Reply on RC2: 'Comment on hess-2023-35', Andrea Galletti,

The authors propose an interesting framework for exploiting satellite data for the calibration of large-scale hydrological model. The steps undertaken to translate satellite-derived information into are explained in thorough detail. Then, the co-dependence between the hydrological model's performance and the calibration of the hydraulic model based on remote-sensed data is analyzed via Global Sensitivity Analysis. Finally, the authors adopt the satellite-derived information to calibrate VIC-Res and validate the predictions against available ground observation.

The concept is promising, however it would benefit from some considerations on its general applicability were drawn. Furthermore, the densely-flowing work sometimes overshadows the connection between the overall proposed framework and the particular tool/result being explained. I recommend acceptance of the paper, provided that the authors address or clarify the following points (minor revision):

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

1. The paper addresses the issue of calibrating macro-scale models in ungauged catchments. The introduction seems to implicitly assume that every large-scale hydrological model is always calibrated, and that this is most times done against available streamflow time series. However, several large scale hydrological models do not actually undergo case-specific calibration. While a calibrated model can provide results that are generally more reliable or realistic, this might not be clear to every reader, or someone could disagree. I recommend addressing the importance and effectiveness of calibration at the beginning (lines up to 15).

Response: We agree with you. We will make this point clear at the beginning of Introduction.

2. The subsequent paragraph (line 17 onwards) could also benefit from a more thorough introduction to models' calibration and what is needed. In particular, it is not clearly explained why and how the presence of hydropower infrastructures for which operations are not known can create pitfalls for model calibration, nor why could it be at all important to include such factors in a calibration of hydrological model's parameters.

Response: We agree with you. We will address these points in the Introduction. Specifically, we will explain how the presence/absence of hydropower infrastructures (reservoir operations) affects model parameterization and the importance of including such factor in the model calibration process.

3. Some references and examples should be provided to support the sentence "Yet, this approach may still partially rely on in-situ data": this sentence opens up to the second class of RS-based works and basically justifies the study, therefore it needs to be backed by clear examples.

Response: We provided some references/examples in the sentence right before (line 34). We realized that our expression could cause a misunderstanding. We will revise it accordingly.

4. Following point #1, I feel like a non-trivial question is whether it is possible and helpful to calibrate large scale hydrological models in ungauged catchments. This opens up to research questions I, II and III, helping to place them in a context that is broader than a mere numerical exercise.

Response: Thank you for this comment, which pushed us to think more about the broader context of our work. As already pointed out in point #1, several large-scale hydrological models do not actually undergo case-specific calibration. We know that this can often lead to poor model performances. Hence, calibration has been suggested as a key avenue for model improvement---whenever computationally feasible (e. g., Bierkens et al., 2015; Samaniego et al., 2017). However, a prerequisite for a "successful" calibration is the availability of "good quality" data (i.e., at a minimum, "sufficiently long" time series of discharge observations and with "sufficiently small" observational errors). In many regions of the world, such good quality data are not available from in-situ observation networks, so the question arises whether remote-sensed data, despite all their uncertainty, are still "good enough" to support model calibration. Hence, the overarching question that this paper addresses is: to what extent is it possible and helpful to calibrate a large-scale hydrological model in ungauged catchments using remote-sensed data? How do we deal with potential interactions between parameters used in data pre-processing (i.e., from remote-sensed data to reconstructed discharge data) and parameters of the hydrological models when doing model calibration? Can we reduce uncertainty in model calibration results by properly taking into account such interactions? We will make sure that the overarching question underpinning our work is clearly illustrated in the Introduction.

