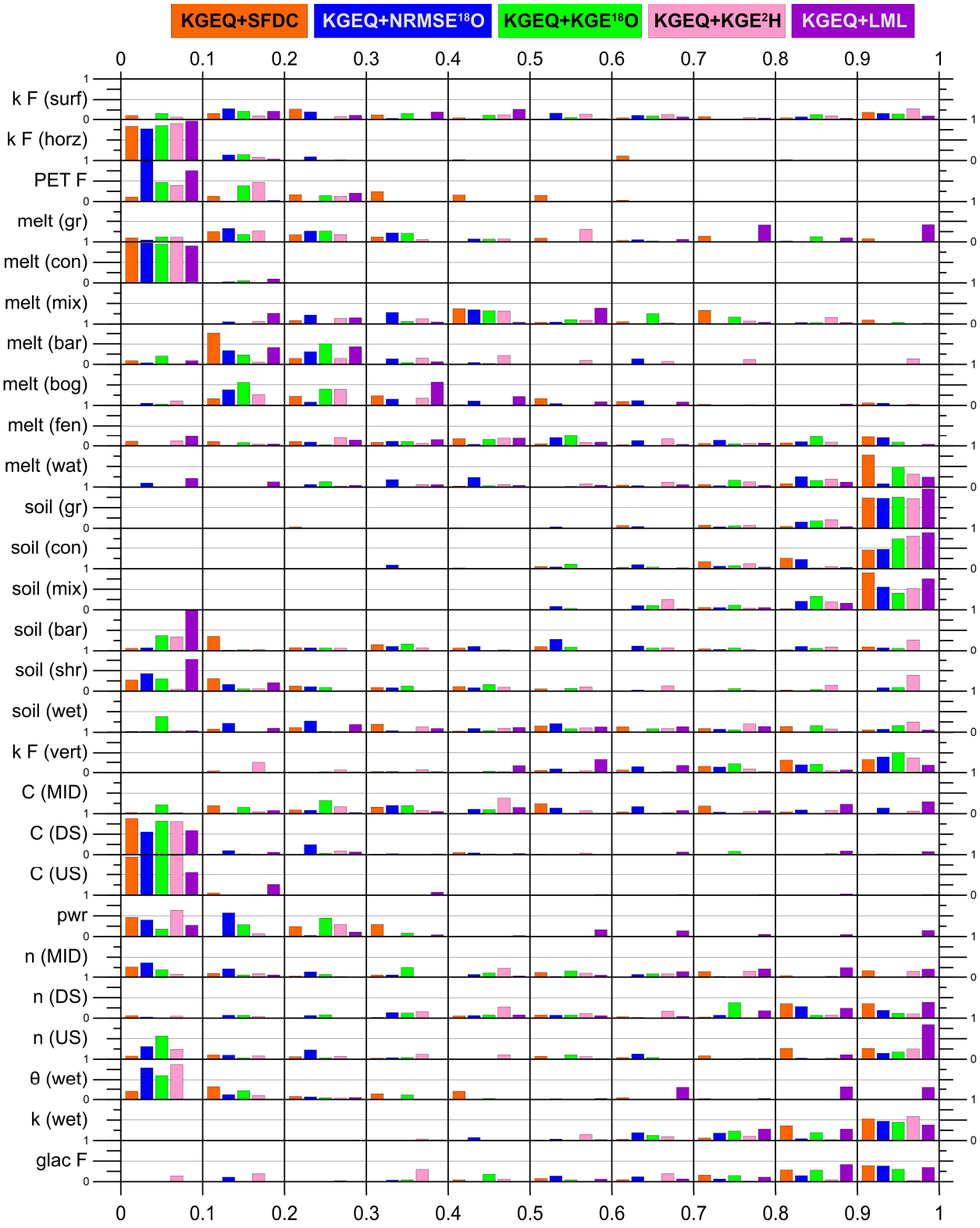

Figure 5-9: Parameter identifiability for behavioral calibrated solutions. All 27 normalized parameters' histograms are shown for the 5 candidate objective metric combinations.

A majority of parameters have values identified to within 30% of the total calibrated parameter range, and a few have genuinely well-identified values (i.e., the horizontal soil conductivity, coniferous forest snowmelt rate and the baseflow coefficient in the upstream region). Soil retention, wetland parameters and the evaporation adjustment factor are generally somewhat identified. All five candidate objective pairs have broadly similar identifiable and unidentifiable parameters, and usually agree on approximately the same parameter values. The two oxygen-18 calibrations (using KGE ^{18}O and NRMSE ^{18}O as second objectives) are extremely similar, and the calibration using KGE ^{2}H as the second objective is similar to both oxygen-18 calibrations, except for the glacier melt factor. The flow-only calibration and the calibration using the LML slope error as the second objective diverge from the other calibrations for some parameters; as an example, the three pairs using an isotope metric including timing error all converge on a low value of wetland porosity, but LML slope error calibration tends toward a high porosity value, and the flow-only calibration does not identify a value for porosity. It is noted, however, that using isotope data in calibration does not consistently improve parameter identifiability (e.g., the evaporation factor and wetland porosity are better identified, but mixed wood upper zone soil water retention has a less identifiable parameter value).

5.4.4. Calibrated ensemble performance

The final stage of evaluating the calibration effectiveness for the various objective function combinations is the assessment of the ensemble performance. All behavioral solutions from the five PA-DDS calibration trials for each candidate objective pair were included, with equal weight, in an ensemble of parameter sets, such that the simulation uncertainty from parameter uncertainty can be seen. The average annual hydrographs for the three most divergent calibrations are shown in Figure 5-10, at the Athabasca River at Fort McMurray, which covers the large majority of the modeled area (other hydrographs can be found in Appendix D).

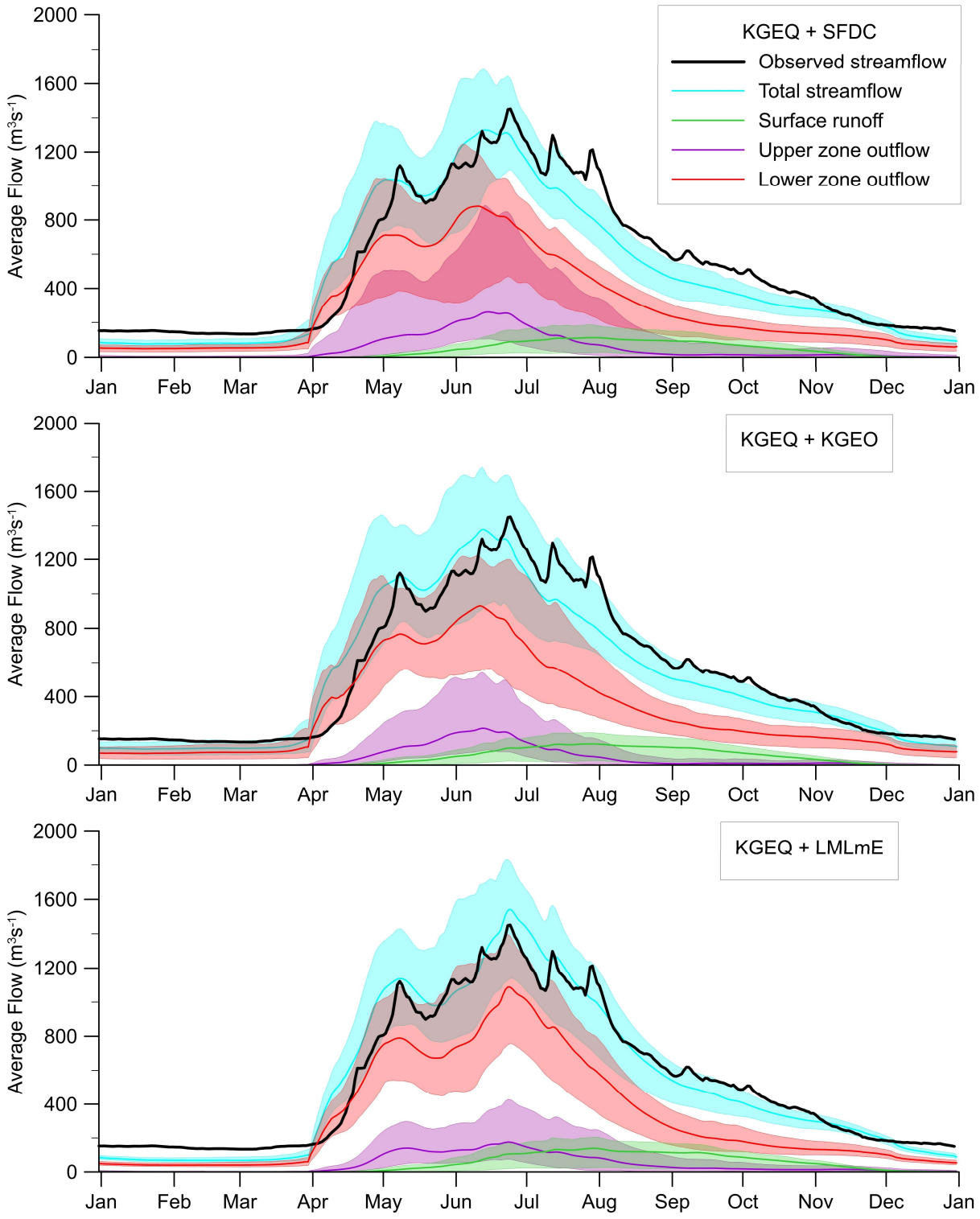

Figure 5-10: Average annual hydrographs for the simulation period (2001 to 2015), with behavioral calibrated ensembles shown, along with surface, upper and lower zone virtual tracer flows at Fort McMurray 07DA001 (mean and total range are included).

Parameter identifiability does not decisively improve when isotope data are included in calibration, but the final ensembles produced are unmistakably different. Mixed flow-isotope calibrations have similar uncertainty in total flow to a flow-only calibration, with slightly improved observation containment (at this gauge at least). On the other hand, the isotope-enabled calibration ensembles have significantly higher confidence in the component contributions generating the total flow. The percent contributions from the upper and lower soil zones are listed in Table 5-5, along with the range in percent contributions.

Table 5-5: Percent contributions from the upper and lower zones to the total streamflow, with the mean, maximum and minimum fractions averaged for all gauges.

		Lower zone (% contribution)			Upper zone (% contribution)		
Objectives		**Average**	**Range**	**(max/min)**	**Average**	**Range**	**(max/min)**
KGEQ	**SFDC**	62.0	26.3	(69/43)	7.6	25.8	(29/3)
KGEQ	**NRMSEO**	61.4	19.9	(71/51)	6.8	12.1	(15/3)
KGEQ	**KGEO**	63.6	15.9	(71/55)	5.4	9.9	(13/3)
KGEQ	**KGEH**	59.7	14.4	(67/53)	6.4	10.7	(13/3)
KGEQ	**LML**	63.4	16.2	(71/55)	5.6	11.5	(14/3)

Average contributions from the two soil zones are consistent for all calibrated ensembles, but the ranges in those contributions are substantially different. All mixed flow-isotope calibrated ensembles have much lower maximum upper zone contributions, and much higher minimum lower zone contributions than the flow-only ensemble. This improvement in flow component identifiability is linked to improvements in process flux identifiability for soils, as shown in Figure 5-11. Median and inter-quartile ranges for soil fluxes are similar for all calibration objective pairs, but the flow-only calibration has substantially long tails than mixed isotope-flow calibrations.

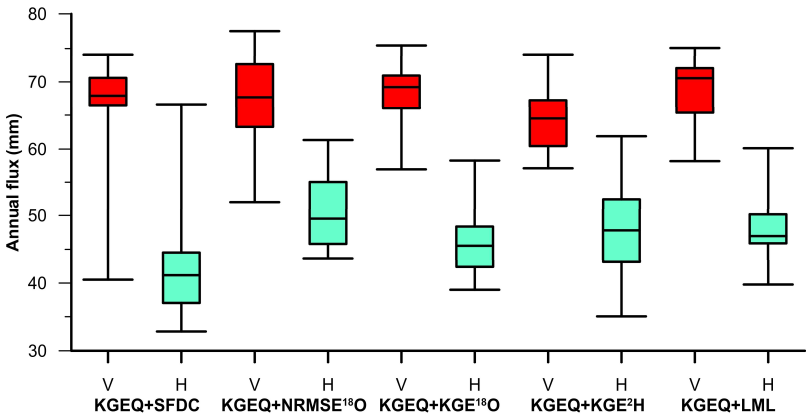

Figure 5-11: Box-whisker plots of average annual vertical (recharge) and horizontal (interflow and overland flow) fluxes from the near surface for behavioral calibrated solutions, with whiskers extending to maximum/minimum.

The similarity in the total flow uncertainty and the improvement in the ensemble observation containment can also be supported quantitatively for gauge sites generally, with the containment ratios and relative bounds in Table 5-6.

Table 5-6: Average containment ratios and relative band-width for all behavioral calibrated ensembles.

		Containment Ratio			Relative Band-Width		
Objectives		Calibration	Validation	Overall difference from flow-only (p-value)	Calibration	Validation	Overall difference from flow-only (p-value)
KGEQ	**SFDC**	0.34	0.29	-	0.86	1.28	-
KGEQ	**NRMSEO**	0.44	0.33	0.001	0.92	1.24	0.809
KGEQ	**KGEO**	0.45	0.31	0.026	0.87	1.01	0.065
KGEQ	**KGEH**	0.42	0.33	0.001	0.96	1.39	0.019
KGEQ	**LML**	0.28	0.23	0.005	0.81	0.81	0.025

The three mixed flow-isotope calibrations using an isotope performance metric that includes timing errors all have containment ratios better than the flow-only calibration for both calibration and validation gauges, and these improvements in the mean containment ratio are statistically significant. There is no significant change in the relative band-width for the ensembles calibrated

with flow and oxygen-18 data; the ensemble calibrated with KGE Q and KGE ^{2}H has wider relative bounds. The final flow-isotope calibrated ensemble, calibrated to maximize KGE Q and minimize LML slope error, has significantly narrower relative band-width, but a lower containment ratio than the flow-only calibration.

5.5.Discussion

5.5.1. Parameter identifiability

Overall, parameter values were not well identified, regardless of metric or data type, with a few exceptions. The small number of well-identified parameter values is unsurprising, given the model complexity, number of calibrated parameters, and large and heterogeneous study area. The best identified parameters, for both flow and isotope data, were the baseflow parameters, which had generally constrained values regardless of the assessment method. That baseflow is the best identified process is entirely unsurprising, as the majority of total streamflow is routed through the lower soil zone as baseflow and the process is controlled by a limited number of parameters. The snow melt rate parameter for coniferous forest (the dominant land class) was well identified by most flow metrics, including KGE Q, but was not well identified by isotope tracer metrics. On the other hand, the PET factor is substantially better identified by isotope tracer metrics than flow metrics alone. In some cases, parameter values may be best identified with a combination of flow and isotope tracer simulation performance; the values of horizontal soil conductivity and wetland porosity are most constrained when both flow and isotopes are optimized. Isotope and flow data only identified different values for the same parameter for upper zone soil water retentions, with the best flow performance associated with very high retention, and the best isotope retention associated with low retention. It is entirely possible to have isotope simulation performance of approximately similar quality with either low or high retention but if flow simulation performance is not considered in parameter value selection, lower retention values would be identified. Altogether, isotope tracer data does not lead to dramatic alterations in parameter identifiability, but the limited changes do tend to be improvements. With so many interacting parameters within the isoWATFLOOD model, individual parameter identifiability is weak, even with additional tracer data.

The addition of isotope performance to the calibration was more adept at constraining the overall process contributions to streamflow (Figures 5-10 and 5-11) than identifying parameter values.

The use of isotope tracer data as a calibration objective avoided many outlying flux simulations and flow component combinations which were included by the calibrations using only flow data. It is not unexpected that overall fluxes are better constrained than parameter values, in models with large numbers of interacting processes and parameters. Given parameter identifiability is desired as a proxy for process identifiability, it is preferable to have better process identification without better parameter value identification than the reverse.

5.5.2. Simulation performance

Both flow-only and mixed isotope-flow multi-objective calibrations resulted in numerous acceptable flow simulations (based on a behavioral threshold of 0.5 flow KGE in calibration), while calibrating to isotope data alone did not. The PA-DDS calibrated solutions which used KGE Q as an objective had flow simulations outperforming any randomly generated parameter set, while the isotope-only calibrations had flow simulation performance similar to the random solutions from the Monte Carlo analysis. Clearly isotopes are not a substitute for hydrometric data. Using only sparse isotope data as a calibration objective may be a better alternative than purely random parameter values within general recommendation ranges, as isotope tracers may rule out some of the least accurate flow simulations. The Monte Carlo analysis was an inefficient method of identifying behavioral parameter sets for both flow and isotope tracer simulations; numerous higher quality solutions were identified using multi-objective calibration using 1% of the computation requirements of the Monte Carlo analysis.

The mixed flow-isotope multi-objective calibrations resulted in solutions with similar flow simulation performance, and better isotope tracer simulation performance, compared to the flow-only calibration. Considered as ensembles, the mixed isotope-enabled calibrations generally outperformed the flow-only calibration, with better observation containment for calibrations with a timed isotope performance metric as the second objective, and similar relative bounds on isotope-enabled calibrated ensembles. The sub-component ensembles benefited most from the inclusion of isotope tracer data in the calibration, with substantially narrower ranges. The additional isotope data excludes many process-contribution combinations which produce acceptable total streamflow simulations. The increased confidence in the flow components may be contributing factor in the correlation between isotope simulation performance in calibration and flow simulation performance in validation. This study has found a similar interrelationship

between isotope tracer performance and flow simulation performance to previous work with the isoWATFLOOD model, but in a different watershed, and using a different calibration methodology (Stadnyk and Holmes, 2020). Overall, isotope tracer data is beneficial to the flow simulation; adding isotope tracer data to the calibration leads to flow simulations comparable or better than simulations calibrated with flow data alone.

5.5.3. Calibration recommendations

Isotope tracer simulations and observation data have demonstrable value as a supplement to hydrometric data in hydrologic model calibration. If possible, calibrating using an isotope simulation performance metric as a secondary objective in a multi-objective calibration is recommended, as the results of this study indicate that mixed isotope-flow calibrations have better streamflow simulations, and better identified streamflow component simulations, than flow-only calibrations. Monte Carlo simulations are not well-suited to identifying parameter values in large-scale process-based models, and multi-objective optimizations for tracer-aided calibrations. It is important that the isotope performance metric include timing error for the isotope tracer simulation. Neither KGE nor NRMSE has a decisive advantage over the other as a calibration objective for tracer simulations. This study found no reason not to use NRMSE to compare isotope simulation quality, as it was equivalent to using KGE for improving isotope and flow model performance. Using KGE to quantify tracer simulation performance has the advantage of aligning with a common flow performance metric, but the metric is more vulnerable to sampling biases than NRMSE. The choice between stable isotope tracers (i.e., oxygen-18 or deuterium) is dependant on the model or the data available; in this study, oxygen-18 is preferrable as it can be simulated independently (the current isoWATFLOOD model code is limited to either oxygen-18 or both isotopes simultaneously), which may not be the case for other modeling software. In summary:

1. Use an isotope tracer performance metric as a second objective in multi-objective calibrations of process-based hydrologic models
2. Choose an isotope tracer simulation performance metric that includes timing error
    - Consider KGE for consistently sampled isotope data sets
    - Use NRMSE for irregularly sampled isotope data sets

5.6.Conclusions

Isotope tracer simulations, and specifically calibrations including tracer model performance optimization, are beneficial for both process representation and streamflow simulations. These benefits, including better ensemble observation containment, higher validation performance scores, and improved process contribution identification, can be realized even in mesoscale watersheds, and with limited isotope observation datasets. Flow data should remain the primary focus of hydrologic model calibration but using an isotope simulation performance metric which includes timing error as a secondary calibration objective leads to more robust streamflow modeling.

Acknowledgments

The authors acknowledge this study occurred within and about Treaty 8 and 6 regions, lands which are, or have historically been, home to no less than nine Indigenous peoples of Canada: the Dane-zaa, Sekani, Secwepemc (Shuswap), Salish, Ktunaxa, Nakoda/Stoney, Woodland Cree, Chipewyan (Denesoline), and Métis. The authors gratefully acknowledge those who have contributed to data collection required to conduct this study, including the Water Survey of Canada, Environment and Climate Change Canada and Innotech Alberta. This research was supported by the Natural Sciences and Engineering Research Council of Canada [CRD 462584-2013], and Global Water Futures [NSERC CFREF-GWF 418474].

5.7.References

Acero Triana, J.S., Chu, M.L., Guzman, J.A., Moriasi, D.N., Steiner, J.L., 2019. Beyond model metrics: The perils of calibrating hydrologic models. Journal of Hydrology 578. https://doi.org/10.1016/j.jhydrol.2019.124032

Alberta Geological Survey, 2013. Bedrock Geology of Alberta [WWW Document]. URL https://open.canada.ca/data/en/dataset/5155d48c-ce34-4493-b4f6-fb4eb94fb348 (accessed 9.20.21).

Asadzadeh, M., Tolson, B., 2013. Pareto archived dynamically dimensioned search with hypervolume-based selection for multi-objective optimization. Engineering Optimization 45, 1489–1509.

Bennett, N.D., Croke, B.F.W., Guariso, G., Guillaume, J.H.A., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, L.T.H., Norton, J.P., Perrin, C., Pierce, S.A., Robson, B., Seppelt, R., Voinov, A.A., Fath, B.D., Andreassian, V., 2013. Characterising performance

of environmental models. Environmental Modelling and Software 40, 1–20. https://doi.org/10.1016/j.envsoft.2012.09.011

Beven, K., Binley, A., 1992. The future of distributed models: Model calibration and uncertainty prediction. Hydrological Processes 6, 279–298. https://doi.org/10.1002/hyp.3360060305

Brooks, J.R., Mushet, D.M., Vanderhoof, M.K., Leibowitz, S.G., Christensen, J.R., Neff, B.P., Rosenberry, D.O., Rugh, W.D., Alexander, L.C., 2018. Estimating Wetland Connectivity to Streams in the Prairie Pothole Region: An Isotopic and Remote Sensing Approach. Water Resources Research 54, 955–977. https://doi.org/10.1002/2017WR021016

Coulibaly, P., Samuel, J., Pietroniro, A., Harvey, D., 2013. Evaluation of Canadian national hydrometric network density based on WMO 2008 standards. Canadian Water Resources Journal 38, 159–167. https://doi.org/10.1080/07011784.2013.787181

Delavau, C., Chun, K.P., Stadnyk, T., Birks, S.J., Welker, J.M., 2015. North American precipitation isotope (δ18O) zones revealed in time series modeling across Canada and northern United States. Water Resources Research 51, 1284–1299. https://doi.org/10.1002/2014WR015687

Delavau, C., Stadnyk, T., Holmes, T., 2017. Examining the impacts of precipitation isotope input (δ18Oppt) on distributed, tracer-aided hydrological modelling. Hydrology and Earth System Sciences 21, 2595–2614. https://doi.org/10.5194/hess-21-2595-2017

Duethmann, D., Blöschl, G., Parajka, J., 2020. Why does a conceptual hydrological model fail to correctly predict discharge changes in response to climate change? Hydrology and Earth System Sciences 24, 3493–3511. https://doi.org/10.5194/hess-24-3493-2020

Efstratiadis, A., Koutsoyiannis, D., 2010. One decade of multi-objective calibration approaches in hydrological modelling: a review. Hydrological Sciences Journal–Journal Des Sciences Hydrologiques 55, 58–78.

Environment and Climate Change Canada, 2020. Historical climate data [WWW Document]. URL https://climate.weather.gc.ca/historical_data/search_historic_data_e.html (accessed 5.26.21).

Environment and Climate Change Canada, 2018. Water Survey of Canada: Historical hydrometric data [WWW Document]. URL https://wateroffice.ec.gc.ca (accessed 5.26.21).

Fatichi, S., Vivoni, E.R., Ogden, F.L., Ivanov, V.Y., Mirus, B., Gochis, D., Downer, C.W., Camporese, M., Davison, J.H., Ebel, B., Jones, N., Kim, J., Mascaro, G., Niswonger, R., Restrepo, P., Rigon, R., Shen, C., Sulis, M., Tarboton, D., 2016. An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. Journal of Hydrology 537, 45–60. https://doi.org/10.1016/j.jhydrol.2016.03.026

Gibson, J.J., Holmes, T., Stadnyk, T.A., Birks, S.J., Eby, P., Pietroniro, A., 2020. 18O and 2H in streamflow across Canada. Journal of Hydrology: Regional Studies. https://doi.org/10.1016/j.ejrh.2020.100754

Gibson, J.J., Yi, Y., Birks, S.J., 2019. Isotopic tracing of hydrologic drivers including permafrost thaw status for lakes across Northeastern Alberta, Canada: A 16-year, 50-lake assessment. Journal of Hydrology: Regional Studies 26. https://doi.org/10.1016/j.ejrh.2019.100643

Gibson, J.J., Yi, Y., Birks, S.J., 2016. Isotope-based partitioning of streamflow in the oil sands region, northern Alberta: Towards a monitoring strategy for assessing flow sources and water quality controls. Journal of Hydrology: Regional Studies 5, 131–148. https://doi.org/10.1016/j.ejrh.2015.12.062

Gupta, H. v., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of Hydrology 377, 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

He, Z., Unger-Shayesteh, K., Vorogushyn, S., Weise, S.M., Kalashnikova, O., Gafurov, A., Duethmann, D., Barandun, M., Merz, B., 2019. Constraining hydrological model parameters using water isotopic compositions in a glacierized basin, Central Asia. Journal of Hydrology 571, 332–348. https://doi.org/10.1016/j.jhydrol.2019.01.048

Holmes, T., 2016. isoWATFLOOD Stable water isotope simulation in the WATFLOOD hydrologic model.

