We thank referee 2 for her/his time and effort and for providing constructive comments.

**RC:** General comment: This study aims at providing tools for predicting discharge from altimetry data in absence of gauge data. In order to achieve this goal, authors: 1- process the raw radar echoes distributed for satellite altimetry missions, 2- run rainfall-discharge models in various configurations 3- compute the Height-Discharge relationships, the so-called rating curves for all the model options, 4- select the best rating curves in the dataset provided by the previous step. Actually, this work is interesting and might be published but not in the present form, mostly because of significant lacks in bibliography, and consequently, lack of comparison of their results and methods with previous ones. Also, many inaccurate statements have to be rewritten. Introduction in particular needs an in-depth rewriting.

**AR:** Thank you for the overview and highlighting the strengths of the manuscript. The bibliography and comparisons have been significantly extended, and several statements have been rewritten to be more focused. We provide detailed responses below.

**RC:** Introduction: line 13: wrong citation. There have been tens of paper dedicated to "developing techniques" (some of them cited in the manuscript). Biancamaria is not one of them. Maybe be acceptable as "see review in Biancamaria et al., 2017"

**AR:** Biancamaria et al. (2017) indeed provides a review. We have modified the citation as suggested.

**RC:** Line 14: Radar altimetry is not "hampered" by the groundtrack spacing of the orbits. Indeed, in many basins, the number of crossings between the satellite orbit and the river drainage network is larger than the number of gauges that have ever existed in these basins. The fact that one cannot choose the location of the measurements is hampering radar altimetry. Radar altimetry is not "hampered" by the large footprint of the altimeters, strictly speaking. For sure, punctual measurements as the ones gained by laser technology (ICESAT) are much easier to process. The size of the footprint has drawbacks, for sure, but it also has significant advantages (the hooking effect for example as it is recalled in the present study, or the averaging effect over a large surface).

**AR:** Whether the groundtrack spacing "hampers" analyses depends on the geographical orientation of the river (East-West is better than North-South) and, of course, on what one considers as "sufficient" spacing. We agree with the reviewer that for our study region the number of crossings is larger compared to the number of gauges, and we have thus removed this statement. With regards to the size of the footprint, there are advantages and disadvantages that come with smaller footprints such as in laser (or SAR) altimetry. Yet, for braided rivers in complex terrain, in the presence of lakes...
and/or inundated areas close to the river, we believe a smaller footprint would be of advantage. We changed the sentence to "Radar altimetry is hampered by the long repeat cycles of the satellites (generally 10 days and longer) and the large footprints of the altimeters renders the processing less straightforward as compared to later altimetry".

RC: Line 15: LRM altimetry has been used to produce hundreds of series over narrow reaches, up to a few tens of meters wide (see for example those in the Amazon basin distributed by hydroweb). I may have missed the publications (and none is given to support the statement) but I did not read that either Cryosat-2 or S3A did much better, up to now. SAR just enables sampling more small reaches than LR Mode does.

AR: The reviewer is correct, LRM altimeters have been able to measure reaches of a few tens of meters width (under favourable conditions). We changed the sentence to "However, recent missions such as CryoSat-2 and Sentinel-3 have been shown to be able to capture more small river reaches due to their improved SAR (Synthetic Aperture Radar) Delay-Doppler measuring systems". We also changed the following sentence (line 17) "For crossings of large rivers . . . via public data bases . . ." to "For crossings of medium and larger rivers . . .".

RC: Line 21: "... it is generally required ...". "generally" is such a vague wording that one cannot say it is wrong but "a real gauge near the virtual gauge" is not required. Water levels have been converted into discharges in many studies, using a variety of methodologies, including the present one.

AR: We changed the sentence to "and the most straightforward way for converting them to discharge requires to have a daily . . .". "Generally" referred to "in most studies" in the original text but we agree this may be debated. But it cannot be doubted that the rating curve approach is most straightforward, see. Paris et al or Tarpanelli et al that we cite in line 28.

RC: Line 29-30: From the sentence "However ....", I understand that the rating curves cannot be applied to the Niger river. Yet, it is the focus of the study ??? I guess that the sentence should be rewritten (maybe stating that "single-polynom" rating curves cannot be applied ... ??

AR: What we meant here is that we don’t have overlap periods between gauge time-series and altimetry and the RC cannot be derived "directly". We clarify this now by modifying the sentence to "However, most of these techniques assume that a stage-discharge (‘rating-curve’) relation can be derived empirically during an overlap period and they can thus not be applied to the Niger river directly."

