



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)
- Rewriting 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 expanded 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  $\delta^{18}\text{O}_{\text{ppt}}$  models and code with interested researchers upon request ([Carly.Delavau@gov.mb.ca](mailto: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 (<http://hdl.handle.net/1993/31946>). 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  $\delta^{18}\text{O}_{\text{ppt}}$ . 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.

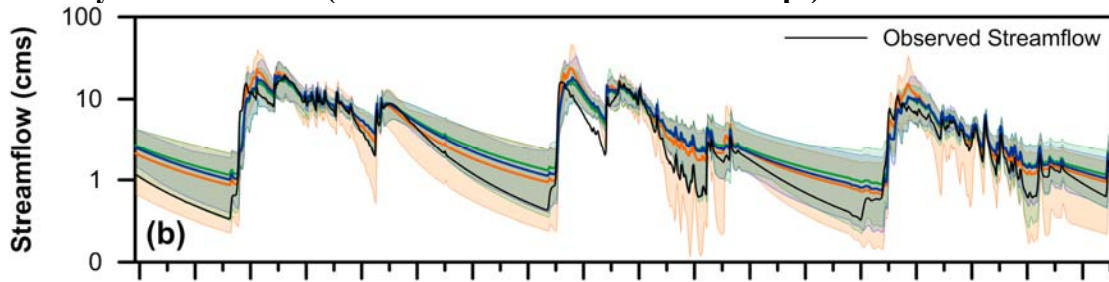
**Thank you for your suggestions, we have revised the discussion to include more specific, in-depth discussion of the simulated streamflow resulting from the three types of precipitation isotope product. And yes, all three precipitation isotope products (three different calibrations) result in almost 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  $\delta^{18}\text{O}_{\text{SF}}$ , 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.**

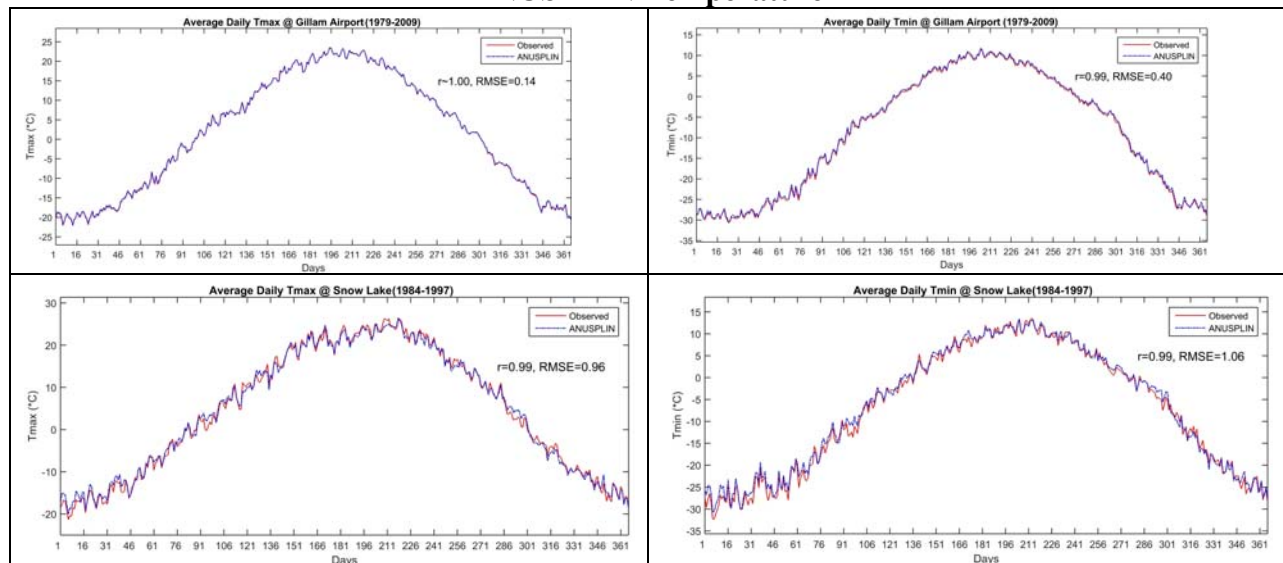
Page 3, Line 27: "...is selected based ON data availability."

