**Manuscript ID** HYDROL61841

**Title** A Novel Remote Sensing-Based Calibration and Validation Method for Distributed Hydrological Modelling in Ungauged Basins

*Journal of Hydrology*

Dear Editor

We are submitting our revised manuscriptentitled “A Novel Remote Sensing-Based Calibration and Validation Method for Distributed Hydrological Modelling in Ungauged Basins” for possible publication in *Journal of Hydrology*.

Following the previous review and your suggestions, we have made a thorough revision to enhance quality of our manuscript. All these comments have been carefully considered and incorporated into the revised version. All revisions to the manuscript were marked up using the “Track Changes” function using MS Word. Below is a summary of the major changes made to the manuscript:

1) We have revised Fig. 2, Fig. 5, and Fig. 8 according to the reviewers' comments and added labels for the sub-basins.

2)We have supplemented the analysis of the reliability of several ET and precipitation products, which has been incorporated into the framework. The results are discussed in Section 5.1, Uncertainty of RS Data, while the related figures and tables are included in the supplementary information.

3)Based on the reviewers' suggestions, we have revised many unclear and ambiguous statements in the manuscript and removed redundant information that does not contribute to the key findings.

4)We have focused on revising Sections 4.3 and 3.3.3 to make the statements clearer and easier to understand.

Thank you for your consideration. I look forward to hearing from you.

