Response to comments on Manuscript hess-2024-230

Title: Achieving water budget closure through physical hydrological processes modelling: insights from a large-sample study

Authors: Xudong Zheng, Dengfeng Liu*, Shengzhi Huang*, Hao Wang, Xianmeng Meng

Manuscript ID: hess-2024-230

Dear Editor and Reviewers,

Please find enclosed our responses to the manuscript assessment entitled “Achieving water budget closure through physical hydrological processes modelling: insights from a large-sample study”.

We would like to express our sincere gratitude to the Editor, the two anonymous reviewers for their invaluable support and constructive suggestions, as well as for the opportunity afforded to us to revise our work. We have given full attention to all comments and suggestions and made all revisions accordingly. It has resulted in an improved manuscript that fully addresses all concerns.

Concerning the revision of the manuscript, the following changes have been made:

(1) The correction of Eq. 7 and the redrawing of Figs. 10 and 15 to enhance readability.
(2) Section. 5.2.1 has been updated to highlight the impact of forcing uncertainty and potential solutions.
(3) Section 5.2.2 has been updated to emphasize the defects of HBV model.
(4) Several spelling errors and grammatical issues have been corrected.

We believe the paper quality has significantly improved through this review process. We are also happy to address any comments that may further strengthen the paper quality. We are uploading our point-by-point response to the comments, an updated manuscript with red highlighting indicating changes, and a clean updated manuscript without highlights.

In the point-by-point responses below, the original comments are displayed in **bold**. **EC-n/**, **RC3-n/** correspond to the Editor, Referee 3, respectively. The corresponding responses begin with **R/** and the revisions is highlighted in red, while important sections are marked in blue. The quoted content is displayed in *italics*.

We thank you for your consideration,

Dengfeng Liu
Email: liudf@xaut.edu.cn

Responses to Editor:

EC-1/ Publish subject to minor revisions (review by editor). Please address the minor comments from referee #3.

R/ We greatly appreciate your timely handling of our manuscript, as well as the recognition of our work and the opportunity to publish after minor revisions. All comments from referee #3 have been addressed point-by-point below. The updated manuscript, which includes a tracked version with changes highlighted in red and a clean version without highlights, will be submitted alongside this response file.

We believe that our manuscript has significantly improved through this review process, and we are open to addressing any comments that may further enhance the quality of our paper. If there are any questions or suggestions, please feel free to contact us.

Responses to RC3:

We sincerely appreciate your time and effort in reviewing our manuscript, and providing valuable suggestions. All comments from RC3 are addressed below with point-by-point responses.

RC3-1/ This is a very interesting and valuable study, but it has to be said that it faces great challenges. I'm trying to understand the whole study, and some confusion and doubts need to be further explained based on the previous reviewers.

R/ Thanks for your positive feedback and recognition of our work. It must be said that, as you pointed out, the goal of this work is indeed challenging. In the era of big data, we believe that effectively integrating our knowledge (water balance) and real-world observation (measurements) holds significant value. This is the primary motivation behind this work. While it may still have some limitations, we are confident that by addressing the issues you kindly raised and making the necessary revisions, it can become a meaningful contribution to the field.

Your comments are valuable for revising and improving our paper. Below, we provide detailed responses to each of your concerns. The corresponding revisions are attached after the responses and have been incorporated into the revised manuscript.

RC3-2/ First of all, the essential purpose of this study is to try to correct the errors of basic data through the principle of water balance using the HBV model, to reduce the balance residual of water in closed basins. The two residual hypotheses seem reasonable, but I found that data correction mainly targets sink terms other than precipitation which is the source term in water balance. Generally speaking, the errors that are difficult to overcome in hydrological model simulation are mainly from precipitation rather than others such as ET, SWE, etc. According to equations 3 and 4, inconsistency residuals of precipitation are attributed to omission residuals, which is also discussed in section 5.2.1. However, Figure 5 shows that the omission residual does not seem to dominate compared to the inconsistency residual. Does this indicate that TRMM data is working well across the COUNS? In my opinion, its inconsistency residual is largely offset by the correction of other data. Moreover, the corrected inconsistency residuals in this study are essentially caused by the inconsistencies of precipitation.

