Response to Editor and Reviewers

By Jin K. L. et al.

Dear Editor and reviewers,

First of all, we appreciate the opportunity to address your concerns and improve the earlier version of the manuscript entitled 'Quantifying evapotranspiration in Nepal using multiple constraints'. We thank you for the very valuable comments and insightful suggestions.

We have made efforts to revise the original manuscript and addressed all the questions point-by-point (see below) hoping our conclusions derived from this revision are more convincing and solid.

In particularly, as the reviewer suggested, we improve its structure, clarify methodology and datasets, and enriches the discussion. We added more interpretation and clarification to discuss the limitations and uncertainties and to reinforce the significance of our findings. We have provided more explanations and descriptions that confirmed that both data quality and method used in this study were carefully considered and appropriate. We are happy to answer any further questions you may have. Please find the detail responses to the review comments below.

Kailun Jin,
On behalf of the authors.

Response to Reviewer #1

General comments:

This study evaluates the applicability of global ET products in Nepal's complex terrain using watershed water balance, the Budyko framework, and ecohydrological modeling. The research finds significant differences between high- and low-elevation regions and identifies water-balance imbalances in certain watersheds. The findings are interesting, but several important issues remain. A substantial revision is recommended.

Response: *We thank you for the very valuable comments and insightful suggestions. We have made efforts to revise the original manuscript and addressed all the questions point-by-point (see below) hoping our conclusions derived from this revision are more convincing and solid.*

Specific comments:

1. Why focus on ET in Nepal? Its importance is not clearly demonstrated. From an ecohydrological perspective, ET contributes only a small fraction in this high-altitude Nepalese region, at least currently and historically.

Response: *In the new revision, we will explain the importance of ET studies. We argue that evapotranspiration (ET) in Nepal is important for several reasons:*

1) At national and basin scales, ET is a major water-balance expenditure, not a negligible residual. Mean annual ET across Nepal ranges from 150 to 1100 mm yr^{-1}, and ET/P >0.5. Thus, ET can consume a large fraction of precipitation, up to about 60%. In addition, an accurate estimation can help identify measurement issues of river flow using an approximate water balance equation Q=P-ET, assuming no change in storage. For example, we found in some instances observed flow data were close to precipitation in several basins and obviously violating the water balance because ET should not be zero.

2) At high elevations mountain regions with limited or no vegetation, ET is small but not in the middle mountains where forests are present.

3) High-altitude regions are warming clearly, with frequent climate-induced *hydrological and geological hazards, while meteorological and hydrological stations are sparse to record them. ET is central to water-energy coupling. Even if historical or current absolute ET is small, under warming conditions, small changes in ET can alter soil water, snow/ice melt, baseflow, and springs, thereby affecting downstream water use. It is, therefore, a sensitive and data-scarce diagnostic variable under environmental change.*

4) For the middle mountain regions and below, from ecohydrology and forest management perspectives, the loss of water through evaporation and transpiration is the most important factor in predicting streamflow response to forest management. Nepal has experienced long-term forest clearance, but it is recovering in recent decades, thanks to community-based forestry management. Local people's experiences also indicate that there is a correlation between forest and water as they recognize flow of water/amount of water in a spring as one indicator in measuring the success of existing forest protection and restoration system. In addition, ET links 'forestry-water supply-human water demand' and is key to understanding water-carbon coupling.

5) Little studies have been done on ET estimates and assessments in Nepal. Fortunately, ET can be estimated from remote sensing with assistance of other observable hydrologic fluxes.

2. Why are only remote sensing ET data used? Multi-source data could be incorporated.

Response: *At the national scale, no other ET data are available. Eddy flux sites are not reliable (For the specific reasons, please refer to the fifth reply in this response.) and not representative. Even the existing weather monitoring network is noticeably sparse relative to WMO standards. The complex mountainous field conditions make it very difficult to quickly and effectively obtain stable, long-term time-series observations of hydro-climate data. From a spatial distribution perspective, the problem is even more prominent. 80% of Nepal's hydrological stations are located in low-altitude areas (<500 m), and there are severe observational gaps in high-altitude regions.*

Therefore, in data-scarce complex terrain regions such as Nepal, remote sensing is currently the most feasible means for ET assessment. Remote sensing ET products (e.g., *PML v2, ETMonitor) have been applied globally across various terrains and climates, including some complex terrain regions. However, their validation quality depends heavily on the availability of ground observational data, and Nepal happens to have systematic shortcomings in this context, resulting in limited validation work on remote sensing ET products to date. This study itself also reveals performance issues of remote sensing products in Nepal. For example, remote sensing models systematically overestimate ET in high-altitude areas and underestimate ET in plain areas, and the temporal trends among different products are inconsistent.*

Under such circumstances, this study integrates multiple remote sensing products with an ecohydrological model (WaSSI) and uses 12 basin-scale water balance datasets constructed locally in Nepal to constrain ET estimates. In addition, the Budyko framework can provide independent physical constraints for ET estimation without relying on dense ground observations. This study introduces the Budyko framework to validate the reasonableness of basin-scale ET estimates, which is also a reasonable strategy for addressing Nepal's data scarcity problem. Therefore, multi-source data fusion approach has actually been applied in this study.

