Reply to comments by anonymous reviewer #3

We thank reviewer 3 for his/her thorough review of our paper, which will help to improve the manuscript and the underlying study for the HESS readership. We will go over his/her comments point by point, with the comments in roman and our reply in italics. The specific actions we intend to perform in order to improve the paper are underlined.

Bierkens et al. present a simplified analytical methodology for first order approximation of the impacts of groundwater pumping on streamflow and ultimately groundwater sustainability. In addition to summarizing the methodology they provide global mappings of streamflow depletions and sustainable pumping limits. While I appreciate that the authors were very clear throughout the manuscript that this methodology is intended to be approximate, I still have very significant concerns and I don't feel that the manuscript in its current form has demonstrated that this approach is adequate to support the types of groundwater sustainability findings that are presented in figures 7-9 for the following reasons:

1. The approach presented here relies on a myriad of simplifying assumptions. While the authors do try to be very transparent in these assumptions, this does not make them less concerning. Specifically, the steady state approach and the distributed well locations are big areas of concern in my opinion. For large scale aggregated analyses of declines this might be okay but for stream aquifer interactions well placement and timing is very important. The key advance of this paper is groundwater surface water interactions and therefore I think the bar is higher for some of these assumptions.

We thank the reviewer for recognizing that we are clear about underlying assumptions. At the same time the reviewer is concerned whether the assumptions are such that no credible large-scale assessments can be made of the impact of groundwater pumping on streamflow. We start by saying that although well placement with respect to surface water is important at local scales, we argue that if many wells are considered, some are closer and other further away from the surface water bodies and that this may be accounted for in a lumped conceptual model aiming at describing the regional response (compared with lumped rainfall-runoff models for entire catchments) as developed in our paper. However, we understand this concern and we will add additional validation datasets (see hereafter), also related to the impacts on surface water, to assess the degree of applicability of our approach.

2. The authors present a sensitivity analysis for their approach which is a helpful illustration of the relationship between variables. However, for me this really only demonstrates that the general interactions are in the correct direction, which follows directly from the equations they used. Much more concerning to me is the uncertainty of the inputs to these equations at the spatial scales presented here and whether reliable estimates for some of the parameters can be generated at all. For example, how accurately can bed slope and bed elevation be captured at this resolution globally? How sensitive are the final results to the uncertainty in these values?

We still feel that the current local sensitivity analysis is insightful, as some relationships may follow from the equations in hindsight, but the behavior can nevertheless be rather complex, which is illustrated by the analysis. Regarding the uncertainties: the slope of the stream is not part of our model, but the stream-bottom elevation is. The uncertainty about stream bottom elevations is however not only a problem for our approach. It is also an issue for groundwater modelling at regional scales even. In most cases, even in regional modelling studies, but certainly for global modelling (Schulze et al., 2006; Sutanudjaja et al., 2018), stream dimensions are taken from geomorphological laws relating stream width and stream depth to bankfull discharge (Leopold and Maddock, 1953), which is the yearly maximum discharge with a return period of 2-3 years. In the parameterization for the global application, we use a stream dimension data-set from PCR-GLOBWB 2 (Sutanudjaja et al., 2018) that is based on this approach. <u>To investigate the impact of uncertainty of bottom elevation d and the other uncertain parameter C, we will add a global sensitivity analysis elucidating the sensitivity of the results in Figures 4 and 6-8 to these parameters</u>.

3. My biggest concern here is that the most important metrics that the authors are highlighting in their findings are not well validated. The authors present primarily comparisons to other global models which rely on similar assumptions and are working at similar spatial resolutions. It seems like it should be expected that the results here would be 'remarkably similar' (line 315-318). Before jumping to a global analysis I would like to see some rigorous evaluation of the model in some of the many heavily studied aquifers across the world comparing to regional models and observations. For example, observational groundwater ranges are reported for a few aquifers (Lines 379-382) but the authors only note that 'our estimates are in the lower end of those observed ranges' I think a much more quantitative comparison is need here.

