Department of Civil Engineering University of Manitoba Winnipeg, MB, Canada Email: tricia.stadnyk@umanitoba.ca Tel: (204) 474-8704 Fax: (204) 474-7516

20 February 2017

Dr. Christine Stumpp Editor Hydrology and Earth System Sciences

### Re: HESS-2016-539

Dear Dr. Stumpp:

Enclosed please find a fully revised, original manuscript now titled "*Examining the impacts of precipitation isotope products (\delta^{18}O) on distributed tracer-aided hydrological modelling*", which is renamed from the previous title "Examining the impacts of estimated precipitation isotope ( $\delta^{18}O$ ) inputs on distributed tracer-aided hydrological modelling" (reference #HESS-2016-539) by Carly J. Delavau, Tricia A. Stadnyk, and Tegan Holmes. We are respectfully submitting our revised manuscript for your consideration in *Hydrology and Earth System Sciences*.

This manuscript evaluates the impact that different spatial and temporal resolutions of precipitation isotope products ( $\delta^{18}O_{ppt}$ ) have on simulated tracer-aided model output, parameters, and uncertainty. We present three different model calibrations, each derived from a different precipitation isotope product, and statistically assess the behavioural simulations, including: the number of parameter sets retained, differences in parameter distributions, and resulting hydrograph separations and associated parameter uncertainty envelopes. Choice of precipitation isotope product influenced parameter distributions, uncertainty envelopes and resulting hydrograph simulations; but had limited impact on resultant total streamflow simulations. This highlights that tracer-aided models are essential in the diagnosis of equifinality, and in quantifying changes to model output and uncertainty resulting from model input. The higher resolution  $\delta^{18}O_{ppt}$  products were able to reproduce the observed streamflow isotopic variability most reliably, and the highest resolution product (REMOiso) had distinct hydrograph separations relative to the KPN43 and static products. Though this study could not confirm the accuracy of the any one product over another (due to a lack of daily  $\delta^{18}O_{ppt}$  observations), it demonstrated that resolution of tracer-aided model inputs directly impacts model parameterization and resulting hydrograph separations.

We have fully revised the paper to take into consideration the constructive comments from the two referees. Given the manuscript required major revision, we have not provided a line-by-line

list of the changes since line numbers have been altered significantly. Instead, we summarize here the major revisions we have made:

- Rewriting of the abstract
- Broader focus on applications to tracer-aided modelling, rather than study-site specific findings
- Additional methodology section on "statistical treatment of data"
- Rewriting of the discussion to focus more specifically on pertinent results and highlight discrepancies within the modelling (based on reviewer feedback)
- Rewritting of the conclusions that highlight take-home messages from the manuscript and that better connect to our broader objectives.

Of note, in response to reviewer feedback and suggestions we have added a supplement (Table S-1 and Figure S-1), a new methodology section (3.4 Statistical treatment of data), renumbered the figures and tables so they appear consecutively, and enhanced the discussion and conclusions sections of the paper. In response to comments from Anonymous Referee #1, we have inserted some detailed text around our assumption that snowpack and snowfall compositions are equivalent, added the snowpack compositions to Fig 5, and expanded our discussion of the results - particularly as they pertain to streamflow and isotopic simulation errors. In response to Dr. Birkel (Referee #2), we have expended the scope and focus away from the study and onto tracer-aided modelling in general and have included spatial maps of the isotope in precipitation input (Figure S-1). We have not run a configuration of the model over 100K iterations, however, because of time constraints and because we did not feel that parameter identifiability was the overall goal of this study. We will however take this advice and apply it to future studies - which are in fact currently underway. We agree that parameter identifiability is important, however, in this paper, we were more interested in input uncertainty and the impact on the range of parameter uncertainty, which we feel we have addressed. A response document provides the full details of the revisions incorporated in the manuscript.

This manuscript has not been previously published in any language nor is it under consideration for publication by another journal. All authors have carefully read the revised manuscript and have agreed to its submission to Hydrology and Earth System Sciences. River discharge and precipitation time series used in this research were from publically available open sources, and all model results and innovations were developed by the authors using the Fortran programming language and Matlab. Figures were generated using a Grapher package. We are willing to share our  $\delta^{18}O_{ppt}$ models and code with interested researchers upon request (Carly.Delavau@gov.mb.ca).

Please note also that the results presented in this paper originate from the lead author's PhD research under the supervision of the second author. The PhD thesis has been published in the University of Manitoba online repository, and is publically available (<u>http://hdl.handle.net/1993/31946</u>). The results from this paper have not been presented at, nor submitted to, any academic conference; and are not currently nor have not been previously submitted for publication in another journal.

Thank you for your consideration of this contribution to HESS, and to the reviewers for their feedback and edits. We look forward to hearing from you.

Sincerely yours,

Dr. Tricia Stadnyk, P.Eng.

Corresponding Author

# **RESPONSE TO THE REFEREES' COMMENTS**

We sincerely thank both referees for their thorough reviews and most constructive comments on our manuscript (Reference # HESS-2016-539). We fully recognize and appreciate the reviewers' efforts in providing these informative reports on our research and their insights have led to an improved interpretation of our results. We have therefore taken into full consideration all of these comments and have prepared responses to these as well as information on how the paper was revised following the referees' suggestions. Our responses to reviewers are provided below **in bold** following the individual comments requiring action from both reviewers, followed by a marked up version of the manuscript (changes highlighted in yellow).

Referee #1:

Specific comments:

The authors should rethink the use of the word "estimated" in the title as well as throughout the whole manuscript. It suggests that the input data was generated specifically for the presented study. It should be clear that (2 of 3) available precipitation isotope product were used to the study. Which is actually an asset for the study and with respect to future studies in other basins.

