Reply on RC1: 'Comment on hess-2023-35', Anonymous Referee #1

Overall, the manuscript could be very helpful in providing a guideline for a calibrating model where almost no in-situ data is available. The authors discussed the in-depth methodology of their proposed calibration, providing a detailed analysis of the impact of parameter tuning on the performance metrics of the model. The authors also provided an analysis of the co-dependence of the hydrological and hydraulic model parameterizations and techniques to break the co-dependency. The central argument presented in the manuscript is using satellite data to infer river discharge against which the model-derived river discharge will be compared and the model recalibrated to achieve the desired accuracy. Sattelite data itself can cause a wide range of uncertainty in the calculation of river cross-section, water surface slope, and so on, especially in the Upper Mekong river basin, which is so complex in topography. The satellite data (used as a proxy of observation for calibration purposes) is prone to uncertainty that eventually impact the parameter tuning. So there's a need to strengthen the discussion by providing a detailed discussion on the uncertainty in the river discharge estimation from satellite data, without which the calibration framework could be questionable. My recommendation is for a major revision with the specific comments and concerns listed below:

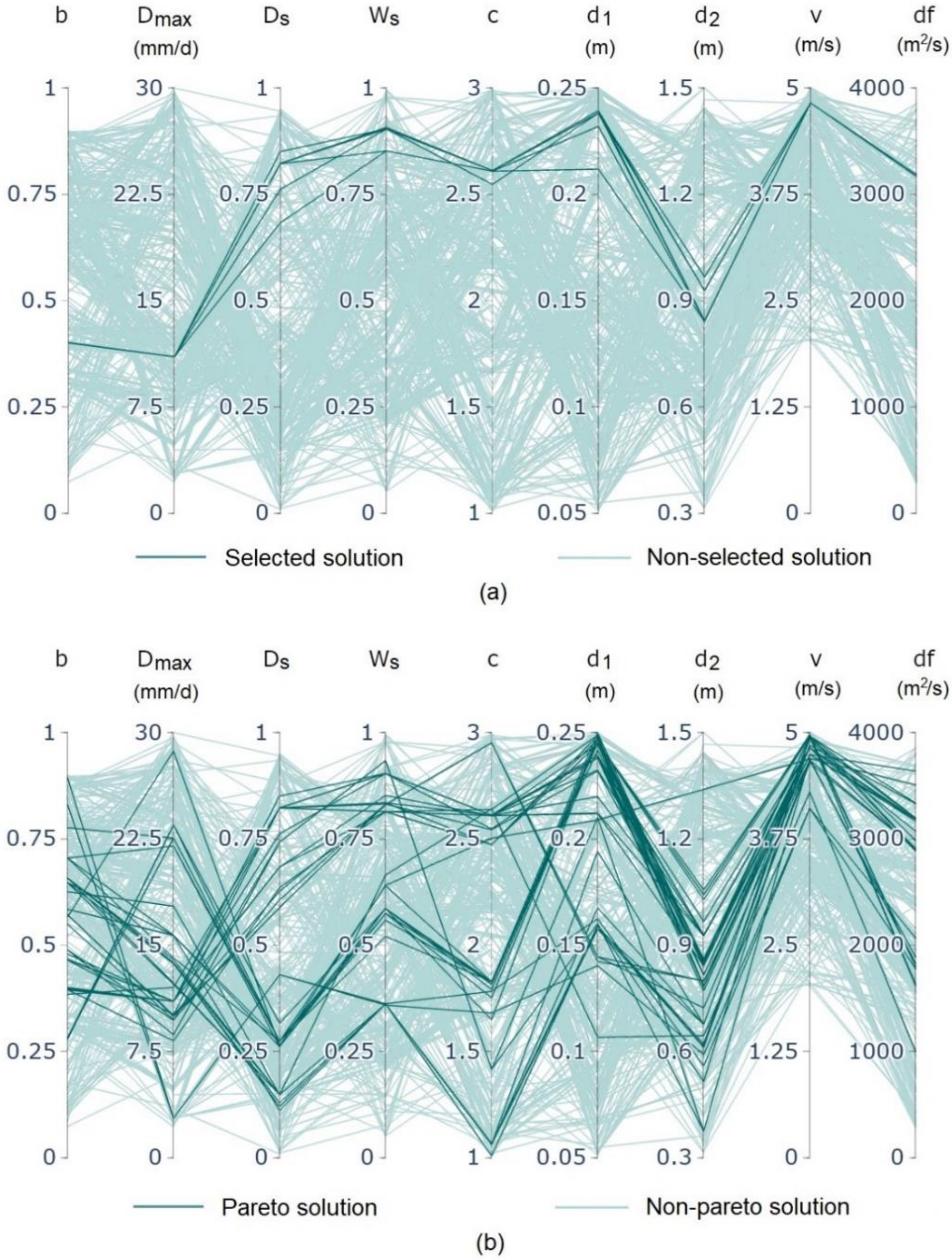
Response: Thank you for the positive feedback as well as the useful comments for improving the paper.