**Correction made.**

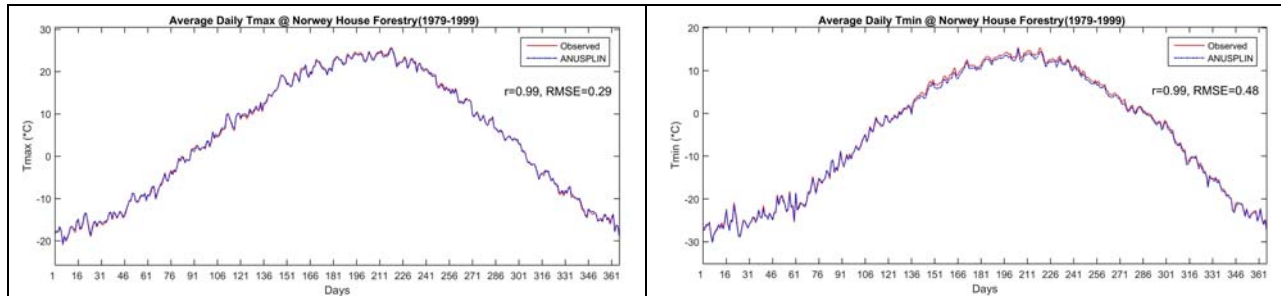
Page 4, Line 20: The study region is not a high elevation region. Please mention correctly why the approach is suitable for the study region.

**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 ( $r \geq 0.98$ ) 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).**