Yours Sincerely,

Dr. 通讯作者名字

**Comments in blue and our response in black**

1. **Reviewer #1:**

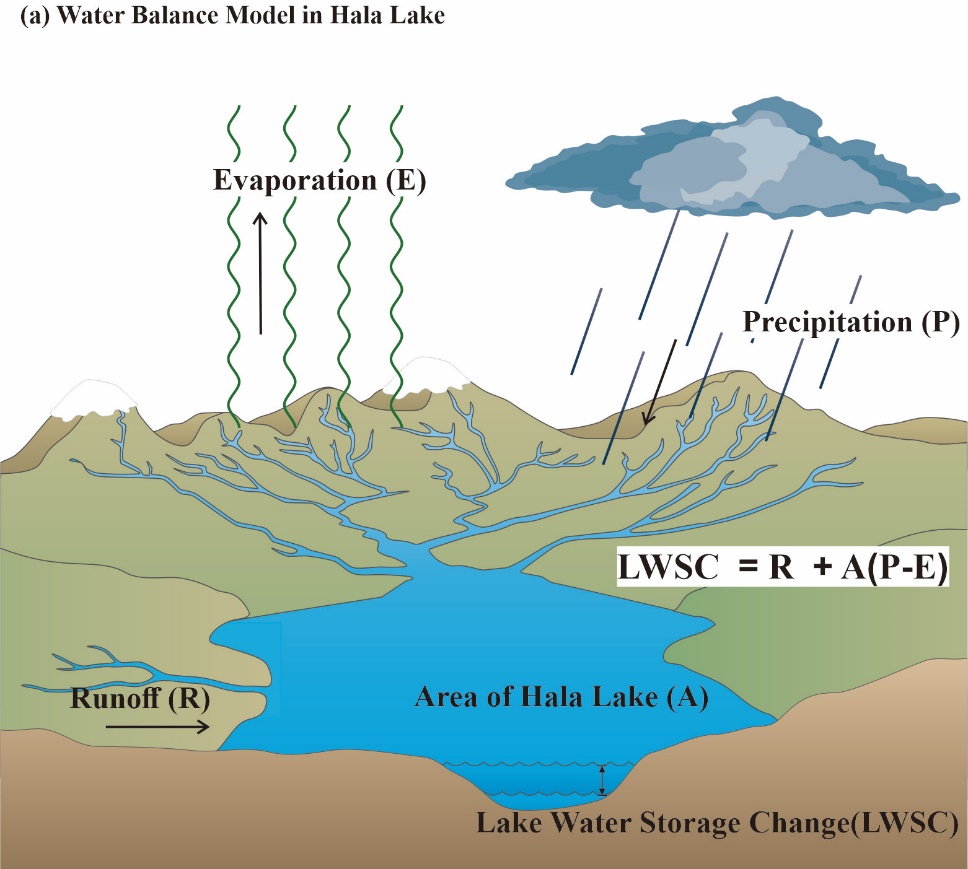
***General comments***

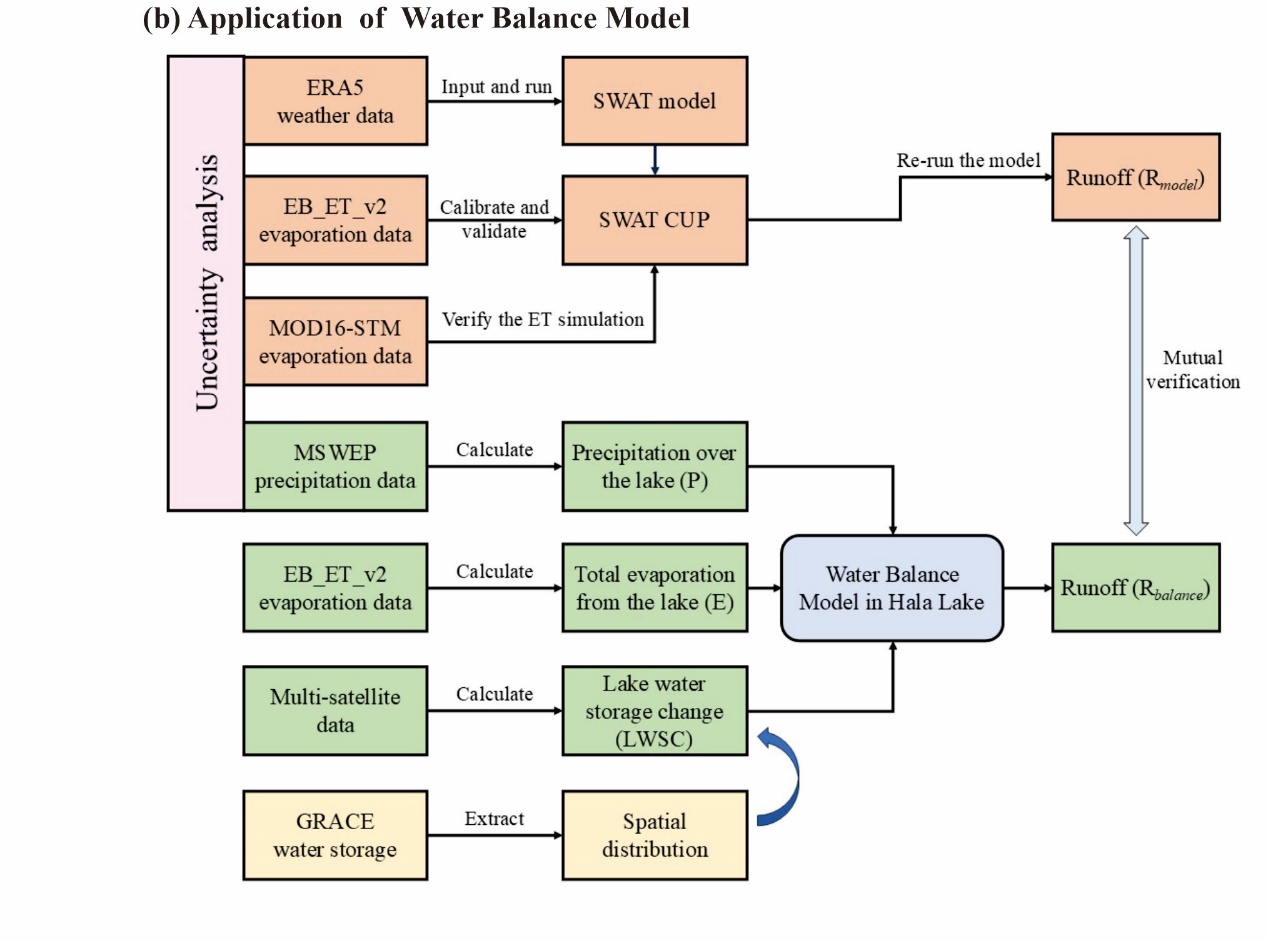
1. *The framework of the complementary validation method presented in Figure 2 is almost impossible to understand. What are R\_s and R\_t? What are LWAC and LWAC\_s? What is the difference between (c) and (d)? This figure is too complicated with too many acronyms. I would suggest presenting the framework as direct and indirect uses of remote-sensing products for hydrological model calibration and validation, using ET products for direct calibration (or hard calibration) and lake level and lake surface area products for indirect calibration (or soft calibration). It should be noted that the indirect use of remote-sensing products through lake water balance calculation only works for a closed basin when the basin is ungauged because, in an open system, the downstream flow wouldn't be available for water balance calculation.*

Response:

Thank you for your valuable suggestions. We acknowledge that Fig. 2 was not sufficiently clear and concise, and we have redrawn it accordingly. In the original Fig. 2, R\_s represented the runoff simulated by the SWAT model after parameter calibration, R\_t denoted the runoff calculated through the water balance equation, while LWAC and LWAC\_s referred to the lake water storage changes obtained directly from satellite data and calculated via the water balance equation, respectively. In the revised Fig. 2 (Fig. R1), we have adopted a more streamlined framework, directly indicating the application of data and models. Additionally, we have incorporated more intuitive graphics to illustrate the establishment of the water balance model.

Thank you for your reminder regarding the characteristics of our study area. Since it is a closed lake, we utilized the water balance calculation method in our analysis.





**Fig. R1.** The framework of the complementary validation methods.

1. *I failed to understand the rationale behind comparing three RS ET products after calibrating and validating SWAT-simulated ET with one RS ET product. Why not use all three ET products to calibrate/validate the SWAT model separately, and then compare the SWAT's runoff simulations resulting from the calibrations using different RS ET products?*

Response:

Thank you for pointing this issue, which is an important question. We now carefully explain this question in the paper (Line xxx). Regarding the SWAT model, the input data driving the model is sourced from ERA5, and the RS-ET data used for model parameter calibration is derived from EB\_ET\_v2. The simulation and calibration of evaporation are part of the process to obtain the simulated runoff results, rather than the focus or outcome of the experiment itself. For the water balance calculations, whether determining runoff or water storage changes, "evaporation (E)" is considered a process variable. To minimize the impact of this process variable, the RS-ET data used for calibration and as input for the water balance equation are kept consistent (both are EB\_ET\_v2). There is no need to employ multiple datasets for model calibration and validation separately. As for the comparison of different RS-ET datasets in the manuscript, it serves to demonstrate the reliability of the evaporation results simulated by the model as well as the credibility of EB\_ET\_v2, which was used for model calibration.