RC: Line 34: What about the works in Penidotti et al. (HESS, 2012), Casé et al. (HESS, 2016), Tourian et al. (HESS 2017) or Fleishman et al. (JoH, 2018) ?? These models are fully available, in particular the discharge series recently computed by Fleichman et al. (2018) using MGB and satellite altimetry in exactly the same area

AR: Thanks for pointing these works out. Indeed we agree that comparing our simple (lumped rainfall-runoff) calibrated simulations to their studies would be beneficial, but we believe it is beyond the scope of the current work. In particular, it is part of our research hypothesis that simple lumped simulation models can be used that do not necessitate an elaborate model setup (in other words that calibrating models in the gauge data period alleviates the inherent model deficiencies). We would further like to point out that some of the models used in the mentioned studies serve other objectives – land surface models are, first of all, developed to reproduce evapotranspiration – which are not required here.

Pedinotti et al (2012) used ISBA-TRIP, a full-fledged continental-scale land surface scheme which was augmented with river routing, floodplain and deep aquifer modules. They evaluate the model with GRACE data, altimetric water heights, in-situ discharge, and satellite-derived flood extent. In Casé et al (we assume the reviewer meant Casse et al, Model-based study of the role of rainfall and land use–land cover in the changes in the occurrence and intensity of Niger red floods in Niamey between 1953 and 2012, in HESS 2016) the same model, ISBA-TRIP is used with various forcing data sets
to study the role of rainfall and land cover on discharge. In both studies, the model appears not to be calibrated to gauge discharge.

In the same spirit, Fleischman et al. (2018) evaluated a two-way coupled hydrological-hydrodynamic model (MGB) in different model structure variants and forced by a single multi-satellite precipitation product. They calibrated their model using observed gauge discharge and validated against altimetric water levels and MODIS flood extent. For those gauges that are used in our study as well (Koulikoro, Ansongo, Dire), they find Nash-Sutcliffe coefficient similar to ours (within 0.1 better or worse). We note they used TMPA precipitation which is close to PERSIANN, the data that we used, but not identical.

Along the same line, we could mention our own recent work in Pomeon et al. (2018). In this work, the SWAT model was set up for a region close to the study region here, calibrated with gauge discharge, and evaluated against GRACE and MODIS evapotranspiration.

In contrast, Tourian et al (we assume the reviewer meant Tourian M.J., Schwatke C., Sneeuw N.: River discharge estimation at daily resolution from satellite altimetry over an entire river basin. Journal of Hydrology, Vol. 546, 230-247, 10.1016/j.jhydrol.2017.01.009, 2017) develop a purely empirical linear cyclostationary dynamic model for discharge, which is learned by ingesting daily gauge time series and which assimilates altimetric discharge (i.e. levels turned to discharge in a preprocessing step via a modified RC approach). This represents an intelligent interpolation approach for discharge rather than a physical modelling study.

As mentioned above, it could make an interesting comparative study to use modelled discharge from any of these studies for generating rating curves within a common time window. However, while we could probably obtain simulated discharge from the authors of these studies, these models are all less straightforward to set up and to transfer to other regions compared to GR4J and HBVlight, and this would thus not add to the objective of the present paper.

As a result of this reasoning, we suggest to modify the sentence on line 32 "Others have proposed to assimilate altimetric levels into elaborate hydrodynamic modelling (Munier et al., 2015); however such models are not always available" to "Others have proposed to simulate discharge using fully-fledged calibrated/validated land surface modelling (Pedinotti et al., 2012, Casse et al., 2016, Fleischmann et al., 2018, or Pomeon et al., 2018), assimilate altimetric levels into elaborate hydrodynamic modelling (Munier et al., 2015), or interpolate discharge based on empirical dynamic models trained on gauge discharge (Tourian et al., 2017); however such models are not always available and less straightforward to transfer to new regions."

RC: Page 2, line 4 : that "radar retracking is key for obtaining meaningful time series..." has been shown in many publications over more than a decade...