Holmes, T., Stadnyk, T.A., Kim, S.J., Asadzadeh, M., 2020. Regional Calibration With Isotope Tracers Using a Spatially Distributed Model: A Comparison of Methods. Water Resources Research 56. https://doi.org/10.1029/2020WR027447

Holmes, T.L., Stadnyk, T.A., Asadzadeh, M., Gibson, J.J., 2022. Variability in flow and tracer-based performance metric sensitivities reveal regional differences in dominant hydrological processes across the Athabasca River basin. Journal of Hydrology: Regional Studies 41, 101088. https://doi.org/10.1016/j.ejrh.2022.101088

Intsiful, A., Ambinakudige, S., 2021. Glacier cover change assessment of the Columbia Icefield in the Canadian rocky mountains, Canada (1985-2018). Geosciences (Switzerland) 11, 1–9. https://doi.org/10.3390/geosciences11010019

Kiang, J.E., Gazoorian, C., McMillan, H., Coxon, G., Le Coz, J., Westerberg, I.K., Belleville, A., Sevrez, D., Sikorska, A.E., Petersen-Øverleir, A., Reitan, T., Freer, J., Renard, B., Mansanarez, V., Mason, R., 2018. A Comparison of Methods for Streamflow Uncertainty Estimation. Water Resources Research 54, 7149–7176. https://doi.org/10.1029/2018WR022708

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. https://doi.org/10.1029/2005WR004362

Knoben, W.J.M., Freer, J.E., Woods, R.A., 2019. Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores. Hydrology and Earth System Sciences 23, 4323–4331. https://doi.org/10.5194/hess-23-4323-2019

Kouwen, N., 2018. WATFLOOD/WATROUTE Hydrological Model Routing & Flood Foresting System [WWW Document]. URL www.watflood.ca

Matott, L.S., 2017. OSTRICH: an Optimization Software Tool, Documentation and User's Guide, Version 17.12.19.

Matott, L.S., Babendreier, J.E., Purucker, S.T., 2009. Evaluating uncertainty in integrated environmental models: A review of concepts and tools. Water Resources Research 45.

Minder, J.R., Mote, P.W., Lundquist, J.D., 2010. Surface temperature lapse rates over complex terrain: Lessons from the Cascade Mountains. Journal of Geophysical Research 115, D14122. https://doi.org/10.1029/2009JD013493

Mizukami, N., Rakovec, O., Newman, A.J., Clark, M.P., Wood, A.W., Gupta, H. v., Kumar, R., 2019. On the choice of calibration metrics for "high-flow" estimation using hydrologic models. Hydrology and Earth System Sciences 23, 2601–2614. https://doi.org/10.5194/hess-23-2601-2019

Moges, E., Demissie, Y., Larsen, L., Yassin, F., 2020. Review: Sources of Hydrological Model Uncertainties and Advances in Their Analysis. Water (Basel) 13, 28. https://doi.org/10.3390/w13010028

Nan, Y., Tian, L., He, Z., Tian, F., Shao, L., 2021. The value of water isotope data on improving process understanding in a glacierized catchment on the Tibetan Plateau. Hydrology and Earth System Sciences 25, 3653–3673. https://doi.org/10.5194/hess-25-3653-2021

Nash, J.E., Sutcliffe, J. v, 1970. River flow forecasting through conceptual models part I—A discussion of principles. Journal of Hydrology 10, 282–290.

Neill, A.J., Tetzlaff, D., Strachan, N.J.C., Soulsby, C., 2019. To what extent does hydrological connectivity control dynamics of faecal indicator organisms in streams? Initial hypothesis testing using a tracer-aided model. Journal of Hydrology 570, 423–435. https://doi.org/10.1016/j.jhydrol.2018.12.066

Oshun, J., Dietrich, W.E., Dawson, T.E., Fung, I., 2016. Dynamic, structured heterogeneity of water isotopes inside hillslopes. Water Resources Research 52, 164–189. https://doi.org/10.1002/2015WR017485

Pechlivanidis, I.G., Jackson, B.M., Mcintyre, N.R., Wheater, H.S., 2011. Catchment scale hydrological modelling: A review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications. Global NEST journal 13, 193–214.

Peralta-Tapia, A., Sponseller, R.A., Tetzlaff, D., Soulsby, C., Laudon, H., 2015. Connecting precipitation inputs and soil flow pathways to stream water in contrasting boreal catchments. Hydrological Processes 29, 3546–3555. https://doi.org/10.1002/hyp.10300

Piovano, T.I., Tetzlaff, D., Maneta, M., Buttle, J.M., Carey, S.K., Laudon, H., McNamara, J., Soulsby, C., 2020. Contrasting storage-flux-age interactions revealed by catchment inter-

comparison using a tracer-aided runoff model. Journal of Hydrology 590, 125226. https://doi.org/10.1016/j.jhydrol.2020.125226

Razavi, S., Gupta, H. v., 2016. A new framework for comprehensive, robust, and efficient global sensitivity analysis: 1. Theory. Water Resources Research 52, 423–439. https://doi.org/10.1002/2015WR017558

Razavi, S., Sheikholeslami, R., Gupta, H. v, Haghnegahdar, A., 2019. VARS-TOOL: A toolbox for comprehensive, efficient, and robust sensitivity and uncertainty analysis. Environmental Modelling & Software 112, 95–107. https://doi.org/https://doi.org/10.1016/j.envsoft.2018.10.005

Rosa, L., Davis, K.F., Rulli, M.C., D'Odorico, P., 2017. Environmental consequences of oil production from oil sands. Earth's Future 5, 158–170. https://doi.org/https://doi.org/10.1002/2016EF000484

Shafii, M., Tolson, B.A., 2015. Optimizing hydrological consistency by incorporating hydrological signatures into model calibration objectives. Water Resources Research 51, 3796–3814. https://doi.org/10.1002/2014WR016520

Shangguan, W., Dai, Y., Duan, Q., Liu, B., Yuan, H., 2014. A global soil data set for earth system modeling. Journal of Advances in Modeling Earth Systems 6. https://doi.org/10.1002/2013MS000293

Song, X., Zhang, J., Zhan, C., Xuan, Y., Ye, M., Xu, C., 2015. Global sensitivity analysis in hydrological modeling: Review of concepts, methods, theoretical framework, and applications. Journal of Hydrology. https://doi.org/10.1016/j.jhydrol.2015.02.013

Spangenberg, J.E., 2012. Caution on the storage of waters and aqueous solutions in plastic containers for hydrogen and oxygen stable isotope analysis. Rapid Communications in Mass Spectrometry 26. https://doi.org/10.1002/rcm.6386

Stadnyk, T.A., Holmes, T.L., 2020. On the value of isotope-enabled hydrological model calibration. Hydrological Sciences Journal 65, 1525–1538. https://doi.org/10.1080/02626667.2020.1751847

Stevenson, J.L., Birkel, C., Neill, A.J., Tetzlaff, D., Soulsby, C., 2021. Effects of streamflow isotope sampling strategies on the calibration of a tracer-aided rainfall-runoff model. Hydrological Processes 35. https://doi.org/10.1002/hyp.14223

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

Tunaley, C., Tetzlaff, D., Birkel, C., Soulsby, C., 2017. Using high-resolution isotope data and alternative calibration strategies for a tracer-aided runoff model in a nested catchment. Hydrological Processes 31, 3962–3978. https://doi.org/10.1002/hyp.11313

Uusitalo, L., Lehikoinen, A., Helle, I., Myrberg, K., 2015. An overview of methods to evaluate uncertainty of deterministic models in decision support. Environmental Modelling & Software 63, 24–31.

Viglione, A., Parajka, J., Rogger, M., Salinas, J.L., Laaha, G., Sivapalan, M., Blöschl, G., 2013. Comparative assessment of predictions in ungauged basins - Part 3: Runoff signatures in Austria. Hydrology and Earth System Sciences 17, 2263–2279. https://doi.org/10.5194/hess-17-2263-2013

Vitt, D.H., Halsey, L.A., Zoltai, S.C., 2000. The changing landscape of Canada's western boreal forest: the current dynamics of permafrost. Canadian Journal of Forest Research 30, 283–287.

Westerberg, I.K., Sikorska-Senoner, A.E., Viviroli, D., Vis, M., Seibert, J., 2020. Hydrological model calibration with uncertain discharge data. Hydrological Sciences Journal 1–16. https://doi.org/10.1080/02626667.2020.1735638

Xiong, L., Wan, M., Wei, X., O'Connor, K.M., 2009. Indices for assessing the prediction bounds of hydrological models and application by generalised likelihood uncertainty estimation. Hydrological Sciences Journal 54, 852–871. https://doi.org/10.1623/hysj.54.5.852

6. Conclusions

6.1. Summary and major findings

The overall aim of this research was to establish guidance for the incorporation of isotope tracer data into hydrologic model calibration, in order to maximize the benefits of isotope-enabled simulations for large scale watershed models. The specific objectives of study were to identify parameter sensitivities and parameter value identifiability for isotope tracer simulations (O1), evaluate differences in simulated process fluxes, internal storage and streamflow between models calibrated with and without isotope tracer data (O2), and develop specific recommendations for tracer-aided calibration objectives and isotope simulation performance metrics to maximize flux simulation benefits (O3). The results presented in this thesis have been organized into three manuscripts, each of which has contributed to addressing multiple objectives, as outlined in Figure 6.1.

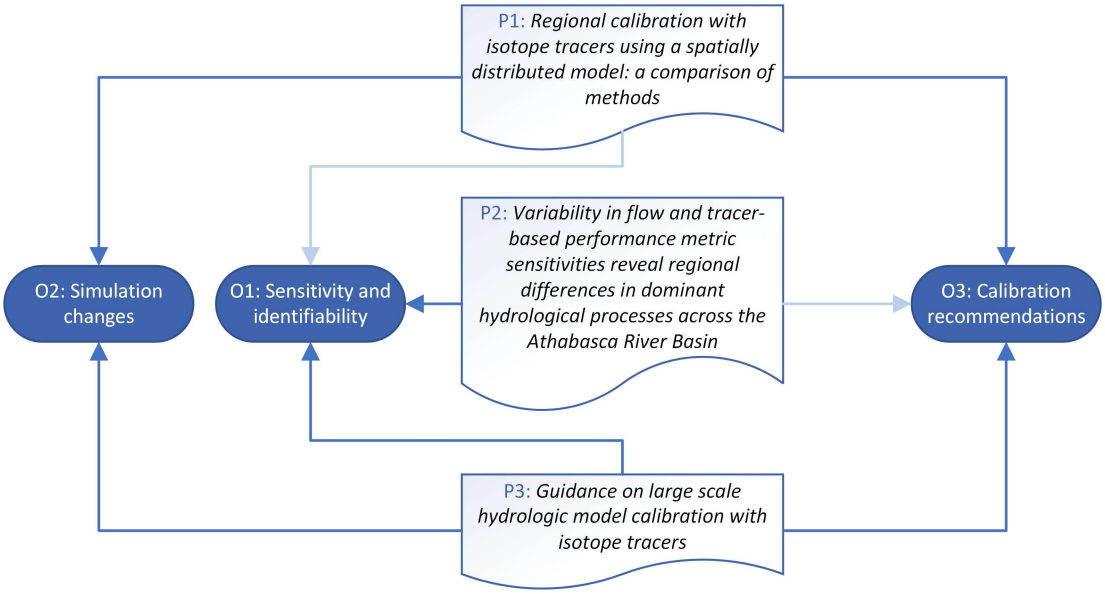

Figure 6-1: Summary of the contributions of each component paper (P1 to P3) to the three research objectives (O1 to O3).

The first paper (Chapter 3, P1) presents a comparison of calibration objective formulations for tracer-aided model calibration. This peer-reviewed journal publication assesses changes in parameter values and modeled hydrologic fluxes for different methods of incorporating isotope data into calibration and provides a recommendation for tracer-aided objective choice. This

chapter contributes to addressing all three research objectives and achieves the first part of Objective 3. In the second paper (Chapter 4, P2), parameter sensitivities for isotope tracer simulation performance metrics are examined. This chapter compares the sensitivities of flow and isotope tracer metrics, examines their temporal and spatial variations, and assesses the link between parameter and process sensitivities. The research in this peer-reviewed journal publication completes the first part of Objective 1 and contributes toward Objective 3. A comparison of isotope simulation performance metrics is presented in the third manuscript (Chapter 5, P3), focused on parameter identifiability and simulations resulting from different metrics as calibration objectives. The results in this chapter are used to support a recommendation for metric choice in tracer-aided hydrologic model calibration, achieving Objective 3, while the analysis of simulation outputs and parameter identifiability finish addressing Objectives 1 and 2.

Sensitivity and identifiability

Objective 1: Understand the sensitivity and identifiability of parameters for isotope tracer simulations and compare them to those for streamflow simulations.

Isotope simulation performance metrics are sensitive to more parameters and processes than flow simulation performance metrics, but these sensitivities are frequently less reliable than flow performance metric sensitivities. Flow metric sensitivities are only sensitive for a sub-set of processes, unless data are broken up and evaluated seasonally. Using flow signatures does not increase the number of sensitive processes compared to using only flow simulation KGE. In contrast, of the isotope metrics evaluated, at least one is somewhat sensitive to all simulated hydrologic processes in the isoWATFLOOD model. Generally, flow metrics are sensitive to processes determining streamflow timing and total volume (i.e., melt, evapotranspiration and baseflow) and isotope metrics are sensitive to processes determining mixing volumes and water age, and fractionating processes (i.e., evaporation, soil storage and flux rates).

Identifiability results were inconclusive, with some differing results depending on the study basin and methods. In the Odei River study, parameter values were generally best identified using a single flow performance metric as a calibration objective; using isotope data in the calibration objective resulted in different identified parameter values for soil parameters, with

somewhat larger ranges of possible parameter values and simulated process fluxes. In the Athabasca River study, there were no major difference in parameter value identified between multi-objective flow-only calibrations and mixed isotope-flow calibration, and minor improvements in identifiability for some parameters when isotope data is included in calibration. Differences in identifiability and identified parameter values were most notable in parameters related to soil storage, evaporation and soil flux rates in both studies.

Simulation changes

Objective 2: Examine differences in simulated process fluxes, internal storage and streamflow between models calibrated with and without isotope tracer data.

Process simulations (i.e., evaporation, transpiration and soil moisture retention) resulting from isotope supplemented calibrations in the Odei River watershed are in closer agreement with other estimates of water storage and loss processes than those from flow-only calibrations. Adding isotope data to model calibration improves confidence in streamflow component simulation and some process simulations in the calibration of the Athabasca River model. The streamflow simulations in both basins were similar in the calibration periods, but when evaluated as an ensemble, calibrating using both isotope tracer and flow data outperformed an ensemble calibrated with flow data alone. Furthermore, ensembles calibrated with isotope data have substantially improved simulation of peak flow in validation periods; particularly when validation periods include flow rates exceeding those previously seen in the historic record.

Calibration recommendations

Objective 3: Develop specific recommendations for tracer-aided calibration objectives and isotope simulation performance metrics to maximize flux simulation benefits for hydrologic models.

Multi-objective optimization was found to be a superior method for tracer-aided calibration in large-scale watersheds compared to the other commonly used methods. Using a Monte Carlo analysis alone in the calibration or evaluation of a complex process-based isotope-enabled model is a poor use of computational resources; solutions of acceptable quality can be identified far more efficiently using optimization algorithms. Multi-objective optimization using a mix of

isotope and flow performance metrics produced solutions of equivalent quality to single objective calibrations with isotope performance included in the objective. The negligible additional computational cost of multi-objective optimization for large-scale models is far outweighed by the production of a range of additional solutions along the tradeoff front between isotope and flow simulation performance, and the use of multi-objective rather than single objective optimization allows for flexibility on the relative weights given to isotope and flow performance in determining final parameter selection.

Including isotope simulation performance in calibration can add valuable information and improve both the flow and isotope simulation performance. In both the Odei and Athabasca River models, flow simulation and isotope simulation performance in calibration are not strongly correlated, but substantial improvements in the quality of the isotope simulation are possible at slight cost to flow simulation performance in calibration. Isotope simulation performance in calibration is closely associated with flow simulation performance in validation and is a better predictor of validation performance than flow performance in calibration. Furthermore, parameter set ensembles calibrated using isotope tracer data outperform an ensemble calibrated with flow data alone. Mixed isotope-flow hydrologic model calibrations are beneficial for flow simulations, however, calibrating a hydrologic model to isotope data alone does not significantly improve flow simulation quality compared to random parameter values within a process-based model. The two isotope simulations have highly correlated performance in calibration, indicating little added value in calibrating to an additional isotope tracer. The greatest benefits to flow simulations in the Athabasca basin came from calibrating to oxygen-18 performance as a secondary objective, but the differences in performance between isotopes were slight and no definite recommendation between isotope tracers can be made given the magnitude of the uncertainties in the results.

An isotope performance metric should be used as an objective in an isotope-enabled calibration. While sensitivity data alone is insufficient for selecting the best metric, as many isotope performance metrics have indistinguishable sensitivities when reliability is considered, GSA results showed that isotope performance metrics can improve the calibration of some modeled processes. Isotope performance should be considered in calibration when process simulation matters, as flow metrics do not respond to changes in some soil-related processes. The best

ensemble flow simulation performance was produced by mixed isotope-flow optimizations using isotope metrics which includes timing error (e.g., NRMSE or KGE), rather than flow-only optimization or mixed optimization using an isotope signature metric.

Conclusions

Isotope tracer simulations are more sensitive to parameters relating to soil water fluxes and storages than streamflow simulations, but calibrating isotope simulations does not significantly improve individual parameter identifiability. However, isotope-aided calibration improves process and streamflow component identifiability, with some modest benefits to streamflow simulations. Multi-objective optimization using an isotope simulation performance metric which includes timing error as a secondary calibration objective is recommended in order to maximize the benefits of isotope-aided hydrologic model calibration for large-scale watershed models.

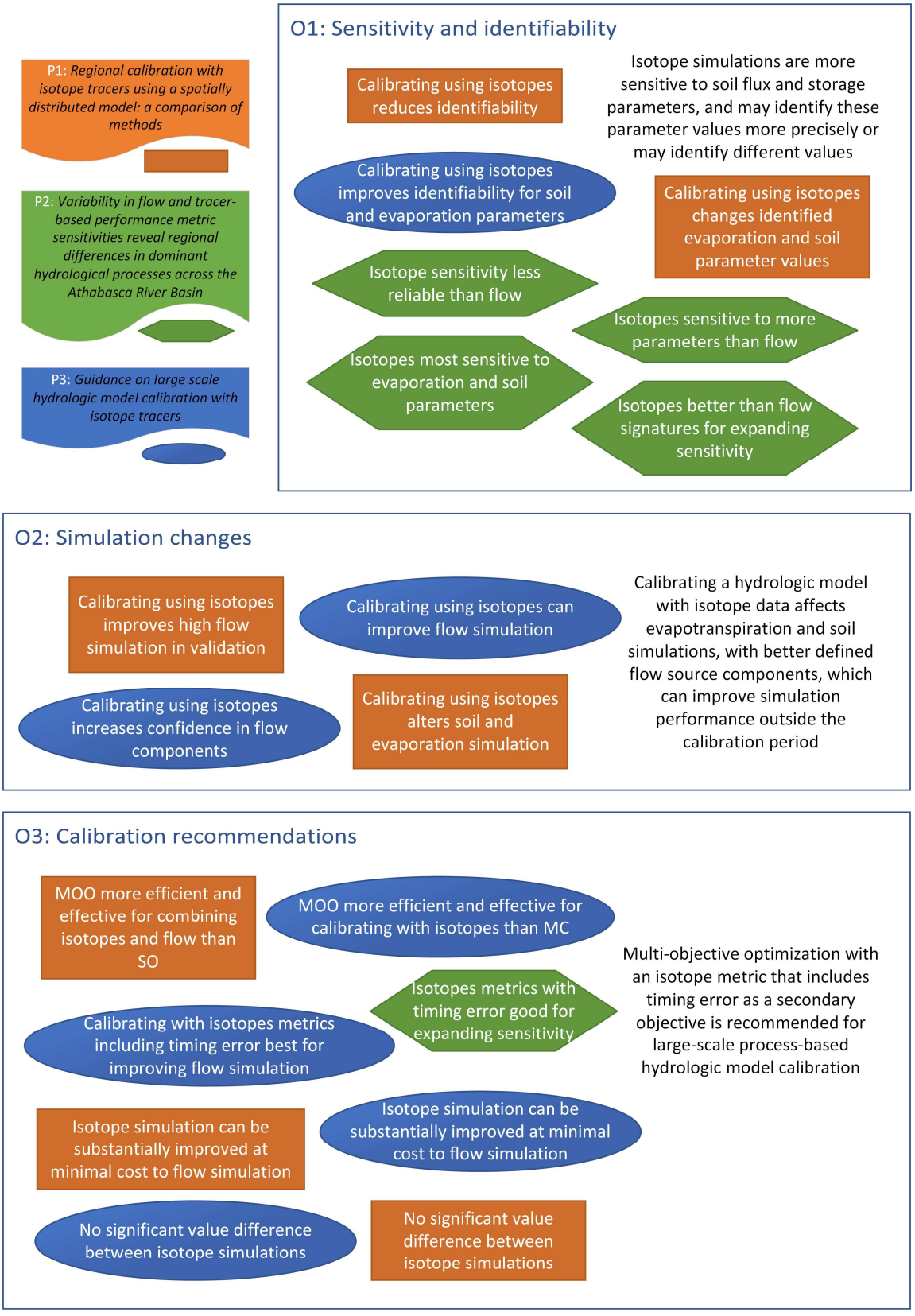

Figure 6-2: Summary of results and conclusions for each research objective

6.2. Future work

This research has been focused on improving hydrologic modeling at the large scale using currently available isotope tracer data, and more specifically on identifying the best methods of incorporating isotope data into model calibration. The work presented here has met the specific research objectives of the thesis, providing evidence-based recommendations for tracer-aided calibration and advancing understanding of the value isotope tracers add to process-based hydrologic modeling at the meso- or large-scale. These results highlight a path forward for the application of isotope tracers in the general practice of hydrology, as the identification of flow sources at large scale and prediction of extreme flow events of immediate and urgent interest to decision-makers. The work also highlights some remaining research gaps in tracer-aided modeling and calibration, and future work in isotope hydrology.

The most significant limitation of this study, and therefore most important opportunity for further research, is the unexamined impact of isotope data sparsity on model calibration. The scope of this research was limited to previously collected isotope data sets for large Canadian watersheds, which have sampling frequencies rarely exceeding monthly, and which often have multi-month gaps and short periods of record. This scope limitation expands the applicability of the recommendations produced from the research, as this type of data set is relatively common and increasingly available. However, the use of exclusively sparse isotope data sets means that it is as yet unknown whether the relatively undramatic improvements from tracer-aided modeling, as compared to those seen in some small-scale studies, are due to the scale of the hydrologic model or the limited isotope data. A high-quality isotope tracer time-series for a meso-scale watershed is needed to disentangle the significance of data sparsity to tracer-aided model calibration. Ideally, there would be a large-scale basin with a complete long-term isotope data record in streamflow at multiple locations; at present, even the best isotope data sets in large basins are short term (rarely over ten years), sparse (approximately monthly sampling, limited numbers of sites) and irregular (lower sampling frequency in winter, sites concentrated in more accessible regions, gaps from intermittent staffing or funding). In the longer term, more data should be collected in a basin with a pre-existing isotope sampling program but there are also opportunities now for experimental studies on the impact of data sparsity on model calibration. Rather than testing calibration on real-world watersheds, initial evaluations could be performed on virtual watersheds, or simBasins. In place of physical field studies and data collection, a

process-based watershed model could be used as the research 'basin'; this would allow for the collection of decades of 'observed' data in minutes, at negligible cost. With a simBasin study site (i.e., a process-based model set up with realistic input data and producing plausible outputs) any number of sampling strategies can be tested, by setting up simSamplers to collect data from the simBasin (e.g., daily sampling at the outlet, weekly samples at limited sub-basins, monthly sampling basin-wide, high-frequency summer-sampling only) and calibrating models of the simBasin using only the collected data. These experiments would not provide new information on real-world watershed function, rather they would point to sampling programs which actual deliver advertised model improvements; in addition to isotope tracers, other hydrologically relevant data (e.g., snowpack levels, vapor flux) could be similarly evaluated.