3. Different PET calculation methods significantly affect the Budyko relationship, potentially leading to biased conclusions in watershed selection and comparative assessment. The Hamon algorithm and Penman-Monteith FAO-56 differ substantially.

Response: *We acknowledge the differences in this study and incorporate them into the uncertainty discussion. This study clearly points out that the reference evapotranspiration (ET_0, Figure S3 in the supplementary file) calculated by the P-M FAO56 method and the Hamon potential evapotranspiration method (Figure 7 in the manuscript) show obvious differences, and this finding corroborates the conclusions of previous studies (Lang et al., 2017). However, in data-scarce regions such as Nepal, the Hamon method is often adopted because of its low input requirements. This study calculates ET_0 (P-M FAO56) and PET (Hamon) using a unified TPMFD dataset, thereby allowing the differences between the two methods to be diagnosed rather than confounded. This study uses the Budyko framework to constrain remote sensing ET* *estimates from both water and energy perspectives, aiming to identify the physical plausibility boundaries of ET estimates rather than to replace the choice of PET method. Finally, we adopt a multi-product evaluation framework to understand the bias of a single PET method against the background of inter-model differences, avoiding overly definitive conclusions caused by method selection. We evaluated eight global remote sensing ET products (GLASS, GLEAM, PMLv2, REA, PEW, ETmonitor, SSEBop, SEBAL) and the WaSSI ecohydrological model. This multi-product framework allows the uncertainty of PET methods to be examined within broader inter-model differences, rather than judging the 'rightness' or 'wrongness' of a particular PET method in isolation. We will address this aspect further in the Discussion section. (Lines 616-626 in revised manuscript)*

4. Are SSEBop and SSEBop V6 the same model in Table 1?

Response: *The SSEBop and SSEBop V6 models are essentially the same in terms of the model structure, but the input data for the two products (PET, and VIIRS for V6) are different, which also leads to notable differences in performance across elevation gradients, especially in high-altitude regions. We have addressed this point in the revised manuscript in Section 2.3.8 (Lines 344–345 in revised manuscript).*

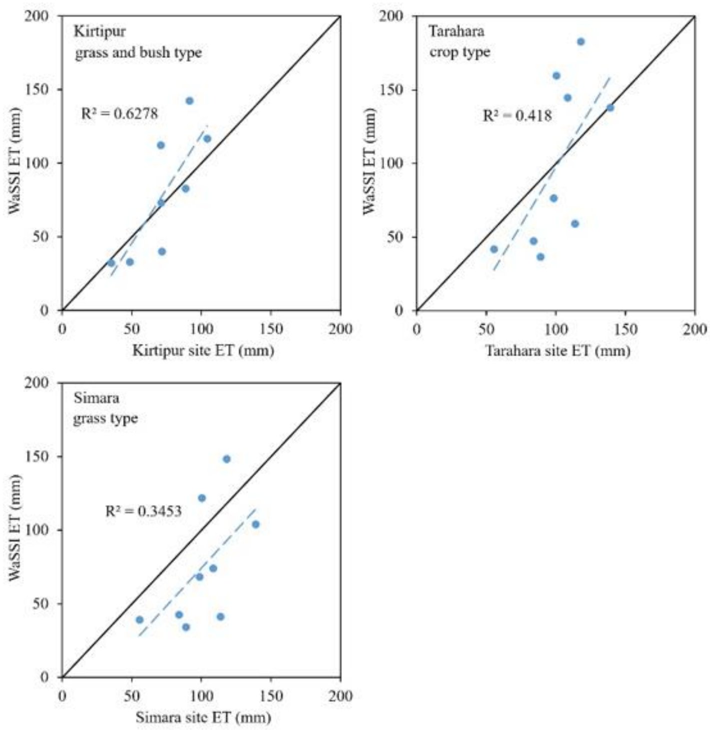
5. WaSSI cannot serve as ground truth, including its national-scale partitioned ET values. Is there no site measurement at all for this region?

