We thank the three reviewers for the positive feedback on the work we have invested in this manuscript and the detailed comments to further improve it. We certainly acknowledge the uncertainties involved in the presented results and recognize that the submitted abstract might create false expectations. We have thus rewritten the abstract as well as the conclusions and are now more precise about the explanation of statistically significant changes. Furthermore, we added a new table (Table 1) that provides an overview on which vegetation processes are implemented in which model, added a new Figure 2 that shows the ensemble mean, and extensively revised Figure 7 (former Fig. 5). We are certain that the presented manuscript is an important study of the capabilities and limitations of global hydrological models.

The following lists all comments of all three reviewers and our rebuttal in *italics*. Attached is a markup document that highlights the changes compared to the initially submitted document. Line numbers refer to the revised version of the document.

Referee #1

1.1

Although the title of the paper and the results emphasize model uncertainties, I think the primary result based on the analysis is that recharge is not very sensitive to climate change as only 15% of cells show significant increases or decreases in recharge based on pre-industrial baseline and only 8% of cells show significant change in recharge from current 1 degree to projected 3 degree condition.

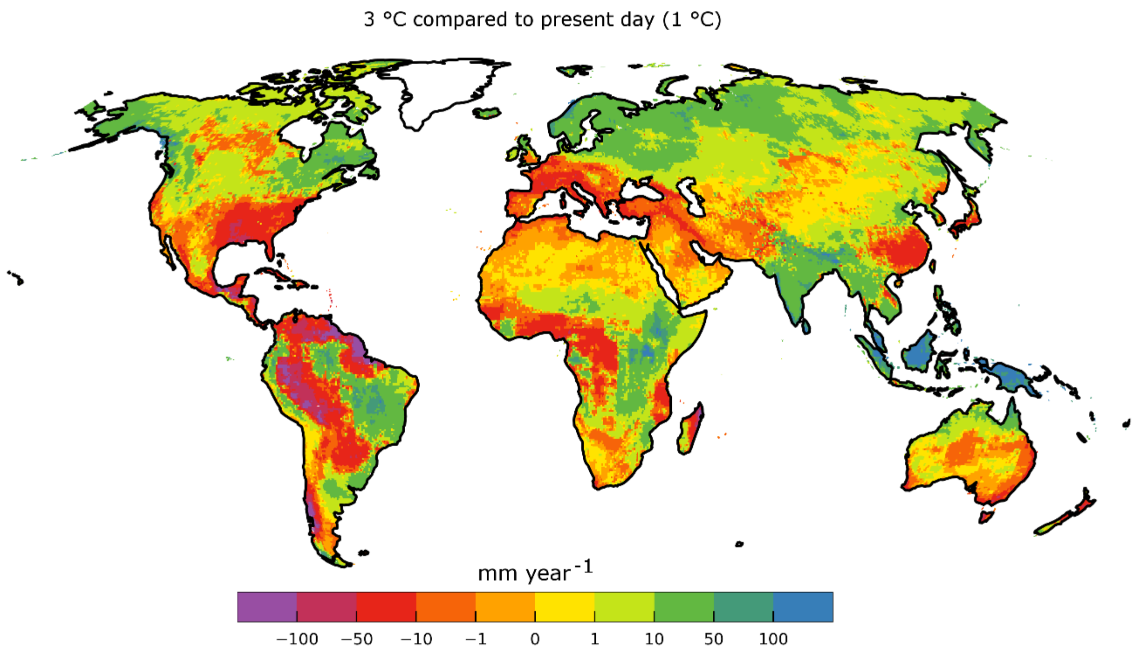
It is not correct to conclude from the fact the only 15% of cells show significant changes that groundwater "recharge is not very sensitive to climate change". This misunderstanding by the reviewer is caused by the various meanings the term "significant" has. In our paper, with "significant" we do not mean that there are "large" changes. We mean a statistical agreement or rather non-agreement between the two ensembles of simulated recharge, the one consisting of recharge computed by the various models under e.g. pre-industrial conditions and the other consisting of recharge computed by the various models at a certain global warming level (as tested here by a Kolmogorov-Smirnov test). To avoid this confusion, we have clarified this fact in Section 2.5 and for multiple references to the significance of the changes. We have also added a new Figure 2 that shows the ensemble mean difference between a 3° C warming and the present day without any statistical tests. We have also strongly modified abstract and conclusions to clearly express that significant refers to statistically significant (see response 3.1).

Section 2.5 now reads (Line 258 ff):

*"*A model ensemble allows us to consider the uncertainty in modeling physical processes as different model use different algorithms and parameters for computing groundwater recharge. To determine whether changes in GWR due to GW computed by the model ensemble are statistically significant, we use the two-sample Kolmogorov–Smirnov (K-S) test to compare the GWR values computed by all GHM-GCM model combinations under e.g., PI conditions with the values at the various GW levels. The use of a two-tailed t-test is not advisable in this setting due to the small sample size (max. 84 in this study). Because the K-S test does not allow to check whether the ensemble agrees on the sign of change in GWR, we apply an additional criterion to determine a significant change similar to Döll et al. (2018). A change is only marked as (statistically) "significant" if the K-S test indicates a significant difference and at least 60% of the model realizations of the ensemble (RCP, GCM and GHM combinations) agree on the sign of change (i.e. a decrease or increase". In case of a low significance, all models may show large responses to climate change while their agreement on the amount or sign of change is low.*"*

The new Figure and its description (Line 270 ff):

“To assess the impact of GW on GWR, Fig. 2 shows the ensemble mean change of GWR between the current 1°C world and a potential 3°C GW. We chose to express changes as absolute change rather than relative change because zero, or close to zero, GWR in some regions of the world leads to not defined or extremely large percentage increases and decreases (Fig. S1 and S2). The model mean shows large decreases of over 100 mm year-1 in South America and in the Mississippi Basin and decreases of up to 50 mm year-1 in the Mediterranean, East China, and West Africa. Increases of over 100 mm year-1 are prominent in Indonesia and East Afrika. Individual GHM-GCM model combinations compute much larger changes.