The scope of this research was also restricted to applications of the isoWATFLOOD^{TM} model, the only large-scale processed-based isotope-enabled hydrologic model available when this study was initiated. With increased interest in isotope tracer simulations embedded in hydrologic models, there are valuable opportunities to extend both the type and scope of future tracer-aided modeling. Possibilities include assessments of model structural uncertainty from multi-model ensemble studies, and tracing freshwater contributions through marine systems with linked isotope-enabled terrestrial and marine system models.

This research has focused on meso- to large-scale watersheds, but the largest studied basin area was less than 200,000 km^2. There is no reason, however, that larger basins should not be examined using isotope-enabled hydrologic model. The Mackenzie River basin, for example, has a pre-existing CHARM model which can be adapted for tracer modeling. With an area over 1,750,000 km^2, an isotope-enabled model of the entire Mackenzie River watershed would permit a comprehensive examination of the effect of upstream area on the informational content of isotope time-series data, as increasing area reduces isotope variabilities.

A temporal extension is arguably more important than a spatial extension of isotope-calibrated hydrologic models. The preliminary application of tracer-aided models to modeling future hydrology using various climate models has indicated the possibility that the small changes in historical process simulation due to isotope calibration can lead to substantial changes with more extreme weather inputs. The differences in simulations under climate between isotope-aided models and those calibrated with only flow data deserves further study.

Prior to using the isoWATFLOOD model for climate change studies at high latitudes, such as pan-Arctic regions, the role and contributions of paleo-water from permafrost and glacier thaw, and regional groundwater systems must be examined. The isotope tracer signals from these old water sources can be distinctive from current precipitation inputs and their existing representation within the isotope model is either unverified or non-existent. If paleo-water fluxes are substantial enough to be detectable in rivers, the isoWATFLOOD should be expanded to include these water sources.

The results of this research have indicated that simulated water storage and mixing volumes are most likely to be affected by the inclusion of isotope data in calibration. An important further step in meso-scale isotope-enabled watershed modeling is a study of water age within a hydrologic model and the effect of isotope-enabled calibration on simulated transit times. This further research could have a direct bearing on contaminant transport in watersheds such as the Athabasca River basin.

The final, and most important, recommendation for future work is to continue isotope sampling of river water in conjunction with hydrometric data collection. This research has demonstrated that even sparse, short isotope datasets can add value to meso-scale hydrologic modeling. By maintaining regular sampling of stream water, the utility of isotope tracers will improve as the time-series extends. An expanded sampling program would extend the potential benefits more widely; more intensive sampling can increase the potential applications of the data and capture more information for smaller watersheds. The benefits of prior investment in isotope tracer sampling can be multiplied by continued sampling at the same locations. The improvements to process-based meso-scale hydrologic modeling identified in this research depend on isotope tracer sampling.

Appendix A: Canadian national isotope sampling program

Maps created by T. Holmes for:

Gibson, J.J., Holmes, T., Stadnyk, T.A., Birks, S.J., Eby, P., Pietroniro, A., 2020. 18O and 2H in streamflow across Canada. Journal of Hydrology: Regional Studies. https://doi.org/10.1016/j.ejrh.2020.100754

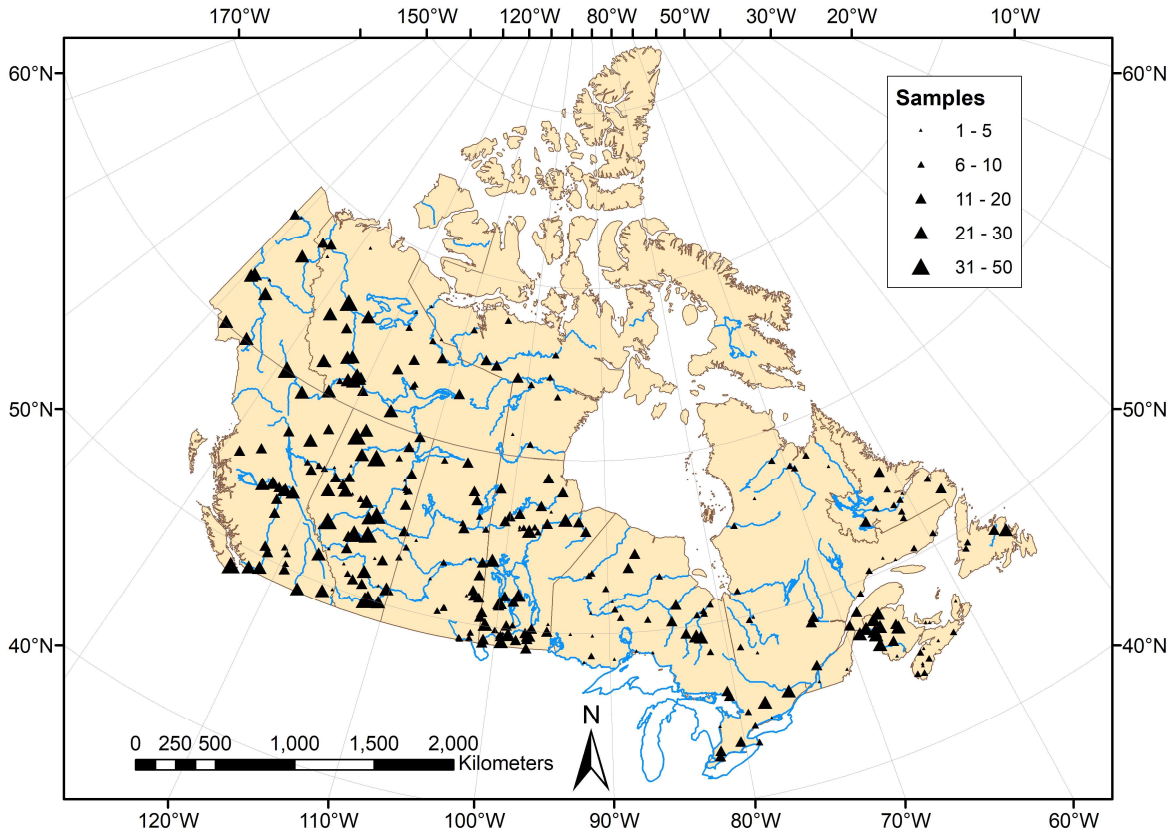

Figure A-1: Map of Canada showing Water Survey of Canada river isotope monitoring network sites, 2013-2019. The size of station symbol is proportionate to the number of samples.

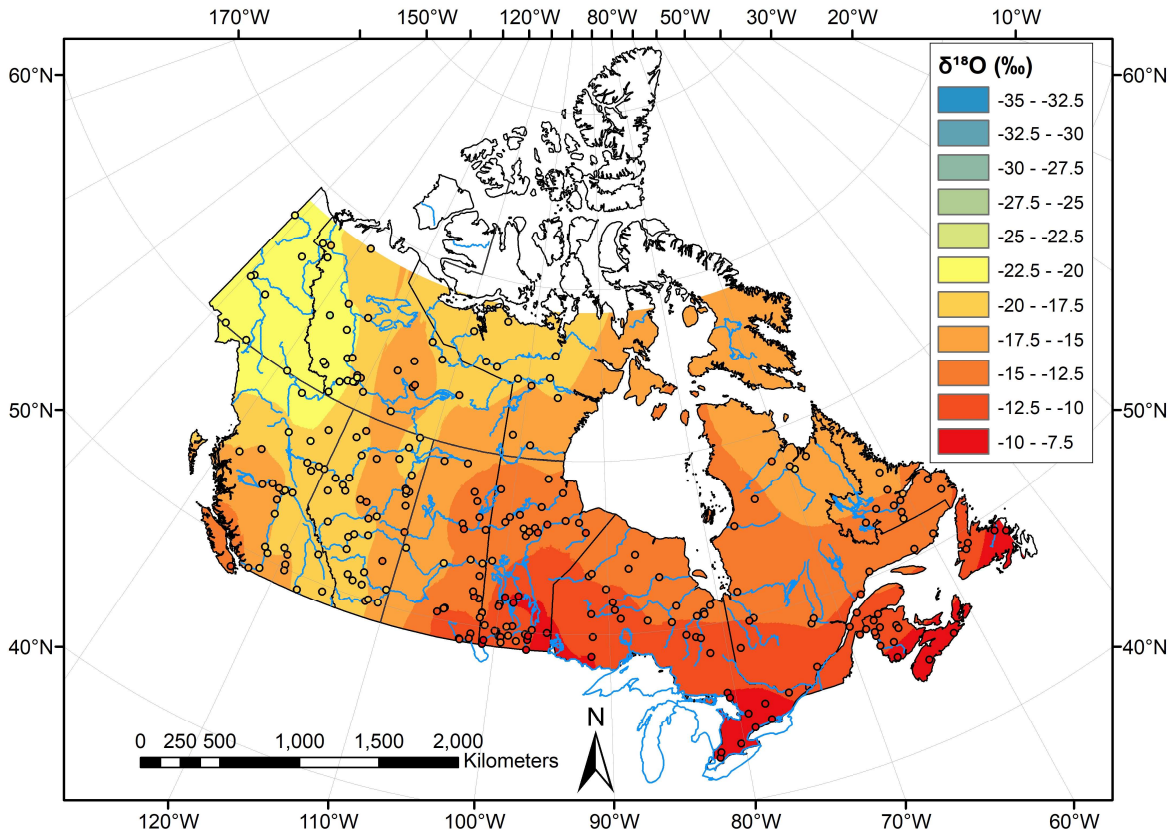

Figure A-2: Spatial distributions of mean $\delta^{18}O$ (‰) in streamflow from flow-weighted samples (2013- 2019). Closed circles indicate locations of gauging stations used in this analysis.

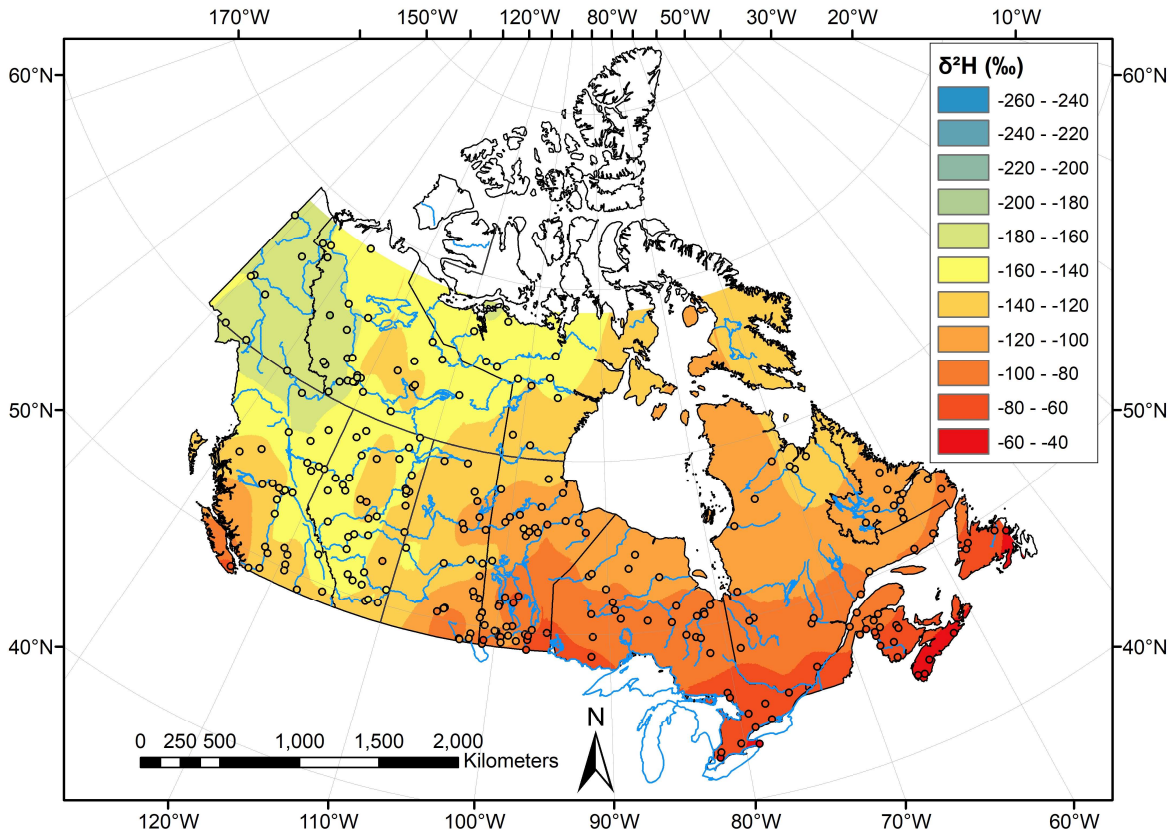

Figure A-3: Spatial distributions of mean δ²H (‰) in streamflow from flow-weighted samples (2013- 2019). Closed circles indicate locations of gauging stations used in this analysis.

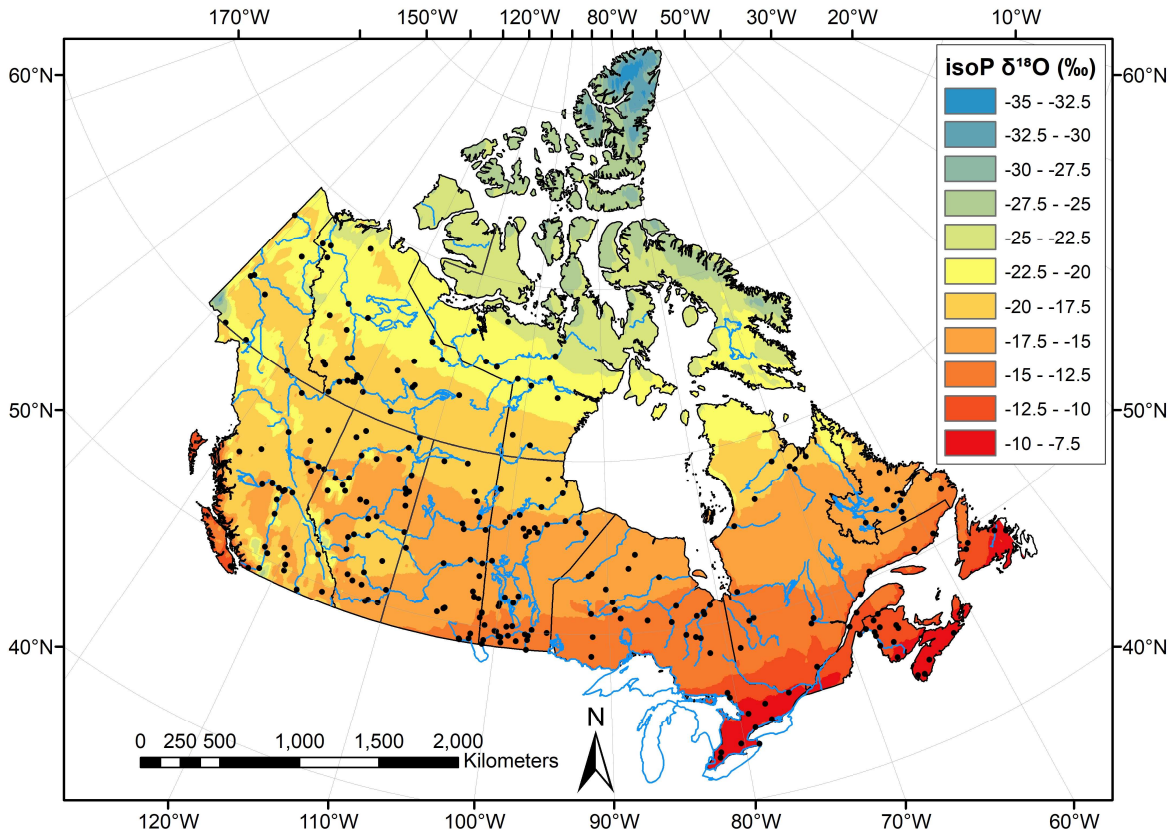

Figure A-4: Spatial distributions of modeled mean $\delta^{18}O$ (‰) in annual precipitation. Closed circles indicate locations of sampled hydrometric gauging stations.

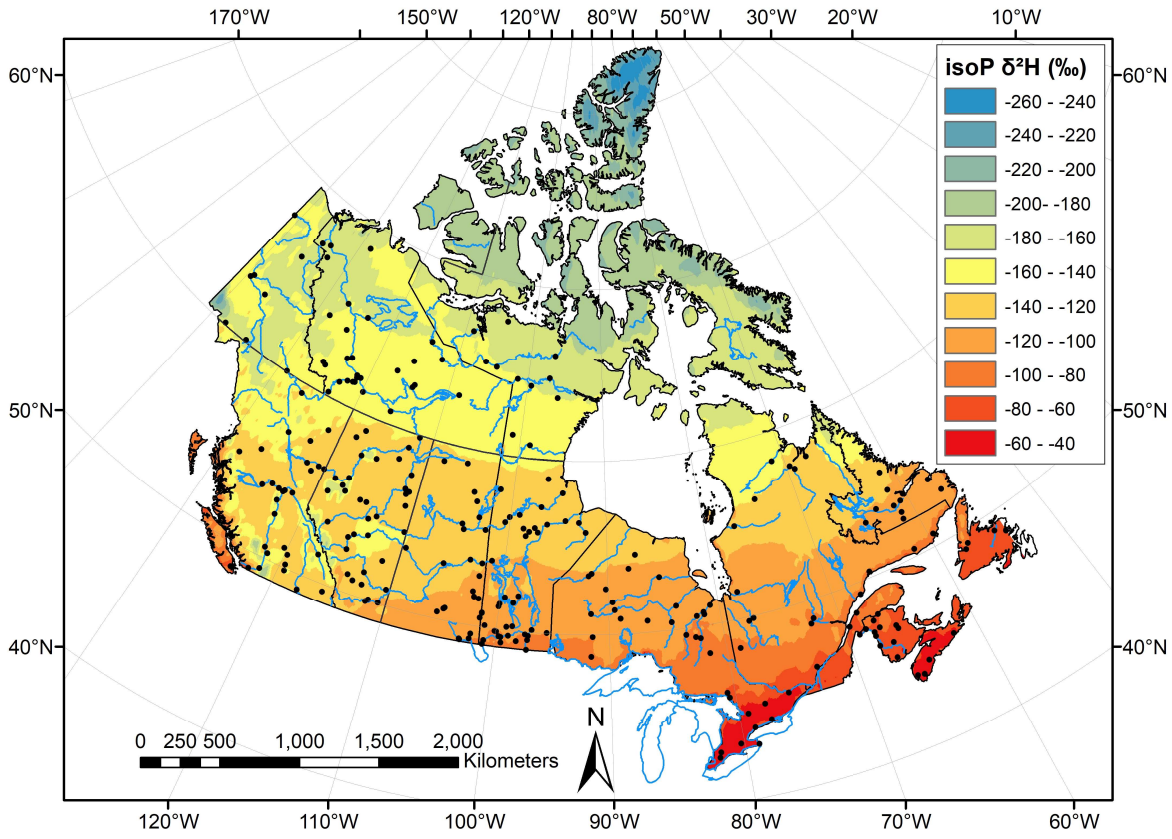

Figure A-5: Spatial distributions of modeled mean δ^2H (‰) in annual precipitation. Closed circles indicate locations of sampled hydrometric gauging stations.

Appendix B: Supporting information for Chapter 3

Auxiliary material from:

Holmes, T., Stadnyk, T.A., Kim, S.J., Asadzadeh, M., 2020. Regional Calibration With Isotope Tracers Using a Spatially Distributed Model: A Comparison of Methods. Water Resources Research 56. https://doi.org/10.1029/2020WR027447

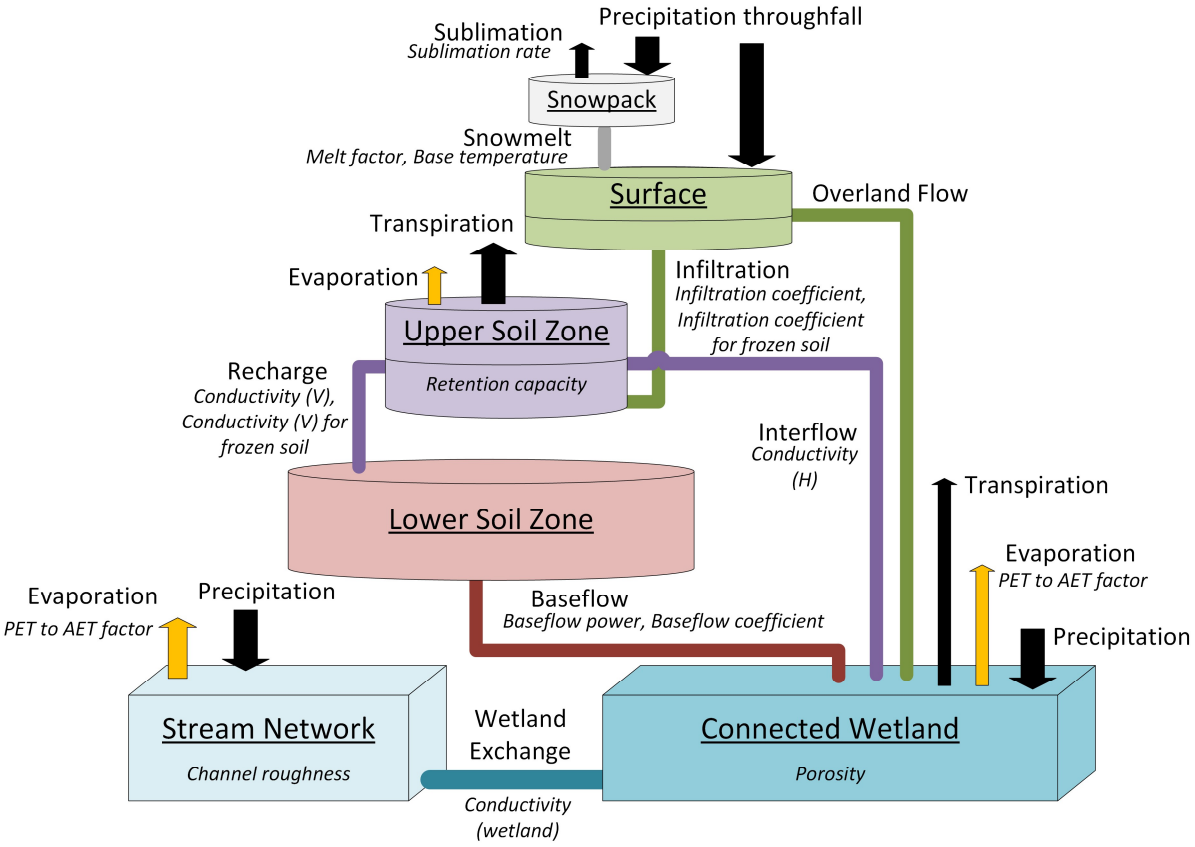

Figure B1. IsoWATFLOOD flux and storage schematic, with calibrated parameters listed in italics. Fractionating model fluxes are shown in orange.

