Dear Anonymous Referee #1,

Thank you very much for your feedback on our paper on "Using altimetry observations combined with GRACE to select parameter sets of a hydrological model in data scarce regions". Hereby we would like to respond to your comments:

Comment 1: I got lost in some of the technical detail around the various alternative calibration strategies tested for calibration. As the abstract seems to suggest insights relating to calibration strategy are the main contribution of this m/s, so I think this needs some more attention. For example, the research hypotheses (l. 109-11) do not address this aspect. In the introduction, can you provide some discussion around the rationale for the different experiments? In fact, to support that, it would be helpful if the authors could provide a table listing, for each variant, the objective function, any transformation of model or observation data (i.e. the observation model), the potential benefits of the variant (i.e., why was it tested), and the empirically-found pros and cons.

Response: This is an excellent suggestion - we agree, the different calibration strategies with respect to altimetry were only introduced in the methods section (Section 3.3.3), but not mentioned in the introduction nor the hypothesis. We will change this in the manuscript by adding a section in the introduction and adjusting the hypothesis. In addition, we agree it would be helpful for the readers to include an overview of the different calibration strategies including their objective functions, discharge – water level conversion techniques and benefits/drawbacks. Therefore, the table shown below will be included.

Comment 2: Please consult Domeneghetti (2016) and Oubanas et al. (2018) and consider whether they may be relevant to your discussion.

Response: Thank you for pointing out these interesting papers! We will include them in the manuscript.

Comment 3: With the caveat that I did not understand all details, I seem to gather that one of the main conclusions of this m/s is that selecting parameters based on rank correlation between discharge and altimetry water level is not sufficient to constrain model parameters, and that altimetry levels need to be converted to actual discharge to provide an appropriate constraint. Is that correct? If so, then that would be expected when evaluating against a performance measure that is extremely bias-sensitive, like Nash-Sutcliffe efficiency (NSE). However, while I know NSE is religiously adhered to by some hydrologists, it is not a relevant performance indicator for all possible uses of river discharge modelling (and indeed many hydrologists have already found a new religion in the more information-rich components of Kling-Gupta Efficiency, KGE). For many practical applications, a high correlation may well be more important than a bias-free estimate, for example in flood and drought applications. Even if volumetric accuracy is more important (e.g. in water resources volume management) then, in this case, you have some gauged data, so provided correlation is high a post-model bias correction would be straightforward. (Although of course station gauge data always have some bias of their own against the unknown truth!). Furthermore, given the almost certainly large uncertainty and bias in the CHIRPS rainfall data for this region, it is likely that a parameter set minimising bias will compensate for the biases and errors I the rainfall data. (Perhaps there are some rain gauge data to test this). In summary, I would recommend not relying on NSE nearly as much, and *also considering correlation measures, perhaps by using the KGE breakdown. At the very least, more discussion is needed.*

Response: We agree the Nash-Sutcliffe efficiency has its limitations just as any other metric. Nevertheless, it still can provide us with valuable information. For many applications in water resources management, it is important to capture both the flow dynamics and volumes correctly; for instance for the management of a dam in the context of flood/drought protection. In that case, a bias-sensitive performance metric such as the Nash-Sutcliffe becomes quite useful.
The Spearman Rank Correlation function only accounts for the dynamical changes and not for the volume which indeed could be taken into account by using information available through gauged data as Referee #1 suggested. This would be possible for this study, but hypothetically, what if no gauged information would be available (which is the reality for the vast majority of river basins world-wide)? This was the assumption made in this study, to answer the question of how well can we do to reproduce river flow in a basin where no flow observations are available and only altimetry data are used for model calibration. Then this study illustrated the added value of converting the water level to discharge using the Strickler-Manning equation to capture the volume better. We agree that the bias in for example the rainfall data can then be compensated through the additional calibration parameters which therefore need to be constrained as much as possible.

Comment 4: Please add some discussion about the performance of the different variants against the different flow signatures introduced in l. 276-280. Rather than referring to Euser et al., why not include the formula in a table and list the performance of each model variant? I note that most of the signatures are sensitive to bias (see below) and the runoff coefficients also to bias in rainfall. That means that the potential bias in the spatial rainfall estimates and station discharge records needs to be discussed.

