We thank the reviewers and the editor for their insightful comments and suggestions to improve the manuscript. Below, we copy reviewer comments in black font and our answers in blue font.

Editor comments

Comments to the author:
Dear Authors,

the referees have overall appreciated your work but some important clarifications on the model and methods (see model structure, water surface identification method and water occurrence thresholds) and more details on the analysis (see uncertainty analysis) are needed.

The paper has been reorganized to make it clearer.

Regarding water surface identification and thresholds, more details have been included in sections 4.1.3 and 4.1.4.

Regarding the uncertainty analysis, more details have been added for the model uncertainty in section 5.2. In addition, we run an additional calibration to test the sensitivity of the model to the distribution of lateral inflows, assuming uniform lateral inflow (see section 4.4, 5.2 and 5.2.3)

You also need to add more background and expand the state-of-the-art, in order to fully clarify the novelty of the work (see Ref#2 comment, that requires also a reference to the ICESat applications, see Bjerklie et al. 2018).

More background and state-of-the-art have been added to the introduction and discussion based on the recommended literature.

Bjerklie et al. 2018 uses ICESat for slope measurement, while in our study we use ICESat-2 for WSE, river width and river cross-section geometry from elevation points in the surroundings of the river main channel. We use the same approach to interpolate the submerged portion of the river that is the one proposed by Dingman.

In your replies you have already described where you are planning to revise your manuscript, and I ask you to address all the referees' comments. In the reply letter, please summarise all the changes you will have made, clarifying exactly how and where, in the revised paper you have addressed each comment.

I look forward to reading your revised manuscript (that I will ask both referees to revise).

Best regards,
Elena Toth

Referee nr1

The paper is a good example of how to use ICEsat-2 satellite altimetry data in a hydraulic model built without in situ topographic surveys of cross-sections. I found the paper interesting and adapt to the HESS journal. I think it deserves to be published after minor changes.

- In the analysis I have only one main concern that relates to the structure of the model without considering large tributaries. My main concern is the fact that you consider the boundary condition as uniformly distributed, but in fact from Figure 4 it is quite evident that there are at least two significant inflows from the two left tributaries (it is not clear whether the right tributary comes upstream or downstream of Jimay station). How did you simulate these significant inflows?

We do not assume a uniform inflow boundary in the HD model. We assume uniform runoff in the contributing area between the two stations. The uniform runoff is then distributed along the chainage according to the flow accumulation map derived from the MERIT DEM. The HD model uses a flow accumulation map along the river stretch to calculate the boundary inflow values at the different chainages in the river. Thus, boundary inflows will be high in chainage intervals that include major tributary junctions.

To compare the performance using the flow accumulation map, we have also added a calibration assuming uniform lateral inflow along the chainage.

- Perhaps the organization of the article can be revised to avoid repetition and short paragraphs. In particular, the description of the ICEsat-2 dataset and the study area follow the method, but they are actually mentioned again and again to explain the different steps. Therefore, I think the best solution is to describe the material first (satellite and in situ dataset, study area) and then the method so that the reader is able to understand why the two satellite products and the hydraulic model scheme are used.

The study area description is moved to a new section (sec. 2 Study Area). The datasets are described in a new section (sec. 3 Data). The methodology is revised to reduce redundancy and described in section 4 Methods.

SPECIFIC COMMENTS:

- Introduction: references on hydraulic simulations should include at least one of the studies conducted by Domeneghetti et al. (2014, 2015, 2020); references on altimetry densification should include the publication of one of the authors Nielsen et al. (2022).

  Domeneghetti et al. 2014 is added as a reference on hydraulic simulations. Nielsen et al. 2022 is added as a reference for altimetry densification methods.

- Lines 119-125: the numbers are rather arbitrary. Please, justify the reason for these thresholds in the main text (93% for the water occurrence; 15 m from the river center-line; 500 m distance between observations and less than 15 times…).