We agree completely, and this was also suggested by the second reviewer too. We have changed the title to "*Examining the impacts of precipitation isotope products* ( $\delta^{18}O$ ) on distributed tracer-aided hydrological modelling" and revised the use of the word 'estimated' (with respect to  $\delta^{18}O_{ppt}$  inputs) throughout the manuscript to "precipitation isotope products", as appropriate.

The first sentence of the abstract is "...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, there is no evidence for that statement. Please include additional information to the introduction section or revise the first sentence of the abstract.

We have subsequently revised the abstract significantly, and agree that it has yet to be proven that the parameter space is constrained by such tools. We have rephrased this sentence as: "*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)*".

The authors highlight the importance of snowmelt in the study region. The stable isotope signature of the snow pack and its melt water is a very challenging topic. Please handle this point very carefully in your publication. On page 5, Line 17 for example you mention that the default method for oxygen-18 input is annual average rainfall and snowfall. In your static approach, however, you used average measurements of rainfall and snowpack from the GEWEX campaign. Please provide the values of snow pack stable isotope signature in figure 5 by the way. Especially during the ablation season the isotopic evolution of the snowpack progresses due to percolating rain water and fractionation caused by processes like melting and sublimation (Zhou et al., 2008; Unnikrishna et al., 2002; Dietermann and Weiler, 2013; Lee et al., 2010). This leads to an increase of heavy isotopes in melt water throughout the freshet period (Taylor et al., 2001, 2002; Unnikrishna et al., 2002). Which is correctly represented by the shown model results. Taylor et al. (2001 and 2002) point out that for hydrological applications (in their case isotope based hydrograph separation) a correct representation of the snow pack melt water is absolutely crucial.

Thank you for your insight, and we couldn't agree more that the isotopic signature of a snowpack and its evolution in snow melt are very challenging processes. We have since revised Figure 5 and added the snowpack data and also included a cautionary note to readers highlighting there is uncertainty surrounding these measurements. For the modelling, as a static input our model would preferably use average annual inputs of rainfall and snowfall. Rainfall and snowpack values were obtained from the GEWEX campaign. There was no data on snowfall composition available – only snowpack compositions - therefore we assume (as model input) that the average annual composition of snowfall is approximately equal to that of the snowpack. We have clarified our assumption in the manuscript.

REMOiso is a distributed dataset and the precipitation amounts are also available spatially distributed over the study area. Why was the precipitation amount weighting only conducted at one location and not spatially distributed?

The only precipitation amount-weighting for REMOiso was done to determine the bias correction at Snare Rapids. There was no need to do this spatially for this purpose as we are comparing CNIP observations (at a point) directly to REMOiso output at a single location corresponding to the location of the CNIP observation station. We averaged the four 6-hourly REMOiso values (at each grid) to arrive at daily compositions that were read into the model as input on a per-grid basis (i.e., no amount-weighting involved – same as for the static and KPN inputs). Based on some (unpublished) analyses we did for a study of

the Mackenzie River Basin (i.e., using the same REMOiso model output), we don't trust the quality of sub-daily REMO precipitation to the point where we would use (sub-daily) precipitation to amount weight REMOiso  $\delta^{18}O_{ppt}$ . If we decided to amount weight, we couldn't use actual observations to amount weight 6-hourly to daily as we only have daily precipitation from Fort Simpson Airport and at various grid locations from the ANUSPLIN product.

The authors mention that "several changes and improvements" (Page 7, Line 16) were carried out in the model version used for the study. In the following only one modification (proportion of bog an fen split) is mentioned. Are there any other modifications? If so, please mention them here.

This was poorly worded on our part. What we meant to say was that the model (isoWATFLOOD) has undergone "several changes and improvements" since it was last published in a study back in 2013 (Stadnyk et al., 2013). These changes and improvements were independent of the current study, and all toward continual improvement of internal dynamics and the model output. We have revised the wording in our manuscript to clarify: "*The isoWATFLOOD model used in this study is based on a previous version used by Stadnyk et al. (2013). The current model, however, uses an updated version of isoWATFLOOD code and the watershed set-up incorporates various model improvements made since 2013, independent of this study.*"

The first two paragraphs of section 4 (Results and discussion) should definitively be revised. There is a lot of content that can be mentioned later in the conclusions section (the last sentence on Line 12-14 for example).

We have significantly revised the results & discussion using the guidance of your questions below to help highlight specific findings related to our key objectives and take-home messages. We have also moved the sentence you reference above to the conclusions.

In section 4.2 (Modelling streamflow) please explain the model results as well as the observed streamflow in much more detail. The three different inputs (and three different calibrations) provide very similar results for the simulated streamflow (Page 15, Lines 5-8). Those results should be discussed in more detail.

Thank you for your suggestions, we have revised the discussion to include more specific, indepth discussion of the simulated streamflow resulting from the three types of precipitation isotope product. And yes, all three precipitation isotope products (three different calibrations) result in almost exactly the same streamflow simulation (i.e., statistically the same according to the Kendall's tau test applied in the paper).

Is there really no discharge in winter (Figure 2 and 3)?

We assume you are referring to Figures 3 & 4 (not 2). And no, observed streamflow does not go zero, but rather becomes very small relative to peak flows: minimum in Jean-Marie from 1997-1999 of 0.194 m<sup>3</sup>/s, or 0.5% of the maximum streamflow, 35 m<sup>3</sup>/s during this same period; and a minimum of 0.043 m<sup>3</sup>/s in Blackstone relative to a maximum flow of 109 m<sup>3</sup>/s, so less than 0.04% of the peak flow. We have added the average ice-on flows over the study period to the study site/background section for clarity. Ice-on winter low flows in

high latitude basins such as this commonly reduce significantly and become near zero due to the long, sustained period frozen ground/soils, lack of mid-winter thaw/melt periods, and accumulation of solid precipitation.