Response: Thank you for pointing this out. The performance of the different variant with respect to the different flow signatures is visualized in Figure 7 in the original manuscript, but not discussed explicitly as the focus was on improving the over model performance with respect to the flow. We will include this in the manuscript together with a table summarizing the formulas for the different flow signatures. In addition, a detailed table summarizing the model performance for each calibration strategy with respect to each flow signatures (as shown in Figure 7) will be added in the supplementary material for the interested reader.

Table 1: Overview of the calibration strategies applied in this study

Calibration strategy name	Calibration data	Objective function	Nr. of calibration parameters	Comments	Discharge – water level conversion method	Benefits (+) & limitations (-)
Discharge (reference)	Discharge (at basin outlet)	D_E	17	Traditional model calibration on observed flow data Combination of 8 different flow signatures	-	-
Seasonal water storage	GRACE	$E_{NS,Stot}$	17	No discharge data used	-	-
Altimetry Strategy 1	Altimetry (at 18 virtual stations) & GRACE	Altimetry: $D_{E,R,WL}$ GRACE: $E_{NS,Stot}$	17	No discharge data used Combination of 18 virtual stations Combined with GRACE	-	+ No extra parameters or data needed + Assumption: monotonic relation between discharge and river water level - Focus on dynamics only, not volume
Altimetry Strategy 2	Altimetry (at 18 virtual stations) & GRACE	Altimetry: $D_{E,NS,RC}$ GRACE: $E_{NS,Stot}$	25	No discharge data used Combination of 18 virtual stations Combined with GRACE	Calibrated Rating curve	+ No extra data needed - Two extra parameters per cross-section
Altimetry Strategy 3	Altimetry (at 18 virtual stations) & GRACE	Altimetry: $D_{E,NS,SM}$ GRACE: $E_{NS,Stot}$	18	No discharge data used Combination of 18 virtual stations Combined with GRACE	Strickler-Manning	+ Only 1 extra parameter - Cross-section data needed - Assumption: constant roughness in space and time
Water level Strategy 1	Water level (at basin outlet) & GRACE	Altimetry: $E_{NS,SM,GE}$ GRACE: $E_{NS,Stot}$	18	No discharge data used Combined with GRACE	Strickler-Manning	+ Only 1 extra parameter - Cross-section data needed - Assumption: constant roughness in space and time
Water level Strategy 2	Water level (at basin outlet) & GRACE	Altimetry: $E_{NS,SM,GE}$ GRACE: $E_{NS,Stot}$	18	No discharge data used Combined with GRACE	Strickler-Manning	+ Only 1 extra parameter - Cross-section data needed - Assumption: constant roughness in space and time

Comment 5: I would like to see some comparison of model vs remotely sensed GRACE and altimetry data, and the performance of the different calibrated variants against it.

Response: In the manuscript, we wanted to find out whether accurate discharge simulations can be obtained when calibrating to altimetry which is why supporting graphs visualizing the flow were mainly shown.

But with Figure 1, we would like to illustrate the difference between the following two strategies with respect to the water level: converting the simulated discharge to observed water levels using 1) calibrated rating curves and 2) the Strickler-Manning equation. Both strategies were applied to Virtual Station 4 (see Figure 2 for its location) as an example. This graph shows that the water level simulations improved significantly when applying the Strickler-Manning equation and including cross-section information. In addition, Figure 3 shows the model simulation results with respect to the total water storage compared to GRACE.