R/ Excellent point! Clearly, you have have gained a profound understanding of the scientific question addressed in this work. We are sincerely grateful for your thoughtful analysis.

The first point we acknowledge is that this framework does not explicitly address the residuals in the precipitation products, which have been considered one of the main sources of residuals in many previous studies (Sahoo et al., 2011; Ansari et al., 2022). As a primary source term and the main forcing of hydrological models, the absence of precipitation correction appears counterintuitive. This was also mentioned in a previous community comment:

CC1-2/ There are PTRMM in both Eq. (5) and (6), then how do we reduce the inconsistency residuals brought by P in the water budget?

However, in our opinion, assuming the uncertainty (residuals) in precipitation product can be neglected is both necessary and feasible. This is also why we explicitly emphasized this assumption at the beginning of the method section. From a necessity perspective, this assumption is made to ensure the operation of the correction framework. Since the current setup does not couple with an atmospheric model, and precipitation, as a forcing, cannot be simulated, it leaves us without a reference within a physically consistent simulation system. From a feasibility perspective, the reason this assumption holds due to the model's inherent calibration compensation capability. Specifically, the calibration compensation capability allows the model to accommodate uncertainties in the forcing (namely, precipitation) through calibration, producing relatively reliable simulations, as mentioned in the introduction (Line 95-100 in the tracked version):

Another distinctive feature of hydrological models, known as error adaptability or calibration compensation capability, underscores their pivotal role as innovative solutions for addressing challenges in achieving water budget closure. The feature emphasizes that hydrological models can, to some extent, compensate for biases in model inputs, outputs and structure, allowing satisfactory performance even when the utilized datasets exhibit certain inaccuracies (Wang et al., 2023). This provides hydrological models with the potential to integrate forcing and evaluation datasets into a unified water balance system under the soft constraint paradigm.

This capability is further enhanced under the constraints of multi-objective calibration. As shown in Fig. C1, despite the potential uncertainty in precipitation, most basins still yield reliable simulation across all variables. All of our analyses exclude basins with unreliable simulations (i.e., the remaining 475 basins), thereby offering a certain level of confidence in this assumption. In other words, given the HBV model, TRMM performs well in these basins.

