Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2016-539-AC3, 2017

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

Interactive comment on "Examining the impacts of estimated precipitation isotope (δ^{18}O) inputs on distributed tracer-aided hydrological modelling" *by* Carly J. Delavau et al.

Carly J. Delavau et al.

Tricia.Stadnyk@umanitoba.ca

Received and published: 14 January 2017

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 and edits to the paper are provided below in bold following the individual comments requiring action from reviewer RC1.

Please do not hesitate to contact us if any of this information is not clear.

Printer-friendly version

Discussion paper

With kind regards,
Tricia Stadnyk (on behalf of co-authors)

Referee 1:
General comments: The authors present an interesting case study about using differing precipitation stable isotope input datasets for distributed hydrological modeling in the Northwest Territories of Canada. Based on three different precipitation stable isotope datasets, three different calibrations of the model isoWATFLOOD were identified based on a Monte Carlo random sampling approach. The results show that modeled streamflow was relatively similar for each of the three used stable isotope input datasets. Whereas the differences in the modeled stable isotope signature in streamflow and the internal apportionment were much more pronounced. However, the study is lacking some important and critical explanation of the presented model outputs. Please find a detailed description of specific comments and technical notes below. The focus of the presented study is in the scope of HESS. Further more I will highlight at this point that the paper is very well written, understandable and the sections are mostly very well structured. However, based on my review below I recommend major revisions prior to a publication in HESS. **Thank you kindly for this summary of our paper. Indeed, we agree based on the reviewers assessment that the discussion can be enhanced with respect to model outputs – we thank you for your detailed assessment and guidance provided. We believe the changes you've suggested have greatly improved the quality of this manuscript.**

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 (δ^{18}O) on distributed tracer-aided hydrological modelling* and revised the use of the word 'estimated' (with respect to δ^{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 are currently conducting such a study within our group (though it has yet to be published), and do have internal evidence that this is the case. That said, 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 aides in model calibration and the 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 been studying this topic for more than five years now, in part through an IAEA coordinated research project experimenting with methods to collect isotopes in snowmelt (Penna et al., 2014), and looking at seasonal changes globally between snowpack composition and snowmelt.

This being said, in this manuscript we need to be diligent in how we handle the topic since we did not collect the isotopes in snow data, and there is no specific legacy of how or where it was collected from (i.e., from what part of the snowpack, averaged depth dependent or composite samples, and unknown spatial variability of the samples). This is one of the reasons why we chose not to include the snowpack compositions on Figure 5 originally. We have since revised the figure and added the snowpack data and also included a cautionary note to readers highlighting there is uncertainty surrounding these measurements. The revised Figure 5 is included in this response (Fig. 1).

For the modelling, as a static input our model would preferably use average annual inputs of rainfall and snowfall. Rainfall values were, as reviewer 1 notes, obtained from the GEWEX campaign. However, there was no data on snowfall composition available – only snowpack compositions - therefore (as in several of other data limited, high latitude tracer-aided modelling studies: Stadnyk et al. 2013; Smith et al., 2015; Smith et al., 2016; Holmes 2016), we assume (as model input) that the average annual composition of snowfall is approximately

equal to that of the snowpack samples from the GEWEX campaign. From our own experiments, we know this is not always the case (typically true only in the short term immediately after a snowfall). With no other snowfall composition data, however, this is an assumption we have been required to make. We have clarified our assumption in the manuscript. We would also like to point out that due to the high latitude of this and several other of our sites, freeze-thaw cycles common in snowpacks are in fact rare in high-latitude (northern Canadian) snowpack where temperatures remain significantly below freezing for the entire winter season – as was found when we compared our (southern in comparison to this field site) Winnipeg, MB, Canada site to the other IAEA study sites included in Penna et al., 2014.

Lastly, we absolutely agree the representation of snowfall, snowpack and snow melt compositions in modelling (particularly high-latitude, seasonal regions) is absolutely crucial, which is why our group is putting extensive resources into resolving some of the uncertainty surrounding these processes and the evolution of isotopic compositions through these processes. Thank you for your feedback, and for reaffirming the importance of this issue!

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? **We are not entirely sure of what you mean by this question, but will attempt to answer it as best we can within the context of what we did in this study. 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. Now, if what you are getting at is why did we not precipitation amount weight**

Printer-friendly version

Discussion paper

REMOiso using REMO-derived precipitation for use in this study, then our answer is as follows. We averaged the four 6-hourly REMOiso values (at each grid) to arrive at daily compositions that were directly 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 δ^{18}O$_{ppt}$ (i.e., in the same way that we are skeptical of sub-daily REMOiso ^{18}O compositions, which is part of the reason that we decided to average output daily to avoid unrealistic variation). 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. We hope we have addressed your question and your concerns.