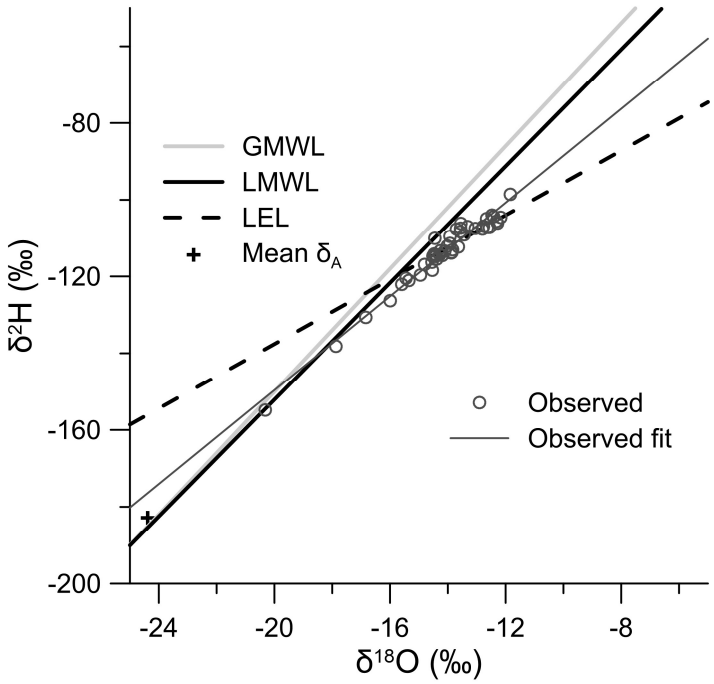

Figure B2. Isotope tracer observations at the Odei River gauge

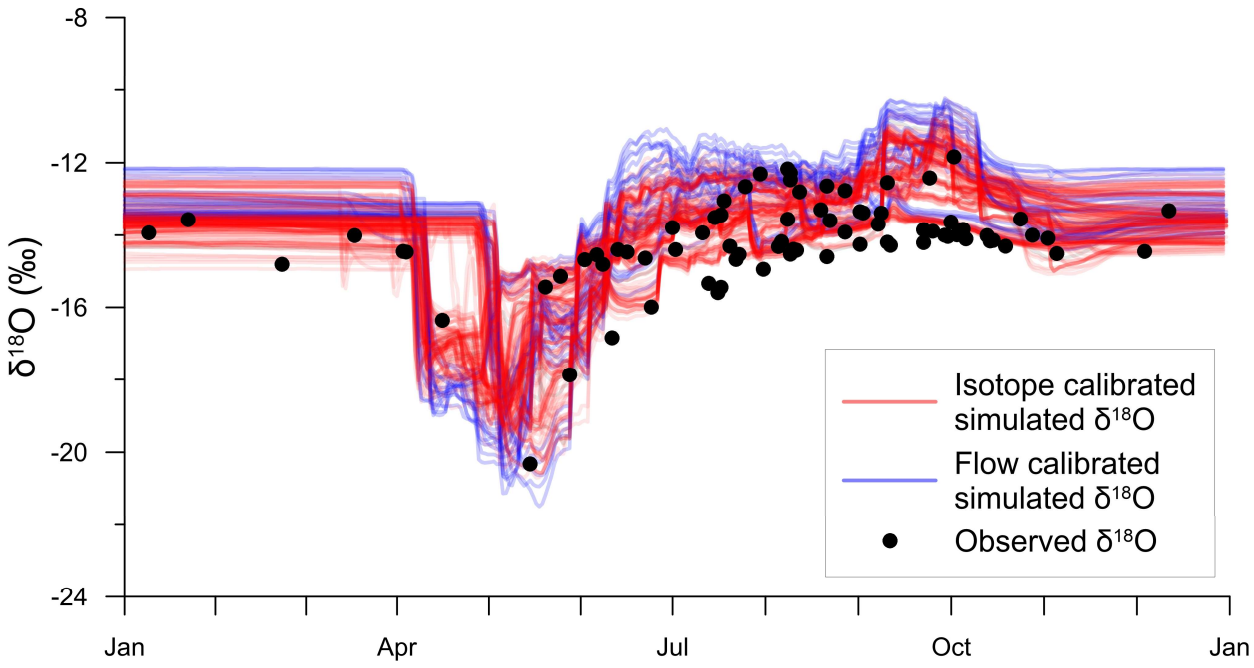

Figure B3. Daily average annual isographs over the calibration period (2009-2015)

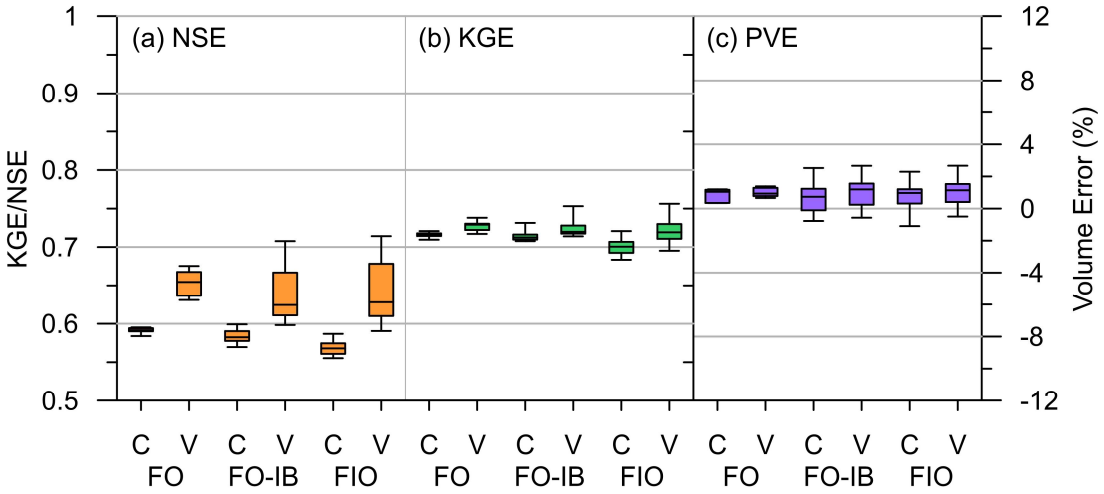

Figure B4. Boxplot comparison of (a) NSE, (b) KGE and (c) PVE for the calibration (C) and validation (V) periods, for the flow-only optimizations (FO), flow optimization with isotope boundaries (FO-IB) and flow and isotope optimizations (FIO) at the Sapochi River gauge, with whiskers extending to maximum and minimum values.

Table B1. Calibrated parameters and parameter bounds

Storage	Parameter	Units	WATFLOOD	OSTRICH	Lower	Upper
Lower zone	Baseflow power	-	pwr	GWpwr	1.5	4
	Baseflow coefficient	-	coeff	Gwcoef	0.0000001	0.005
Channel	Channel roughness	-	r2n	STRr2n	0.0005	0.05
Connected wetland	Porosity	-	theta	WETpore	0.15	0.9
	Conductivity	m/s	kcond	WETk	0.1	0.25
Surface	Infiltration, class 1	mm/h	ak	SW_ak_1	1	50
	Infiltration, class 2	mm/h		SW_ak_2	1	50
	Infiltration, class 3	mm/h		SW_ak_3	1	50
	Infiltration, class 4	mm/h		SW_ak_4	1	200
	Infiltration 1, frozen	mm/h	akfs	akfs1	0.1	50
	Infiltration 2, frozen	mm/h		akfs2	0.1	50
	Infiltration 3, frozen	mm/h		akfs3	0.1	50
	Infiltration 4, frozen	mm/h		akfs4	0.1	200
Upper zone	Retention capacity 1	mm	retn	retn1	20	150
	Retention capacity 2	mm		retn2	20	150
	Retention capacity 3	mm		retn3	10	100
	Retention capacity 4	mm		retn4	100	500
	Conductivity (H) 1	-	rec	rec1	0.1	4
	Conductivity (H) 2	-		rec2	0.1	4
	Conductivity (H) 3	-		rec3	0.1	4
	Conductivity (H) 4	-		rec4	0.1	4
	Conductivity (V) 1	h^{-1}	ak2	IF_ak2_1	0.001	0.2
	Conductivity (V) 2	h^{-1}		IF_ak2_2	0.001	0.2
	Conductivity (V) 3	h^{-1}		IF_ak2_3	0.001	0.2
	Conductivity (V) 4	h^{-1}		IF_ak2_4	0.001	0.2
	Conductivity, frozen	h^{-1}	ak2fs	IFfs_ak2	0.0001	0.2
Snowpack	Melt factor, class 1	mm/°C/h	fm	fm1	0.03	0.25
	Melt factor, class 2	mm/°C/h		fm2	0.03	0.25
	Melt factor, class 3	mm/°C/h		fm3	0.03	0.25
	Melt factor, wetland	mm/°C/h		fmwet	0.03	0.25
	Melt factor, water	mm/°C/h		fmopen	0.03	0.25
	Base T, class 1	°C	Base	base1	-3	2
	Base T, class 2	°C		base2	-3	2
	Base T, class 3	°C		base3	-3	2
	Base T, wetland	°C		basewet	-3	2
	Base T, water	°C		baseopen	-3	2
	Sublimation rate 1	mm/d	sublim_rate	sublim1	0.05	0.5
	Sublimation rate 2	mm/d		sublim2	0.05	0.5
	Sublimation rate 3	mm/d		sublim3	0.05	0.5
	Sublimation, wetland	mm/d		sublimwet	0.05	0.5
	Sublimation, water	mm/d		sublimopen	0.05	0.5
Open water	PET to AET factor	-	fpet	openEf	0.4	1

Table B2. Estimated transpiration ratio, total transpiration, evaporation, and soil water content from the various model calibrations, in comparison with GLEAM data (Martens et al. 2017; Miralles et al. 2011), across the Odei River basin (2002-2015).

	GLEAM (Range)	FO (Range)	FO-IB (Range)	FIO (Range)
T/total ET	0.68 (0.56-0.79)	0.58 (0.57-0.59)	0.58 (0.55-0.60)	0.60 (0.58-0.62)
T (mm/year)	297.3 (246.3-344.7)	159.0 (156.4-161.9)	159.7 (151.0-163.6)	164.7 (159.4-169.6)
E (mm/year)	12.8 (9.5-16.4)	22.2 (20.5-24.3)	18.4 (17.1-20.9)	18.3 (17.3-18.9)
Soil moisture (m^3/m^3)	0.18 (0.15-0.19)	0.07 (0.06-0.08)	0.09 (0.07-0.13)	0.14 (0.13-0.16)

Appendix C: Supporting information for Chapter 4

Auxiliary material from:

Holmes, T.L., Stadnyk, T.A., Asadzadeh, M., Gibson, J.J., 2022. Variability in flow and tracer-based performance metric sensitivities reveal regional differences in dominant hydrological processes across the Athabasca River basin. Journal of Hydrology: Regional Studies 41, 101088. https://doi.org/10.1016/j.ejrh.2022.101088

Flow and Isotope Data

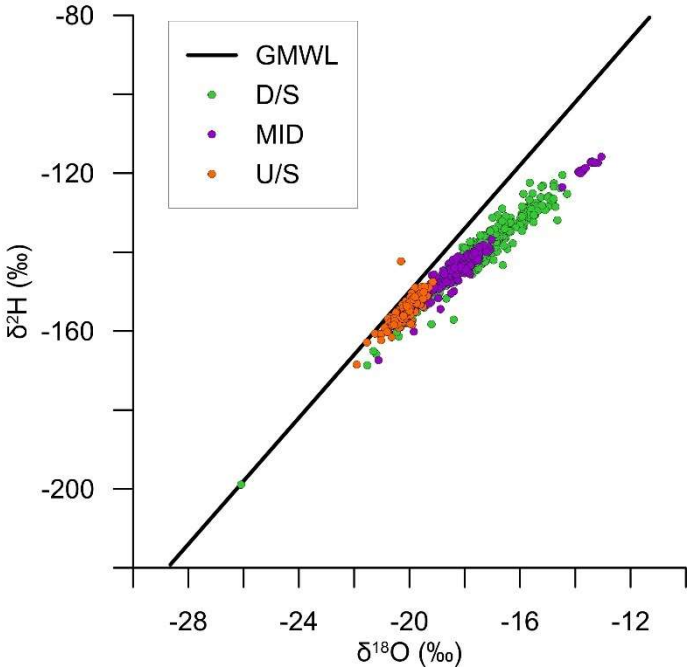

Figure C1: All isotope samples for the downstream (D/S), mid-reach (MID) and upstream (U/S) regions of the Athabasca River watershed used in this study, with the global meteoric water line (GWML) shown for reference.

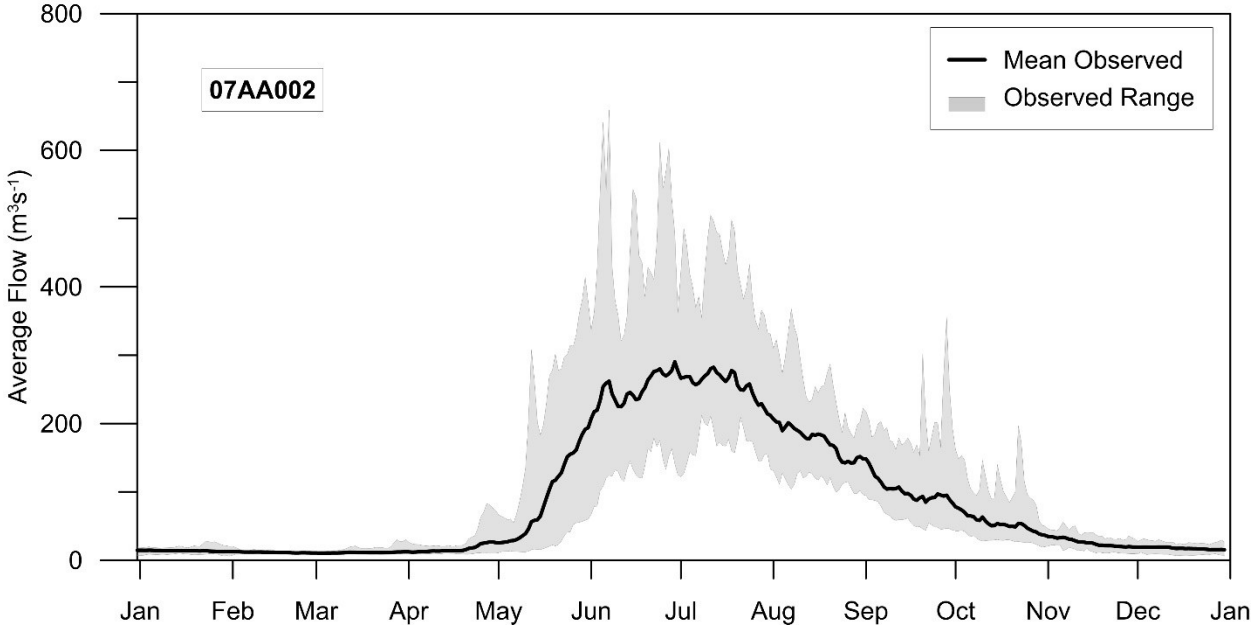

Figure C2: Average annual hydrograph for WSC gauge 07AA002 for 2002-2015, with observed ranges for each day in gray.

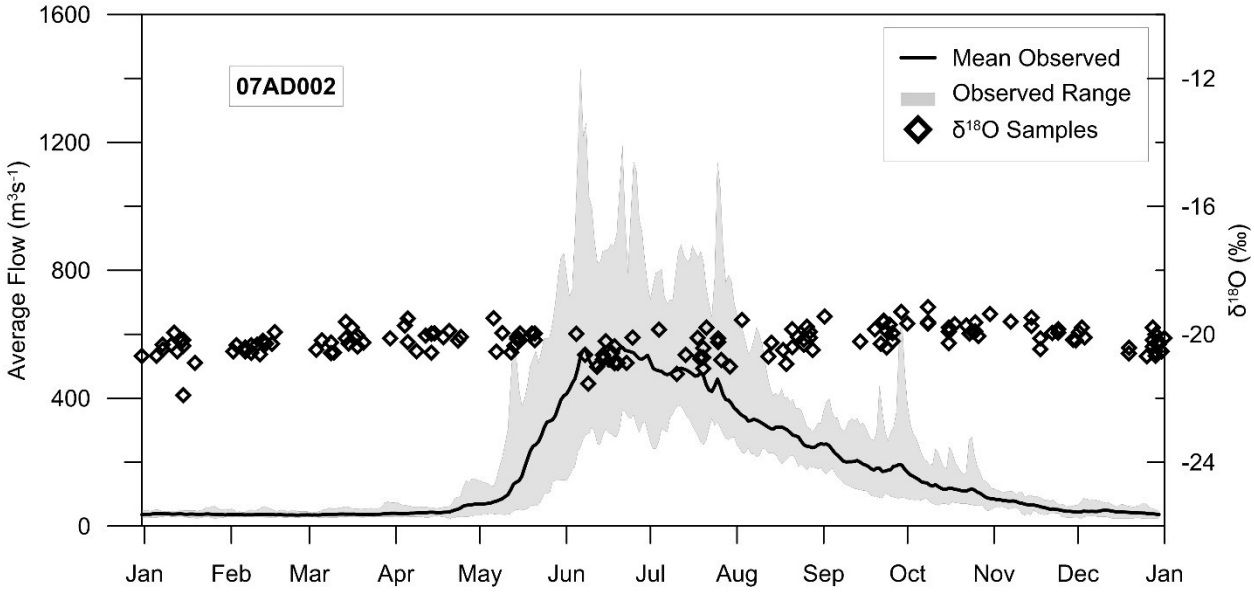

Figure C3: Average annual hydrograph for WSC gauge 07AD002 for 2002-2015, with observed flow ranges for each day in gray. All isotope samples in the same period taken at the site are shown as points.

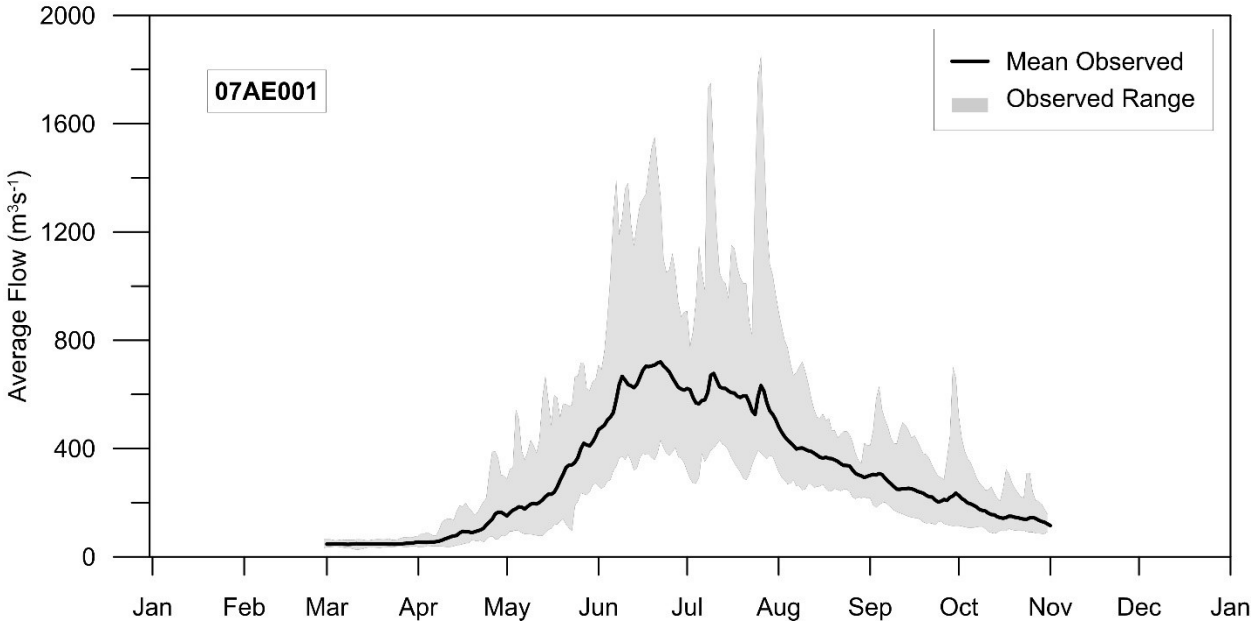

Figure C4: Average annual hydrograph for WSC gauge 07AE001 for 2002-2015, with observed ranges for each day in gray.

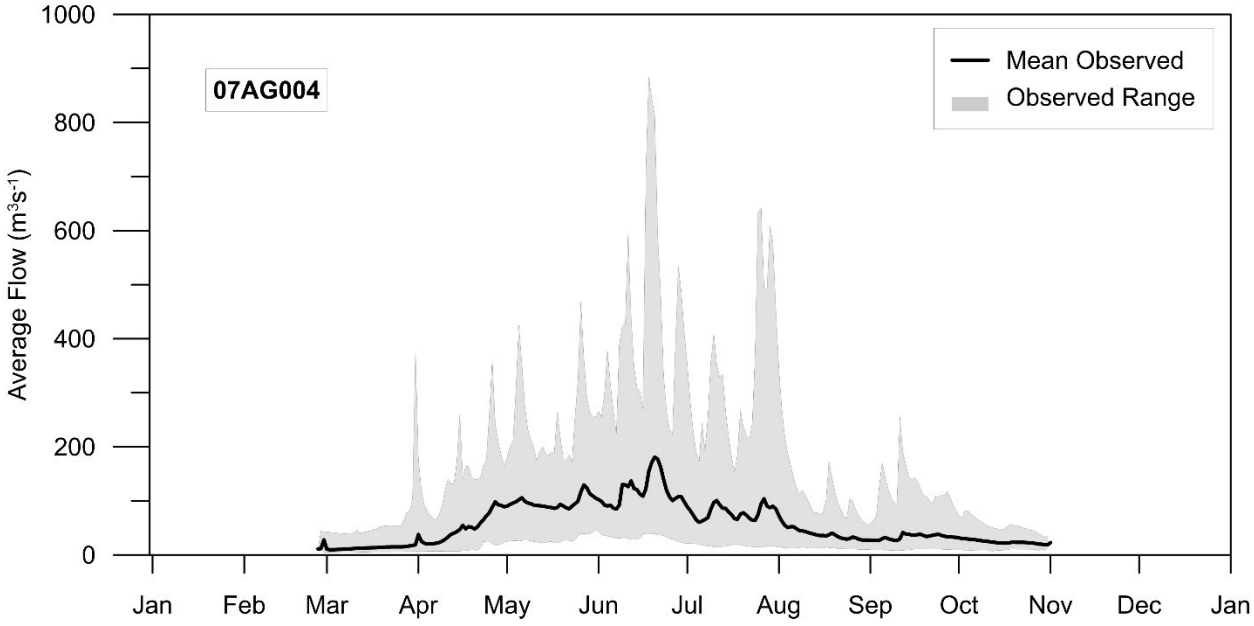

Figure C5: Average annual hydrograph for WSC gauge 07AG004 for 2002-2015, with observed ranges for each day in gray.

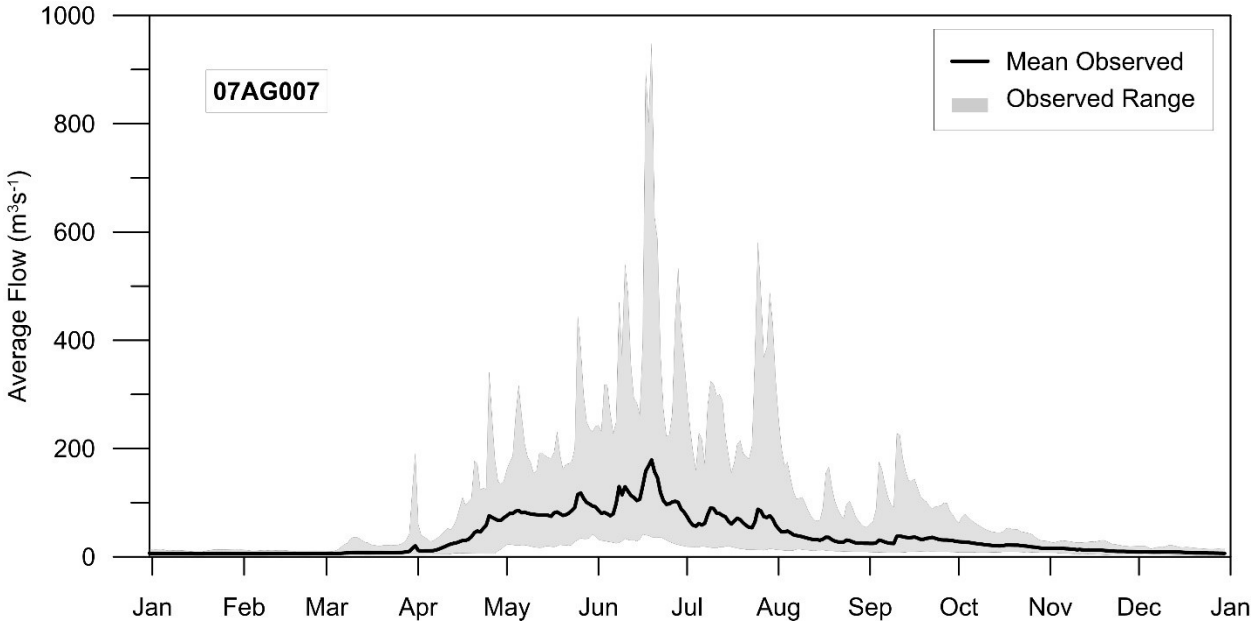

Figure C6: Average annual hydrograph for WSC gauge 07AG007 for 2002-2015, with observed ranges for each day in gray.

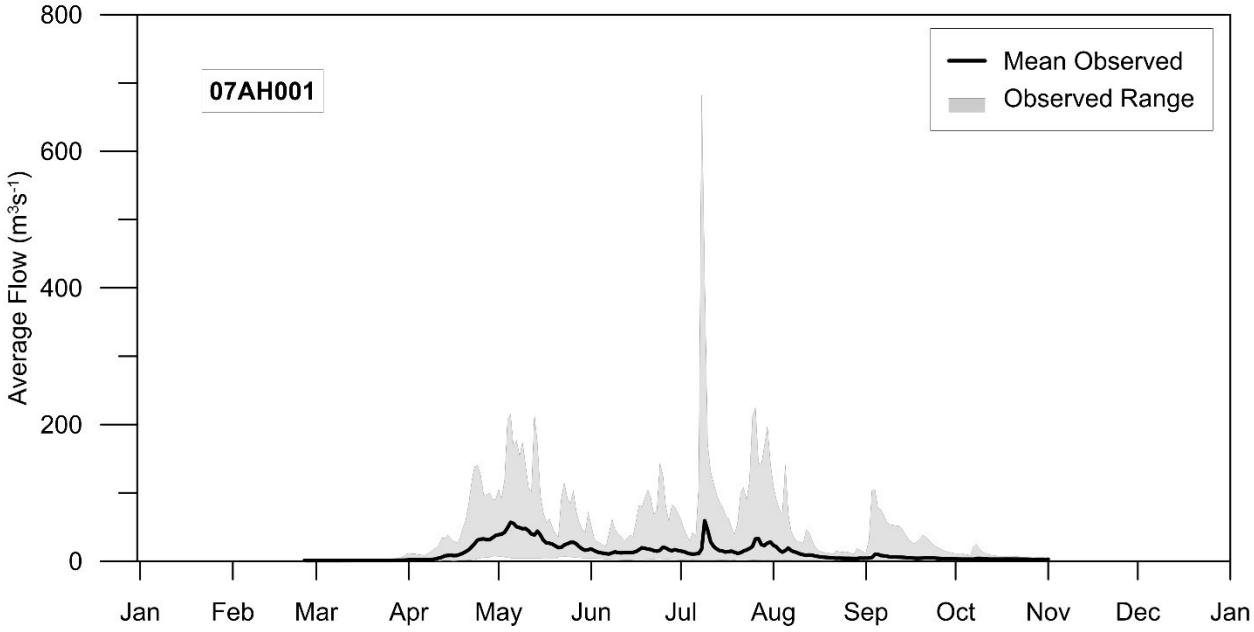

Figure C7: Average annual hydrograph for WSC gauge 07AH001 for 2002-2015, with observed ranges for each day in gray.