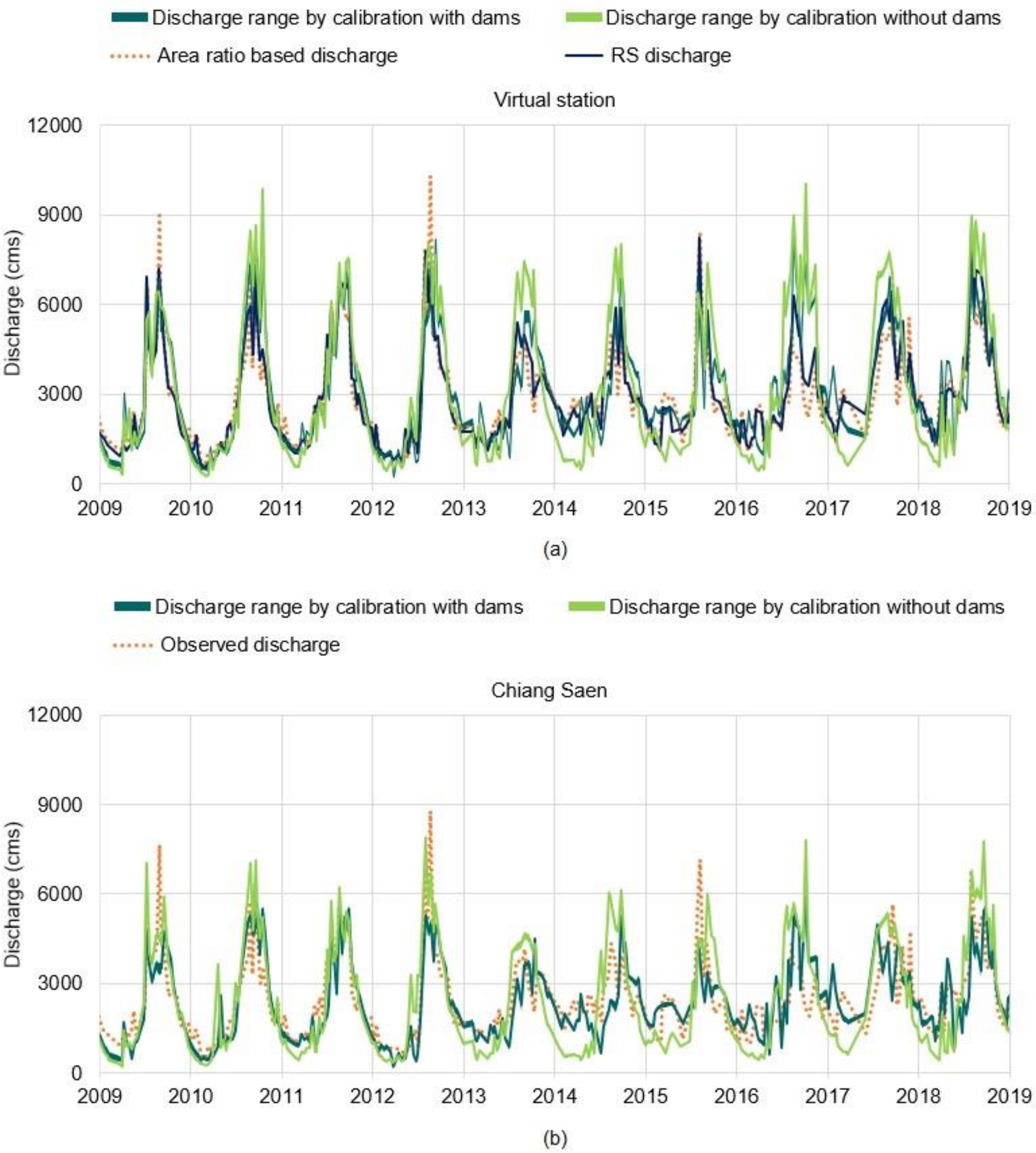
5. The choice of the 2009-2018 time window is not motivated as of line 84. I suspect the decision was taken backwards for compatibility with data coverage and quality (Figure 3). If this is the case, it should be stated clearly and the corresponding statements (line 96 and caption of fig 3) adjusted. If the choice was driven by something else, it should be pointed out.

Response: Yes, the reason is not only to include the filling period of the two largest reservoirs (as stated in line 83-84) but also to account for coverage and temporal resolution of altimetry data. We will make these points clearer.