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 model used in this study (isoWATLOOD) is based on the version used by Stadnyk et al. (2013), noting that a different version of isoWATFLOOD and the Fort Simpson watershed model were used here that incorporates various model improvements made since 2013, independent of this study*

Printer-friendly version

Discussion paper

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). This is the core definition of equifinality, illustrated here in this study! Despite there being significant differences in the parameters, the net result of the simulation remains almost identical (driven by the requirement for the model to meet specific efficiency criteria). This can happen by changing how and where water is stored internally in the model – greatly affecting the transit time of water through the model and into the stream – but ultimately not impacting the total flow simulated by the model (because various internal processes trade-off in their respective contributions). Again, this was highlighted in this study by the fact that upon closer examination of the model simulations, internal apportionment of water was significantly altered from one input (calibration) to the next, particularly when comparing REMOiso to the KPN and static calibrations. This the very essence of our study! Your specific questions regarding this section of the discussion are answered below.**

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^3/s, or 0.5 percent of the maximum streamflow, 35 m^3/s during this same period; and a minimum of 0.043 m^3/s in Blackstone relative to a maximum flow of 109 m^3/s, so less than 0.04 percent of the peak flow. 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. E.g., Figure 3 panel b in log-scale, which is not included in the revised manuscript but is provided in this response for your reference (Fig. 2). We have added some text regarding low flow, ice-on streamflow values to the manuscript though.

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). We would argue that groundwater is not as influential as the bog complexes (or bog cascades as Connon et al., 2015 defined them), which depending on wetness levels, interconnect and disconnect seasonally and inter-annually. 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. It is likely that the timing of the simulated snowmelt contribution to runoff from the model resulted in what appears to be an overlap between the most depleted simulated isotopic composition of streamflow with observational data that is enriching (i.e., which is in fact post-**

freshet in 1998 and 1999 due to the presence of a strong El Nino event, noted by St. Amour et al 2005, and occurring due to evaporative fractionation). Similarly, in 1999, note the gap between ice-on observed compositions of streamflow and the enriching values (i.e., again post-freshet and occurring due to evaporative enrichment) relative to the most depleted simulated isotopic composition of streamflow, occurring around the same time due to snowmelt-driven runoff (i.e., and a later-than-observed snowmelt period in the model, more clearly distinguishable in 1999, Fig 3, panel b). These differences are likely occurring as a result of differential warming (rate and onset) caused by the El Nino event in 1998 and somewhat in 1999 that results in contrasting behaviours (from 1997, but also between observations and our calibrated model). Regarding the baseflow or groundwater composition comment you had, please see our response below (next comment/response).

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 (Stadnyk et al. 2005) or isoWATFLOOD (Stadnyk-Falcone, 2008) 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. Regarding groundwater isotopic composition: if groundwater was sampled, we do not have the data. It was our understanding (from speaking with Natalie St. Amour) that her 2005 paper used ice-on low flow to define baseflow or groundwater contribution (Table IV, St. Amour et al., 2005). Her paper suggests groundwater compositions are -20.5+/- 0.8 per mille (in 1998) and -20.4 +/- 1.0 per mille (in 1999). Uncertainty is due to averaging across all the five Fort Simpson basins. Modelled groundwater compositions in isoWATFLOOD were found to be 40-70 percent and 60-70 percent for Jean Marie and Blackstone, respectively during the post-freshet (JJASON) period.

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've 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. Our conclusions now start with the following numbered take-home conclusions, which are further elaborated on in the conclusions section; specifically, *that choice of precipitation isotope product: 1. Does not impact simulation of total streamflow; 2) Impacts model parameterization, and therefore modelling uncertainty; 3) Impacts internal apportionment of water in the model (through model parameterization), impacting resultant hydrograph separation -***

and therefore simulated transit times of water; and 4) impacted δ^{18}O$_{sf}$ most significantly when event composition differed significantly from streamflow composition (e.g., snowmelt and large rainfall events).

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 δ^{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 AGU (Lilhare, 2016). In this study, both the seasonality and amount of precipitation from ANUSPLIN were found to match well with observations from three nearby (within**

the Nelson River watershed) Environment Canada meteorological station observations. 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 percent Dv, log(percent 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. Moreover, we don't believe the term reliable captures what we are doing here. Multiple simulations can all have the same statistical likelihood, therefore all reliably predict a given result (statistical likelihood, or efficacy criteria). But some may do so with parameter values that are unrealistic and not representative of the environmental system (i.e., non-behavioural).

