Examining the impacts of precipitation isotope input ($\delta^{18}O_{ppt}$) on distributed, tracer-aided hydrological modelling

Carly J. Delavau[1], Tricia Stadnyk [1], Tegan Holmes[1]

[1]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 assist with process understanding and source separation; which facilitates model calibration and diagnosis of model uncertainty (Tetzlaff et al. 2015; Klaus & McDonnell, 2013). Data availability in high-latitude regions, however, proves to be a major challenge associated with this type of application (Tetzlaff et al., 2015). Models require a time series of isotopes in precipitation ($\delta^{18}O_{ppt}$) to drive simulations, and throughout much of the world - particularly in sparsely populated high-latitude regions - these data are not widely available. Here we investigate the impact that choice of precipitation isotope product ($\delta^{18}O_{ppt}$) has on simulations of streamflow, $\delta^{18}O$ in streamflow ($\delta^{18}O_{SF}$), resulting hydrograph separations and model parameters. In a high-latitude, data sparse, seasonal basin (Fort Simpson, NWT, Canada), we assess three precipitation isotope products of different spatial and temporal resolution (i.e., semi-annual static, seasonal KPN43, and daily bias corrected REMOiso), and apply them to force the isoWATFLOOD tracer-aided hydrologic model. Total simulated streamflow is not significantly impacted by choice of $\delta^{18}O_{ppt}$ product, however, simulated isotopes in streamflow ($\delta^{18}O_{SF}$) and the internal apportionment of water (driven by model parameterization) are impacted. The highest resolution product (REMOiso) was distinct from the two lower resolution products (KPN43 and static), but could not be verified as correct due to a lack of daily $\delta^{18}O_{ppt}$ observations. The resolution of $\delta^{18}O_{ppt}$ impacts model parameterization and seasonal hydrograph separations, producing notable differences among simulations following large snowmelt and rainfall events when event compositions differ significantly from $\delta^{18}O_{SF}$. Capturing and preserving the spatial variability in $\delta^{18}O_{ppt}$ using distributed tracer-aided models is important because this variability impacts model parameterization. We achieve an understanding of tracer-aided modelling and its application in high-latitude regions with limited $\delta^{18}O_{ppt}$ observations, and the value such models have in defining modelling uncertainty. In this study, application of a tracer-aided model is able to identify simulations with improved internal process representation, reinforcing that tracer-aided modelling approaches assist with resolving hydrograph component contributions and work towards diagnosing 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 for calibration and

validation of hydrological models when there is: (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) issues with equifinality affecting model parameterization; and (3) uncertainty in model results (e.g., Beven and Binley, 1992; Kirchner, 2006; Fenicia et al., 2008; Dunn et al., 2008).

5

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.
10 Conservative tracer data provides insights into the dominant hydrological processes and integrated runoff response (in northern catchments: Birks and Gibson, 2009; Tezlaff 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,
15 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^2). This has resulted in distributed studies at the regional-scale ($>10^3$ km^2) left largely unexplored, with the exception of a few, select applications (Stadnyk et al.,
20 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.

25

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 ($\delta^{18}O_{ppt}$) 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
30 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 $\delta^{18}O_{ppt}$ observations highlights the need for alternative methods to provide estimates of stable isotopes in precipitation for tracer-aided model input (termed '$\delta^{18}O_{ppt}$ products'). Options include empirically-based models generating

gridded time series of precipitation isotopes (e.g., Lykoudis et al., 2010; Delavau et al., 2015), and isotope-enabled climate model output (for a comprehensive overview, see: Noone and Sturm, 2010; Xi 2014).

Small-scale catchment studies rely on continuous records of $\delta^{18}O_{ppt}$ 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 $\delta^{18}O$ 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 input is spatially invariant is not preferable, as $\delta^{18}O_{ppt}$ 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).

This study aims to explore how varying spatial and temporal resolutions of precipitation isotope products, or $\delta^{18}O_{ppt}$ input, impact regional tracer-aided model simulations and parameterization. Forcing a tracer-aided, distributed hydrological model (isoWATFLOOD) with three precipitation isotope products, we examine how the different $\delta^{18}O_{ppt}$ products impact the:

    a)   simulation of total streamflow and its isotopic variability ($\delta^{18}O_{SF}$);

    b)   internal apportionment of water, namely the seasonality of hydrograph separation; and,

    c)   model parameterization and simulation uncertainty.

We explore the impact that varying the resolution of $\delta^{18}O_{ppt}$ inputs has on the capability of the model to reproduce observed $\delta^{18}O_{SF}$ variability; and the usefulness of a tracer-aided modelling approach to help inform and quantify simulation equifinality.

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 $\delta^{18}O_{ppt}$ input. The study period of 1997–1999 is selected based on data availability.

This study considers two sub-basins of the greater Fort Simpson basin: the Jean-Marie (1310 km^2) and Blackstone River (1390 km^2) sub-basins (Fig. 1). The basins vary in relief from 0.3 % in the Jean-Marie sub-basin to 0.63 % for the Blackstone sub-basin, on average. 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 $\delta^{18}O_{ppt}$ input selection on these processes can be examined.

5   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.

10

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

15   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

20   surface fitting method that has been found well suited to the interpolation of various climate variables, particularity in data-sparse, high-elevation regions as the method accounts for spatially varying dependencies on elevation (McKenney et al., 2011). We have validated ANUSPLIN against independent station observations (precipitation and temperature) across the Prairies and Boreal regions of Canada as a precipitation forcing for hydrologic modelling. It has been found adequate ($r \geq 0.98$) for the purpose of short-term modelling studies. An inverse-distance weighting approach is used to spatially distribute the daily

25   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 (Petrone et al., 2000; St Amour et al., 2005).

30   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. Winter (ice-on) flows tend to be very low given highly seasonal, high-latitude hydrology, underlying discontinuous permafrost, and the absence of mid-winter melt (St. Amour et al., 2005). Averaged winter ice-on flows from 1997-1999 were 0.194 m^3/s and 0.034 m^3/s for the Jean Marie and Blackstone sub-basins, respectively. A statistical summary of observations used in this study is provided in Table 2.

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 δ^{18}O and δ^2H of streamflow, rainfall, snowpack, and surface waters (wetlands and lakes) during the open water season (May to October). During ice-on 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 δ^{18}O compositions in streamflow in the Jean-Marie (n = 71) and Blackstone (n = 69) sub-basins for model calibration. Although δ^{18}O$_{ppt}$ compositions (n = 27) were collected as part of the GEWEX sampling campaign, these data are not preferred for tracer-aided hydrologic model input due to their spatial uniformity and poor temporal resolution. Observations are incorporated into this study as the 'static' δ^{18}O$_{ppt}$ input, and as a means to validate the KPN43 and REMOiso products and to inform the static precipitation product. The number of measurements and their statistical properties are summarized in Table 2. Isotopic compositions of δ^{18}O are expressed in delta (δ) notation as a deviation from VSMOW (Vienna Mean Standard Mean Ocean Water) in units of per mille (‰),such that $\delta_{water} = (R_{water}/R_{VSMOW} - 1)$ x 1000 ‰, where R is ^{18}O/^{16}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 analytical uncertainties of ± 0.1 ‰ for δ^{18}O.

2.4 Precipitation oxygen-18 input

The precipitation isotope products evaluated in this study represent a variety of spatial and temporal scales, and were selected because they are commonly available for all tracer-aided hydrologic modelling applications. The first type of input used in this study is annual average δ^{18}O$_{ppt}$ compositions of rainfall and snowfall for each year of simulation (i.e., yearly resolution). Values for the FSB were obtained by averaging observations of δ^{18}O in rainfall and the snowpack obtained from the GEWEX study (Table 2; Table 3). δ^{18}O$_{ppt}$ compositions were assumed constant throughout the study domain (i.e., spatially uniform). Due to a lack of snowfall data collected during this study, we assumed the average annual isotopic composition of the snowpack was representative of the snowfall composition, as has been done in other data sparse, high-latitude tracer-aided modelling studies (Smith et al., 2015; Smith et al., 2016; Holmes, 2016; Stadnyk et al., 2013). It is well established in the literature that the isotopic composition of snowfall is not necessarily equal to the average annual composition of the snowpack (due to sublimation and snow metamorphism; Zhou et al., 2008; Taylor et al., 2001; 2002). The high latitude of our study site, however, makes freeze/thaw cycling during the winter rare, making this assumption more reasonable. Due to the averaged values and lack of spatial variability, this product is referred to as 'static' throughout the remainder of the manuscript, and consists of two

constant $\delta^{18}O_{ppt}$ values (rain and snow) for each year. This product is specifically designed and evaluated for remote regions that lack spatially and temporally varying $\delta^{18}O_{ppt}$ observations.

Times series simulations obtained from the KPN43 model created by Delavau et al. (2015) are used as the second type of
5 $\delta^{18}O_{ppt}$ product in this study. The KPN43 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 time step. This product is generated at a 10 km resolution (to mirror model set-up), and varies spatially throughout the study domain due to the variation in the climatic predictors and geographic information required to produce simulations.
10

The third $\delta^{18}O_{ppt}$ product included 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 $\delta^{18}O_{ppt}$ output is available at a 55 km spatial resolution and a 6h time step. REMOiso output is averaged in this study, however, to a daily time step, as the range and variability of sub-daily $\delta^{18}O_{ppt}$ are erroneously large, and the resolution of streamflow oxygen-18 calibration data do not warrant a temporal
15 frequency of input finer than daily.