6. I am dubious about the reliability of interpolating (linearly) daily reservoir operations between two monthly values. Hydropower often operates at the daily scale (or lower). The configuration of cascading dams might help masking errors in this procedure and their repercussions on the performance metrics. Did the authors compare their results with a fully-natural setup in order to evaluate the impacts of the reconstructed reservoir operation?

Response: To answer this comment, let us first elaborate on how reservoirs in the Upper Mekong River have been designed and operated. Xiaowan and Nuozhadu are the two largest reservoirs: they have a massive capacity (~36 km^3) and account for about 85 % of the total system's storage. Because of their size, their role is not to follow inter and intra-daily electricity demand variability, but rather to ensure a stable supply of power and to minimize the variability in the production of the other dams composing the hydropower system. This goal is reflected by their operating patterns. In the wet season (June-November), Xiaowan and Nuozhadu reservoirs gradually store water until reaching their maximum operational level (and release extra water if necessary). The other reservoirs run at their normal operational level (full capacity for power generation). In the dry season (December-May), Xiaowan and Nuozhadu gradually release water to the downstream reservoirs to ensure that the other reservoirs can run at their normal operational level. Putting this information together, we believe it is fair to state that Xiaowan and Nuozhadu are characterized by slow-varying dynamics (International Rivers, 2014; Vu et al., 2022). Hence, basing our analysis on interpolated monthly values is reasonable, although not ideal, of course. Our argument is partially supported by the analysis carried out in Vu et al. (2022), where we successfully compared the monthly storage of Xiaowan and Nuozhadu derived from Landsat images against the storage derived from Jason altimetry data (10-day temporal resolution) and Sentinel-1/2 images (6-day temporal resolution). Because of the spatial and temporal coverages of those data, we use the result derived from Landsat images for this study. We understand that all the information mentioned above is important. We will add it to Section 3.1.2.

A comparison between the river discharge simulated with and without reservoirs is technically feasible (see the figure below). However, we do not fully understand how this comparison can help evaluate the process of interpolating daily reservoir operations between two monthly values, and thus we believe that the explanation provided above is more compelling. Hence, we suggest including this in the response to reviewers but not in the revised manuscript.