1. *It may be because I am not very familiar with the GRACE products. I don't understand how the GRACE products were incorporated into the framework to verify the spatial distributions of SWAT runoff simulations. Further explanation is needed.*
2. *L324-328: I don't understand how spatial analysis of runoff was done.*

Response:

We sincerely appreciate your insightful questions and have made corresponding revisions to the relevant section (3.3.3). (**Line 340-349**)

The earlier validations primarily focused on the temporal validation of runoff results, while the incorporation of GRACE data aims to evaluate the spatial distribution results. Specifically, the simulated runoff, lake precipitation, and lake evaporation were input into the water balance equation to calculate the LWSC results, which were subsequently visualized for six sub-basins. Likewise, the LWSC results derived from GRACE data were visualized within the same sub-basins, facilitating a comparison of the spatial distributions between the two datasets. As the LWSC is calculated based on runoff, this approach also serves as an indirect validation of the spatial distribution of runoff.

图示

AI 生成的内容可能不正确。

你怎么不用这张图告诉他，GRACE在其中起的什么作用？

你说的还不是很清晰

首先你告诉他GRACE计算的水储量包括个啥，水重力卫星，就是陆地的所有水，尤其对于这个相对封闭的系统比较合适

首先在第一阶段：通过LWSC = R + P – E，其中LWSC是GRACE，就能通过P、E反推R了

在第二阶段，拿率定好的SWAT，能推出一个TWSC（用水量平衡方程），GRACE也能计算一个TWSC，空间一对比若符合，则就成为互补验证了

都没说清楚

***Specific comments:***

1. *L118: What simulation? Do you mean SWAT runoff simulation?*

Response:

Thank you for your thoughtful question. Indeed, the term "simulation" specifically refers to the SWAT runoff simulation results.

1. *L136: What do you mean by "diurnal variations"? Do you mean daily or seasonal variations?*

Response:

Thank you for your insightful question. In the manuscript, the rivers surrounding Hala Lake are described as some seasonal rivers, but their daily variations are also significant. The ambiguous part of this statement has been revised to make it more precise and easier to understand. (**Line 137-139**)

1. *L155: Please define LWSC when it appears for the first time.*

Response:

Thank you for your suggestion. The term "LWSC" has already been defined when it first appears in the manuscript. (**Line 157**)

1. *L157-158: This sentence is very confusing. Please rewrite it. I believe that the "simulated runoff results" are the results from water balance calculations based on lake level and lake area RS products, not SWAT simulation. Am I correct?*

Response:

Thank you for your thoughtful suggestion. Your understanding is absolutely correct, and I have revised the sentence accordingly to avoid any potential ambiguity in the expression. (**Line 158-162**)

1. *L158-161: How LWSC was incorporated into the model, and which model? Do you mean SWAT? What is the second LWSC as a hydrological variable? What do you mean by that? What do you mean by "observed data"? I thought it was an ungauged basin.*

Response:

Thank you for your insightful question. This sentence serves as a summary of the overall framework and approach; however, it lacks conciseness and is not essential to this section. To improve the clarity and brevity of the manuscript, I have removed this unnecessary statement. (**Line 162-166**)

1. *Fig 2. This figure is way too complicated. Please see my general comments.*

Response:

Thank you for your valuable suggestion. The figure has been revised in accordance with your general comments (Fig. R1).

1. *L194: A parenthesis is missing.*

Response:

Thank you for your careful suggestion. Due to my oversight, one side of the parenthesis was missing, which has now been corrected. (**Line 201**)

1. *L197 & L224: Please define GSW for its first appearance in the paper.*

Response:

Thank you for your suggestion. I have added full definitions for all product names when they first appear in this section. (Line 203-208)

1. *L242: Do you mean "six subbasins" by "six sections"?*

Response:

Thank you for your reminder. The term "six sections" refers to the six delineated sub-basins. I have revised the wording to ensure clarity and precision. (**Line 254-255**)

1. *L276: What do you mean by "varied normally"? Also, do you mean "for each variable", not "for each condition"?*
2. *L277-278: Do you mean "the impact of each meteorological factor on ET", not "runoff"?*

Response:

Thank you for your two questions. Each meteorological variable is expected to vary over the study period; for example, daily temperatures are different. The method involves keeping one variable constant (e.g., setting the daily temperature to 20°C) while allowing other variables to vary according to natural patterns, hence the term "varied normally." I have replaced "condition" with "variable" for greater accuracy and revised the sentence to eliminate any ambiguity.

Additionally, regarding the sentence you mentioned, "the impact of each meteorological factor on ET" actually refers to its impact on runoff. At this stage, the model has already been calibrated using evapotranspiration (ET) data, and the newly derived parameters are reintroduced into the model. As described in the previous response, this process allows for assessing changes and differences in runoff results, meaning the focus is on runoff impacts rather than ET. In this study, ET serves only as an intermediate variable for calibrating the model, and no factors influencing the ET process are explored. I have revised this sentence to avoid any lack of clarity in the expression. (**Line 286-293**)

1. *L288: What do you mean by "non-rate-dependent ET products"?*

Response:

Thank you for pointing out this error. It was indeed a typographical mistake on my part. The correct meaning should be that the ET products not used for calibration were referred to. I have made the necessary corrections in the manuscript. (**Line 303**)

1. *L301-302: I believe the units for P and E should be m (meters), not mm.*

Response:

Thank you for identifying this error. You are absolutely correct that the formula should use meters (m) to maintain unit consistency. During the simulation of ET and the extraction of rainfall from products, the units were in millimeters (mm), leading to an oversight in unit conversion. This issue has now been corrected. (**Line 316 and 317**)