  The water occurrence threshold is selected to ensure the consistency of the retained data. The threshold is conservative because we only accept data over very high water occurrence pixels. Water occurrence of 95% is only exceeded in very few pixels.

The 15 m from the river centerline are taking into account the minimum river widths to be sure the points fall on the water surface (min river width is 80 m).

The 500 meters distance is selected taking into account the spatial resolution of the HD model, which currently is 300 m and 100 meters for selected areas.

This has been better explained in the section, justifying the selection of the different thresholds. This part corresponds to section 4.1.3

- Lines 132-138: I'm not sure I understand these lines. What is meant by "the reference water surface elevation of the cross-section changes "? If it is a reference, it should be fixed. And what is meant by "the change in flow rate is added to the corresponding depth of the cross section"? How can a flow discharge be added to a depth? Please, rephrase the sentences so that they are clear.

This paragraph has been rewritten for clarification in section 4.1.4: "Cross-sectional data from ATL03 is selected for the available dates in the low-flow season. Depending on the acquisition dates, the discharge varies, and accordingly, the WSE measured by the ATL03 product. This can lead to unreasonable variations in the bottom elevation when constant depth is assumed leading to simulation errors. To avoid such errors and have a realistic reference bottom level, we define a reference WSE. The reference WSE is taken from the ATL03 cross-section in which the acquisition date has the lowest discharge value from the in-situ data. The discharge variation is calculated between the lowest discharge value and the discharge values for ATL03 cross-section acquisition dates. Since the discharge variation should be proportional to the WSE variation, we define a correction factor α relating discharge variation to WSE variation as $\Delta WSE = \alpha \cdot \Delta Q$. The correction factor α is a calibration parameter, and we add $\alpha \cdot \Delta Q$ to the corresponding cross-section depth value."

- Line 141: it is not clear why the two products ATL03 and ATL13 are shifted of 41 cm. Is this explained somewhere in the manuscript? If not, please can the authors add the reason of this bias?

The reason for the bias is explained in 4.1.4 more clearly: "The method we use to detect the WSE in the ATL03 product is different from the one used to produce ATL13 product, hence we can find variations between our detected WSE and the one provided by ATL13 product for the same acquisition date. An example of the difference between WSE detected in ATL03 and ATL13 WSE can be seen in figure 4 (e), where there is an offset between the two products of about 0.41 m."

- Lines 142-148: Please, explain the concept better. It is not clear why you are removing the red dots in Figure 3a that correspond to the zero change in discharge, or the dot at 400. In fact, I do not understand the logic of these analyses. Perhaps, they deserve more detail in the text.

In this analysis, we look at the WSE observations from ATL13 and the extracted WSE from ATL03 using our method. We compare the ATL13 and the ATL03 values when they are within 300 meters. If the acquisition date for the compared ATL13 and ATL03 products are the same, we expect the WSE variation to be zero and the same with the variation of discharge, however, we observe a bias in some cases in the WSE variation, since we use a different method to extract the WSE from ATL03 from what is used in the ATL13 product. If the ATL13 and ATL03 products are from different acquisition dates, we check whether the variation of WSE makes sense with respect to the variation of discharge. They should be related monotonously. If we observe low variation of discharge we expect low variation in WSE and vice-versa.

This part has been better explained and connected also to the previous comment since it is related to the analysis. The new text can be seen in section 4.1.1

- Line 212: remove “the”

  It was removed

- Looking at Figure 5 the shape of some cross-sections looks rather unrealistic (c,d,e,f). The river bottom looks high (shallow) and this could affect all the analysis. Do you have information on the topographic survey of some cross sections that could help to understand how much error is in the bottom estimate?

  The data corresponds to the low flow season, where the depths in this portion of the Yellow River varies between 1-2 meters. Unfortunately, we do not have access to in-situ surveys from the region.

- Please, define all acronyms: e.g. RHS, UPA, Obj

  They refer to Right-hand-side, upstream area, and objective function.

