The authors investigated the assimilation of satellite altimetry to improve discharge. The article is excellent. The authors presented 3 types of altimetry assimilation, which are 1. direct, 2. anomaly, and 3. normalized assimilation, and concluded that to improve discharge it is better to just assimilate the water surface dynamics (3.) given that the model is relatively accurate. If the model is completely corrupted, the authors concluded that anomaly and/or direct assimilation are more effective.

My main complaint about this manuscript is in section 3.1 (specifically 3.1.1, 3.1.2, and 3.1.3) where the figures (fig 4, 5 and 6) showed hydrographs that were different from the locations discussed in the text. Besides that, I have just some small comments that are specified below:

Line 40. I think it is better to change GHM definition only to Global Hydrodynamic Models instead of Hydrological due to some features the authors discuss further such as "runoff as a forcing factor", "discretized river", "surface water dynamics", etc.. Line 77 and 519 could be GHM instead of global hydrodynamic models.

Line 52. The authors could also mention Laser altimetry. The ICESat missions are also very used in academic research. Maybe instead of a radar pulse, can be a radar/light pulse or even an electromagnetic pulse.
As an alternative to the semi-variogram analysis to determine the spatial dependency weights, the authors could have used "backwater lengths in rivers" studied by Samuel (1989). It can give an idea of which river reaches are affected by WSE variations at the VS locations. It would be a good idea to compare both approaches in future studies (not now).


Line 156. using a power law dependent on what? Width? Upstream drainage area? Is it the same power law parameters for the whole basin?


Line 229. So, the mean and standard deviation were calculated based on the open loop simulation?

Line 231 to 239. Some of your readers might be unfamiliar with the Amazon Basin. It would be interesting to write a short and objective section about this basin, presenting a DEM map at least (a mean Precipitation map would be nice too).
Line 275. How do you measure the relative sharpness and the difference in reliability? Line 383 should be here.

Several wrong references in section 3.1.

Line 309. The authors said that the Santos Dumont gauge is in the Purus River, but in Figure 4 it says Jurua River.

Line 332. "Figure 5c–e displays hydrographs of the Jurua (Gaviao), Amazon (Manacapuru), and Negro (Serrinha) rivers" but in Figure 5 it says Manicore on the Madeira River, Aruma on the Purus River, and Sao Felipe on the Negro River.

Line 351. "The lower panels of Figure 6 illustrate flow dynamics along the Amazon mainstem (Sao Paulo De Olivenca; Figure 6c) and Japura (Vila Bittencourt; Figure 6d) and Negro (Curicuriari; Figure 6e) rivers." but in Figure 6 it is written "Hydrographs recorded at Humaita on the Madeira River, Santos Dumont on the Jurua River, and Canutama on the Purus River are presented on panels c, d, and e, respectively."

Line 321. Saying that the "direct DA generally improved flow dynamics" is very optimistic. Based on these results, I'd probably say that the direct DA maintained or even degraded the general performance, at least for discharge.

Line 340. Once again, I think it is an optimistic conclusion. In the last sentence, the authors just said: "although NSE and ISS values worsened slightly." So how can the authors say afterward that "discharge estimates improved moderately"? I don't think that improvements in the correlation coefficient are enough for such a statement given that the NSE has become worst. But I reckon that seasonality got better as correlation got higher. Maybe the authors should clarify what they try to achieve with DA assimilation.
Figure S2 should be in the main manuscript. It could be together with Figure 7 as 7c, 7d, and 7e. Figures 7a and 7b don't need to be so large.

Line 427. "However, the direct DA experiments efficiently improved sharpness, thereby increasing confidence in the assimilated river discharge." I would say "FALSELY increasing confidence" as the authors just observed that the reliability drops more than 50% for direct DA experiments. What is the point of being narrower if the observation falls out of the confidence interval? I think the authors should be careful with that.

On tables 3 and 4 it would be nice to see the Open Loop and the CaMa VIC BC performances for comparison.

Line 541. HTESSEL not HTEESSEL.

Line 624. Which experiment is that? The one in section 4.2.? Or the one with VIC BC (section 3.3)?