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 **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 δ^{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, but instead reported only the range of the values by selecting these specific pairings. Moreover, Since static represents δ^{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 δ^{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, H.V., and H. 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 a tabular summary that includes a percentage breakdown for seasonal (summer/fall, or JJASON and winter/spring, or DJFMAM) snowfall and rainfall during our study period (1997-1999) in this response (Fig. 3). 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 (Fig 4).
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 on Fig. 5 are the revised 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, where Ptotal represents the sum of snow water equivalent (SWE) and rainfall, used in the model based on: δP = (δ_{rain} x RAIN + δ_{snow} x SWE)/Ptotal. Since both distinct compositions can be used/needed in the same time step, we feel it is important to distinguish the time-series' and show them independently.**

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 Table S-1).**

References cited in our response:

Beven K, and Freer J: Equifinality, data assimilation and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. J Hydrology 249: 11-29, 2001.

Beven, K., and A. Binley: The future of distributed modelsâĂŤModel calibration and uncertainty prediction, Hydrol. Processes, 6(3), 279– 298, 1992.

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. Davison B, Chun K-P, Fortin V, LeConte R, Liu A, Mekonnen M, Stadnyk T, Wheater H. Verification of Gridded Rainfall Products in Central-Western Canada. In preparation.

Holmes, T. L. (2016), Assessing the values of stable water isotopes in hydrologic modeling: A dual-isotope approach. M.Sc. Thesis, University of Manitoba, Winnipeg, 198 pp.

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

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.

Lilhare, R.: High-resolution hydrological modelling of the Lower Nelson River Basin, Manitoba, Canada. Poster, ArcticNet, Winnipeg MB, Canada 5-9 December 2016.

Penna, D., M. Ahmad, S. J. Birks, L. Bouchaou, M. Brenčič, S. Butt, L. Holko,G. Jeelani, D. E. Martínez, G. Melikadze, J. B. Shanley, S. A. Sokratov, T. Stadnyk, A. Sugimoto, and P. Vreča (2014), A new method of snowmelt sampling for water stable isotopes, Hydrol. Process., 28, pages 5367–5644, doi: 10.1002/hyp.10273.

Smith A., Delavau C., and Stadnyk., T., 2015. 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.
Smith A., Welch C., and Stadnyk., T., 2016. Assessment of a lumped coupled flow-isotope model in data scarce Boreal catchments. Hydrological Processes, 30: 3871-3884.
Stadnyk, T. A., C. Delavau, N. Kouwen, and T. W. D. Edwards (2013), 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.
Stadnyk-Falcone, T. A., (2008), 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.
Tetzlaff, D., J. Buttle, S. K. Carey, K. McGuire, H. Laudon, and C. Soulsby (2015), 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.
Tolson BA, and Shoemaker CA: Efficient prediction uncertainty approximation in the calibration of environmental simulation models. Water Resources Research, 44:W04411, 2008.
Zak, S. K., and K. J. Beven: Equifinality, sensitivity and predictive uncertainty in the estimation of critical loads, Sci. Total Environ., 236(1–3), 191–214, 1999.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-539, 2016.

