



# Examining the impacts of estimated precipitation isotope ( $\delta^{18}\text{O}$ ) inputs on distributed tracer-aided hydrological modelling

Carly J. Delavau<sup>1</sup>, Tricia Stadnyk<sup>1</sup>, Tegan Holmes<sup>1</sup>

<sup>1</sup>Department of Civil Engineering, University of Manitoba, Winnipeg, R3T 5V6, Canada.

Correspondence to: Carly J. Delavau (Carly.Delavau@gov.mb.ca)

**Abstract.** Tracer-aided hydrological models are becoming increasingly popular tools as they have documented utility in constraining model parameter space during calibration, reducing model uncertainty, and assisting with selection of appropriate model structures. However, the issue of data availability, particularly input data, proves to be a major challenge associated with this type of application. Tracer-aided hydrological modelling typically requires a time series of isotopes in precipitation ( $\delta^{18}\text{O}_{\text{ppt}}$ ) to drive model simulations, but unfortunately, throughout much of the world, and particularly in sparsely populated high-latitude regions, these data are not widely available. This study uses the isoWATFLOOD tracer-aided hydrological model to investigate the usefulness of three types of estimated  $\delta^{18}\text{O}_{\text{ppt}}$  for model input, and the impact that these data have on model simulations and parameterization in the remote Fort Simpson Basin, NWT, Canada. This study showed that although total simulated streamflow was not significantly impacted by choice of  $\delta^{18}\text{O}_{\text{ppt}}$  input, isotopes in streamflow ( $\delta^{18}\text{O}_{\text{SF}}$ ) simulations and the internal apportionment of water (and therefore, model parameterizations) were impacted, particularly during large precipitation and snowmelt events. This finding highlighted the importance of estimated  $\delta^{18}\text{O}_{\text{ppt}}$  to capture both the variability and seasonality in precipitation isotopes as critical for tracer-aided hydrological modelling, especially when precipitation events displayed distinctly different isotopic compositions than that of streamflow. This study achieves an understanding of how isoWATFLOOD can be used in regions with a limited number of  $\delta^{18}\text{O}_{\text{ppt}}$  observations, and that the model can be of value in such regions. This study reinforces that a tracer-aided modelling approach assists with resolving hydrograph component contributions, and works towards diagnosing the issue of model equifinality.

## 1 Introduction

Hydrological models are critical tools for the planning, development, design, operation and sustainable management of water resources (Singh and Frevert, 2006). These models provide insight into applications such as the prediction of floods, droughts and water availability, and the effects of climate and land use change on water resources. Problems arise in calibrating and validating hydrological model predictions, primarily due to: (1) a lack of available data at sufficient resolutions to force and validate model simulations - especially in remote, high-latitude locations (in Canada: Coulibaly et al., 2013); (2) model parameterization and issues with equifinality; and (3) the associated uncertainty in model results (Beven and Binley, 1992; Kirchner, 2006; Fenicia et al., 2008; Dunn et al., 2008).



It is now widely accepted that calibration and validation of hydrological models based solely on streamflow is not a sufficient evaluation measure (Kuczera, 1983; Beven and Binley, 1992; Kuczera and Mroczkowski, 1998; Seibert and McDonnell, 2002; Kirchner, 2006; Fenicia et al., 2008; Dunn et al., 2008). Modellers are focusing on a model's ability to correctly partition, store and release water from hydrologic compartments, in addition to adequately simulating total streamflow response. Conservative tracer data provides insights into the dominant hydrological processes and integrated runoff response (in northern catchments: Birks and Gibson, 2009; Tezloff et al., 2015), and such data assist with constraining model parameter space during calibration, reducing model uncertainty, and assisting with selection of appropriate model structures (e.g., Tetzlaff et al., 2008; Birkel et al., 2010a; McMillian et al., 2012; Birkel et al., 2014; Smith et al., 2016). An increasing number of studies have investigated the utility of tracer-aided modelling approaches, especially over the past decade (for a comprehensive overview, see Birkel and Soulsby, 2015).

Although greatly informative, previous tracer-aided modelling studies have generally been conducted using lumped conceptual rainfall-runoff models in highly instrumented small-scale experimental catchments ( $<10^2$  km<sup>2</sup>). This has resulted in distributed studies at the regional-scale ( $>10^3$  km<sup>2</sup>) left largely unexplored, with the exception of a few, select applications (Stadnyk et al., 2013). Modelling at the regional-scale typically requires a distributed approach to capture the heterogeneity in meteorological inputs, basin characteristics, and runoff response, resulting in more complex, highly parameterized models (e.g., Michaud and Sorooshian, 1994; Carpenter and Georgakakos, 2006; Her and Chaubey, 2015). Because it is at these larger scales where models are applied operationally and management decisions are based, there is a critical need to understand the abilities, limitations, and uncertainties associated with distributed tracer-aided modelling at the regional scale.

Although there is an identified need, the issue of data availability, particularly input data, proves to be a major challenge associated with this type of application (Birkel and Soulsby, 2015). Tracer-aided hydrological modelling typically requires a time series of isotopes in precipitation ( $\delta^{18}\text{O}_{\text{ppt}}$ ) to drive model simulations. Unfortunately, throughout much of the world, and particularly in sparsely populated high-latitude regions (such as the vast majority of Canada), these data are not widely available. Although automatic samplers are becoming increasingly common, watersheds in which snow accumulation is substantial will continue to be fraught with difficulties surrounding the collection and characterization of precipitation isotopes, particularly during the winter months (Dietermann and Weiler, 2013; Penna et al., 2014). The lack of spatial and temporal density of  $\delta^{18}\text{O}_{\text{ppt}}$  observations highlights the need for alternative methods to provide estimates of stable isotopes in precipitation for tracer-aided model forcing. Options include empirically-based models generating gridded time series estimates of precipitation isotopes (e.g., Lykoudis et al., 2010; Delavau et al., 2015), in addition to isotope-enabled climate model output (for a comprehensive overview, see: Noone and Sturm, 2010; Xi et al., 2014).



Small-scale catchment studies rely on continuous records of  $\delta^{18}\text{O}_{\text{ppt}}$  observations at high temporal frequencies (typically daily, and less commonly, weekly) for model input. At the larger scale, tracer-aided modelling completed by Stadnyk et al. (2013) in the remote Fort Simpson region of northern Canada used annual average compositions of rainfall and snowfall  $\delta^{18}\text{O}$  to drive model simulations. Their results suggested that utilizing annual, spatially static oxygen-18 in precipitation forcing has the potential to significantly impact simulations and consequently, model parameterization as well. The assumption that model forcing is spatially invariant is not preferable, as  $\delta^{18}\text{O}_{\text{ppt}}$  can vary drastically over small space and time scales due to changes in moisture sources and transport processes, rainout history and seasonality (e.g., in Canada: Gat et al., 1994; Moran et al., 2007; Birks and Edwards, 2009).

Utilizing estimates of  $\delta^{18}\text{O}_{\text{ppt}}$  to drive tracer-aided model simulations has not yet been investigated, particularly at the regional-scale where the lack of observed data have identified a need for estimates of  $\delta^{18}\text{O}_{\text{ppt}}$ . Therefore, this study aims to explore how estimated  $\delta^{18}\text{O}_{\text{ppt}}$  inputs of differing spatial and temporal variability impact regional tracer-aided modelling within two remote Canadian basins, where limited  $\delta^{18}\text{O}_{\text{ppt}}$  observations exist. Our specific research objectives are to:

- a) force a regional tracer-aided hydrological model (isoWATFLOOD) with three types of estimated  $\delta^{18}\text{O}_{\text{ppt}}$  input of varying temporal and spatial resolution;
- b) examine how choice of  $\delta^{18}\text{O}_{\text{ppt}}$  input affects the simulation of total streamflow and isotopes in streamflow ( $\delta^{18}\text{O}_{\text{SF}}$ ), while accounting for parameter uncertainty, and;
- c) explore if choice of  $\delta^{18}\text{O}_{\text{ppt}}$  input affects the internal apportionment of water, namely the seasonality of hydrograph components.

## 2 Study area and data

### 2.1 The Fort Simpson Basin

The Fort Simpson Basin (FSB) is located within the Lower Liard River valley close to the town of Fort Simpson, Northwest Territories, Canada (61°45 N; 121°14 W; Fig. 1). This region has been the focus of several tracer-aided hydrological studies (e.g., St Amour et al., 2005; Stadnyk et al., 2005; 2013; Stadnyk-Falcone, 2008). The FSB is selected for this study to build upon previous modelling work conducted within the region, and follow up on recommendations from Stadnyk et al. (2013) suggesting further analysis and improvement of isoWATFLOOD  $\delta^{18}\text{O}_{\text{ppt}}$  input. The study period of 1997–1999 is selected based data availability.

This study considers two sub-basins of the greater Fort Simpson basin: the Jean-Marie and Blackstone River sub-basins (Fig. 1). Differences in wetland distribution and function, basin physiography and land cover make-up between the two watersheds (Table 1) are the primary reasons in selecting these sub-basins for this study. These marked differences ensure



that watersheds of varying dominant hydrological processes are represented in the modelling, and therefore the impacts of  $\delta^{18}\text{O}_{\text{ppt}}$  input selection on these processes can be examined.

The land cover classification breakdown (Table 1) shows the primary land cover type within the sub-basins as transitional, consisting of shrubs, deciduous varieties and early generation spruce. The region has a high proportion of wetlands, with the total wetland percentage in Table 1 representing both bogs (disconnected drainage) and fens (connected drainage); although the amount of each type within each respective sub-basin varies. Aylsworth and Kettles (2000) state that Jean-Marie is predominately fen peatlands, while Blackstone is bog-dominated peatlands, with very few or no fen peatlands present.

The Ecoregions Working Group (1989) classifies the FSB as a sub-humid mid- to high-boreal ecoclimatic region (Hbs), classified by cool summers approximately five months in length, with moderate (300-500 mm) annual precipitation. Winters are very cold with persistent snow cover. The hydrological response is dominated by snowmelt during late April to early May, while summer and fall runoff events are due to major rainfall, with a return to baseflow occurring during dry summer periods or towards the beginning of the ice-on season in October.

## 2.2 Meteorological and hydrometric data

Daily total precipitation, mean daily temperature, and hourly relative humidity data are obtained from Environment Canada's Fort Simpson Airport weather station. Observed precipitation is supplemented with ANUSPLIN-derived daily precipitation extracted at eight locations throughout the Fort Simpson region (Fig. 1). ANUSPLIN is a multidimensional non-parametric surface fitting method that has been found well suited to the interpolation of various climate variables, particularly in data-sparse, high-elevation regions as the method accounts for spatially varying dependencies on elevation (McKenney et al., 2011). An inverse-distance weighting approach is used spatially distribute the daily ANUSPLIN and observed precipitation time series across the model domain (Kouwen, 2014). Rainfall that occurred over the study period, particularly in 1997, was significantly higher than normal. Additionally, 1998 was above average in temperature, which is especially prevalent in the first portion of the year. Other researchers have attributed the increased rainfall and warmer temperatures to a strong El Niño influence from mid-1997 to mid-1998 (Petroni et al., 2000; St Amour et al., 2005).

Hydrometric records are obtained from Water Survey of Canada. Jean Marie was gauged at Highway No.1 in 1972 with a period of record of 44 years, whereas Blackstone was gauged at Highway No.7 in 1991 having a record length of 25 years. Neither sub-basin is regulated, therefore all flows are considered to be natural. During the study period, mean annual discharge was above normal in both sub-basins in 1997, normal in Jean Marie and slightly below normal in Blackstone in 1998, and below normal in both sub-basins in 1999. A statistical summary of observations used in this study is provided in Table 4.