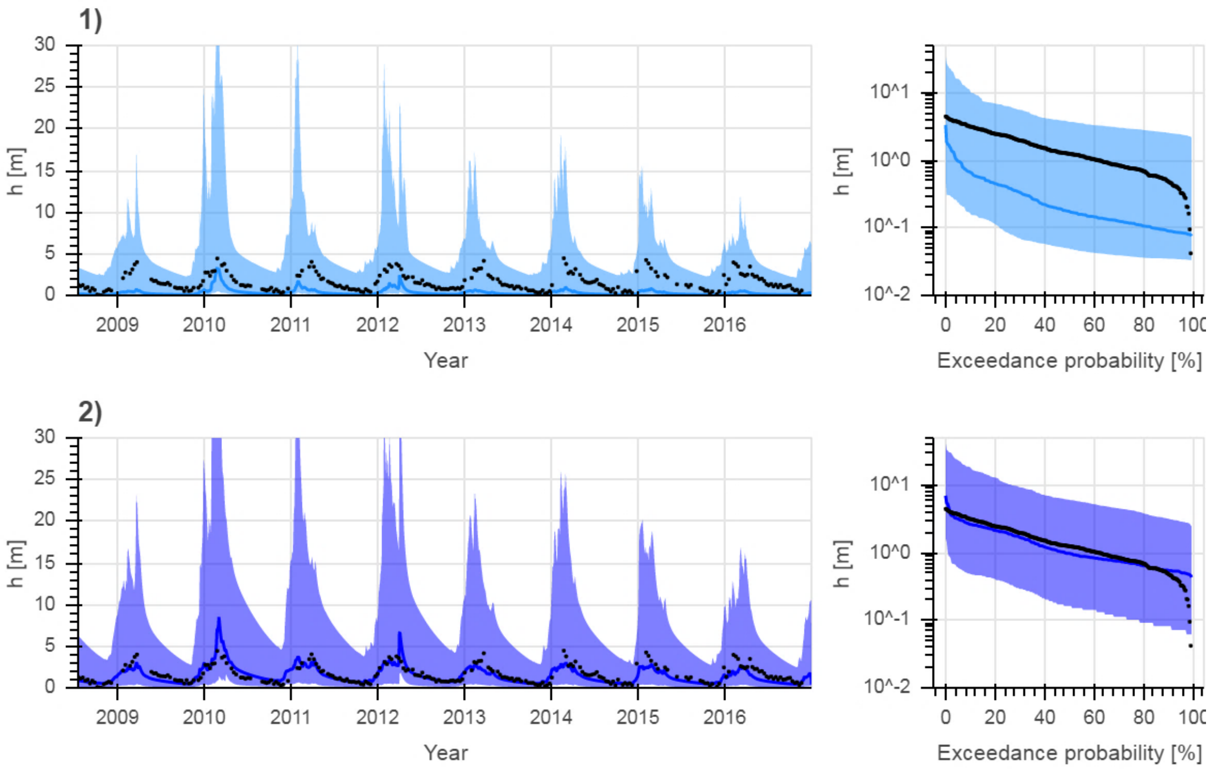

Figure 1: Range of model solutions for Virtual Station 4 (see Figure 2 for its location). The left panel shows the time series and the right panel the exceedance probability graph of the recorded (black) and modelled water level: the line indicates the solution with the highest calibration objective function and the shaded area the envelope of the solutions retained as feasible. Solutions retained as feasible based on altimetry observations using all virtual stations within the basin and 1) calibrated rating curves for the discharge – water level conversion or 2) the Strickler-Manning equation with cross-section information retrieved from Google Earth.

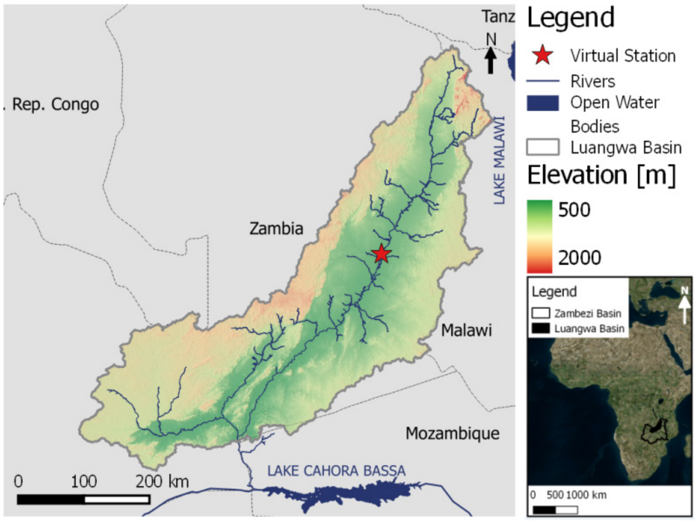

Figure 2: Map of the Luangwa River Basin illustrating the location of Virtual Station 4