(new) Figure 2 Ensemble mean change in GWR [mm year-1] between conditions of present day warming of 1 °C GW and at 3 °C GW, averaged over the GWR changes of all GHM-GCM model combinations.

Ensemble mean changes as shown in Figure 2 may be low in some areas, but this could be due to large positive changes compute by some GHM-GCM model combinations being canceled by large negative changes by other model combinations. To assess the changes which show a high statistical agreement in-between the model combinations, we determine where computed changes of GWR are statistically significant (Section 2.5).”

1.2

It would be good to acknowledge that recharge is likely the most difficult component of the water budget to simulate because it is modeled as a residual, accumulating uncertainties in other water budget components.

We agree and accordingly have changed the abstract and introduction. (Line 27 and 60):
“Groundwater recharge is an important indicator for groundwater availability, but it is a water flux that is difficult to estimate as uncertainties in the water balance accumulate, leading to possibly large errors in particular in dry regions.” And “The simulation of GWR is possibly one of the most challenging components

of the water budget as it accumulates the uncertainties of all other components of the budget. Especially in semiarid regions, uncertainties in precipitation and evapotranspiration lead to considerable uncertainty in recharge. An additional factor in estimating groundwater recharge is the simulation of the groundwater table and thus capillary rise and focused recharge."

1.3

In addition, it is extremely difficult to simulate in semiarid regions because small uncertainties in precipitation and evapotranspiration can result in large uncertainties in recharge.

Has been addressed together with 1.2.

1.4

Many studies suggest that climate change will result in increased climate extremes (floods and droughts) that may result in increased recharge from focused rather than diffuse recharge; however, it seems that few of the models consider focused recharge.

We agree, however none of the models includes a reliable implementation of focused recharge. Current developments of global gradient-based groundwater models will improve the implementation of these processes further but currently we focus on diffuse GWR in this study.

This is a limitation now stated more clearly (Line 125):
"We do not consider focused recharge in this study as no model offers a reliable implementation of these processes until now."

1.5

The authors refer to groundwater levels throughout the paper with respect to temperature levels; however, this is confusing as groundwater levels are generally considered water table levels. It might be good to include temperature when referring to these.

The submitted manuscript makes no assumptions about the effect of changes in groundwater recharge on groundwater levels. We assume that the referee misread parts of the manuscript as the abbreviation GW (Global Warming) can easily be confused with GroundWater in GWR (Groundwater Recharge). While we realize this is not a perfect choice for an abbreviation we chose to keep it.

1.6

I agree that it is good to focus on absolute changes in recharge rather than relative changes. The authors suggest that underestimating runoff would result in increased GWR; however, this would not be the case if GWR is focused and derived from runoff as in semiarid regions (L. 74).

We are happy to hear that the reviewer agrees with our choice in using absolute changes. This is something we have debated extensively when writing this manuscript. We agree with the comment, however as stated in 1.4 this study mainly focuses on defuse groundwater recharge.

1.7

The authors repeatedly use present tense to refer to work that was done for this study. I think it would be more appropriate to use past tense.

The manuscript was heavily revised in this regard at multiple places and should now contain a more precise use of tenses. Please see the attached markup document.

1.8

The model CLM-5 has been upgraded substantially relative to CLM-4.5. It might be good to consider CLM-5 rather than CLM-4.5.

It is true that CLM-5 is an improved version of the model, however our analysis is based on the available ISIMIP 2.b outputs, where only the modeling team of CLM-4.5 has submitted groundwater recharge. Running additional simulations with another model would compromise the reproducibility of our findings and clear link to rigorous ISIMIP protocol. This study is not an investigation of a specific model but of a consistent model ensemble.

1.9

Soil thickness varies substantially among the models (e.g. LPJ 13 m thick). It would be good to comment on the impact of varying soil thickness on model results.

While we agree that such an analysis is of interest it is clearly out of scope for this paper. It would require to modify and rerun all the complex models in a sensitivity analysis. The corresponding author is not the developer of these models, which are very complex to setup and run and even harder to compare, this is why we are using the ISIMIP framework, which enables a baseline that allows for the complex model comparison shown in this manuscript. To allow for a more in-depth comparison on how different processes are implemented in GHMs there is currently a complex manuscript under development targeting the implementation differences in these models.

Referee #2

2.1

One of the main conclusions is that dynamic vegetation has a strong impact on the estimated GWR, which is far from being as obvious based on the results presented.

While we agree that there is a significant uncertainty on how changes in CO2 levels impact the water balance Figure 7 and 8 clearly shows substantial differences between the model types. Of course, this does not necessarily mean that this difference can only be explained through the simulation of dynamic vegetation. Nevertheless, it seems likely and supports that further research on this topic is necessary.
We added a sentence to make this clearer in our conclusions (Line 579 ff):

*"*However, consideration of these processes does not lead to a uniform increase of groundwater recharge, as might be expected from the physiological effect of increasing atmospheric CO_2 concentration. In some regions with decreasing groundwater recharge, where groundwater availability is a major concern, models that include these processes show the largest differences among themselves. Further research is necessary to understand GWR on large scales, and how it is affected by climate. Simulation of groundwater recharge by global hydrological models needs to be analyzed in more detail, and the benefit of integrating gradient-based groundwater flow models in GHMs should be assessed.*"*

2.2

In addition, some figures are difficult to read and seem to support only partially the comments.

We have improved the font size in multiple figures. See also comment 2.6.

