

Dear Authors,

The report from the second round of the review by an expert shows that your manuscript requires more improvements. Actually the reviewer strongly recommended that the manuscript should be rejected. However, based on my own assessment, I recommend reconsideration of your manuscript after major revisions. I request you to carefully address each of the comments generated from the second round of review of your manuscript.

Dear Prof. Charles Onyutha,

We sincerely thank you and the reviewer for the time and effort devoted to the second round of review. We appreciate your careful Editorial assessment and the opportunity to resubmit our manuscript after major revisions.

We note that the comments focus primarily on the presentation of the results, particularly the clarity and quality of the figures. In response, we made considerable efforts to improve the presentation of the manuscript, especially the figures, following the reviewer's suggestions. In retrospect, we acknowledge that some figures would benefit from further refinement, and we appreciate the reviewer's emphasis on this aspect. We have therefore carefully and thoroughly addressed all comments raised in the second-round review. Where some requests regarding figure improvements were not entirely specific, we have made our best efforts to interpret and implement them, and we remain open to making further adjustments if further guidance is provided.

The major revisions that we have undertaken include:

- Substantial improvements in figure quality, including reconstruction or revision of Figures 1–7 and the addition of new Figures S1–S4.
- Minor restructuring of the Methods section, where closely related subsections were merged to improve clarity and coherence.
- Clearer discussion of uncertainties associated with remote sensing and reanalysis products, as well as a more explicit explanation of the rationale and benefits of using these products for model validation.

A detailed, point-by-point response to all reviewer comments is provided below, with corresponding changes clearly indicated in the revised manuscript.

We hope that, with these revisions, the manuscript now meets the standards of *Geoscientific Model Development*, and we would be grateful for your further consideration.

Yours sincerely,

Songjun Wu, on behalf of all co-authors

Reviewer #1

The authors have responded part of my concern, while significantly revisions are required but it seems lacking. First, as I mentioned several times in previous rounds, the figure quality is low and should be improved in substantial way. The research outcome should be both amiable for researchers but also for general readers. The revisions on figures significantly lack.

**** We thank reviewers for the feedback. We acknowledge that figure quality is critical for both scientific communication and accessibility to a broader audience. In response to this comment, we have substantially revised all figures in the manuscript.**

In addition, we have made minor adjustments to the manuscript structure to enhance clarity, and we have expanded the discussion of uncertainties associated with remote sensing and reanalysis products. A detailed, point-by-point response describing these changes is provided below.

For example, Figure 1: why do authors change the colormap to black and white, and some important land marks are removed? No panel numbers hinder the readability of the manuscript.

**** Thank you for the comments for Figure 1.**

First, we would like to clarify the reason for updating color codes in Panel (a). We used black and white because some previous colors were identical to the ones used in Panel (b) and (c). However, color codes in Panel (a) represent elevation, while those in subplot (b) and (c) are used to differentiate hydrological, isotopic, and nitrogen processes. Mixing those colors across subplots would result in ambiguity. Therefore, we retain the color codes (black and white) for Panel (a).

Then, regarding the removed landmarks in Panel (a), they are the monitoring sites of surface water and groundwater in this catchment (Demnitz Mill Creek). However, the motivation of presenting this catchment in subplot (a) is to show how the spatial discretization of watersheds is realized in EcoTWIN, while the monitoring sites are irrelevant to this topic. However, we realize the lack of scale bars. They are added to Panel (a).

Lastly, we agree that the lack of panel numbers hinders the readability, and they have been added. The figure caption has also been updated accordingly.

Figure 2: No river inputs marks have been added, which is important for the hydrological

model; also no panel names/numbers;

** We follow the suggestion to add river networks into Figure 2. The panel names and numbers have also been added with figure caption updated accordingly.

Figure 3: it seems the revised manuscript 'with track of changes' is not consistent with that 'no track change' version, at least for the subtitle.

** Thank you for the comment. We have carefully re-checked Figure 3. The subtitle is now consistent between the tracked-changes version and the clean (no tracked changes) version of the revised manuscript.

Figure 4-7: the figures are very hard to identify, e.g., time series etc.

** Thank you for the comment. We agree that the time series in subplots are hard to identify due to their limited size. Therefore, major revisions have been implemented for Figure 4-7.

Figure 4 to Figure S1: The original Figure 4 was removed due to its complexity and unidentifiability of the time series. The simulated and observed time series were moved to supplementary material (Figure S1). Due to the relatively large number of sampling sites included in model calibration, we only show poorer simulations with $KGE < 0.2$ for clarity and diagnostic purposes.

Figure 5-7 to Figure 4 and Figure S2-4: The original Figure 5-7 were removed due to the limited size of time series. The most important information – the distributed grid-to-grid validation of three internal states/fluxes – was merged into new Figure 4. The time series between simulation and remote sensing/reanalysis products were moved to Figure S2-4 for diagnostic purposes.

Relabeling of subsequent figures: The original Figures 8–10 have been relabeled as Figures 5–7, and the original Figures S1–S3 have been relabeled as Figures S5–S7 to reflect the changes described above.

The structures of this work should be carefully adjusted: e.g., '3.2 Model calibration' and '3.4 Calibration performance' should be merged, and '3.3 Model validation' and '3.5 Validation performance' should be merged.

** We agree that merging the calibration-related sections and the validation-related sections improves the readability and logical flow of the manuscript. Accordingly, Sections 3.2 and 3.4

have been merged, as well as Sections 3.3 and 3.5. The associated text has been slightly revised to ensure a smooth and coherent integration of the merged sections.

Table 2: the statistic terms and values should be written in a formal way. For example, it is 'Pearson correlation coefficient' NOT 'Correlation', RMSE should have units.

** Thank you for the comment. We have changed all abbreviations in Table 2 to formal statistic terms. The unit of each statistic was also added as a new column.

The other important part should be improved is the fact that ERA5 is the reanalyzed data product, not the real observations, it should be carefully considered that whether it can be used as a standard/reference to examine/decide that the model result is accurate and reliable.

** Yes, we agree that as reanalysis products, ERA5 snow depth is uncertain. Similarly, MODIS evapotranspiration and GRACE water storage also have uncertainty due to the used model/algorithm, data quality, and spatial/temporal resolution. This has been acknowledged in section 3.3 Model validation (see *italic text* below). We also avoid using the term "observation" for any of the remote sensing/reanalysis products.

However, remote sensing and reanalysis products have long been for validation of hydrological models. The good agreement between simulations and remote sensing records can validate the spatial and temporal patterns of model simulations, although it does not necessarily imply accurate representation of absolute magnitudes. This has also been clarified in section 3.3.

Note that all three products may contain considerable uncertainties. ERA5 is a reanalysis product that combines historical observations into global estimates using modelling and data assimilation approaches, therefore inevitably embeds uncertainties associated with model structure and observational coverage (Hersbach et al., 2020). MODIS evapotranspiration is derived from remotely sensed spectral information, energy partitioning approaches and the Penman–Monteith framework, whose uncertainty may exceed 30% depending on spatial scale and environmental conditions (Mu et al., 2011). GRACE infers changes in terrestrially stored water masses from spatial and temporal variations in the Earth's gravity field; however, its coarse spatial resolution can introduce substantial uncertainty when used for hydrological validation, particularly at basin or sub-basin scales (Tapley et al., 2004). Nevertheless, good agreement between simulations and remote sensing or reanalysis products can enhance confidence in the robustness of simulated spatial and temporal patterns, although it does not necessarily imply accurate representation of absolute magnitudes.