Reply to comments by anonymous reviewer #1

We thank reviewer 1 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.

Summary

This study investigates the sensitivity of a linear reservoir-based model for groundwater pumping from unconfined aquifers and streamflow depletion, as well as applies this model for discretized cells of ∼100 km2 globally. Steady-state hydrologic parameters are input into the model with the outputs tracking the changing groundwater and surface water heads, presumably only considering a single model cell. A definition of sustainability is applied to the conceptual framework to explore the global spatial distribution of various model outputs and metrics.

Thanks for this summary that is mostly correct, except for the fact that it is stated that we apply this model to a single model cell. This is not correct. When we apply the framework globally, cells are connected in the sense that inflow in each cell depends on streamflow coming in from upstream cells.

1. I strongly disagree with the definition, and connected implications, of physical sustainability in this manuscript. According to the definition that groundwater pumping is sustainable so long as it doesn't cause the water table to disconnect from a surface waterbody is extreme. This definition means that nearly the entire flow of the river (i.e., Qi) could be extracted from the groundwater pumping and still be considered sustainable. A similar argument could be made that any streamflow lost due to groundwater pumping would be not sustainable, but the opposite is not necessarily the definition of a sustainable pumping regime (physical or otherwise). A dry or reduced flow river is not emblematic of sustainable abstractions in my opinion. Figure 1b can represent a physically unsustainable system, as the lost streamflow could lead to negative environmental effects downstream and could also cause feedbacks with downstream groundwater-surface water interactions. It would be fair to state that qcrit in this analysis is indicative of certain unsustainable hydrologic conditions, but it is not the threshold between sustainable and unstainable in either this conceptual framework or the real world.

*The notion of sustainable groundwater withdrawal is indeed complex and highly debatable and its definition still in development, as shown in a recent review by Gleeson et al. (2020). We maintain that as long as streams are connected with the phreatic surface and pumping is such that it will not lead to a groundwater-stream disconnection, that this is a **physically** sustainable system, where an equilibrium water table will develop. This notion of physical sustainability is also used in the Gleeson et al. (2020) paper. We agree however, that this may still lead to damage to ecosystems or downstream effects on groundwater depth etc. Also, we agree that due to the simplicity of our lumped model approach (<u>we will introduce the term lumped model in the abstract and introduction in a next version of the paper</u>), we* *do not account for the fact that disconnection may occur first higher up in the stream network and that disconnection is a spatiotemporally heterogenous process. We state this in our paper (lines 102-103). Regardless, the aim of this paper was not to pose an analytical framework for groundwater sustainability. Thus, the suggestion of the reviewer to delete references to sustainability and focus on the framework inspecting large-scale effects of groundwater withdrawal on surface water and groundwater including stream-aquifer disconnection is taken to heart. <u>We will remove the term sustainability from the manuscript and refer to physically stable and unstable pumping regimes instead.</u> The term stable then refers to pumping rates resulting in an equilibrium water table decline and instable pumping rates resulting in disconnection between groundwater and surface water and persistent decline of groundwater heads and groundwater depletion.*

> The opportunistic simplification of "capture" in this study is not complete, and the water budget and simplicity of the approach do not address capture in a sufficiently meaningful way to allow the application at the global scale to inform pumping management plans.

*Indeed, it is not complete. We do not take account of the impact of water table decline on evaporation and **diffuse** groundwater recharge. However, it is safe to say that the impact of water table decline on diffuse recharge is second order compared to the impacts on streamflow, in particularly in more semi-arid to semi-humid regions where soil saturation by shallow water tables is limited. Of course, our model does include the impact of water table decline on groundwater discharge to the stream and, in case $h < h_s$, recharge from the stream to the aquifer (**concentrated** recharge). So capture is taken into account in essence, certainly at the larger-scales that we state that our lumped model is said to operate on. We therefore, do not see any compelling argument why it cannot be applied at the global scale. Obviously, it will not inform pumping management plans for a single well or multiple wells at a local scale (we do not claim this), but may be informative for regional-scale effect studies of many wells.*

> As a somewhat connected note on this topic, the study does not need any definition or use of sustainability. If the study were instead posed on the potential disconnection of groundwater from surface water, then there would be no need for the value-loaded aspect of sustainability definitions. The "critical" outputs could be relabeled as "disconnection" or extreme flow reversal outputs.