2.3

Introduction: The choice is made to estimate ground water recharge throughout the continental part of the world. However, many areas have no extended groundwater. Whymap.org provides map of the extension of the aquifer. What does it mean to estimate GWR where there is no or local and shallows aquifer? Won't it be more interested to estimate the recharge on the aquifer domains?

Groundwater recharge is also of interest outside of the major global aquifers represented in WHYMAP in blue. In each 0.5° grid cell there are very likely local aquifers (e.g. in alluvial valleys, and confined aquifers are also affected by groundwater recharge. There are other studies which use WHYMAP e.g. Gleeson et al. 2012 Nature, or Taylor et al. 2013 NCC, however our analysis here firstly focuses on a grid-based analysis (allowing for a better understanding on the model differences), which could be in a follow-up study be applied on an aquifer scale.

2.4

Line 200: It is necessary to provide information on the bias-adjusted method? What is the assumption? What is the reference climate used?

A detailed explanation of the method is out of scope of the paper. We have added, to the existing reference to additional information on the climate inputs, a reference of the used method (line 213):

"The bias adjustment method used for the GCMs in ISIMIP2b is using a trend preserving algorithm (Frieler et al., 2017) with EWEMBI (Lange 2018) as baseline (reference) climate condition."

2.5

Section 2.1: I suggest to provide a table that summarizes if GHM includes or not in the GWR another part that a partition of the precipitation, especially, which GHM includes river to aquifer exchange, since this may be a very important difference in alluvial regions. Moreover, this table should also summarize which GHM integrates direct effect of CO2, explaining clearly if they account for stomatal aperture sensitivity to CO2 and/or for vegetation dynamic (LAI) (so far, this is not clear).

As explained in 1.4 we do not consider focused recharge or transmission losses in this study (see 1.4). Also, it might be difficult to distinguish specific alluvial regions on this coarse resolution. A new Table 1 has been added to summarize the implementation of CO2 related vegetation processes more clearly (Line 126 ff).

Table 1 Overview which models are able to simulate the impact of evolving CO_2 concentrations on vegetation and how it is implemented.

GHM	Considers CO_2	Summary of considered vegetation processes in ISIMIP2b	**Reference**
WaterGAP2	No	-	-
CLM4.5	Yes	Photosynthesis depends on root zone soil moisture availability. The description is similar to LPJmL listed below.	(Di Liu and Mishra, 2017)

		The area a population of plant functional types (PFTs) takes up is prescribed and only changes if the input data does.	
H08	No	-	-
JULES-W1	Yes	Evapotranspiration is considered from five PFTs and four non-vegetative surface types. Each grid cell is composed of different fractions of those 9 surface types. Transpiration occurring from vegetation is based on photosynthetic process, which is subject to stomatal conductance regulated by the CO_2 concentration. Furthermore, transpiration is also controlled by the soil moisture availability in the root zone.	(Best et al., 2011; Clark et al., 2011)
LPJmL	Yes	Vegetation composition is determined by the fractional coverage of PFTs at the grid scale. PFTs are defined to account for the variety of structure and function within a stand and are therefore simulated as average individuals competing for light and water according to their crown area, LAI, and rooting profiles. The vegetation dynamics component of LPJmL includes carbon allocation to different PFT tissue compartments, PFT interaction, and establishment and mortality processes. Photosynthesis and stomatal response are simulated following Farquhar et al. (1980) and the generalization by Collatz et al. (1991) for global modelling, based on the function of absorbed photosynthetically active radiation, temperature, day-length, and canopy conductance for each PFT present in a grid cell.	(Schaphoff et al., 2018)
PCR-GLOBWB	No	-	-
CWatM	No	-	-
MATSIRO	Yes	The consideration of CO_2 effects is functionally similar to that in CLM, and there is no dynamic vegetation scheme. CO_2 is prescribed in the model, which is used in the photosynthesis scheme to calculate stomatal conductance, among other parameters, following Farquhar et al. (1980). Soil moisture stress on photosynthesis is considered using moisture availability in the root zone with root distribution fraction in each soil layer. All of that is done for different vegetation or plant functional types.	(Takata et al., 2003)

2.6

Figure 2 is of bad quality: Is it necessary to have all the extremes? Is it reasonable to have a range over 2000mm/year? most aquifers recharge maps stop before 1000mm/year and often under 500. Are GWR values above 1500 in Figure 2 located in capacitive aquifers or in very local and shallow aquifer as defined by Whymap (see comment 1)? As this figure is difficult to read, it is impossible to check comments line 317-322

We think it is necessary to include all extremes even if it reduces the readability of the figure. Only by showing the outliers we are able to openly discuss what the models compute no matter if the values are reasonable or not. Again, the focus on specific aquifers is not something we target in this manuscript (See also 2.3).

2.7

Nice to disentangle the impact of GCM and GHM, but, fig3a includes 76 cases while fig b includes only 36 cases. How does this compare? In order to try to understand what the impact of GHM is, and what the impact of the response of GHM to GW is, it seems required to show the variance of the 8 GHMs on preindustrial case. Figure 7 shows that there are important change, but, variance will be helpful to compare with fig3a.

*As explained in 2.3 not all RCPs and GCM combinations may lead to a warming of 3°, thus the number of involved ensemble members changes. The simulation of PI GWR per GHM is already shown in S1 and S4. We now also refer to these figures in this paragraph: (Line 393) "*For the simulated variance at PI see Fig. S1 and S4.*"*

2.8

Figure 4 includes only one realization of one GHM and one GCM. Why this GCM? Why this GHM? Is this GCM includes dynamic vegetation? Is the dynamic vegetation of the GCM consistent with the one of the GHM? In any case, it would be nice to have some information on LAI changes;

This is stated at the end of the paragraph "Unfortunately, no other GHM-GCM combinations with these alternative CO_2 concentration variants are available in the framework of ISIMIP2b." Yes, GFDL includes dynamic vegetation as well. We agree that an assessment of the differences between the GHM and GCM implementation of the dynamic vegetation would be of interest, however, it is clearly out of scope of in this study. It is also unclear to us how information on LAI changes would provide more insights on the presented results.