AR: This is correct. We modify "we will show" to "we will confirm"

RC: Methods and Data Page 5 line 5 : Capability in retrieving water levels has considerably evolved from TP, Jason1 to Jason 2. Hence, they cannot be qualified by a single "70-80cm RMSE". Note that much better results than those by Tourian et al. (2016) can be found in the literature for Jason 2 (see for instance Seyler et al., 2012)

And

RC: Line 8: Tourian et al. (2016) estimated their bases over the ocean, which is of limited meaning for rivers because the echoes are too different. other estimates are available in the literature

AR: The "70-80cm RMSE" is the Tourian et al estimate for Jason-2, i.e. not a single value for TP, J1 and J2. But we are aware that better results have been reported, e.g. in Seyler et al. (2013) though the authors of this study appear reluctant ("0.35 m . . .could be then the error estimated on the water stage derived from Jason-2 ranges, when no other validation is available."). Papa et al. (2012) suggest 0.28 and 0.19 m for Ganges
and Brahmaputra – clearly this depends on river width. We modify the sentence to "Relative altimeter errors (…) are thought to be at the level of 20-80 cm Root Mean Square Error (RMSE) e.g. for Jason-2 dependent on river width (Papa et al., 2012, Seyler et al., 2013, Tourian et al., 2016)". For Envisat and Saral/Altika: "Relative Errors are believed to be at the 15-70 cm range (Sridevi et al., 2016, Tourian et al., 2016, Bogning et al., 2018)".

**RC:** Line 9-10: "biases […] reference systems effects". Could the authors explain which "reference system " may turn into significant bias effect ?

**AR:** The reviewer is right here. What we meant here is that reference system effects (e.g. realization of the vertical datum, connection of gauge to datum, geoid model error) add to observed systematic discrepancies and to the estimated bias, but this is of course not the altimeter’s fault. We suggest to remove "reference systems effects" for clarity.

**RC:** Page 6 lines 4-5: Roscher et al. (2017) retracked altimetry data on the coastal domain. I may have missed the publication but I am not aware of any publication showing that their STAR retracking performs better over rivers (both in height accuracy and number of data) than the other-usual algorithms, such as ICE1, ICE3, … If so, authors should add a reference.

**AR:** The reviewer is right; the present study is the first publication where the Roscher et al STAR method is used over rivers. Please note that we do not claim STAR to be the only method which provides decent heights over rivers. Other approaches, such as the Multi-Subwaveform Retracker (MSR, Boergens et al., 2016) often lead to very similar results compared to our STAR approach. From Fig. 1 it becomes obvious that simple retrackers such as ICE1 in this case can lead to unrealistic peaks over rivers, especially during low water level conditions. However, please note that we do not want to provide an in depth discussion of retracking methods here since this is not the main topic of this paper. We modify the sentence to "which lead to much more … in coastal applications. Here, we make use of the point-cloud by-product of STAR in order to derive improved river heights. ".

**RC:** Line 8 : Berry presented the hooking effect extensively in meetings in the early 2000’s but as far as I know, the first publication presenting the use of these parabolas to gain accuracy in retrieving water levels is Frappart et al. (RSE 2006).

**AR:** True. We add the reference to Frappart et al. (2006).

**RC:** Line 8: "which masks water levels". Actually, since the onboard tracking is locked on the river echo, the bank topography is masked by the river, not the opposite.

**AR:** Yes, this is actually explained with the previous sentence. We shorten this sentence to "leads to the so-called ‘hooking effect’ (…), a spurious parabolic profile …".

**RC:** Page 7 line 8 : "… even over narrow river crossing …". It may be a question of my own limitations in english but I think that "even" should be replaced by "in particular" since the hooking effect is particular visible -and useful- over the narrow reaches.

**AR:** This depends on the point of view. The hooking effect is particularly pronounced over narrower crossings and it is thus easier to detect/remove it. In this sense the reviewer is correct (of course if there would be no topography i.e. the river much wider, there would be less hooking effect but nevertheless better altimetric accuracy; this was what we meant to say). To remove this ambiguity, we decide to remove the word "even" from the sentence.

**RC:** Line 20 : Authors mention that the S3A data are distributed with 2 retracking algorithms, the OCOG and the Samosa ones. I suggest that both algorithms will be tested in their study.

**AR:** Actually both have been included already, please see results in Fig. 7 and the explanation on page 7 lines 12-15. GPOD uses Samosa and Copernicus used OCOG.

