RC1: 'Comment on hess-2023-308', Anonymous Referee #1, 25 Jan 2024

This paper presents an interesting work on the diagnose of NWM model with the aid of isotope data. The isotope simulation identifies different bias characteristics in high and low elevation regions. The bias of isotope simulation in low regions is attributed to the contributions of irrigation return flows, and some statistical analysis and literature reviews are conducted to support this hypothesis. Overall, the logic is clear, the analysis is solid, and the written is good, making this paper worth publishing in HESS. However, I would like to point out some major concerns related to the isotope dataset and mass balance calculation.

Thank you. We appreciate your comments and ideas.

1. The calculation of surface water isotope: From a hydrological viewpoint, the calculation surface water isotope ratio (equation 1) is confusing. The authors determine the isotope ratio through dividing the summed isotope fluxes by the summed runoff and groundwater fluxes. However, the term "runoff" usually refers to the sum of surface runoff and subsurface runoff. I don't know whether the "runoff" provided by NWM model refers to surface runoff or the total runoff. According to the equation, it seems that the runoff is actually only surface runoff. If this is the case, I suggest the authors to make it clear in the main text. Otherwise, the isotope ratio of surface water should be [Rgw*Fgw+Rp*(Fro-Fgw)]/Fro.

Thanks for this clarifying point. You're right – runoff in this case is only surface runoff, as it is defined and calculated that way by the NWM (variable is 'runoff from terrain routing', qSfcLatRunoff, m3 s-1). The model also includes runoff from bottom of soil to bucket (qBtmVertRunoff, m3) and flux from gw bucket (qBucket, m3 s-1). See https://water.noaa.gov/about/output_file_contents for output file contents.

In the text I retained the more general definition of runoff in the introduction since it's more generally applicable but called out the potential for runoff to be from the surface or subsurface. In the methods section, specifically where I first introduce the NWM variables we use, I have added clarification to the definitions of the NWM variables. Later on, when the mass balance calculation is introduced, I clarified that the runoff is surface runoff. Throughout the results and discussion, I've updated 'runoff' to 'surface runoff' to help retain the distinction in the minds of the readers.

2. Choice of isotope dataset:

The authors adopted the monthly long term average precipitation and groundwater isotope data as the input data. This is okay for groundwater because its isotope composition is rather stable. However, the precipitation isotope usually has very strong temporal variation, especially during wet season. Given that a high-resolution dataset of hydrological fluxes produced by NWM was adopted, it might be better to use a high-resolution precipitation isotope dataset, such as the output of isotope enabled general circulation models (iGCM, such as https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2008JD010074). The reliability of the precipitation and groundwater isotope dataset itself also need to be evaluated, at least citing some descriptions about their accuracies in published papers. Otherwise, it would be hard to determine whether the simulation biases come from the NWM data or the isotope input data.

Besides, I find that the authors evaluated the isotope simulation performance by comparing the long term simulated results with the measurement surface water isotope sampled at a specific time. Such kind of measurement data would be highly dependent on the specific precipitation events and the corresponding isotope composition before sampling. So it might be more reasonable to compare the measurement data with the corresponding simulated result at the sampling time. We can observe a smaller range of simulated isotope compared to measurement in Figure 3. I think the aggregation of precipitation input and simulated isotope could be the reason of this.

Nonetheless, I understand that replacing the input data and repeating the whole calculation process is challenging. If this is difficult to achieve, please consider addressing these issues in a discussion section.

The authors agree that the differences in the time integrations represented by the model data and the observations are a limitation of this study, and also agree that future studies should consider higher temporal variability in the precipitation input value to help address temporal variability in observations arising from recent precipitation inputs to rivers.

Unfortunately, it is outside the capacity of the authors to re-do the study at a finer temporal scale at this point. However, we have plans to address this issue in future regional studies and fully agree that a more accurate treatment of the temporal component of the issue is a critical part of pushing this kind of study forward and improving our ability to evaluate water models using tracers.

To that end, thank you for the reference to the iGCM data – this dataset, or another precipitation isotope-reanalysis approach (maybe with WRF+NADP datasets), or a statistical approach like Finkebeiner et al., 2021 (https://doi.org/10.1175/JHM-D-20-0142.1) could be an excellent method to deploy during a second stage (higher resolution, smaller region studies) of this project. We will take these ideas under consideration as we propose and develop continuations of this project.

Nonetheless, we have taken your suggestion to highlight the accuracies of the isoscapes we did use in the analysis. Considerations have been included in the methods section. Likewise, we have adjusted in the presentation of the evaluation of variability in the observation-model differences, choosing to highlight the strength of our dataset – spatial variability - and present the evaluation of interannual variability as insurance that the spatial variability doesn't co-vary with temporal variability due to interannual variability in sampling patterns. Hopefully this helps clarify for readers the strengths and weaknesses of this specific approach.

Specific issues:

- L142: Provide the full name of HUC2 at the place it appears for the first time.

  Added.

- L208~L215: How were the 10 random draws generated, and how were they used in specify to evaluate uncertainties? There seems to be results related to uncertainty in the result section.

  Both the groundwater and precipitation 'isoscapes' have uncertainty layers available. The uncertainty estimates, alongside the mean estimates from these products were leveraged for the uncertainty analysis. The random draws assumed a normal distribution of variance around the mean estimate. I have reworded this section, hopefully clarifying our approach.

- L400: Figure S8 should be S10?

  Thank you, good catch. The SI figures were in the wrong order so the reference in the main text has remained the same, but the order of SI figures has been updated.

- L442-443: This is a strong statement and is the basic of following analysis. Consider providing more explanation on it.

  We have added a few sentences of explanation to help readers follow our logic.

- L449: There is an additional letter "d"

  Fixed. Also fixed the superscripting of 18 in the same line.

- Please provide the r2 and p values in the scatter figures such as Figure 8 and 10

  Adding an r2 and p value isn't appropriate for Figures 8 and 10 because they aren't intended to show a linear relationship between two variables. They are intended to show non-linear relationships between multiple variables (i.e., d_{diff} to elevation, stream order, aridity/climate and d_{diff} to elevation, and water use characteristics of upstream area). Adding the r2 and p values are likely to confuse readers as to the focus of the figures. The linear regression statistics for other scatterplots (i.e., Figure 3, Figures S4-7) are available from tables in the text, which are now referenced in the figure captions. I also updated Table 1 to include the R2 of the regressions.

- Table 3: There is an additional symbol "+" in the last row

  Thank you, the extra + sign has been removed.