We considered providing panel b on Fig 3 & 4 in log-scale to emphasize that there are in fact low-flow values; but this greatly diminished peak flow analysis and peak flow uncertainty, which was a key point in our study. We have included those log-scale figures here for your assessment (not included in revised manuscript):



What are the influences of groundwater on the hydrology of the region? The same holds for section 4.3. Explain the results in more details.

Given the region resides within the discontinuous to semi-permafrost region of Canada, the influence of sub-surface contributions to runoff would be sporadic and is difficult to define (as several studies in the region have shown, Connon et al., 2015). The model we use in this study (isoWATFLOOD) has the capability to raise/lower wetland water table levels, connecting and/or disconnecting with channel runoff, which is a reasonable analogy to this complex interaction.

There is especially the time of the spring freshet that needs much more carefully discussed. We have incorporated an analysis of the results during spring freshet into our discussion.

The model results show a sharp drop of streamflow stable isotope signature, while the observed values are getting more and more enriched at that time. This completely opposed development may be related that the contribution of snowmelt water to total streamflow during that time is too high (or the signature of the snow melt signal is wrong, please remind here my suggestions above) and the contribution of baseflow too low.

We assume you are referring to the freshet period in 1998. Note that we did not have continuously observed isotopes in streamflow during the peak freshet (i.e., high flow sampling is not always feasible), and as a result there are some missing observations during this time of year (mostly in 1999), despite this being our most frequent period of sampling overall (relative to other seasons). Moreover, as we've explained the model assumes snowfall composition to be equal to snowpack composition, and then can apply a constant offset or fractionation from snowpack composition/accumulation to snowmelt. In this study, that offset was set =0 given the lack of snowpack to snowmelt observations from which to calibrate to. Therefore, it is most likely that, in this year, the assumed fractionation from pack to melt water was wrong and not well defined. Again – without observed data to compare to, it is impossible for us to adjust this factor to improve results;

# however, adding a snowmelt dynamics module to the model would be a great asset, one which has been recognized by our group and that we are working toward. We have added some text in the revised manuscript to discuss this discrepancy.

The contributions of baseflow (groundwater) to total streamflow during the post-freshet are especially for Jean Marie River much lower in the present model study compared to the results of St Amour et al. (2005). Furthermore, please provide the stable isotope signature of groundwater. And compare those observed values with the values generated by the models in the groundwater routine after the spin-up period.

As Stadnyk et al. (2005) and Stadnyk-Falcone (2008) pointed out, contributions of "groundwater" from the model (isoWATFLOOD) cannot be directly compared to those derived by St. Amour et al. (2005) owing to the definition of what groundwater is considered in the two modelling methodologies. In St. Amour et al (2005), a mixing model is used that separates old and new water contributions over time – which means that groundwater is defined as old water, or that is water that is existing pre-event. Whereas using WATFLOOD or isoWATFLOOD to perform hydrograph separation in the same region, lower contributions of groundwater are derived by the (iso)WATFLOOD model since the model separates soil water (upper zone storage) from baseflow or groundwater (lower zone storage) and wetland storage -- all of which would constitute 'old' (pre-event) water using traditional two-component mixing models. We have added text in the revised manuscript to describe this.

Please check the manuscript for repetitive information. The sentence on Page 13, 34+35 for example appears almost identical on the next page again (Page 14, Lines 20-22). This would be an excellent take-home sentence for the conclusions section by the way.

Thank you for pointing this out. We have re-read the manuscript and removed any apparent redundancies, particularly the ones you have pointed out to us. We have moved the sentence you highlighted to the conclusions section.

In general, I am missing some distinct conclusions in the conclusions section of the submitted manuscript. There are a lot of recommendations and speculations but no clear take-home messages.

Agreed. In re-reading the manuscript, we too realized that we can write better conclusions that highlight the take-home messages this manuscript presents. Also elaborated on in the conclusions now is the take-home message that precipitation isotope products of higher resolution (e.g., REMOiso, daily resolution) better capture event-specific compositions that, when significantly different from  $\delta^{18}O_{SF}$ , tend to cause significant deviations from seasonal and semi-annual (i.e., static) inputs. Though we cannot verify the correctness of the higher resolution product (REMOiso) in this study due to monthly observed precipitation, it is clear that temporal resolution plays a significant role in model parameterization and resulting hydrograph separations. We have also added a separate "*Future Directions*" section (based on Reviewer #2 feedback) that is comprised of the future work discussion from our original conclusions.

Technical notes:

Page 1, Line 18: "...to capture both the variability and seasonality". There it would be better to write "spatial variability and seasonality" or "spatial and temporal variability", since the seasonality is also a variability (temporal).

### We have made this correction.

Page 1, Line 31: (e.g. Beven and Binley, 1992; Kirchner....)
Correction made.
Page 3, Line 22 and Line 29: Please provide size and elevation characteristics of the basins here.
We have added this information.

Page 3, Line 27: "...is selected based ON data availability." Correction made.

Page 4, Line 20: The study region is not a high elevation region. Please mention correctly why the approach is suitable for the study region.

From another project our research group is working on, a detailed analysis of ANUSPLIN's suitability for high-latitude, Boreal regions (i.e., specifically the Nelson River) was done by a PhD student (Rajtantra Lilhare) and presented recently in a poster at ArcticNet (Lilhare, 2016). In this study, both the seasonality and amount of precipitation from ANUSPLIN were found to match well (r>=0.98) with nearby observations (3 for precipitation, 6 for temperature; all within the Nelson River watershed) from Provincial and Environment Canada meteorological station observations (shown here, but not included in our paper).