We agree with the reviewer (see our answer above).

2. What are the hydrologic restrictions of the constant hydrologic inputs? Importantly, it appears that the streamflow velocity remains constant while the depth and discharge can change. This suggests that the Q was not connected between cells, such that the pumping analysis was only providing information for each cell individually, such that Qi is constant and unaffected by pumping. This was not stated clearly in the text. This is important for then later calculations of depletion, comparisons with observational data (i.e., GRACE depletion rates), the delineation of "sustainable" vs "unsustainable" areas or watersheds, and the "global limit to sustainable gw pumping". These calculations

represent nearly all of the location-specific results, and the lack of hydrologic connectivity is especially concerning for the calculation of qeco (Eq 4).

Streamflow velocity is indeed assumed constant. This is an assumption made to keep the relation between stream discharge Q and stream water elevation h_s linear, resulting in linear ordinary differential equation to be solved. This assumption is further supported by the fact that streamflow between (larger size) rivers and streams and for the same streams/rivers over time is surprisingly constant, often varying between 0.5-1.5 m/s (see e.g., Figure 2. In Schulze et al., 2005). It is also based on the fact that for a rectangular channel it follows from Manning's equation that the derivative $dV/dh_s \sim h_s^{-1/3}$ which results in small changes in velocity with water depth for larger water depths (even more so for a trapezoidal channel). This does not mean however, that discharge in our model does not change as a result of groundwater pumping. It does! So, there is certainly a connection between pumping and streamflow and the impacts on environmental flow. The approach produces large-scale changes to downstream discharge due to groundwater pumping in an area given upstream inflow to this area. What is not done in the global application is propagating the accumulated effects of pumping, i.e., by analyzing cell-by-cell following the large-scale streamflow network from upstream to downstream, although we could have been done this. Instead, in our global analysis, upstream withdrawals from surface water and groundwater are included as they come from PCR-GLOBWB. They would also be implicitly included in case an observation-based streamflow dataset (e.g., Barbarossa et al., 2019) would have been used. <u>We will make this more clear when introducing the global application and discussing its connection with previous PCR-GLOBWB results</u>.

3. This study needs to connect more clearly with the Zipper et al. (2019) paper rather than an offhanded statement on "a single well-network" method. This study is also applying a one well one stream methodology that fits within the levels of complexity tested by Zipper et al. Treating the aquifer as an infinitely deep linear reservoir with uniform drawdown is less informative when applied to real locations (i.e., in the spatial analysis in this study) than the analytical approaches in Zipper et al. The distance a pumping well is from a stream is critical to calculating the streamflow and aquifer depletion, and the Zipper paper certainly serves as a foundation for global hydrologic studies that already have basically all of the information needed. Similarly, superposition was not mentioned in this study, but it could surely provide a very simple but powerful tool for calculating more realistic drawdowns. Forcing all drawdown across the model cell to be equal with this conceptualization also sets a very optimistic limit for what is being inappropriately labeled critical metrics for "sustainability".