### 2.3 Isotope data

During 1997 to 1999, intensive sampling took place in the Fort Simpson Basin as part of the Mackenzie Study of the Global Energy and Water Experiment (GEWEX; Stewart et al., 1998). The campaign sampled  $\delta^{18}\text{O}$  and  $\delta^2\text{H}$  of streamflow, rainfall, snowpack, and surface waters (wetlands and lakes) during the open water season (May to October). During ice-on  
5 conditions, the isotope stratigraphy of river ice extracted during late March in 1998 and 1999 was used to reconstruct the isotopic composition of winter streamflow (Gibson and Prowse, 1999; Prowse et al., 2002; St Amour et al., 2005). This study uses measured  $\delta^{18}\text{O}$  compositions in streamflow ( $\delta^{18}\text{O}_{\text{SF}}$ ) in the Jean-Marie ( $n = 71$ ) and Blackstone ( $n = 69$ ) sub-basins for model calibration. Although  $\delta^{18}\text{O}_{\text{ppt}}$  compositions ( $n = 27$ ) were collected as part of the GEWEX sampling campaign, their spatial and temporal resolutions are not adequate for model forcing. These data are incorporated into the  
10 study as a means to validate simulated  $\delta^{18}\text{O}_{\text{ppt}}$  forcing input, when available. The number of measurements and their statistical properties are summarized in Table 4. Isotopic compositions of  $\delta^{18}\text{O}$  are expressed in delta ( $\delta$ ) notation as a deviation from VSMOW (Vienna Mean Standard Mean Ocean Water) in units of per mille (‰), such that  $\delta_{\text{water}} = (R_{\text{water}}/R_{\text{VSMOW}} - 1) \times 1000$  ‰, where  $R$  is  $^{18}\text{O}/^{16}\text{O}$  in the sample and standard, respectively. Isotope samples were analyzed at the Environmental Isotope Laboratory at the University of Waterloo, and St Amour et al. (2005) indicated maximum  
15 analytical uncertainties of  $\pm 0.1$  ‰ for  $\delta^{18}\text{O}$ .

### 2.4 Precipitation oxygen-18 input

The default method for  $\delta^{18}\text{O}_{\text{ppt}}$  input in isoWATFLOOD is annual average compositions of rainfall and snowfall  $\delta^{18}\text{O}$  for each year of the simulation. Values for the FSB are obtained through averaged measurements of  $\delta^{18}\text{O}$  in rainfall and snowpack from the GEWEX study (Table 2; Table 4).  $\delta^{18}\text{O}_{\text{ppt}}$  compositions are assumed constant throughout the watershed  
20 domain. Due to the averaged values and lack of spatial variability, this input type is referred to as *static* throughout the remainder of the manuscript.

Times series simulations obtained from the KPN43 model created by Delavau et al. (2015) are used as a secondary  $\delta^{18}\text{O}_{\text{ppt}}$  input. The KP43 model uses North American Regional Reanalysis (NARR; Mesinger et al., 2006) climate variables, teleconnection indices, and geographic information to produce gridded time series of oxygen-18 in precipitation at a monthly  
25 time step. This input is generated at a 10 km resolution (to mirror model set-up), and varies spatially throughout the study area due to the variation in the climatic predictors and geographic information required to produce simulations.

The third  $\delta^{18}\text{O}_{\text{ppt}}$  input assessed in this study is regional climate model output from the isotope-enabled climate model, REMOiso (Sturm et al., 2005; Sturm et al., 2007). Raw REMOiso  $\delta^{18}\text{O}_{\text{ppt}}$  output is available at a 55 km spatial resolution and  
30 a 6h time step. However, REMOiso output is averaged to a daily time step, as the range and variability of sub-daily  $\delta^{18}\text{O}_{\text{ppt}}$



are erroneously large, and the resolution of streamflow oxygen-18 calibration data do not warrant a temporal frequency of input finer than daily.

#### 2.4.1 REMOiso bias correction

Due to a lack of published studies evaluating REMOiso performance within Canada, a comparison between REMOiso  
5 output and Canadian Network for Isotopes in Precipitation observations (CNIP; Birks and Gibson, 2009) is completed to  
determine if REMOiso simulations require a regional bias correction. CNIP data are now part of the Global Network for  
Isotopes in Precipitation (GNIP) database and can be accessed at: <http://www.iaea.org/water> (IAEA, 2014). This analysis is  
completed at Snare Rapids, NWT, the closest CNIP station to the FSB, for the years of 2000 and 2001. Snare Rapids is  
located approximately 330 km northeast of Fort Simpson and has monthly  $\delta^{18}\text{O}_{\text{ppt}}$  observations spanning the years of 1997–  
10 2010. A longer time frame of comparison between CNIP and REMOiso is not possible due to the short overlapping period of  
REMOiso simulations and CNIP observations. Daily REMOiso simulations are averaged to monthly compositions for direct  
comparison to CNIP data using the precipitation amount-weighting approach in Eq. (1):

$$\delta^{18}\text{O}_{\text{ppt monthly}} = \sum P_i * (\delta^{18}\text{O}_{\text{ppt}})_i / \sum P_i \quad (1)$$

15 where  $P_i$  is the amount of daily precipitation (mm) obtained from the Snare Rapids Canadian Air and Precipitation  
Monitoring Network (CAPMoN) station operated by Environment Canada, where isotopic compositions are also sampled.

Uncorrected REMOiso simulations exhibit a positive bias in this region (Fig. 2), which is expected based on the ECHAM4  
20 mean annual  $\delta^{18}\text{O}_{\text{ppt}}$  output (Noone and Sturm, 2010) and personal communications with S. J. Birks and K. Sturm (2016).  
Therefore, a seasonal bias correction is applied to daily REMOiso simulations. The bias correction is calculated as the  
average seasonal difference between the monthly amount-weighted REMOiso output and the CNIP observations (Table 3).  
Corrected monthly and daily REMOiso output at Snare Rapids are displayed on Fig. 2 as the dashed red and solid orange  
lines, respectively. For the current study, daily REMOiso output for the Fort Simpson region is bias corrected with the  
25 seasonal values in Table 3. The statistical properties of the corrected daily REMOiso simulations, alongside the KPN43  
monthly simulations and the static seasonal averages are summarized in Table 4.

### 3 Methods

#### 3.1 Background and set-up

The tracer-aided hydrological model used in this study is isoWATFLOOD (Stadnyk-Falcone, 2008; Stadnyk et al., 2013).  
30 isoWATFLOOD is an extension of the WATFLOOD hydrological model, whereby water and oxygen-18 are simultaneously  
budgeted throughout the modelled hydrologic cycle. WATFLOOD is a distributed model that uses grouped response units



(GRUs) to simulate streamflow in hydrologically-distinct land cover units (Kouwen et al., 1993; Kouwen, 2014). Process representation within WATFLOOD is considered to be a combination of both conceptual and physical, as certain algorithms are conceptually-based (e.g., evaporation and snowmelt), while others are more based in physics (e.g., channel routing). Due to the coupling of isotopes to each hydrological processes simulated in WATFLOOD, simulation of isotopic composition does not introduce any additional parameters. A more comprehensive description of isoWATFLOOD's model structure and governing equations can be found in Stadnyk et al. (2013) and select descriptions are provided in Table 6.

isoWATFLOOD requires  $\delta^{18}\text{O}$  of rainfall and snowfall and hourly distributed relative humidity to force the model. Additionally,  $\delta^{18}\text{O}$  compositions for hydrologic storages of river/fen water, soil water, baseflow, and snowpack are needed for model initialization, which can be obtained from field data or estimated. Here, regional isotopic initializations are derived from measured data (St Amour et al., 2005), and are summarized in Table 5. Sensitivity analyses have shown that within one month of simulation isoWATFLOOD spin-up is complete and, past this point, initialization values have no bearing on model output. All other data required by isoWATFLOOD (e.g., distributed precipitation, temperature, evaporation, inflows, etc.) are passed from WATFLOOD forcings or computations.

The watershed model set-up in this study is based off the version used by Stadnyk et al. (2013), with several changes and improvements. Based on findings from Aylsworth and Kettles (2000), we implemented a 90 % bog and 10 % fen split in Blackstone and a 30 % bog and 70 % fen split in Jean-Marie. The entirety of the FSB is modelled at a 10 km spatial resolution, and the model is run continuously from January 1996 to December 1999; whereby 1996 is utilized as spin-up to set initial hydrologic and isotopic storage conditions.

### 3.2 Calibration and parameter uncertainty

Being a distributed model, WATFLOOD has a large number of parameters requiring calibration. For this reason, a sensitivity analysis is first conducted to identify which parameters have the largest influence on both streamflow and  $\delta^{18}\text{O}_{\text{SF}}$ . A subset of parameters are identified for inclusion in the calibration based on this sensitivity analysis, including nine hydrological parameters from each of the five most prominent land classes (mixed/deciduous, coniferous, transit, bogs and fens), and four routing parameters from each of the two modelled sub-basins. This results in 53 parameters that are incorporated in the parameter uncertainty assessment (Table 6; Table 8). Allowable ranges for each parameter are determined based on published values alongside personal communications with N. Kouwen (Kouwen, 2014) (Table 8).

This study uses a multi-criteria, multi-objective approach to model calibration, with the procedure summarized as follows:

- i. A Monte Carlo random sampling approach, assuming uniform parameter distributions, is used to individually select each parameter from its allowable range (Table 8). Random parameter sampling is completed 30,000 times, generating 30,000 unique parameter sets for isoWATFLOOD model evaluation.



- ii. For each of the three  $\delta^{18}\text{O}_{\text{ppt}}$  inputs (KPN43, REMOiso and static), streamflow and  $\delta^{18}\text{O}_{\text{SF}}$  are simulated from 1996 to 1999 for all 30,000 parameter sets (as defined in (i)).
- iii. Modelled streamflow and  $\delta^{18}\text{O}_{\text{SF}}$  are assessed statistically over the period of study (1997–1999, excluding the 1996 spin-up year), and regionally across the Jean Marie and Blackstone sub-basins. Simulations are classified as behavioural or non-behavioural based on the following set of efficiency criteria thresholds for streamflow and  $\delta^{18}\text{O}_{\text{SF}}$ :
- 5
- a. Streamflow:
- $\text{NSE} \geq 0.5$ ;
- $|\%Dv| \leq 20 \%$ , and;
- $|\log(\%Dv)| \leq 20 \%$ .
- 10
- b.  $\delta^{18}\text{O}_{\text{SF}}$ :
- $\text{RMSE} \leq 2.5 \%$ , and;
- $\text{KGE} \geq 0.3$ .
- iv. All reported results, such as simulation means, percentiles and parameter values, are derived from behavioural simulations corresponding to each  $\delta^{18}\text{O}_{\text{ppt}}$  input. The presented uncertainty bounds are the 5<sup>th</sup> and 95<sup>th</sup> percentiles of simulated streamflow,  $\delta^{18}\text{O}_{\text{SF}}$ , and snowpack  $\delta^{18}\text{O}$ , drawn from the behavioural simulations at each time step for each  $\delta^{18}\text{O}_{\text{ppt}}$  input.
- 15

The behavioural thresholds used in this study are subjectively defined, however, are arrived at through a review of methods employed in similar studies (e.g., Moriasi et al., 2007; Birkel et al., 2010a; 2010b; 2011; Smith et al., 2016), measurement error, and an iterative process exploring the sensitivity between the set thresholds and resulting behavioural simulations for each input type. Based on this analysis, the Nash-Sutcliffe efficiency (NSE; Nash and Sutcliffe, 1970), volumetric error criteria (%Dv), root mean square error (RMSE), and the Kling-Gupta efficiency criterion (KGE; Gupta et al., 2009; Kling et al., 2012) are selected. A multi-criteria model evaluation approach is used as each criterion places emphasis on different statistical properties of a simulation. For example, NSE has a documented bias towards peak flow, and conversely, log (%Dv) is more appropriate evaluation measure for periods of low flow. The NSE, %Dv, and log(%Dv) efficiency are not considered suitable metrics for  $\delta^{18}\text{O}_{\text{SF}}$  assessment due to the temporal discontinuity of the isotope observations, therefore RMSE and KGE are used as isotopic simulation statistics. It should also be noted that  $\delta^{18}\text{O}_{\text{SF}}$  observations are not equally distributed through time, whereby the highest concentration of observations occurs during snowmelt in the month of May (~25 %), and the fewest observations during the six month ice-on period from November to April (~23 %). The sporadic distribution of observations may result in the calibrations more highly weighted to certain periods of the year and the dominate processes occurring at that time; therefore having the potential to impact model parameterization.

20

25

30





## 4 Results and discussion

Results of the three calibrations indicate that  $\delta^{18}\text{O}_{\text{ppt}}$  input influences the number simulations that meet the behavioural criteria thresholds. The KPN43 input results in an increased number of behavioural simulations ( $n = 321$ ) in comparison to the REMOiso ( $n = 268$ ) or static ( $n = 216$ ) input types (Table 7). This suggests that the choice of  $\delta^{18}\text{O}_{\text{ppt}}$  input may potentially impact internal apportionment of water (i.e., the modelled proportion of water entering, stored, and released from the ground surface, upper and lower zones) through model parameterization. A summary of behavioural parameter set characteristics is provided in Table 8. Land cover parameters are reported as weighted averages for the modelled region.