  The acronyms have been defined in the manuscript

- It is not clear why the paragraph 2.4.2 is described here and what is the role after. Try to explain why you are using the MIKE Hydro model.

  This is presented as one possible application of the presented cross-section retrieval workflow. The calibrated cross-sections can be used in a full-hydrodynamic simulation. In our case, the hydrodynamic simulation is made with MIKE hydro using the cross-section and calibrated parameters. We have improved this discussion in the revised manuscript

  This has been better explained. MIKE hydro model is the hydrodynamic model used to demonstrate the performance of the calibrated parameters in the transient state which is a use case of the method. Section 2.4.2 (4.4.2 in the revised manuscript) is related to the results in section 5.2.4.

- Line 290-294: For a reader who is not thoroughly familiar with the satellite product, this sentence is difficult to understand. Please try to explain what is the difference between weak and strong beam data. Also, since this is a product feature, I think it can be moved to a methods section.

  The sentence has been reformulated and moved to method section 4.1.1. In addition, clarification of the ICESat-2 ground track configuration has been added in material section 3.1.

- Line 296-297: please, add the references for these distances (e.g. cross-section chainage and longitudinal distance.

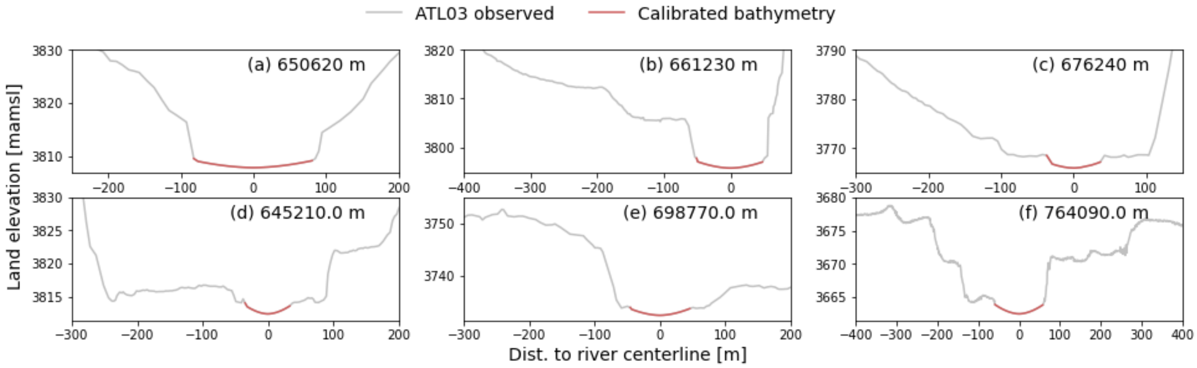
  These distances are referenced to the river chainage. Clarification is added in section 5.1.1

- Line 320-321 the sentence is not clear. Please, reformulate it.

  The sentence has been reformulated in section 5.2: “The calibration of uniform low flow depth and Manning's roughness is performed against ATL13 WSE observations for. A forward run of the model takes 39 seconds and the hydraulic model calibration converges after about 1000runs. To speed up the calibration, we define discharge classes, in which we group the WSE by the corresponding discharge value in the acquisition date. Acquisition dates with the same discharge value belong to the same class. This way, each iteration includes individual runs for 29 different discharge classes.”

- Line 333: the biggest errors are in the downstream sections because I think the estimated bottom of the river is too high. Can I see some cross-sections in the stretch from 65000 to 68000?

  I include cross-sections between chainage 65000 and 68000



  The bottom of the river is higher in this stretch and might explain the underestimation in depth. However, the distributed depth calibration did not improve the general RMSE in our model

- Figure 13 a: please specify the Depth coming from Mike 11 and Depth coming from ATL13 in the x-axis and y-axis. Are you able to explain the differences between the simulated and satellite observed WSE in the July-August 2020?