It was not our attention to supply an offhanded statement and dismiss the work of Zipper et al (2019). We certainly see its value and 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. Our approach is indeed different in scale and less informative for local impacts. This manuscript does not present a single well single stream method. Instead, it is a lumped model (as opposed to a spatially explicit model of Zipper et al. (2019)) applicable to larger scales where all wells are lumped into a diffuse sink, assuming indeed the same drawdown. That this would lead to an *optimistic limit is not clear to us. It underestimates the effects of the wells that are relatively close to a water course and underestimates the effects of wells further away.*

4. Also of concern, relating to the Zipper paper, is the rather haphazard definition of the interaction term in this study, F. It includes the streambed conductance, a very difficult to constrain and important parameter, while also adding other geometries. If depletion is "often highly heterogeneous and incorrect estimates can lead to errors in estimated streamflow depletion (Fleckenstein et al., 2006; Irvine et al., 2012; Lackey et al., 2015)", as stated in Zipper et al., then I have a lot of trouble trusting the two versions of F (and J) used in this study, as neither sources were meant to provide such information on streambed connectivity to an aquifer. As such, the two sets of maps are pretty samples from an unknown distribution with unknown uncertainty. Also, the maps only show the actual values and never provide any information on the relative similarity/dissimilarity of the two calculations (other than being "striking", but not explained which is more realistic). Subtracting the two datasets and providing a map and histogram would give a sense of how important the unknown response time input for J and F is. These inherently include a length that may be inconsistent with the way this study was discretized. This again makes me question the utility of the many global outputs in this study.

*Two points are made here, and we will answer them consecutively. First, the statement that the C or J value is haphazard. Well, it is not actually. In classical drainage theory, the flow geometry-related resistances and streambed resistance are often lumped into a single parameter called drainage resistance akin to our C-parameter (see e.g. Ernst (1956) and Kraijenhoff van de Leur (1958)) that can in fact be related to the domain geometry and hydraulic parameters and thus have a semi-physical basis. The drainage resistance parameter is also related to the characteristic response time of Cuthbert et al (2019). We therefore follow previous approaches of lumping groundwater flow. It is evident that water table and streamflow depletion decline due to pumping are sensitive to local heterogeneities. The fact that we use a lumped model does not mean that we negate the existence of such local heterogeneities. We do not resolve them because we aim to model large-scale (aquifer-scale, 100 km^2 grid cells) average responses to large-scale pumping. This is analogous to a lumped rainfall-runoff model of a catchment: using it does not mean that one denies that within a catchment heterogeneities of runoff response exist. Instead, it chooses to model a catchment total or average runoff response, often precisely because the local heterogeneities **cannot** be resolved. Also, the analytical depletion formulas used in e.g., Zipper et al. (2019) equally assume homogenous hydrogeology. Finally, streambed conductance is poorly constrained indeed, and this affects any groundwater modelling effort, both analytical as well as numerical.*

The second issue is the maps comparing the depletion rates with the two datasets of C. <u>It is a good idea to have a difference map between the PCR-GLOBWB C results and those obtained from Cuthbert et al (2019). We will do this in a next version of the paper.</u>

5. More information on the input datasets would be useful. For example, a description of the dataset used to apply "realistic" pumping rates for the unconfined aquifers needs to be at least stated rather than requiring the reader to track it down elsewhere. The

validity of these pumping rates sets the validity of all of the spatial results. Uncertainty in these pumping rates and resulting uncertainty in the results would also be useful, as the focus on mapped outputs implies the targeted impact of the global analysis is site-specific rather than global.

We thank the reviewer for pointing this out, and we are happy to comply. <u>Apart from Table 3, we will provide a Supplementary Information file where we will provide maps of the input and parameter files used for the global analysis.</u>

6. The connection between the PCRGLOB-WB (2) model needs to be stated in the beginning rather than in the discussion. The differences and novelty of this study needs to be presented at the beginning with the full context, rather than stating the similarity between this analysis and the previous modeling work "is not as surprising as it seems". The differences need to be VERY CLEARLY presented. Along this reasoning, the comparison of the depletion rates between this study and the former work needs to be more detailed. How many of the inputs between the models were different? How many of the equations? Are the integrated depletion rates for the globe smoothing over larger differences?