1. *L366: What do you mean by "ET data was used for simulation and rate determination"?*

Response:

Thank you for pointing out the issue. It was a typo on my part. It should be "calibration and validation" instead of "rate determination." The correction has been made. (**Line 387-388**)

1. *L371: It should be "the model simulation is less accurate than the average of the observed data". The word "average" is missing.*

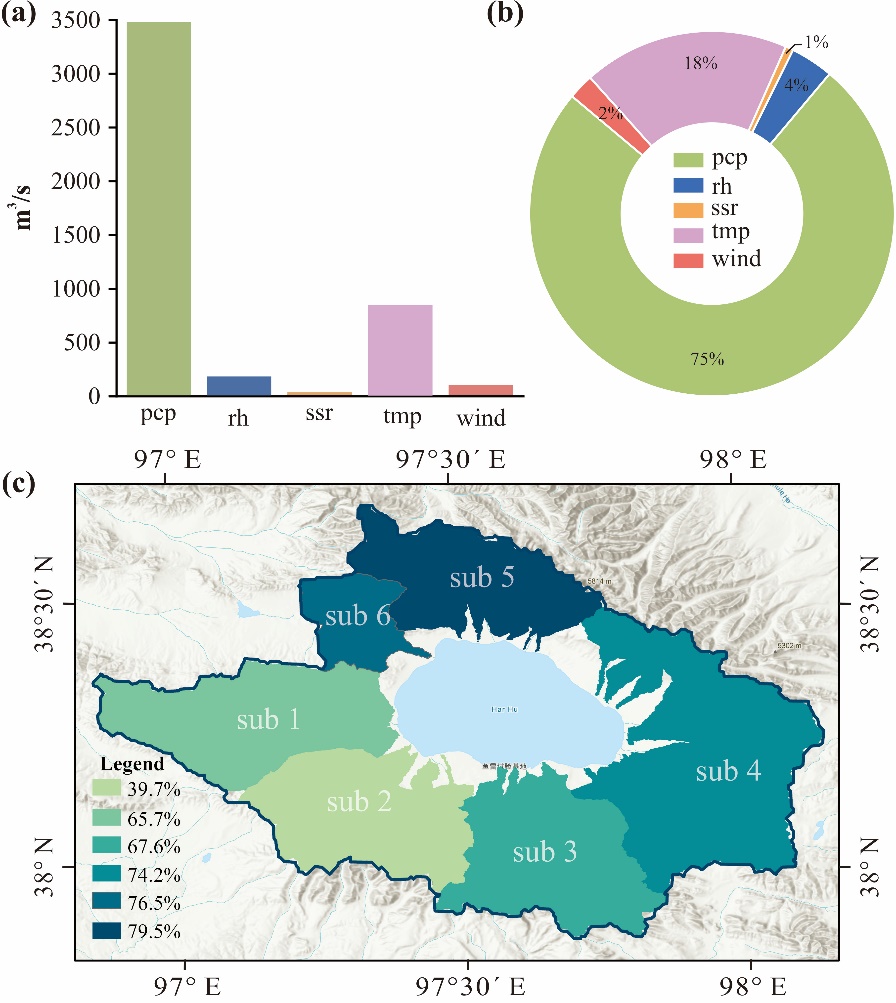
Response:

Thank you for your thorough review. I have added the missing words. (**Line 393**)

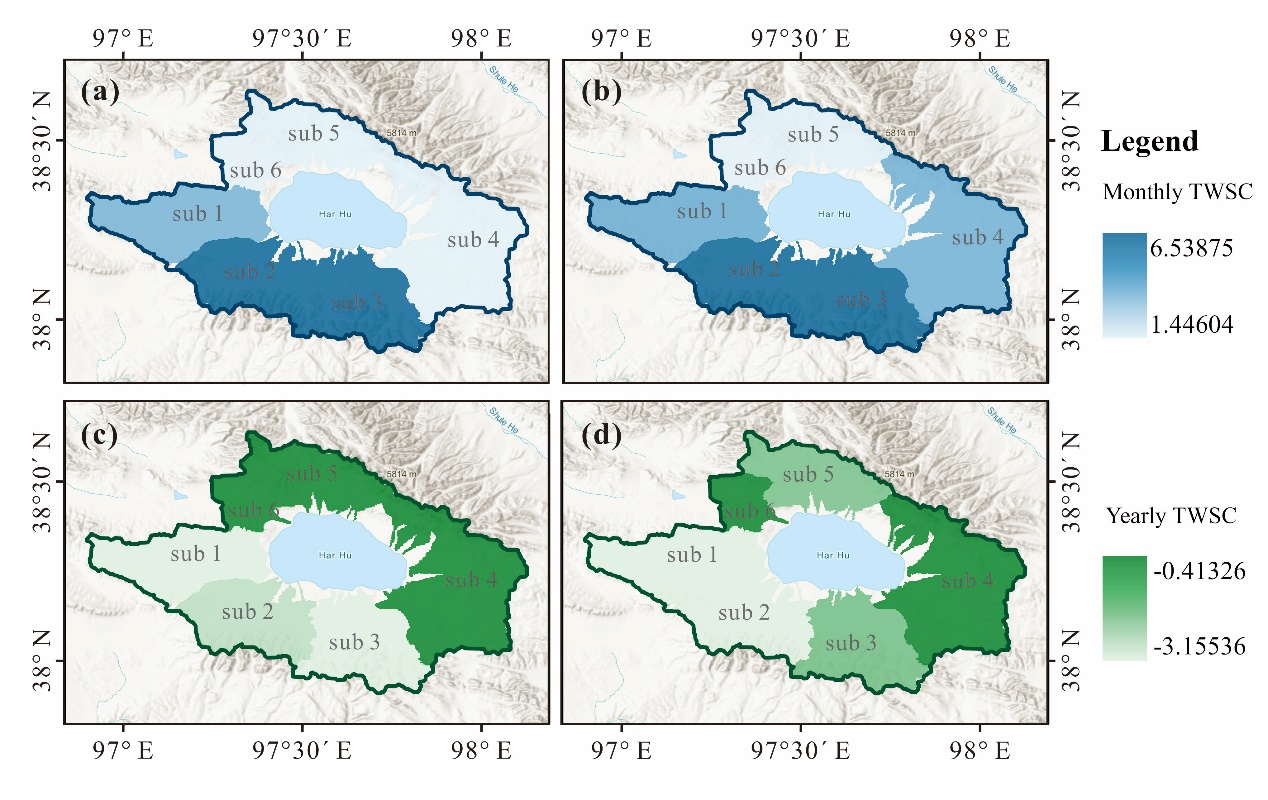
1. *L386-388 and other places: Sub-basins are not labeled in Fig. 1, 5, or 8.*
2. *Fig 5c & Fig 8 & Fig 1: Please label subbasins.*