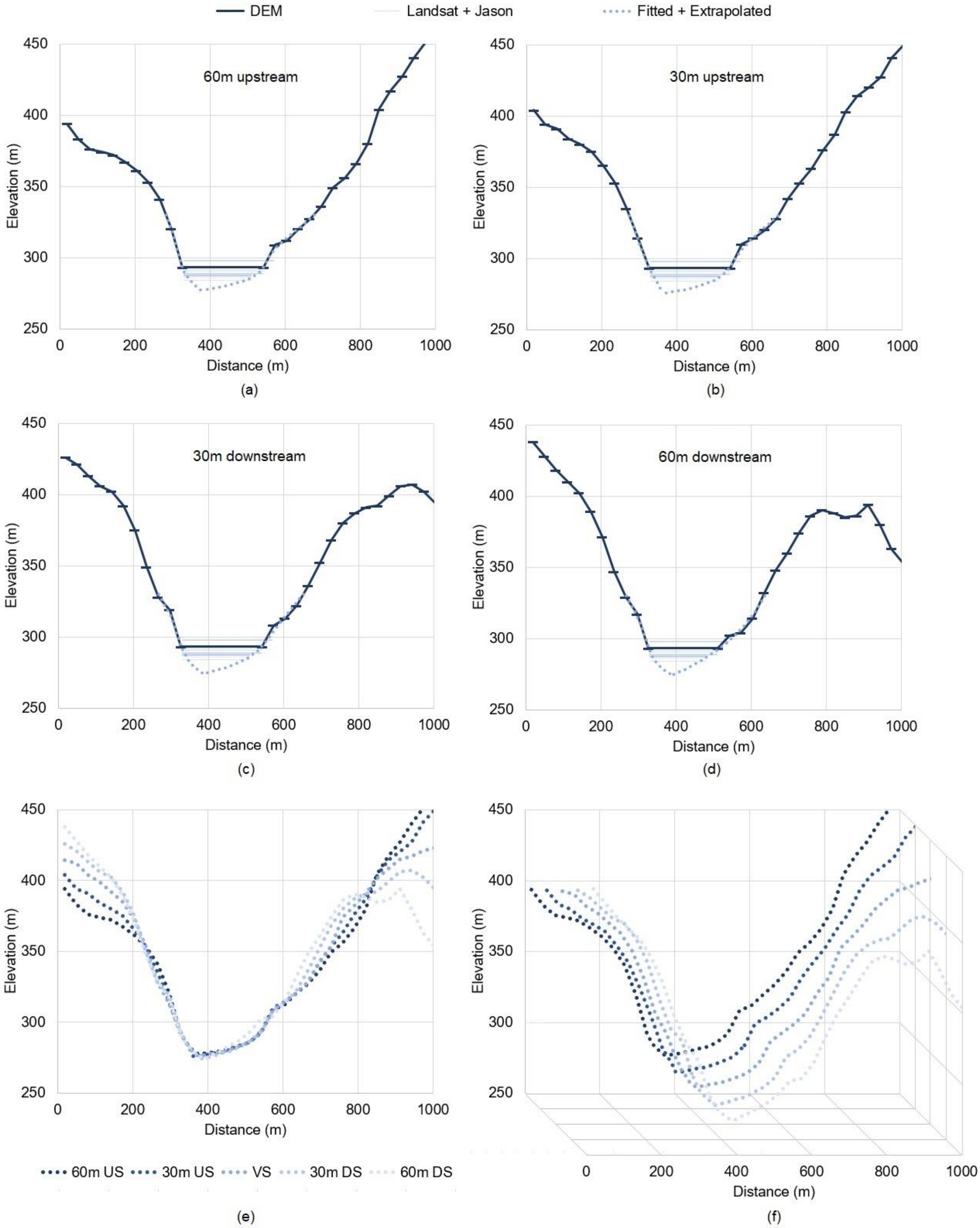
Comparisons between simulated discharges with dams (dark green) and without dams (yellow green) at the virtual station (a) and Chiang Saen station (b). The variation ranges of both two settings (with and without dams) are corresponding to the twelve selected solutions in calibration (explained in Section 4.3).

7. The estimation of the river cross section represents an important step in the framework, influencing the outcomes of the rating curve and all subsequent analyses. With the river width ranging from 400m to 200m depending on water level, and with DEM and Landsat images having a resolution of 30m, one could think that the combination of cells taken as the cross section is not univoque. Exploring multiple alternatives for the chosen cross section could help assessing the variability of this extrapolation (if at all present). Otherwise, a plane view or schematic of the chosen pixels would help understanding the decision taken. Furthermore, river bed might be subjected to vertical evolution (i.e., erosion or sediment accumulation) which is hardly grasped by this methodology since it relies on a static DEM. Performing a similar extrapolation (of the river cross section) e.g.,at location 812 which is closer to Chiang Saen and then comparing the so-obtained discharge values with ground observations would have provided a rather solid, yet not exhaustive, base for any assumption made at Virtual Station, reinforcing the implicit validation described at lines 267-269.

Response: Thank you for the suggestion of exploring multiple alternatives for the chosen cross-section. In the panels a-d of the figure below, we show the results of four alternative cross-sections, created by moving the one at the location of the virtual station (reported in the manuscript, marked by VS in the panel e and f) 30 and 60m (1 and 2 cells) both upstream and downstream. The alternatives are well in agreement with the one reported in the manuscript. Specifically, riverbed elevations are 277.2, 275.6, 276, 274.5, and 274.3 m a.s.l. (from upstream to downstream). We plot the five cross-sections together in panel e, and show a 3D visualization in panel f. We plan to add this part to the Supplement.