  The specified depth has been included in figure 14. Also, labels are changed to clarify, instead of depth, we use WSE – inferred river BL. For the differences in July-August 2020, we don’t have ATL13 observations for the dates in which the WSE increases significantly, only for 02/08/2020 and 26/08/2020 in which the difference with respect. the MIKE results are 0.3 m and 0.4 m respectively. The increase in WSE in the MIKE simulation corresponds to an increased value in the in-situ discharge measurements that are input in the model.

- Lines 384-385: do these studies refer to the same study areas? Please, specify.

These studies refer to different study areas. The study areas of these studies have been added to the text, Songhua River for Jiang et al. 2019 and Zambezi Catchment for Kittel et al. 2021.

- Line 439: please clarify this sentence because the paper does not show any comparison with other satellite missions to be stated that it "performs better than previous altimetry missions"

  We have included a comparison with available Hydroweb VS in the AOI. The comparison can be seen in figure 12 and details have been added to the discussion.

Referee nr2

This work presents the calibration of a 1D hydraulic model on ICESat-2 altimetric data, for a river portion with unknown bathymetry friction and given discharge upstream and downstream, which is an interesting topic. The calibrated parameters lead to fair performances in terms of fit to observed water surface elevations (WSE). The parameter space representation and sensitivity analysis are pertinent.

However the scientific novelty of the study and scientific positioning is not sufficiently clear, the authors omit lots of recent works about inverse hydraulic-hydrological modeling with current altimetric and water extent data, and forthcoming SWOT data.

The novelty of the study and its position has been clarified. References to recent works have been added to the discussion and introduction.

The main novelty of the study relies on the use of ICESat-2 data in a hydraulic calibration framework. ICESat-2 ATL03 products demonstrate the ability to map the topography surrounding the river at a high spatial resolution (0.7 m in the photon cloud and 3.5 m after processing the data). In addition, the ICESat-2 ATL03 product offers inland WSE with a variable resolution of a few meters, providing observations in narrower river streams (<250 m) that were not available or of low quality with previous altimetry missions.

I recommend the authors thoroughly rework the manuscript, rewrite the introduction and sharpen the analysis of results, reoganize some parts to ease the reading. I provide some elements below.

This work relies on:

- ICEsat-2 data, preprocessed using water masks and an effective parameterization (from Dingman) for unobserved low flow cross-section bathymetry. (Which has been used with altimetry and hydraulic modeling in Bjerklie et al. 2018).

The work by Bjerklie et al. 2018 uses ICESat data, which has a footprint of 70 m, while ICESat-2 has a footprint of 12 meters. The improved resolution of the data product offered by ICESat-2 which was launched in 2018 creates an added value when measuring cross-section geometry.

Background on Bjerklie et al. 2018 has been added in the introduction. This study uses ICESat data to validate Jason-2 slopes and uses the same approach to interpolate the river-submerged portion (from Dingman). Our study uses ICESat-2 land and water surface elevation from ATL03 and ATL13 products.

- a 1D steady state Saint-Venant shallow water model with 3 spatially uniform parameters (Manning friction, low flow depth, power shape) used in the calibration process. Rerun with calibrated parameters and unsteady solver in the MIKE platform. (Other 1D hydraulic models are used with altimetric data and effective bathymetry parameterization in references provided below).

- a global calibration algorithm and a global sensitivity analysis algorithm, both from literature are used.

I require clarifications and improvements on these points:

- No real analysis about the hydraulic inverse problem from satellite data, and of the scientfic difficulties related to it as the "bathymetry-friction" equifinality (Garambois and Monnier 2015), existing elaborate algorithms including variational data assimilation used for high dimensional calibrations and adapted to satellite hydraulics (cf. Larnier et al. 2020, Garambois et al. 2020 and references therein).

The new aspect in this study is the use of ICESat-2 data, and not the hydraulic inverse modeling workflow as such, which is essentially equivalent to the one presented in https://doi.org/10.1029/2020WR029261.

In our study, the uniform roughness calibration gives better results than the distributed roughness calibration (see Appendix A).