2.9

Line 422 : Decrease of precipitation lead to a decrease of vegetation productivity) I guess it is more complicated than that, you may correct

Agreed it very much depends. We rephrased the sentence and added another relevant reference in this matter.

Now reads (Line 458 ff):

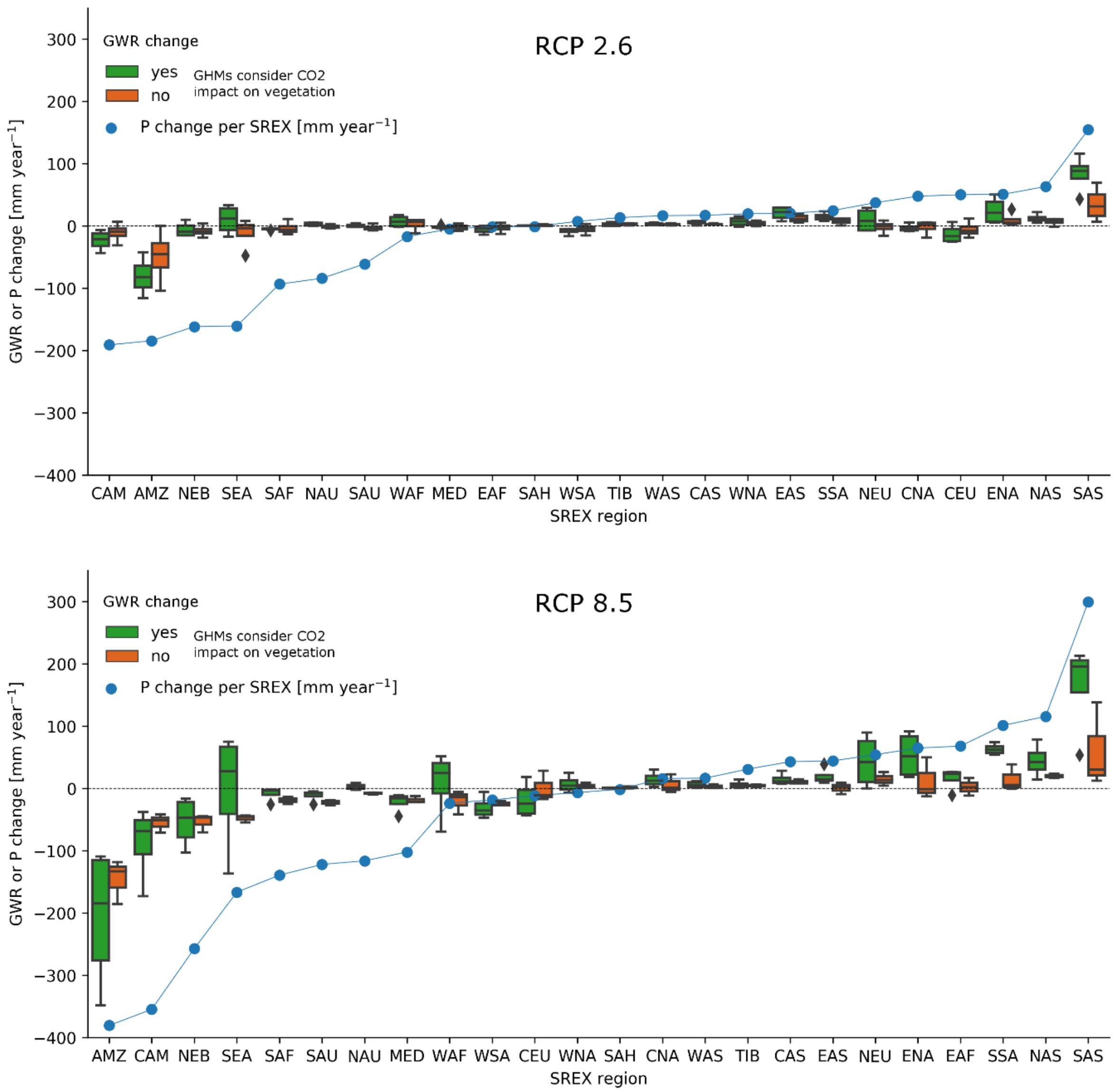
"Decreases in precipitation may lead to a decrease in vegetation productivity (if not counteracted by an increased water-use efficiency due to elevated CO_2 concentrations (Singh et al.; 2020))) and thus to a decrease in transpiration."