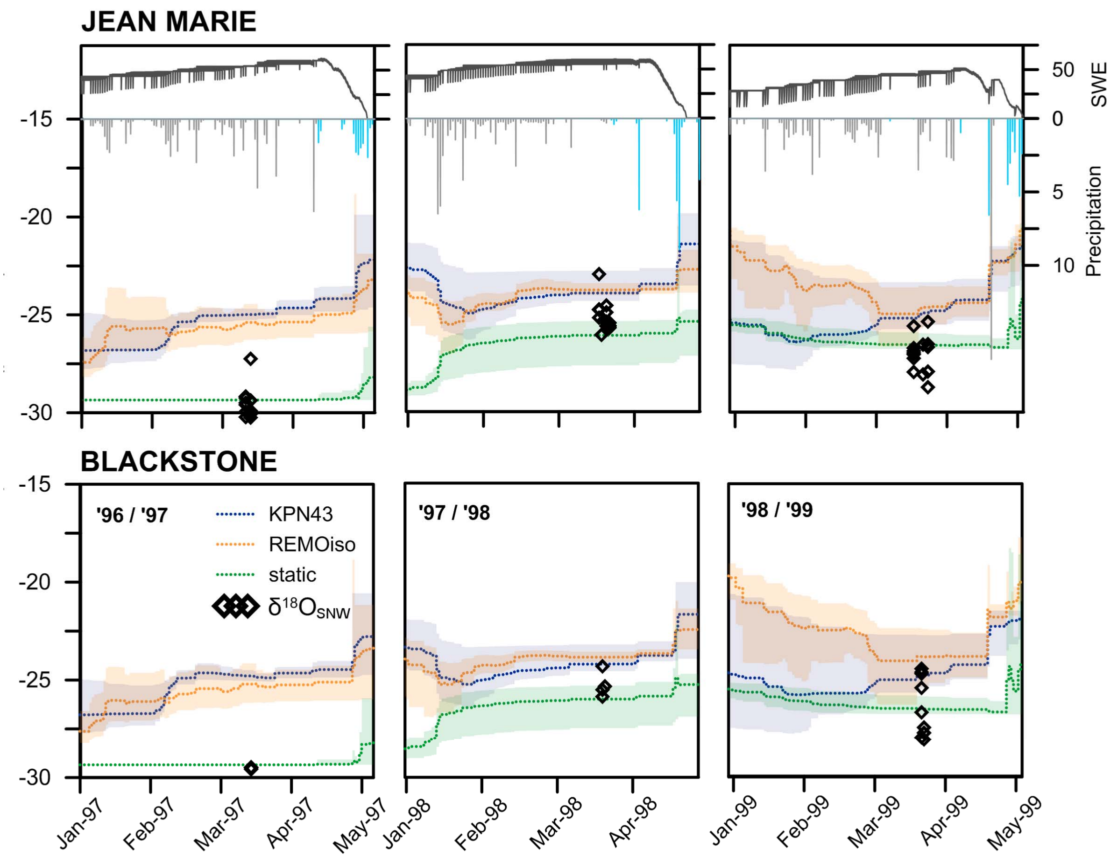

Fig. 1. Revised Fig.5 from manuscript, showing observations of snowpack isotopes

Printer-friendly version

Discussion paper

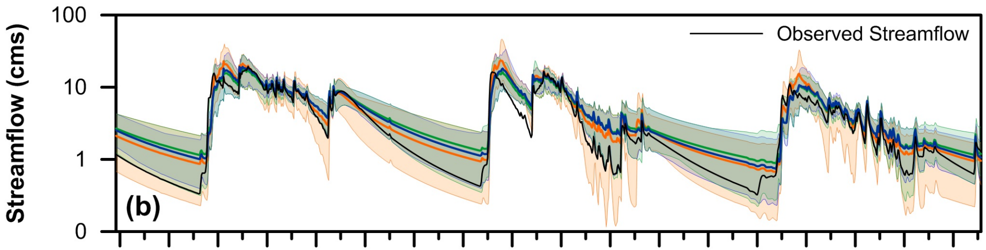

Fig. 2. Fig. 3 panel b in log-scale

Printer-friendly version

Discussion paper

Study Period (1997-1999)	Dec-May	June-Nov	TOTAL
Precipitation (TOTAL) (mm)	350.7	956.3	1307
Precipitation (% of total)	***27%***	***73%***	
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%***	***89%***	

Fig. 3. Tabular summary of percent breakdown for seasonal rain and snowfall

Printer-friendly version

Discussion paper

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%***	***70%***	
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%***	***90%***	

Fig. 4. Tabular summary of climate normal (1981-2010) rain and snowfall

Printer-friendly version

Discussion paper

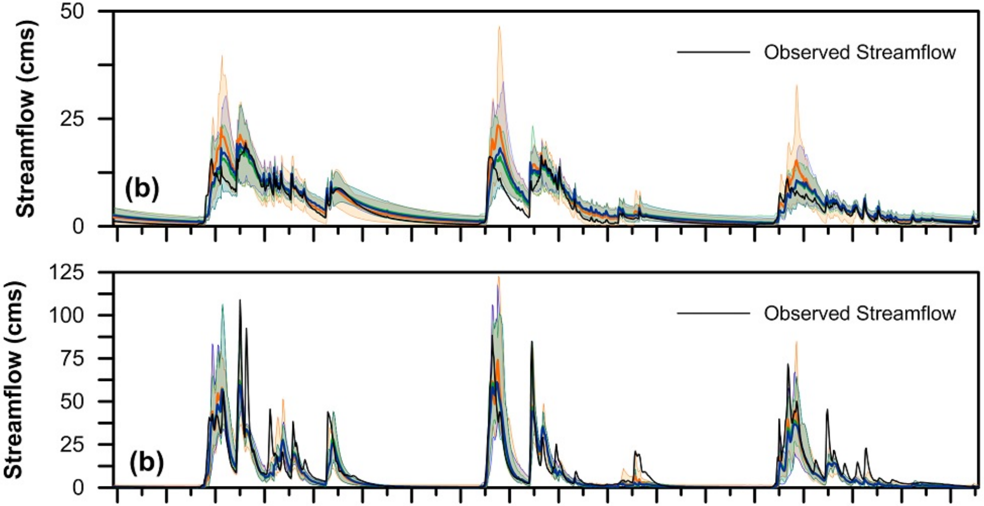

Fig. 5. Fig 3 & 4 panel b revised for manuscript

Printer-friendly version

Discussion paper