Response: *We agree that WaSSI cannot serve as ground truth. WaSSI, like other models, produces simulated results. However, WaSSI integrates water balance, vegetation, and soil conditions, and may, therefore, provide the most plausible ET estimates. Like many other models, it has limitations and is constrained by the accuracy of model inputs. To better reflect the real on-the-ground conditions, we searched extensively through the literature, and found the currently available eddy covariance (EC) flux data mainly come from the following sources: Joshi et al. (2020) provided flux observation data from three sites: Kirtipur, Simara, and Tarahara, covering the period from March to November 2016 at a monthly temporal resolution. The surface energy balance closure ratios at these three sites were approximately 56%, 61%, and 64%, respectively. In this response, we provided a comparison between WaSSI and the observational results*

(Response Figure. 1) of Joshi et al. However, the Simara site is located beside an airport runway, and the underlying surface at the Tarahara site is rotation cropland (paddy June–October, wheat October–March, maize March–June), with significant anthropogenic disturbance, which limits its representativeness. Overall, the above limited site data have systematic deficiencies in the following dimensions, making it difficult for them to serve as an independent validation benchmark for WaSSI or remote sensing products. First, spatial representativeness is insufficient. The three low-altitude sites are concentrated with two in the Terai Plain and one (Kirtipur) in central hill and cannot represent the high mountain and Himalayan regions that account for most of Nepal's land area. Second, temporal continuity is poor. Joshi et al. (2020) covered only March–November 2016, the Langtang data cover only one year, and the Lirung data cover only 15 days. Without multi-year continuous observations, interannual variability cannot be assessed. In addition, anthropogenic disturbance is significant at all three sites. The underlying surfaces at the Simara and Tarahara sites are strongly affected by the airport and agricultural rotation, making it difficult to extrapolate the observation results to natural ecosystems.

Precisely because a single data source, whether site observations like eddy covariance flux or the WaSSI model, cannot provide a reliable 'ground truth,' this study adopted a multiple-constraint strategy, including eight remote sensing ET products (GLASS, GLEAM, PMLv2, REA, PEW, ETmonitor, SSEBop, SEBAL) to provide constraints on inter-model differences, the WaSSI model to provide constraints on mechanistic consistency, 12 basin-scale water balance datasets to provide water balance constraints, and the Budyko framework to provide physical boundary constraints on energy–water coupling. The core purpose of this framework is not to determine 'which ET value is absolutely correct,' but to identify which estimates fall within physically plausible boundaries. The study found that medium-sized basins exhibit severe water imbalance (observed runoff exceeds precipitation), and ET/P values deviate significantly from the theoretical Budyko curve. This deviation itself is regarded as an important diagnostic signal.

Therefore, Nepal is not entirely without site measurements, but the existing *observational data have fundamental deficiencies in spatial coverage, temporal continuity, and underlying surface representativeness, and cannot serve as an independent ground validation. Under the premise of acknowledging the "unavailability of ground truth," we used multiple constraints from remote sensing products, an ecohydrological model, basin water balance, and the Budyko framework to narrow the uncertainty range of ET estimates as much as possible. The essence of this strategy is a shift from "seeking the truth" to "identifying a plausible range," which is a pragmatic choice for conducting ET research in data-scarce regions.*



Response Figure 1. Comparison of WaSSI model results with the observations from Joshi et al. (2020)

6. REA is not a satellite product; it is a data fusion product.

Response: *Thank you for your suggestion. We have revised the relevant statements (Lines 323–326), and we also identified and corrected an expression error: REA is a fusion product derived from three model-based ET products.*

7. Equation 2 is missing the variable i.

Response: *Thank you for your reminder. We have annotated "i" in the text and specified its meaning. (Line 405 and 409 in revised manuscript)*

8. The time periods for water balance data, different remote sensing products, and WaSSI do not align.

Response: *Thank you for your valuable comments. The time periods indicated in Table 1 actually represent the time series covered by each dataset we collected. However, in the error analysis, the water balance results we calculated were matched according to the available time series. That is, the precipitation data cover the period 1980–2014, and the runoff data cover 1980–2015, so the water balance results we obtained are for 1980–2014. We then used this as the baseline to calculate the bias for each remote sensing product. In our bias analysis, we ensured that as many years of remote sensing data as possible were included. For example, the GLEAM data have a time series of 1980–2022, so we used the data from 1980–2014 to compare with our calculated water balance results. We will add an explanation of this in the revised manuscript to make our analytical process clearer. (Line 195-201 in revised manuscript)*

9. Spatial scales are mismatched. It is recommended to uniformly aggregate all data to either the watershed scale or the coarsest common resolution.