1. In the abstract, the authors mentioned that their approach could be readily transferable to another basin. However, the authors did not provide any convincing discussion on how the same framework can be used for another basin. For example, what cautions should other researchers follow when applying the same technique to a highly complex basin with rugged terrain or complex topography? As the estimation of river cross-section and water surface slope could be challenging/more uncertain in some other basin.

Response: We agree with you. We will add this point to Section 5. Specifically, we will discuss about two key elements concerning the application of our approach (to other basins), namely river cross-section estimation, and water surface slope estimation.

2. In section 3.3.1 authors discussed the calibrated model parameters and presented the calibration outcomes later. However, one would expect to see a discussion of the calibration of the most sensitive parameters.

Response: In Section 3.1.1 (Sensitivity analysis), we stated that we carry out a Global Sensitivity Analysis to study the relationship between the performance of VIC-Res and the parameterization of the rating curve (Manning's coefficient). To do that, we investigated a total of 10 model parameters, including 7 soil parameters of the rainfall-runoff module, 2 parameters of the routing module, and the Manning's coefficient. Therefore, in the corresponding result section (4.2, Sensitivity analysis), we focused on providing results related to Manning's coefficient and VIC-Res performance metrics, as they present the co-dependence we are interested in. However, we also showed the results of VIC-Res parameters (Figure 7) and discussed them in the second paragraph of Section 4.2.2. We will expand *the discussion on the model parameters in this sub-section. We will also include an analysis of the parameters after the calibration process in Section 4.3 (please refer below for a preliminary version of the analysis).*



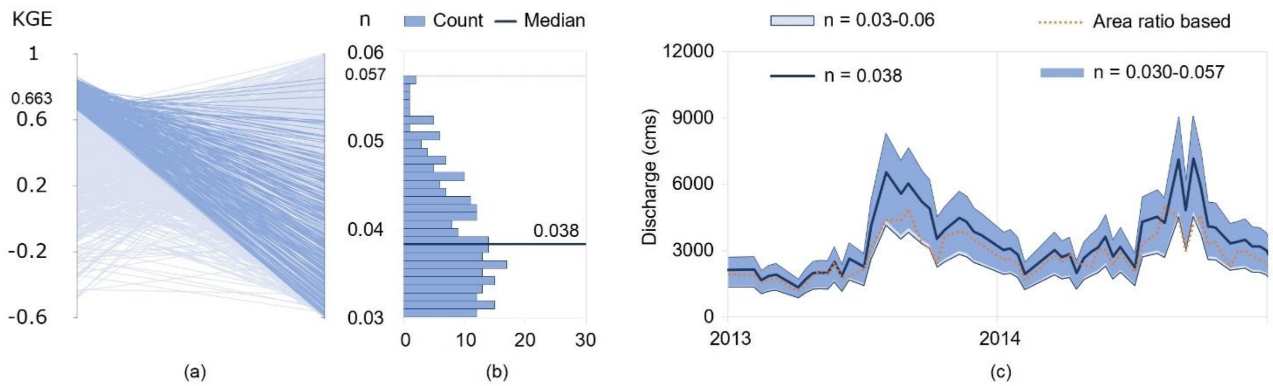
Parameters of the solutions from model calibration. In panel a, solutions were selected by intersecting the four top 25% parameterizations for each performance metric. In panel b, selected solutions are the Pareto solutions.

3. As the discussion is so central to the simultaneous calibration of the using RS Discharge (from satellite data). Thus an uncertainty analysis of the estimation of the river cross-section or uncertainty in the RS Discharge from satellite data is extremely necessary. It directly impacts the RS Discharge estimation against which VIC-RES discharge is compared to calculate performance matrices. Thus uncertainty in RS Discharge can substantially impact the calibration process and parameter tuning, fundamentally questioning the Novel technique the authors suggested in that manuscript. I recommend a discussion on uncertainty in RS Discharge estimation and how it may affect the calibration process. Although the author provided some insights in section 4.2.3

Response: Besides some insights provided in sections 4.1.3 (we believe that you meant Section 4.1.3 instead of 4.2.3), we discussed the uncertainty in RS discharge in the third paragraph of Section 5. This said, we agree with you that this point is important, so we will proceed by expanding and strengthening it.