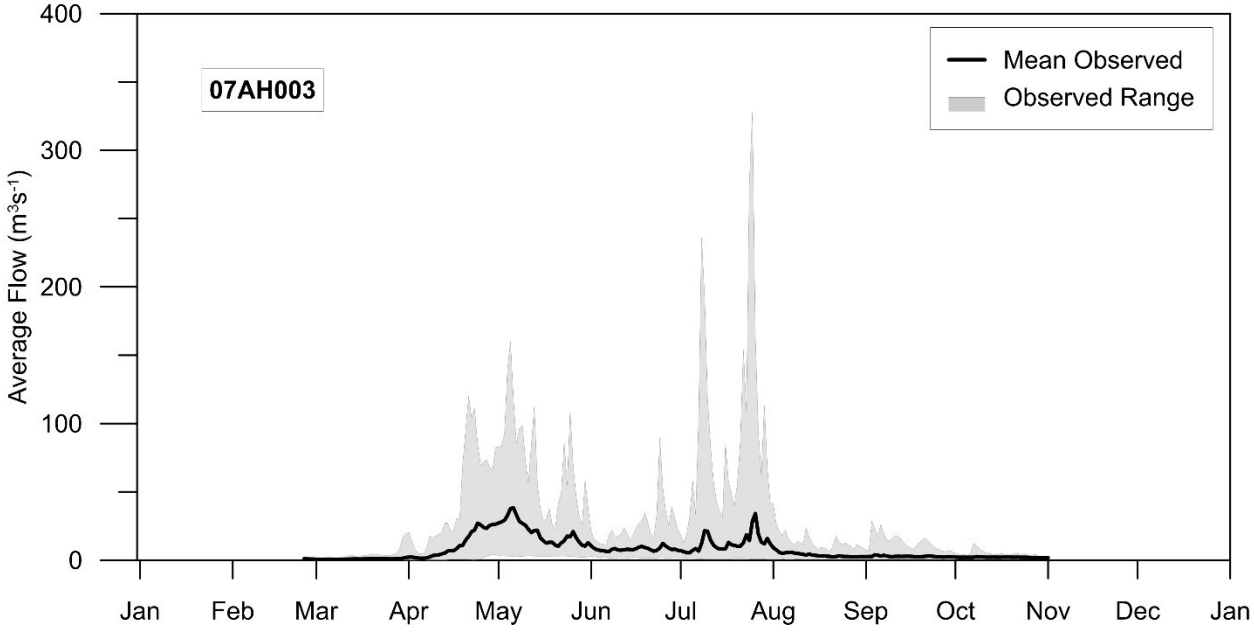

Figure C8: Average annual hydrograph for WSC gauge 07AH003 for 2002-2015, with observed ranges for each day in gray.

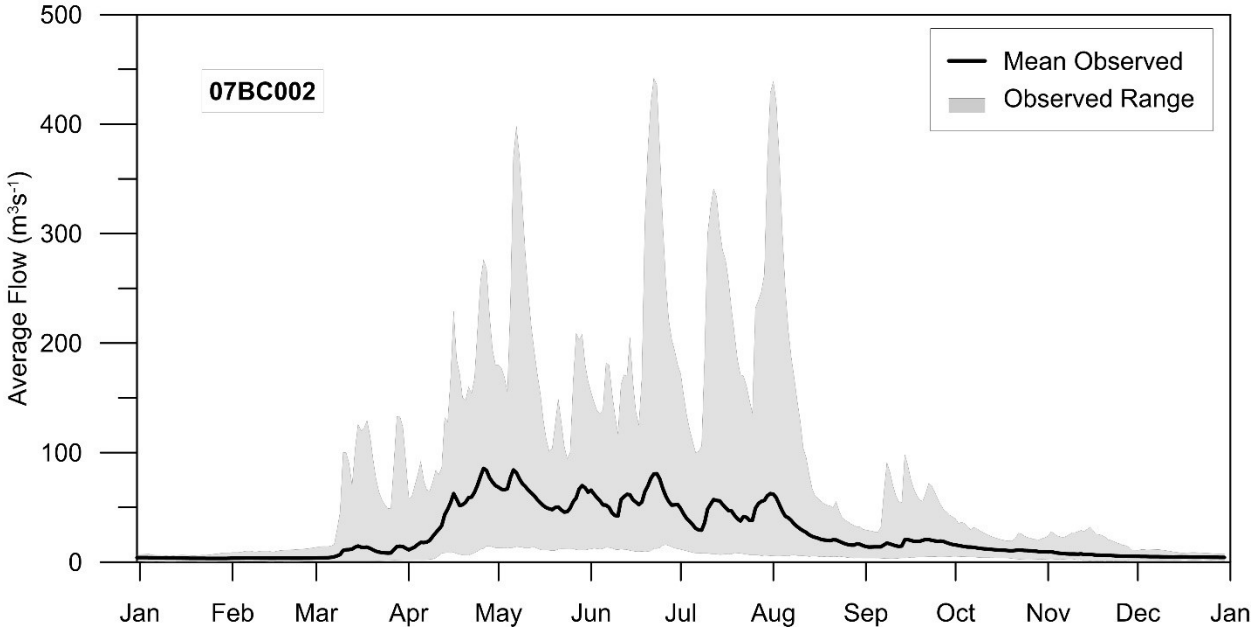

Figure C9: Average annual hydrograph for WSC gauge 07BC002 for 2002-2015, with observed ranges for each day in gray.

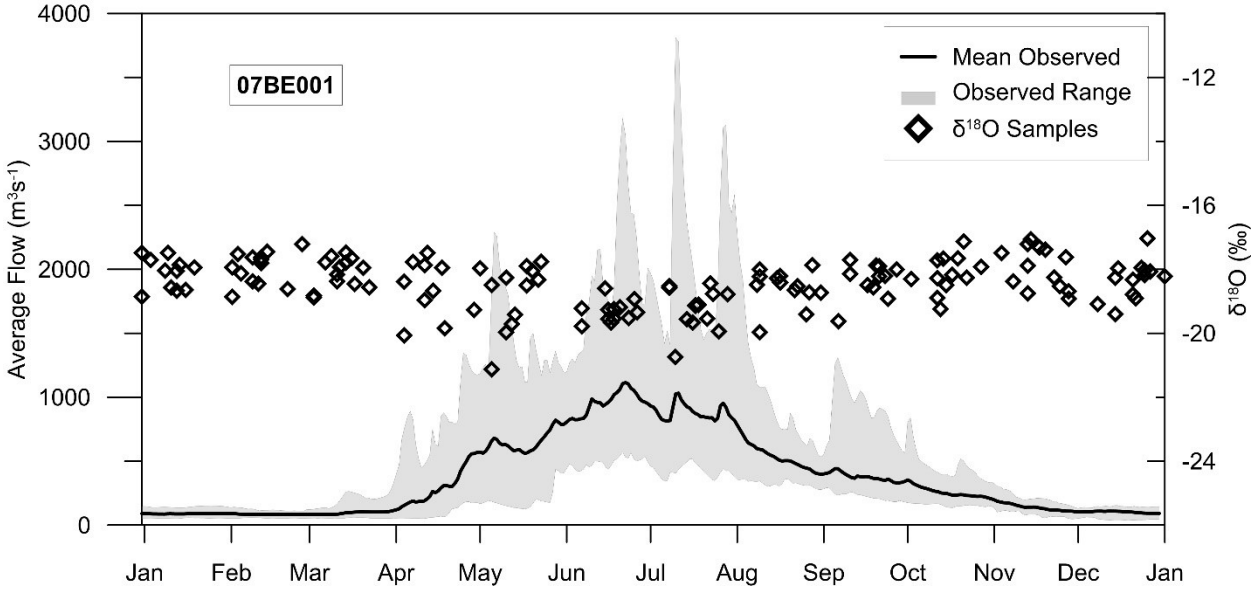

Figure C10: Average annual hydrograph for WSC gauge 07BE001 for 2002-2015, with observed flow ranges for each day in gray. All isotope samples in the same period taken at the site are shown as points.

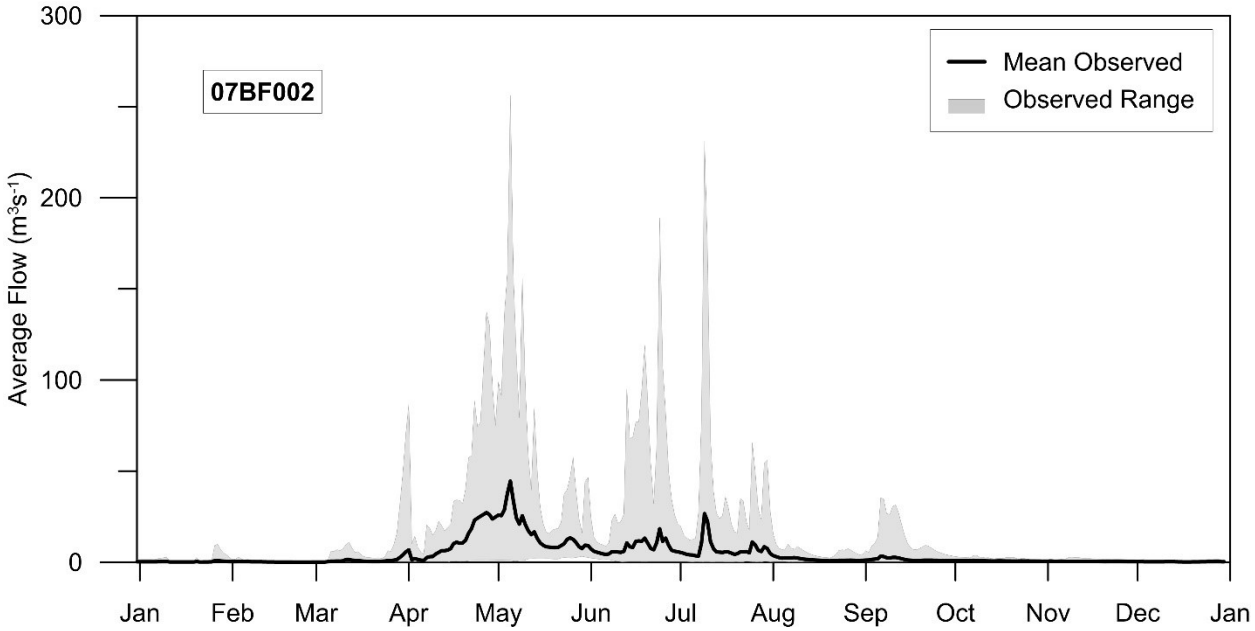

Figure C11: Average annual hydrograph for WSC gauge 07BF002 for 2002-2015, with observed ranges for each day in gray.

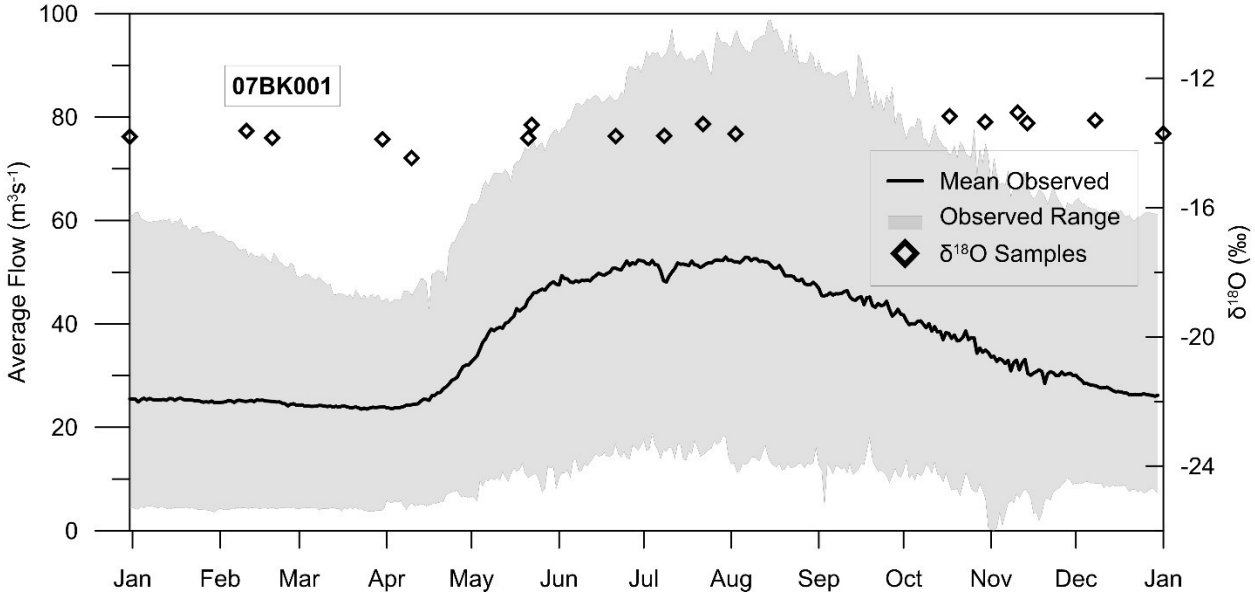

Figure C12: Average annual hydrograph for WSC gauge 07BK001 for 2002-2015, with observed flow ranges for each day in gray. All isotope samples in the same period taken at the site are shown as points.

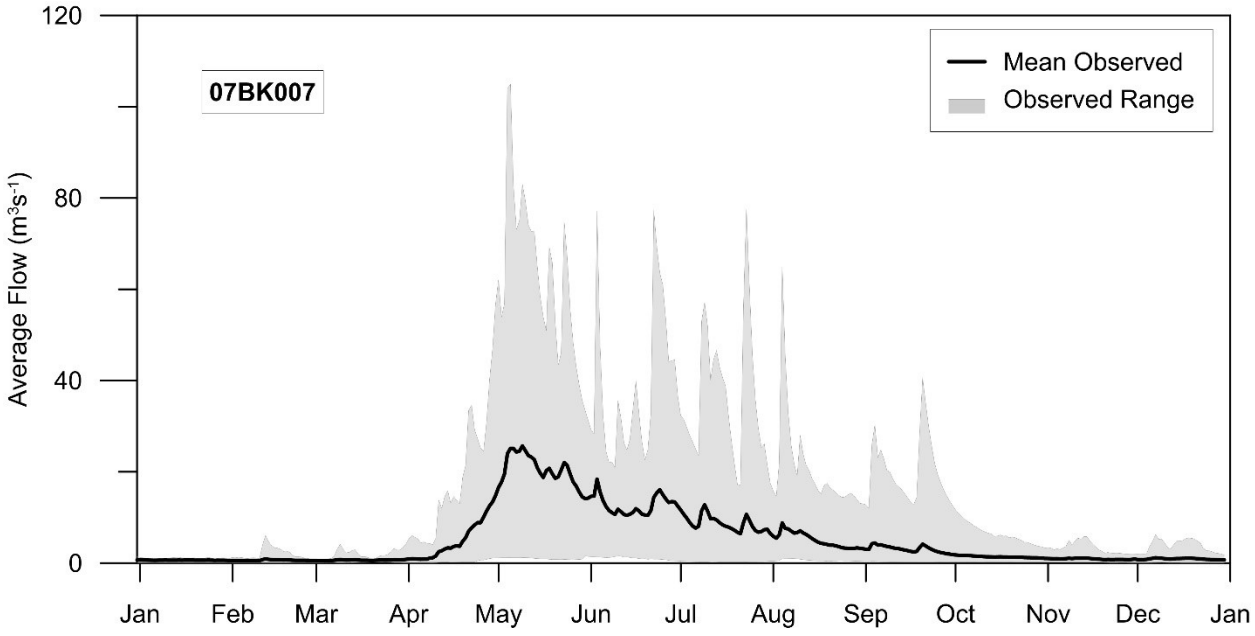

Figure C13: Average annual hydrograph for WSC gauge 07BK007 for 2002-2015, with observed ranges for each day in gray.

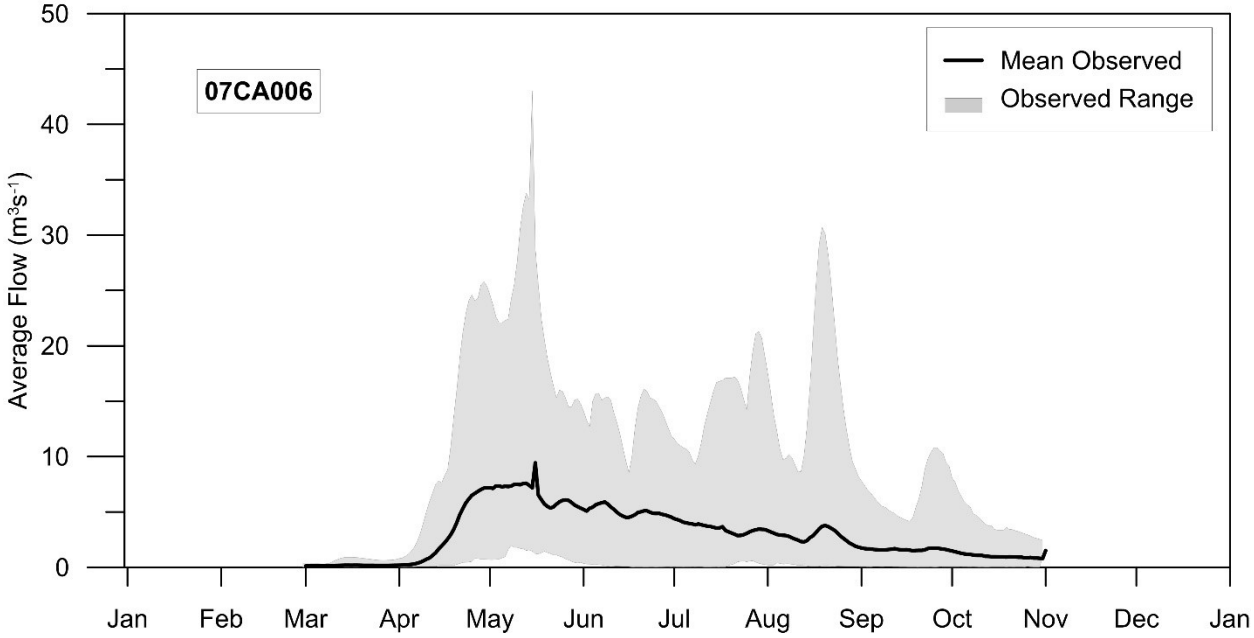

Figure C14: Average annual hydrograph for WSC gauge 07CA006 for 2002-2015, with observed ranges for each day in gray.

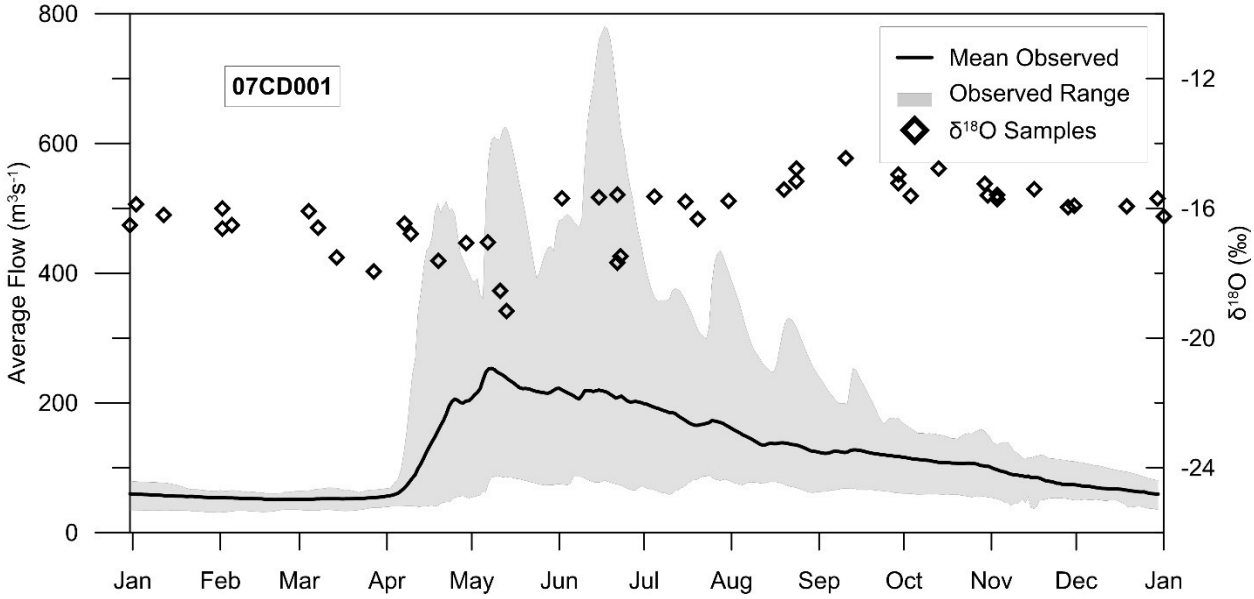

Figure C15: Average annual hydrograph for WSC gauge 07CD001 for 2002-2015, with observed flow ranges for each day in gray. All isotope samples in the same period taken at the site are shown as points.

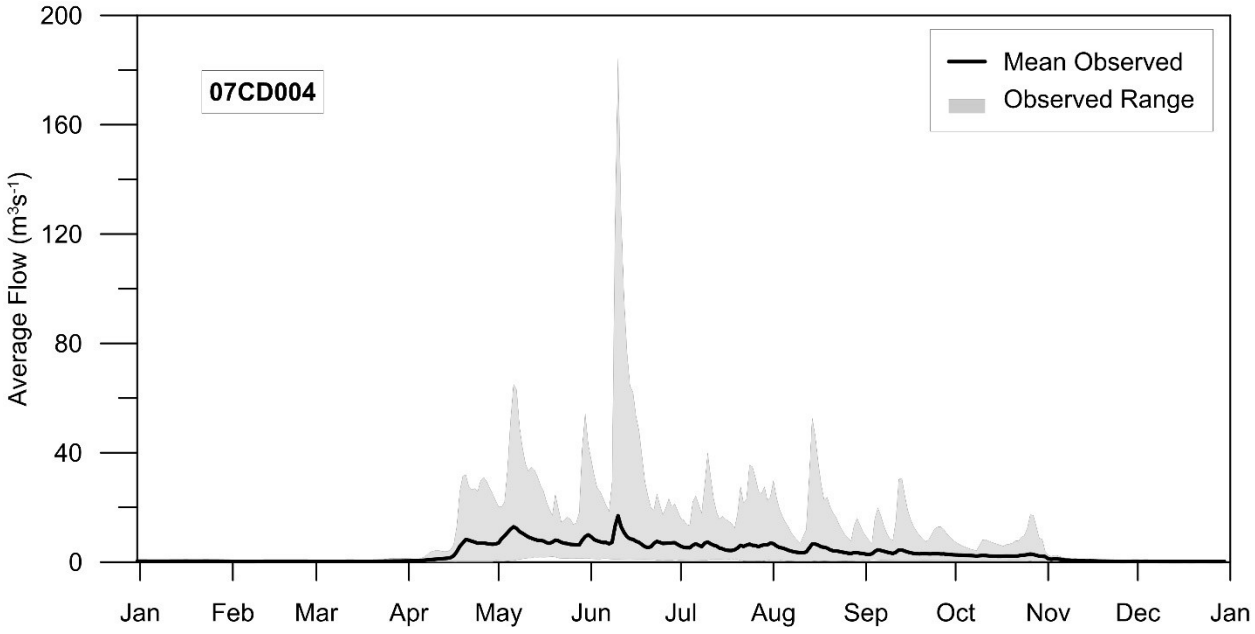

Figure C16: Average annual hydrograph for WSC gauge 07CD004 for 2002-2015, with observed ranges for each day in gray.