**ANUSPLIN Temperature** 



#### **ANUSPLIN Precipitation** Average Daily Precipitation @ Gillam Airport(1979-2009) Mean Monthly Precipitation @ Gillam Airport(1979-2009) 100 Oh 90 Observed ANUSPLIN ANUSPLIN 80 =0.99. RMSE=2.0 r= 0.93, RMSE=0.34 50 40 20 10 Jar Fet 196 211 226 241 256 271 286 30 181 316 331 Average Daily Precipitation @ Snow Lake(1984-1997) Mean Monthly Precipitation @ Snow Lake(1984-1997) 100 90 80 ANUSPLIN ANUSPLIN =0.98, RMSE=5.94 r= 0.88, RMSE=0.54 70 60 50 40 30 20 10 Jan Feb 196 211 226 241 Mar Ap 1.60 Auc Oct New erage Daily Precipitation @ Norwey House Forestry(1979-1999) thly Precipitation @ Norwey House Forestry (1979-1999) 100 Observed ANUSPLIN 90 ANUSPLIN 80 0.99, RMSE=2.6 r= 0.96, RMSE=0.30 70 60 50 40 30 20 10 Jan Feb Mar Apr May Jul Aug Sep Oct Nov Dec Jun Days

Simultaneously, we have been involved in an assessment of precipitation datasets and reanalysis products across the Canadian Prairies and Boreal region for the purposes of hydrological modelling applications. ANUSPLIN was included in this comparison, where data products were evaluated against independent station data (not used in the derivation of each product). A manuscript summarizing this comparison is currently in preparation by Dr. Bruce Davison, who found that ANUSPLIN scored well in terms of accuracy (relative to station observations), but showed some bias over the long-term. Based on our knowledge of ANUSPLIN for our study area, we believe that it is adequate to describe daily

## precipitation over the short term, but this decision would need to be reconsidered should the study length be extended.

Page 4, Line 21: "...is used TO spatially..." Correction made.

Page 5, Line 9: Why are they not adequate for model forcing? This is the input data used and referred STATIC in the study, right? Please revise this sentence **Our apologies. We have revised this sentence to instead state:** "...*their spatial and temporal resolutions are not preferred for tracer-aided hydrologic model forcing due the observations being uniform in space, and their poor temporal resolution.*"

Page 5, Line 12: "such that" appears twice. Corrected.

Page 5, Line 24: KP43 instead of KPN43. Corrected – thank you for noticing this!

Page 6, Line 4: From my point of view the section 2.4.1 is a description of methods and should therefore be moved in the appropriate section. We agree and have moved this section to a new section in study methods.

we agree and have moved this section to a new section in study methods.

Page 6, Line 16: Please mention that Snare Rapids is a CNIP station for clarity. We have added this information and clarified.

Page 6, Line 7: IAEA (2014) this citation is listed in the references section as IAEA/WMO (2014). Please adapt. **This has been corrected.** 

Page 7, Line 16: based on instead of based off. **Corrected.** 

Page 8, Line 5: The authors should reconsider the terms "behavioural" and "non-behavioural" for the model outputs of streamflow and stable isotope signature of streamflow. From my point of view those terms are not appropriate in this context. Reliable and non-reliable are terms coming to my mind here.

These terms are not our own and are taken from the modelling literature referring to whether or not a simulation meets the threshold criteria value (based on efficiency criteria for each study – and defined here as a combination of %Dv, log(%Dv), NSE, KGE, and RMSE) to remain "included" in the final analysis. The term behavioural refers to the fact that the simulation (and therefore parameters driving the simulation) are adequately describing the behaviour of the environmental system (i.e., hydrological response). Since this terminology is historically well defined in the model calibration and equifinality literature (e.g., Tolson & Shoemaker, 2008; Beven & Freer, 2001; Zak & Beven, 1999; Beven & Binley, 1992, ...), we would prefer not to deviate from the accepted terminology.

Page 8, Line 13: KGE. This abbreviation is introduced later (Line 23). Would be nice to have the explanation earlier.

Though we see your point, it would clutter the step-by-step methodology and we feel it would be out of place to put the statistic description further up. We have instead noted that the statistic is described below for readers who are unfamiliar with it.

Page 8, Line 30: Please mention for completeness that the other circa 52 % were sampled during the summer months.

## We have added this for clarification.

Page 10, Line 11-18: Please explain clearer that you are talking about the average streamflow simulations of the three calibrations used in this paragraph. The reader will otherwise think you are talking about an average streamflow simulation (Line 12) of all model runs. Further more please precise which model you are talking about at the end of line 12 and beginning of line 13 ("The model also has…").

Thank you for pointing this out. We agree and have revised this portion of the discussion to be much more specific to which runs we are referring (i.e., all models, the range and/or mean of the models, or a specific model derived from a particular  $\delta^{18}O_{ppt}$  input).

Page 10, Line 13: difficulty instead of difficultly Corrected – again, impressive that you noticed this! Many thanks.

Page 10, Line 20-27: Please explain shortly why you have compared REMOiso vs. static and KNP43 vs. static for calculating the Kendall's tau coefficient. We in fact calculated Tau for all possible comparisons (ie. KPN vs. REMOiso, KPN vs. static, REMOiso vs. static) for both basins, but did not report all values in the manuscript-instead reporting only the range of the values by selecting these specific pairings. Moreover, since static represents  $\delta^{18}O_{ppt}$  observations (annual average), by comparing REMOiso and KPN43 directly to static, we are in essence comparing them to simulations derived from mean annual  $\delta^{18}O_{ppt}$  observations.

Page 10, Line 29: Please revise the title of section 4.3 to Modelling delta oxygen-18 in streamflow). **Done.** 

Page 11, Line 14-16: Please check the literature and provide a reference here. We have provided the following reference where the authors looked a comparison of a decomposition of the NSE and KGE stats: Kling and Gupta (2009).