4. In section 3.3.2- The authors discussed that they used multiple performance matrices to cover a different aspect of modeling accuracy. However, the use of KGE as a performance metric is also suggested, as it considers bias, correlation, and variability.

Response: Thanks for your suggestion. Here, we provide the result using KGE. Noticeably, using KGE yields results similar to those obtained with NSE. In particular, the top 25% parameterizations w.r.t. KGE (250 samples) have 208 parameterizations in common with those of NSE, thereby resulting in similar narrowed ranges of Manning's coefficient (n) and RS discharge. The reason for this result is probably due to the fact that KGE is based on a decomposition of NSE into correlation, variability bias, and mean bias components (Gupta et al., 2009). Because of the similarity in results obtained with NSE and KGE, we decide to exclude the ones associated to KGE in our analysis.



In panel a, the dark blue lines highlight the parameterizations yielding the top 25% performance (highest KGE). The histogram in panel b illustrates the frequency distribution of n corresponding to the top 25% parameterizations with the median depicted by the dark blue line. In panel c, the light blue envelop is the range of variability of the discharge estimated with n ∈[0.03,0.06], while the dark blue envelop is the range corresponding to the top 25% performance. The black lines are the discharge corresponding to the median values of n, while the orange dotted line is the discharge estimated from observations at Chiang Saen via the area-ratio method.

5. In Figure 8: could you discuss why there is less variability in 2009-2012 and after that, there is considerable variability, particularly in the low flows?

Response: This is a very good point. First, note that the cascade dam system in the Upper Mekong modified the natural flow downstream, it increased low flows (Vu et al., 2022). The change can be seen most clearly since 2013, when Nuozhadu—the largest reservoir in the Upper Mekong—became operational. This change is captured in the altimetry water level, which we use to convert to RS discharge through the rating curve. On the other hand, as shown in Figure 5b, when converting from water depth (calculated from water level) to discharge, the higher the value of water depth, the wider the discharge variability. That explains the considerable variability in RS discharge of 2013-2018 compared to the one of 2009-2012. We will add this point to Section 4.1.3.

6. In Figure 9: the timing of the peak is missed in some of the years, e.g., 2007. You can just add a discussion on the sensitivity of different parameter tuning in capturing timing/seasonality or peak. Or which is the most sensitive parameter?

Response: Overall, the RS, simulated, and area-ratio discharges at the virtual station (Figure 9a) show similar behaviors in the time-to-peak, so we believe you refer to the comparison against the observed discharge at Chiang Saen (Figure 9b), where some discrepancies in the time-to-peak emerge (e.g., in 2014 and 2017). These discrepancies could be due to different factors. First, and most important, note that we calibrated our model with RS discharge at the virtual station and then validated it with observed discharge at Chiang Sean. Another reason could be the uncertainty due to the use of gridded precipitation data (Kabir et al., 2022). We will elaborate on these points in Section 4.3.

7. Also. How can the hydrological response unit's resolution or size impact calibration? It can impact the calibration substantially. For example, the Lancang river basin is so narrow and elongated. Thus, the use of a coarse hydrological response unit of Coarse-resolution may accurately impact the identification of river grid cells.

Response: The hydrological response unit's resolution (or size of distributed/grided hydrological models) could affect the rainfall-runoff and routing estimations, and thus affect simulated discharge and model calibration (Egüen et al., 2012). Looking at the existing distributed models for the Mekong region, we note that our cell size (i.e., 0.0625°) somewhat falls in between w.r.t. what is currently being adopted—for example, Costa-Cabral et al. (2007) and Tatsumi & Yamashiki (2015) adopted a resolution of 1/12° and 0.25°, while Du et al. (2020) and Bonnema and Hossain (2017) used a resolution of 90m/900m and 0.01°, respectively. We agree that this point is important, but also believe that it applies to 'any' modelling exercise, not only to those relying on remotely-sensed data, like ours. Because of this reason, we feel that adding a thorough analysis on the impact of model resolution would go beyond the goal of our study—and potentially confuse its main message. We thus suggest including a discussion on this point in Section 5.

8. Figure 2 and 3 can be merged together. Having two figures does not add much value to the discussion.

Response: We agree with you on this point. We will condense Figure 2 and 3 into a new figure.