2.10

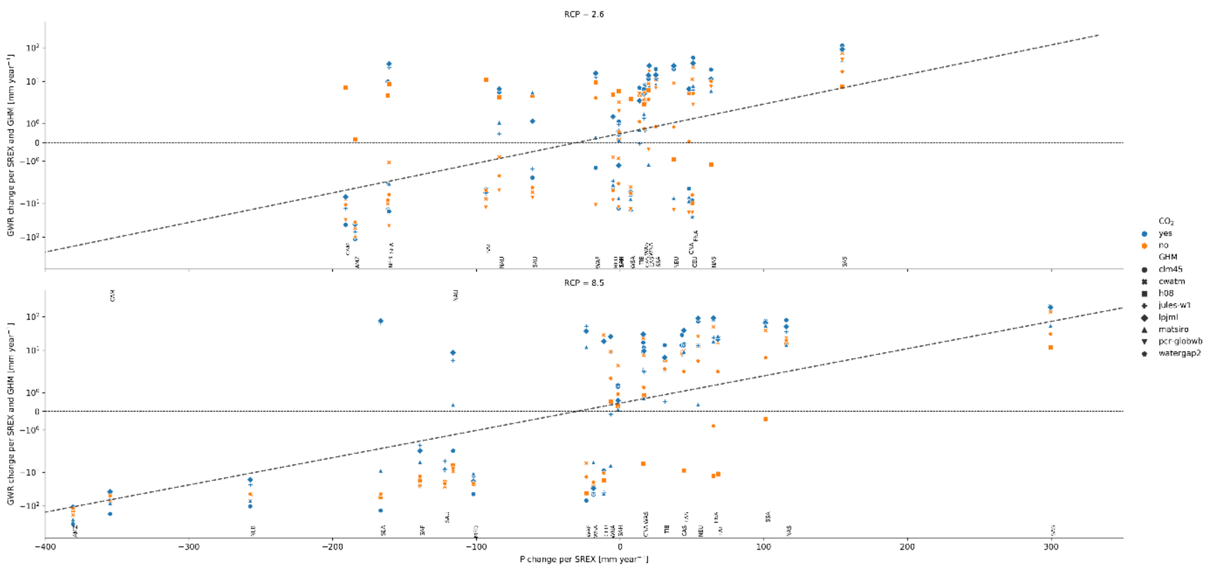
Figure 5: is this figure correct? a same GHM can appear twice for a subregion with the same colors. A change of precipitation of 100mm/year can lead to an increase of GWR of the same amount. Is the vegetation dead? It would be nice to better understand which process occurs in detail in one subregion?

Yes, the figure is correct. However, we realize that it might be difficult to distinguish between the amounts of markers. To allow for a more comprehensive figure we revised in heavily and summarized the models in a bar chart (see new Figure 7 below). We have furthermore improved the original plot and added it to the supplement to allow the interested reader to still investigate the model differences.

Concerning the large increases of GWR linked to large increases of precipitation, it is unclear which process is mainly responsible for this feedback. Possibly the increased water use efficiency of the vegetation allows for a large increase in GWR linear to the increase in P.



(heavily revised) Figure 1 Relation of changes in precipitation (P) (mean(1981-2010) – mean(2070-2099)) to changes in GWR (mean(1981-2010) – mean(2070-2099)) depending on the model type (with or without CO_2; see also Table 1) per SREX (selection as in Table 3) for RCP 2.6 and RCP 8.5 for the GCM HadGEM2-ES.



(new) Figure S10 Relation of changes in precipitation (P) (mean(1981-2010) – mean(2070-2099)) to changes in GWR (mean(1981-2010) – mean(2070-2099)) depending on the model type per SREX (selection as in Table 1) for RCP 2.6 and RCP 8.5 for the GCM HadGEM2-ES. Y-Achis is log-scaled. The dashed line is the 1:1 line.

2.11

Fig 6 compares 4 GHM with dynamic vegetation (change in LAI?) with the 4 others who assumes constant vegetation (but do they account sensitivity of the stomatal aperture to CO2?). The difference on the physics of the GHMs is large, but the impact seems reduced. How can we be sure that the difference is linked to the dynamic vegetation? Similar difference may well exist between this different GHMs without change in vegetation dynamic. Moreover, are the significant changes located where the aquifers are extended and capacitive, or where the aquifers are very shallow and local?

No, the other 4 models with constant vegetation do not account for the stomatal aperture. To clarify we now define the terminology much more clearly in 2.1 to avoid any confusion. Additionally, we added Table 1 to provide an overview on which model implements what kind of vegetation.

Now reads in S2.1:
"In the following, we use the term *active vegetation* for models that consider the physiological effect of changes in CO_2 on vegetation and the term *dynamic vegetation* for the models that allow for a changing vegetation regarding LAI and/or vegetation type."

We agree that we cannot be sure without a further extended sensitivity analysis, which unfortunately is not possible at this point, that the differences we are seeing are solely due to the inclusion of vegetation. However, the only analysis possible in this regard with the available data shown in Figure 6 (based on new version) supports that assumption. We are now stating this more clearly:

Line 493: "It is likely that the shown differences are due to the implementation of dynamic vegetation in the GHMs (compare Fig. 6), however it is possible that other model peculiarities and processes are relevant as well."

For your comment on specific aquifers see 2.3.

2.12

Assessment of the GHM should not appear in the discussion. It should be earlier, or in supplement: GWR estimates by Mohan is only the GWR from precipitation (no river inputs). How does this compare to the numerous GHM?

We don't think that an earlier appearance is merited. The assessment of the models themselves is not the focus of this paper, thus a discussion of the figure at the beginning of the paper is not useful to convey our central messages. Moving them to the supplement is also not helpful as Fig. 9 transports important messages to understand the relevance of the presented results and conclusions correctly. We think that it is a helpful figure to discuss the limitations of the approach without skewing the central message of the paper. The discussion is thus the right place even though it might be unusual.
Concerning the question if the estimates are comparable see also our answer in 1.4.

2.13

Line 539: "Despite the uncertainties, this study shows that climate change will impact groundwater availability in many regions of the world". this is naïve: it was already shown by numerous regional studies.

Additional evidence is always a good thing. We further provide three important additional messages that regional studies are not able to show: 1) We provide global patterns of change – also for regions that have not been studied yet and consistent patterns on much larger scales. 2) We present a multi-model ensemble approach, which is also not common in regional studies. 3) The used models operate at the coarse spatial resolution closer to climate models without the requirement to downscale uncertain climate input to specific regions.