Page 11, Line 16: functions or function(s)? Functions. We have corrected this. Page 11, Line 21+22: Please provide some values (and percentages related to total annual precipitation) from mean annual precipitation for the mentioned periods (summer and fall, winter and spring).

We are including here a percentage breakdown for seasonal (summer/fall, or JJASON and winter/spring, or DJFMAM) snowfall and rainfall during our study period (1997-1999). We have provided some values in our revised manuscript.

Study Period (1997-1999)	Dec- May	June- Nov	TOTAL
Precipitation (TOTAL) (mm)	350.7	956.3	1307
Precipitation (% of total)	27%	<b>73%</b>	
Snowfall (mm)	257.4	196.6	454
Snowfall (% of total precip)	20%	15%	35%
Snowfall (% of total snowfall)	57%	43%	
Rainfall (mm)	93.3	759.7	853
Rainfall (% of total precip)	7%	58%	65%
Rainfall(% of total rainfall)	11%	<b>89%</b>	

In comparison to the long-term climate normal (1981-2010) at Fort Simpson Airport, we can see that our study period is reasonably representative of long-term conditions for this region – certainly within any observation error.

Climate Normal (1981-2010) Fort Simpson A	Dec- May	June- Nov	TOTAL
Precipitation (TOTAL) (mm)	117.4	270.2	387.6
Precipitation (% of total)	30%	<b>70%</b>	
Snowfall (cm)	119.9	67.2	187.1
Snowfall (mm)	93.6	55.5	149.1
Snowfall (% of total precip)	24%	14%	38%
Snowfall (% of total snowfall)	63%	37%	
Rainfall (mm)	23.8	214.7	238.5
Rainfall (% of total precip)	6%	55%	62%
Rainfall(% of total rainfall)	10%	<b>90%</b>	

Page 14, Lines 29-31: This sentence is a bit confusing. Please revise. We have edited this sentence in the process of revising the discussion.

Page 14, Line 31: isoWATFLOOD or WATFLOOD? isoWATFLOOD. This has been clarified.

Page 15, Line 4: isoWATFLOOD or WATFLOOD?

Actually, upon re-reading, we feel this pertains to hydrological models in general and have therefore revised our text to be more general.

# Page 15, Line 10: isoWATFLOOD or WATFLOOD? WATFLOOD. This has been corrected.

Page 16, Line 13: kpn or KPN43?

Modified to KPN43. Thank you.

Please check the citations carefully. Pietroniro et al. (1996) and Töyra et al. (1997) are listed in the references section (Page 18, Line 46 and Page 19, Line 34) but appear not in the manuscript itself.

Thank you for noticing this – we have gone through each reference and ensured there is a corresponding citation in-text. We have removed the references you noted were missing citations.

In general, I liked the style and the coloring of the figures. However, figure 2 and 3 are a bit unclear. It is a real asset to show the uncertainty bounds of the different calibrations. The authors should rethink the presentation of this data, especially the streamflow results (panel b). You have raised a really interesting perspective here! When we wrote the manuscript and prepared the figures, our interest was in how and where the uncertainty bounds overlapped and were NOT different – but we recognize that to some readers, where they differ is of more interest. Therefore, we have darkened and shaded the lines defining each uncertainty envelope so that readers can pick out the uncertainty bands related to each model, and their overlap/differences. (shown here are panel (b) for Figure 3 Jean Marie and Figure 4 Blackstone, respectively):



Further more I would suggest indicating periods with snowfall and rainfall, if possible. At this point it would also make sense to combine the two times series (static-rainfall and static-snowfall) to one static-precipitation input time-series.

Regarding rainfall and snowfall being combined into one time-series, we respectfully disagree since these are two distinct inputs in isoWATFLOOD that can be both used at the

# same time when there are rain-on-snow events – meaning that both compositions are needed to define the mixed composition of precipitation.

Figure 6: Are here shown the mean or median values (circle symbols)? We are showing mean values here, and have clarified in figure caption and in methods section.

Figure 7: Please refer to Table 6 (were the parameters are explained) in the figure caption. We have added this citation for Table 6.

The order of the table numbering in the text is sometimes were confused (Page 4, Line 4: Table 1; Page 4, Line 32: Table 4, for example). Please order them correctly.

This has been corrected and tables are now numbered in the order in which they are cited in text.

In general, I suggest reducing the amount of tables. Table 3 for example is not needed. The applied average correction values (and the range) can be mentioned in the text. Table 5 is also unnecessary. You can mention the values in the text. However, it would be very relevant to explain in more detail how these values were selected.

We have removed Tables 3 and 5 and included this information in the text instead.

Table 8 is also unnecessary from my point of view.

Given one of the primary goals of this study is to assess the impact of input choice (precipitation isotope product) on the model parameterization, we feel Table 8 contains highly valuable information for tracer-aided modellers tackling the same issues. Therefore, we are inclined to keep it included in our study, but have decided to include it as supplemental information instead of in the manuscript.

## Referee #2

Specific comments: My main point would be that the paper is in parts very much focussed on the particularities of the study site and also the presented model characteristics. However, the results and potential impact of this paper go in my opinion beyond this case study and this could be better emphasized to maximize impact particularly in the hydrological modeller community. I therefore, suggest the following:

We also agree that the findings presented in this manuscript go beyond our specific application to the Fort Simpson region and are therefore more general and impactful than we have conveyed them. We have edited the manuscript in a way that conveys our findings in a more general sense, specifically with respect to a range of study sites (particularly those that have seasonality as this one), isotope-enabled models, and modelling applications. Thank you for this feedback.

- Title and Abstract: You could consider substituting the term "estimated" with e.g. "precipitation isotope product" throughout the manuscript to emphasize the different origins of the input functions.