9. In section 3.2.1: The authors used a regression technique (sixth-degree polynomial) to fit the data point best. However, the author said that is best works for the natural condition of rivers. AS MANUSCRIPT TITLE, the authors mentioned "Heavily Regulated Basin." One would like to know the author's novel technique for heavily regulated basins. In the suggested numerical framework, I think I do not see any strong linkage of the reservoir operation (like heavy regulation) with the calibration/parameter tuning. Or could you provide an explanation of how your technique is mainly applicable to the heavily regulated basin? Or it may be more justifiable to say Novel calibration technique for the poorly gauged basin.

Response: Thank you for raising this point. Let us begin by clarifying a point that is perhaps at the origin of this comment: As stated in the title of the manuscript, our framework is developed to calibrate models in heavily regulated basins (that is, where river discharge is modified from its natural flow by man-made reservoirs). Part of this framework includes a method to construct the river cross-section at a virtual station (Section 3.2.1). We stated that this specific method works best for riverbanks (topography) in natural conditions. We do not see any conflict between our overall intent and the specific method used to infer discharge, because the term "heavily regulated" is used to describe flow regime while the term "natural conditions" is used to refer to the river topography at a specific location.

Moving to the choice of the manuscript title (as well as our contribution), we would like to begin by noticing that reservoir operations could affect the model parameterization. Calibrating hydrological models with and without the representation of reservoirs could result in different sets of model parameters, even though they both have good model performances (i.e., comparable simulated and observed discharges). This is because, when reservoir operations are not included, model calibration adopts parameterizations that compensate for the absence of the reservoirs (Dang et al., 2020). In our case study, the discharge (observed or inferred from satellite data) includes the flow modification caused by reservoirs, so the reservoir operations must be included in the model. Specifically, we do that by integrating the reservoir operations estimated from satellite data (to replace the measured data due to their unavailability) into the hydrological model. To the best of our knowledge, this specific approach is still at its infancy in the hydrological modelling domain, so this is why we would like to retain the term "heavily regulated basins" in the title.

Finally, we note (and agree with the reviewer) that our approach is not only applicable to regulated basins. In fact, the reservoir operation module could be switched off when working on a natural catchment. We will elaborate on this point in Section 5.

References

Bonnema, M. and Hossain, F.: Inferring reservoir operating patterns across the Mekong Basin using only space observations, Water Resources Research, 53, 3791-3810, https://doi.org/10.1002/2016wr019978, 2017.

Costa-Cabral, M. C., Richey, J. E., Goteti, G., Lettenmaier, D. P., Feldkötter, C., and Snidvongs, A.: Landscape structure and use, climate, and water movement in the Mekong River basin, Hydrological Process, 22, 1731-1746, https://doi.org/10.1002/hyp.6740, 2007.

Dang, T. D., Chowdhury, A. F. M. K., and Galelli, S.: On the representation of water reservoir storage and operations in large-scale hydrological models: Implications on model parameterization and climate change impact assessments, Hydrology and Earth System Sciences, 24, 397–416, https://doi.org/10.5194/hess-24-397-2020, 2020.

Du, T. L. D., Lee, H., Dui, D. D., Arheimer, B., Li, H., Olsson, J., Darby, S. E., Sheffield, J., Kim, D., and Hwang, E.: Streamflow prediction in "geopolitically ungauged" basins using satellite observations and regionalization at subcontinental scale, Journal of Hydrology, 588, 125016, https://doi.org/10.1016/j.jhydrol.2020.125016, 2020.

Egüen, M., Aguilar, C., Herrero, J., Millares, A., and Polo, M. J.: On the influence of cell size in physically-based distributed hydrological modelling to assess extreme values in water resource planning, Natural Hazards Earth System Sciences, 12, 1573-1582, https://doi.org/10.5194/nhess-12-1573-2012, 2012.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of Hydrology, 377, 80-91, https://doi.org/10.1016/j.jhydrol.2009.08.003 2009.

Kabir, T., Pokhrel, Y., and Felfelani, F.: On the precipitation-induced uncertainties in process-based hydrological modeling in the Mekong River Basin, Water Resources Research, 58, e2021wr030 828, https://doi.org/10.1029/2021wr030828, 2022.

Tatsumi K. and Yosuke Yamashiki, Y.: Effect of irrigation water withdrawals on water and energy balance in the Mekong River Basin using an improved VIC land surface model with fewer calibration parameters, Agricultural Water Management, 159, 92-106, http://dx.doi.org/10.1016/j.agwat.2015.05.011, 2015.

Vu, D. T., Dang, T. D., Galelli, S., and Hossain, F.: Satellite observations reveal 13 years of reservoir filling strategies, operating rules, and hydrological alterations in the Upper Mekong River basin, Hydrology and Earth System Sciences, 26, 2345-2364, https://doi.org/10.5194/hess-26-2345-2022, 2022.