Response:

Thank you for your kind reminder. You are correct that I overlooked this detail. I have added sub-basin annotations to Fig. 5 and 8 (Fig. R2 and R3). However, adding them to Fig. 1 may not be an ideal adjustment, as it could disrupt the overall layout of the figure. Since Fig. 1 is intended to present the basic information of the study area, I have opted to include the sub-basin annotations in other relevant figures instead. I hope for your understanding.



**Fig. R2.** (a) Meteorological sensitivity results for the entire basin, (b) the extent to which each meteorological factor affects runoff, and (c) the sensitivity of precipitation in each sub-basin. In the figure, pcp means daily precipitation, rh means relative humidity, ssr means solar radiation, temp means daily maximum and minimum temperature, and wind means wind speed. Sub 1 to sub 6 represent the six sub-basins.



**Fig. R3.** Insights into monthly and annual average distribution of LWSC based on GRACE and water balance calculations: (a) monthly average water storage change from GRACE data, (b) monthly average water storage change from water balance calculations, (c) annual average water storage change from GRACE for the ten-year period from 2005 to 2015, and (d) annual average water storage change from water balance calculations for the same period.

1. *Fig 3 caption: Which remote sensing ET product?*

Response:

Thank you for your reminder. The text refers to the EB\_ET\_v2 product, and I have added it accordingly. (**Line 414**)

1. *L406: It should be Fig. 5c, not Fig. 5b.*

Response:

Thank you for your reminder. Additionally, the earlier references should also be Fig. 5a and Fig. 5b, and I have corrected them as well. (**Line 424 and 429**)

1. *L454: This statement contradicts the statement in L251, which states that only EB\_ET\_V2 was used for model calibration and validation.*
2. *L455: What do you mean by "surpassing the SWAT simulation results"?*
3. *L456-457: I don't understand what you are saying here.*

Response:

Thank you very much for your thoughtful question. In response to your three questions, I have the following answers and revisions.

For the first question, you are absolutely correct that only EB\_ET\_V2 is used for model calibration. My previous wording was unclear. What I intended to convey is: "Fig. 6 presents the comparison of SWAT-simulated ET and the calibration product EB\_ET\_V2 with MOD16-STM and ERA5 data, showing the fit of both SWAT and EB\_ET\_V2 with MOD16-STM and ERA5, respectively." I have revised the original text and opted for more concise language.

As for the comment of L455, this was a case of unclear wording on my part. "Surpassing the SWAT simulation results" means that EB\_ET\_V2 shows a better fit with MOD16-STM compared to the SWAT simulation.

In terms of the last one, what this means is that both the SWAT simulation results and EB\_ET\_V2 are compared with ERA5 evaporation products. The former shows significantly better performance than the latter. Upon analysis, this may be because SWAT is driven by ERA5 data, making it more likely to exhibit correlation and consistency with ERA5 products. In contrast, EB\_ET\_V2 is a different product and may differ from ERA5. (**Line 474-487**)

1. *L472-473 and Figure 7: Do you mean the "total" runoff of the six sub-basins?*

Response:

Thank you very much for your insightful question and your understanding is correct. Dividing the six sub-basins is merely a step in the calculation process. Ultimately, it is the total runoff from these six sub-basins that is used.

1. *L474: What do you mean by "observed data?*

Response:

Thank you for your reminder. This sentence contained some ambiguity and has been revised to: Using the LWSC derived from satellite data, we calculated the annual and monthly runoff based on the developed water balance equation. These runoff values were then compared to those simulated by the SWAT model to assess the model's accuracy. (**Line 502-504**)

1. *Section 4.2.3: I don't understand how GRACE data was used to determine the spatial distribution of model simulated runoff.*

Response:

Thank you very much for raising this question. Although the spatial resolution of the GRACE satellite is quite coarse, the six sub-basins delineated happen to align with the location of one grid cell of the GRACE satellite. By calculating the LWSC for each grid cell in the satellite product and mapping it to each sub-basin, then displaying the values with colors ranging from dark to light, the spatial distribution of water storage can be determined. Similarly, runoff simulated by the SWAT model, after calculation through the water balance equation, also yields LWSC, which is displayed using varying color intensities. The spatial distributions of the two are then compared. Since LWSC and runoff can be interconverted using the water balance equation, the spatial distribution of runoff can be indirectly determined.

