RC3: 'Comment on hess-2023-308', Anonymous Referee #3, 06 Feb 2024

This is a well-written, comprehensive study that applies a novel evaluation framework using isotopes to diagnose potential error sources in hydrologic models. The study offers useful analyses and conclusions for hydrologic model evaluation, with a focus on processes, which are supported by multiple lines of evidence. I believe the study merits publication, but I have several main comments that I offer for the authors to consider:

Thank you for your helpful review!

- **Isotope tracer expertise is assumed**. I suspect that many HESS readers are not experts in isotope tracers, and there is a lot of assumed expertise and jargon. The manuscript would benefit from more explanations up front and throughout so that readers unfamiliar can benefit from this novel approach. I point out examples in my specific comments.

  Based on your suggestions, I've made substantial changes to the way I introduce and describe isotope systematics in the paper, with an eye to how a non-expert may read the work. I hope these changes are effective and improve reader clarity and interest.

- **The main conceptual method is hard to pick out from all the details**. The authors do an excellent job of explaining their methods in great detail, but I found myself missing the forest for the trees in my first read through. Having never read a tracer study, the main approach – comparing the observed isotopes with the NWM derived estimates (which comes from both NWM fluxes AND gridded datasets) was hard to follow. I think this should be clarified earlier, and I offer a few suggestions to improve the Method section organization in my minor comments. For example, Figure 1 is very detailed, but hard to follow all that's going on at the outset of the Methods, so I suggest adding a general conceptual overview and/or simple flow chart to guide the reader (including what's in Equation 1, otherwise it appears much too late for the reader to follow what's happening). I recognize that this is a subjective suggestion, but I think it would help to increase the reach of your paper.

  Thank you for this feedback. Based on your suggestions, we have attempted to clarify the approach in the abstract, intro, and methods. Based on your comments, it seems that much of the confusion can be cleared up by emphasizing the observation-model comparison nature of the study first in all places, and following up with the specifics around producing the 'modeled' river isotope data. This emphasis has been applied throughout and is described in response to specific comments.

- **The key results are hard to pick out from the supporting results**. I appreciate the comprehensive results and multiple lines of evidence presented, but will caution the authors that it can make it difficult for readers to focus on the key results on which the main conclusions are based. The authors may want to review all the results and see if there are any that they might like to include in the Supplemental (I make a few suggestions). I recognize that this is a subjective suggestion, and that there is a tradeoff to including too few versus too many results, but I think a slightly more curated results section would help to increase the reach of your paper. (Also, I will note that the authors do a nice job of summarizing the key results in the Abstract and Conclusions, so this is just a suggestion for the Results themselves).

  I am hopeful that our re-arrangement of the presentation of the latter part of the results helps with this challenge in extracting the main results from the large amount of information presented. At your suggestion, I removed Section 3.3, which covered interannual variability to the supplemental. I then swapped Section 3.4 and 3.5 so that the spatial analysis comes first, and the seasonal/growing season evolution section comes second and can used as support for the conclusions of Section 3.5 (now 3.3). Hopefully this highlights the main objective of the paper, which is about agricultural return flows, rather than assessing all modes of variability in the dataset.

Specific comments:

Abstract:

Line 3: Is “fidelity” the right work here?

The Merriam-Webster dictionary defines fidelity as “the quality or state of being faithful or loyal”, “exactness in details”, and “the degree to which an electronic device (as a record player, radio, or television) correctly reproduces its effect (as sound or a picture)”. The word is present in reference to models in a similar way in which we’ve used it in such papers as:

- “Fidelity of WRF model in simulating heat wave events over India” (https://www.nature.com/articles/s41598-024-52541-2)
- “Advancing Process Representation in Hydrological Models: Integrating New Concepts, Knowledge, and Data” (see first line of abstract, https://doi.org/10.1029/2021WR030661)

- "Learning from hydrological models' challenges: A case study from the Nelson basin model intercomparison project" (see abstract, https://doi.org/10.1016/j.jhydrol.2023.129820)

I believe our usage is consistent with other descriptions of model performance elsewhere in the literature, so will leave it as is.

Line 4: In parenthesis, I am not familiar with the delta 18 Oxygen and delta 2 hydrogen notation. Is there a way to define them here for non-experts? Maybe just remove the parenthetical all together from this sentence? You may need to define them later (line 8, etc).