Among input types there are potentially significant differences in several parameters, which will be explored further throughout the remainder of the manuscript. In almost all instances, the range of the parameters was not significantly constrained from the allowable parameter range. Due to the wide range of behavioural parameter values (Table 8), we are confident that the approach used is sufficient to characterize parameter uncertainty. However, and not unexpectedly, this finding also indicates that 30,000 model evaluations are not sufficient to quantify parameter identifiability for isoWATFLOOD.

### 4.1 Precipitation oxygen-18 input

Time series of  $\delta^{18}\text{O}_{\text{ppt}}$  inputs for Jean Marie and Blackstone are displayed on panel (a) of Fig. 3 and Fig. 4, respectively. On average, KPN43 input is the most enriched (-20.48 ‰), followed by REMOiso (-21.78 ‰), with static being the most depleted (-22.82 ‰). KPN43 and static inputs have similar variation about their mean values, with CV's equal to 0.19 and 0.20, respectively. Conversely, REMOiso has a higher CV (0.25) and much larger range, which is, in part, due to the finer daily time step of this input. Spatial variability between Jean Marie and Blackstone sub-basins is zero for the static input; however some variation among sub-basins is seen for KPN43 and REMOiso. Interestingly, both the KPN43 and REMOiso inputs show, on average, more depleted  $\delta^{18}\text{O}_{\text{ppt}}$  values within Blackstone (-20.79 ‰ and -22.01 ‰, respectively) in comparison to Jean Marie (-20.17 ‰ and -21.54 ‰, respectively), in addition to increased variability. This is likely caused by the higher elevations present in the headwaters of the Blackstone relative to the Jean Marie (a maximum difference of ~215 m).

Although there are only 19 rainfall  $\delta^{18}\text{O}$  observations collected over the study period for Jean Marie, and eight within Blackstone (hollow black diamonds on Fig. 3 and Fig. 4), these limited data provide some information regarding accuracy of the estimated  $\delta^{18}\text{O}_{\text{ppt}}$  inputs. By visual inspection, each of the three inputs produces reasonable estimates of  $\delta^{18}\text{O}_{\text{ppt}}$ . This is expected for the static input as the seasonal compositions are derived from observations; however, this comparison provides some level of validation for KPN43 and REMOiso. REMOiso is the only input that can somewhat replicate the event-scale variability in  $\delta^{18}\text{O}_{\text{ppt}}$  due to its daily time step. Both KPN43 and static inputs appear to generally capture the average



magnitude of summer rainfall events and overall seasonal variability; however there are insufficient observations to statistically support this statement within the Fort Simpson basin.

#### 4.2 Modelling streamflow

5 All calibrations adequately capture variations in total streamflow in both sub-basins, as emphasised by the regional calibration statistics (Table 7). On average, behavioural streamflow simulations have a NSE of 0.68, and %Dv of 13.8 %. Mean daily streamflow and uncertainty bounds for the KPN43, REMOiso and static model calibrations are displayed on panel (b) of Fig. 3 and Fig. 4. Differences in hydrograph characteristics between Jean Marie and Blackstone are due to variations in basin physiography, storage mechanisms and land cover composition, specifically large differences in average basin slope and wetland dynamics (St Amour et al., 2005).

10

Within the Jean Marie sub-basin, both the timing and volume of peak flows derived from snow melt and early summer rains are well captured in 1998, however, volume is under predicted in 1997 and 1999 for the average streamflow simulation. The model also has difficulty capturing the volume of the recession limb, which may be attributed to the parameterization of baseflow and fen response in this sub-basin. In the Blackstone, the recession limb of the hydrograph and low flow volume are well modelled, however peak flows (with the exception of the 1997 snow melt) are under estimated. Within both sub-basins, flows from 1997 fall rain events are well captured, except for an October 1998 rainfall event that generated a limited streamflow response in both sub-basins. This may point to inadequate precipitation forcing due to the climate station proximity and high spatial variability of rainfall, or possibly inadequate soil moisture parameterization.

15

An interesting finding is the similarity of mean streamflow simulation despite contrasting  $\delta^{18}\text{O}_{\text{ppt}}$  inputs. Kendall's tau coefficient ( $\tau$ ) is used to determine the level of correlation between streamflow simulations generated by the three  $\delta^{18}\text{O}_{\text{ppt}}$  input methods (and associated model parameterizations). In Jean Marie,  $\tau$  ranges between 0.92 (REMOiso versus static) to 0.97 (KPN43 versus static). In Blackstone  $\tau$  is more tightly constrained, ranging from 0.96 (REMOiso versus static) to 0.98 (KPN43 versus static). All  $\tau$  values are statistically significant. It should be noted that small deviations between mean streamflow simulations occur during spring melt, where REMOiso driven streamflow consistently shows higher peaks than KPN43 and static driven simulations. However, these differences in mean streamflow fall within overlapping uncertainty bounds and therefore are not deemed significant outside of parameter uncertainty.

25

#### 4.3 Modelling $\delta^{18}\text{O}$

Mean daily  $\delta^{18}\text{O}_{\text{SF}}$  simulations and uncertainty bounds for KPN43, REMOiso and static model calibrations are displayed on panel (c) of Fig. 3 and Fig. 4. Each calibration produces mean simulations that capture many of the trends (but not particularly the magnitudes) present in the observed  $\delta^{18}\text{O}_{\text{SF}}$  record. Observed  $\delta^{18}\text{O}_{\text{SF}}$  show depletion in streamflow  $\delta^{18}\text{O}$  due to large influxes of snowmelt during the spring freshets, with  $\delta^{18}\text{O}_{\text{SF}}$  gradually enriching over the summer months due to the

30



influence of evaporative enrichment of surface and soil waters, occasionally punctuated by rainfall events that may enrich or deplete  $\delta^{18}\text{O}_{\text{SF}}$ . During late fall and throughout the winter,  $\delta^{18}\text{O}_{\text{SF}}$  tends toward a more depleted, stable groundwater composition (St Amour et al., 2005).

5 Though each of the calibrations demonstrate many of the same trends as the observed  $\delta^{18}\text{O}_{\text{SF}}$  record, there are notable differences. As simulated  $\delta^{18}\text{O}_{\text{SF}}$  uncertainly envelopes associated with each input type are, at times, non-overlapping, differences in  $\delta^{18}\text{O}_{\text{SF}}$  simulations can be attributed to  $\delta^{18}\text{O}_{\text{ppt}}$  input and therefore are not just an artefact of parameter uncertainty. The dissimilarities between  $\delta^{18}\text{O}_{\text{SF}}$  simulations are also reflected in the RMSE statistic (Table 7). The RMSE is larger for static-derived  $\delta^{18}\text{O}_{\text{SF}}$  simulations due to increased emphasis on periods with observed data (i.e., spring freshet),  
 10 where larger offsets between simulated and observed  $\delta^{18}\text{O}_{\text{SF}}$  exist. The KGE statistic does not mirror RMSE, as it shows only minor differences between  $\delta^{18}\text{O}_{\text{SF}}$  simulations. The nature of the KGE statistic is to put less emphasis on error offsets derived from peak flows (i.e., spring freshet) by providing a more balanced approach where error is summed first and squared at the end, preserving the sign of the error and enabling a trade-off of error throughout the simulation (Gupta et al., 2009). Therefore, this statistic better reflects the fit of the overall simulation throughout the study period, however, further research  
 15 is required to better understand the impacts of sporadic sampling resolution (of  $\delta^{18}\text{O}_{\text{SF}}$  observations) on efficiency criteria, and consequently the objective functions(s).

Differences in  $\delta^{18}\text{O}_{\text{SF}}$  simulations within each sub-basin are due to a combination of: (1) the markedly different  $\delta^{18}\text{O}_{\text{ppt}}$  compositions between inputs entering the system during large precipitation events, and (2) the way in which new water  
 20 flushes through the system via the various hydrological compartments. For this study area, large precipitation events can be further separated into: (1) major rainfall events occurring in post-freshet (summer and fall) months, and (2) the accumulation of winter snowfall and corresponding spring freshet

Post-freshet  $\delta^{18}\text{O}_{\text{SF}}$  simulations are impacted by rainfall amount and composition, as well as the offset between simulated  
 25  $\delta^{18}\text{O}_{\text{SF}}$  and  $\delta^{18}\text{O}_{\text{ppt}}$  input at the time of rainfall. As rainfall amount and/or the offset increases, the resulting impact on simulated  $\delta^{18}\text{O}_{\text{SF}}$  also increases. This highlights the importance of capturing the variability in the  $\delta^{18}\text{O}_{\text{ppt}}$  input, particularly for large and isotopically distinct (from streamflow) rainfall events. The threshold defining a large rainfall event will vary depending on factors such as basin physiography, land cover, storage capacity and antecedent conditions. St Amour et al. (2005) estimate this threshold to be  $\geq 40$  mm within the Fort Simpson region. An example of a large, yet isotopically distinct,  
 30 rainfall event is June 11–12, 1998 when approximately 70 mm fell over two days with an observed  $\delta^{18}\text{O}_{\text{ppt}}$  composition of -22.7 ‰. Both REMOiso and static inputs reasonably captured the event (-20.9 ‰ and -20.1 ‰, respectively), however, the KPN43 input predicted a composition of -17.6 ‰. This single event resulted in a significant offset between KPN43  $\delta^{18}\text{O}_{\text{SF}}$  compared with REMOiso and static  $\delta^{18}\text{O}_{\text{SF}}$  which was maintained throughout the remainder of 1998, until the 1999 freshet.



Throughout much of Canada and other regions experiencing a high-latitude climate, a substantial portion of annual streamflow (and typically the peak flow) is generated during the spring freshet when the accumulation of solid precipitation from the winter season melts over the period of a few weeks. Therefore, it is important to understand how differences in  $\delta^{18}\text{O}_{\text{ppt}}$  input impact snowpack and snowmelt isotopic compositions in isoWATFLOOD. Figure 5 shows the evolution of precipitation-weighted snowpack oxygen-18 ( $\delta^{18}\text{O}_{\text{SNW}}$ ) throughout each winter season of the study period.

Comparison of like-forcing pairs between Jean Marie and Blackstone reveals subtle spatial differences in simulated  $\delta^{18}\text{O}_{\text{SNW}}$ . There are, however, significant differences between KPN43, REMOiso and static snowpack compositions within each sub-basin. Interestingly, REMOiso and KPN43 simulations of snowpack show similar end of winter precipitation-weighted  $\delta^{18}\text{O}_{\text{SNW}}$ , differing by less than 0.5 ‰ in 1997–1998 and 1998–1999. REMOiso and KPN43 inputs also consistently generate significantly more enriched snowpack compositions throughout the study period in comparison to static  $\delta^{18}\text{O}_{\text{SNW}}$  input. On average, KPN43 is 3.3 ‰ more enriched, and REMOiso is 3.1 ‰ more enriched than end of season static  $\delta^{18}\text{O}_{\text{SNW}}$ . These differences may also stem from insufficiencies in modelled snowpack due to fractionation during sublimation, melting and refreezing of the snowpack that is unaccounted for in the current isoWATFLOOD model. The static input may inadvertently account for some of these processes, as the specified compositions are from snow pack sampling conducted towards the end of winter (in late March). Research into defining snow fall, pack and melt offsets (from field studies), and refining isoWATFLOOD's cryospheric dynamics and processes is currently ongoing.

These significant differences in simulated snowpack composition are one of the primary causes for the offsets between KPN43, REMOiso and static  $\delta^{18}\text{O}_{\text{SF}}$  simulations (Fig. 3 and Fig. 4). Once a  $\delta^{18}\text{O}_{\text{SF}}$  simulation has been offset, it is not possible to 'reset' the isotopic composition in late fall when streamflow decreases to near-zero since there is still mass remaining in the system. This can result in compounding isotopic error over a continuous simulation period, thus highlighting the sensitivity of this tracer as a calibration tool. This compounding error is also observed for rainfall events, but generally to a lesser extent due to the relatively smaller durations and magnitudes (volume contributions) of most rainfall events in high latitude regions.