We agree with the second point that riverbeds might be subjected to vertical evolution, which is hardly grasped by a static DEM. We will discuss this point as one of the limitations of the methodology (in Section 5).

As for the last point, we believe that applying our methodology to location 812 and other locations which are close to Chiang Saen station may not be possible. This is due to a number of issues: (1) the temporal resolution (35 and 27 days for Envisat and Sentinel-3A respectively) and coverage (see Figure 3) of altimetry data is too coarse; (2) Manning's equation works best for straight river segments with limited discharge variations due to tributaries and distributaries nearby (Przedwojski et al., 1995), while most locations close to Chiang Saen are located in curved river segments, where many tributaries join the mainstream of the Mekong River; and (3) modifications of the riverbank topography (for road and agriculture/farming) around most of those locations affect river cross-section estimations. We also plan to elaborate on these points in Section 2.3 and Section 5, where we will expand the discussion on the limitations of our methodology.



Panels a-d show the results of four alternative cross-sections, created by moving the one at the location of the virtual station (reported in the manuscript, marked by VS in the panel e and f) 30 and 60m (1 and 2 cells) both upstream and downstream. The five cross-sections are plotted together in panel e, and a 3D visualization is provided in panel f.

8. Line 338 onwards: the selected 25% is said to be the best according to all metrics: does this mean that the (i assume normalized, as defined in the methods) metrics were averaged and the best 25% averages were taken? This concept could be made a bit clearer. Furthermore, it is unclear what determined the choice of the 12 and 58 stations presented later in this section.

Response: To select the solutions we first created four groups (each containing the top 25% solutions w.r.t. a single performance metric) and then took their intersection. This approach yields 12 solutions. The other approach we considered is the definition of Pareto efficiency, which yields 58 solutions. We will make sure this point is crystal clear to avoid any misunderstandings.

9. Line 359: Phrasing here is a bit misleading, and a more thorough introduction could be useful: GSA was used to determine the co-dependence between n and the model's performance, while the potential bias intrinsic to the co-dependence was not really assessed, although one could reasonably suppose that it is present. Providing further evidence (references) that joint estimation of discharge and parameters leads to a biased calibration could help strenghten the need to break the co-dependence.

Response: Yes, we agree with you that this part is potentially misleading, so we will revise it accordingly. We note that the existing literature does not look specifically at the problem of biased calibration (when co-calibrating a hydraulic and hydrological model). In fact, this is, to the best of our knowledge, the first study that looks into the pitfalls of such co-calibration process.

TECHNICAL CORRECTIONS:

Line 227: is 2008-2018 correct? it seems it should be 2009
Response: Yes. You are right. It is 2009-2018.

Line 297: the value in parentheses should be 0.045
Response: Yes. You are right. It is 0.045.

Line 382-383: do not contribute to (reducing) modelling uncertainty?
Response: It is "could contribute to modelling uncertainty".

Line 393: One often recurring... one should be removed?
Response: Yes, we will improve the readability of this sentence.

References

Bierkens, M. F. P.: Global hydrology 2015: State, trends, and directions, Water Resources Research, 51, 4923-4947, https://doi.org/10.1002/2015wr017173, 2015.

International Rivers: The Environmental and Social Impacts of Lancang Dams, https://archive.internationalrivers.org/sites/default/files/attached-files/ir_lancang_dams_researchbrief_final.pdf, last access 11 April 2023, 2014.

Przedwojski, B., Blazejewski, R., Pilarczyk, K.: River training techniques: Fundamentals, design and applications, Netherlands: Taylor & Francis, 1995.

Samaniego, L., Kumar, R., Thober, S., Rakovec, O., Zink, M., Wanders, N., Eisner, S., Schmied, H. M., Sutanudjaja, E. H., Warrach-Sagi, K., and Attinger, S.: Toward seamless hydrologic predictions across spatial scales, Hydrology Earth System Sciences, 21, 4323-4346, https://doi.org/10.5194/hess-21-4323-2017, 2017.

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.