Reply to comments by reviewer Grant Ferguson.

We thank Grant Ferguson for his thorough review of our paper and his insights, which will help to improve the manuscript and the underlying study for the HESS readership. We will go over his 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.*

Bridging the gap between global models that have used a water budget approach and more detailed numerical models at the regional scale by considering depletion and capture relationships is an important step forward. The approach used in this study holds promise in addressing the problem of the water budget myth (Bredehoeft, 2002) at the largest scales.

We thank the review for his encouraging words.

However, there are some issues with how this work is framed in terms of sustainability and renewability along with some technical issues with the model. The authors of this study cite papers that are inconsistent in how they define renewable groundwater resources. Bierkens and Wada (2019) use a mean residence time as a measure of sustainability (line 50) while Wada (2016) uses a recharge-based approach (line 52). These definitions of renewability are problematic for multiple reasons. First, as pointed out by Bredehoeft (2002) and shown in the current study, renewal of groundwater is not restricted to background recharge but can also come from the reversal of hydraulic gradients at groundwater-stream interfaces once pumping begins. Second, mean residence time under background conditions is a function of flow system size is not connected to declines in water levels or streamflow in a simple way (Ferguson et al., 2020). Furthermore, when pumping a well to steady-state conditions (i.e. 100% capture) there is inevitably a portion of that groundwater that is non-renewable due to the cone of depression that develops to draw water towards the pumping well. The definitions of renewability used here are not useful in the context of groundwater management and are not necessary to support the ideas put forward in this manuscript. Removing discussion of these ideas will help to keep the focus on the problem of capture and depletion.

The reviewer refers to an important point in that renewability in terms of recharge or mean renewal time may be debatable, as indicated by one of his recent commentaries in nature Geoscience (Ferguson et al. (2020). We do not want to engage in a debate about the proper definition of non-renewable groundwater use here, as it is not necessary for this paper as the reviewer rightfully states. The introduction was to make the point that groundwater overuse leads to groundwater depletion, which may be seen as non-renewable groundwater use if it takes a long time of the water taken out of storage to recover (viz. recovery times as suggested by the Ferguson and others in their papers). *Therefore, we have changed the phrasing the introduction* *to: "This has greatly intensified the dependence of irrigated crops on groundwater withdrawal (Wada et al., 2012) and caused a steady increase of groundwater depletion rates (Wada and Bierkens, 2019). Recent estimates of current groundwater withdrawal range approximately between 600-1000 km^3 yr^{-1} leading to estimated depletion rates of 150-400 km^3 yr^{-1} (Wada, 2016)."*

The definition of sustainability is problematic because of its specificity. Complete disconnection of water tables from streams as described in lines 87-92 is without a doubt problematic in humid and sub-humid areas but serious issues that would also be deemed unsustainable may occur before this happens, notably dry wells. This also creates issues with using groundwater in semi-arid and arid areas where losing and ephemeral streams exist, and groundwater flow systems exist on a larger scale than the 5 arc-minutes considered here. There are a variety of different conditions that need to be met to ensure sustainable development. Less rigid metrics for sustainable development of groundwater are likely more appropriate. The conditions put forth by Gleeson et al. (2020) that require maintaining water levels and flows above critical flow is vague but points to the need to understand disparate goals from various stakeholders and the unlikeliness of solving this problem with global models and one-size-fits-all metrics. As a community, we need to stop thinking in black and white in terms of sustainability. The authors can resolve this by focusing more explicitly on water tables and streams disconnect as an undesirable outcome rather than linking this disconnect to a definition of sustainability.

We concur. Reviewer #1 made the same objections against our use of the term physical sustainability. Therefore, we will not use the term in the next version of the paper and focus on stream-groundwater disconnection. We do however need to use a term to distinguish between the two regimes to avoid lengthy descriptions and have decided to use "stable" ($q <= q_{crit}$) and unstable ($q > q_{crit}$) withdrawal regimes.