We agree that a further critical evaluation of the global results would be beneficial to scrutinize our global results. <u>In the following version of the paper we intend to add a separate paragraph that compares the global results to: a) observed time series of groundwater levels and streamflow for part of an aquifer that is known for heavy groundwater exploitation (e.g., part of the Ogallala aquifer); b) to a regional high-resolution model study that produces both streamflow depletion estimates as well as groundwater taken out of storage as a result of pumping; c) to groundwater depletion rates and streamflow depletion rates as calculated by a global groundwater model (de Graaf et al., 2019).</u>

4. Furthermore the validation that is provided here is really focused on groundwater depletions and I think the validation of the stream aquifer interactions or sustainable limits to groundwater withdrawals (the highlight of the paper) is lacking. If the main purpose of this work is to get to sustainability estimates and to connect to streamflow then these are the parts of the methodology which must be most thoroughly evaluated. I realize that this information is not available globally (hence the novelty of this work). However, I don't see any reason why these behaviors cannot be rigorously and quantitatively evaluated in some example locations for which data or models are available.

Directly validating the limits, e.g,. q_{crit}, t_{crit}, q_{eco} is not possible, regardless which approach is used to assess these. These limits will always be based on model estimates, regardless of the model used. Perhaps a local pumping experiment could provide local estimates, but otherwise they are non-observables that nevertheless provide guidelines to sustainable groundwater withdrawal when obtained with local models and threshold values to compare regional-scale pumping rates with in large-scale groundwater status assessment. However, we are able to provide sensitivity analyses (see answer to point 2 above) and additional evaluation of both groundwater and streamflow depletion or depletion rates (see point 3 above) to evaluate the underlying (lumped) conceptual model framework.

5. Finally, the sustainability language in this paper is concerning to me. First of all because sustainability is a very subjective topic and it's not clear that the first order type approximations used here can really get at true sustainability. Second of all because I think these results can easily be misinterpreted based on how they are presented here. The authors do try to specify that this approach is only for first order approximations, but if that is the goal here then I think they should focus on using this methodology to provide ranges of potential groundwater depletions and stream interactions, and not be using this to present things link groundwater limits which can very easily be misinterpreted and miss-used.

In accordance with the objections of the other reviewers about our limited notion of sustainability (mostly physically defined), <u>we will refrain from using the term in our paper in relation to stream-groundwater disconnection.</u> We believe that the limits presented in Figure 9 are still of use, because they are not to guide local withdrawal rates, but rather can be used in global studies on the state of groundwater use. <u>We will provide this caveat in the next version of the paper.</u>

Overall, I do think this is a well written paper that is clearly presented and easy to follow. Unfortunately, I am not convinced about the validity of the approach, and as a result the findings that are presented. I think a much more rigorous evaluation of the methodology is needed including quantitative analysis of every metric that is going to be presented in the findings. I completely understand that this methodology is intended to be approximate and will not perform as well as regional integrated models or intensive observational studies. However, I think these should still be the bar for comparison so that users of this approach can fully understand its strengths and weaknesses of the simplified approach, and so that any metrics that are too uncertain are not included

We thank the reviewer for the complements and acknowledge his/her concerns that echo what has been stated under points 1-5. We are certain that by providing additional sensitivity analyses, scrutinizing the global results with a more rigorous evaluation with data and the results of regional models and by removing the frame of groundwater sustainability, we will be able to address his/her concerns successfully.

References

de Graaf, I.E.M., Gleeson, T., van Beek, L.P.H., Sutanudjaja, E.H. and Bierkens, M.F.P. (2019). Environmental flow limits to global groundwater pumping. *Nature* 574, 90-108.

Leopold, L. and Maddock, (1953). *The hydraulic geometry of stream channels and some physiographic implications*: Professional Paper 252, United States Geological Survey.

Schulze, K., Hunger, M and Doll, P. (2006). Simulating river flow velocity on global scale. *Adv. Geosci*. 5, 133–136.

Sutanudjaja, E. H., van Beek, R., Wanders, N., Wada, Y., Bosmans et al. (2018). PCR-GLOBWB 2: a 5 arcmin global hydrological and water resources model. *Geosci. Model Dev*. 11, 2429–2453., 2018.

.