We rephrased the conclusion and it now reads:
"Potential GWR changes due to climate change require increased attention from the scientific community as well as from decision-makers because they affect future water availability in many regions and thus the wellbeing of billions of people. This study shows that simulated global-scale estimates of GWR vary strongly among GHMs, which contribute more strongly to the overall uncertainty of future groundwater recharge than the applied GCM output. However, statistically significant increases and decreases of GWR could be identified in specific regions per GW level. The presented inter-model ranges of GWR changes are an important input for processes aiming at developing strategies for climate change adaptation, as risk-averse decision-makers may want to orient their strategies towards adapting to the worst-case GWR change and not to the projected ensemble mean change.

This study shows that including vegetation processes in GHMs can change projected GWR changes substantially. However, consideration of these processes does not lead to a uniform increase of groundwater recharge, as might be expected from the physiological effect of increasing atmospheric CO_2 concentration. In some regions with decreasing groundwater recharge, where groundwater availability is a major concern, models that include these processes show the largest differences among themselves. Further research is necessary to understand GWR on large scales, and how it is affected by climate. Simulation of groundwater recharge by global hydrological models needs to be analyzed in more detail, and the benefit of integrating gradient-based groundwater flow models in GHMs should be assessed."

2.14

Line 545: “Moreover, this study shows that including dynamic vegetation processes in GHMs can change the results substantially”: this is not that clear from your results. I was expected more impact indeed.

This relates to 2.5, 2.8, 2.9, 2.11 and should be much clearer now.

2.15

A map of the extension of the subdomain is required

We assume that the referee is referring to the SREX regions, which are provided in Fig S6.

2.16

Numbering of figures is wrong

Figure 1 existed twice in the manuscript. We apologize for any caused inconvenience. This was corrected.

2.17

Line 287: Masson-Delmotte et al

This has been corrected.

2.18

Figure S3 should be in the text since it is discussed

We disagree. It is only shortly mentioned and not necessary to understand to message of the paper. It would be possible to move it to the Appendix but we don’t think that it adds any valuable information that is not already transported through the text.

Referee #3

3.1

I do appreciate the tremendous amount of work the authors put into the study and already apologize for not being able to be more positive. The study has a flaw in that it assumes/postulates that GHMs are able to simulate groundwater recharge processes (line 61). This remains to be shown (see specific comment below). The authors also compare to Mohan et al. (2018), a data set, which is also highly uncertainty itself, and show that all models essentially have no skill. Thus, the study is hypothetical and should be seen as a model sensitivity study, which does not necessarily reflect reality. The large uncertainty in the results supports that notion. The authors discuss the limitations in detail and come to the conclusion (line 538) “Despite the uncertainties, this study shows that climate change will impact groundwater availability in many regions of the world. ” Yes, that‘s probably true. But it’s sad that due to the large uncertainty, no additional concrete conclusions can be drawn from the results.

Because the detailed analyses and numbers have very low confidence, I am not able to comment on the simulation results. In summary, I find that GHMs have not been tested comprehensively especially with regard to complex processes such as groundwater recharge. Thus, in my opinion, the study is too early (first testing, then analyses). Looking at the description of the models, I would not even call the estimated flux groundwater recharge. If GHMs are applied I suggest to study the major fluxes of the water balance including their impact on the residuals of the balance equation in the models. In a way that's what the authors are doing, yet the presentation does not show these results.

We appreciate the recognition of the amount of work that went into this study and strongly disagree with the conclusion of reviewer #3 that "the study is too early". We think it is timely to do a study that aims at understanding the best information we have on potential impacts of climate change on groundwater at the global scale, which is provided by the multi-model ensemble output that is analyzed in this study. The merit of our study is similar to that of many climate change studies done with global climate models. It is well known that different global climate models project very different future climatic changes in response to the same greenhouse gas emissions scenario; nevertheless, their results including an understanding of their uncertainties are of interest to many and impact decision-making. We explicitly wanted to focus on uncertainties and not (impossible) predictions, which is also reflected by the title of our manuscripts: ""Uncertainty of simulated groundwater recharge …".

*It is definitely justified to criticize the ability of GHMs to simulate groundwater recharge, which we openly discuss in this publication. The referee himself/herself admits that "*The authors discuss the limitations in detail". *While the analysis has flaws the study discusses them openly and it provides new information and understanding. That no precise and certain prediction of future groundwater recharge are currently possible is certainly, using the word of the referee "sad". However, the study does identify statistically significant increases and decreases in recharge in multiple regions of the world. And it provides a range of potential future recharge changes for each world region that persons in charge of climate change adaptation should take into account for lack of better knowledge.*

*The statement that "*that GHMs have not been tested comprehensively" *contradicts a substantial number of studies that have been published in the recent years devoted to the evaluation of GHMs, for example: (Scanlon et al., 2018; Müller Schmied et al., 2014; Döll and Fiedler, 2008; Döll and Flörke, 2005) and many more. Still, additional testing and improvement is still necessary and this study provides important pointers for future research. In any case, we agree that groundwater recharge is a complex process that needs to be developed further in GHMs. Our manuscript offers first insights on what these improvements might be by presenting a novel comparison of a large ensemble of models in this regard. The suggested approach of comparing other components of the water balance is something that has been done in the framework of ISIMIP in multiple other publications ("*e.g., flood risk (Willner et al., 2018; Thober et al., 2017; Alfieri et al., 2017), low flows in Europe (Marx et al., 2018), evapotranspiration (Wartenburger et al., 2018), runoff and snow in Europe (Donnelly et al., 2017) or multi-sectoral impacts (Byers et al., 2018)."*) that are cited in the introduction of this manuscript.*

In reaction to the reviewer comment, we made substantial changes to both abstract and conclusions to focus less on specific numerical values than on explaining the potential benefits and applicability of the study results.