Provided that both  $\delta^{18}\text{O}_{\text{SF}}$  and  $\delta^{18}\text{O}_{\text{SNW}}$  are significantly different among  $\delta^{18}\text{O}_{\text{ppt}}$  inputs, internal water apportionment (determined by model parameterization) may also be influenced by  $\delta^{18}\text{O}_{\text{ppt}}$  input type. Therefore, hydrograph component contributions are further explored to determine the effect that the differences in  $\delta^{18}\text{O}_{\text{ppt}}$  input has on these contributions.

#### 4.4 Hydrograph component analysis and parameter distributions

Percent of volume contributing to total streamflow from surface runoff, interflow and baseflow storage for each season (DJF: December-January-February; MAM: March-April-May; JJA: June-July-August; and, SON: September-October-November) and each of the three  $\delta^{18}\text{O}_{\text{ppt}}$  inputs are shown on Fig. 6. Jean Marie and Blackstone sub-basins generally display similar



trends in internal water apportionment throughout the year, indicating generally similar model parameterizations. Some seasonal differences are visible, which can be linked to variations in basin physiography, land cover, and storage characteristics reflected by differences in the baseflow (Izf and pwr) and wetland parameters (kcond and theta) between basins (Table 8). Overall, the freshet and post-freshet percent volume contributions to total streamflow in this study are in  
5 general agreement with those reported in previous studies. For example, St Amour et al. (2005) also found groundwater contributions to be significant throughout the year, and estimated post-freshet contributions to total streamflow at 71 % ( $\pm 9$  %) and 64% ( $\pm 10$  %) within Jean Marie and Blackstone, respectively. Snowmelt contributions were estimated to be 21 % ( $\pm 2$  %) and 40 % ( $\pm 4$  %) of total streamflow volume for Jean Marie and Blackstone. Additionally, Jasechko et al. (2016) estimate that annually 80 – 90 % of the Mackenzie River streamflow is “old” water (i.e., water that has not entered the  
10 stream within the last  $2.3 \pm 0.8$  months). Their findings also suggest that the annual percentage of old streamflow can be higher in mountainous watersheds with steeper slopes, such as the FSB, than lower-gradient watersheds.

Comparison of seasonal volume contributions derived from each  $\delta^{18}\text{O}_{\text{ppt}}$  input reveal that during spring (MAM), REMOiso-driven simulations show significantly more surface flow contribution to total streamflow, with the mean volume lying above  
15 the 95<sup>th</sup> percentile for both the KPN43 and static input simulations (Fig. 6). On average, REMOiso simulations contribute almost twice as much surface runoff to total streamflow as KPN43 and static simulations during MAM (39 % versus 25 % and 22 %, respectively, for Jean Marie; and similar, yet slightly larger, percent contributions for Blackstone).

Based on the averaged seasonal analysis, no other significant differences in component contributions outside of parameter  
20 uncertainty can be attributed to  $\delta^{18}\text{O}_{\text{ppt}}$  input. It is important to note, however, that each  $\delta^{18}\text{O}_{\text{ppt}}$  input results in differing amounts of parameter uncertainty, both seasonally and overall, as represented by differing widths of uncertainty bounds (cross symbols) on Fig. 6. The variation in uncertainty bounds between  $\delta^{18}\text{O}_{\text{ppt}}$  inputs is also visible on Fig. 3 through Fig. 5. The REMOiso input yields the largest amount of uncertainty in total streamflow, also reflected in the relatively larger amounts of uncertainty in surface water and baseflow component contributions (Fig. 6). Conversely, KPN43 and static  
25 inputs generate similar or slightly larger uncertainty in interflow (soil water) contributions relative to REMOiso and lower uncertainty surrounding surface and baseflow contributions, and overall total streamflow. These differences in uncertainty are attributed to the number, and characteristics of the behavioural parameters retained for each  $\delta^{18}\text{O}_{\text{ppt}}$  input, which originate due to distinctions in magnitude and variability (both spatial and temporal) among  $\delta^{18}\text{O}_{\text{ppt}}$  inputs.

30 Further demonstrated by parameter probability distributions (Fig. 7), the three calibrations resulted in noteworthy differences in behavioural parameters. We do not display these distributions to comment definitively on parameter identifiability because, as previously noted, the number of evaluations was insufficient for that purpose. But rather, we introduce this analysis for select parameters to reinforce and explain the findings from Fig. 3 through Fig. 5, and to highlight that within this study, model parameterization is impacted by  $\delta^{18}\text{O}_{\text{ppt}}$  input. The selected parameters influence evaporation (f-ratio),



surface runoff during snowmelt (akfs, base), upper and lower zone storage (retn), interflow (retn), and baseflow (lzf, pwr). Results show that more often than not, REMOiso parameter distributions are different than KPN43 or static parameter distributions. Although there are dissimilarities between KPN43 and static parameter distributions, however these are typically not as prevalent. This echoes the findings from Fig. 7 that KPN and static derived contributions to total streamflow are more similar than contributions arising from REMOiso; which may very well be an artefact of the increased temporal and spatial variability in the REMOiso  $\delta^{18}\text{O}_{\text{ppt}}$  input relative to that of the KPN43 and static inputs.

Differences in surface water contribution during snowmelt between REMOiso, KPN43 and static inputs are likely explained by differences in the akfs and base parameters. Parameter distributions derived from REMOiso are significantly different (as verified through Kolmogorov–Smirnov testing of distributions) than those from the KPN43 and static inputs for these parameters (Fig. 7, panels (b) and (f)). Lower akfs values represent decreased infiltration and increased surface runoff during snowmelt, which corresponds to REMOiso's increased surface water contributions to total streamflow during spring (MAM). Differences in baseflow contribution and uncertainty between  $\delta^{18}\text{O}_{\text{ppt}}$  inputs are attributed, in part, to differences in the lzf and pwr parameters (Fig. 7, panels (c-d) and (g-h)), which have a large impact on the quantity of baseflow and the slope of the recession limb of the hydrograph. Wider uncertainty bounds for REMOiso relative to KPN43 and static within Blackstone (Fig. 6, panel (f)), and for all  $\delta^{18}\text{O}_{\text{ppt}}$  inputs during fall and winter within Jean Marie (Fig. 6, panel (c)), are likely due to the wider range of behavioural values for the pwr parameter, specifically the inclusion of lower values which results in longer, more drawn out recession limbs.

Although more work is required towards assessing and understanding parameter identifiability for WATFLOOD, the above analysis shows that selection of  $\delta^{18}\text{O}_{\text{ppt}}$  input has direct implications on model parameterization, and this source of uncertainty should be considered in future studies.

## 5 Conclusions and recommendations

This study uses three types of estimated  $\delta^{18}\text{O}_{\text{ppt}}$  input to force a tracer-aided hydrological model, isoWATFLOOD, and investigates the impact that  $\delta^{18}\text{O}_{\text{ppt}}$  inputs of differing spatial and temporal variability have on total streamflow, isotopic composition of streamflow, and the seasonality of individual hydrograph components. This work informs the over-arching goal of quantifying and reducing uncertainty (and equifinality) in isoWATFLOOD streamflow simulations.

This study demonstrated that although total simulated streamflow is not significantly affected by choice of  $\delta^{18}\text{O}_{\text{ppt}}$  input,  $\delta^{18}\text{O}_{\text{SF}}$  simulations and the internal apportionment of water (surface flow, interflow, and baseflow contributions) in WATFLOOD can be significantly impacted, especially during large precipitation and snowmelt events. The ability of estimated  $\delta^{18}\text{O}_{\text{ppt}}$  to capture both the variability and seasonality in precipitation isotopes, especially when precipitation events



display distinctly different isotopic compositions than that of streamflow, is critical for tracer-aided model forcing. Differences in  $\delta^{18}\text{O}_{\text{SF}}$  and water partitioning between compartments are driven by differences in model parameterization, as witnessed by variations in the amount of uncertainty and parameter distributions between  $\delta^{18}\text{O}_{\text{ppt}}$  input types. This suggests that choice of  $\delta^{18}\text{O}_{\text{ppt}}$  input impacts parameterization of WATFLOOD, and for this reason, if estimates of  $\delta^{18}\text{O}_{\text{ppt}}$  are used in  
5 modelling, modellers should account for this input uncertainty in overall uncertainty assessments. Findings also show that simulations of total streamflow did not show significant differences between the three  $\delta^{18}\text{O}_{\text{ppt}}$  inputs and corresponding parameterizations, despite  $\delta^{18}\text{O}_{\text{SF}}$  simulations displaying significant differences. This reinforces to the utility of tracer-aided models to diagnose issues with equifinality in model simulations.

10 As WATFLOOD is a complex model with a large amount of parameters, it is important to work towards conducting a comprehensive study focusing on  $\delta^{18}\text{O}_{\text{ppt}}$  input uncertainty and parameter identifiability. Ideally, further studies should be conducted in a watershed that is adequately instrumented to characterize observed  $\delta^{18}\text{O}_{\text{ppt}}$  input both spatially, but more importantly, temporally (i.e., daily or weekly sampling resolution). This will facilitate a ‘baseline’ model calibration (using observed  $\delta^{18}\text{O}_{\text{ppt}}$ ) from which deviations in parameter distributions due to estimated  $\delta^{18}\text{O}_{\text{ppt}}$  input can be more intensely  
15 explored. This type of study would allow further investigation of several key questions: first, if these pseudo-forcings are adequate alternatives in place of  $\delta^{18}\text{O}_{\text{ppt}}$  observations; second, if there is a specific subset of model parameters that are more sensitive to estimated  $\delta^{18}\text{O}_{\text{ppt}}$  input, and how (if at all) these parameters compensate for compounding error stemming from estimated  $\delta^{18}\text{O}_{\text{ppt}}$  input. Unfortunately, at least within Canada, a well instrumented watershed at the regional scale does not yet exist. This again points to the importance of implementing additional (or enhancing current) hydro-meteorological  
20 monitoring networks. If observation networks allow, this type of study should also be conducted in watersheds of differing dominant hydrological processes (e.g., rainfall-dominated versus snowmelt-dominated) to better understand  $\delta^{18}\text{O}_{\text{ppt}}$  input uncertainty on parameterization across the range of modelled hydrologic processes.

As expected, REMOiso exhibits some bias and will continue to need correction before application within Canadian  
25 watersheds. More studies are needed to examine the differences between  $\delta^{18}\text{O}_{\text{ppt}}$  observations and REMOiso simulations throughout Canada to better understand the nature of this bias, and the most appropriate bias correction methods; which can be done using observations from the CNIP database. This feedback regarding REMOiso performance across Canada is also beneficial to model developers. Additionally, the suitability and performance of other isotope-enabled RCM’s for use in Canada should be explored. Regarding the usefulness of the inputs within regions of limited  $\delta^{18}\text{O}_{\text{ppt}}$  observations, both the  
30 static and REMOiso inputs require existing  $\delta^{18}\text{O}_{\text{ppt}}$  observations (i.e., from CNIP) to either define or bias correct the input, which may limit their use for certain applications. If these data are not available, the KPN43 input provides reasonable results without the need for any additional observations. For all inputs, the existence of CNIP (and other isotopes in precipitation networks) observations is crucial to the development, validation, and bias correction of estimated  $\delta^{18}\text{O}_{\text{ppt}}$  inputs.



Understanding how estimated  $\delta^{18}\text{O}_{\text{ppt}}$  input affects model parameterization and subsequently internal distribution of water is important to understanding overall uncertainty of the model and the model's abilities and limitations without access to observed  $\delta^{18}\text{O}_{\text{ppt}}$  times series data. Canada, alongside many other countries, continues to have significant regions with sparse hydro-meteorological observations (Coulibaly et al., 2013). This study achieves an understanding of how isoWATFLOOD can be used in regions with a limited number of  $\delta^{18}\text{O}_{\text{ppt}}$  observations, and that the model can be of value in such regions. This study reinforces that a tracer-aided modelling approach assists with resolving hydrograph component contributions, and works towards diagnosing the issue of model equifinality and knowledge that modellers are achieving the 'right answers for the right reasons' (Kirchner, 2006). Attaining this understanding of  $\delta^{18}\text{O}_{\text{ppt}}$  input uncertainty on simulated model output is especially important when calibrated models are used as tools to assess how changes in climate or land-use effect future predictions of streamflow.