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). isoWATFLOOD is an extension of the WATFLOOD hydrological model, whereby water and oxygen-18 are simultaneously
20 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
25 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 4.

isoWATFLOOD requires the $\delta^{18}O$ of precipitation (either rain and snow separately, or total precipitation) and can utilize (though does not require) distributed relative humidity inputs to force the model. Additionally, $\delta^{18}O$ compositions for
30 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 storage initializations are derived from measured data obtained during the GEWEX campaign and reported by St Amour et al. (2005). These include streamflow (-13.52 ‰), interflow (soil

water; -14.60 ‰), baseflow (-20.00 ‰), and snowpack (-22.00 ‰) background compositions. 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 isoWATFLOOD model used in this study is based on a previous version reported by Stadnyk et al. (2013). The current version used here is an updated version of isoWATFLOOD code, and the watershed set-up incorporates various model improvements made since 2013, independent of this study. 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 $\delta^{18}O_{SF}$. 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 4; Table S-1). Allowable ranges for each parameter are determined based on published values alongside personal communications with N. Kouwen (Kouwen, 2014) (Table S-1).

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 S-1). Random parameter sampling is completed 30,000 times, generating 30,000 unique parameter sets for isoWATFLOOD model evaluation.

  ii. For each of the three $\delta^{18}O_{ppt}$ inputs (KPN43, REMOiso and static), streamflow and $\delta^{18}O_{SF}$ are simulated from 1996 to 1999 for all 30,000 parameter sets (as defined in (i)).

  iii. Simulated streamflow and $\delta^{18}O_{SF}$ 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) (Beven & Binley, 1992) based on meeting (or not) the following set of efficiency criteria thresholds, defined in detail below, for simulated streamflow and $\delta^{18}O_{SF}$:

    a. Streamflow:

         $NSE \geq 0.5$;

         $|\%Dv| \leq 20$ %, and;

$$|\log(\%Dv)| \le 20\ \%.$$

b. $\delta^{18}O_{SF}$:

$$RMSE \le 2.5\ ‰,\ and;$$

$$KGE >= 0.3.$$

5   Behavioural thresholds used in this study are subjectively defined, but 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.

10   A multi-criteria model evaluation approach 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 $\delta^{18}O_{SF}$ assessment due to the
temporal discontinuity of the isotope observations, therefore RMSE and KGE are used as isotopic simulation statistics. The
KGE statistic puts less emphasis on peak flow differences by providing a more balanced approach where error is first summed

15   and then squared at the end, preserving the sign of the error and enabling a trade-off of error throughout the simulation period
(Gupta et al., 2009). It should also be noted that $\delta^{18}O_{SF}$ 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 %), with the remaining 52 % of observations sampled during
summer. The sporadic distribution of observations may result in the calibrations more highly weighted to certain periods of

20   the year and the dominate processes occurring at that time; therefore having the potential to impact model parameterization.

3.3 REMOiso bias correction

Due to a lack of published studies evaluating REMOiso performance within Canada, a comparison between REMOiso 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

25   Precipitation (GNIP) database and can be accessed at: http://www.iaea.org/water (IAEA/WMO, 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 $\delta^{18}O_{ppt}$ observations spanning the years of 1997–2010. 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. For bias-correction purposes, daily REMOiso simulations are averaged to monthly

30   compositions for direct comparison to CNIP data using the precipitation amount-weighting approach in Eq. (1):

$$\boldsymbol{\delta^{18}O_{ppt\ monthly} = \sum P_i * (\delta^{18}O_{ppt})_i\ /\ \sum P_i} \tag{1}$$

8

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 under the Canadian Network for Isotopes in Precipitation (CNIP).

5 Uncorrected REMOiso simulations exhibit a positive bias in this region (Fig. 2), which is expected based on the ECHAM4 mean annual $\delta^{18}O_{ppt}$ 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. Corrected monthly and daily REMOiso output at Snare Rapids are displayed on Figure 2 as the dashed red and solid orange lines, respectively.

10 For the current study, daily REMOiso output for the Fort Simpson region is bias corrected with the seasonal correction values, ranging from -4.5 ‰ (NDJF) to -8.9 ‰ (MAM), with an average of -7.0 ‰.

The statistical properties of the corrected daily REMOiso simulations, alongside the KPN43 monthly simulations and the static seasonal averages are summarized in Table 2.

15 **3.4 Statistical treatment of data**

For discussion and analysis purposes (Section 4.2 to 4.4), results represent only the behavioural simulations derived from each $\delta^{18}O_{ppt}$ product. Uncertainty bounds are the 5th and 95th percentiles drawn from the ensembles of behavioural simulations; denoted as the shaded bounds around each model's mean simulation.

20 Kendall's tau coefficient (τ) is a non-parametric test used to compare the level of correlation between two variables. We compute Kendall's tau for the mean daily streamflow and $\delta^{18}O_{SF}$ simulations derived from the three inputs. By computing τ coefficients for pairs of simulated time series (i.e., REMOiso versus KPN43, REMOiso versus static, and KPN43 versus static), we can statistically evaluate the similarity of model output derived from different precipitation isotope products.

25 Parameter probability distributions (Table 4) are arrived at by first weighting behavioural parameters for each land cover type to their corresponding percent coverage within the modelled sub-basins. Land cover weighted parameter values are then ranked and non-exceedance probabilities determined. Routing parameter distributions for each sub-basin are arrived at using a similar approach, but are not weighted by coverage. The non-parametric Kolmogorov–Smirnov (K-S) test is used to assess if behavioural parameter distributions are considered to be from the same distribution.

30

Spatially distributed precipitation isotope product maps (Fig. S-1) represent daily precipitation isotope averaged across seasons (DJF, MAM, JJA, SON), and are precipitation amount-weighted using WATFLOOD precipitation input (interpolated Environment Canada Canadian Daily Climate Data, housed in WATFLOODs radcl.r2c files; Kouwen 2014). Maps are

generated overlapping the model grid (10k) for the entire FSB domain, which includes the Jean Marie and Blackstone sub-basins.

4 Results and discussion

Results of the three calibrations indicate that choice of $\delta^{18}O_{ppt}$ input influences the number of simulations that meet behavioural criteria thresholds. The KPN43 product results in more behavioural simulations (n = 321) relative to the REMOiso (n = 268) or static (n = 216) products (Table 5). This also implies that choice of $\delta^{18}O_{ppt}$ input influences the models internal apportionment of water (i.e., hydrograph separations) via model parameters. Among input types, potentially significant differences in several parameters were noted (Table S-1), and is discussed in Section 4.4. In almost all instances, the ranges of the parameters were not significantly constrained from the allowable parameter ranges, yielding confidence in our simulated parameter uncertainty envelopes.

4.1 Precipitation oxygen-18 input

Of the three $\delta^{18}O_{ppt}$ products, KPN43 input is on average the most enriched (-20.48 ‰), followed by REMOiso (-21.78 ‰), and static as the most depleted (-22.82 ‰) (panel (a), Fig. 3 and 4). The KPN43 and static products show similar variation about their means, with CVs 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 is zero for the static product as annual snow and rainfall compositions are spatially averaged across the domain. Spatial variation among sub-basins is noted in the KPN43 and REMOiso products. Both the KPN43 and REMOiso products show, on average, more depleted $\delta^{18}O_{ppt}$ values within Blackstone (-20.79 ‰ and -22.01 ‰, respectively) in comparison to Jean Marie (-20.17 ‰ and -21.54 ‰, respectively), likely caused by the higher elevation headwaters of Blackstone relative to Jean Marie (a maximum difference of ~215 m). Figure S-1 provides seasonally averaged, spatially distributed maps for each product. Averaged spatial variability is greatest for the KPN43 forcing, followed by REMOiso, and is constant for the static product. REMOiso shows less long-term average variability because its temporal variability is greater, resulting in more chaotic (randomized) signals of $\delta^{18}O_{ppt}$ that produce weaker long-term signals when averaged over time. KPN43, on the other hand, exhibits more consistent spatial patterning of $\delta^{18}O_{ppt}$ variability, resulting in stronger signals of long-term variability on a per-grid basis (Fig. S-1). REMOiso input is derived on a 55 km grid, meaning that approximately 5 isoWATFLOOD grids are equivalent to 1 REMOiso grid, which also results in a terrain (variability) smoothing effect. The static input exhibits seasonal variability caused by the different compositions of rain and snow, and mixed shoulder season compositions (MAM and SON) when both rain and snow occur.

Although there are only 19 rainfall $\delta^{18}O$ observations collected over the study period in Jean Marie, and eight within Blackstone (hollow black diamonds on Fig. 3 and Fig. 4, panel (a)), these limited data provide some information regarding the accuracy

10

of the products. By visual inspection, each of the three products generates reasonable estimates of $\delta^{18}O_{ppt}$. This is expected for the static input, which is derived directly from these observations; however, this provides qualitative validation for KPN43 and REMOiso. REMOiso is the only product that can somewhat replicate event-scale variability in $\delta^{18}O_{ppt}$ due to its daily time step. The KPN43 product appears to represent the average composition of summer rainfall events, and displays reasonable seasonal variability. There are insufficient observations to statistically validate these products within the Fort Simpson basin. The semi-annual static input perhaps does a reasonable job of reflecting $\delta^{18}O_{ppt}$ seasonality because of the high-latitude location of the basin, where much shorter shoulder seasons exist.

