Text in bold: Reviewer comments; *Text in italic: Authors response*

Guy Schumann - Referee #2

This paper describes in detail a much needed continental-scale cross-comparison study at the continental scale of the typical regional MGB model using not only different global scale models but also observed or observation inferred variables (e.g. TWS from GRACE, satellite altimetry). The paper is technically very sound and strong and I did not see any problems with the methods employed.

We thank Dr. Schumann for dedicating his time to reviewing our manuscript and for highlighting the need of such a study to the scientific community.

I really enjoyed reading this paper and although it is fairly long in places, I think it is written in a very comprehensive way and very well organized and presented - I applaud such work and writing. Well done! This said, there are some main points I would like to highlight and see addressed before publication.

Thank you very much for this motivating comment. We did our best efforts to draw the attention of a broad public, as well as to extend a regional model to the continental domain using interesting approaches of global-scale modeling.

I think it would benefit the paper a lot by listing a number of steps or recommendations to follow for large-scale hydrologic model assessment or validation

Thanks for the suggestion. We will add some recommendations. It may be also interesting to highlight South American basins that can serve as "stress tests" for hydrological models, linking with the comments from Reviewer #1.

It looks to me as though generally speaking the headwaters are difficult to get right or better said "to agree with other models", which means to me that they are generally very difficult to model correctly. This is of course not surprising given that the topographic complexity and hydrological processes in these regions are not well represented in the models. It would be useful if the authors could comprehensively outline the reasons for those "problem areas".

Thanks for the suggestion. We will add some additional discussion about these problem areas. Indeed, there is first an issue of scale. Global or even continental models are not designed to provide estimates for headwater catchments and small rivers due to the resolution of these models and datasets used. Methods for downscaling / interpolation of forcing data can also impact model results, in addition to limitations of satellite estimates.

As far as I understand the authors, model calibration is still challenging and therefore could also be responsible for explaining some or even most of the differences observed between different models. Logically it follows that there

should be the general recommendation to define and build a set of data that should be used for calibration of large scale models, so that comparison studies later are even more valuable. I think the authors, if they can agree, should call for such a data set in their section of "Model adjustment" (section 3.3) or later in the conclusion is maybe even a better place.

Thanks for the suggestion. We can add a recommendation on joining efforts to set up a continental dataset for South America, which can facilitate the intercomparison / validation of models with scales ranging from regional to global.