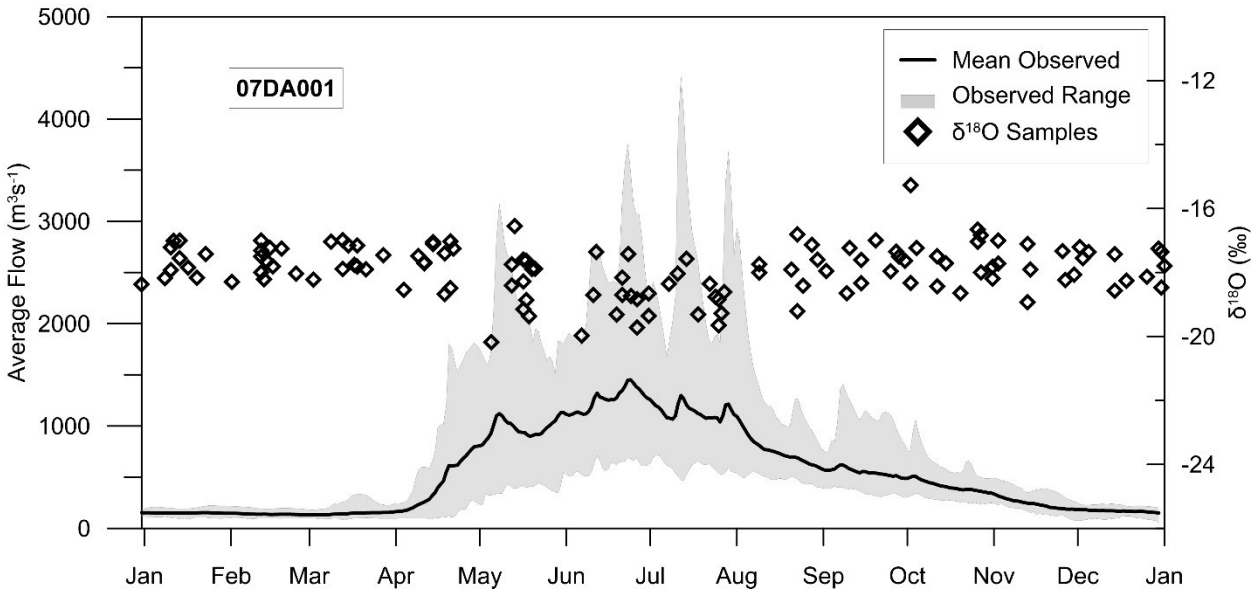

Figure C17: Average annual hydrograph for WSC gauge 07DA001 for 2002-2015, with observed flow ranges for each day in gray. All isotope samples in the same period taken at the site are shown as points.

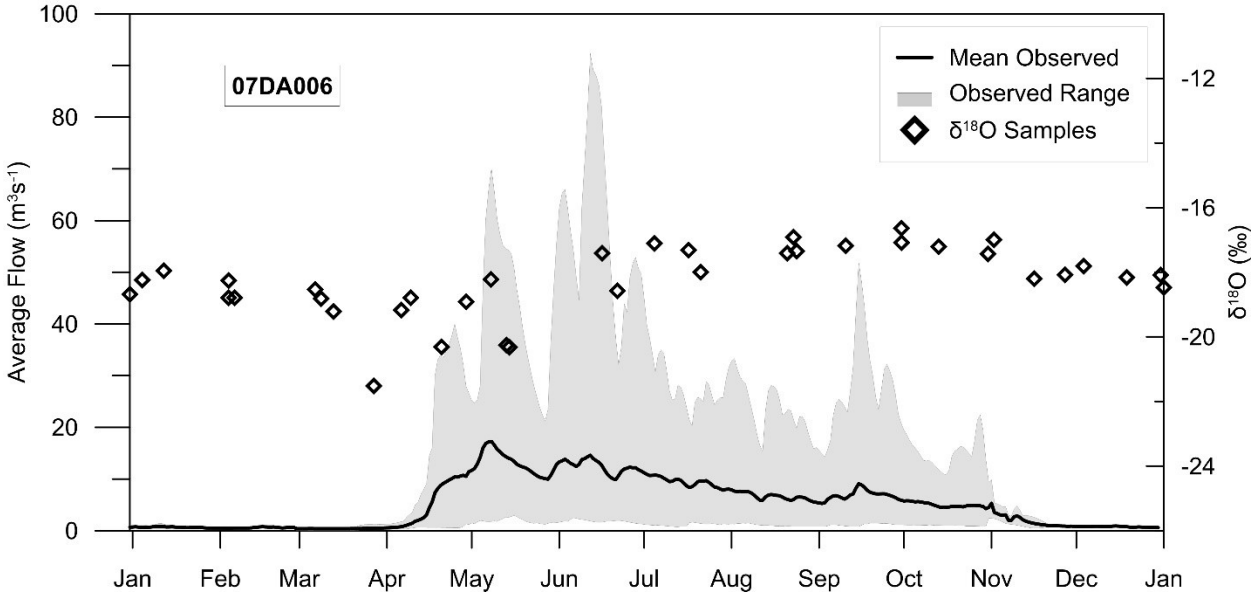

Figure C18: Average annual hydrograph for WSC gauge 07DA006 for 2002-2015, with observed flow ranges for each day in gray. All isotope samples in the same period taken at the site are shown as points.

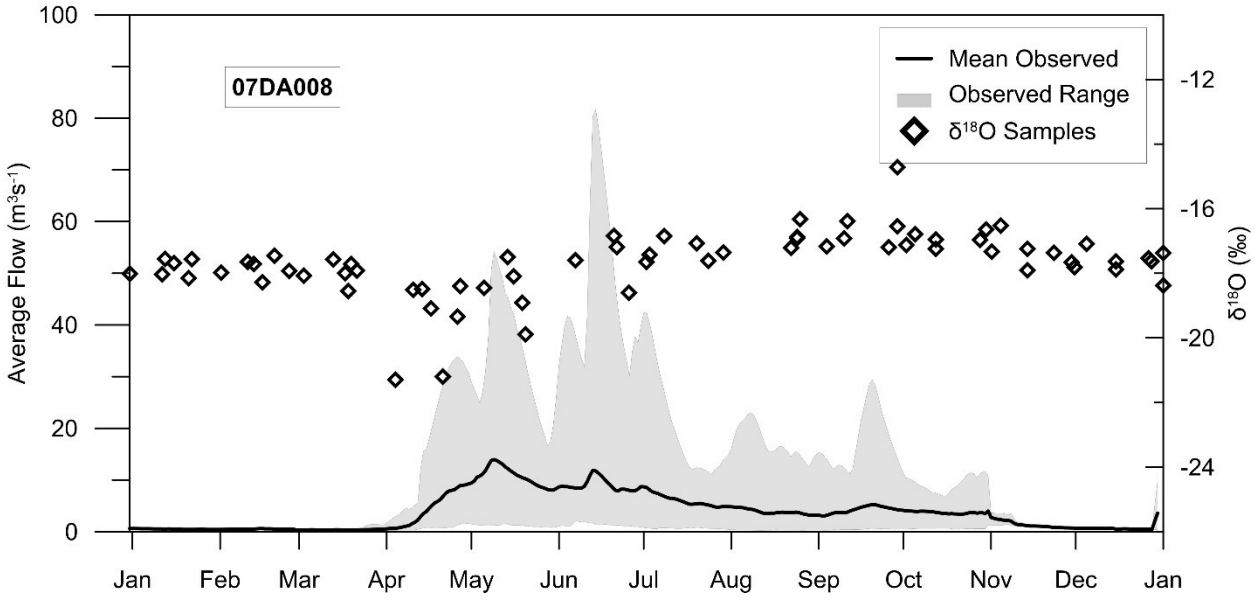

Figure C19: Average annual hydrograph for WSC gauge 07DA008 for 2002-2015, with observed flow ranges for each day in gray. All isotope samples in the same period taken at the site are shown as points.

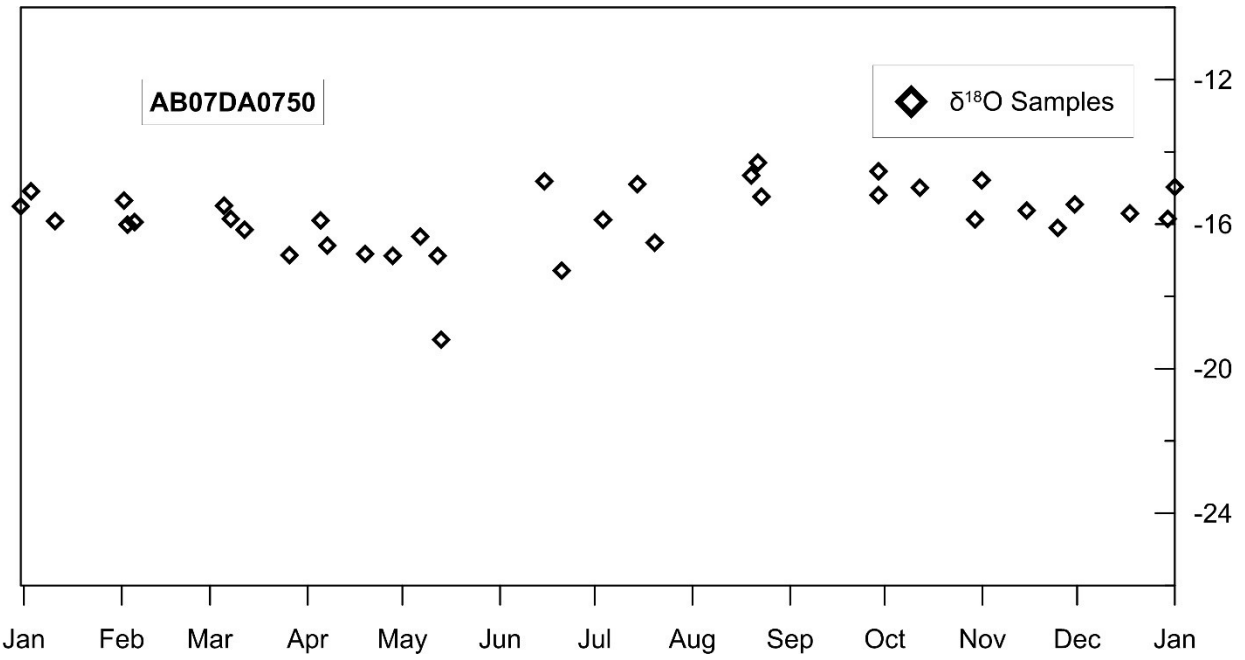

Figure C20: All isotope samples from 2002-2015 for gauge AB07DA0750 are shown as points.

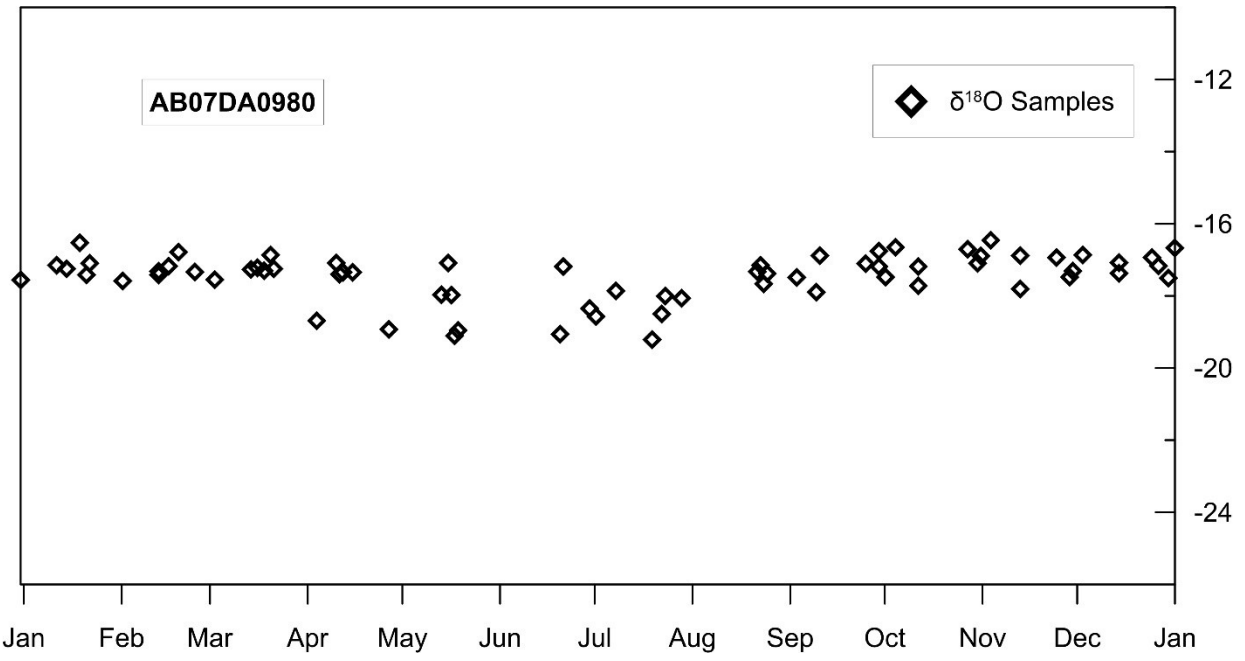

Figure C21: All isotope samples from 2002-2015 for gauge AB07DA0980 are shown as points.

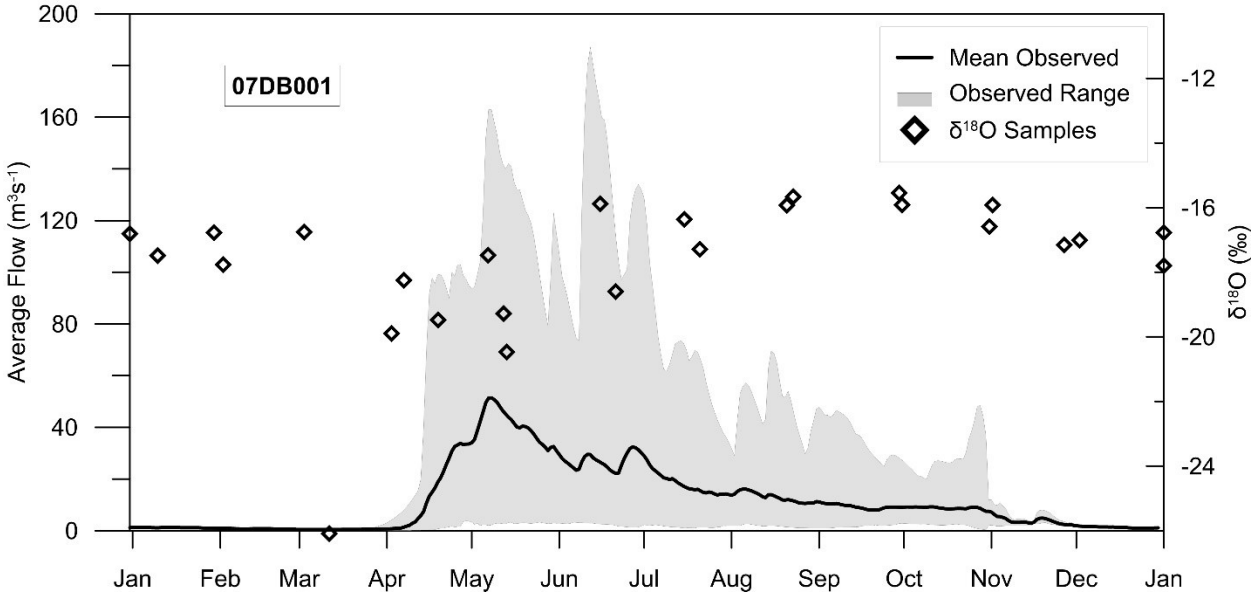

Figure C22: Average annual hydrograph for WSC gauge 07DB001 for 2002-2015, with observed flow ranges for each day in gray. All isotope samples in the same period taken at the site are shown as points.

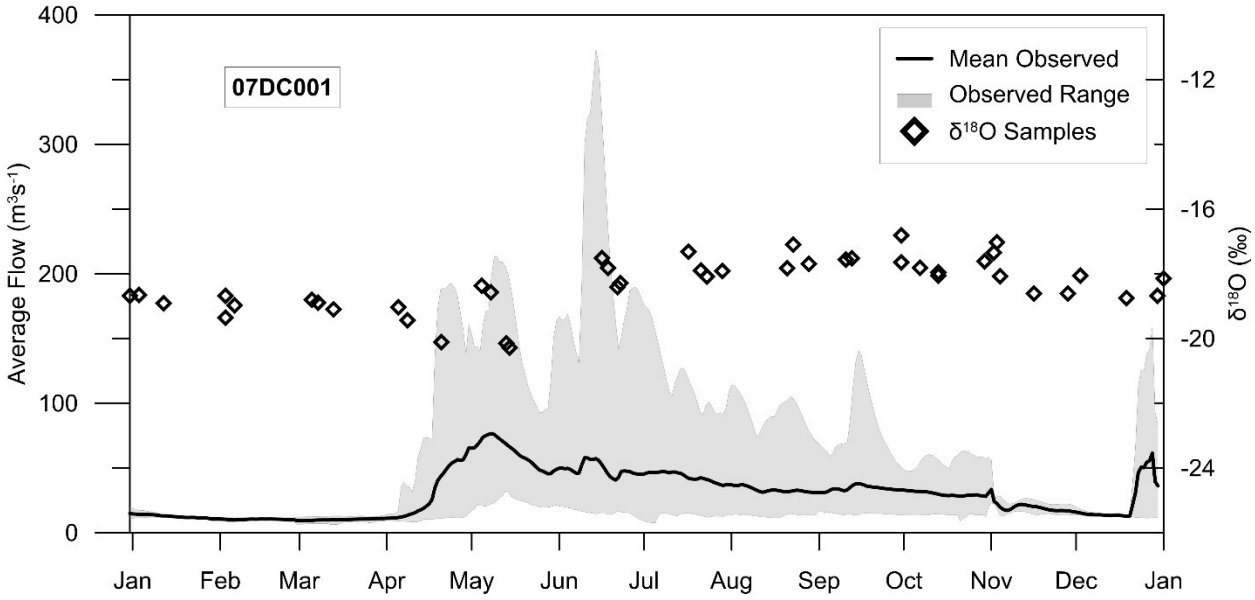

Figure C23: Average annual hydrograph for WSC gauge 07DC001 for 2002-2015, with observed flow ranges for each day in gray. All isotope samples in the same period taken at the site are shown as points.

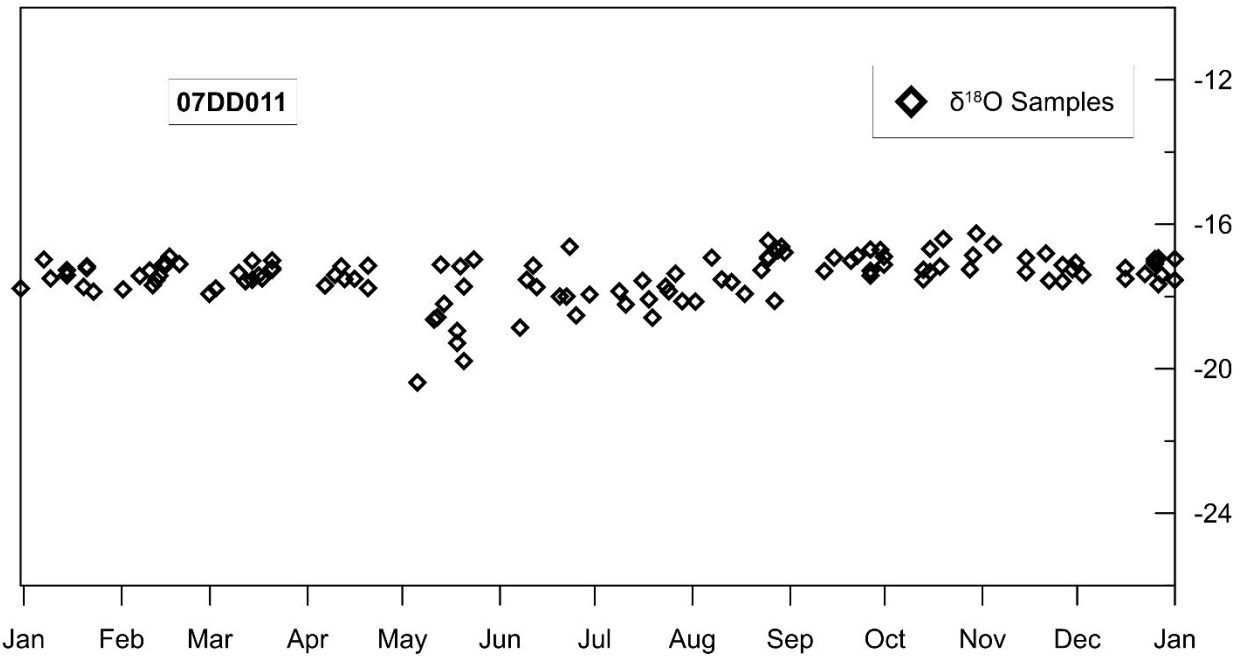

Figure C24: All isotope samples from 2002-2015 for gauge 07DD011 are shown as points.

VARS Configuration

The VARS program for this study was run with a Δh of 0.1 and the following parameter bounds:

Table C1: Parameter bounds for VARS analysis. Parameter ranges were selected based on the model developers' recommendations.

Factor#	LB	UB	Name	Units	Description
1	0	1	akF	-	Surface infiltration coefficient factor
2	0	1	recF	-	Horizontal soil conductivity factor
3	0.1	1	fpet	-	Open water evaporation
4	0.01	0.25	fmF1	mm/°C/h	Snow melt factor- grass, shrub and agriculture
5	0.01	0.25	fmF2	mm/°C/h	Snow melt factor- coniferous forest
6	0.01	0.25	fmF3	mm/°C/h	Snow melt factor- mixed forest
7	0.01	0.25	fmF4	mm/°C/h	Snow melt factor- barren ground and glacier
8	0.01	0.25	fmF5	mm/°C/h	Snow melt factor- disconnected wetland
9	0.01	0.25	fmF6	mm/°C/h	Snow melt factor- connected wetland
10	0.01	0.25	fmF7	mm/°C/h	Snow melt factor- water and impervious
11	20	150	retn1	mm	Soil retention capacity- grass and agriculture
12	30	200	retn2	mm	Soil retention capacity- coniferous forest
13	30	200	retn3	mm	Soil retention capacity- mixed forest
14	20	150	retn4	mm	Soil retention capacity- barren ground
15	20	200	retn5	mm	Soil retention capacity- shrub
16	50	500	retn6	mm	Soil retention capacity- disconnected wetland
17	0	1	ak2F	-	Vertical soil conductivity factor
18	0.000001	0.001	flz1	-	Groundwater coefficient- central
19	0.000001	0.001	flz2	-	Groundwater coefficient- boreal shield
20	0.000001	0.001	flz3	-	Groundwater coefficient- mountains
21	1.5	3.99	pwr	-	Groundwater power
22	0.012	0.04	r2n1	-	Channel roughness- central
23	0.012	0.04	r2n2	-	Channel roughness- boreal shield
24	0.012	0.04	r2n3	-	Channel roughness- mountains
25	0.1	0.95	theta	-	Connected wetland porosity
26	0.1	0.8	kcond	m/s	Connected wetland conductivity
27	0.5	2	gladjust	-	Glacier melt factor

VARS results

Table C2: Reliability ranges from VARS for all metrics

	akF	recF	fpet	fmF1	fmF2	fmF3	fmF4	fmF5	fmF6	fmF7	retn1	retn2	retn3	retn4	retn5	retn6	ak2F	flz1	flz2	flz3	pwr	r2n1	r2n2	r2n3	theta	kcond	gladjust
NRMSE	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
NSE	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
logNSE	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
KGE	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
β (bias)	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
α (var)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Correlation	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
SFDC	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Q5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Q95	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
O NRMSE	0.0	0.3	0.0	0.0	0.1	0.2	0.3	0.2	0.0	0.0	0.2	0.1	0.1	0.0	0.1	0.0	0.0	0.2	0.1	0.1	0.2	0.0	0.1	0.0	0.1	0.0	0.0
O KGE	0.0	0.0	0.0	0.0	0.0	0.0	1.4	0.0	0.0	0.0	0.0	0.0	0.7	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0
O β (bias)	0.0	0.4	0.0	0.0	0.0	0.2	0.4	0.2	0.0	0.0	0.3	0.2	0.1	0.0	0.2	0.0	0.0	0.2	0.1	0.0	0.2	0.0	0.1	0.0	0.1	0.0	0.0
O α (var)	0.0	0.0	0.0	0.0	0.0	0.0	1.4	0.0	0.0	0.0	0.0	0.0	0.7	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0
O Correlation	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
O SDC	0.0	0.0	0.0	0.0	0.0	0.0	0.8	0.0	0.0	0.0	0.0	0.0	1.1	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0
O Q5	0.0	0.2	0.0	0.0	0.0	0.1	0.2	0.1	0.0	0.0	0.2	0.1	0.0	0.0	0.1	0.0	0.0	0.1	0.1	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0
O Q95	0.0	0.2	0.0	0.0	0.1	0.1	0.2	0.1	0.0	0.0	0.2	0.1	0.1	0.0	0.1	0.0	0.0	0.2	0.1	0.1	0.2	0.1	0.1	0.0	0.1	0.0	0.0
H NRMSE	0.0	0.3	0.0	0.0	0.1	0.2	0.3	0.2	0.0	0.0	0.3	0.1	0.1	0.0	0.1	0.0	0.0	0.2	0.2	0.1	0.2	0.0	0.1	0.0	0.1	0.0	0.0
H KGE	0.0	0.1	0.1	0.0	0.1	0.0	0.2	0.1	0.0	0.0	0.0	0.0	0.8	0.0	0.1	0.1	0.0	0.3	0.1	0.1	0.3	0.1	0.1	0.0	0.0	0.1	0.0
H β (bias)	0.0	0.4	0.0	0.0	0.0	0.2	0.4	0.2	0.0	0.0	0.3	0.2	0.1	0.0	0.2	0.0	0.0	0.2	0.1	0.0	0.2	0.0	0.1	0.0	0.1	0.0	0.0
H α (var)	0.0	0.1	0.1	0.0	0.1	0.0	0.2	0.1	0.0	0.0	0.0	0.0	0.8	0.0	0.1	0.1	0.0	0.3	0.1	0.1	0.3	0.1	0.1	0.0	0.0	0.1	0.0
H Correlation	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
H SDC	0.0	0.1	0.0	0.0	0.1	0.0	0.3	0.1	0.0	0.0	0.1	0.1	0.2	0.0	0.1	0.0	0.0	0.3	0.2	0.2	0.3	0.0	0.0	0.0	0.0	0.0	0.0
H Q5	0.0	0.5	0.0	0.0	0.0	0.2	0.5	0.2	0.0	0.0	0.4	0.1	0.0	0.0	0.2	0.0	0.0	0.2	0.1	0.0	0.2	0.0	0.0	0.0	0.2	0.0	0.0
H Q95	0.0	0.2	0.0	0.0	0.1	0.1	0.3	0.2	0.0	0.0	0.2	0.1	0.1	0.0	0.1	0.0	0.0	0.2	0.1	0.1	0.2	0.1	0.1	0.0	0.1	0.1	0.0
LML mE	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
LML bE	0.0	0.5	0.0	0.0	0.0	0.2	0.4	0.2	0.0	0.0	0.3	0.1	0.0	0.0	0.1	0.0	0.0	0.1	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.1	0.0
LML RE	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table C3: VARS sensitivity correlations for all metrics