4.2 Modelling streamflow

All calibrations adequately capture variations in total streamflow in both sub-basins, as emphasised by the regional calibration statistics (Table 5). 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 Figure 3 and Figure 4. Differences in hydrograph characteristics between Jean Marie and Blackstone result from 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).

Within the Jean Marie, both the timing and volume of peak flows derived from snow melt are well captured in 1998, however, volume is under predicted in 1997 and 1999 for the average streamflow simulation. The parameter uncertainty bounds generally enclose the observed spring melt hydrograph, except in 1999 where the timing of the melt peak is simulated later than was observed. Snowmelt is controlled by a degree-day snowmelt function in WATFLOOD, using temporally constant snowmelt parameters. Parameter stationarity likely results in an inadequate description of the inter-annual variability in energy balance and snowpack ripening dynamics within the model. All simulations have difficulty capturing the volume of the snowmelt recession limb, which may be caused by the parameterization of baseflow and fen responses in this sub-basin. Based on previous studies (Connon et al., 2015), it has been suggested that bog and fen complexes are likely one of the primary drivers of hydrograph timing and shape due to their ability to dynamically alter drainage pathways, particularly in this region. In 1997, following a significant melt event, all simulations in Jean-Marie exhibit higher than observed recession limb flows; indicating runoff was slow to drain and storages were too high. This could be an indication of WATFLOOD's inability to capture the dynamic flow paths occurring within Jean Marie's extensive fen network. This same recession limb discrepancy does not occur in Blackstone, where there are much fewer fens, and bogs would remain hydraulically isolated even during wetter conditions (Connon et al., 2015). In the Blackstone, the recession limb hydrograph is well simulated across all inputs, however, peak flows (with the exception of the 1997 snow melt) are generally under estimated. Post freshet, simulations adequately capture the timing of rainfall events; however (with the exception of 1997 in the Jean Marie) consistently underestimate the magnitude of the peaks. This underestimation is most evident when all simulations generated a very limited streamflow response to an early October rainfall event in 1998, underestimating the observed peak flow by approximately 50

11

% (Jean Marie) and 75 % (Blackstone). These results may point to inadequate precipitation forcing due to the climate station proximity and high spatial variability of rainfall, inadequate soil moisture parameterization, or could be an unintended side effect of using NSE in our calibration (Gupta et al., 2009).

5   Most interesting is the similarity of the streamflow simulations among the different $\delta^{18}O_{ppt}$ products, further assessed by Kendall's tau coefficient (τ). In Jean Marie, τ ranges between 0.92 (REMOiso versus static) to 0.97 (KPN43 versus static). In Blackstone τ is more tightly constrained, ranging from 0.96 (REMOiso versus static) to 0.98 (KPN43 versus static). All τ values are statistically significant. It should be noted that small deviations between mean streamflow simulations occur during spring melt, where REMOiso-derived streamflow consistently results in higher peaks than KPN43 and static-driven

10   simulations. These differences in mean streamflow, however, fall within overlapping uncertainty bounds and are not significant outside of parameter uncertainty. Despite significant changes to model parameters (Table S-1), the resultant efficiency statistics among the mean streamflow simulations remain nearly identical (Table 5). Based on this analysis, we find that the three precipitation isotope products generate statistically similar streamflow simulations. Given the insignificant differences in streamflow response, it is only through analysis of $\delta^{18}O_{SF}$ that the impact of different model parameterizations is assessed.

15   **4.3 Modelling $\delta^{18}O$ in streamflow**

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

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

Though each of the model calibrations result in similar trends relative to the observed $\delta^{18}O_{SF}$ record, there are notable departures. As simulated $\delta^{18}O_{SF}$ uncertainly envelopes associated with each $\delta^{18}O_{ppt}$ product are, at times, non-overlapping,

25   differences in $\delta^{18}O_{SF}$ simulations can be attributed to $\delta^{18}O_{ppt}$ product and, therefore, are not just an artefact of parameter uncertainty (unlike streamflow). The dissimilarities between $\delta^{18}O_{SF}$ simulations are also reflected in the RMSE statistic (Table 5); the RMSE is larger for static-derived simulations due to increased emphasis on periods with a higher observation density (i.e., spring freshet), where larger offsets between simulated and observed $\delta^{18}O_{SF}$ exist. The KPN43 and REMOiso calibrations produce comparable RMSE. The KGE statistic shows only minor differences between $\delta^{18}O_{SF}$ simulations given the statistic

30   puts more emphasis on long-term bias (Gupta et al., 2009), therefore reflecting the fit of the overall simulation throughout the study period for this highly seasonal basin (Kling and Gupta, 2009). Further research is required to better understand the impacts of sporadic sampling resolution (for $\delta^{18}O_{SF}$ observations) on efficiency criteria, and consequently the objective functions. It is noted that sampling during peak freshet was, at times, limited by accessibility during high water stage; therefore,

some temporal gaps exist in the observed $\delta^{18}O_{SF}$ record (particularly in 1999) during the period that streamflow compositions are generally most depleted.

Differences in $\delta^{18}O_{SF}$ simulations within each sub-basin are due to a combination of: (1) markedly different $\delta^{18}O_{ppt}$ input compositions during large precipitation events amongst precipitation isotope products, and (2) how new water transits through the system via the model's hydrological compartments. For this study area, large precipitation events can be separated into: (1) the accumulation of winter snowfall and corresponding spring freshet (approximately 35 to 40 % of annual precipitation), and (2) major rainfall events occurring post-freshet (summer and fall) (with rainfall representing approximately 60 to 65 % of annual precipitation).

No single model calibration produces consistently strong simulations of $\delta^{18}O_{SF}$ during the snowmelt period. The KPN43 calibration best captures the timing and magnitude of spring freshet, however overestimates $\delta^{18}O_{SF}$ (i.e., is too enriched) during the 1997 melt in Blackstone. Conversely, the static and REMOiso calibrations capture the large depletion during the 1997 melt in the Blackstone, but produce overly depleted simulations during the 1998 and 1999 freshets - most notably within the Jean Marie. There is a tendency for all models to simulate relatively depleted spring freshet $\delta^{18}O_{SF}$ compositions. This can be attributed to several factors: (1) overly enriched $\delta^{18}O_{ppt}$ during the winter months, (2) unaccounted for snow metamorphism processes, (3) an overestimation of direct snowmelt runoff (i.e. streamflow volume), and (4) inaccurate antecedent composition of $\delta^{18}O_{SF}$ simulated by the models just prior to the spring melt.

Post-freshet, $\delta^{18}O_{SF}$ simulations are impacted by rainfall amount and composition, and the offset between simulated $\delta^{18}O_{SF}$ and $\delta^{18}O_{ppt}$ input at the time of rainfall. As rainfall amount and/or this offset increases, the resulting impact on simulated $\delta^{18}O_{SF}$ increases. This highlights the importance of capturing the spatial and temporal variability in rainfall $\delta^{18}O$, particularly for large and isotopically distinct (from streamflow) events. The threshold defining a large rainfall event varies depending on basin physiography, land cover, storage capacity, and antecedent conditions. St Amour et al. (2005) estimate this threshold to be ≥40 mm within the Fort Simpson region. Such a large, isotopically distinct rainfall event occurred June 11–12, 1998 when approximately 70 mm fell over two days with an observed $\delta^{18}O_{ppt}$ composition of -22.7 ‰. Both the REMOiso and static products reasonably capture this event (-20.9 ‰ and -20.1 ‰, respectively, across the study domain); however, the KPN43 product predicted an average $\delta^{18}O_{ppt}$ composition of -17.6 ‰. In the Jean Marie, where large fen networks help to moderate rainfall-runoff response, the observed $\delta^{18}O_{SF}$ did not deplete in response to this event, but rather maintain a similar pre-event composition around -19.17 ‰ (Fig 3, panel (c)). KPN43-driven simulations most closely match observed $\delta^{18}O_{SF}$ due to the antecedent composition of $\delta^{18}O_{SF}$ prior to the event, even though the KPN43 input generated the least accurate estimate of the depleted rainfall $\delta^{18}O_{ppt}$. Conversely, in the Blackstone the June 11–12 rainfall generated a much different response in observed $\delta^{18}O_{SF}$: a sharp depletion from -19.11 ‰ to -20.98 ‰ (Fig 4, panel (c)). In this instance, the REMOiso and static calibrations most closely match the observed $\delta^{18}O_{SF}$ due to their closer representations of the rainfall event composition. In the Blackstone,

13

this single event results in a significant offset between KPN43-driven $\delta^{18}O_{SF}$ simulations relative to those driven by REMOiso and static products, maintained throughout 1998 and up until the 1999 freshet resets the $\delta^{18}O_{SF}$.