I recognize that the water isotope notation was not introduced yet when these were referenced. The idea here is that skipping the notation may be ok in the abstract, as long as the full definition is produced as soon as the information appears in the introduction. This is typically done (it occurs in other water isotope-focused and tracer-focused publications in general journals) because the abstract is too short to allow for a full definition, yet the symbology is included to stimulate reader recognition and clarity about which isotope systems will be discussed. I changed the initial parenthetical statement slightly to indicate the isotope ratios we are using and mention that they are expressed in delta notation.

Line 4-7: I had a hard time understanding generally what you did from these sentences until I read the manuscript and thoroughly studied Figure 1. Here's a suggestion of what might help a reader like me (at least in terms of laying out the general conceptual framework of what was done – please fix details if I have them wrong):

"In this study, we compare observational river isotope data with estimates of river isotopes derived from the NWM. The evaluation is done in 5 basins in the western US in summer from 2000 and 2019. In terms of observations, we use 4503 in-stream water isotope observations in 877 reaches. In terms of the corresponding estimates of river isotopes, these are calculate using a mass balance equation based on NWM-fluxes and estimates of isotope ratios from long term mean gridded precipitation and groundwater datasets."

I have made some changes to the abstract wording that are similar to those suggested.

Introduction

Line 18-26. This paragraph is quite general in scope and I don't think is needed. I think you could delete and start with paragraph 2 (i.e., line 27), using something like the first line of your abstract, i.e., : "Hydrologic models, such as the National Oceanic and

Atmospheric Administration's National Water Model (NWM), provides critical analyses and predictions of streamflow that support water management decisions. The NWM is an application of the WRF-Hydro model (Gochis et al., 2018), and is fully routed with high spatial and temporal resolution, providing short and medium term streamflow … " etc.

The initial paragraph lays out some important motivation for why accurate streamflow estimates are important, and why pursuing improving accuracy of operational hydrologic models is crucial. I plan to retain the paragraph.

Line 27: The operational NWM is based on the WRF-Hydro model (not its data); the NWM is an application of the WRF-Hydro model. Perhaps you are confusing WRF-Hydro (a hydro model) with WRF (a meteorological model). Suggest saying, "…like the National Oceanic and Atmospheric Administration's National Water Model (NWM) which is an application of the WRF-Hydro model (Gochis et al., 2018)" or you could say "is based on the WRF-Hydro model (Gochis et al., 2018)."

Thank you for this correction. We have updated the text.

Line 34: "fidelity" does not seem like the right word here.

As above, I believe our usage is consistent with other descriptions of model performance elsewhere in the literature, so will leave it as is.

Line 84: Sentence that starts with "tracers" – can you add a general sentence on tracers and isotopes (before getting into the 16O, etc), for the non-expert? When you get to the parenthetical 16 oxygen, etc, please define these. I start catching your drift a bit later when you define 2H/18O as "heavy" and 1H and 16O as "light", though as a non-expert I'm not sure to what it is compared.

This paragraph has been updated, including additional clarity around the concept of tracers, with general examples of tracer studies.

Line 89: The expression: "The secondary parameter, deuterium excess.." is not clear to a non-expert.

I have expanded this section and hopefully with the changes to the whole section, is clearer to a non-expert reader.

This paragraph (lines 84-92) is very dense, especially for a non-expert (of isotopes), try rereading as someone who doesn't know about isotopes and generalizing a bit as possible, to help the reader understand this powerful evaluation tool.

I have made substantial changes to this paragraph to help lead non-experts through the basics of isotope systematics, while have also made an effort to keep the paragraph condensed for clarity and readability of the entirety of the introduction. Hopefully the clarifications help.

Line 93-94. What is isotopic fractionation? Can you say this another way for the non-expert?

I have updated this section to add a cursory definition of fractionation.

Line 110: What is evopoconcentrated and evaporative enrichment?

I have updated these to more general descriptive language.

Line 115-116: Same as in the Abstract, I wasn't really sure what you did until I read through and studied Figure 1, although this is easier for me to understand than the Abstract. Could you start most generally, saying: "In this study, we compared stream water isotope observations with estimates of water isotopes derived from an isotope mass balance. The isotope mass balance is from xyz NWM and lmnop gridded long term, etc".

I flipped the first and second sentences to highlight the comparison aspect over the mass balance aspect, as you suggested, and made minor editorial changes to the sentences for clarity.