	NRMSE	NSE	logNSE	KGE	β (bias)	α (var)	Correlation	SFDC	Q5	Q95	O NRMSE	O KGE	O β (bias)	O α (var)	O Corr	O SDC	O Q5	O Q95	H NRMSE	H KGE	H β (bias)	H α (var)	H Corr	H SDC	H Q5	H Q95	LML mE	LML bE	LML RE
NRMSE	1.00	0.99	-0.09	0.73	0.17	0.48	0.96	0.67	0.69	-0.02	0.06	-0.02	0.05	-0.02	0.20	-0.01	0.02	0.14	0.06	0.02	0.04	0.02	0.22	0.15	-0.01	0.13	0.09	0.01	0.23
NSE	0.99	1.00	-0.09	0.67	0.14	0.39	0.97	0.65	0.63	-0.06	0.05	-0.02	0.03	-0.02	0.20	-0.02	0.01	0.12	0.05	0.01	0.03	0.01	0.23	0.15	-0.02	0.11	0.09	0.00	0.23
logNSE	-0.09	-0.09	1.00	0.23	-0.01	0.18	-0.06	0.02	-0.08	0.31	-0.09	-0.04	-0.11	-0.03	0.34	-0.02	0.27	-0.09	-0.15	-0.06	-0.14	-0.06	0.00	-0.14	-0.11	-0.13	0.92	0.16	0.83
KGE	0.73	0.67	0.23	1.00	0.50	0.92	0.64	0.65	0.92	0.06	0.08	0.06	0.07	0.06	0.39	0.03	0.14	0.10	0.06	-0.04	0.05	-0.04	0.28	0.11	0.01	0.08	0.35	0.11	0.42
β (bias)	0.17	0.14	-0.01	0.50	1.00	0.43	0.17	0.69	0.53	-0.15	0.35	0.72	0.38	0.71	0.72	0.48	0.23	0.31	0.34	0.12	0.35	0.12	0.75	0.40	0.38	0.32	0.05	0.38	-0.02
α (var)	0.48	0.39	0.18	0.92	0.43	1.00	0.33	0.40	0.91	0.11	0.07	-0.04	0.06	-0.04	0.23	-0.03	0.13	0.05	0.06	-0.07	0.06	-0.07	0.12	0.01	0.02	0.04	0.25	0.08	0.32
Correlation	0.96	0.97	-0.06	0.64	0.17	0.33	1.00	0.72	0.55	-0.02	-0.02	0.02	-0.04	0.01	0.24	-0.02	-0.06	0.07	-0.02	-0.02	-0.05	-0.02	0.27	0.14	-0.07	0.06	0.10	-0.04	0.22
SFDC	0.67	0.65	0.02	0.65	0.69	0.40	0.72	1.00	0.58	0.13	0.17	0.50	0.18	0.49	0.65	0.27	0.11	0.22	0.16	0.03	0.16	0.03	0.70	0.34	0.18	0.22	0.11	0.21	0.14
Q5	0.69	0.63	-0.08	0.92	0.53	0.91	0.55	0.58	1.00	0.00	0.18	0.11	0.19	0.10	0.27	0.07	0.14	0.16	0.17	0.00	0.17	0.00	0.27	0.12	0.15	0.16	0.05	0.14	0.14
Q95	-0.02	-0.06	0.31	0.06	-0.15	0.11	-0.02	0.13	0.00	1.00	0.00	-0.09	0.00	-0.09	-0.01	-0.08	0.10	0.01	0.00	-0.02	0.01	-0.02	-0.03	-0.06	0.04	0.00	0.10	0.03	0.12
O NRMSE	0.06	0.05	-0.09	0.08	0.35	0.07	-0.02	0.17	0.18	0.00	1.00	0.43	0.99	0.43	0.36	0.37	0.81	0.95	1.00	0.34	0.99	0.34	0.34	0.70	0.90	0.95	0.03	0.87	0.01
O KGE	-0.02	-0.02	-0.04	0.06	0.72	-0.04	0.02	0.50	0.11	-0.09	0.43	1.00	0.43	1.00	0.69	0.87	0.30	0.48	0.42	0.57	0.40	0.58	0.71	0.62	0.36	0.48	0.07	0.37	0.01
O β (bias)	0.05	0.03	-0.11	0.07	0.38	0.06	-0.04	0.18	0.19	0.00	0.99	0.43	1.00	0.42	0.35	0.34	0.83	0.90	0.98	0.26	1.00	0.26	0.35	0.61	0.95	0.90	-0.01	0.91	-0.03
O α (var)	-0.02	-0.02	-0.03	0.06	0.71	-0.04	0.01	0.49	0.10	-0.09	0.43	1.00	0.42	1.00	0.69	0.88	0.30	0.48	0.42	0.60	0.40	0.60	0.70	0.62	0.36	0.48	0.07	0.36	0.02
O Correlation	0.20	0.20	0.34	0.39	0.72	0.23	0.24	0.65	0.27	-0.01	0.36	0.69	0.35	0.69	1.00	0.51	0.50	0.41	0.33	0.27	0.32	0.27	0.91	0.45	0.31	0.39	0.50	0.44	0.49
O SDC	-0.01	-0.02	-0.02	0.03	0.48	-0.03	-0.02	0.27	0.07	-0.08	0.37	0.87	0.34	0.88	0.51	1.00	0.24	0.46	0.37	0.86	0.33	0.86	0.49	0.62	0.20	0.45	0.13	0.22	0.10
O Q5	0.02	0.01	0.27	0.14	0.23	0.13	-0.06	0.11	0.14	0.10	0.81	0.30	0.83	0.30	0.50	0.24	1.00	0.74	0.77	0.18	0.81	0.18	0.36	0.38	0.82	0.71	0.41	0.90	0.42
O Q95	0.14	0.12	-0.09	0.10	0.31	0.05	0.07	0.22	0.16	0.01	0.95	0.48	0.90	0.48	0.41	0.46	0.74	1.00	0.95	0.49	0.91	0.49	0.37	0.82	0.75	1.00	0.08	0.74	0.09
H NRMSE	0.06	0.05	-0.15	0.06	0.34	0.06	-0.02	0.16	0.17	0.00	1.00	0.42	0.98	0.42	0.33	0.37	0.77	0.95	1.00	0.35	0.99	0.35	0.33	0.73	0.88	0.96	-0.03	0.84	-0.04
H KGE	0.02	0.01	-0.06	-0.04	0.12	-0.07	-0.02	0.03	0.00	-0.02	0.34	0.57	0.26	0.60	0.27	0.86	0.18	0.49	0.35	1.00	0.28	1.00	0.21	0.60	0.05	0.49	0.13	0.09	0.16
H β (bias)	0.04	0.03	-0.14	0.05	0.35	0.06	-0.05	0.16	0.17	0.01	0.99	0.40	1.00	0.40	0.32	0.33	0.81	0.91	0.99	0.28	1.00	0.27	0.33	0.64	0.94	0.92	-0.05	0.88	-0.06
H α (var)	0.02	0.01	-0.06	-0.04	0.12	-0.07	-0.02	0.03	0.00	-0.02	0.34	0.58	0.26	0.60	0.27	0.86	0.18	0.49	0.35	1.00	0.27	1.00	0.21	0.60	0.05	0.48	0.13	0.08	0.16
H Correlation	0.22	0.23	0.00	0.28	0.75	0.12	0.27	0.70	0.27	-0.03	0.34	0.71	0.35	0.70	0.91	0.49	0.36	0.37	0.33	0.21	0.33	0.21	1.00	0.45	0.34	0.36	0.13	0.36	0.15
H SDC	0.15	0.15	-0.14	0.11	0.40	0.01	0.14	0.34	0.12	-0.06	0.70	0.62	0.61	0.62	0.45	0.62	0.38	0.82	0.73	0.60	0.64	0.60	0.45	1.00	0.42	0.82	0.01	0.38	0.00
H Q5	-0.01	-0.02	-0.11	0.01	0.38	0.02	-0.07	0.18	0.15	0.04	0.90	0.36	0.95	0.36	0.31	0.20	0.82	0.75	0.88	0.05	0.94	0.05	0.34	0.42	1.00	0.76	-0.07	0.94	-0.10
H Q95	0.13	0.11	-0.13	0.08	0.32	0.04	0.06	0.22	0.16	0.00	0.95	0.48	0.90	0.48	0.39	0.45	0.71	1.00	0.96	0.49	0.92	0.48	0.36	0.82	0.76	1.00	0.04	0.74	0.04
LML mE	0.09	0.09	0.92	0.35	0.05	0.25	0.10	0.11	0.05	0.10	0.03	0.07	-0.01	0.07	0.50	0.13	0.41	0.08	-0.03	0.13	-0.05	0.13	0.13	0.01	-0.07	0.04	1.00	0.23	0.97
LML bE	0.01	0.00	0.16	0.11	0.38	0.08	-0.04	0.21	0.14	0.03	0.87	0.37	0.91	0.36	0.44	0.22	0.90	0.74	0.84	0.09	0.88	0.08	0.36	0.38	0.94	0.74	0.23	1.00	0.19
LML RE	0.23	0.23	0.83	0.42	-0.02	0.32	0.22	0.14	0.14	0.12	0.01	0.01	-0.03	0.02	0.49	0.10	0.42	0.09	-0.04	0.16	-0.06	0.16	0.15	0.00	-0.10	0.04	0.97	0.19	1.00

Appendix D: Supporting information for Chapter 5

Table D1: Parameter bounds for calibration. Parameter ranges were selected based on the model developers' recommendations.

Name	Lower bound	Upper bound	Units	Description
akF	0	1	-	Surface infiltration coefficient factor
recF	0	1	-	Horizontal soil conductivity factor
fpet	0.1	1	-	Open water evaporation
fmF1	0.01	0.25	mm/°C/h	Snow melt factor- grass, shrub and agriculture
fmF2	0.01	0.25	mm/°C/h	Snow melt factor- coniferous forest
fmF3	0.01	0.25	mm/°C/h	Snow melt factor- mixed forest
fmF4	0.01	0.25	mm/°C/h	Snow melt factor- barren ground and glacier
fmF5	0.01	0.25	mm/°C/h	Snow melt factor- disconnected wetland
fmF6	0.01	0.25	mm/°C/h	Snow melt factor- connected wetland
fmF7	0.01	0.25	mm/°C/h	Snow melt factor- water and impervious
retn1	20	150	mm	Soil retention capacity- grass and agriculture
retn2	30	200	mm	Soil retention capacity- coniferous forest
retn3	30	200	mm	Soil retention capacity- mixed forest
retn4	20	150	mm	Soil retention capacity- barren ground
retn5	20	200	mm	Soil retention capacity- shrub
retn6	50	500	mm	Soil retention capacity- disconnected wetland
ak2F	0	1	-	Vertical soil conductivity factor
flz1	0.000001	0.001	-	Groundwater coefficient- central
flz2	0.000001	0.001	-	Groundwater coefficient- boreal shield
flz3	0.000001	0.001	-	Groundwater coefficient- mountains
pwr	1.5	3.99	-	Groundwater power
r2n1	0.012	0.04	-	Channel roughness- central
r2n2	0.012	0.04	-	Channel roughness- boreal shield
r2n3	0.012	0.04	-	Channel roughness- mountains
theta	0.1	0.95	-	Connected wetland porosity
kcond	0.1	0.8	m/s	Connected wetland conductivity
gladjust	0.5	2	-	Glacier melt factor

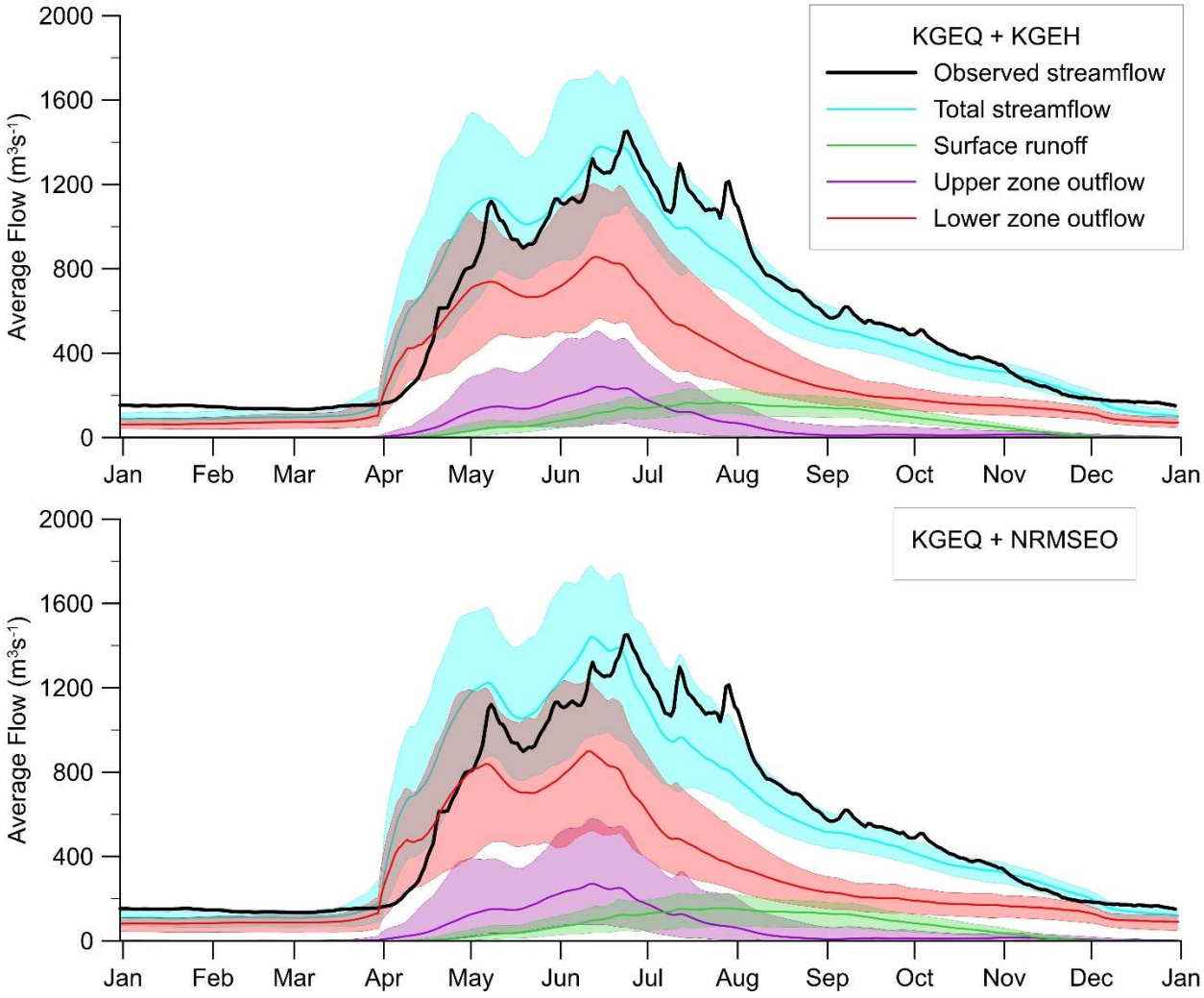

Figure D1: Average annual hydrographs for the simulation period (2001 to 2015), with behavioral calibrated ensembles shown, along with surface, upper and lower zone virtual tracer flows at Fort McMurray 07DA001 (mean and total range are included).

Appendix E: Model equations and assumptions for isoWATFLOOD

The isoWATFLOOD model simulates isotope mass balances for all of the hydrologic storages in the WATFLOOD model. The stable isotope volume in every grouped response unit, in every grid cell, is calculated based on the fluxes in and out of the storage unit. The stable isotope volume in each storage compartment is computed each time interval based on the stored volume at the beginning of the time interval, and the inflow and outflow over the time interval. Volumes within WATFLOOD and isoWATFLOOD are measured in cubic meters, while isotopic concentrations are unitless (m^3/m^3), but are later converted into delta values for output relative to Vienna Standard Mean Ocean Water (VSMOW). The δ value is defined as:

$$\delta‰=(R/R_{std} - 1) \times 10^3 \qquad (E.1)$$

Where R is the relative abundance of D/H or $^{18}O/^{16}O$, and R_{std} comes from the world standard (Vienna Standard Mean Ocean Water, VSMOW2, or previously VSMOW) (Gat, 1996). Isotopes are transferred between compartments based on the isotopic concentration within the storage and the water flux calculated by WATFLOOD. This appendix is a summary of the equations used to simulate stable isotopes in each type of hydrologic storage, and in evaporated water. More detailed information can be found in:

Holmes, T. (2016). isoWATFLOOD Stable water isotope simulation in the WATFLOOD hydrologic model. http://www.civil.uwaterloo.ca/watflood/downloads/isoWATFLOOD%20manual.pdf

Holmes, T. (2016). Assessing the value of stable water isotopes in hydrologic modeling: a dual isotope approach. (http://hdl.handle.net/1993/31724). Winnipeg, MB: University of Manitoba.

Evaporation

To model fractionation due to evaporation the concentration of stable isotopes in the evaporated water vapor, C_E^{iso}, is needed; this concentration is dependent on both atmospheric conditions and

the concentration of the evaporating water body. There are two significant components to fractionation: equilibrium and kinetic separation (Gibson, et al., 2008). Equilibrium isotopic fractionation is temperature dependent, with the equilibrium isotopic separation between liquid and water, ε^*, calculated as (Gibson, et al., 2008):

$$\varepsilon^* = \alpha^* - 1 \qquad (E.2)$$

The equilibrium fractionation factor for the liquid to vapor transition, α^*, is calculated using the empirically derived equations from Horita and Wesolowski (1994):

$$\alpha^*_{18O} = e^{\left(-7.685+6.7123*(10^3/\mathrm{T})-1.6664*(10^6/\mathrm{T}^2)+0.35041*(10^9/T^3)\right)/1000} \qquad (E.3)$$

$$\alpha^*_{2H} = e^{(1158.8*(T^3/10^9)-1620.1*(\mathrm{T}^2/10^6)+794.84*(\mathrm{T}/10^3)-161.04+2.9992*(10^9/T^3))/1000} \qquad (E.4)$$

Where α is in decimal format, and T is air temperature in degrees kelvin. The kinetic separation of isotope species is dependent on both surface properties and relative humidity, with kinetic isotopic separation, ε_k, defined as (Gat, 1996):

$$\varepsilon_k = C_D n\theta(1-h) \qquad (E.5)$$

Where C_D is 28.55‰ for ^{18}O and 25.115‰ for D, h is the relative humidity, n is a turbulence parameter, and θ is a transport resistance factor reduced from unity for large open water bodies (Gibson, et al., 2008). In isoWATFLOOD, n is equal to 2/3 for wetlands, 1/2 for lakes, 1 for soil water, and 0 for river channels (the water is assumed to be too turbulent for kinetic fractionation). The θ parameter is equal to 1, except for lakes, which are assumed to have some additional resistance, with θ equal to 0.88.

The isotopic composition of the evaporated water vapor, δ_E, is modeled using a simplified version of the Craig-Gordon equation describing the composition of the water vapor as a

function of the composition of the evaporating water body, δ_L, is used in the simulation (Gonfiantini, 1986):

$$\delta_E = \frac{\frac{(\delta_L - \varepsilon^*)}{\alpha^*} - h\delta_a - \varepsilon_k}{1 - h + \varepsilon_k} \quad (E.6)$$

Where δ_a is the isotopic composition of the ambient atmospheric moisture; h, ε and δ are in decimal notation. The evaporating water composition, δ_L, is dependent on the composition of the water source undergoing evaporation, which can be different for each storage compartment in the model, and is assumed to be the last computed (previous time step) isotopic composition of any given evaporating compartment. The isotopes in atmospheric moisture are often assumed to be in equilibrium with those in precipitation, although if evaporated water constitutes a significant fraction of the vapor, this assumption may not be valid (Gibson, et al., 2008). With the assumed relationship between the isotopic composition of atmospheric moisture and the isotopic composition of precipitation, δ_P, which depends on there being no recirculated moisture, δ_a can be calculated as (Gibson, et al., 2016):

$$\delta_A = \frac{\delta_P - \varepsilon^*}{1 + \varepsilon^*} \quad (E.7)$$

with ε^* and δ in decimal notation.

Enrichment is limited by atmospheric conditions, such that a desiccating water body has a limiting isotopic composition, δ^*, reached as the water volume approaches zero (Gibson, et al., 2016):

$$\delta^* = \frac{h\delta_a + \varepsilon_k + \varepsilon^*/\alpha^*}{h - \varepsilon_k - \varepsilon^*/\alpha^*} \quad (E.8)$$

The simulation uses the composition from the previous time interval in calculating δ_E for any storage undergoing evaporative fractionation, including soil storages, connected wetlands, rivers and lakes. The concentration of stable isotopes in the evaporated water vapor, C_E^{iso}, is calculated simply as the concentration form of δ_E (which is in δ form). The simulated evaporative fractionation assumes that the evaporation source is well mixed, and that the source water composition and all atmospheric conditions remain constant through the simulation time interval (i.e., one hour, or potentially less for wetlands and lakes).

Evapotranspiration

Evaporative fractionation is assumed to apply only to the evaporation portion of evapotranspiration. Therefore, the evapotranspiration flux calculated by WATFLOOD must be partitioned into evaporation and transpiration. The split between evaporation and transpiration depends on the water held in the upper zone of the soil, quantified by the upper zone storage indicator, *UZSI* in WATFLOOD which is a function of the soil moisture. Evapotranspiration in WATFLOOD ranges from the potential rate when the soil is saturated, to zero when water content reaches the permanent wilting point, *PWP.* The upper zone indicator is computed as (Neff, 1996):

$$UZSI = \left(\frac{UZS - PWP}{SAT - PWP}\right)^{1/2} \quad (E.9)$$

Where *UZSI* is the upper zone storage and *SAT* is the saturated soil storage.

As the soil dries, the transpiration fraction of evapotranspiration increases and the evaporation fraction decreases. The proportion of evaporation, *E*, to evapotranspiration, *ET*, is estimated using a power law function (Stadnyk-Falcone, 2008):

$$\frac{E}{ET} = a \cdot UZSI^b \quad (E.10)$$

Where *a* is the maximum evaporation to evapotranspiration fraction for the land class (typically between 0.3 and 0.5), and *b* is the rate at which the evaporative component of evapotranspiration increases with soil water content. Evapotranspiration in connected wetlands is not partitioned in this way, as they do not have an upper zone, or *UZSI*. Instead, the *a* value is used as a constant value for E/ET.