Throughout much of Canada and in other high-latitude climates, the spring freshet generates a substantial portion of annual streamflow (and typically peak annual flow) when the accumulation of solid precipitation from the winter season melts in late spring over a few week period. It is therefore important to understand how differences among the products impact snowpack (and subsequently snowmelt) isotopic compositions in isoWATFLOOD. Figure 5 shows the evolution of precipitation-weighted snowpack oxygen-18 ($\delta^{18}O_{SNW}$) throughout each winter of the study period relative to the observed snowpack compositions (hollow black diamonds). Not surprisingly, the static snowpack compositions closely match with observed $\delta^{18}O_{SNW}$, and we note that KPN43 and REMOiso snowpacks are more enriched. Caution should be used when comparing modelled versus observed data here as there is little inter-annual consistency in the number of samples and the location where sampling took place, and no information was provided as to how the $\delta^{18}O_{SNW}$ were collected (i.e., depth-integrated or depth-dependent samples). Comparison of like-forcing pairs between Jean Marie and Blackstone reveal subtle spatial differences in simulated $\delta^{18}O_{SNW}$. Dissimilarities between the three products within each basin are, however, significant. Interestingly, REMOiso and KPN43 end of winter precipitation-weighted $\delta^{18}O_{SNW}$ compositions differ by less than 0.5 ‰ in 1997–1998 and 1998–1999. REMOiso and KPN43 inputs consistently generate significantly more enriched snowpacks relative to the static $\delta^{18}O_{SNW}$ product (and much of the observed data). On average, KPN43 is 3.3 ‰ more enriched, and REMOiso is 3.1‰ more enriched than end of season static $\delta^{18}O_{SNW}$. Differences in simulated $\delta^{18}O_{SNW}$ among the products are partially attributed to the poor representation of snowpack physics (i.e., fractionation resulting from sublimation and snow metamorphism) in the current version of the isoWATFLOOD model. The static input inadvertently accounts for some of these processes, as the specified compositions are derived from snowpack observation near end of winter (in late March). Uncertainty in simulated $\delta^{18}O_{SNW}$ among the products is notable as well, with static $\delta^{18}O_{SNW}$ uncertainty remaining relatively constant over the winter relative to REMOiso, and particularly KPN43 where uncertainty decreases as snowpack depth increases (Fig. 5). This is an artefact of the parameterization of sublimation in the models. As the winter progresses, the snowpack grows and sublimated volumes become a smaller fraction of the total snowpack, thus decreasing the effect (and uncertainty) that sublimation has on the volume-weighted $\delta^{18}O_{ppt}$ of the snowpack. This is observed during periods when the simulated snowpack and snow water equivalent (SWE) are larger, for example, 1998 relative to 1999 (Fig. 5).

These significant differences in simulated snowpack composition are one of the primary reasons for offsets between KPN43, REMOiso and static $\delta^{18}O_{SF}$ simulations (Fig. 3 and Fig. 4, panel(c)). Once a $\delta^{18}O_{SF}$ simulation has been offset, it is not possible to 'reset' the composition in late fall as streamflow decreases to near-zero and mass retained in the system. This can result in compounding isotopic error (if the offset deviates from observed data) during continuous simulation, thus highlighting the sensitivity of the tracer as a calibration tool. Compounding error is also observed for rainfall events, but generally to a lesser

14

extent due to the relatively smaller durations and magnitudes (volume contributions) of most rainfall events (relative to snowmelt) in high-latitude regions.

Since both $\delta^{18}O_{SF}$ and $\delta^{18}O_{SNW}$ are significantly different among $\delta^{18}O_{ppt}$ products, internal water apportionment (determined by model parameterization) is also likely impacted. Differences in hydrograph separations among the calibrated models are explored to determine the impact $\delta^{18}O_{ppt}$ has on internal water apportionment and simulation uncertainty.

4.4 Hydrograph component analysis and parameter distributions

Component contributions to total streamflow from surface runoff, interflow and baseflow storage in each season (DJF: December-January-February; MAM: March-April-May; JJA: June-July-August; and, SON: September-October-November) derived from each $\delta^{18}O_{ppt}$ product are shown on Figure 6. Jean Marie and Blackstone display similar trends in internal water apportionment throughout the year, indicating generally similar model parameterizations and hydrograph separations among the two basins. Some seasonal differences in component separations exist, however, which are linked to variations in basin physiography, land cover, and storage characteristics reflected by differences in the baseflow (lzf and pwr) and wetland parameters (kcond and theta) among basins (Table S-1). Freshet and post-freshet percent contributions to total streamflow in this study are in agreement with those reported in previous studies. St Amour et al. (2005) reported significant post-freshet groundwater contributions (71 % ± 9 % and 64 % ± 10 % for Jean Marie and Blackstone, respectively), compared to the mean post-freshet (JJASON) contributions we report on Figure 6 (40 – 70 % and 60 – 70 % for Jean Marie and Blackstone, respectively). In agreement with this, 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 stream within the last 2.3 ± 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 in the FSB and specifically Blackstone, relative to lower-gradient watersheds. Groundwater as defined by St. Amour et al (2005) and Jasechko et al (2016) denotes 'old water', or water residing in the system prior to an event. In our study, groundwater is defined as baseflow in isoWATFLOOD (Stadnyk et al. 2005) and is separate from interflow (soil water in the unsaturated zone) and wetlands. Baseflow contributions in this study are therefore slightly lower than those estimated from the two-component hydrograph separation methods. Snowmelt contributions from St. Amour et al (2005) were 21 % (± 2 %) and 40 % (± 4 %) of total streamflow for Jean Marie and Blackstone, respectively; which are in agreement with mean spring (MAM) surface runoff contributions in our study (20 – 40 %) for both basins.

Comparison of seasonal volume contributions derived from each $\delta^{18}O_{ppt}$ product reveal that during spring (MAM), REMOiso-driven simulations show more surface flow contribution to total streamflow, with the mean volume lying above the 95th percentile volumes 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 the Jean Marie; and similar, yet slightly larger, percent contributions for the Blackstone).

15

From the seasonal analysis, no other significant differences in component contributions outside of parameter uncertainty can be attributed to $\delta^{18}O_{ppt}$ product. It is important to note, however, that each $\delta^{18}O_{ppt}$ product results in different amounts of parameter uncertainty, both seasonally and overall, as represented by width of the uncertainty bounds (cross symbols on Fig. 6). The variation in uncertainty bounds between $\delta^{18}O_{ppt}$ products is also visible on Figure 3 through Figure 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 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 behavioural parameters retained for each $\delta^{18}O_{ppt}$ input, which originate due to distinctions in magnitude and variability (both spatial and temporal) among $\delta^{18}O_{ppt}$ products.

Further demonstrated by parameter probability distributions (Fig. 7), the three calibrations result 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. Rather, we introduce this analysis to further explore how model parameterization is impacted by $\delta^{18}O_{ppt}$ input. The selected parameters (Table 4) influence evaporation (f-ratio), surface runoff during snowmelt (akfs, base), upper and lower zone storage (retn), interflow (retn), and baseflow (lzf, pwr). REMOiso parameter distributions more often than not differ from KPN43 and static parameter distributions. Although dissimilarities between KPN43 and static parameter distributions exist, these are typically not as prevalent as differences with REMOiso-derived distributions. This echoes the findings from Figure 7 that KPN and static-derived component contributions are more similar than those derived from REMOiso; which may very well be due to the increased spatial and temporal variability of the REMOiso $\delta^{18}O_{ppt}$ product. Though we cannot verify correctness of the REMOiso $\delta^{18}O_{ppt}$ given the absence of daily precipitation isotope observations, differences among inputs imply that temporal resolution of $\delta^{18}O_{ppt}$ plays a role in the parameterization of a model, and resultant hydrograph separation.

Differences in surface water contributions during snowmelt between REMOiso, KPN43 and static inputs are likely derived from differences in the akfs and base parameters. Parameter distributions from REMOiso are significantly different (as verified through Kolmogorov–Smirnov testing of distributions) than the KPN43 and static input distributions for these parameters (Figure 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). Dissimilarities in baseflow contributions among $\delta^{18}O_{ppt}$ inputs are influenced by 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 calibrations within Blackstone (Fig. 6, panel (f)), and for all models during fall and winter within Jean Marie (Fig. 6, panel (c)), are likely due to the wider range of

16

behavioural values for the pwr parameter, specifically the inclusion of lower values which results in longer, more drawn out recession limbs. It appears that choice of precipitation isotope product influences parameter distributions in isoWATFLOOD, which in turn alters internal water apportionment. In the tracer-aided modelling community, this has significant implications for hydrograph separation and any associated transit time analyses; both of which will be influenced by choice (resolution) of

5    $\delta^{18}O_{ppt}$ product.

5 Conclusions

This study used three types of precipitation isotope products as $\delta^{18}O_{ppt}$ input to a tracer-aided hydrological model (isoWATFLOOD) to investigate the impact differing spatial and temporal resolutions have on simulation of streamflow, isotopic composition of streamflow, internal hydrograph separations, and model parameterization and corresponding

10   parameter uncertainty. Our study found that choice of precipitation isotope product ($\delta^{18}O_{ppt}$):

1. did not impact simulation of total streamflow, or the achieved efficiencies of streamflow simulation;
2. impacted the internal apportionment of water, thereby impacting hydrograph separations;
3. impacted model parameterization, and therefore simulation uncertainty; and
4. impacted the variability of simulated $\delta^{18}O_{SF}$, most noticeably when event compositions differed significantly from

15   streamflow composition (e.g., snowmelt and large rainfall events).