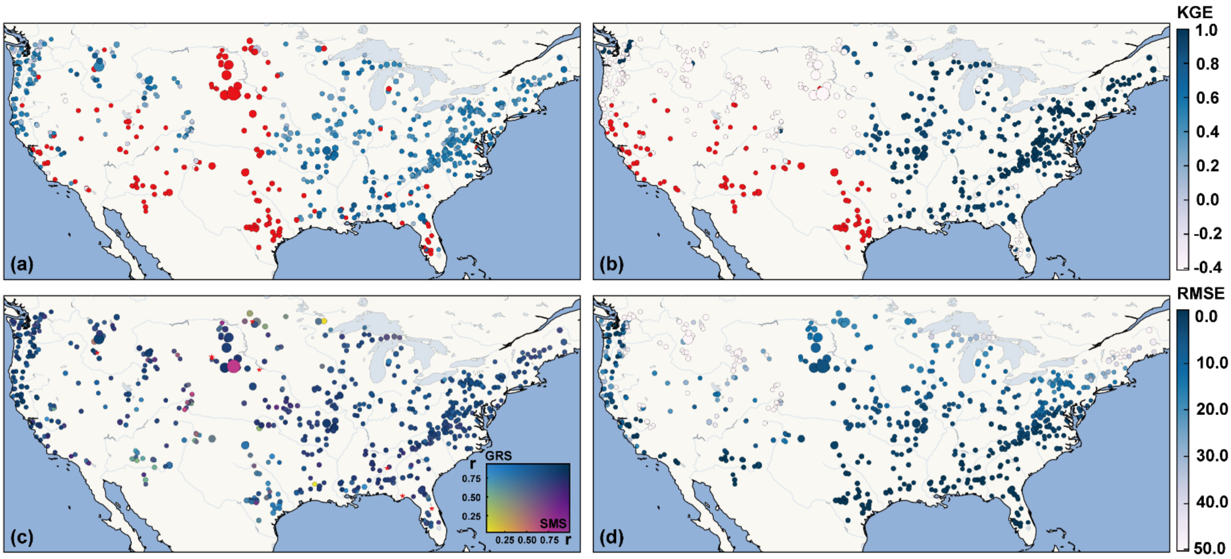

Figure C1. The multi-objective simulation performances of the HBV model across the CAMELS basins. Results are based on (a) runoff, (b) evaporation, (c) soil moisture storage and groundwater reservoir storage, and (d) snow water equivalent. Red dots represent unreliable simulation performance, and the size of points is proportional to the basin area. The unit of RMSE is "mm".

Therefore, under this assumption, the residual in precipitation included in Eq. 4 is eliminated through model calibration. In other words, the TRMM precipitation forms a physically consistent simulation system with the remaining simulated variables (Eq. 6), although it still contains uncertainty. In extreme case, when the measurement system is corrected to be identical to the simulation system, all measurements would become physically consistent. The correction amount, which is the sum of the changes occurring in the measurements (without precipitation), corresponds to the inconsistency residuals. This process can be seen as a collapse from Eq. 5 to Eq. 6.

Returning to the uncertainty inherent in the precipitation products, it is undoubtedly always present. However, when different precipitation products are used, the correction process ensures that the final corrected results are both similar and physically consistent (Fig. 15). This is achieved by maintaining similar inconsistency residuals—corresponding to a similar correction amount—as long as differences in precipitation do not result in substantial variations in the hydrological processes. In general situation, the impact of the uncertainty in precipitation is mitigated through the calibration process, resulting in similar hydrological process. It is worth highlighting that the iterative calibration process (applying a correction rate less than 1) also plays a crucial role (see the response to RC3-4).

On the other hand, the omission residuals are dynamically adjusted. Since the precipitation serves as the source term in the water balance equation, it will lead to different overall water amounts. In the context of a relatively complete water balance equation, the impact of omission residuals is minimal, as their source is solely precipitation. In contrast, the inconsistency residuals arise from uncertainty in the other five terms (Eq. 5). This is why the overall residuals in Fig. 5 are predominantly driven by inconsistency residuals.

Nevertheless, the results in Sect. 5.2.1 suggest that, in this study, the impact of precipitation on the correction of other variables in minimal (likely because both precipitation products are relatively accurate). The difference in omission residuals can provide guidance for the assessment and correction of precipitation uncertainty, as discussed in this section (Line 721-728).

Clues to this possibility are hidden in the above analysis. Systematic biases in precipitation products are directly reflected in the water budget equation, leading to different total input water volumes. Consequently, with the inconsistency residuals of other variables unchanged, maintaining the water balance would require an increase in omission residuals (Fig. 15e). Therefore, it can be inferred that, with other variables unchanged, TRMM demonstrates superior water budget closure compared to Daymet, which contains smaller inconsistency residuals. In other words, the difference in the two omission residuals reflects the discrepancy in inconsistency residuals contained within the two precipitation products. This portion of the omission residuals difference can be directly corrected in the precipitation.

The inability to correct precipitation forcing is a major limitation of this framework. We discuss this issue in Sect. 5.2.1, and in response to your comment, we have added the following points based on our reflections.

Line 292-293 in the tracked version:
In extreme case, when the measurement system is corrected to be identical to the simulation system, all measurements would become physically consistent. This process can be seen as a collapse from Eq. 5 to Eq. 6.

Line 713-715 in the tracked version:
This is achieved by maintaining similar inconsistency residuals—corresponding to a similar correction amount—as long as differences in precipitation do not result in substantial variations in the hydrological processes.

Line 731-733 in the tracked version:
Another strategy is to couple an atmospheric model with this framework to generate simulated precipitation, allowing for the correction of precipitation products. In subsequent work, we will explore these approaches and try to extend the PHPM-MDCF based on these ideas.

