Dear Editor,

Thank you for the opportunity to revise our manuscript according to the remarks of two additional referees. They both raise valid points and we have changed the manuscript accordingly. This means that we:

- Reframed the manuscript in accordance with the suggestions of Referee 4, i.e., we now state explicitly in the abstract, the introduction and the discussion that the analytical framework should be paired with a more complex global hydrological model to be used as a fast and first-order screening tool.
- Added more discussion on the cause for differences between the critical transition times obtained from Sutanudjaja et al (2018) and Cuthbert et al (2019) and the validation results (Referee 5).
- Corrected the minor issues noted by Referee 5.

In the following we will respond point-by-point to the comments by the referees. The comments are denoted in Roman, our response in Italics and quotes of added text in the manuscript in Roman red. Line numbers refer to the newly revised manuscript.

We think that the two rounds of reviews have greatly improved the manuscript and hope that it is now ready for publication in HESS.

We await your decision with interest.

With kind regards,

Marc Bierkens (on behalf of my co-authors).

**Anonymous Referee 4**

*I am not a groundwater modeler. I have read the discussion of the paper (comments of the reviewers and the response of the authors) and the revised manuscript. This is an interesting discussion, and the authors have addressed many of the comments of the reviewers. Hopefully one or more of the previous reviewers will be satisfied with these revisions. I have taken a higher level view of the paper, and feel that a simple twist to the presentation of the paper will make the paper much more acceptable and appealing, and will could provide avenues for further extension of the work. This change will require some moderate revision of the paper.*

*My own view is that the simple model they present here is so simple and is such a caricature of groundwater theory that it cannot form the building block for a bottom-up distributed groundwater model. In fact, the application of the simple model in this paper draws its power from observations or detailed predictions by a more detailed and sophisticated model. In other words, the simple model can at best be a good screening tool to elaborate on controls on large-scale sensitivities of groundwater and surface water to groundwater withdrawal. In other words, the simple model is not a standalone model, but is a top-down model that is used in combination with a more detailed, physically based model such as PRC-GLOBWEB2 in this case.*

*When I first read the paper, I got the wrong impression that this is about development and validation of a simple groundwater model. The discussion and the response of the reviewers have clarified to me that this is wrong interpretation of the paper. However, a reframing of the paper along the lines I suggested above would give a more correct interpretation of the paper, and will bring out more the novelty and usefulness of the simple model as a screening tool to assess groundwater sustainability at regional scales. In this new reframing the pairing of the simple model and the more complex model (PRC-GLOBWEB) presents a analysis framework for assessment of groundwater sustainability, and there is every opportunity to refine and improve both kinds of models as more data and observations become available.*

*In any case, I hope the authors and the previous reviewers agree with or appreciate this interpretation and this can contribute to a further moderate revision of the paper, which will make it much more appealing to reviewers and readers alike.*

We thank anonymous referee 4 for the supportive review and valuable suggestions to stage the paper differently. We respectfully disagree with referee 4 that the analytical framework is a caricature of groundwater theory, as there have been previous publications showing that at larger scales surface-water groundwater interaction behaves as a (piecewise) linear reservoir (see e.g., Savenije (2018), Hydrol. Earth Syst. Sci., 22, 1911–1916) and a similar parameterization is used in MODFLOW and several global hydrological models. Nevertheless, its lumped nature makes it difficult to directly link the C-value to flow geometry and hydraulic properties of the aquifer (although our appendix shows an example on how it can be). That being said, we agree with the suggestions of the reviewer, and we have reframed the analytical framework and present is a screening tool to be used together with a more complex model to quickly perform sensitivity studies focussed on regional-scale groundwater withdrawal impacts and sustainability. We have added the following lines:

Abstract lines 34,35: After a local sensitivity analysis, the framework is combined with parameters and inputs from a global hydrological model and subsequently used to provide global maps of critical withdrawal rates and timing, the areas where current withdrawal exceeds critical limits, and maps of groundwater depletion and streamflow depletion rates that result from groundwater withdrawal.

Abstract lines 39-41: Pairing of the analytical framework with more complex global hydrological models presents a screening tool for fast first-order assessments of regional-scale groundwater sustainability, and for supporting hydroeconomic models that require simple relationships between groundwater withdrawal rates and the evolution of pumping costs and environmental externalities.

Introduction lines 132-134: We envision that such an analytical framework, when parameterized with parameters and inputs from a more complex global-scale hydrological model, can be used as a screening tool for fast first-order assessments of regional-scale groundwater sustainability, and for supporting hydroeconomic models that require simple relationships between large-scale groundwater withdrawal rates and the evolution of pumping costs and environmental externalities.