Of the 30,000 simulations performed for each precipitation isotope product forcing, only 10 % or less were behavioural for each input type. Due to the wide range of behavioural parameter values (Table S1), however, we are confident that the approach used was sufficient to characterize parameter uncertainty. Not unexpectedly, this finding also indicates that 30,000 model

20   evaluations were not sufficient to quantify parameter identifiability in this study.

Although total simulated streamflow was not significantly affected by choice of $\delta^{18}O_{ppt}$ input, $\delta^{18}O_{SF}$ simulations and the internal apportionment of water (surface flow, interflow, and baseflow) were significantly impacted here. Significant differences in internal water apportionment among the models were diagnosed via $\delta^{18}O$ uncertainty. Variation between models

25   was greatest during the freshet period, where significantly different snowpack compositions were simulated among the different precipitation isotope products. The highest resolution (REMOiso, daily) input resulted in significantly different parameter distributions and seasonal hydrograph separations than the other two (monthly and semi-annual) inputs. These findings have direct implications for hydrograph separation, and simulated water transit times. In this study, we found that choice of $\delta^{18}O_{ppt}$ input directly impacted model parameterization, and for this reason, studies should account for both input

30   and parameter uncertainty. Also highlighted was the significance of the snowpack and melt dynamics in tracer-aided models applied in high-latitude regions, resulting in high seasonal uncertainty and indicating more research is warranted to improve

17

process representation. Use of a tracer-aided model afforded an examination of internal model dynamics resulting from specific parameterizations, allowing us to assess the realism of individual simulations as opposed to their efficacy alone.

This study demonstrated that direct quantification of model equifinality was possible using tracer-aided models, and furthermore, we demonstrated that this equifinality was not diagnosable via simulation of streamflow. We have achieved an understanding of how tracer-aided models, like isoWATFLOOD, can be used in data sparse regions, with limited input data (including $\delta^{18}O_{ppt}$ observations), and that despite these limitations, these models can still be of value. Regarding the usefulness of precipitation isotope products in regions with limited observations, both the static and REMOiso inputs require existing $\delta^{18}O_{ppt}$ observations (i.e., from CNIP) to either define or bias correct the input, limiting their use for certain applications. If these data are not available, the KPN43 input provided reasonable results without the need for additional observations. The existence of CNIP (and other precipitation isotope networks) was crucial to the development, validation, and bias correction of existing $\delta^{18}O_{ppt}$ products. Attaining an understanding of how $\delta^{18}O_{ppt}$ input uncertainty impacts simulated model output is important when calibrated models are used to assess climate-driven or land-use-driven impacts on streamflow in remote, data sparse, high-latitude regions.

For use in tracer-aided modelling, precipitation isotope products should capture both the event-based variability and seasonality of precipitation isotopes to reproduce realistic $\delta^{18}O_{SF}$ variability. Higher resolution $\delta^{18}O_{ppt}$ inputs (e.g., REMOiso, daily) were able to capture event-specific compositions that, when significantly different from $\delta^{18}O_{SF}$, tended to cause significant deviations from the $\delta^{18}O_{SF}$ derived from monthly and semi-annual (i.e., static) inputs. Unfortunately, we could not verify the correctness of the higher resolution product (i.e., REMOiso) in this study due to the sporadic sampling of isotopes in precipitation observations. Static and seasonal precipitation isotope products missed event-specific isotopic variation occurring as a result of heavy rainfall events, which require increased temporal resolution (e.g., daily $\delta^{18}O_{ppt}$ inputs from REMOiso; but perhaps weekly input would suffice) to resolve rainfall event compositions. In seasonal environments, precipitation isotope products must capture the transition from rainfall to snowfall, and from snow accumulation to snowmelt to sufficiently model $\delta^{18}O_{SF}$. In this study, both static and monthly inputs adequately captured $\delta^{18}O_{SF}$ variability at the basin outlet, perhaps the result of the unique seasonality in high-latitude regions. Spatial variability was detected across the study region in $\delta^{18}O_{ppt}$ inputs, and can be represented by distributed tracer-aided models, like isoWATFLOOD. There is reason to suspect that the variability in (both spatial and temporal) precipitation isotope inputs influences model parameterization, therefore spatial variability should be preserved to derive the most representative model of a given region.

This work highlighted the power of tracer-aided modelling to inform and quantify equifinality in hydrological simulation, helping modellers to work towards reducing modelling uncertainty. Although more work is required to assess and understand

parameter identifiability, our analysis showed that selection of precipitation isotope ($\delta^{18}O_{ppt}$) product had direct implications on model parameterization, and that input uncertainty should be considered in future studies.

6 Future directions

Distributed hydrological models, such as WATFLOOD, are complex with large numbers of parameters, therefore it is important as a community to work toward conducting comprehensive studies that focus on input data uncertainty and parameter identifiability. In the tracer-aided modelling community, this includes uncertainty from precipitation isotope products and their varying spatial and temporal resolutions. Ideally, further studies should be conducted in well-instrumented basins where $\delta^{18}O_{ppt}$ input can be better characterized using observed data at higher spatial, and most importantly, temporal resolutions. Several key questions warranting more detailed investigation include: (1) are precipitation isotope products adequate alternatives in place of $\delta^{18}O_{ppt}$ observations; (2) are there a specific subset of model parameters that are more sensitive to choice of precipitation isotope product; and (3) how do (if at all) parameters compensate for compounding model error. Unfortunately, at least within Canada, a well instrumented watershed at the regional scale does not yet exist, pointing to the importance of implementing additional (or enhancing current) iso-hydro-meteorological monitoring networks.

Not unexpectedly, the RCM-driven precipitation isotope product in this study, REMOiso, exhibited some bias and needed correction prior to application. More studies are needed 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 at a monthly resolution. Due to the lack of high-resolution $\delta^{18}O_{ppt}$ observations in Canada, however, daily or weekly validation is not yet possible. Additionally, the suitability and performance of other isotope-enabled RCM's for use in Canada, and elsewhere, should be explored.

Lastly, as a tracer-aided hydrologic community we need to push for the sustained monitoring of isotopes in precipitation and streamflow that are required to inform our models and improve uncertainty assessment. This study elucidated the impact that discontinuous observations can have on quantifying model uncertainty; which would only be further exasperated by the absence of observations all together. In Canada, a concerted effort is needed by the Government to protect and sustain our observation networks, which are required for improved prediction in remote regions for climate and hydrologic change detection.

Author contribution

C. Delavau developed model code to generate Kpn43 $\delta^{18}O_{ppt}$ 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.

19

Stadnyk. C. Delavau prepared the manuscript with contributions from all co-authors. T. Stadnyk edited the manuscript based on reviewer comments with help from C. Delavau. C. Delavau and T. Holmes completed amendments to figures.

Competing interests

The authors declare that they have no conflict of interest.

References

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

20 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.

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,

25 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

30 models using multi-criteria calibration and tracers in temperate, upland catchments, Hydrol. Process., 24, 260–275, doi: 10.1002/hyp.7478, 2010b.

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.

5   Connon, R.F., W.L. Quinton, J.R. Craig, J. Hanisch, and O. Sonnentag: The hydrology of interconnected bog complexes in discontinuous permafrost terrain. Hydrol. Process., 29: 3831-3847. 2015.

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
10   revealed in time series modelling across Canada and northern United States, Water Resour. Res., 51, doi:10.1002/2014WR015687, 2105.

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.:
15   Conceptualization in catchment modelling: simply learning?, Hydrol. Process., 22: 2389–2393, doi: 10.1002/hyp.7070, 2008.

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.

20   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.

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–
25   2548, doi: 10.1002/(SICI)1099-1085(199911)13:16<2537::AID-HYP940>3.0.CO;2-A, 1999.

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.

Her, Y., and Chaubey, I.: Impact of the numbers of observations and calibration parameters on equifinality, model
30   performance, and output and parameter uncertainty, Hydrol. Process., 29, 4220–4237, doi: 10.1002/hyp.10487, 2015.

Holmes, T.L.: Assessing the Value of Stable Water Isotopes in Hydrologic Modeling: A Dual-Isotope Approach. MSc. Thesis, Department of Civil Engineering, University of Manitoba, 2016.

IAEA/WMO: Global Network of Isotopes in Precipitation. The GNIP Database. Accessible at: http://www.iaea.org/water, 2014.

35   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, 2016.

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.

Klaus, J., and McDonnell, J. J.: Hydrograph separation using stable isotopes: Review and evaluation. *Journal of Hydrology,*
40   Volume 505, pp. 47-64, 2013.

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.

Kling, H.V., and H. Gupta: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J. Hydrol., 377: 80-91, 2009.

45   Kouwen, N.: WATFLOOD/WATROUTE Hydrological Model Routing and Flood Forecasting System. User's Manual, University of Waterloo, Waterloo, ON. 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.

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.

5 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.

10 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.

Mesinger, F., et al.: North American regional reanalysis, Bull. Am. Meteorol. Soc, 87, 343–360, doi: 10.1175/BAMS-87-3-343, 2006.

15 Michaud, J., and Sorooshian, S.: Comparison of simple versus complex distributed runoff models on a midsized 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.

20 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:16/0022-1694(70)90255-6, 1970.