Methods

127-128. Same as previous comments, I initially had trouble understanding what was done here. Even though I like Figure 1, it is very dense. I'm wondering if you could have a very simple conceptual figure first to ground the reader before Figure 1, where you just show the 3 main pieces: (1) Direct obs, (2a) NWM, (2b) gridded ratios, as well as equation 1 and where the pieces fit in on the model side. This could be in a section called "2.1 Conceptual Framework", and could include the general figure, which would introduce sections 2.4 and 2.5, and it would absorb the current specs listed in "2.1 Temporal domain", "2.2 Spatial…" and "2.3 Data assim". Then Figure 1, with all the details, could come later.

My intent was for the first paragraph of the methods to function as the 'conceptual framework'. Following other suggestions by the reviewer, I altered the text in this initial paragraph to highlight the comparison part of the work first, and follow with the description of the modeling approach. I also simplified some of the methods so they might be more clear, allowing a reader to absorb the conceptual framework before

diving into the methods. I think an addition, simplified conceptual framework figure is not necessary with the improvements to the methods section and text throughout.

Figure 1: This is a very nice figure, but dense… See previous comments. One note: For NWM data feeding into Equation 1: maybe have the notation (Fgw) and (Fro) and for obs data have (Rgw) and (Rro) there to link with Equation 1. This might be better suited to a more general conceptual figure though (see previous comment).

We have updated this figure to include that notation to help link the conceptual figure to the equation and the overall methodological concept.

2.3. Data assimilation: This seems minor to be a full section, and I wasn't sure what this was related to. Can this be part of 2.2 Spatial domain or just Supplemental? Or if you decide to have a Conceptual Framework section, it could be absorbed in that.

I absorbed this section into the Spatial Domain section, as suggested, and modified the text for simplicity and clarity.

Line 160. Correction, the NWM is based on the WRF-Hydro model – which is an open source, community hydrologic model, it is not based on inputs from it (I think you might be confusing WRF-Hydro with WRF, where WRF would provide inputs): "The operational hydrologic model is based on the open-source, community hydrologic model, WRF-Hydro (Gochis et al., 2020b, a)…"

Thanks for that clarification. Text has been updated.

Line 170: Do you mean Figure 1 here?

Yes, that's correct. Thank you for the catch.

Equation 1: Seeing this equation helped me to see how the pieces fit together. If you decide to include a conceptual model, I suggest having this equation in it to see how each piece fits in (you could do it generally, for just one reach, as a demonstration, so it was a simpler equation without the subscripts).

Thank you for this suggestion and insight.

Figure 2 – I like this figure and how it showed the way to interpret. I often had to look back at this figure to interpret later results.

Thank you. That was the intent of the figure – to help with interpretation of the results.

Line 371: What is a meteoric water line? What is a surface water line?

A description of WLs is now in the introduction, which should help clear up this section.

Line 380: Is Table 1 needed for the main text or could it go into the supplement?

We will retain it in the main text.

Line 386. What is an isotopologue?

Isotopologue has now been introduced in the introduction, which hopefully clears up this comment.

Line 405: "The strongest signal in our data is that of evaporation, evidenced by combinations of positive δ18Odiff and negative ddiff in arid regions." <- This is an important conclusion, but there are so many results it's hard to quickly see what evidence this is from – I had to really go back and study all the figures and tables to realize I needed to imagine all the points from Figure 4 as if they were on Figure 2. Can you add something to that effect to guide the reader? Or add the colored quadrants to remind the reader? Or maybe just say in the caption of Figure 4, "see Figure 2 for what the different locations on the x- y- axis mean"?

I added a sentence to the caption of Figure 4 that refers the reader to Figure 2 for interpretation of the scatter plot. I think adding annotation to the plot would make it too busy and would make Figure 2 redundant.

Table 2: Is this need for the main text or could it go into the supplement?

We will retain it in the main text.

Section 3.3. This section and Figure 5 did not seem particularly important to the results/conclusions, and one suggestion would be to put it in the Supplemental (so that the other key results are less buried). If so, you could just have one sentence at the end of the previous section saying something like "There was little interannual variability, which we interpret to mean there was pervasive presence of eval… etc.. see Supplemental xxx".

I took your advice and put this section and figure into the supplemental data. I also moved Section 3.4, which was concerned with temporal evolution of stream waters over the growing season in different basins and used that as support for Section 3.5, which is the primary result of the manuscript. I updated the transition paragraph to

reference the interannual variability analysis in the SI as a basis for the robustness of the spatial analysis in the context of the dataset.