1. *Section 4.3: What software or program do you use for Bayesian inference? I am not sure how this section serves the objectives of the paper. If you decide to include this section in the paper, more explanation and discussion are needed for Table 4.*

Response:

Thank you for your insightful question and suggestions. Please find the responses below:

***Software or Program for Bayesian Inference:***

The Bayesian inference and joint distribution analysis were conducted using the Stan programming language within the R statistical environment. The package was used for parameter estimation, while additional packages such as and were utilized for visualization of posterior distributions.

***Relevance to the Study Goals:***

The Bayesian analysis in this study aims to quantify the uncertainty in SWAT-simulated runoff (SWAT\_Q) and water balance-derived runoff (Water\_Balance\_Q). By exploring the variability and correlation between these datasets, the analysis provides a more robust validation of the SWAT model outputs. This enhances the reliability of the model in simulating hydrological processes, a key goal of the study.

***Explanation and Discussion of Table 4:***

To address your concern, I have revised the manuscript to expand the explanation of Table 4. The table provides a detailed summary of the posterior statistics, including the mean, standard deviation, and the 95% highest density interval (HDI) for key parameters, along with Monte Carlo standard errors and effective sample sizes. These statistics are crucial for evaluating the quality of the Bayesian inference results and their implications for uncertainty in the hydrological model. I have also added a discussion linking these findings to the broader objectives of the study. (**Section 4.3**)

1. *Section 5.4: I don't want to pretend I understand this section much.*

Response:

Thank you for your question. I will provide a detailed explanation of this section.

This section discusses the issues observed after plotting the runoff curves obtained from the calibrated SWAT model in chronological order, along with related considerations. As shown in Figure 9, during the lowest runoff periods (baseflow) in many years, the runoff does not display a simple upward trend. Instead, there is a sudden drop followed by a sharp rise, forming an inflection point that contradicts my expectations. This prompted further exploration and analysis.

First, I examined the initial results from the SWAT model before any parameter calibration was performed. Next, I reviewed the simulation results using different calibration parameters. Additionally, I analyzed the curves of the precipitation and evaporation data used to drive the model. Based on these investigations, I concluded that such anomalies likely originate from the input precipitation data itself. Precipitation influences the final runoff results through certain parameters, and the parameter calibration process effectively reduces the impact of these anomalies on runoff.

I have also considered whether this section was expressed unclearly, making it difficult to understand. Adjustments have been made to improve clarity and ease of interpretation.

1. **Reviewer #2:**

***General comments***

1. *Did the authors rely exclusively on remote sensing ET data for model calibration? Given the uncertainties typically associated with remote sensing ET data, achieving robust results solely with these data would be noteworthy. It would be helpful if the authors provided additional details regarding the accuracy and reliability of the ET data used.*
2. *The discussion of uncertainties in the multiple remote sensing datasets is somewhat general. Since improving model accuracy is a key aim of this study, it would strengthen the methodology if these uncertainties were considered and explicitly addressed from the outset.*

Response:

Thank you for your two highly valuable suggestions. In this study, the model calibration indeed relied solely on RS-ET products. However, in further validation, several other ET products and precipitation dataset were also used. Regarding the issue of uncertainty that you mentioned, I agree that this was an oversight on my part. I have added relevant content in the discussion section (**5.1. Uncertainty of RS data**).

Specifically, although there are no in-situ observation stations within the study area, some observational stations exist in regions with the same latitude and similar climatic conditions. These stations provide short-term ET data (2003–2005) and precipitation data (2021–2023). While the time series is not long enough to support model calibration, it can be used to validate the reliability of the remote sensing data. The ET values extracted from the corresponding grids of the three ET products were compared with evaporation data from the CN-Ha2 station near the Hala Lake basin. Additionally, MSWEP precipitation data were extracted from the corresponding grids of four precipitation stations near the study area for comparison. Detailed information about the five stations is provided in Table R1 of the supplementary information. The reliability analysis results for the ET and precipitation products are presented in Fig. R4 and R5. The results show that all three ET products exhibit good agreement with station data (R > 0.7), with the EB\_ET\_v2 product performing the best. Similarly, the MSWEP precipitation product aligns well with observed rainfall data from the four stations (R > 0.6). This indicates that, despite some uncertainties, the evaporation and precipitation products are relatively reliable. I have added the tables and figures for this section to the supplementary materials.