# We like this terminology and have adopted it for the revised title "*Examining the impacts of precipitation isotope products* ( $\delta^{18}O$ ) on distributed tracer-aided hydrological modelling", as well as throughout the paper. Thank you for the suggestion.

From Line 17 in the abstract, I suggest to revise these sentences, as they do not really reflect the key findings. For example, the statement that the model is only as good as its input function is rather trivial and could be changed to some more specific statement such as which temporal resolution is needed (hourly, daily, weekly...) to adequately simulate stream isotope signatures and which product is the best?

Thank you for this suggestion, and we also agree. We have reworded the abstract to instead state "Here we investigate the impact that choice of model 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". And perhaps more importantly, we have revised our discussion and conclusions to comment specifically on the impact that precipitation isotope product resolution has on model output. This has become one of our key take-home messages.

I also suggest to more specifically mention that the coupled simulation of flow and isotopes actually allowed you to constrain the simulations towards a better internal representation of the dominating processes.

We agree and have revised the last sentence in our abstract to state: "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."

- 2.2, Line 21:...is used "to" spatially distribute... Corrected, thank you.

- Page 7, Line 16:...based "on"? Corrected.

- Page 9, Line 14: Would it be feasible to test this for one model configuration and run it over let's say 100K iterations to be able to check for differences compared to 30K runs? Feasible, absolutely. In the time we have for edits to be submitted for this manuscript – no (we estimate it would take minimum 1 month, perhaps longer). That said, we are in the process of doing 100k runs with (iso)WATFLOOD in another northern basin to look at parameter identifiability with and without the use of isotopes in model calibration and nearing the end of those runs. We are planning to submit this manuscript for peer review within the next couple of months, where we will more definitively tackle the issue of parameter identifiability. Though we think this is a critical issue, it is not the intended focus of this manuscript, but rather follow up work that we now (more clearly) describe in the new "Future Directions" section of this manuscript. - Results and discussion: The results could be better linked to the wider literature. E.g. why not include the mean monthly precipitation isoscapes from Bowen and Revenaugh (2003) as a means of evaluation?

This is an interesting suggestion, however, this would only further evaluate KPN43 and REMOiso "products" and not  $\delta^{18}O_{sf}$  or other types of simulation output that are our intended focus. Bowen and Revenaugh's 2003 isoscapes are derived from long term average global models that did not include any CNIP data within their formulation, so we aren't convinced this would be a good dataset from which to further validate our REMOiso or KPN43 estimates of  $\delta^{18}O_{ppt}$  over the Fort Simpson region.

I am missing a more concise attempt to generalize the results concerning model uncertainty and the value of tracer data in hydrological modelling.

We agree and have revised the discussion section of the manuscript – and conclusions – extensively to help draw these generalized results into take-home conclusions for the broader tracer-aided modelling community.

- Page 10, Line 1: How is the static approach with a single annual isotope value able to capture seasonal variability?

The static approach is actually two annual isotope values: one for rainfall and one for snowfall. Therefore, technically speaking, the static approach is capable of capturing some seasonality. This is a point we have more clearly (and in more detail) described in the manuscript. The fact that the static input captures "sufficient seasonality" is likely more a function of our high-latitude study site than the value of a static input alone. Namely, in high-latitude environments, particularly Fort Simpson, there is no mid-winter freeze/thaw/melt – resulting in snowpack accumulation throughout the entire winter season and one significant freshet in late spring. Similarly, soils freeze up as does any soil moisture that may in other regions contribute to baseflow and/or streamflow throughout the winter. In high-latitude regions, seasonality is more binary than quarterly, therefore the two annual static inputs do a reasonable job of capturing the seasonality.

- Conclusions and recommendations: I suggest to summarize the key points and present them in a numbered order. I also think it would be better to present the outlook as a separate section. We have taken your suggestion to mean a numbered summary of the key take-home messages, which we have better aligned with the objectives and numbered accordingly in the conclusions section. With regards to "outlook", we assumed you mean future work to be done with the modelling, and have added a "Future Directions" section to this manuscript.

- Would it be possible to include gridded maps of the different mean annual (and seasonal min/max) isotope products over the study area in relation to the observed data for comparison purposes?

Thank you for this suggestion. Though we don't feel another figure is warranted in the manuscript, we see the value in these figures and the presentation of our precipitation isotope products for the modelling community and have decided to add it as a supplement to our manuscript (Figure S-1). To generate the spatially distributed precipitation isotope

products maps, daily isotope in precipitation input used to drive the distributed traceraided model was averaged daily across each season (DJF, MAM, JJA, SON) for each source (static, REMOiso, KPN43). Maps were generated using the model grid (10k) and entire modelling domain (includes both Jean-Marie and Blackstone), and isotope compositions were flux-weighted using daily distributed (10 k) precipitation input to WATFLOOD (interpolated Environment Canada station observation, housed in WATFLOODs radcl .r2c files; Kouwen 2014).

The resultant maps indicate clear differences in spatial variability among the inputs. Static – not surprisingly – is spatially constant (as it should be!), but seasonally variant resulting from the mixture of rain and snowfall events on the shoulder seasons (MAM and SON). REMOiso has less variability than the KPN43 input, resulting from REMOiso's 55 km grid resolution (i.e., ~5 of the isoWATFLOOD grids shown on our Figure) which would act to smooth topographical and land cover differences that are, in part, driving changes in precipitation isotopic composition. We've added a brief discussion to the paper and reference to Figure S-1.