Penna, D., Ahmad, M., Birks, S. J., Bouchaou, L., Brenčič, M., Butt, S., Holko, L., Jeelani, G., Martínez, D. E., Melikadze,
25 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.

Philip, J. R.: An infiltration equation with physical significance, J. Soil Sci., 77(2), 153-157, 1954.

30 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.

Singh, V.P. and Frevert, D. K..: Watershed models, CRC Taylor & Francis, Boca Raton, 2006.

35 Smith A., Delavau C., and Stadnyk., T.: Identification of geographical influences and flow regime characteristics using regional water isotope surveys in the lower Nelson River, Canada. Canadian Water Resource Journal, 40 (1): 23-35, 2015

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-Falcone, T. A.: Mesoscale Hydrological Model Validation and Verification using Stable Water Isotopes: The
40 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.

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

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.

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.

5   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.

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.

Taylor, S., Feng, X., Kirchner, J. W., Osterhuber, R., Klaue, B., and Renshaw, C. E.: Isotopic evolution of a seasonal snowpack and its melt, Water Resources Research, 37, 759-769, doi:10.1029/2000WR900341, 2001.

10   Taylor, S., Feng, X., Williams, M., and McNamara, J.: How isotopic fractionation of snowmelt affects hydrograph separation, Hydrological Processes, 16, 3683-3690, doi:10.1002/hyp.1232, 2002.

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.

Tetzlaff, D., Uhlenbrook, S., Eppert, S., and Soulsby, C.: Does the incorporation of process conceptualization and tracer data

15   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.

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.

Zhou, S., Nakawo, M., Hashimoto, S., and Sakai, A.: The effect of refreezing on the isotopic composition of melting

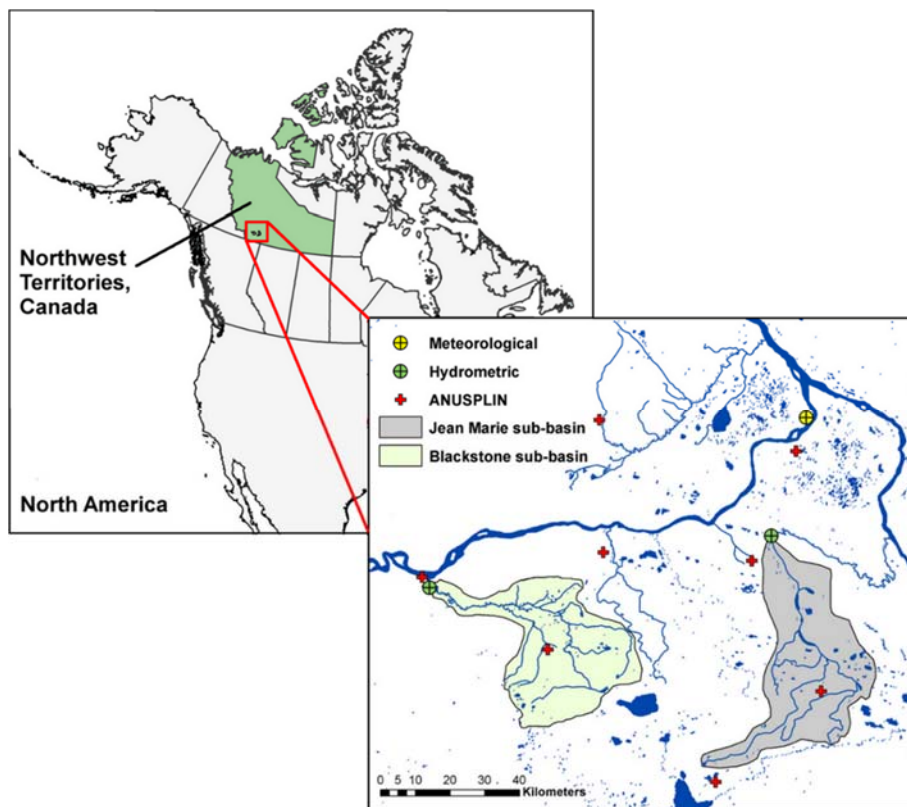20   snowpack, Hydrological Processes, 22, 873-882, doi:10.1002/hyp.6662, 2008

Figure 1: Fort Simpson River Basin (all other tributaries of the Liard and Mackenzie Rivers have been removed for ease of viewing).
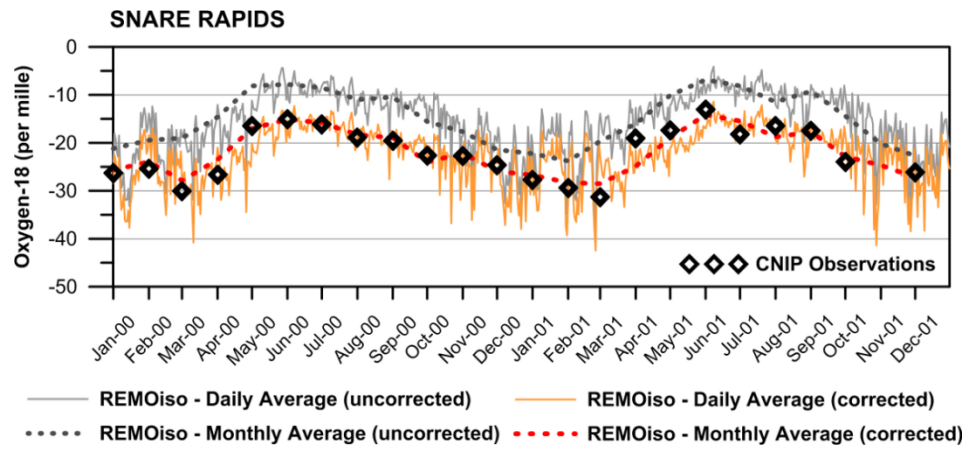
23

Figure 2: Comparison of raw and corrected REMOiso $\delta^{18}O_{ppt}$ output with CNIP monthly compositions at Snare Rapids, NWT.
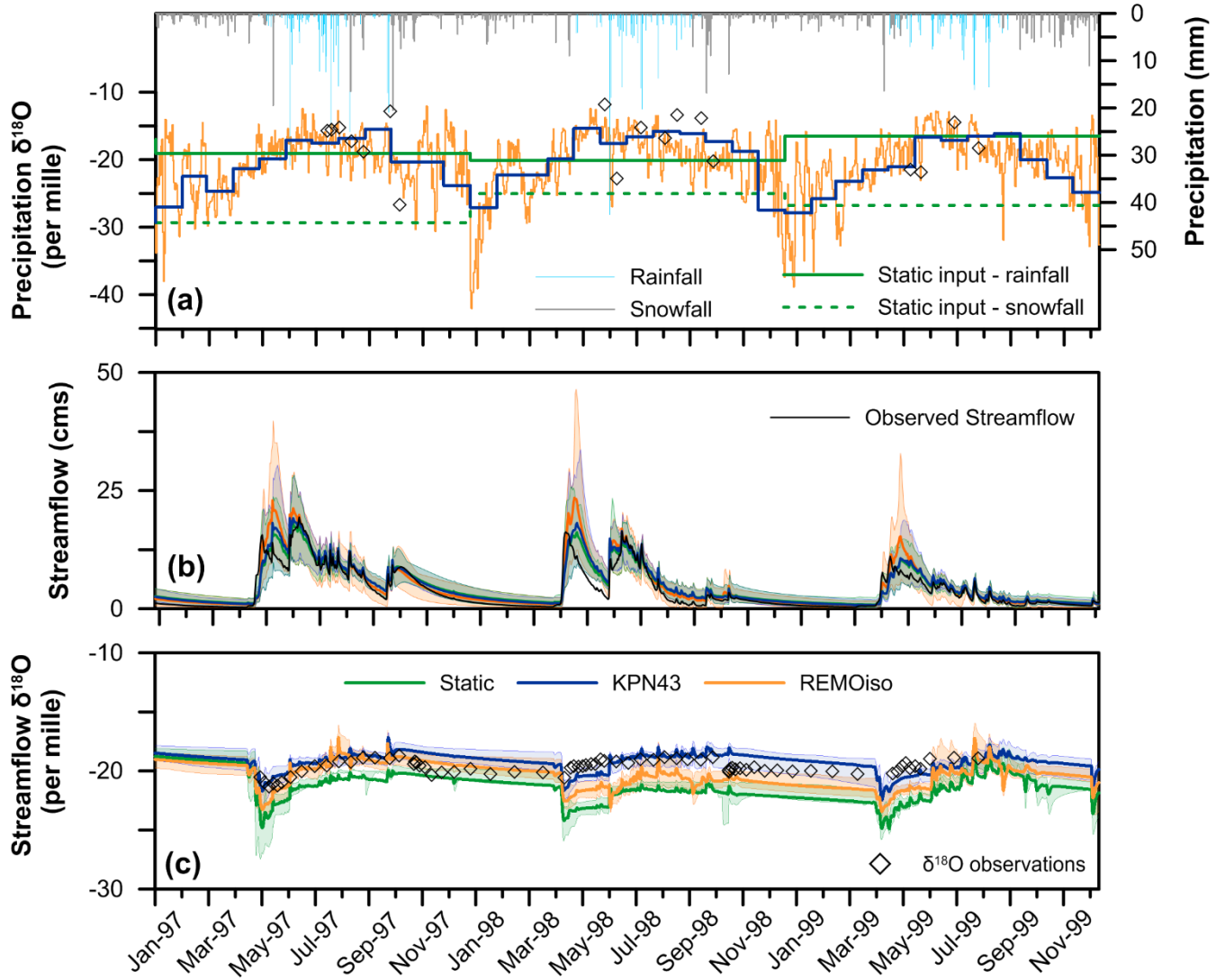
5