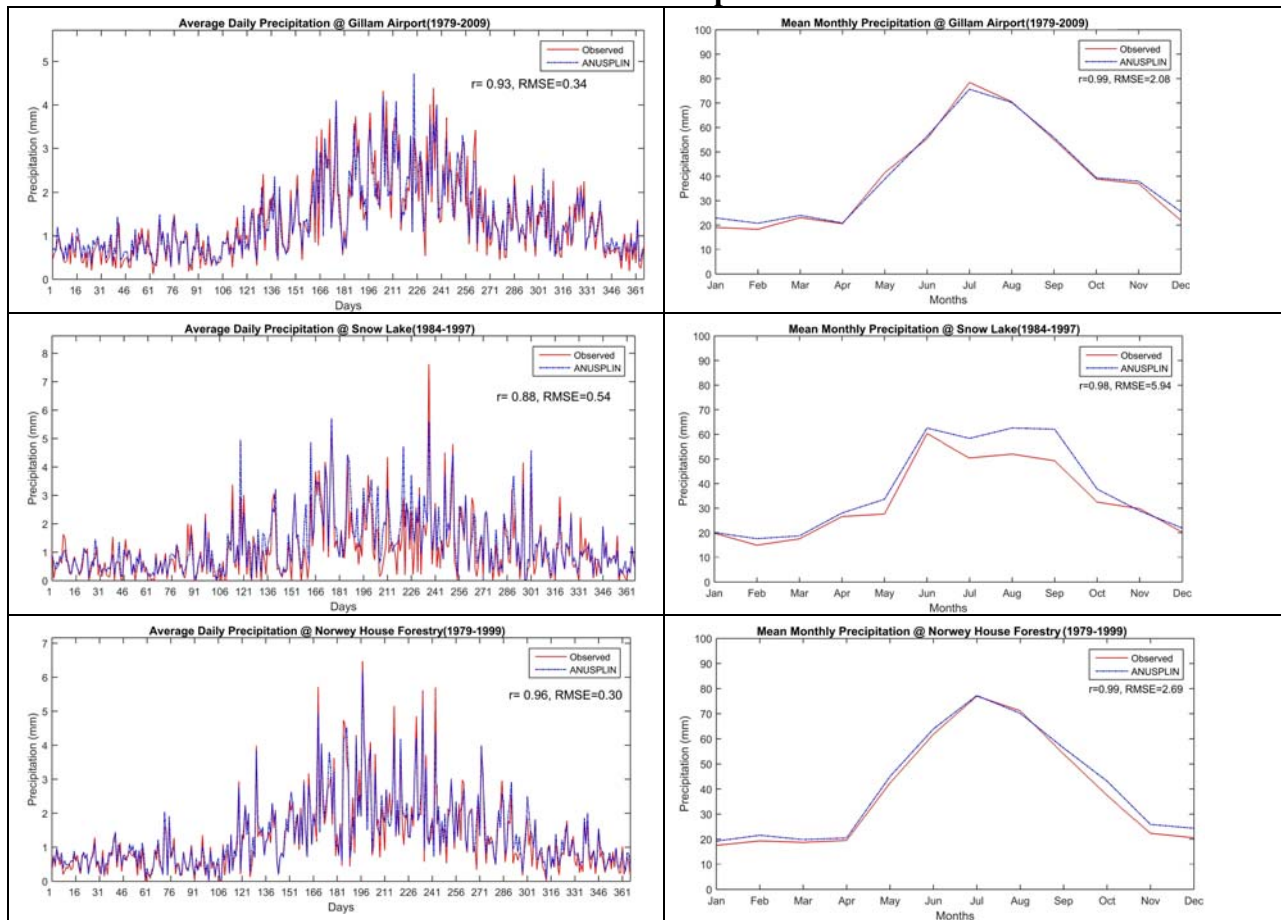
### ANUSPLIN Temperature







### ANUSPLIN Precipitation



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.**

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  $\delta^{18}\text{O}_{\text{ppt}}$  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  $\delta^{18}\text{O}_{\text{ppt}}$  observations (annual average), by comparing REMOiso and KPN43 directly to static, we are in essence comparing them to simulations derived from mean annual  $\delta^{18}\text{O}_{\text{ppt}}$  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
<b>Precipitation (% of total)</b>	<b>27%</b>	<b>73%</b>	
Snowfall (mm)	257.4	196.6	454
<b>Snowfall (% of total precip)</b>	<b>20%</b>	<b>15%</b>	<b>35%</b>
<b>Snowfall (% of total snowfall)</b>	<b>57%</b>	<b>43%</b>	
Rainfall (mm)	93.3	759.7	853
<b>Rainfall (% of total precip)</b>	<b>7%</b>	<b>58%</b>	<b>65%</b>
<b>Rainfall(% of total rainfall)</b>	<b>11%</b>	<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
<b>Precipitation (% of total)</b>	<b>30%</b>	<b>70%</b>	
Snowfall (cm)	119.9	67.2	187.1
Snowfall (mm)	93.6	55.5	149.1
<b>Snowfall (% of total precip)</b>	<b>24%</b>	<b>14%</b>	<b>38%</b>
<b>Snowfall (% of total snowfall)</b>	<b>63%</b>	<b>37%</b>	
Rainfall (mm)	23.8	214.7	238.5
<b>Rainfall (% of total precip)</b>	<b>6%</b>	<b>55%</b>	<b>62%</b>
<b>Rainfall(% of total rainfall)</b>	<b>10%</b>	<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?