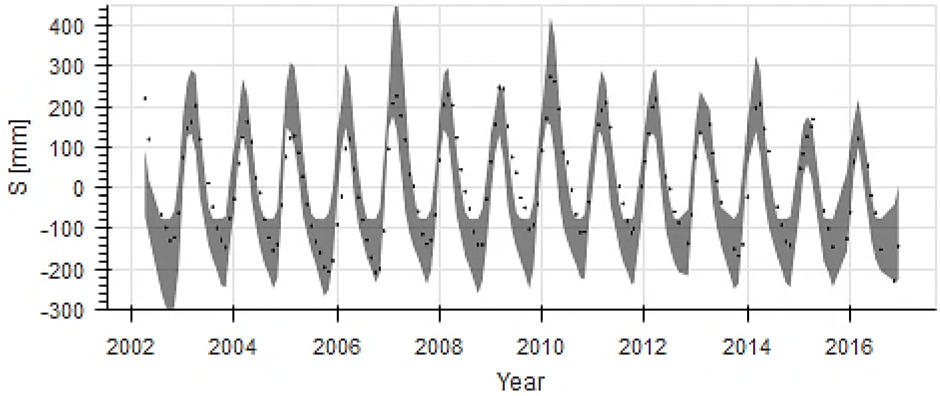

Figure 3: Range of random model realizations with respect to the total water storage (grey) including the observation according to GRACE (black).

Comment 6: GRACE observations are coarse and subject to various uncertainties. To better understand uncertainty relating to calibration to GRACE, can you discuss the contributions of the different storage terms to the temporal variation? This would help to understand where the main uncertainties might be, e.g., how important surface water storage variations are. Also, given the proximity of lakes, dams and wetlands (Cahora Bassa, Lake Malawi, Bangwelu wetlands), they may well have had an influence on GRACE water storage variations. There is no question they are sufficiently close to affect the signal, but perhaps their water level variations haven't been very large during the analysis period. Please discuss this and provide some evidence. For example, you could look at their water level changes (e.g. from altimetry) and you could map the temporal correlation of each GRACE pixel to the respective pixels over each of these 3 areas. Finally, please discuss the SEE between model and GRACE water storage in comparison to the random noise in the GRACE solutions.

Response: Thank you for this comment. There are indeed quite some uncertainties and random noise in the GRACE observations; this will be included in the discussion as it is still missing. Within the hydrological system, there are several components that contribute to the total water storage such as the water stored on the surface, in the shallow subsurface zone and in the groundwater. The temporal variation of the first two components is relatively high whereas groundwater levels change slowly. The temporal variation in the monthly total water storage is dominated by the slow variations in the groundwater level. In addition, there are strong seasonal variations in this region due to a very clear wet and dry season which is reflected in all storage components.

As pointed out by Referee #1, there are several lakes/reservoirs and wetlands in the area that could affect GRACE observations. For example the water level variations at Cahora Bassa are significantly larger than the variations in GRACE focusing on the pixel where the virtual station is located (Figure 4A). This influence decreases when focusing on a larger area for instance the area within a 300 km radius of the virtual station (Figure 4B) which is the same distance used to smooth the data and filter out noise. Similar results were found for the other open water bodies and swamps mentioned by Referee #1. In this study, the smallest distance between the basin and a large open water body or swamp was 51 km for Lake Malawi, 72 km for Kafue Flats, 74 km for Cahora Bassa, 135 km for Kariba, 173 for Bangweulu and 210 km for Tanganyika (Figure 5). Hence, large open water bodies and swamps can indeed affect GRACE observations considerably especially for small areas.