RC3-3/ In addition, there are some issues that need to be improved or resolved:

R/ Thank you for your careful review, which has been extremely valuable in enhancing the quality of our manuscript. We have addressed and made revisions for all of these issues, which are provided below.

Specific comments

RC3-4/ "Here, an initial correction rate of 0.5 is set to gradually correct the multisource datasets, thereby avoiding potential uncertainties that arise from excessive correction." Why is there an initial correction? Isn't it calibrated according to the equation 7.

R/ Thank you for pointing out the discrepancies in our equations. It has been revised to (Line 274-277 in the tracked version):

$$M_c^v = M_o^v - Res_i \times \frac{d_v}{d_{all}} \times \alpha, \quad (7)$$

where M_c^v is the corrected measurements of variable v, and M_o^v is the original measurements; d_v is the difference between simulation and measurement of variable v, and d_{all} represents the aggregate of differences for all variables; α is the correction rate, with an initial value of 0.5.

In addition, it should be noted that setting a correction rate effectively helps mitigate uncertainty caused by over-correction. This correction rate can be adjusted based on the simulation results in correction step 3.

RC3-5/ The study compared the effects of Daymet and TRMM. It seems that R is consistent, but E is very different (Figure 15). However, E is a corrected result and should not theoretically differ greatly. I'm not sure if this is due to the calibration method or the data itself.

R/ Thank you for pointing out the potential confusion in Fig. 15. In fact, evaporation shows high consistency between the two correction. In subplot Fig. 15a, the results of the two correction results are represented by the blue line and red dots, respectively, which may give the appearance of inconsistency. In addition, the evaporation points are hidden by other points in Fig. 15b, which further leads to a misunderstanding.

Recognizing this, we have modified Fig. 15b to position the evaporation points in the top layer, as shown below (Line 734-738 in tracked version). It can be seen that the scatter points from both corrections are located near the 1:1 line, indicating that the results of the two evapotranspiration corrections are consistent.