The ability of the model to reproduce observed depletion rates is debatable because the time to full capture isn't properly considered in the model application. The simulation assumes that steady-state conditions existed before 2000 but depletion issues were known well before this time (Konikow, 2013). The match with GRACE (lines 339- 350, Figure 5) data is coincidental because many regions should be on a later portion of the capture trajectory shown by Konikow and Leake (2014). Testing the model against observations would require more careful consideration of initial conditions and choice of simulated period. This may not be possible given the data available. However, presenting the simulation as an illustration of what would happen if pumping started in the year 2000 with no prior development is still a powerful demonstration of the capabilities of this model.

We apologize for the misunderstanding that we seem to assume that not depletion occurred before 2000. This is not so. We have compared with GRACE only for the areas where we have that $q > q_{crit}$, while assuming that in that case $t > t_{crit}$ (exactly because we assume previous groundwater development). Thus, we compare depletion rates under this assumption with observed depletion rates from GRACE. We will make this more clear in the next version of the paper.

It is not surprising that this approach reproduces similar patterns to other global models of groundwater depletion (lines 315-320). The assumptions and approaches are not that different in the models mentioned. A comparison of the results presented here to large-scale numerical models may provide a better test of model performance. Condon and Maxwell's (2019) model examining the impacts of groundwater pumping on streamflow over a large section of the USA at a 1 km resolution provides such an opportunity. There are assuredly some differences in computation times but the numerical approach will likely be superior in resolving hydraulic gradients and could likely be done at a global scale in the near future.

We believe it will be some time before the results of the model of Condon and Maxwell (2019) will be applied globally. Also, our approach is meant to make quick inferences and be used in e.g., regional hydroeconomic simulation and optimization approaches, requiring close to instantaneous results when applied. Nevertheless, our approach should be sufficiently trustworthy. Therefore, also following the suggestions of reviewer #3, we will extend the validation of our results. <u>We intend to compare the aquifer and streamflow depletion (rates) with a) observations from and affected aquifer (e.g., Ogallala) and 2) with large-scale numerical models</u>. <u>We intend to compare to the results to a global groundwater model (de Graaf et al., 2019). We will contact Laura Condon to see if we can obtain the aquifer and streamflow reduction results (Figure 2A and 2B) of Condon and Maxwell (2019)</u>.

Furthermore, the analytical approaches reviewed by Zipper et al. (2019) are not restricted to single wells, as suggested in line 116. Invoking superposition with some of those concepts may provide another path forward to study capture and depletion at large scales.

This is a valid point, and we agree that it could be used by analyzing multiple wells by assuming superposition. <u>We will acknowledge this in the next version of the paper</u>. It remains however not possible to include the change in groundwater-surface water interaction from connected to disconnected in their approach.

It is unclear that the approach used in the current study is "likely the simplest analytical form that can be devised to describe the effects of groundwater pumping at the larger scale" (lines 437-438). Objectively deciding the level of detail that effects of pumping need to be captured does not seem possible. Rather than making such claims, a more in-depth consideration of how the global approach presented here compares to numerical models or analytical techniques at local and regional scales might provide important context for this work. Such a comparison may help to guide future efforts in advancing large-scale groundwater modelling.

Reviewer #1 also took issue with this this claim and <u>we will delete this sentence from the next version of the paper</u>. We hope that a more extensive validation in the following version of the paper will help to decide whether our lumped conceptual model is not too simple to provide reasonable estimates of aquifer and streamflow depletion at larger scales.

This is a potentially important study in understanding large-scale groundwater depletion. While there are unresolved questions on the effectiveness of this approach is due to issues with initial conditions in the simulation, qualitatively it looks promising. The relationship between the results presented here and the threshold between sustainable and unsustainable development of groundwater is debatable. However, disconnection of water tables and streams is a clear indicator that groundwater pumping has resulted in an undesirable outcome and other thresholds may have already been passed.

We thank the reviewer for this encouraging final paragraph of his review.

References

Condon, L. E. and Maxwell, R. M.: Simulating the sensitivity of evapotranspiration and streamflow to large-scale groundwater depletion, Science advances, 5(6), eaav4574, 2019.

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.

Ferguson, G., Cuthbert, M. O., Befus, K., Gleeson, T., & McIntosh, J. C. (2020). Rethinking groundwater age. *Nature Geoscience*, 13(9), 592–594