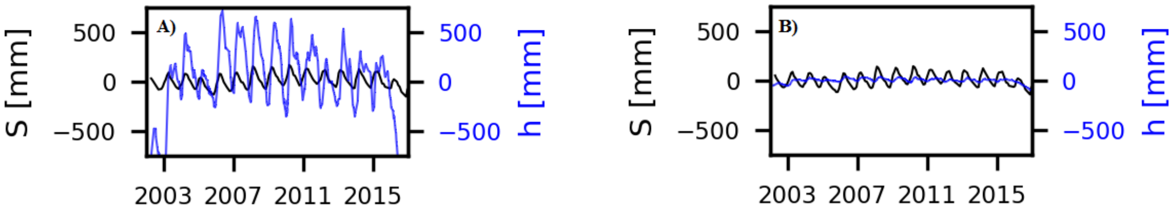

Figure 4: Altimetry observations at Cahora Bassa (blue) and average total water storage according to GRACE (black) for the following areas of interest: A) GRACE cell in which the virtual station is located, and B) area surrounding the virtual station with a radius of 3 degree (GRACE area of influence). The altimetry was multiplied with the area-weighted contribution of open water within the area of interest.

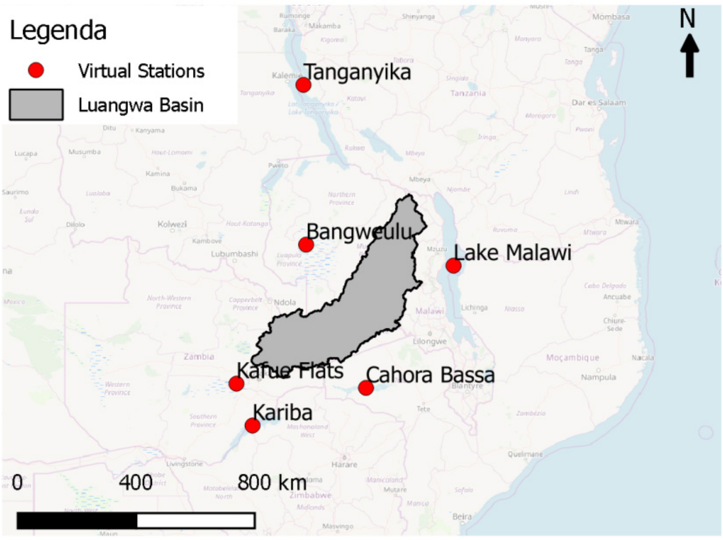

Figure 5: Map of virtual stations

The temporal correlation of each GRACE cell relative to a specific cell is illustrated in Figure 6. For this purpose, the GRACE observations for the cell in which the virtual station for Cahora Bassa is located is plotted against all cells within an area surrounding the virtual station with a radius of 3 degree (GRACE area of influence). This figure shows that there is a relatively strong temporal correlation between the GRACE cells which could be a result of for example the strong seasonality in this area. However, the temporal correlation between GRACE and the altimetry observation is significantly weaker for this example (blue dots in Figure 6). This indicates that the Cahora Bassa reservoir had a limited impact on the GRACE observations within its representative cell despite the large fluctuations.

Unfortunately, we are not sure what Referee #1 is referring to when mentioning the abbreviation “SEE”.

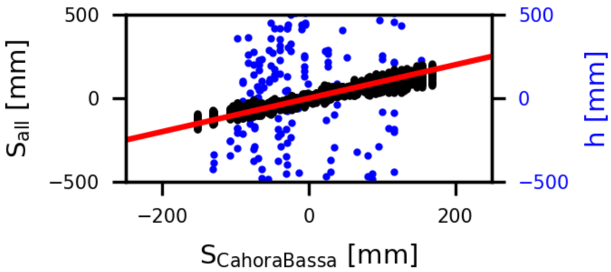

Figure 6: Temporal correlation of the GRACE observations for the cell in which the virtual station for Cahora Bassa is located (horizontal axis) and for A) all cells within an area surrounding the virtual station with a radius of 3 degree (GRACE area of influence, vertical axis, black), and B) the altimetry observation at Cahora Bassa (vertical axis, blue). The 1:1 line is visualised in red.