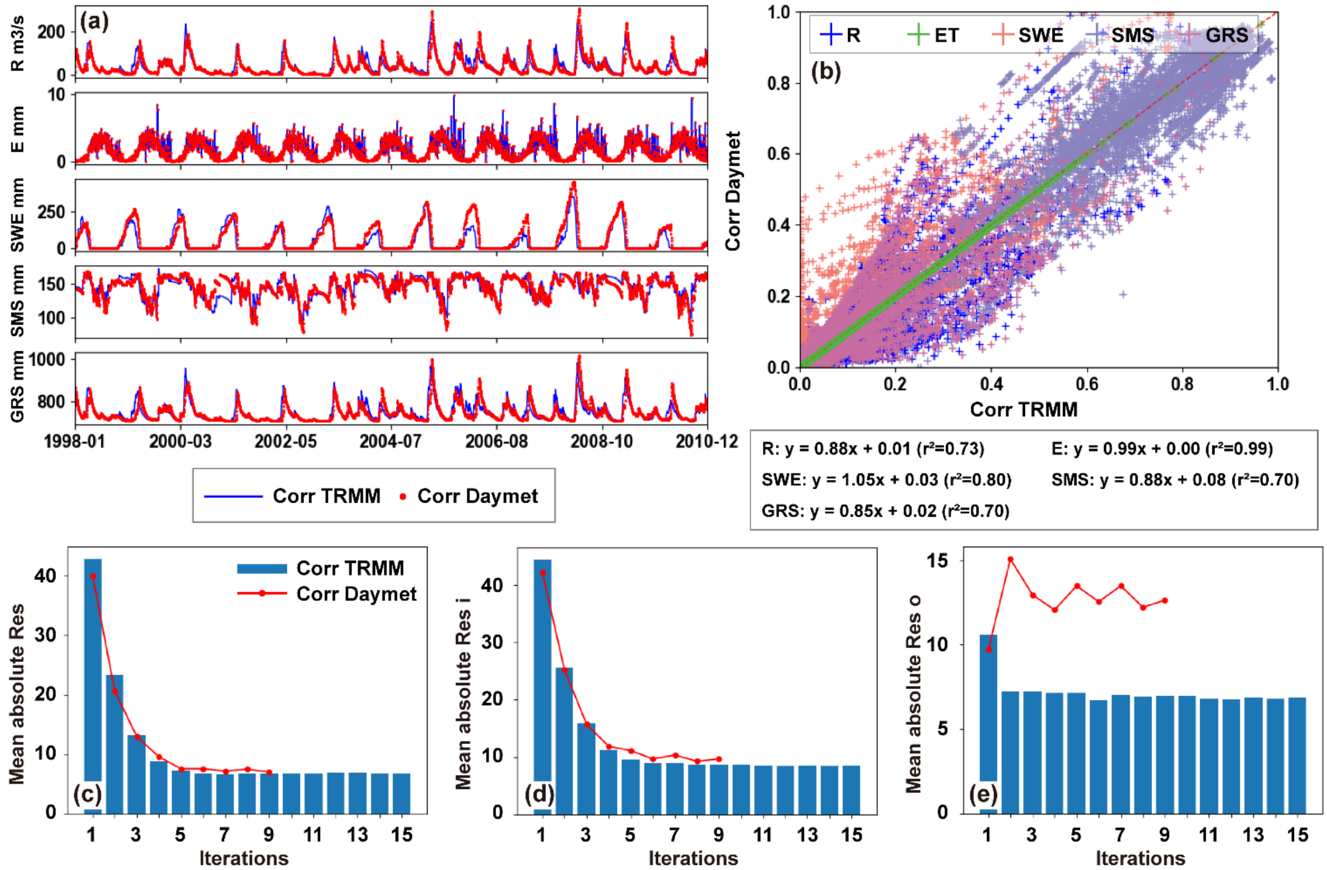

Figure 15. Comparison of correction results based on different forcing datasets (TRMM and Daymet) at basin 1013500. (a-b) Corrected time series of five water budget variables. (c-e) Variation of long-term mean absolute values of three residuals with correction iterations at the monthly scale. The unit of residuals is “mm”.

RC3-6/ Section 5.2.2 mentions the uncertainty of the model structure, but does not cover the defects of HBV. HBV has some shortcomings in both snowmelt simulation and groundwater simulation.

R/ Thank you for your suggestion. According to your suggestion, we have included the shortcoming of the HBV model in Sect. 5.2.2 of the revised manuscript (Line 775-778 in the tracked version). For your convenience and review, we present the revised content below.

It is worth noting that, while we have validated the reliability of the HBV model in the current study, its simplistic physics and lumped design structure lead to significant limitations in simulating several processes such as snow and groundwater (Brunner et al., 2021). In other words, the HBV model may not be suitable for accurately representing the reality of these specific processes.

RC3-7/ The spatial resolution of the gridded data used in this study is very rough, and some of the basins selected are very small. It will cause great uncertainty if the basic data is not reliable.

R/ Thank you for your comment. We agree with this point; the disparity between scale of data and that of research object is an important source of uncertainty. This may be the potential cause of the scale effect in the residuals discussed in Sect. 4.4.1 (Fig. 12, we present it below).