24

Figure 3: Input and behavioural simulations for Jean Marie, including: (a) KPN43, REMOiso and static $\delta^{18}O_{ppt}$ input time series and daily precipitation; and simulated (b) mean daily streamflow and uncertainty bounds and (c) mean daily $\delta^{18}O_{SF}$ and uncertainty bounds, for KPN43, REMOiso and static driven model calibrations. $\delta^{18}O_{ppt}$ input-specific uncertainty bounds are represented as the shaded regions, with shading colour corresponding to $\delta^{18}O_{ppt}$ type.
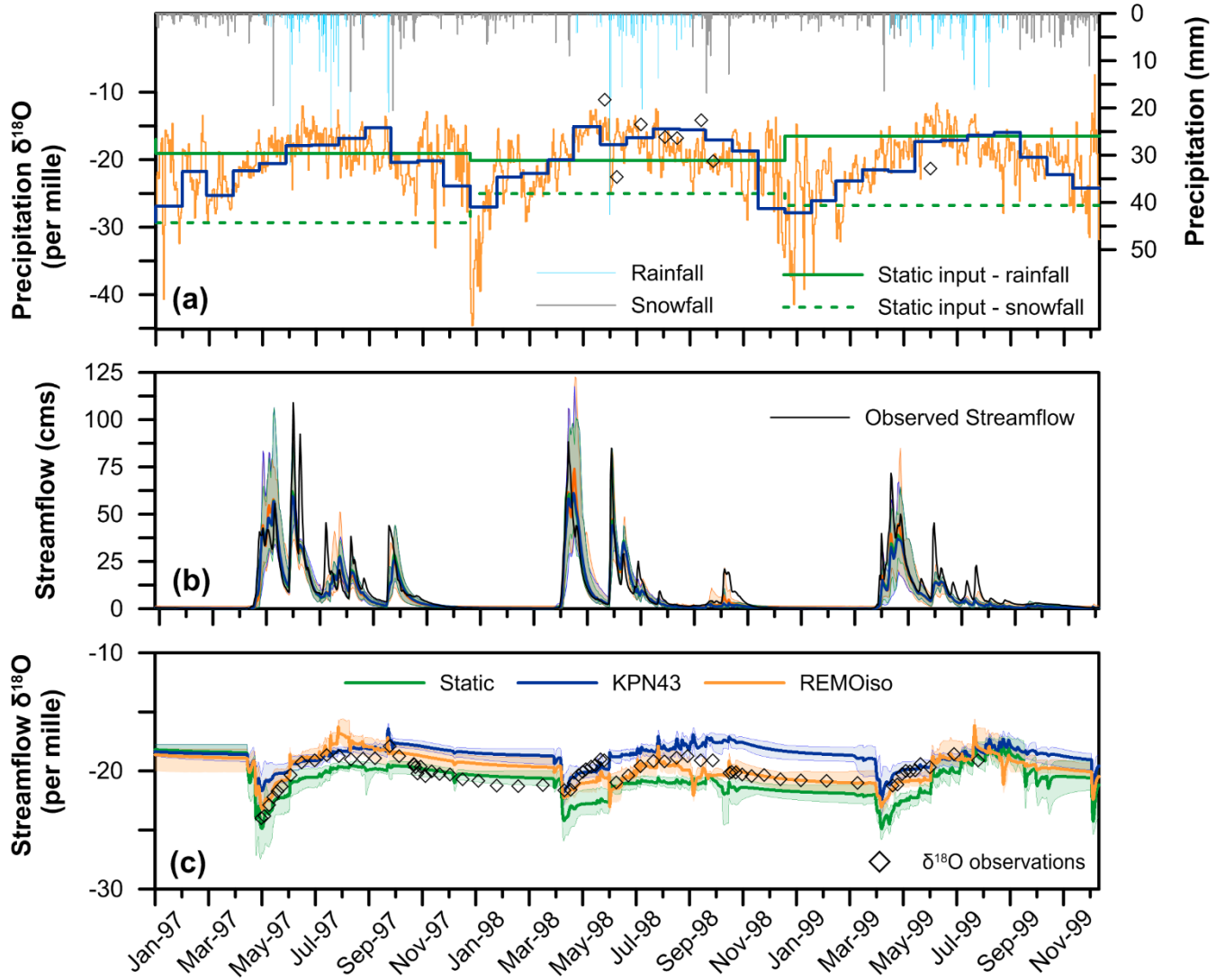
Figure 4: Input and behavioural simulations for Blackstone, including: **(a)** KPN43, REMOiso and static $\delta^{18}O_{ppt}$ input time series and daily precipitation; and simulated **(b)** mean daily streamflow and uncertainty bounds and **(c)** mean daily $\delta^{18}O_{SF}$ and uncertainty bounds, for KPN43, REMOiso and static driven model calibrations. $\delta^{18}O_{ppt}$ input-specific uncertainty bounds are represented as the shaded regions, with shading colour corresponding to $\delta^{18}O_{ppt}$ type.
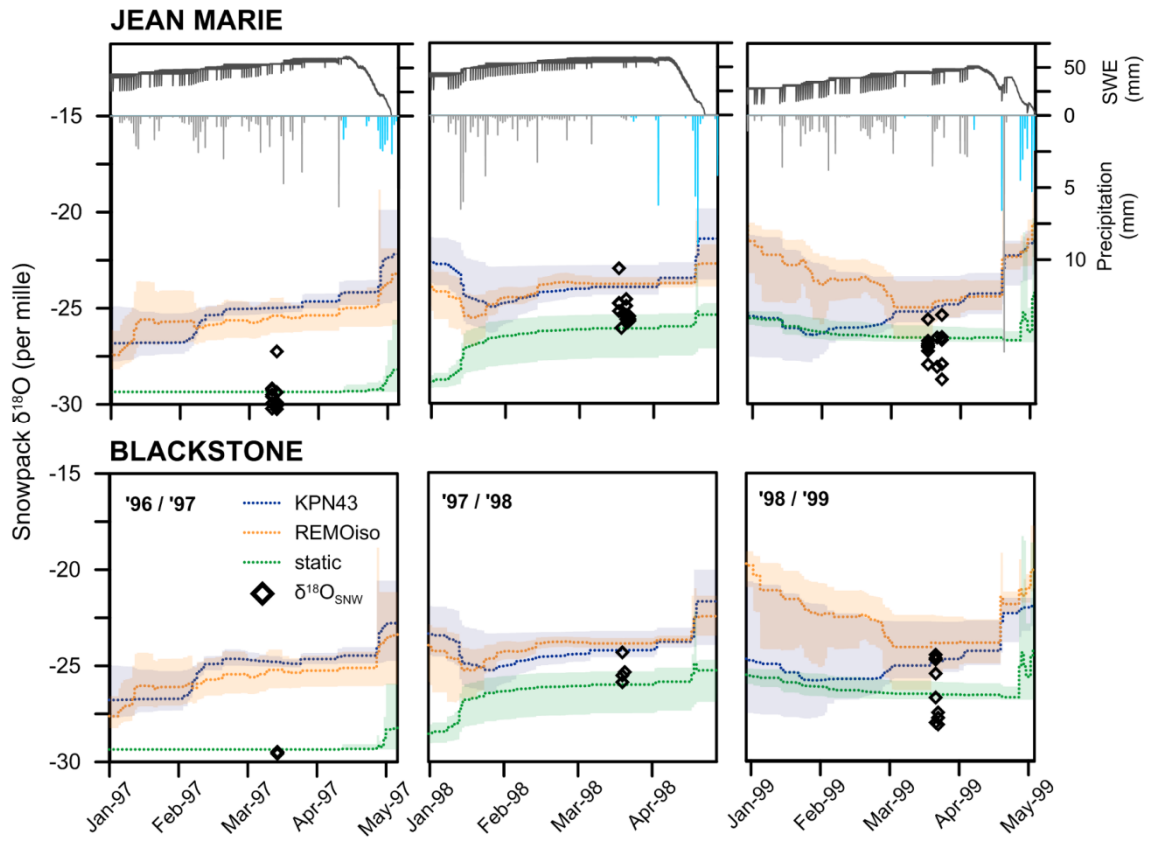
Figure 5: Precipitation-weighted $\delta^{18}O$ of snowpack ($\delta^{18}O_{SNW}$) for KPN43, REMOiso and static inputs from January to the end of melt for each year of the study period. Snow water equivalent (SWE), snowfall (gray line), and rainfall (blue line) are also shown. $\delta^{18}O_{ppt}$ input-specific uncertainty bounds are represented as the shaded regions. Diamond symbols represent $\delta^{18}O_{SNW}$ observations sampled within each respective sub-basin during the GEWEX campaign.
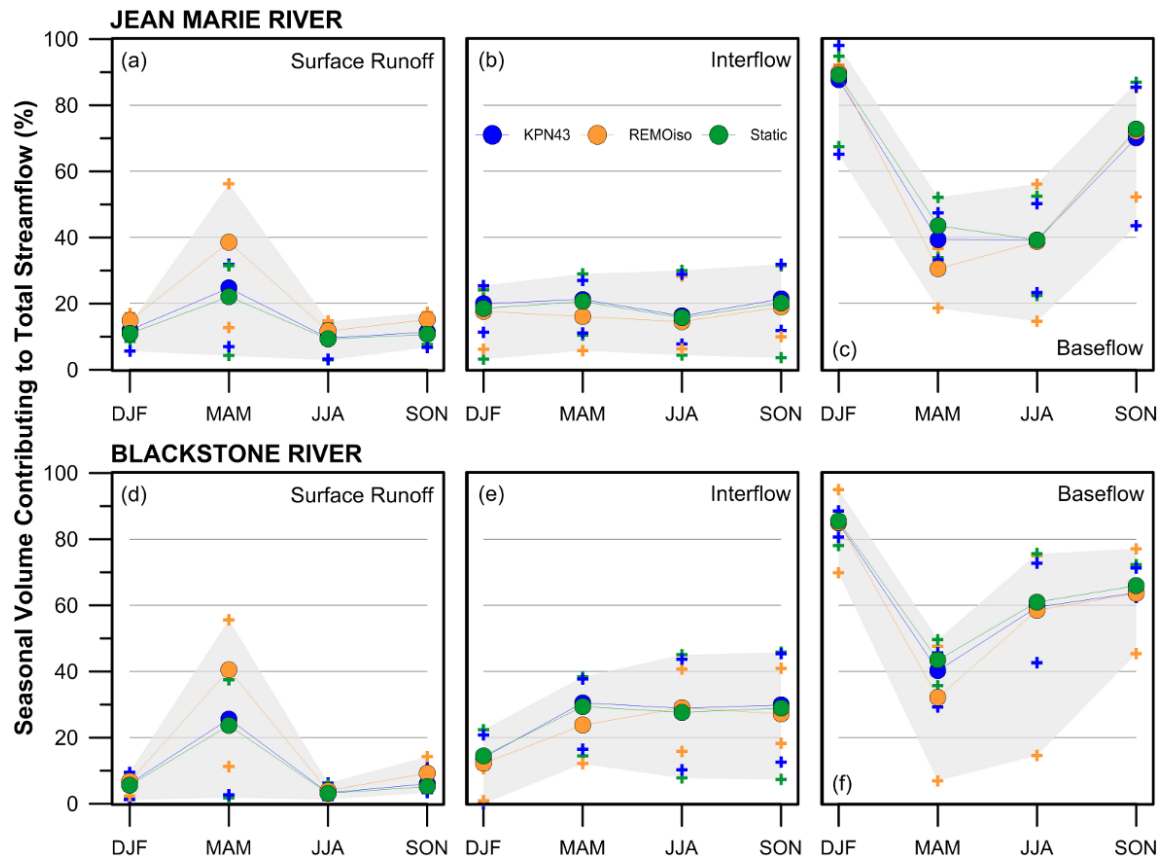
Figure 6: Percent seasonal volume contributing to total streamflow from surface runoff, interflow and baseflow storages for each season. Cross symbols represent the 5th and 95th percentiles for each forcing method, and circle symbols signify the mean values. The combined uncertainty bounds representing the 5th and 95th simulations from all three δ18Oppt input types are shaded in gray.