For your interest and review – we also generated a figure (not included in the manuscript) averaged across the entire study period (1997-1999) for each model input:



This confirms the enhanced spatial variability from the KPN43 model, followed by REMOiso (derived from a 55km RCM), and the spatially constant Static input. Because of the high-latitude of the study region, the static input shows that snowfall prevails over rainfall for this site (in terms of isotopic composition), and that the 3-year annual average is more depleted than the temporally (and spatially) variable inputs. KPN43 variability is enhanced in the 3 year average because it is more consistent from grid-to-grid in each year (driven by the KPN43 regionalization) than REMOiso, which would vary temporally and spatially daily and from year to year.

We could not generate an observed isotope in precipitation map because we did not have enough observed data to so.

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

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

<sup>5</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 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

- 10 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.,
- 15 semi-annual static, seasonal KPN43, and daily bias corrected REMOiso), and apply them to force the isoWATFLOOD traceraided 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
- 20 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
- <sup>25</sup>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

30 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

10

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; 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, Di tale a 10, 2016).

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<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.,

- 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

30

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 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,

- 5 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
- 10 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 δ<sup>18</sup>O<sub>ppt</sub> input, impact regional tracer-aided model simulations and parameterization. Forcing a tracer-aided, distributed hydrological model

- 15 (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

20

25

#### 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.
- 30 This study considers two sub-basins of the greater Fort Simpson basin: the Jean-Marie (1310 km<sup>2</sup>) and Blackstone River (1390 km<sup>2</sup>) 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
  - 3

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 or towards the beginning of the ice-on season in October.

15

#### 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 datasparse, 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 ≥ 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<sup>3</sup>/s and 0.034 m<sup>3</sup>/s for the Jean Marie and Blackstone sub-basins, respectively. A statistical summary of observations used in this study is provided in Table 2.

#### 5 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  $\delta^{18}$ O and  $\delta^{2}$ H 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

- 10 composition of winter streamflow (Gibson and Prowse, 1999; Prowse et al., 2002; St Amour et al., 2005). This study uses measured  $\delta^{18}$ O compositions in streamflow in the Jean-Marie (n = 71) and Blackstone (n = 69) sub-basins for model calibration. Although  $\delta^{18}$ O<sub>ppt</sub> 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'  $\delta^{18}$ O<sub>ppt</sub> input, and as a means to validate the KPN43 and REMOiso products and to
- 15 inform the static precipitation product. The number of measurements and their statistical properties are summarized in Table 2. Isotopic compositions of  $\delta^{18}$ 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  $\delta_{water} = (R_{water}/R_{VSMOW} - 1) \times 1000$  ‰, 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  $\pm 0.1$  ‰ for  $\delta^{18}$ O.

#### 20 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  $\delta^{18}O_{ppt}$  compositions of rainfall and snowfall for each year of simulation (i.e., yearly resolution). Values for the FSB were obtained by averaging observations of  $\delta^{18}O$  in rainfall and the snowpack obtained from the GEWEX study

- 25 (Table 2; Table 3). δ<sup>18</sup>O<sub>ppt</sub> 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
- 30 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 δ<sup>18</sup>O<sub>ppt</sub> values (rain and snow) for each year. This product is specifically designed and evaluated for remote regions that lack spatially and temporally varying δ<sup>18</sup>O<sub>ppt</sub> observations.

Times series simulations obtained from the KPN43 model created by Delavau et al. (2015) are used as the second type of  $\delta^{18}O_{ppt}$  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

15

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 subdaily  $\delta^{18}O_{ppt}$  are erroneously large, and the resolution of streamflow oxygen-18 calibration data do not warrant a temporal 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 hydrologic storages of river/fen water, soil water, baseflow, and snowpack are needed for model initialization, which can be

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.

5

10

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 δ<sup>18</sup>O<sub>SF</sub>. 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).

20

25

30

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

- 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 δ<sup>18</sup>O<sub>SF</sub> 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 δ<sup>18</sup>O<sub>SF</sub>:
    - a. Streamflow:

NSE  $\geq$  0.5; |%Dv|  $\leq$  20 %, and;

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

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

RMSE  $\leq 2.5$  ‰, 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 δ<sup>18</sup>O<sub>SF</sub> 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 Precipitation (GNIP) database and can be accessed at: http://www.iaea.org/water (IAEA/WMO, 2014). This analysis is

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):

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

(1)

where P<sub>i</sub> 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 δ<sup>18</sup>O<sub>ppt</sub> 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 5<sup>th</sup> and 95<sup>th</sup> percentiles drawn from the ensembles of behavioural simulations; denoted as the shaded bounds around each model's mean simulation.

- 20 Kendall's tau coefficient ( $\tau$ ) 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  $\tau$  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 radel.r2c files; Kouwen 2014). Maps are

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

#### 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

- 5 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
- 10 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

- 15 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
- 20 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
- 25 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.
- 30 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 refelecting  $\delta^{18}O_{ppt}$  seasonality because of the high-latitude

location of the basin, where much shorter shoulder seasons exist.

#### 4.2 Modelling streamflow

wetland dynamics (St Amour et al., 2005).

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
 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

15

5

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

- 20 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.
- 25 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,
- 30 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

% (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 ( $\tau$ ). 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-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 δ<sup>18</sup>O<sub>SF</sub> record, there are notable departures. As simulated δ<sup>18</sup>O<sub>SF</sub> uncertainly envelopes associated with each δ<sup>18</sup>O<sub>ppt</sub> product are, at times, non-overlapping,
  differences in δ<sup>18</sup>O<sub>SF</sub> simulations can be attributed to δ<sup>18</sup>O<sub>ppt</sub> product and, therefore, are not just an artefact of parameter uncertainty (unlike streamflow). The dissimilarities between δ<sup>18</sup>O<sub>SF</sub> 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 δ<sup>18</sup>O<sub>SF</sub> exist. The KPN43 and REMOiso calibrations produce comparable RMSE. The KGE statistic shows only minor differences between δ<sup>18</sup>O<sub>SF</sub> 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 δ<sup>18</sup>O<sub>SF</sub> simulations within each sub-basin are due to a combination of: (1) markedly different δ<sup>18</sup>O<sub>ppt</sub> 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).