# Table R1. Details of flux stations and meteorological stations.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Site code** | **Site name** | **Latitude** | **Longitude** | **Time scale** | **Resource** |
| CN-Ha2 | Haibei Shrubland | 37.609 | 101.327 | 2003-2005 | (Li, 2016) |
| 52643 | Sunan | 38.833 | 99.617 | 2021-2023 | China Meteorological Administration |
| 52836 | Dulan | 36.300 | 98.100 | 2021-2023 |
| 52825 | Nuomuhong | 36.433 | 96.433 | 2021-2023 |
| 52533 | Jiuquan | 39.767 | 98.483 | 2021-2023 |  |

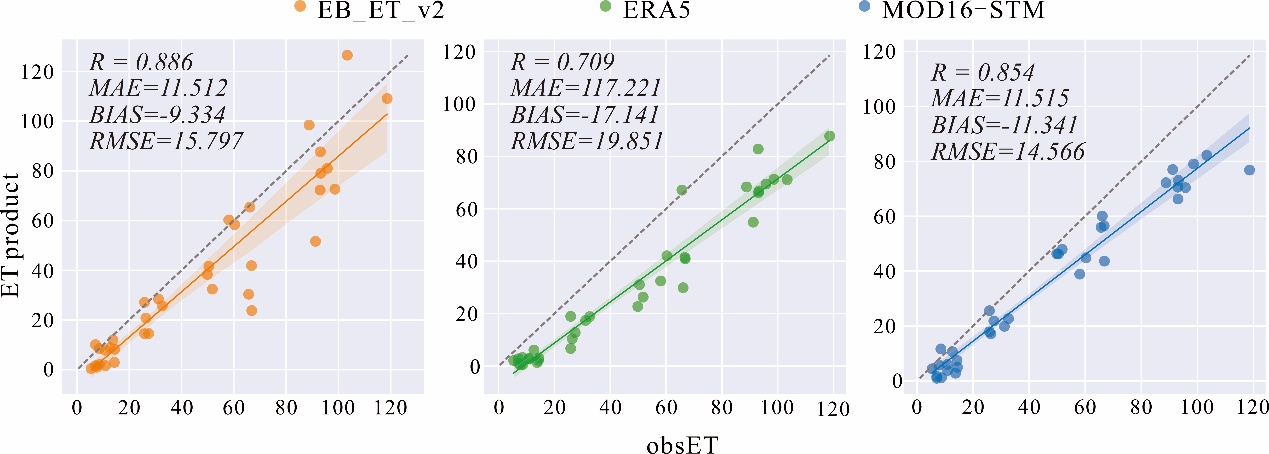


Fig. R4. Scatter plots between the observed flux station against three ET products. The orange, green and blue points represent EB\_ET\_v2, ERA5, and MOD16-STM product, respectively.

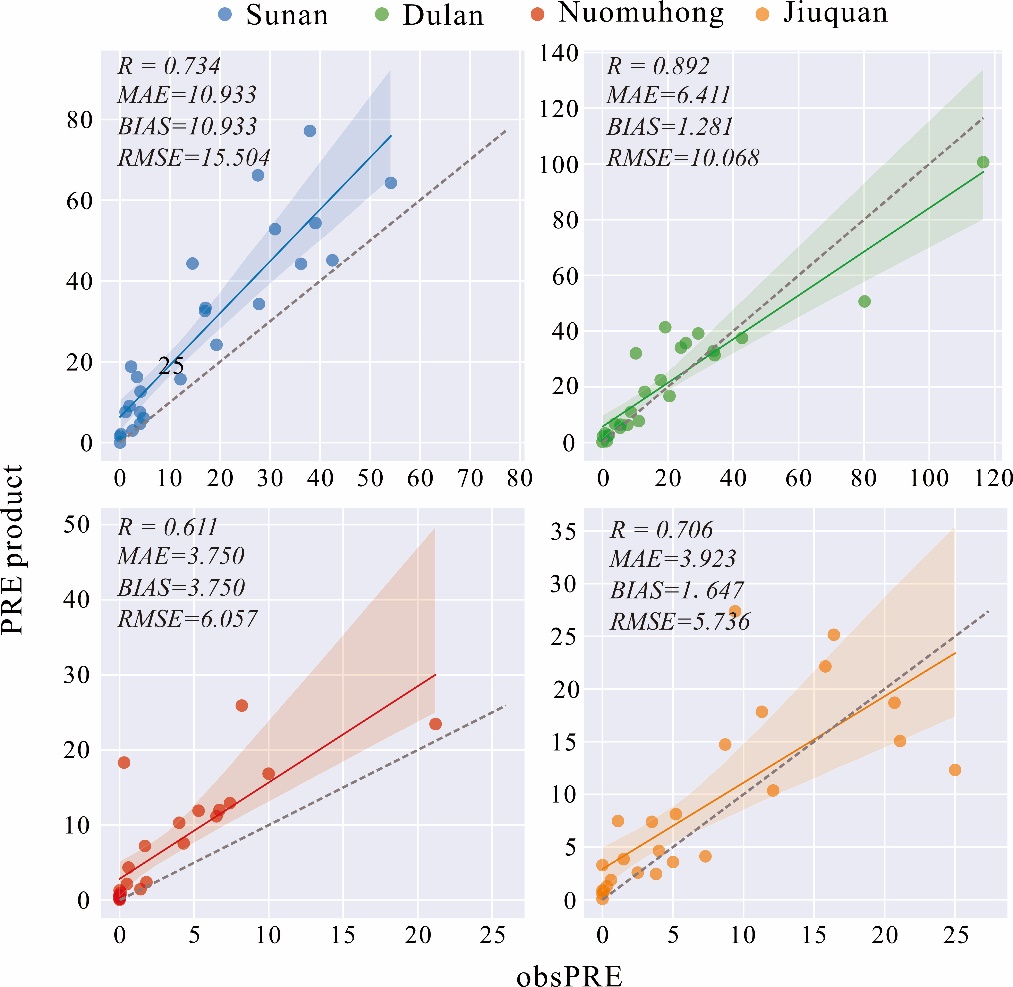


Fig. R5. Scatter plots between the four observed precipitation stations against MSWEP data extracted at the corresponding grid locations of the stations. The blue, green, red and orange points represent Sunan (52634), Dulan (52836), Nuomulan (52825), and Jiuquan (52533) site, respectively.

