**Response to reviewer's comments**
***Manuscript number (hess-2023-25)***

**Note: Below are the reviewer's comments in black, and our responses in blue.**

## Referee#1 comments:

The scientific importance of the article is high. Congratulations to the authors! They covered a very interesting and important topic. As a novelty, they tried to link ecosystem services derived from integrated hydrological model results to monetary evaluation.

Grammatical and technical errors are not typical, the article is of high quality from a stylistic point of view. The written part of the publication is fine.

Thank you for taking the time to review our paper and for your positive evaluation! We are glad to hear that you found our topic interesting and important. We appreciate your acknowledgment of the scientific significance of our work, and your compliments on the quality of our writing. This feedback is very valuable to us as we continue to refine our research. Thank you again for your review.

I feel it necessary to place some supporting literature references in some places. I marked them in the attached document.

There are parts in the description of the modeling work that are not completely understandable, and it is essential to clarify them. I marked them in the attached document.

After the clarifications and suggested references have been replaced, the article can definitely be recommended for publication.

Thank you for your feedback. We will review the manuscript thoroughly and make the necessary changes based on your suggestions. We will also add the suggested literature references. Furthermore, we will revise the model description section to make it more comprehensive, with inclusion of additional detail and examples.

Line 18: You should define what green water means in your article. It can be a bit confusing in this form.

Thank you for your valuable comment. We apologize for any confusion caused by not providing a clear definition of the term "green water". We appreciate your input and the revised manuscript will include a clear definition of the term.

Line 67: Maybe you should mention the most simplest approaches like matrix models as well.

Thank you for your suggestion. In line with your suggestion, we will revise the text to include a mention of matrix models, as you correctly pointed out that these represent one of the simplest approaches for modeling complex systems.

Lines 70-77: You should emphasize the uncertainties of these tools from a hydrologic point of view. There are studies that highlighted their limitations.

We have taken your suggestion into consideration and in the revised manuscript we will include a discussion on the limitations and uncertainties associated with hydrologic models.

Line 86-89: Perhaps the best support for the weakness and unreliability of these models is when they yielded the same results as simple matrix models (without any hydrological calculations). Maybe this article raises your interest: https://doi.org/10.1016/j.ecolind.2022.109143
Maybe you should refer to the model.
This one seems to be appropriate:
https://doi.org/10.1111/j.1745-6584.2011.00882.x

or this:
https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=HydroGeoSphere:+A+Three-Dimensional+Numerical+Model+Describing+Fully-Integrated+Subsurface+and+Surface+Flow+and+Solute+Transport&btnG=

Are there other models which would be capable to handle hydrology similar? E.g.: MIKE SHE..
You may pay attention for this:
https://www.frontiersin.org/articles/10.3389/feart.2021.721009/full#B3

Thank you very much for your valuable suggestions and for providing us with the links to the relevant papers. We have thoroughly reviewed the articles you recommended. We will take into account the information contained in these articles and incorporate it into the revised manuscript. We will also more clearly emphasize our rational for using fully-integrated groundwater – surface water models, in the sense that our research objectives necessitate that groundwater, and groundwater – surface water interactions are quantified, which is beyond the capability of simple matrix models. We will also explicitly mention and cite other common integrated models (i.e., Parflow and MikeSHE).

Line 120-122: Maybe you should refer to- or describe in a nutshell the definition Strahler order to help readers from other disciplines (economy, ecology).

Thank you for the suggestion. We will add a brief description of the Strahler order in the revised version of the manuscript.

Line 153-163: This paragraph should be supported with some reference related to the topic of applicable models on different spatial scales.

Thank you for this suggestion. We will include the relevant references in the revised text.

Line 192: Did you carry out some kin of harmonization on input spatial data regarding to their resolution?

In the revised manuscript we will describe that land surface and subsurface hydraulic properties were mapped into the HGS model's unstructured FEM using a dominant component approach, meaning that if two or more property classes exist within the input dataset for a single element, the majority class is represented.

Line 217-219: How do these stations operate? At what intervals is data recorded?

In the revised manuscript we will clearly note that the surface water flow monitoring stations provide daily temporal resolution, while the groundwater monitoring network data provide hourly temporal resolution. We will also embed the URL for the monitoring data sources in the text.

Line 220-221: Why did you use other metrics to evaluate groundwater performance?
Why did you used this one? Maybe you should take into account other statistical evaluation tools. In the results section we can see that, the coefficient of determination is almost perfect, but the difference between the mean GWLs are significant. Maybe you should bring in the RMSE as well.

We agree that it would be valuable to incorporate additional statistical evaluation tools in our analysis of groundwater performance and will include RMSE values for the simulated vs observed groundwater levels in the revised manuscript.

Line 266-267: Good accuracy, according to what? You should cite a reference.

In the revised manuscript, we will provide the reference (Moriasi et al., 2007) to support our model performance interpretation. We will also reword the text in this section so that it is less subjective.

Line 268-271: What was the temporal resolution of the compared data (daily, weekly, monthly, yearly)? What does the R2 value refer to? The large difference between the observed and modeled average groundwater level depth is worrisome. Especially knowing that in L107-L109 you wrote about a shallow GW depth of 1-3 m. This can also significantly affect the modeled actual evapotranspiration values.

Regarding the temporal resolution of the compared data, the observed groundwater level data was collected at a hourly resolution from the nine WSC hydrometric stations across the SNW; however, the hourly data was aggregated to daily average prior to being used for model performance evaluation. Both groundwater and surface water simulation performance was evaluated based on daily temporal frequency. In the revised manuscript we will rewrite this section to be more clear in regards to the temporal frequency.

The $R^2$ value refers to the proportion of the variance in the observed groundwater level that is predicted from the simulated groundwater level. A high $R^2$ value indicates a good fit between the observed and simulated data.

We acknowledge your concern regarding the large difference between the observed and modeled average groundwater level depth. However, we would actually regard the average difference of 2.8 m between simulated and observed groundwater levels to be very good for two primary reasons. Firstly, the model covers 3830 $km^2$ and has element edge lengths that vary from ~100 m to 300 m, hence subtle variabilities in local topography (from which groundwater depths are referenced) are not perfectly captured in the model geometry. Secondly, because groundwater extractions were not represented in the model, simulated groundwater levels are biased higher, and this bias will be most pronounced in groundwater production areas, where the monitoring wells tend to be placed.

In the revised manuscript we will include supplemental material that graphically depicts surface water and groundwater simulation performance.

Line 366-368: What about the limitations of the fine-grained models? Are they applicable anywhere with any spatial scale? Data needs, other requirements (resource, financial, expert, so on).

We appreciate your input and acknowledge the importance of discussing the limitations of fine-grained models in our paper. In the revised manuscript we will include a section that outlines the limitations of fine-grained models along with their data requirements, expert knowledge, and computational requirements. We will also highlight the applicability of these models across different spatial scales and discuss their potential limitations in certain contexts.

Finally, we thank the reviewer for dedicating their time to reviewing our manuscripts. Their valuable suggestions and feedback have greatly contributed to enhancing the quality of our research work.