We respectfully disagree with the notion that the connection between PCR-GLOBWB needs to be stated up front. As we state in the discussion, a global application of the approach could also have been parameterized with the outputs of other global hydrological models or even global datasets based on observations and remote sensing. So, apart from us authors also being responsible for building and maintaining PCR-GLOBWB, there is no intended connection.

What is true is that the stream-aquifer interaction equation (Equation 2) is similar to what is used in PCR-GLOBWB, but also in other global hydrological models such as WGHM and even in the parameterization of the river package of MODFLOW. This is exactly what we state in Chapter 2 right at the beginning. Otherwise, PCR-GLOBWB is very different. It does not use any of the analytical solutions shown in Table 1, but rather uses a spatio-temporal discrete approach (time explicit) to solve the water balance equations. The analytical expressions are based on time-invariant forcing of the system and thus simplified. Still, they provide similar results close to instantaneously, instead of after days of numerical integration. In hindsight, this similarity can indeed be explained by the linearity of the groundwater reservoir that is also present in PCR-GLOBWB. <u>We agree that this discussion is best done earlier and we will move the discussion about the similarity in results between our model and PCR-GLOBWB to the results section.</u> We will also provide a pixel-by-pixel difference map with PCR-GLOBWB depletion in the Supplementary Information to add more detail and additionally with depletion rates from a global groundwater model (De Graaf et al., 2019) (also upon a request of Reviewer #2).

7. The comparison with GRACE data needs further development. How were the averages of depletion upscaled for these aquifers and some identification of the target areas would be useful? What are the unlabeled dots in Fig 5? What areas do they represent? What do the large misfits between the depletion rates, especially for the low rates from this study, indicate about the model performance and limitations? The issue of total

water storage changes and an infinitely thick unconfined aquifer could be discussed in more detail.

We thank the reviewer for these suggestions. <u>We will add a map with shape files of the aquifer systems identified in the scatter plot to the Supplementary Information. We will also identify the unexplained dots and add more information on how the average depletion rates were calculated.</u> We stress that the thickness of the aquifer is not an issue in our lumped conceptual model. We only present rate of storage change and do not presume to make predictions of when an aquifer becomes depleted without knowledge and inclusion of aquifer thickness or maximum pumping depth.

8. The focus of the discussion of uncertainty on confining conditions is not allencompassing, nor does it even assuage my concerns on the way the aquifer system was developed. Insufficient description of the various geometries and model inputs make it difficult to fully question the role of confined vs unconfined aquifers. An infinite depth unconfined aquifer system as the domain with an area the size of the grid cell is somewhat clear. Are the pumping rates only for the unconfined aquifer? If so, then why compare to GRACE TWS, as those are heavily tied to confined aquifer pumping in many areas? Justifications are lacking and explorations of the uncertainty of the effect of unconfined aquifer with infinite depth/storage on the results is missing from the analysis.

We acknowledge that we do not take into account that many aquifers are confined. Ignoring that an aquifer is actually confined, like we do, would have a big effect on groundwater-surface water interactions and would likely underestimate storage decline. Still, we can compare with GRACE to see how "wrong" we are. <u>We will extend the discussion around the possible effects of ignoring confined aquifers when discussing Figure 5 about the comparison with GRACE.</u>

9. ". . .likely the simplest analytical form that can be devised" is amazingly pompous and immediately false. Bragging at its finest. (Line 448).

This is not a very courteous way of saying that we overstate our case. <u>We stand corrected and will remove the sentence.</u>

10. The definition of F is different between Figure 2 and Equation 2. Reversing the inequality with a negative sign in Eq 2 results in problems. Figure 2 appears to be the correct definition, where negative F represents streamflow depletion and positive as baseflow. With Eq 2, h > hs leads to –F whereas hh > hs leads to –F whereas h < hs leads +F. In Eq 3, it appears that +F should lead to more streamflow, such that Fig 2 has the correct definition of F. A statement that +F is inflow into the surface water or something to that effect could help the reader follow this definition. Fig 2 should match the equations in the text and be consistent with the rest of the math. Similarly, some variables in Table 2 are capitalized when they are not in the text.