The references are added to the introduction and discussion. In this study we focus on the use of ICESat-2 data, Garambois and Monnier 2015 analysis is based on estimations of bathymetry and friction from observed WSE, and provide a deep analysis on "bathymetry-friction" equifinality. In our study, we put the focus on showing the performance of the model, given the observed elevation in the surroundings of the low flow main channel.

In the case of the data assimilation algorithms, we decided to use the approach presented by Kittel et al. 2021 for uniform calibration. The reduced computational cost of this algorithm is an advantage when using large data-sets as ICESat-2 cross-sections, and the uniform calibration provides good results. The proposed literature is added to introduction and discussion when relevant.

In addition, we analyze the sensitivity of the model to the lateral inflow definition, adding a uniform runoff distribution calibration. This is the main source of uncertainty in the prior information since we use in-situ discharge at the upstream and downstream end. This can be seen in section 4.4 and 5.2.3

- Model derived rating curves (and even stage fall discharge relationships, with WS slope...) have already been presented and thoroughly analyzed in Malou et al. with a hydraulic model calibrated on altimetry and water extents.

The point here is to show how the suggested workflow can combine spatio-temporally distributed ICESat-2 observations into time series of discharge and/or WSE at specific chainage points.

References to Malou et al. relating rating curves are added to the introduction and discussion and interpolation method in section 4.5.1.

- It is not clear to me how much snapshots of WS are used, what about temporal variability? Is there any nadir altimetric time series available on this study zone? The only validation regarding temporal variability is thus at gauging stations used as boundary conditions for the hydraulic models?

The WSE snapshots correspond to 236 ATL13 products which contain 3199 water surface elevation points. After filtering and averaging for spatially close observations, we calibrate with 81 WSE observations (see section 5.1.2). The temporal variability includes measurements from November 2018 to September 2020 for the calibration period, and from December 2020 to November 2021 for the validation period, with data covering the low-flow and high-flow season. This is better explained in section 5.1.2.

There are two of VS available in the Hydroweb database that have been included to compare with the simulated WSE from the model in section 5.2.3 figure 12.

- Hydraulic analysis are not deep enough, regarding forward-inverse hypothesis and resulting misfit wrt observations and river morphological features, regarding also "WS interpolation at any river point" as mentioned in the flowchart of Fig 1. Detailed discussions about hydraulic extrapolation in altimetry context can be found in Pujol et al. 2020, Malou et al. 2021.

We would like to emphasize again that the new aspect of this study is the use of ICESat-2 datasets and not the hydraulic inverse modeling workflow as such.

A new analysis regarding the lateral inflows has been added since this is the main source of uncertainty (see section 4.4, 5.2 and 5.2.3). The proposed references have been added to the introduction and discussion when relevant.

- An algorithm enabling Bayesian uncertainty analysis is used "To study the uncertainty of the model", the uncertainty is provided on infered parameters but not on other estimates in the rest of the paper ; analysis are not deep. What is the sensitivity of the parameter inference to first guess, to pdf choices and other calibration algorithm parameters?

The model uncertainty from the inferred parameters has been added to figs. 8 and 9. An analysis of the sensitivity to the lateral inflow is added since it is the main source of uncertainty and the results are also compared with the prior parameter guess. The pdf choice for the prior distribution is a Uniform distribution, this has been added in section 4.3.

- Is "80-180 meters in low flow season" corresponding to a narrow river? This corresponds to rivers visible with current nadir altimeters and by the future swot mission, which should be properly discussed with a literature review.

These values correspond to low flow season. The quality of the nadir altimetry time series available for this river reach is relatively limited (see analysis of Hydroweb VS), which shows that ICESat-2 can make a significant contribution here. SWOT may deliver good data on this reach in the future, but such data is currently not available.

Discussion relating to SWOT mission studies from the literature has been added to the introduction. In addition, we add the available Hydroweb data to compare to the results from the interpolated WSE from the model (see fig. 12).

- Clarify UPA_{x} in Eq. 7.

The acronym has been clarified before the equation.