The new abstract:

"Billions of people rely on groundwater as an accessible source for drinking water and irrigation, especially in times of drought. Its importance will likely increase with a changing climate. It is still unclear, however, how climate change will impact groundwater systems globally and thus the availability of this vital resource. Groundwater recharge is an important indicator for groundwater availability but it is a water flux that is difficult to estimate as uncertainties in the water balance accumulate, leading to possibly large errors in particular in dry regions. This study investigates uncertainties in groundwater recharge projections using a multi-model ensemble of eight global hydrological models (GHMs) that are driven by the bias-adjusted output of four global circulation models (GCMs). Preindustrial and current groundwater recharge values are compared with recharge for different global warming (GW) levels as a result of three representative concentration pathways (RCPs). Results suggest that projected changes strongly vary among the different GHM-GCM combinations, and statistically significant changes are only computed for few regions of the world. Statistically significant GWR increases are projected for Northern Europe and some parts of the Arctic, East Africa and India. Statistically significant decreases are simulated in southern Chile, parts of Brazil, central USA, the Mediterranean, and southeast China. In some regions, reversals of groundwater recharge trends can be observed with global warming. Because most GHMs do not simulate the impact of changing atmospheric CO_2 and climate on vegetation and thus evapotranspiration, we investigate how estimated changes in GWR are affected by the inclusion of these processes. In some regions, inclusion leads to differences in groundwater recharge changes of up to 100 mm $year^{-1}$. Most GHMs with active vegetation simulate less severe decreases of groundwater recharge than GHMs without active vegetation and in some regions even increases instead of decreases. However, in regions where GCMs predict decreases in precipitation and groundwater availability is most important, model agreement among GHMs with active vegetation is lowest. Additional research on simulating groundwater processes in GHMs is necessary."

The new conclusions:

"Potential GWR changes due to climate change require increased attention from the scientific community as well as from decision-makers because they affect future water availability in many regions and thus the wellbeing of billions of people. This study shows that simulated global-scale estimates of GWR vary strongly among GHMs, which contribute more strongly to the overall uncertainty of future groundwater recharge than the applied GCM output. However, statistically significant increases and decreases of GWR could be identified in specific regions per GW level. The presented inter-model ranges of GWR changes are an important input for processes aiming at developing strategies for climate change adaptation, as risk-averse decision-makers may want to orient their strategies towards adapting to the worst-case GWR change and not to the projected ensemble mean change.

This study shows that including vegetation processes in GHMs can change projected GWR changes substantially. However, consideration of these processes does not lead to a uniform increase of groundwater recharge, as might be expected from the physiological effect of increasing atmospheric CO_2 concentration. In some regions with decreasing groundwater recharge, where groundwater availability is a major concern, models that include these processes show the largest differences among themselves. Further research is necessary to understand GWR on large scales, and how it is affected by climate. Simulation of groundwater recharge by global hydrological models needs to be analyzed in more detail, and the benefit of integrating gradient-based groundwater flow models in GHMs should be assessed."

3.2

Only a couple specific comments Abstract: The reported percent increases/decreases of GWR suggest an accuracy that is simply not there; especially given the huge uncertainty in the results. Thus, the abstract sends the wrong message, especially to water managers and decision makers.

We agree that there should be no confusion on the main message of this manuscript and that the results should be considered with care. To attract the reader's attention on this matter and avoid any confusion for readers that might misinterpret the results we have greatly adapted the abstract. See also 3.1..

3.3

54: One of the most important factors is missing: Depth and dynamics of the free water table.

This is now mentioned (Line 63 ff):

*"*An additional factor in estimating groundwater recharge is the simulation of the groundwater table and thus capillary rise and focused recharge. This has not been achieved yet in GHMs, however, recently, global hydrological models (GHMs) started integrating gradient-based groundwater models to better estimate the flows between surface water and groundwater as well as the impact of humans and the changing climate on the groundwater system (de Graaf et al., 2019; Reinecke et al., 2019). Neglecting capillary rise may lead to an overestimation of decreases and increases of GWR due to a changing climate."

3.4

61: This is true, but has never been tested. I suggest to compare the GHMs against fully integrated hydrologic models, such as Cathy, Hydrogeosphere, OpenGeosys, ParFlow, etc. in order to test the ability of GHMs to simulate recharge processes. This is one of many tests that GHMs should undergo in my opinion.

The models currently under development have been compared to these models (e.g. ParFlow in Reinecke et al. 2019). We agree that this comparison needs to continue in the future.

We added a sentence to the conclusions to reflect that. See 3.1.

3.5

Section 2.2: Porbably I missed these details: what's the time step, the spatial resolution, etc?

The spatial resolution is 0.5° x 0.5° (described in section 2.1) and temporal resolution of the original GWR data is monthly, which was averaged to yearly values in this study (also described in 2.1). Time step is a term one might use in the context of numerical models, such as gradient-based groundwater models, which are not part of this study.

References

Alfieri, L., Bisselink, B., Dottori, F., Naumann, G., Roo, A. de, Salamon, P., Wyser, K., and Feyen, L.: Global projections of river flood risk in a warmer world, Earth's Future, 5, 171–182, doi:10.1002/2016EF000485, 2017.

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Ménard, C. B., Edwards, J. M., Hendry, M. A., Porson, A., Gedney, N., Mercado, L. M., Sitch, S., Blyth, E., Boucher, O., Cox, P. M., Grimmond, C. S. B., and Harding, R. J.: The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes, Geosci. Model Dev., 4, 677–699, doi:10.5194/gmd-4-677-2011, 2011.