10

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

- 15 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.
- 20 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  $\geq 40$
- 25 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
- 30 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,

# this single event results in a significant offset between KPN43-driven δ<sup>18</sup>O<sub>SF</sub> simulations relative to those driven by REMOiso and static products, maintained throughout 1998 and up until the 1999 freshet resets the δ<sup>18</sup>O<sub>SF</sub>.

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 (δ<sup>18</sup>O<sub>SNW</sub>) 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

- 10  $\delta^{18}O_{SNW}$ , 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 depthdependent 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,
- 15 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
- 20 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
- 25 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, 30 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

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

Since both δ<sup>18</sup>O<sub>SF</sub> and δ<sup>18</sup>O<sub>SNW</sub> are significantly different among δ<sup>18</sup>O<sub>ppt</sub> 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 δ<sup>18</sup>O<sub>ppt</sub> 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 δ<sup>18</sup>O<sub>ppt</sub> 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 %  $\pm$  9 % and 64 %  $\pm$  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  $\pm$  0.8 months). Their findings also
- 20 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-
- 25 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 δ<sup>18</sup>O<sub>ppt</sub> product reveal that during spring (MAM), REMOisodriven simulations show more surface flow contribution to total streamflow, with the mean volume lying above the 95<sup>th</sup> 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).

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.

- 5 6). The variation in uncertainty bounds between δ<sup>18</sup>O<sub>ppt</sub> 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
- 10 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 δ<sup>18</sup>O<sub>ppt</sub> input. The selected parameters (Table 4) influence evaporation (fratio), 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.
- 25

30

15

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

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

#### 5 Conclusions

5

15

20

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

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 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

<sup>30</sup> 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

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 δ<sup>18</sup>O<sub>ppt</sub> 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 δ<sup>18</sup>O<sub>ppt</sub> observations (i.e., from CNIP) to either define or bias correct the input, limiting their use for certain applications. If
- 10 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.
- 15

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

- 20 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
- 25  $\delta^{18}O_{SF}$ . 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.
- 30

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

- 5 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
- 10 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.
- 15 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.
- 20

30

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

25 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 δ<sup>18</sup>O<sub>ppt</sub> 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.

#### **Competing interests**

The authors declare that they have no conflict of interest.

#### 5 Acknowledgements

The authors would like to acknowledge the contributions of the late Dr. Peter Rasmussen to this work, who's invaluable advice and mentorship is sorely missed but largely shaped the outcome of this research. 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

- 10 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 Water Survey of Canada for the hydrometric data, and 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
- 15 Engineering Research Council (NSERC) Alexander Graham Bell Canada Graduate Scholarship (CGS-D).

#### References

25

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.
  - 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
- 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.
- 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
   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.:
   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– 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
   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.

- 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.

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

- 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 d180 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,
   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 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:
  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. 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
   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

5



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



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



Figure 3: Input and behavioural simulations for Jean Marie, including: (a) KPN43, REMOiso and static  $\delta^{18}O_{ppt}$  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<sub>SNW</sub>) 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<sub>ppt</sub> input-specific uncertainty bounds are represented as the shaded regions. Diamond symbols represent  $\delta^{18}$ O<sub>SNW</sub> 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 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  $\delta^{18}O_{ppt}$  input types are shaded in gray.



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

Variable (gauge ID)	Unit	Number of	Mean	CV	SP Range	
(9	2	Measurements	(SP, PoR)	(SP, PoR)	(min, max)	
	Hydroi	metric/Meteorological	Data			
Daily Average Streamflow Jean Marie (10FB005)	m <sup>3</sup> /s	1095	4.66, 5.25	1.24, 2.06	0.19, 34.9	
Daily Average Streamflow Blackstone (10ED007)	m <sup>3</sup> /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	
	Ise	otopic Measurements*				
Streamflow δ <sup>18</sup> O - Jean Marie	‰	71	-19.70	0.03	-21.34, -18.72	
Streamflow δ <sup>18</sup> O - Blackstone	‰	69	-20.17	0.06	-24.01, -17.92	
Rainfall δ <sup>18</sup> Ο Jean Marie and Blackstone	‰	27	-17.55	0.23	-26.70, -11.12	
Precipitation δ <sup>18</sup> O Forcing*						
KPN43 δ <sup>18</sup> O <sub>ppt</sub> input	‰	1800 (36 values at 50 grid points)	-20.48	0.19	-28.86, -13.91	
REMOiso δ <sup>18</sup> O <sub>ppt</sub> input	‰	54750 (1095 values at 50 grid points)	-21.78	0.25	-42.82, -10.68	
Static δ <sup>18</sup> O <sub>ppt</sub> input	‰	300 (6 values at 50 grid points)	-22.82	0.20	-29.35, -16.52	

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.

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

Year	δ <sup>18</sup> O rainfall	δ <sup>18</sup> O snowfall
	(‰)	(‰)
1996	-17.00	-29.35
1997	-19.10	-29.35
1998	-20.10	-25.03
1999	-16.52	-26.79

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

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:					
	Lower zone drainage function	$QLZ = LZF * (LZS)^{PWR}$					
pwr	exponent	Where: LZS is lower zone storage					
	exponent	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-					
akfs	Surface permeability	based (Philip, 1954)					
rec	Interflow coefficient	Interflow is represented by a simple storage-discharge relation:					
		DUZ = REC * (UZS-RETN)*Si					
retn	Upper zone retention [mm]	Where: UZS = upper zone storage 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 (10	95 observations for	r performance eval	luation)		
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		