#### Author contribution

C. Delavau developed model code to generate Kpn  $\delta^{18}\text{O}_{\text{ppt}}$  input, perform Monte Carlo simulations, and process the corresponding output. T. Stadnyk and T. Holmes developed and enhanced isoWATFLOOD code for the version of isoWATFLOOD used in this study. C. Delavau performed the analysis presented in this manuscript, with assistance from T. Stadnyk. C. Delavau prepared the manuscript with contributions from all co-authors.

#### Competing interests

The authors declare that they have no conflict of interest.

#### Acknowledgements

The authors would like to acknowledge Dr. K. Sturm for provision of the REMOiso data utilized in this study. Additional thanks go to Dr. N. Kouwen for direction and input on WATFLOOD modelling. CNIP is made possible through the help of the Canadian Air and Precipitation Monitoring Network (CAPMoN) for sample collection - Kaz Higuchi and Dave MacTavish in particular, the Environmental Isotope Laboratory at the University of Waterloo for sample analysis, and Tom Edwards for initiating and maintaining the network. We would also like to acknowledge Dan McKinney for provision of the ANUSPLIN data used in this study. Finally, we would like to acknowledge the contributions of our reviewers whose valuable input has improved this manuscript significantly. This research was partially funded by Natural Sciences and Engineering Research Council (NSERC) Alexander Graham Bell Canada Graduate Scholarship (CGS-D).





## References

- Aylsworth, J. M. and Kettles, I. M.: Distribution of peatlands, in: *The Physical Environment of the Mackenzie Valley, Northwest Territories: A Base Line for the Assessment of Environmental Change*, edited by: Dyke, L. D., and G. R. Brooks, Geological Survey of Canada Bulletin 547, Geological Survey of Canada, Ottawa, 49–55, 2000.
- 5 Beven, K. and Binley, A.: The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Process.*, 6, 279–298, doi: 10.1002/hyp.3360060305, 1992.
- Birkel, C., and Soulsby, C.: Advancing tracer-aided rainfall–runoff modelling: a review of progress, problems and unrealised potential, *Hydrol. Process.*, 29, 5227–5240, doi: 10.1002/hyp.10594, 2015.
- Birkel, C., Soulsby, C., and Tetzlaff, D.: Developing a consistent process-based conceptualization of catchment functioning using measurements of internal state variables, *Water Resour. Res.*, 50, 3481–3501, doi:10.1002/2013WR014925, 2014.
- 10 Birkel, C., Tetzlaff, D., Dunn, S. M., and Soulsby, C.: Using time domain and geographic source tracers to conceptualize streamflow generation processes in lumped rainfall-runoff models, *Water Resour. Res.*, 47, W02515, doi:10.1029/2010WR009547, 2011.
- Birkel, C., Dunn, S. M., Tetzlaff, D., and Soulsby, C.: Assessing the value of high-resolution isotope tracer data in the stepwise development of a lumped conceptual rainfall–runoff model, *Hydrol. Process.*, 24, 2335–2348, doi: 10.1002/hyp.7763, 2010a.
- Birkel, C., Tetzlaff, D., Dunn, S. M., and Soulsby, C.: Towards a simple dynamic process conceptualization in rainfall–runoff models using multi-criteria calibration and tracers in temperate, upland catchments, *Hydrol. Process.*, 24, 260–275, doi: 10.1002/hyp.7478, 2010b.
- 20 Birks, S. J., and Gibson, J. J.: Isotope hydrology research in Canada, 2003–2007, *Can. Water Resour. J.*, 34(2), 163–176, doi:10.4296/cwrj3402163, 2009.
- Birks, S. J., and Edwards, T. W. D.: Atmospheric circulation controls on precipitation isotope-climate relations in western Canada, *Tellus Ser. B*, 61B(3), 566–576, doi: 10.1111/j.1600-0889.2009.00423.x, 2009.
- Carpenter, T. M., and Georgakakos, K. P.: Intercomparison of lumped versus distributed hydrologic model ensemble simulations on operational forecast scales, *J. Hydrol.*, 329(1–2), 174–185, doi: 10.1016/j.jhydrol.2006.02.013, 2006.
- 25 Coulibaly, P., Samuel, J., Pietroniro, A., and Harvey, D.: Evaluation of Canadian National Hydrometric Network density based on WMO 2008 standards, *Can. Water Resour. J.*, 38(2), 159–167, doi: 10.1080/07011784.2013.787181, 2013.
- Delavau, C., Chun, K. P., Stadnyk, T., Birks, S. J., and Welker, J. M.: North American precipitation isotope (d18O) zones revealed in time series modelling across Canada and northern United States, *Water Resour. Res.*, 51, doi:10.1002/2014WR015687, 2105.
- 30 Dietermann, N. and Weiler, M.: Spatial distribution of stable water isotopes in alpine snow cover, *Hydrol. Earth Syst. Sci.*, 17, 2657–2668, doi:10.5194/hess-17-2657-2013, 2013.
- Dunn, S. M., Freer, J., Weiler, M., Kirkby, M. J., Seibert, J., Quinn, P. F., Lischeid, G., Tetzlaff, D., and Soulsby, C.: Conceptualization in catchment modelling: simply learning?, *Hydrol. Process.*, 22: 2389–2393, doi: 10.1002/hyp.7070, 2008.
- 35 Ecoregions Working Group: Ecoclimatic regions of Canada, first approximation. In *Ecoregions Working Group of the Canada Committee on ecological land classification, Ecological Land Classification Series 23. Sustainable Development Branch, Canadian Wildlife Service, Environment Canada: Ottawa, Ontario, 1989.*
- Fenicia, F., McDonnell, J. J., and Savenije, H. H. G.: Learning from model improvement: On the contribution of complementary data to process understanding, *Water Resour. Res.*, 44, W06419, doi: 10.1029/2007WR006386, 2008.
- 40 Gat, J. R., Bowser, C. J., and Kendall, C.: The contribution of evaporation from the Great Lakes to the continental atmosphere: estimate based on stable isotope data, *Geophys. Res. Lett.*, 21, 557–560, doi: 10.1029/94GL00069, 1994.
- Gibson J. J., and Prowse, T. D.: Isotopic characteristics of ice cover in a large northern river basin, *Hydrol. Process.*, 13, 2537–2548, doi: 10.1002/(SICI)1099-1085(199911)13:16<2537::AID-HYP940>3.0.CO;2-A, 1999.
- 45 Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, *J. Hydrol.*, 377(1–2), doi: 10.1016/j.jhydrol.2009.08.003, 2009.
- Hargraeves, G. H., and Samani, Z. A.: Estimating potential evapotranspiration, *ASCE, J. Irrig. and Drain*, 108(3), 225–230, 1982.



- Her, Y., and Chaubey, I.: Impact of the numbers of observations and calibration parameters on equifinality, model performance, and output and parameter uncertainty, *Hydrol. Process.*, 29, 4220–4237, doi: 10.1002/hyp.10487, 2105.
- IAEA/WMO: Global Network of Isotopes in Precipitation. The GNIP Database. Accessible at: <http://www.iaea.org/water>, 2014.
- 5 Jasechko, S., Kirchner, J. W., Welker, J. M., and McDonnell, J. J.: Substantial proportion of global streamflow less than three months old, *Nat. Geosci.*, 9, 126–129, doi: 10.1038/ngeo2636, 2106.
- Kirchner, J.: Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, *Water Resour. Res.*, 42(3), W03S04, doi:10.1029/2005WR004362, 2006.
- Kling, H., Fuchs, M., and Paulin, M.: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios, *J. Hydrol.*, 424–425, 264–277, doi: 10.1016/j.jhydrol.2012.01.011, 2012.
- 10 Kouwen, N.: WATFLOOD/WATROUTE Hydrological Model Routing and Flood Forecasting System. User's Manual, University of Waterloo, Waterloo, ON. [www.watflood.ca](http://www.watflood.ca), 2014.
- Kouwen, N., Soulis, E. D., Pietroniro, A., Donald, J., and Harrington, R. A.: Grouped response units for distributed hydrological modelling, *J. Water Resour. Plann. Manage.*, 119, 289–305, doi: 10.1061/(ASCE)0733-9496(1993)119:3(289), 1993.
- 15 Kuczera, G.: Improved parameter inference in catchment models: 2. Combining different kinds of hydrologic data and testing their compatibility, *Water Resour. Res.*, 19(5), 1163–1172, doi:10.1029/WR019i005p01163, 1983.
- Kuczera, G., and Mroczkowski, M.: Assessment of hydrologic parameter uncertainty and the worth of multiresponse data, *Water Resour. Res.*, 34(6), 1481–1489, doi:10.1029/98WR00496, 1998.
- 20 Lykoudis, S. P., Argiriou, A. A., and Dotsika, E.: Spatially interpolated time series of d18O in Eastern Mediterranean precipitation, *Global Planet Change*, 71, 150–159, doi: 10.1016/j.gloplacha.2009.09.004, 2010.
- Noone, D., and Sturm, C.: Comprehensive dynamical models of global and regional water isotope distributions, in: *Isoscapes: Understanding movement, pattern, and process on Earth through isotope mapping*, edited by: West, J. B., G. J. Bowen, T. E. Dawson, and K. P. Tu, Springer, London, 195–219, 2010.
- 25 McKenney, D. W., Hutchinson, M. F., Papadopol, P., Lawrence, K., Pedlar, J., Campbell, K., and Owen, T.: Customized Spatial Climate Models for North America, *Bull. Am. Meteorol. Soc.*, 92(12), 1611–1622, doi:10.1175/2011BAMS3132.1, 2011.
- McMillan, H., Tetzlaff, D., Clark, M., and Soulsby, C.: Do time-variable tracers aid the evaluation of hydrological model structure? A multimodel approach, *Water Resour. Res.*, 48, W05501, doi:10.1029/2011WR011688, 2012.
- 30 Mesinger, F., et al.: North American regional reanalysis, *Bull. Am. Meteorol. Soc.*, 87, 343–360, doi: 10.1175/BAMS-87-3-343, 2006.
- Michaud, J., and Sorooshian, S.: Comparison of simple versus complex distributed runoff models on a mid-sized semiarid watershed, *Water Resour. Res.*, 30(3), 593–605, doi:10.1029/93WR03218, 1994.
- Moran, T. A., Marshall, S. J., Evans, E. C., and Sinclair, K. E.: Altitudinal gradients of stable isotopes in lee-slope precipitation in the Canadian Rocky Mountains, *Arctic Antarctic Alpine Research*, 39(3), 455–467, <http://www.jstor.org/stable/20181717>, 2007.
- 35 Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., and Veith, T. L.: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, *Trans. ASABE*, 50(3), 885–900, 2007.
- Nash, J. E., and Sutcliffe, J. V.: River flow forecasting through conceptual models part I—A discussion of principles, *J. Hydrol.*, 10(3), 282–290, doi:10.1016/0022-1694(70)90255-6, 1970.
- 40 Penna, D., Ahmad, M., Birks, S. J., Bouchaou, L., Brenčić, M., Butt, S., Holko, L., Jeelani, G., Martínez, D. E., Melikadze, G., Shanley, J. B., Sokratov, S. A., Stadnyk, T., Sugimoto, A., and Vreča, P.: A new method of snowmelt sampling for water stable isotopes, *Hydrol. Process.*, 28, pages 5367–5644, doi: 10.1002/hyp.10273, 2014.
- Petrone, R. M., Griffis, T. J., Rouse, W. R.: Synoptic and surface climatology interactions in the central Canadian subarctic: normal and El Niño seasons, *Phys. Geog.*, 21, 368–383, doi: 10.1080/02723646.2000.10642715, 2000.
- 45 Pietroniro, A., Prowse, T. D., Hamlin, L., Kouwen, N., and Soulis, R.: Application of a grouped response unit hydrological model to a northern wetland region, *Hydrol. Process.*, 10(10), 1245–1261, doi: 10.1002/(SICI)1099-1085(199610)10:10<1245::AID-HYP457>3.0.CO;2-0, 1996.
- Philip, J. R.: An infiltration equation with physical significance, *J. Soil Sci.*, 77(2), 153–157, 1954.