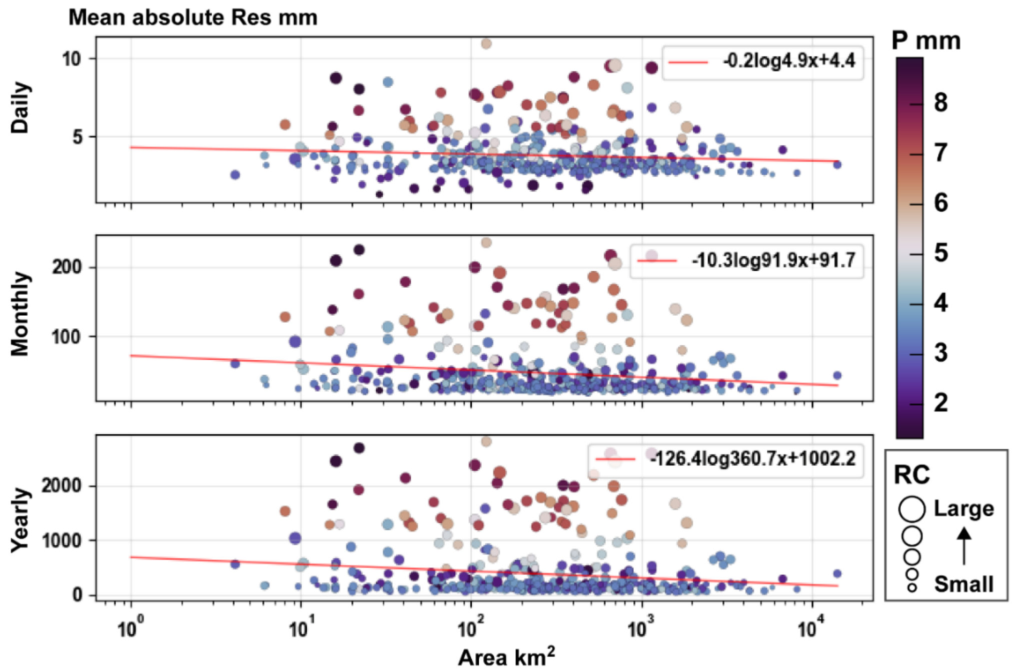

Figure 12. Relationship between the mean absolute of water budget residuals, basin area, long-term average daily precipitation, and runoff coefficient (RC) over 475 CAMELS basins with reliable simulations. The respective red lines represent the linear regression of residuals with basin area for each timescale.

In this study, to establish the water balance equation and drive the hydrological model at the daily scale, while ensuring a certain level of data reliability, we conducted a data selection process from a broad range of available datasets. With reference to previous studies, we ultimately selected the datasets presented in Table. 1. While this does introduce some uncertainty, it is a decision made by balancing data availability with the research objective.

Nevertheless, we believe that the uncertainty introduced by scaling process remains limited within a certain range. This is for the following two reasons. First, all datasets are aggregated to the basin scale in depth units. This is to say, the grid volume data are divided by the grid area, which helps reduce the impact of the mismatch between the grid and basin boundaries, even for very small basins. Second, as we mentioned in the response to RC3-2, the hydrological model's calibration compensation capability further reduces the impact of the mismatch. Since the lumped hydrological model do not account for spatial heterogeneity, the impact of grid distribution is minimal. These factors significantly reduce the uncertainty caused by scale differences. The accurate simulation of runoff at basin outlet support this, as it effectively represents the basin scale, while the forcing is at the grid scale (Fig. C1).

Building on the discussion of this issue, we have added the following points in the discussion section (i.e., Sect. 5.2.1) to highlight the potential uncertainty arising from scale mismatch.

Line 685-687 in the tracked version:

First, the uncertainty in the forcing may arise from two aspects, one is the inaccuracy of the datasets themselves, and the other is the uncertainty introduced by the scaling process (i.e., the conversion from grid scale to basin scale).

Line 699-700 in the tracked version:
Thirdly, the uncertainty caused by the mismatch between the grids and basin boundaries is effectively alleviated through the unit conversion (i.e., from volume to depth units).

Reference

Sahoo, A., M. Pan, T. Troy, R. Vinukollu, J. Sheffield, and E. Wood (2011), Reconciling the global terrestrial water budget using satellite remote sensing, Remote Sensing of Environment, 115, 1850-1865.

Ansari, R., M. U. Liaqat, and G. Grossi (2022), Evaluation of gridded datasets for terrestrial water budget assessment in the Upper Jhelum River Basin-South Asia, Journal of Hydrology, 613, 128294.

Finally, we would like to once again thank the Editor and all the Reviewers for their thorough review and support of our paper. If you have any questions, suggestions, or discussions, please feel free to contact us.