**RC:** Line 26 : Authors should mention that the unit of a is $m^4 s^{-1}$ only if $b = \frac{3}{2}$ ( if b is
different from $\frac{1}{3}$, the units would not equilibrate on both sides of Eq 1

AR: This is correct. We suggest to modify the sentence from "In Eq. (1), a (in $m^\frac{1}{3}s^{-1}$) and the dimensionless parameter b describe the hydraulic behaviour. These parameters can be computed from Manning's equation under idealized conditions (Paris et al., 2016)." To "In Eq. (1), a and b describe the hydraulic behaviour. These parameters can be computed from Manning's equation under idealized conditions (Paris et al., 2016), then usually $b = \frac{1}{3}$ (dimensionless) and a would be in units of $m^\frac{1}{3}s^{-1}$."

RC: Line 30: "river width is difficult to observe from space". This sentence should be reworded since river width is definitely not difficult to observe from space and it is commonly derived from existing satellite imagery products

AR: River width needs to be known or observed at the time of the altimeter overflight, and for optical sensors this is may be challenging in tropical regions due to cloud coverage. We agree with the new Sentinel products this problem may be largely alleviated, but around 2000 the situation was clearly different from now. We suggest to modify the sentence to "river width has been difficult to observed in the past".

RC: Page 8, line 1: "assuming gauge and virtual gauge data [...] area available...". Replace "gauge" by "discharge" ??

AR: Thanks for pointing this out. We suggest to modify "Assuming gauge and virtual gauge data from altimetry are available" to "Assuming observed discharge and virtual gauge level data from altimetry are available".

RC: Page 9 Eq 3: I suggest that authors mention that Getirana Peters-Lidar (HESS 2013) showed that this methodology based on minimum of squared residuals of Log regressions can converge towards physically meaningless values (which may explain the final results in terms of rating curves found in the present study

AR: Getirana and Peters-Lidard (HESS 2013) indeed show that minimizing the residual of Eq. (3) (their Eq. (2) is identical to our Eq. (3)) will not necessarily lead to a converging solution for the parameter $Z_0$. However, they also state that non-convergence does not imply bad fitting (i.e., inability to simulate discharge). It simply mirrors the fact that even $Z_0$ values which are false can produce a better fit than the true one. This leads to a high uncertainty for the final result of $Z_0$ and thus indeed explains our final results in terms of rating curves. We suggest to add an explanation in the discussion rather than in the methodology: "Getirana and Peters-Lidard (2013) point out that the used procedure of rating curve fitting may not necessarily converge for $Z_0$, which may lead to inaccurate values for this parameter." (section 3.3, after "Noticeable is that the rating curve is almost parallel . . .").

RC: Line 9: I suggest that the authors present an example of "breakpoint" in a figure.

AR: We add a reference here to Fig. 8 (GR4J, Envisat with BP).

Page 17: all these results, although interesting, lack analysis with respect to literature, in particular with respect to the recent study by Tourian et al. (HESS 2017) and by Fleischman et al. (Joh, 2018).

AR: As discussed earlier, these studies had somewhat different objectives and so it is not easy to compare. We suggest to focus here on (1) skill of simulated discharge w.r.t. gauge discharge, (2) altimetric discharge w.r.t. simulated discharge. We add a paragraph that compares our results to the studies of Tourian et al. (2017) and Fleischmann et al. (2018) as far as this is possible.

RC: Page 18: Figure 8 is quite difficult to read and I suggest that authors redo it. Actually, it would be more clear to plot the discharge versus the true altimetric height instead of versus the water depth (given that, in addition, no information is provided about the Zo parameters). Plotted this way one would have only 2 discharge values for a given altimetric height. Also, information about the Zo parameters could be presented simply by the intersection of the rating curves with the height axes.

AR: We changed the figure in the proposed way (Fig. 2) and adapted the discussion
to it.

**RC:** Page 21, line 21: Sentence "we find that simulated discharge may aid..." is really over-selling the findings of the present study since many studies already showed such a result. The sentence has to be reworded.

**AR:** We suggest to replace "We find that simulated discharge may aid in creating empirical altimetry-discharge rating curves..." by "We find that discharge simulated by simple lumped rainfall-runoff models may aid in creating empirical altimetry-discharge rating curves...".

**RC:** Line 28: ".... most likely...". In absence of any evidence presented in the study, I think that" presumably" would be better adapted.

**AR:** We agree and we have modified the sentence to "presumably when the river inundates its banks."

**References**

- Sridevi, T., Sharma, R., Mehra, P., Prasad, K.V.S.R., 2016. Estimating discharge from the Godavari River using ENVISAT, Jason-2, and


Fig. 1. STAR vs. ICE1 retracker
Fig. 2. Rating curves with altimetric heights on x-axis