- Prowse, T. D., Conly, F. M., Church, M., and English, M. C.: A review of hydroecological results of the Northern River Basins study, Canada. Part 1. Peace and Slave Rivers, *River Res. Appl.*, 18(5), 429–446, doi: 10.1002/rra.681, 2002.
- Seibert, J., and McDonnell, J. J.: On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration, *Water Resour. Res.*, 38(11), 1241, doi:10.1029/2001WR000978, 2002.
- 5 Singh, V.P. and Frevert, D. K.: *Watershed models*, CRC Taylor & Francis, Boca Raton, 2006.
- Smith A., Welch, C., and Stadnyk, T.: Assessment of a lumped coupled flow-isotope model in data scarce Boreal catchments, *Hydrol. Process.*, doi: 10.1002/hyp.10835, 2016.
- Stadnyk, T. A., Delavau, C., Kouwen, N., and Edwards, T. W. D.: Towards hydrological model calibration and validation: Simulation of stable water isotopes using the isoWATFLOOD model, *Hydrol. Process.*, 27, 3791–3810, doi:10.1002/hyp.9695, 2013.
- 10 Stadnyk-Falcone, T. A.: *Mesoscale Hydrological Model Validation and Verification using Stable Water Isotopes: The isoWATFLOOD Model*. Ph.D. Thesis, University of Waterloo, Waterloo, 386 pp., <http://hdl.handle.net/10012/3970>, 2008.
- Stadnyk, T., St Amour, N. A., Kouwen, N., Edwards, T. W. D., Pietroniro, A., and Gibson, J. J.: A groundwater separation study in boreal wetland terrain: The WATFLOOD hydrological model compared with stable isotope tracers, *Isot. Environ. Health Stud.*, 41(1), 49–60, 2005.
- 15 St Amour, N. A., Gibson, J. J., Edwards, T. W. D., Prowse, T. D., and Pietroniro, A.: Isotopic time-series partitioning of streamflow components in wetland dominated catchments, lower Liard river basin, Northwest Territories, Canada, *Hydrol. Process.*, 19(17), 3357–3381, doi: 10.1002/hyp.5975, 2005.
- 20 Stewart, R. E., Leighton, H. G., Marsh, P., Moore, G. W. K., Ritchie, H., Rouse, W. R., Soulis, E. D., Strong, G. S., Crawford, R. W., and Kochtubajda, B.: The Mackenzie GEWEX Study: the water and energy cycles of a major North American river basin, *Bull. Am. Meteorol. Soc.*, 79, 2665–2684, doi: 10.1175/1520-0477(1998)079<2665:TMGSTW>2.0.CO;2, 1998.
- Sturm, K., Hoffmann, G., and Langmann, B.: Simulation of the stable water isotopes in precipitation over South America: Comparing regional to global circulation models, *J. Clim.*, 20(15), 3730–3750, doi: 10.1175/JCLI4194.1, 2007.
- 25 Sturm, K., Hoffmann, G., Langmann, B., and Stichler, W.: Simulation of d18O in precipitation by the regional circulation model REMOiso, *Hydrol. Process.*, 19(17), 3425–3444, doi: 10.1002/hyp.5979, 2005.
- Tetzlaff, D., Buttle, J., Carey, S. K., McGuire, K., Laudon, H., and Soulsby, C.: Tracer-based assessment of flow paths, storage and runoff generation in northern catchments: a review, *Hydrol. Process.*, 29, 3475–3490, doi: 10.1002/hyp.10412, 2015.
- 30 Tetzlaff, D., Uhlenbrook, S., Eppert, S., and Soulsby, C.: Does the incorporation of process conceptualization and tracer data improve the structure and performance of a simple rainfall-runoff model in a Scottish mesoscale catchment?, *Hydrol. Process.*, 22, 2461–2474, doi: 10.1002/hyp.6841, 2008.
- Töyryä J.: Generation of a digital elevation model and establishment of a database for the Fort Simpson study area, No. KW504-7-0074, National Hydrology Research Institute, Saskatoon, Saskatchewan, 1997.
- 35 Xi, X.: A Review of Water Isotopes in Atmospheric General Circulation Models: Recent Advances and Future Prospects, *Int. J. Atmos. Sci.*, 250920, doi:10.1155/2014/250920, 2014.

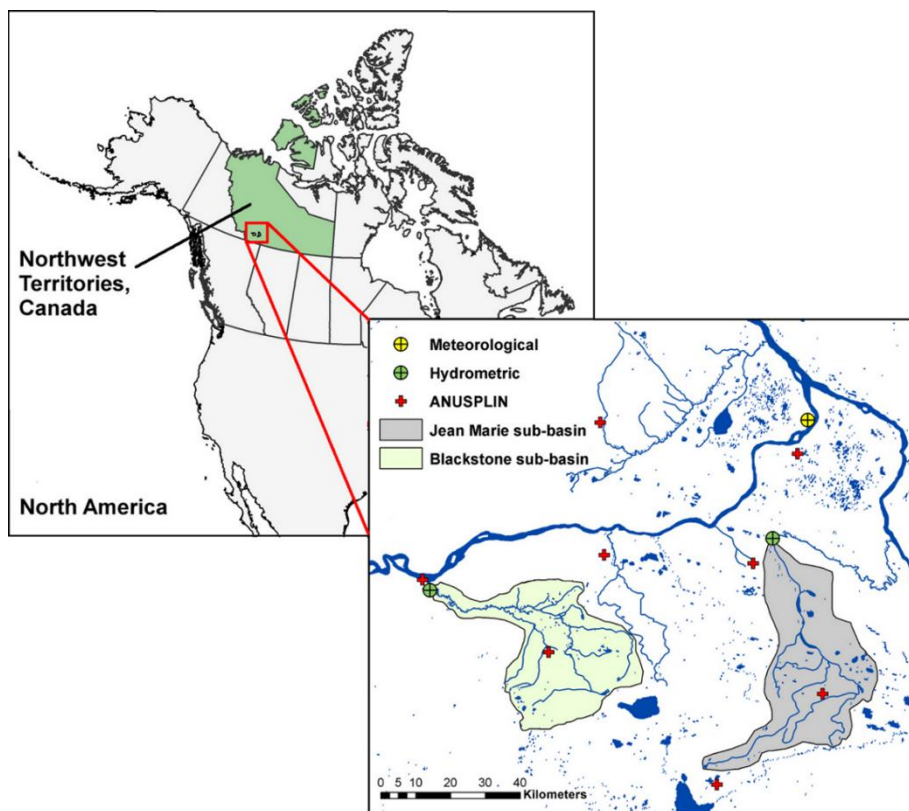
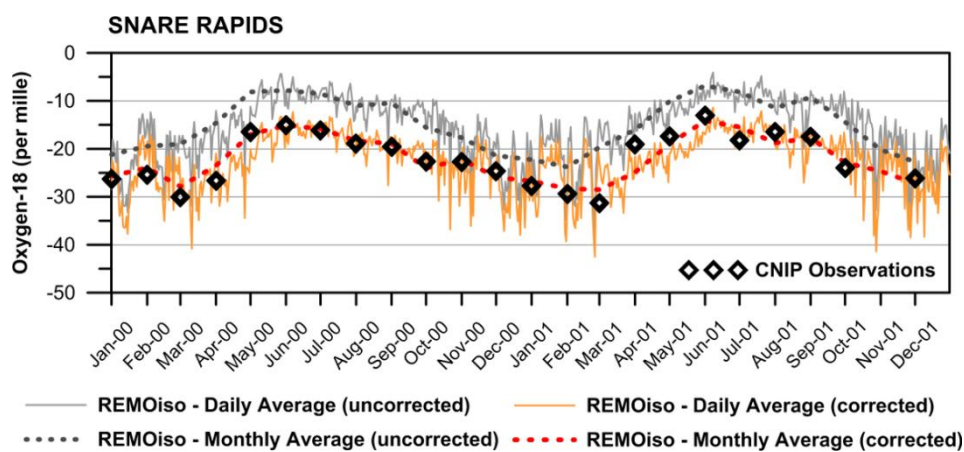


Figure 1: Fort Simpson River Basin (all other tributaries of the Liard and Mackenzie Rivers have been removed for ease of viewing).



5

Figure 2: Comparison of raw and corrected REMOiso  $\delta^{18}\text{O}_{\text{ppt}}$  output with CNIP monthly compositions at Snare Rapids, NWT.

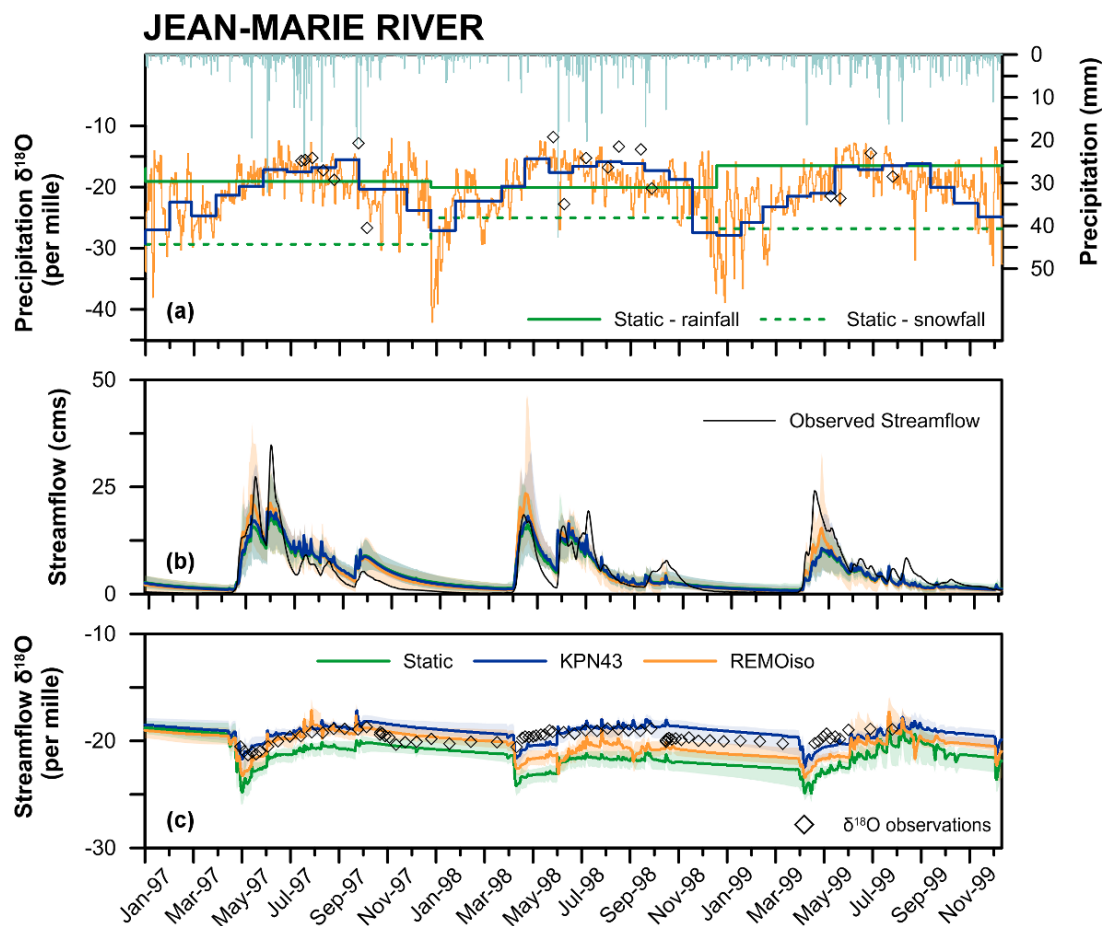


Figure 3: Input and behavioural simulations for Jean Marie, including: (a) KPN43, REMOiso and static  $\delta^{18}\text{O}_{\text{ppt}}$  input time series and daily precipitation; and simulated (b) mean daily streamflow and uncertainty bounds and (c) mean daily  $\delta^{18}\text{O}_{\text{SF}}$  and uncertainty bounds, for KPN43, REMOiso and static driven model calibrations.  $\delta^{18}\text{O}_{\text{ppt}}$  input-specific uncertainty bounds are represented as the shaded regions, with shading colour corresponding to  $\delta^{18}\text{O}_{\text{ppt}}$  type.