1. *The methodology section could be made clearer. For example, when evaluating simulation results, the inclusion of mathematical formulas with consistent notation aligned with earlier equations would enhance clarity and precision.*

Response:

Thanks very much for your constructive suggestions. I have revised the methods section according to your suggestions, particularly regarding the symbols used in the equations.

***Specific comments:***

1. *Lines 71-72: For hydrological simulations with specific optimization objectives, deep learning methods might outperform traditional model calibration approaches. While the literature review touches on deep learning, its inclusion may seem unnecessary unless tied directly to the study. This could raise questions about why deep learning methods were not considered or applied in this research.*

Response:

Thank you very much for your positive comments and constructive suggestions. Although deep learning has its advantages, it relies heavily on large volumes of input data, which makes it unsuitable for the data-scarce region of this study. Additionally, its unclear physical mechanisms make attribution analysis challenging. This naturally leads to the introduction of hydrological models that are more suitable for this context. Therefore, it is still necessary to discuss deep learning to some extent. Taking your feedback into account, I agree that deep learning is not directly related to this research and is not its core focus. As such, I have reasonably removed some redundant content on this topic. (**Line 59-64)**

1. *Line 194: When introducing the three ET products, it would be helpful to clearly articulate their respective roles in the study. Why was* *EB\_ET\_V2 chosen for calibration while others were used for validation? Additionally, their accuracies and uncertainties should be detailed. It's worth noting that some ET products are influenced by precipitation data during their generation, which might introduce dependencies or conflicts with the precipitation data used in this study. Has this been considered?*

Response:

Thank you for pointing out this important aspect. EB\_ET\_V2 features high temporal and spatial resolution (0.05°, 1 day), making it particularly suitable for the study area, which has a relatively small size and finely divided sub-basins. It significantly outperforms other ET products and is therefore used for model calibration. (**Line 221-223**)

Indeed, some ET products may be influenced by precipitation data during their generation, which could potentially introduce dependencies or conflicts with the precipitation data used in this study. To address this, we carefully selected ET products that are generated independently of the precipitation data used in our model calibration and validation. Additionally, we conducted a cross-comparison between the ET products and precipitation datasets to ensure minimal overlap or bias. While this mitigates the risk to some extent, we acknowledge that this issue may still contribute to uncertainty in the results and have discussed this in the section on "**5.1 Uncertainty of RS Data.**"

1. *Line 196: The full name of MSWEP should be spelled out at its first mention. Precipitation data is crucial for hydrological simulations, and its accuracy deserves thorough discussion. Several studies have highlighted significant biases in existing precipitation datasets for the northern Tibetan Plateau. How might these biases affect the study's conclusions, and has the quality of the precipitation data been carefully evaluated?*

Response:

Thank you for your valuable feedback. You are absolutely right that MSWEP should be spelled out in full when first mentioned. Regarding the importance of precipitation data in hydrological simulations, we agree that its accuracy is crucial and warrants detailed discussion. Several studies have indeed highlighted significant biases in existing precipitation datasets for the northern Tibetan Plateau. To address this, we carefully evaluated the quality of the precipitation data used in this study. While we acknowledge that biases in precipitation data could potentially affect the conclusions, we have attempted to minimize their impact by validating the dataset against observational data from nearby regions with similar climatic conditions. Additionally, we have discussed the uncertainties associated with precipitation data in the section on "**5.1 Uncertainty of Input Data**".

1. *Line 566: Since this paper centers on using remote sensing data for hydrological model calibration and validation, addressing the uncertainty in these datasets is fundamental. Rather than presenting this discussion as a separate section near the end, it would be more effective to incorporate uncertainty considerations into the research framework itself.*

Response:

Thank you for your valuable suggestion. After making revisions, I have incorporated uncertainty analysis into the flowchart. This primarily addresses the uncertainty analysis of ET products and precipitation datasets. Detailed information can be found in the supplementary materials, and the figures and tables I added are included in Fig. R1, R4 and R5 and Table R1.

1. *While GRACE data plays a key role in validation, the study does not sufficiently address its limitations, particularly its coarse spatial resolution (300 km). This limitation may significantly affect accuracy in smaller basins like Hala Lake. A more detailed discussion of this issue would improve the study's overall robustness.*

Response:

Thank you for highlighting this important point. We acknowledge the coarse spatial resolution of GRACE data (300 km) as a significant limitation, particularly for smaller basins such as Hala Lake. However, it is worth noting that the six sub-basins delineated in our study happen to align with the location of individual grid cells of the GRACE satellite. By calculating the LWSC for each grid cell in the satellite product and mapping it to each corresponding sub-basin, we were able to represent the spatial distribution of water storage, with values displayed using colors ranging from dark to light. To further address the resolution limitation, we incorporated additional data sources with higher spatial resolution, such as EB\_ET\_V2 and SWAT-simulated runoff, to supplement the GRACE data and mitigate the impact of this constraint.

Additionally, we have discussed the spatial resolution issue in detail in the revised manuscript, particularly in the "Uncertainty of Input Data" section, emphasizing how this limitation was considered during the model validation and how the results were interpreted in light of this constraint. We appreciate your suggestion, as it has helped us strengthen the robustness of the study.