Response: *We agree that the different spatial scales used in the study were not sufficiently distinguished in the original manuscript. However, the different spatial levels were intentionally used to address different scientific questions rather than as interchangeable units of comparison. To make this framework clearer, we have revised the manuscript to explicitly define the purpose of each spatial level and to clarify where direct quantitative comparisons are made.*

The study uses four complementary spatial levels. Specifically, (1) 12 monitored watersheds were used to assess hydrological observation consistency and identify suitable watersheds; (2) 5 watersheds with acceptable water-balance closure were used for quantitative evaluation of the eight ET products and WaSSI against water-balance ET; (3) WaSSI simulations across 385 watersheds were used to examine broader watershed-scale spatial behavior and consistency with the Budyko framework; and (4) national gridded analyses were used to characterize spatial patterns, elevation-dependent differences, and long-term ET trends across Nepal. We have clarified this spatial hierarchy in the revised Methods and Results and have made the spatial unit *explicit in the relevant figure captions and descriptions (Please revise the manuscript accordingly). This clarification also prevents the results from the different spatial levels from being interpreted as direct when they serve different purposes.*

10. The time span used for trend analysis is too short.

Response: *Thank you for your valuable comments. The time spans of the data we collected are not uniform. For example, the REA data cover the period 1980–2017, while the SSEBop data cover 2003–2020. Given this inconsistency, we combined the time spans of multiple datasets to make the temporal coverage as uniform and as long as possible. Most of the remote sensing data can be unified starting from 2001, with the exception of SSEBop, which starts in 2003, and REA, which ends in 2017. Therefore, we chose 2001–2017 as the period for trend analysis. At the same time, when running the model simulations, we simulated the complete time series of evapotranspiration results according to the temporal sequence of the input data. For the SEBAL model simulations, we also applied linear interpolation between the results of two remote sensing image dates to obtain daily time-series data at a daily temporal resolution. Overall, we made our best efforts to ensure the completeness and consistency of the data.*

11. Numerous unit and language errors exist: "elevation <1500 mm" should be m; "annual ET 600–1100 m" should be mm. Other spelling errors include "Himalya," "Lessor Himalya," "Mann-Kendell," and "eitht."

Response: *Thank you for your reminder. We have conducted a comprehensive review and revision, including units and spelling, throughout the full text.*

12. It is suggested to rewrite the abstract to clearly distinguish observational results, model outputs, and mechanistic assumptions.

Response: *Thank you very much for your suggestions. We have thoroughly revised the abstract and restructured it into three parts: observational results, model output, and mechanistic hypotheses. This content will be available in the revised version.*

13. Quantitative analysis of bias sources is lacking.

Response: *We agree that a quantitative attribution of bias sources remains insufficient.*

In this study, we compared ET estimates primarily with basin-scale water balance estimates. We also considered quantitatively separating the contributions from model forcing, parameter calibration, and model physics. However, reliable in-situ observations are extremely scarce in Nepal, and the evaluated products differ in forcing data, model structure, and calibration strategies, making it difficult to isolate the internal causes of inter-product discrepancies. We, therefore, provide a qualitative attribution based on previous bias-analysis studies and on our own comparisons of forcing data and model mechanisms. A clear example is the contrast between SSEBop and SSEBop V6: SSEBop shows pronounced overestimation at high altitudes, whereas SSEBop V6 improves high-altitude ET performance by changing the input PET data, illustrating how input data accuracy can strongly affect product results. Nevertheless, we acknowledge that this does not constitute a full quantitative decomposition. Future work could include factorial sensitivity experiments, variance decomposition, or ensemble-based attribution to quantify the relative contributions of forcing, parameters, and model structure. We have added a statement in the Discussion to clarify this limitation and to outline these potential directions. (Lines 838-851 in revised manuscript)

14. GLEAM and SSEBop results appear to show little overestimation in high-altitude regions, which differ from the authors' conclusion.

Response: *Admittedly, under the Budyko framework constraints at the watershed scale, the results of SSEBop, SSEBop V6, and GLEAM products show little overestimation; however, compared with the other two products, SSEBop exhibits significant overestimation at the pixel scale. The main reason for this discrepancy is that, in mid-to high-altitude regions, some river basins have large vertical elevation ranges, with a single basin encompassing both high- and low-altitude areas. The basin-scale results are based on the mean values of pixels extracted according to different hydrological response units, which smooths out the evapotranspiration anomalies at high elevations within the basin. In contrast, the pixel-scale results are extracted pixel by pixel based on elevation data. Figures 4 and 5 in manuscript also show that the product results in these high-altitude areas exhibit notable fluctuations and overestimation. We believe* *that a corresponding explanation should be added to the Discussion section to improve the manuscript. (Line* 677-687 in revised manuscript)

Reference:

Lang, D., Zheng, J., Shi, J., Liao, F., Ma, X., Wang, W., Chen, X., & Zhang, M. (2017). A Comparative Study of Potential Evapotranspiration Estimation by Eight Methods with FAO Penman–Monteith Method in Southwestern China. Water, 9(10), 734. https://doi.org/10.3390/w9100734

Badayar Joshi, B., Ma, Y., Ma, W., & Wang, B. (2020, May). Seasonal and Diurnal Variations of Carbon Dioxide and Energy Fluxes over Three Land Cover Types of Nepal. In EGU General Assembly Conference Abstracts (p. 1820).