Results lines 361-365: It should be noted that our results are obtained at only a fraction of the computational costs of global hydrological models: a few minutes at a single PC compared to 2 days on a 48-core machine with PCR-GLOBWB at 5 arc-minutes. Thus, the sensitivity to changing pumping rates or changes in recharge under climate change can be quickly evaluated.

Discussion and conclusions lines 609-615: We have introduced an analytical framework based on a lumped conceptual model that intents to describe to what extent groundwater withdrawal affects groundwater heads and streamflow under changing regimes of groundwater-surface water interaction. By feeding the framework with the parameters and inputs from a more complex hydrological model (i.e., PCR-GLOBWB), it can be used as a screening tool for regional-scale groundwater sustainability. i.e., by providing a rich tableau of hydrologically and ecologically relevant outputs at very limited computational costs.

We hope that with this addition, the value of our approach becomes more clear and attractive to the reader, as was intended by the reviewer.

**Anonymous Referee 5**

*Summary*
*Bierkens et al. present a novel analytical approach to assess the impact of groundwater withdrawals on the disconnection of streamflow and groundwater, that can be applied globally. The approach further facilitates the calculation of critical pumping rates that are needed to maintain the surface-groundwater connection. Results include an assessment of model parameters and qualitative comparisons to other studies.*

*Recommendation*
*I am a new reviewer to this submission, which has already undergone a first round of major reviews. Thus, my comments focus on how well the authors have addressed the reviewer comments - but I am adding a few comments that I deem important and may require a few (minor) adjustments.*
*In general, I think that the authors addressed most (if not all) major comments from the reviewers. Major changes include, e.g., a reformulation of the manuscript focus (from sustainability to the disconnection of streamflow-groundwater interactions), additional validation and comparison exercise using other data sets (originating outside the group), additional supplementary materials were added, among other minor revisions to address the reviewer's concerns. There is just one overarching concern, raised by all reviewers, that could be addressed in a slightly better way. I would thus recommend that the authors revise their response to this concern in another minor revision.*

*Issues with validation/Intercomparison*
*All reviewers requested additional validation/comparisons to other studies. In response, the authors added (i) a comparison to other input data sets (i.e., critical transition times from Cuthbert et al), and (ii) qualitative intercomparisons to other global (de Graaf et al.), continental (Condon and Maxwell), and regional/local (Wen and Chen) studies. The latter includes a local validation with streamflow and groundwater head trends for the Republican River basin in the US.*
*While the authors went through a lot of effort employing these other data sets for the validation and intercomparison - which is appreciated -, I believe this is the only important part that may need clarification. While I understand that these intercomparisons are difficult, especially since most of these variables cannot be measured, I think a bit more effort in explaining the differences and the expected shortcomings of the approach is needed.*

Thank you for the clear summary of our approach and the appreciation of the efforts we did to evaluate the approach during the last revision. We have added additional text trying to explain the differences between our approach and other models in the Results section and also additional notes about the applicability of the approach.

*This concerns, for example, the differences in critical transition times in Fig. 5 derived from the two data sets (Sutanudjaja et al.; and Cuthbert et al.). The differences here are between a factor 100 and a factor 1000. I am not an expert on this, but it is indeed "quite striking" (l. 399). The text does not really explain what the impact of these differences on the further* *results in the manuscript is (and if these are used at all). The explanation (l. 399-404) is not very convincing to me (but again, I am not an expert).*

We have tried to explain in more depth what is the cause of the difference between the Sutanudjaja et al, (2018) and Cuthbert et al. (2019) results. We have also extracted a conclusion about the limitations of using the approach in estimating critical transition and e-folding times.

Results lines 374-382: These differences can even add up to 2-3 orders of magnitude, which is extremely large. The reason is that the characteristic response times based on Cuthbert et al. (2018) are much larger (also up to 2-3 orders of magnitude) than those based on PCR-GLOBWB. Since the e-folding time in the stable regime is close to proportional to the *C*-value (e.g., Figure 3g), this is also true for the critical transition time. The very large differences in response times between these two datasets reveals that our method is only as good as its inputs and that critical transition times and times to full capture calculated with our approach should be interpreted with care and as order of magnitude estimates at best.