5

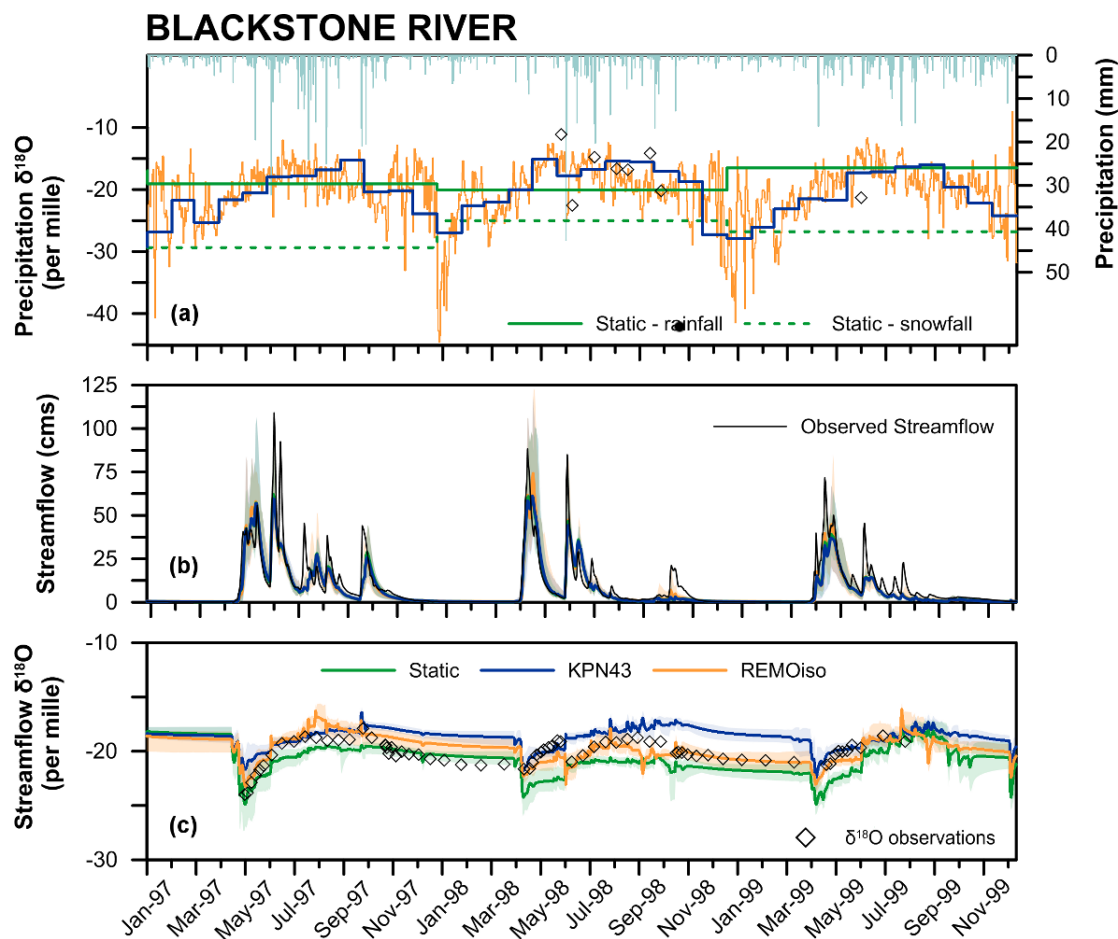


Figure 4: Input and behavioural simulations for Blackstone, including: (a) KPN43, REMOiso and static  $\delta^{18}\text{O}_{\text{ppt}}$  input time series and daily precipitation; and simulated (b) mean daily streamflow and uncertainty bounds and (c) mean daily  $\delta^{18}\text{O}_{\text{SF}}$  and uncertainty bounds, for KPN43, REMOiso and static driven model calibrations.  $\delta^{18}\text{O}_{\text{ppt}}$  input-specific uncertainty bounds are represented as the shaded regions, with shading colour corresponding to  $\delta^{18}\text{O}_{\text{ppt}}$  type.

5

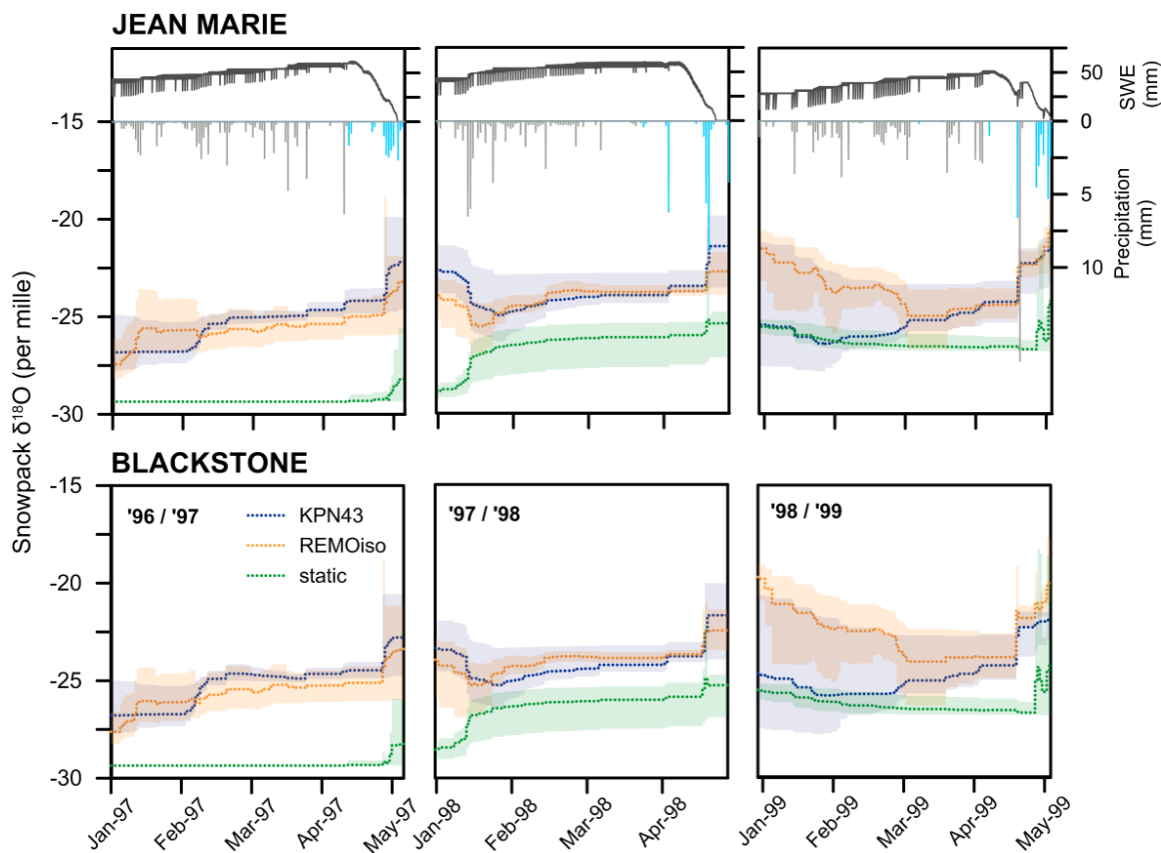
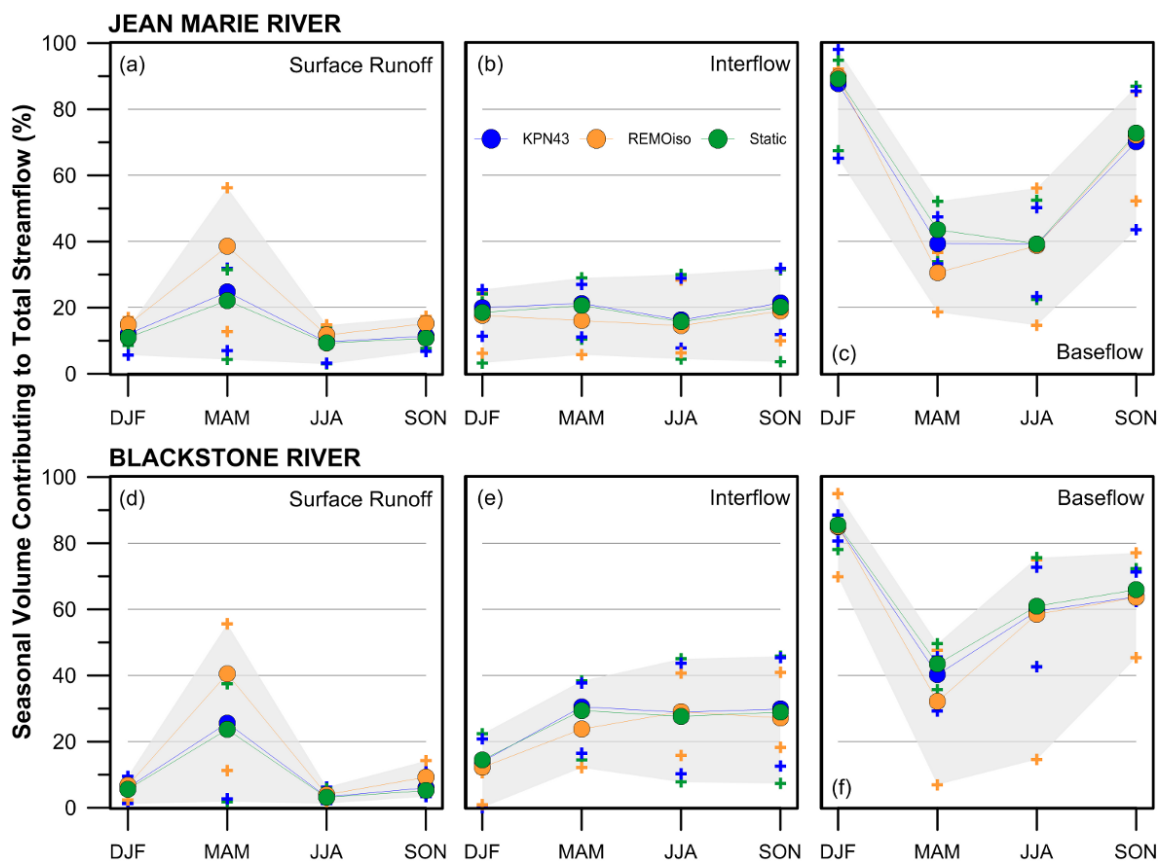


Figure 5: Precipitation-weighted  $\delta^{18}\text{O}$  of snowpack ( $\delta^{18}\text{O}_{\text{SNW}}$ ) for KPN43, REMOiso and static inputs from January to the end of melt for each year of the study period. Snow water equivalent (SWE), snowfall (gray line) and rainfall (blue line) are also shown.  $\delta^{18}\text{O}_{\text{ppt}}$  input-specific uncertainty bounds are represented as the shaded regions.



**Figure 6:** Percent seasonal volume contributing to total streamflow from surface runoff, interflow and baseflow storages for each season. Cross symbols represent the 5<sup>th</sup> and 95<sup>th</sup> percentiles for each forcing method, and circle symbols signify the mean values. The combined uncertainty bounds representing the 5<sup>th</sup> and 95<sup>th</sup> simulations from all three  $\delta^{18}\text{O}_{\text{ppt}}$  input types are shaded in gray.