Byers, E., Gidden, M., Leclère, D., Balkovic, J., Burek, P., Ebi, K., Greve, P., Grey, D., Havlik, P., Hillers, A., Johnson, N., Kahil, T., Krey, V., Langan, S., Nakicenovic, N., Novak, R., Obersteiner, M., Pachauri, S., Palazzo, A., Parkinson, S., Rao, N. D., Rogelj, J., Satoh, Y., Wada, Y., Willaarts, B., and Riahi, K.: Global exposure and vulnerability to multi-sector development and climate change hotspots, Environ. Res. Lett., 13, 55012, doi:10.1088/1748-9326/aabf45, 2018.

Clark, D. B., Mercado, L. M., Sitch, S., Jones, C. D., Gedney, N., Best, M. J., Pryor, M., Rooney, G. G., Essery, R. L. H., Blyth, E., Boucher, O., Harding, R. J., Huntingford, C., and Cox, P. M.: The Joint UK Land Environment Simulator (JULES), model description – Part 2: Carbon fluxes and vegetation dynamics, Geosci. Model Dev., 4, 701–722, doi:10.5194/gmd-4-701-2011, 2011.

Collatz, G.J., Ball, J.T., Grivet, C., and Berry, J. A.: Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration: a model that includes a laminar boundary layer, Agricultural and Forest Meteorology, 54(2-4), 107-136, doi:10.1016/0168-1923(91)90002-8, 1991.

de Graaf, I. E.M., Gleeson, T., van Rens Beek, L. P. H., Sutanudjaja, E. H., and Bierkens, M. F. P.: Environmental flow limits to global groundwater pumping, Nature, 574, 90–94, doi:10.1038/s41586-019-1594-4, 2019.

Di Liu and Mishra, A. K.: Performance of AMSR_E soil moisture data assimilation in CLM4.5 model for monitoring hydrologic fluxes at global scale, Journal of Hydrology, 547, 67–79, doi:10.1016/j.jhydrol.2017.01.036, 2017.

Döll, P. and Fiedler, K.: Global-scale modeling of groundwater recharge, Hydrol. Earth Syst. Sci, 863–885, 2008.

Döll, P. and Flörke, M.: Global-Scale Estimation of Diffuse Groundwater Recharge: Model Tuning to Local Data for Semi-Arid and Arid Regions and Assessment of Climate Change Impact, Frankfurt Hydrology Paper - 03, 2005.

Döll, P., Trautmann, T., Gerten, D., Schmied, H. M., Ostberg, S., Saaed, F., and Schleussner, C.-F.: Risks for the global freshwater system at 1.5 °C and 2 °C global warming, Environ. Res. Lett., 13, 44038, doi:10.1088/1748-9326/aab792, 2018.

Donnelly, C., Greuell, W., Andersson, J., Gerten, D., Pisacane, G., Roudier, P., and Ludwig, F.: Impacts of climate change on European hydrology at 1.5, 2 and 3 degrees mean global warming above preindustrial level, Climatic Change, 143, 13–26, doi:10.1007/s10584-017-1971-7, 2017.

Farquhar, G. D., Caemmerer, S. von, and Berry, J. A.: A biochemical model of photosynthetic CO2 assimilation in leaves of C 3 species, Planta, 149, 78–90, doi:10.1007/BF00386231, 1980.

Marx, A., Kumar, R., Thober, S., Rakovec, O., Wanders, N., Zink, M., Wood, E. F., Pan, M., Sheffield, J., and Samaniego, L.: Climate change alters low flows in Europe under global warming of 1.5, 2, and 3 °C, Hydrol. Earth Syst. Sci., 22, 1017–1032, doi:10.5194/hess-22-1017-2018, 2018.

Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F. T., Flörke, M., and Döll, P.: Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration, Hydrol. Earth Syst. Sci., 18, 3511–3538, doi:10.5194/hess-18-3511-2014, 2014.

Reinecke, R., Foglia, L., Mehl, S., Trautmann, T., Cáceres, D., and Döll, P.: Challenges in developing a global gradient-based groundwater model (G^3M v1.0) for the integration into a global hydrological model, Geosci. Model Dev., 12, 2401–2418, doi:10.5194/gmd-12-2401-2019, 2019.

Scanlon, B. R., Zhang, Z., Save, H., Sun, A. Y., Müller Schmied, H., van Beek, L. P. H., Wiese, D. N., Wada, Y., Di Long, Reedy, R. C., Longuevergne, L., Döll, P., and Bierkens, M. F. P.: Global models underestimate large decadal declining and rising water storage trends relative to GRACE satellite data, Proceedings of the National Academy of Sciences of the United States of America, 115, E1080-E1089, doi:10.1073/pnas.1704665115, 2018.

Schaphoff, S., Bloh, W. von, Rammig, A., Thonicke, K., Biemans, H., Forkel, M., Gerten, D., Heinke, J., Jägermeyr, J., Knauer, J., Langerwisch, F., Lucht, W., Müller, C., Rolinski, S., and Waha, K.: LPJmL4 – a dynamic global vegetation model with managed land – Part 1: Model description, Geosci. Model Dev., 11, 1343–1375, doi:10.5194/gmd-11-1343-2018, 2018.

Takata, K., Emori, S., and Watanabe, T.: Development of the minimal advanced treatments of surface interaction and runoff, Global and Planetary Change, 38, 209–222, doi:10.1016/S0921-8181(03)00030-4, 2003.

Thober, S., Kumar, R., Wanders, N., Marx, A., Pan, M., Rakovec, O., Samaniego, L., Sheffield, J., Wood, E. F., and Zink, M.: Multi-model ensemble projections of European river floods and high flows at 1.5, 2, and 3 degrees global warming, Environ. Res. Lett., 13, 14003, doi:10.1088/1748-9326/aa9e35, 2017.

Willner, S. N., Levermann, A., Zhao, F., and Frieler, K.: Adaptation required to preserve future high-end river flood risk at present levels, Science advances, 4, eaao1914, doi:10.1126/sciadv.aao1914, 2018.