We thank the reviewer for noticing this inconsistency. <u>We have aligned the sign of F in Figure 2 with the equations in the text</u>. If F is positive it contributes to groundwater (depletes *streamflow) and when negative to streamflow (groundwater discharge). <u>We have added a sentence to this effect to the text. We have also corrected the inconsistency in low-upper case between Table 2 and the text.</u>*

11. Numerous typos and misspellings throughout the paper. Lines 65, 68, 85, 101, 139, 149, 263, 283 (? or are tenths of years impressive?), 347, 388, 486, 717. 12. Ln 299 – inflow is flow in or out of the stream? Unclear here and elsewhere as this depends on perspective (towards surface water or towards groundwater?).

We thank the reviewer for noticing. <u>We have corrected the typos</u>. "tenths of years" is "Dunglish" for "decades". <u>We will also better clarify what inflow means</u>.

11. Ln 277 – Eq A30 mainly states that these fluxes negate each other, but the relationship of the ratio of these components is not known as q appears in this equation twice, unless additional assumptions are made (i.e., the ratio of the non-q components are equal to zero).

We don't think so. The capture part can be calculated (which is actually q_{crit}), which is always smaller than the pumping rate q in case of $q > q_{crit}$. Once that is known, the remaining part comes out of storage, which also follows directly from the groundwater decline rate (A24). The ratio can be calculated if pumping rate q is known.

13. Ln 806 – distance, not difference

Thank you for noticing. <u>Corrected</u>.

14. Ln 812 – it can also be set to other elevations, such as is implied in this study where pumpable groundwater exists below the streambed elevation.

This is a correct observation. In that case the analogy with the Kraijenhoff van de Leur (1958) solution breaks down because the latter does not account for disconnected streams. We do assume that J remains the same though. <u>We will state this assumption in the revised paper.</u>

15. All map figures are clipped to middle latitudes in the pdf I reviewed. I am unsure if this was intentional or not, but it seems arbitrary given the global extent of the analysis.

This clipping was done intentionally. The reason is that all the major groundwater pumping and depletion occurs between 60^o north and 60^o south. This allows us to show the major features while saving space. All global numbers are based on integrating across the entire globe however.

16. Separately on Qi, depending on the size of the watershed/catchment of interest, it seems strange to attribute the need for these to mountainous areas. Zero-order watersheds seem to also be depicted in Fig A1, which is absolutely not expected.

Figure A1 is just a schematic and the tributaries do not represent first order catchments at the scale of a lower river basin. Mentioning mountainous areas as source of inflow comes *from the fact that mountain front recharge is an important source of recharge in many of the heavily irrigated semiarid regions of the world.*

References

Barbarossa, V., Huijbregts, M., Beusen, A. et al. (2018). FLO1K, global maps of mean, maximum and minimum annual streamflow at 1 km resolution from 1960 through 2015. *Sci Data* 5, 180052.

Cuthbert MO, Gleeson T, Moosdorf N, Befus KM, Schneider A, et al. (2019). Global patterns and dynamics of climate-groundwater interactions. *Nat. Clim. Change* 9:137–41

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.
[1]

Ernst, L.F. (1956). Calculation of the steady flow of groundwater in vertical cross-sections. *Neth. J. of Agric. Sci*. 4, 126–131.

Gleeson,T., Cuthbert, M.O., Ferguson, F. and Perrone (2020). Global groundwater sustainability, resources, and systems in the Anthropocene. *Annu. Rev. Earth Planet. Sci.* 2020. 48:17.1–17.33

Kraijenhoff van de Leur, D. A. (1958). A study of non-steady ground-water flow with special reference to the reservoir-coefficient, *De Ingenieur* 19, 87–94.

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