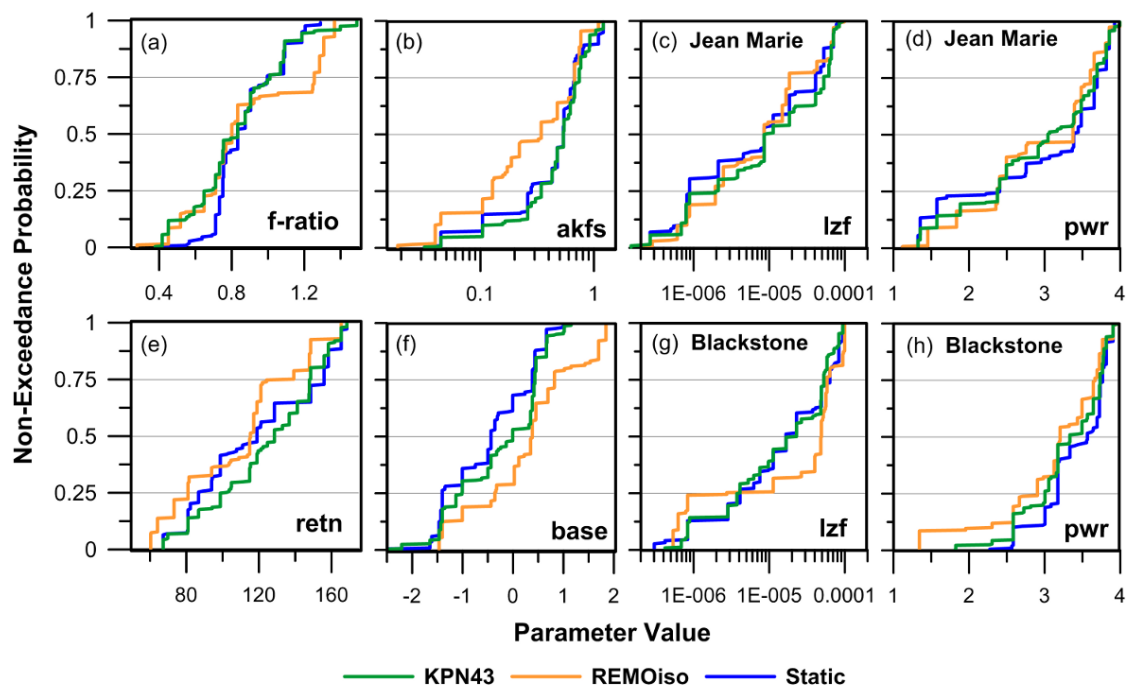


Figure 7: Probability distributions for select parameters, as indicated in the bottom right corner of each panel. Parameters are from behavioural simulations, and (a), (b), (e) and (f) have been weighted to the land cover distribution within Jean Marie and Blackstone, as outlined in Table 1. Panels (c) and (d) and river class parameters within Jean Marie, and panels (g) and (h) contain river class parameters for Blackstone.



**Table 1: Basin characteristics, including land cover classification, area, and average basin slope (recreated from data provided in St Amour et al., 2005)**

Sub-basin	Land Cover Classification (%)						Area (km <sup>2</sup> )	Basin Slope (%)
	Deciduous	Mixed	Coniferous	Transitional	Wetland	Water		
Jean- Marie River	5	22	23	31	14	1.3	1310	0.3
Blackstone River	7	17	14	39	21	0.7	1390	0.63



**Table 2: Static  $\delta^{18}\text{O}_{\text{ppt}}$  input compositions of annual rainfall and snowfall oxygen-18 for isoWATFLOOD.**

<b>Year</b>	<b><math>\delta^{18}\text{O}</math> rainfall (‰)</b>	<b><math>\delta^{18}\text{O}</math> snowfall (‰)</b>
<b>1996</b>	-17.00	-29.35
<b>1997</b>	-19.10	-29.35
<b>1998</b>	-20.10	-25.03
<b>1999</b>	-16.52	-26.79



**Table 3: Seasonal REMOiso bias correction.**

<b>Season</b>	<b>Correction (%)</b>
<b>November – December – January – February</b>	-4.5
<b>March – April – May</b>	-8.9
<b>June – July – August</b>	-7.3
<b>Septemer – October</b>	-8.5
<b>Average Correction Applied:</b>	-7.0



**Table 4: Data summary for the study period (SP) and period of record (PoR). The coefficient of variation (CV) is calculated as the ratio of the standard deviation to the mean.**

Variable (gauge ID)	Unit	Number of Measurements	Mean (SP, PoR)	CV (SP, PoR)	SP Range (min, max)
<b>Hydrometric/Meteorological Data</b>					
<b>Daily Average Streamflow Jean Marie (10FB005)</b>	m <sup>3</sup> /s	1095	4.66, 5.25	1.24, 2.06	0.19, 34.9
<b>Daily Average Streamflow Blackstone (10ED007)</b>	m <sup>3</sup> /s	1095	8.96, 10.76	1.65, 2.17	0.04, 109
<b>Mean Daily Air Temperature Fort Simpson (2202101)</b>	°C	1093	-1.5, -3.02	N/A	-40.8, 25.3
<b>Daily Precipitation Fort Simpson (2202101)</b>	mm	1088	1.12, 1.01	3.04, 3.19	0.0, 43.0
<b>Hourly Relative Humidity* Fort Simpson (2202101)</b>	%	26280	73.9	0.24	14, 100
<b>Isotopic Measurements*</b>					
<b>Streamflow δ<sup>18</sup>O - Jean Marie</b>	‰	71	-19.70	0.03	-21.34, -18.72
<b>Streamflow δ<sup>18</sup>O - Blackstone</b>	‰	69	-20.17	0.06	-24.01, -17.92
<b>Rainfall δ<sup>18</sup>O Jean Marie and Blackstone</b>	‰	27	-17.55	0.23	-26.70, -11.12
<b>Precipitation δ<sup>18</sup>O Forcing*</b>					
<b>KPN43 δ<sup>18</sup>O<sub>ppt</sub> input</b>	‰	1800 (36 values at 50 grid points)	-20.48	0.19	-28.86, -13.91
<b>REMOiso δ<sup>18</sup>O<sub>ppt</sub> input</b>	‰	54750 (1095 values at 50 grid points)	-21.78	0.25	-42.82, -10.68
<b>Static δ<sup>18</sup>O<sub>ppt</sub> input</b>	‰	300 (6 values at 50 grid points)	-22.82	0.20	-29.35, -16.52

\* Provided only for the study period, 1997 – 1999.



**Table 5: Initialization values for FSB isoWATFLOOD simulations.**

<b>Variable</b>	<b>Description</b>	<b>Value (‰)</b>
$\delta^{18}\text{O}_{\text{sf}}$	Background delta for river water initialization	-13.52
$\delta^{18}\text{O}_{\text{IF}}$	Background delta for soil water initialization	-14.60
$\delta^{18}\text{O}_{\text{GW}}$	Background delta for groundwater initialization	-20.00
$\delta^{18}\text{O}_{\text{SNW}}$	Background delta for snow initialization	-22.00



**Table 6: Parameters included in the Monte Carlo calibration, alongside a description of what the parameter represents and the algorithm it is used within.**

Name	Description	Algorithm
<b>Routing Parameters</b>		
<b>flz</b>	Lower zone drainage function	An exponential ground water depletion function that gradually diminishes the base flow. Ground water is replenished by drainage of the UZS:
<b>pwr</b>	Lower zone drainage function exponent	$QLZ = LZS * (LZS)^{PWR}$ Where: LZS is lower zone storage QLZ is the baseflow flux
<b>theta</b>	Wetland porosity	Physically-based wetland routing algorithm (McKillop et al., 1999)
<b>kcond</b>	Conductivity parameter	
<b>Hydrologic Parameters</b>		
<b>f-ratio</b>	Interception capacity multiplier	Conceptual evaporation algorithm based on Hargreaves and Samani (1982). f-ratio is a multiplier for the interception capacity for each land class.
<b>ak</b>	Surface permeability (bare ground)	Conceptual infiltration algorithm (similar to Green and Ampt, 1911); but based on Richard's equation which is physically-based (Philip, 1954)
<b>akfs</b>	Surface permeability	
<b>rec</b>	Interflow coefficient	Interflow is represented by a simple storage-discharge relation: $DUZ = REC * (UZS - RETN) * Si$ Where: UZS = upper zone storage
<b>retn</b>	Upper zone retention [mm]	DUZ = depth of upper zone storage released as interflow Si = internal land surface slope
<b>ak2</b>	Recharge coefficient (bare ground)	Upper zone to lower zone drainage is represented by a simple storage-discharge relation: $DRNG = AK2 * (UZS - RETN)$ Where: DRNG is the drainage from UZS to LZS
<b>mf</b>	Melt factor [mm/°C/hr]	$M = MF (T_a - base)$ Anderson (1976)
<b>base</b>	Base Temperature [°C]	
<b>sub</b>	Sublimation factor	Sublimation is modelled by a static sublimation factor. Amount of sublimation is a fraction of the observed snowfall. For new model setups, the sublimation factor has been replaced by a static sublimation rate.



**Table 7: Average simulation statistics from n behavioural simulations for streamflow and  $\delta^{18}\text{O}_{\text{SF}}$  for the three model calibrations (using KPN43, REMOiso, and static inputs).**

<b>Average statistics from n behavioural simulations</b>	<b>KPN43</b>	<b>REMOiso</b>	<b>Static</b>
<b>n</b>	321 / 30000	268 / 30000	216 / 30000
<b>Streamflow (1095 observations for performance evaluation)</b>			
<b>NSE</b>	0.68	0.68	0.69
<b> % Dv </b>	13.9	13.4	14.2
<b> Log(% Dv) </b>	11.5	8.9	11.6
<b><math>\delta^{18}\text{O}_{\text{SF}}</math> (140 observations for performance evaluation)</b>			
<b>RMSE (‰)</b>	1.39	1.32	2.09
<b>KGE</b>	0.36	0.33	0.35





**Table 8: Allowable parameter ranges and resulting averaged parameter characteristics for behavioural simulations retained from the three model calibrations (KPN43, REMOiso, and static inputs). Parameters are summarized as: median (minimum, maximum).**

Parameter	Allowable Range	KPN43	REMOiso	static
<b>Jean Marie Sub-basin</b>				
<b>flz</b>	[ $1 \times 10^{-7}$ , $1 \times 10^{-4}$ ]	$8.60 \times 10^{-6}$ ( $1.49 \times 10^{-7}$ , $9.32 \times 10^{-5}$ )	$8.60 \times 10^{-6}$ ( $2.03 \times 10^{-7}$ , $9.86 \times 10^{-5}$ )	$8.60 \times 10^{-6}$ ( $2.03 \times 10^{-7}$ , $9.86 \times 10^{-5}$ )
<b>pwr</b>	[1, 4]	3.05 (1.33, 4.00)	3.38 (1.12, 4.00)	3.39 (1.12, 3.96)
<b>theta</b>	[0.1, 1.0]	0.60 (0.13, 0.96)	0.61 (0.13, 1.0)	0.60 (0.13, 0.98)
<b>kcond</b>	[0.1, 1.5]	0.79 (0.17, 1.46)	0.68 (0.17, 1.46)	0.86 (0.17, 1.46)
<b>Blackstone Sub-basin</b>				
<b>flz</b>	[ $1 \times 10^{-7}$ , $1 \times 10^{-4}$ ]	$2.30 \times 10^{-5}$ ( $4.19 \times 10^{-7}$ , $9.41 \times 10^{-5}$ )	$4.91 \times 10^{-5}$ ( $5.36 \times 10^{-7}$ , $9.95 \times 10^{-5}$ )	$1.67 \times 10^{-5}$ ( $3.02 \times 10^{-7}$ , $9.86 \times 10^{-5}$ )
<b>pwr</b>	[1, 4]	3.34 (1.82, 3.96)	3.21 (1.35, 3.96)	3.57 (2.28, 3.91)
<b>theta</b>	[0.1, 1.0]	0.55 (0.11, 1.0)	0.60 (0.13, 1.0)	0.52 (0.24, 1.0)
<b>kcond</b>	[0.1, 1.5]	0.69 (0.11, 1.41)	0.80 (0.21, 1.48)	0.69 (0.17, 1.49)
<b>Land Cover Weighted-Average Parameter results: median (minimum, maximum)</b>				
<b>fratio</b>	[0.1, 2.5]	0.70 (0.12, 2.22)	0.70 (0.15, 2.16)	0.80 (0.12, 2.23)
<b>ak</b>	[1, 50]	21.6 (2.1, 47.6)	25.8 (2.4, 47.5)	23.6 (1.2, 46.9)
<b>akfs</b>	[0.005, 2]	0.212 (0.006, 1.878)	0.059 (0.006, 1.724)	0.203 (0.006, 1.850)
<b>rec</b>	[0.05, 1]	0.47 (0.08, 0.90)	0.46 (0.08, 0.88)	0.43 (0.09, 0.90)
<b>retn</b>	[10, 200]	122 (18, 189)	119 (23, 181)	114 (20, 186)
<b>ak2</b>	[.001, 0.2]	0.013 (0.001, 0.188)	0.008 (0.001, 0.172)	0.021 (0.001, 0.184)
<b>fm</b>	[0.075, 0.2]	0.117 (0.076, 0.189)	0.119 (0.078, 0.190)	0.112 (0.076, 0.189)
<b>base</b>	[-3.5, 3.5]	-0.20 (-3.28, 3.14)	0.34 (-2.81, 3.13)	-0.35 (-3.11, 2.85)
<b>sub</b>	[0.1, 1.1]	0.53 (0.11, 1.05)	0.43 (0.13, 1.05)	0.50 (0.11, 1.05)