Snow

The isotope volume added to the snowpack, I_{SNW}^{iso}, is a function of the precipitation over the time interval:

$$I_{SNW}^{iso} = \left(C_{RS}^{iso} q_{RS} + C_S^{iso} q_S\right)\Delta t \quad (E.11)$$

Where C_S^{iso} and C_{RS}^{iso} are the isotopic concentrations of snow and rain and q_S and q_{RS} are the inflows from snow and rain on the snowpack. The water volume removed from the snowpack, Q_{SNW}, includes snowmelt, q_{melt}, and sublimation, q_{sub}:

$$Q_{SNW} = (q_{melt} + q_{sub})\Delta t \quad (E.12)$$

The volume of water stored in the snowpack at the end of the time interval, $S_{SNW,2}$, is known because WATFLOOD`s calculations for the time interval are completed prior to the isotope module run. The volume of isotopes in the snowpack at the end of the time interval, $S_{SNW,2}^{iso}$, can therefore be directly computed as:

$$S_{SNW,2}^{iso} = \frac{S_{SNW,1}^{iso} + I_{SNW}^{iso}}{1 + Q_{SNW}/S_{SNW,2}} \quad (E.13)$$

Where $S_{SNW,1}^{iso}$ is the volume of isotopes in the snowpack at the beginning of the time interval. The concentration of isotopes in the snowpack, C_{SNW}^{iso}, is calculated using the storage volumes at the end of the time interval:

$$C_{SNW}^{iso} = \frac{S_{SNW,2}^{iso}}{S_{SNW,2}} \quad (E.14)$$

The concentration of the snowpack is also the concentration of the snowmelt and sublimation fluxes removed from the snowpack based on the assumption that fractionation from melt and sublimation is negligible. The snowpack composition is constant through the pack depth, and through the time interval, based on an assumption of instant and complete mixing of the snowpack through its depth. While this is not physically possible, as the pack is solid water and does not readily mix, the implication of this assumption is that melt and sublimation occur evenly through the entire snowpack depth.

Surface

The isotope volume added to the surface compartment, I_{SW}^{iso}, includes rain on bare ground, q_R, snowmelt, q_{melt}, and glacial melt, q_{gl}:

$$I_{SW}^{iso} = \left(C_R^{iso} q_R + C_{SNW}^{iso} q_{melt} + C_{gl}^{iso} q_{gl}\right)\Delta t \quad (E.15)$$

Where C_R^{iso} is the isotope concentration of rain, C_{SNW}^{iso} is the isotope concentration of snowmelt and C_{gl}^{iso} is the isotope concentration of glacial melt. Surface storage is highly transient, so evaporation is not included in the WATFLOOD model. Therefore, the volume of water removed from surface storage, Q_{SW}, includes only infiltration from bare and frozen soil, q_1 and q_{1fs}, and lateral runoff into the channel or wetlands from bare and frozen soil, q_{df} and q_{dffs} :

$$Q_{SW} = \left(q_1 + q_{1fs} + q_{df} + q_{daffs}\right)\Delta t \quad (E.16)$$

The volume of water in the surface compartment at the end of the time interval, $S_{SW,2}$, is known from the calculations completed in WATFLOOD before the isotope module is run. Two separate cases exist for the surface isotope mass balance, depending on the water volume retained in the surface compartment. If the water volume is greater than zero, the isotope volume in the surface compartment at the end of the time interval, $S^{iso}_{SW,2}$, is calculated using the isotope volume in the surface compartment at the beginning of the time interval, $S^{iso}_{SW,1}$:

$$S^{iso}_{SW,2} = \frac{S^{iso}_{SW,1} + I^{iso}_{SW}}{1 + Q_{SW}/S_{SW,2}} \quad (E.17)$$

The concentration of the outflow from surface storage, C^{iso}_{SW}, is then:

$$C^{iso}_{SW} = \frac{S^{iso}_{SW,2}}{S_{SW,2}} \quad (E.18)$$

In the case where there is no water volume remaining at the end of the time interval (i.e., outflow exceeds storage), $S_{SW,2}$ will be equal to zero and the isotope volume will be:

$$S^{iso}_{SW,2} = 0 \quad (E.19)$$

However, the outflow from the surface compartment may still be positive, either from the infiltration of old storage, new inflows, or some combination of the two. The concentration of the outflow from surface storage is then:

$$C^{iso}_{SW} = \frac{S^{iso}_{SW,1} + I^{iso}_{SW}}{S_{SW,1} + I_{SW}} \quad (E.20)$$

Where $S_{SW,1}$ is the water volume in the surface compartment at the beginning of the time interval, and I_{SW} is the volume of water added over the time interval:

$$I_{SW} = (q_R + q_{melt} + q_{gl})\Delta t \quad (E.21)$$

The surface water isotope model assumes instant and complete mixing of the surface water and constant isotopic composition within the hourly time interval. These assumptions are reasonable, as the surface water storage volume is small and very shallow and likely turbulent, such that any new inputs (which would be of constant composition across the GRU area) would be mixed throughout the surface water depth immediately.

Upper zone

The isotope volume entering the upper zone, I_{UZ}^{iso}, includes both infiltration from bare soil, q_{df}, and from frozen soil, q_{daffs}:

$$I_{UZ}^{iso} = C_{SW}^{iso}(q_{df} + q_{daffs})\Delta t \quad (E.22)$$

The isotopic concentration of the inflow, C_{SW}^{iso}, is the concentration of the surface storage. The volume of water moved from the upper zone to other compartments, Q_{UZ}, is the sum of vertical drainage and lateral interflow:

$$Q_{UZ} = (q_{int} + q_{intfs} + q_{drng} + q_{drngfs})\Delta t \quad (E.23)$$

Where q_{int} and q_{intfs} are interflow from bare and frozen soil, and q_{drng} and q_{drngfs} are drainage from bare and frozen soil. The volume of water both transpired, T_{UZ}, and evaporated, E_{UZ}, from the upper zone over the time interval are also removed. Transpiration is assumed to be non-fractionating in isoWATFLOOD, with transpired water having the same isotopic concentration as the upper zone storage. Evaporated water is given the isotopic concentration $C_{UZ,E}^{iso}$. The volume of isotopes in the upper zone at the end of the time interval, $S_{UZ,2}^{iso}$, can be calculated directly, since all other variables; including $S_{UZ,1}^{iso}$, the volume of isotopes in the upper

zone at the beginning of the time interval and $S_{UZ,2}$, the water volume in the upper zone at the end of the time interval; are known:

$$S_{UZ,2}^{iso} = \frac{S_{UZ,1}^{iso} + I_{UZ}^{iso} - C_{UZ,E}^{iso} E_{UZ}}{1 + (Q_{UZ} + T_{UZ})/S_{UZ,2}} \quad (E.24)$$

The isotopic concentration of the drainage and interflow depends on the concentration of the upper zone storage:

$$C_{UZ}^{iso} = \frac{S_{UZ,2}^{iso}}{S_{UZ,2}} \quad (E.25)$$

The upper zone isotope model also assumes instant and complete mixing through the upper zone depth in a given time step (i.e., 1 hour), such that the source water for the evaporated water isotope composition is fully mixed. This assumption is unrealistic when new water is added to the storage in a given time step since the rate of movement through the soil will be less than needed for complete mixing. The isotopic composition of the upper zone storage and fluxes are assumed to be constant through the one hour simulation time interval, and meteorological conditions affecting evaporative fractionation are also assumed to be constant through this period. The isotopic composition of the lower zone storage is assumed to have no effect on the upper zone composition.

It is also assumed that there will always be water in the upper zone (at least, based on WATFLOOD, the PWP depth). In rare cases, when there is no precipitation for extended periods, low upper zone retention and high interflow coefficients, this may be violated, resulting in no simulated values from then on (i.e., simulation failure).

Lower zone

All drainages originating in upper zone storages, including bare soil drainage, q_{drng}, and frozen soil drainage, q_{drngfs}, are included in the isotope volume added to the lower zone, I_{LZ}^{iso}:

$$I_{LZ}^{iso} = \sum_{i=1}^{c} C_{UZ,i}^{iso}(q_{drng,i} + q_{drngfs,i})\Delta t \quad (E.26)$$

Where c is the number of land classes with upper zone storages and C_{UZ}^{iso} is the isotopic concentration of the upper zone. The only outflow from the lower zone is the baseflow, q_{lz}, so the volume of water removed from the lower zone storage, Q_{LZ}, is:

$$Q_{LZ} = q_{lz}\Delta t \quad (E.27)$$

The volume of isotopes in the lower zone at the beginning of the time interval, $S_{LZ,1}^{iso}$, and the volume of water in the lower zone at the end of the time interval, $S_{LZ,2}$, are known from WATFLOOD, so the volume of isotopes in the lower zone storage can be calculated directly:

$$S_{LZ,2}^{iso} = \frac{S_{LZ,1}^{iso} + I_{LZ}^{iso}}{1 + Q_{LZ}/S_{LZ,2}} \quad (E.28)$$

The baseflow concentration, C_{LZ}^{iso}, depends on the lower zone storage at the end of the time interval:

$$C_{LZ}^{iso} = \frac{S_{LZ,2}^{iso}}{S_{LZ,2}} \quad (E.29)$$

The lower zone storage is assumed to be instantly and completely mixed regardless of the storage volume, and the isotopic composition of the lower zone storage and outflow are assumed to be unvarying through the one hour simulation time interval. The instant and completing mixing assumption for the lower zone will be less accurate in the timestep where new inflows are added to the stored volume, as the groundwater volume can be very large.

Connected wetland

The first step for calculation the isotopic composition of the connected wetland is to find the isotopic concentration of the wetland without evaporative enrichment. The inflow and outflow from wetlands depend on the direction of flow between the wetland and the channel. If the flow, q_{WET}, is positive, water is moving from the wetland to the channel. The volume of isotopes added to the wetland storage over the current time interval, $I^{iso}_{WET,2}$, includes isotopes from direct rainfall, q_R, snowmelt from the wetland snowpack, q_{melt}, interflow from bare and frozen soil, q_{int} and q_{intfs}, direct runoff from bare and frozen soil, q_1 and q_{1fs}, and baseflow, q_{lz}:

$$I^{iso}_{WET,2} = \left(C^{iso}_R q_R + C^{iso}_{SNW} q_{melt}\right)\Delta t + \sum_{i=1}^{c}\left(C^{iso}_{UZ,i}\left(q_{int,i} + q_{intfs,i}\right) + C^{iso}_{SW,i}\left(q_{1,i} + q_{1fs,i}\right)\right)\Delta t + C^{iso}_{LZ} q_{lz}\Delta t - C^{iso}_{WET,1}(E_{WET} + T_{WET})\Delta t \quad \textit{(E.30)}$$

Where c is the number of terrestrial land classes, C^{iso}_R is the isotopic concentration of rainfall, C^{iso}_{SNW} is the isotopic concentration of snowmelt, C^{iso}_{UZ} is the isotopic concentration of the upper zone, C^{iso}_{SW} is the isotopic concentration of surface water, and C^{iso}_{LZ} is the isotopic concentration of the lower zone. Evaporation, E_{WET}, and transpiration, T_{WET}, from the wetland are removed from the inflow; both are assumed to have the isotope concentration of the wetland storage at the beginning of the time interval, $C^{iso}_{WET,1}$ in the initial mass balance. Because evapotranspiration is removed from the inflow, the volume of water removed from the wetland storage over the current time interval, $Q_{WET,2}$, depends only on the flow into the channel:

$$Q_{WET,2} = q_{WET}\Delta t \quad \textit{(E.31)}$$

WATFLOOD calculates the storage in connected wetlands using the inflow and outflow from the current and the previous time interval. The isotope storage calculation must match the method used for water, but because the water mass balance has already been calculated for the time interval, the storage of isotopes in the wetland at the end of the time interval, $S^{iso}_{WET,2}$, can be calculated directly:

$$S^{iso}_{WET,2} = \frac{S^{iso}_{WET,1} + \left(I^{iso}_{WET,1} + I^{iso}_{WET,2} - Q^{iso}_{WET,1}\right)/2}{1 + Q_{WET,2}/2S_{WET,2}} \quad (E.32)$$

where $S^{iso}_{WET,1}$ is the storage of isotopes in the wetland at the beginning of the time interval, $I^{iso}_{WET,1}$ is the volume of isotopes added to the wetland in the previous time interval, $Q^{iso}_{WET,1}$ is the volume of isotopes removed from the wetland in the previous time interval, and $S_{WET,2}$ is the storage of water in the wetland at the end of the time interval.

If the flow, q_{WET}, is negative, water is moving from the channel into the wetland. In this case, the volume added to the wetland also includes the wetland flow, which is subtracted since it is negative:

$$\begin{aligned} I^{iso}_{WET,2} = {} & \left(C^{iso}_{R} q_R + C^{iso}_{SNW} q_{melt}\right)\Delta t \\ & + \sum_{i=1}^{c} \left(C^{iso}_{UZ,i}\left(q_{int,i} + q_{intfs,i}\right) + C^{iso}_{SW,i}\left(q_{1,i} + q_{1fs,i}\right)\right)\Delta t + C^{iso}_{LZ} q_{lz}\Delta t \\ & - C^{iso}_{WET,1}(E_{WET} + T_{WET})\Delta t - C^{iso}_{STR,1} q_{WET}\Delta t \end{aligned} \quad (E.33)$$

The flow from the channel is assumed to have the same concentration as the streamflow in the previous time interval, $C^{iso}_{STR,1}$. Both evapotranspiration and the wetland-channel flow are included in $I^{iso}_{WET,2}$, therefore:

$$Q_{WET,2} = 0 \qquad (E.34)$$

The equation for $S^{iso}_{WET,2}$ is then simplified to:

$$S^{iso}_{WET,2} = S^{iso}_{WET,1} + \left(I^{iso}_{WET,1} + I^{iso}_{WET,2} - Q^{iso}_{WET,1}\right)/2 \qquad (E.35)$$

For either case the current isotopic concentration of the connected wetland storage, $C^{iso}_{WET,2}$, is:

$$C^{iso}_{WET,2} = \frac{S^{iso}_{WET,2}}{S_{WET,2}} \qquad (E.36)$$

The second step, which is only needed if there is evaporation occurring over the time interval, is to calculate the enrichment of the wetland due to evaporation. The isotopic concentration in delta format of an evaporating water body is calculated using the isotopic concentration of the connected wetland in delta format, $\delta_{WET,2}$:

$$\delta_{WET,E} = \delta_S - \left(\delta_S - \delta_{WET,2}\right)\left(\frac{S_{WET,2}}{S_{WET,1}}\right)^{-\left(1+m\frac{E}{I}\right)/\left(1-\frac{E}{I}-\frac{Q}{I}\right)} \qquad (E.37)$$

Where E is the volume of water evaporated over the time interval, I is the volume of water added to the wetland over the time interval from land outflows, precipitation, snowmelt and flow from the channel and Q is the volume of water removed from the wetland either into the stream channel or through transpiration. $S_{WET,2}$ is the volume of water at the end of the time interval, and $S_{WET,1}$ is the water volume at the beginning of the time interval. The value m is:

$$m = \frac{h - \varepsilon^*/\alpha^* - \varepsilon_k}{1 - h - \varepsilon_k} \qquad (E.38)$$

Where h is the fractional relative humidity, and ε^*, α^* and ε_k are calculated from equations E.2 to E.5. δ_S is the steady state isotopic concentration of the evaporating water body:

$$\delta_S = \frac{\delta_I + m\frac{E}{I}\delta^*}{1 + m\frac{E}{I}} \quad (E.39)$$

Where δ_I is the isotopic concentration of *I* in delta format, and δ^* is calculated from equation E.8.

The enriched isotopic concentration is then used to update the isotope volume in storage, and the isotope concentration of the storage:

$$S^{iso}_{WET,2} = C^{iso}_{WET,E} S_{WET,2} \quad (E.40)$$

$$C^{iso}_{WET,2} = C^{iso}_{WET,E} \quad (E.41)$$

Regardless of whether the wetland was enriched, the isotope outflow volume $Q^{iso}_{WET,2}$ is calculated for the mass balance calculation in the subsequent time interval:

$$Q^{iso}_{WET,2} = C^{iso}_{WET,2} Q_{WET,2} \quad (E.42)$$

The connected wetland isotope model assumes that all inflows, whether precipitation or groundwater, are instantly and completely mixed into the wetland storage. Meteorological conditions affecting evaporative fractionation and the isotopic composition of the wetland outflow are assumed to be constant over the simulation time interval, which may be hourly or sub-hourly depending on the routing time interval.

Channel

The flow of isotopes matches that of water, with the change in storage dependent on the inflow and outflow at the beginning and end of the time interval. There are three possible cases for the calculation of the incoming volume of isotopes for the current time interval, $I^{iso}_{STR,2}$. In the most

common case, $I^{iso}_{STR,2}$ includes all isotopes coming from upstream grids, $Q^{iso}_{STR,2,U/S}$, and inflows from the connected wetland, rainfall and snowmelt, with evaporation removed:

$$I^{iso}_{STR,2} = \sum Q^{iso}_{STR,2,U/S} + \left(C^{iso}_{WET,2} q_{WET} + C^{iso}_{R} q_R + C^{iso}_{SNW} q_{melt} - C^{iso}_{STR,E} E_{STR}\right)\Delta t \quad (E.43)$$

Where $C^{iso}_{WET,2}$ is the isotopic concentration of the wetland at the end of the time interval, q_{WET} is the flow between the channel and wetland, C^{iso}_{R} is the isotopic concentration of rainfall, q_R is the rainfall over the channel, C^{iso}_{SNW} is the isotopic concentration of snowmelt, q_{melt} is the snowmelt from the channel snowpack, $C^{iso}_{STR,E}$ is the isotopic concentration of the evaporating moisture and E_{STR} is the volume of water evaporated from the channel over the time interval. The flow from the connected wetland can also be negative, such that channel water is transferred to the connected wetland, with the wetland flow having the concentration of the channel at the beginning of the time interval:

$$I^{iso}_{STR,2} = \sum Q^{iso}_{STR,2,U/S} + \left(C^{iso}_{STR,1} q_{WET} + C^{iso}_{R} q_R + C^{iso}_{SNW} q_{melt} - C^{iso}_{STR,E} E_{STR}\right)\Delta t \quad (E.44)$$

Where $C^{iso}_{STR,1}$ is the isotopic concentration of the channel at the beginning of the time interval. Finally, if connected wetlands are not used in the WATFLOOD model, water from the land area is added to the channel directly:

$$I^{iso}_{STR,2} = \sum Q^{iso}_{STR,2,U/S} + \sum_{i=1}^{c} \left(C^{iso}_{UZ,i}\left(q_{int,i} + q_{intfs,i}\right) + C^{iso}_{SW,i}\left(q_{1,i} + q_{1fs,i}\right)\right)\Delta t + C^{iso}_{LZ} q_{lz}\Delta t + \left(C^{iso}_{R} q_R + C^{iso}_{SNW} q_{melt} - C^{iso}_{STR,E} E_{STR}\right)\Delta t \quad (E.45)$$

Where c is the number of terrestrial land classes, q_{int} and q_{intfs} are interflow from bare and frozen soil, C^{iso}_{UZ} is the isotopic concentration of the upper zone, C^{iso}_{SW} is the isotopic concentration of surface water, q_1 and q_{1fs} are direct runoff from bare and frozen soil, C^{iso}_{LZ} is the

isotopic concentration of the lower zone and q_{lz} is baseflow. Only the outflow to the downstream grid, q_o, is included in the outflow volume in the current time interval, $Q_{STR,2}$:

$$Q_{STR,2} = q_o \Delta t \quad \textit{(E.46)}$$

The isotope storage in the channel at the end of the time interval, $S^{iso}_{STR,2}$, is calculated from the isotope storage at the beginning of the time interval, $S^{iso}_{STR,1}$; the isotope volume added in the current and previous time intervals, $I^{iso}_{STR,1}$ and $I^{iso}_{STR,2}$; the isotope volume removed in the previous time interval, $Q^{iso}_{STR,1}$; and the water stored in the channel at the end of the time interval, $S_{STR,2}$:

$$S^{iso}_{STR,2} = \frac{S^{iso}_{STR,1} + \left(I^{iso}_{STR,1} + I^{iso}_{STR,2} - Q^{iso}_{STR,1}\right)/2}{1 + Q_{STR,2}/2S_{STR,2}} \quad \textit{(E.47)}$$

The concentration of isotopes in the channel storage and outflow is:

$$C^{iso}_{STR,2} = \frac{S^{iso}_{STR,2}}{S_{STR,2}} \quad \textit{(E.48)}$$

The volume of isotopes removed in the current time interval is also calculated for the mass balance in the subsequent interval:

$$Q^{iso}_{STR,2} = C^{iso}_{STR,2} Q_{STR,2} \quad \textit{(E.49)}$$

All channel storage in a grid cell is assumed to be instantly and completely mixed, and the outflow composition is assumed to be constant for the simulation time interval, which may be hourly or sub-hourly depending on the routing time interval.

Lake or reservoir

Evaporative fractionation is modeled using the equations for the time-dependent model lakes presented in Gibson (2002). Firstly, the isotopic concentration of the lake prior to evaporative enrichment is calculated from the mass balance in the lake. The isotope volume added into the

lake in the current time interval, $I_{LK,2}^{iso}$; including upstream flows, $Q_{STR,2,U/S}^{iso}$; and flows generated from land in the lake grids, snowmelt, and rainfall:

$$I_{LK,2}^{iso} = \sum Q_{STR,2,U/S}^{iso} + \sum_{r=1}^{n}\left(\sum_{i=1}^{c}\left(C_{UZ,i,r}^{iso}\left(q_{int,i,r} + q_{intfs,i,r}\right) + C_{SW,i,r}^{iso}\left(q_{1,i,r} + q_{1fs,i,r}\right)\right)\Delta t + C_{LZ,r}^{iso} q_{lz,r}\Delta t\right) + \left(C_R^{iso} q_R + C_{SNW}^{iso} q_{melt} - C_{LK,1}^{iso} E_{LK}\right)\Delta t \quad (E.50)$$

Where c is the number of non-water land classes, n is the number of grids in the lake, q_{int} and q_{intfs} are interflow from bare and frozen soil, C_{UZ}^{iso} is the isotopic concentration of the upper zone, C_{SW}^{iso} is the isotopic concentration of surface water, q_1 and q_{1fs} are direct runoff from bare and frozen soil, C_{LZ}^{iso} is the isotopic concentration of the lower zone, q_{lz} is baseflow, C_R^{iso} is the isotopic concentration of rainfall, q_R is the rainfall over the lake, C_{SNW}^{iso} is the isotopic concentration of snowmelt and q_{melt} is the snowmelt from the lake snowpack. Evaporation over the time interval, E_{LK}, is assigned the concentration of the lake at the beginning of the time interval, $C_{LK,1}^{iso}$, in the initial mass balance. Only the outflow from the lake to the downstream receiving grid, q_o, is included in $Q_{LK,2}$, the volume of water removed from the lake storage over the current time interval:

$$Q_{LK,2} = q_o\Delta t \quad (E.51)$$

The isotope storage in the lake at the end of the time interval, $S_{LK,2}^{iso}$ is first calculated from the isotope storage at the beginning of the time interval, $S_{LK,1}^{iso}$, the isotope volume added in the current and previous time interval, $I_{LK,1}^{iso}$ and $I_{LK,2}^{iso}$, the isotope volume removed in the previous time interval, $Q_{LK,1}^{iso}$, and the water stored in the lake at the end of the time interval, $S_{LK,2}$:

$$S_{LK,2}^{iso} = \frac{S_{LK,1}^{iso} + \left(I_{LK,1}^{iso} + I_{LK,2}^{iso} - Q_{LK,1}^{iso}\right)/2}{1 + Q_{LK,2}/2S_{LK,2}} \quad (E.52)$$

The isotope concentration of the lake without evaporative enrichment is then calculated as:

$$C_{LK,2}^{iso} = \frac{S_{LK,2}^{iso}}{S_{LK,2}} \quad (E.53)$$

If any evaporation occurs over the time interval, a new enriched isotope concentration in delta notation, $\delta_{LK,E}$, is calculated using the equations from Gibson (2002):

$$\delta_{LK,E} = \delta_S - \left(\delta_S - \delta_{LK,2}\right)e^{-\left(1+m\frac{E}{I}\right)\frac{It}{V}} \quad (E.54)$$

Where E is the volume of water evaporated over the time interval, I is the volume of water added to the lake over the time interval from land outflows, precipitation, snowmelt and flow from upstream and V is the volume of water in the lake at the beginning of the time interval. This equation assumes that the lake volume is constant over the simulation time interval. $\delta_{LK,2}$ is the isotope concentration of the lake without evaporative enrichment in delta notation. The value m is defined by Welhan and Fritz (1977) (see equation E.37). The enriched isotopic concentration is then used to update the isotope volume in storage, and the isotope concentration of the storage:

$$S_{LK,2}^{iso} = C_{LK,E}^{iso} S_{LK,2} \quad (E.55)$$

$$C_{LK,2}^{iso} = C_{LK,E}^{iso} \quad (E.56)$$

Regardless of whether the lake was enriched, the isotope volume removed in the current time interval, $Q_{LK,2}^{iso}$ is calculated for the mass balance calculation in the subsequent time interval:

$$Q_{LK,2}^{iso} = C_{LK,E}^{iso} Q_{LK,2} \quad (E.57)$$

The lake isotope simulation assumes instant and complete mixing of the lake, and constant isotopic compositions for the lake storage and outflow over the simulation time interval, which

may be hourly or sub-hourly depending on the routing time interval. The meteorological conditions affecting evaporative fractionation are assumed to be constant over the simulation time interval, and evaporation is assumed to have no effect on the isotopic composition of atmospheric moisture.