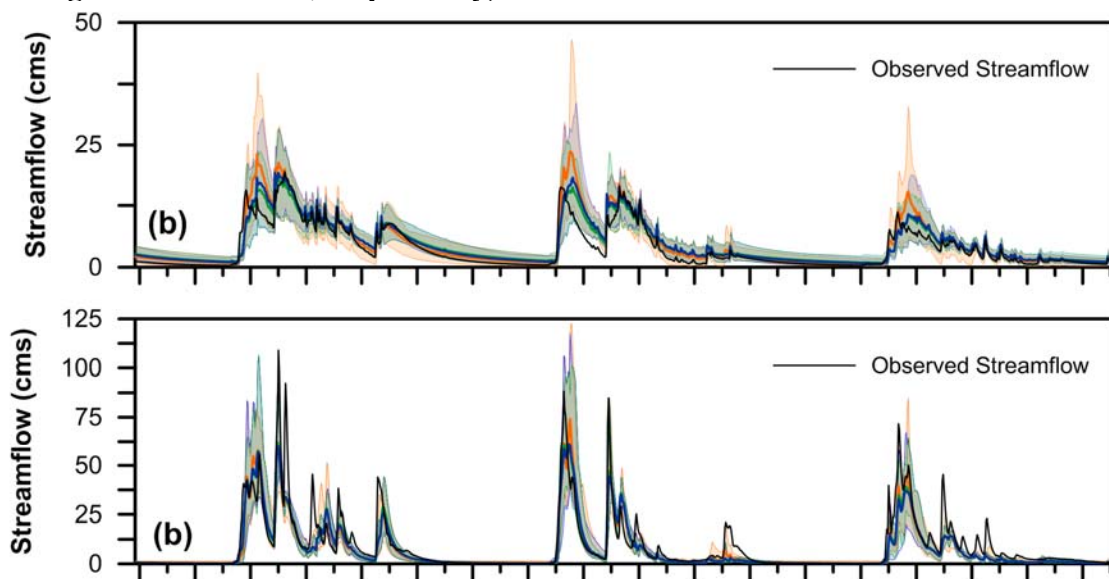
**Modified to KPN43. Thank you.**

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 time 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 ( $\delta^{18}\text{O}$ ) 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 ( $\delta^{18}\text{O}_{ppi}$ ) has on simulations of streamflow,  $\delta^{18}\text{O}$  in streamflow ( $\delta^{18}\text{O}_{SF}$ ), 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  $\delta^{18}\text{O}_{\text{sf}}$  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  $\delta^{18}\text{O}_{\text{ppt}}$  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?

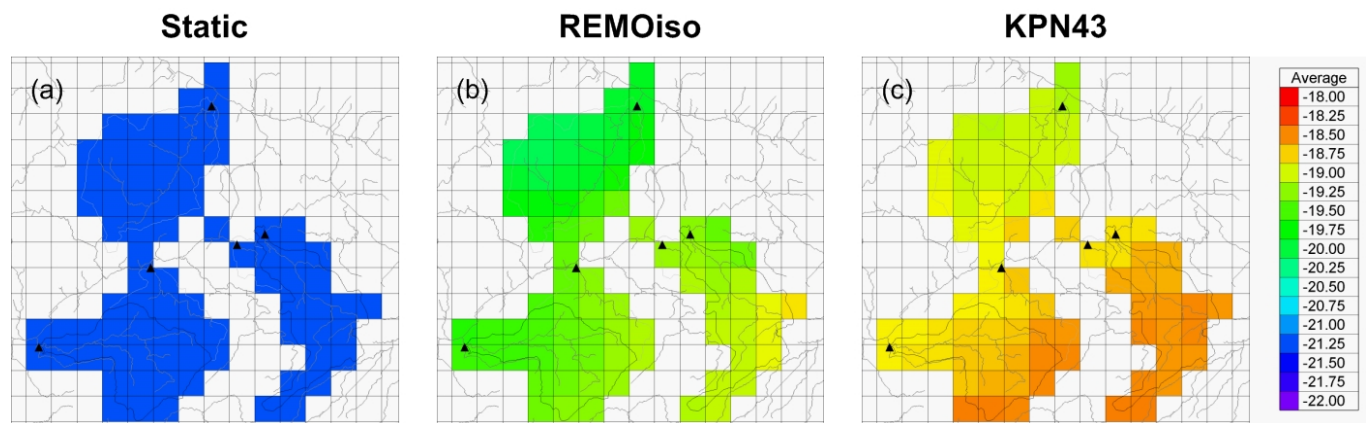
**Thank you for this suggestion. Though we don't feel another figure is warranted in the manuscript, we see the value in these figures and the presentation of our precipitation isotope products for the modelling community and have decided to add it as a supplement to our manuscript (Figure S-1). To generate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



products maps, daily isotope in precipitation input used to drive the distributed tracer-aided 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.