5
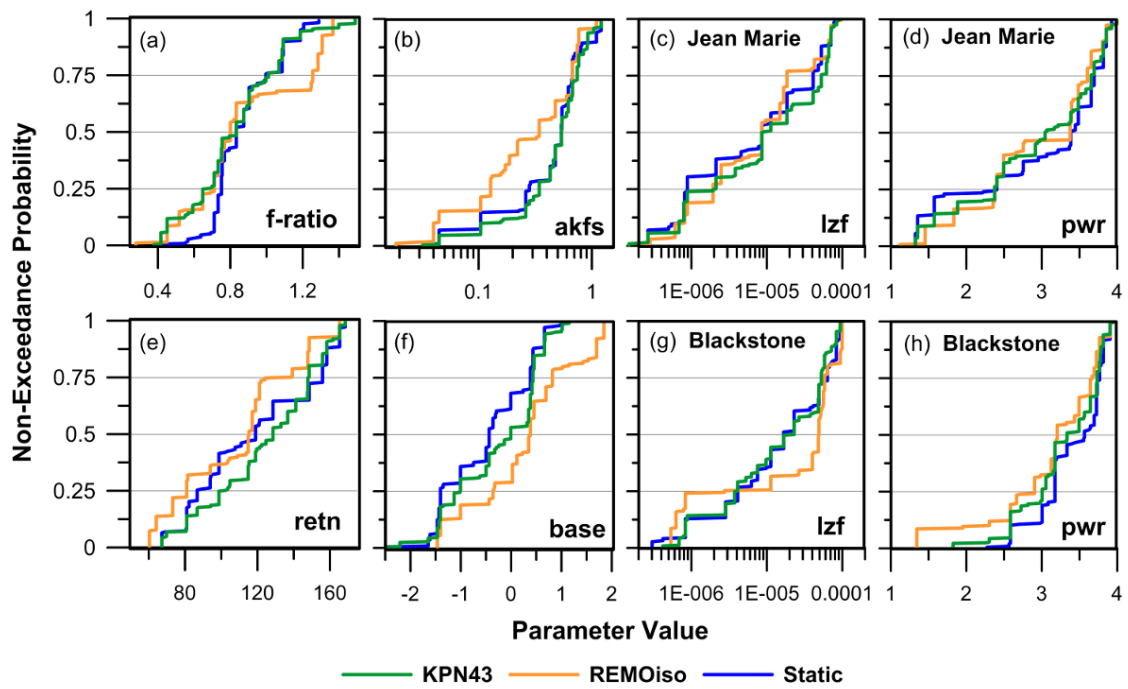
Figure 7: Probability distributions for select parameters (Table 5), as indicated in the bottom right corner of each panel. Parameters are from behavioural simulations, and (a), (b), (e) and (f) have been weighted to the land cover distribution within Jean Marie and Blackstone, as outlined in Table 1. 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^2)	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: 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)
Hydrometric/Meteorological Data					
Daily Average Streamflow Jean Marie (10FB005)	m^3/s	1095	4.66, 5.25	1.24, 2.06	0.19, 34.9
Daily Average Streamflow Blackstone (10ED007)	m^3/s	1095	8.96, 10.76	1.65, 2.17	0.04, 109
Mean Daily Air Temperature Fort Simpson (2202101)	°C	1093	-1.5, -3.02	N/A	-40.8, 25.3
Daily Precipitation Fort Simpson (2202101)	mm	1088	1.12, 1.01	3.04, 3.19	0.0, 43.0
Hourly Relative Humidity* Fort Simpson (2202101)	%	26280	73.9	0.24	14, 100
Isotopic Measurements*					
Streamflow δ^{18}O - Jean Marie	‰	71	-19.70	0.03	-21.34, -18.72
Streamflow δ^{18}O - Blackstone	‰	69	-20.17	0.06	-24.01, -17.92
Rainfall δ^{18}O Jean Marie and Blackstone	‰	27	-17.55	0.23	-26.70, -11.12
Precipitation δ^{18}O Forcing*					
KPN43 $\delta^{18}O_{ppt}$ input	‰	1800 (36 values at 50 grid points)	-20.48	0.19	-28.86, -13.91
REMOiso $\delta^{18}O_{ppt}$ input	‰	54750 (1095 values at 50 grid points)	-21.78	0.25	-42.82, -10.68
Static $\delta^{18}O_{ppt}$ input	‰	300 (6 values at 50 grid points)	-22.82	0.20	-29.35, -16.52

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

Table 3: Static $\delta^{18}O_{ppt}$ input compositions of annual rainfall and snowfall oxygen-18 for isoWATFLOOD.

Year	$\delta^{18}O$ rainfall (‰)	$\delta^{18}O$ snowfall (‰)
1996	-17.00	-29.35
1997	-19.10	-29.35
1998	-20.10	-25.03
1999	-16.52	-26.79

Table 4: 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
Routing Parameters		
flz	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:
pwr	Lower zone drainage function exponent	$QLZ = LZF * (LZS)^{PWR}$ Where: LZS is lower zone storage QLZ is the baseflow flux
theta	Wetland porosity	Physically-based wetland routing algorithm
kcond	Conductivity parameter	(McKillop et al., 1999)
Hydrologic Parameters		
f-ratio	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.
ak	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)
akfs	Surface permeability	
rec	Interflow coefficient	Interflow is represented by a simple storage-discharge relation: $DUZ = REC * (UZS-RETN)*Si$ Where: UZS = upper zone storage
retn	Upper zone retention [mm]	DUZ = depth of upper zone storage released as interflow Si = internal land surface slope
ak2	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
mf	Melt factor [mm/°C/hr]	$M = MF (T_a - base)$
base	Base Temperature [°C]	Anderson (1976)
sub	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 5: Average simulation statistics from n behavioural simulations for streamflow and $\delta^{18}O_{SF}$ for the three model calibrations (using KPN43, REMOiso, and static inputs).

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