*Further, the intercomparisons to other, e.g., numerical models such as ParFlow-CLM, remain rather qualitative. I would be okay with that if it was intuitive, but the text barely touches upon the reasons for differences. Instead, the comparisons are lacking behind. E.g., l. 546f: "Figure S12 shows again that the analytical approach yields larger depletion estimates than ParFlow, but the results are more similar than with de global model of De Graaf et al (2019)." ... but likely for different reasons, no? Also, for the comparison to de Graaf, the differences appear (maybe only) larger outside of the US.*

Although it remains speculation without deep insights into the models of Condon and Maxwell (2019) (which we do not have) to explain the differences, also more quantitatively. Nonetheless, we have added additional text providing a possible explanation on why the results of Condon and Maxwell are closer than those of the De Graaf et al. (2019). The most likely explanation is that, apart from neglecting the lateral flow between cells, which is taken into account by both Condon and Maxwell (2019) and De Graaf et al. (2019), our approach also neglects the falling dry of water courses, which is taken into account by Condon and Maxwell (2019) but not by the De Graaf et al. (2019). So, the first omission results in overestimation of the groundwater level decline and the second omission by an underestimation, which therefore partly offsets the overestimation in case of Condon and Maxwell (2019).

Results lines 511-518: It is speculative at best to explain why the results of Condon and Maxwell (2019) are more similar. One possible explanation may be that the overestimation of decline rates due to ignoring lateral flow between cells in our approach is partly offset by the neglect of headwater streams falling dry under continuous pumping. This effect is included in ParFlow-CLM, which results in larger head decline rates that are closer to ours. The global groundwater model of De Graaf et al (2019) does not include this effect as streams in this model remain water carrying, even if the groundwater level drops below the stream bottom elevation.

*Along these lines, personally, I think that a few more notes related to the shortcomings of the approach and its expected performance may be required at the end / in the summary. This is especially, because applications of this analytical approach with other data sets / models are encouraged. So I think the reader needs to know (i) why differences to the aforementioned studies appear, and (ii) where the model is expected to perform well and where not. I think this concern is in line with concerns from all reviewers (esp. #1 and #3), who criticize some of the assumptions - even though they are explicitly mentioned throughout the text.*

Comparison of our results to other global and regional results does not reveal geographic differences between our framework's accuracy. We can however say something about the minimum scale (resolution) the approach is still producing reasonable results and also about the type of variables that can be estimated at what accuracy. For an explanation of the cause of differences with other approaches we refer to the earlier comments.

Discussion lines 621-630: The estimated global groundwater and surface water depletion rates were compared with observations and model results at various scales (support and extent), with mixed but overall favourable results up to the sub-basin scale. Results show that the analytical framework provides similar results to that of global hydrological models, but tends to overestimate the groundwater depletion rates when compared to groundwater flow models that account for lateral flow between cells. Also, without calibration, the critical transient times, i.e., the time from commencement of pumping till the detachment of the water table from the stream, as well as the related time to full capture, are order-of-magnitude estimates at best. Finally, when using global datasets, the analytical framework is limited to the sub-basin scale and too coarse for local-scale estimates.

<u>Some more minor notes</u>

- *Please check the consistency of units throughout the text (e.g. use SI unit "d" instead of "day" and unify "yr", "y" and "year"*

Thanks for noticing this. We have corrected it.

- *l. 657-670: this should maybe be moved to the methods*

We have considered moving it to the Methods, but, since this is actually a further elaboration of the results and an example of application, we feel it better fits the results section. Therefore prefer to leave it where it is.

- *l. 104-105: check sentence; "these transitions do not occur" and "is that ... is that"*

Thank you for noticing. We have corrected this.

- *l. 255: It should be "h_s(infinity)" I think*

Line 255 refers to Table 1. It already shows h_s(infinity) in the table.

- *l. 283: check sentence; "that are of interest to show" or remove "of"*

Corrected

- *l. 360: and?*

We are sorry, but it is not clear what is meant by this note.

- *l. 374: delete one "between"*

Corrected

- *Fig. 6+7: could the authors add what is considered 'negligible' ?*

We have added “(< $10^{-4}$)” to quantify what we consider to be negligible.

- *l. 545: "de" --> "the" (a little bit more Dunglish ;))*

We corrected this piece of Dutch creeping in.

- *Table S1: small v instead of large V for consistency with main manuscript?*

The capital V stands for any of the named output variables in Table S1, so it is kept as is.

*Concluding remark*

*In general, I am very appreciative of the work and fully support the notion towards large scale hydrology. In particular, I welcome the broad applicability of the approach, e.g. at various scales and with other models/observations - and concur with the authors that this approach may be useful for (i) bridging the time gap until global numerical approaches are ready, but (ii) also for benchmarking, intercomparisons and uncertainty analyses.*

We thank the reviewer for the kind words and for the effort to make this a better paper.