Comment 7: The apparent benefit of having accurate river cross-section data along with the altimetry data is an interesting one, and could be the most important contribution of this m/s. Can you explore opportunities to build on this insight a bit more please? For example, it is my understanding that profiles can be derived from the altimetry measurements. I am not a radar altimetry specialist and appreciate the authors are not either, but I am sure insights can be found in the literature. Secondly, given the importance of river geometry, can you discuss whether river width and pseudobathymetry from optical remote sensing might help you (see Sichangi et al., 2016; Hou et al., 2018), particularly now there are such data globally at Landsat resolution. In fact, a simple and useful addition would be to add a map of each virtual and actual gauge derived from the Global Surface Water Dataset which is a great resource (Pekel et al., 2016; https://global-surface-water.appspot.com/map). Finally, one of the other reviewers will probably already suggest you mention the SWOT mission. While not seeing inherent merit in arm-waving, in this case, it is interesting to discuss to what extent the SWOT observations might provide richer and/or more accurate data (e.g. on river cross-section and profile) than the current crop of altimeters.

Response: We agree, showing the benefit of detailed cross-section information when combining it with altimetry observations is the main major finding of this manuscript. This approach has a lot of potential. For example, it would be very interesting to combine altimetry observations with river width estimates derived from Landsat or Sentinel-1/2 (Huang et al., 2018). Alternatively, altimetry observations could be combined with CryoSat based altimetry observations which provide water level information at lower temporal resolution (every 369 days), but higher spatial resolution (equatorial inter-track distance of 7.5 km) providing valuable information to estimate the river slope (Schneider et al., 2017;Jiang et al., 2017). In addition, with the upcoming SWOT (Surface Water Ocean Topography) mission, more accurate altimetry observations should be available as also river slope observations and cross-sections; the repeat cycle will be 21 days and across-track resolution between 10 m and 60 m increasing the number of observation points available within a specific area (Biancamaria et al., 2016;Langhorst et al., 2019). Also, it would be very useful to improve cross-section estimates with respect to the submerged part as already explored in previous studies

(Domeneghetti, 2016). Furthermore, drone observations could be used to obtain more accurate cross-section information and estimates of the river slope and roughness (Entwistle and Heritage, 2019). We will add a more extensive discussion on these points in the revised manuscript.

Literature

Biancamaria, S., Lettenmaier, D. P., and Pavelsky, T. M.: The SWOT Mission and Its Capabilities for Land Hydrology, Surveys in Geophysics, 37, 307-337, 10.1007/s10712-015-9346-y, 2016.

Domeneghetti, A.: On the use of SRTM and altimetry data for flood modeling in data-sparse regions, Water Resources Research, 52, 2901-2918, 10.1002/2015WR017967, 2016.

Entwistle, N. S., and Heritage, G. L.: Small unmanned aerial model accuracy for photogrammetrical fluvial bathymetric survey, Journal of Applied Remote Sensing, 13, 1-19, 19, 2019.

Huang, Q., Long, D., Du, M., Zeng, C., Qiao, G., Li, X., Hou, A., and Hong, Y.: Discharge estimation in high-mountain regions with improved methods using multisource remote sensing: A case study of the Upper Brahmaputra River, Remote Sensing of Environment, 219, 115-134, https://doi.org/10.1016/j.rse.2018.10.008, 2018.

Jiang, L., Schneider, R., Andersen, O. B., and Bauer-Gottwein, P.: CryoSat-2 altimetry applications over rivers and lakes, Water (Switzerland), 9, 10.3390/w9030211, 2017.

Langhorst, T., Pavelsky, T. M., Frasson, R. P. d. M., Wei, R., Domeneghetti, A., Altenau, E. H., Durand, M. T., Minear, J. T., Wegmann, K. W., and Fuller, M. R.: Anticipated Improvements to River Surface Elevation Profiles From the Surface Water and Ocean Topography Mission, Frontiers in Earth Science, 7, 102, 2019.

Schneider, R., Godiksen, P. N., Villadsen, H., Madsen, H., and Bauer-Gottwein, P.: Application of CryoSat-2 altimetry data for river analysis and modelling, Hydrol. Earth Syst. Sci., 21, 751-764, 10.5194/hess-21-751-2017, 2017.