Response to comments from reviewer #3

We thank the reviewer for this positive response to our paper and thoughtful comments that significantly improved the paper. We have modified the manuscript accordingly, and detailed corrections are listed below. In the following the reviewer's comments are given in italics and our reply in normal font.

Reviewer 3

Review of "A high resolution global scale groundwater model" By I. De Graaf et al.

Scientific Significance:
Does the manuscript represent a substantial contribution to scientific progress within the scope of Hydrology and Earth System Sciences (substantial new concepts, ideas, methods, or data)?

The paper addresses a very interesting topic which is definitely suitable for the journal. I believe it is important to have a better consideration of groundwater at the global scale: this will help understanding and refining the results of the global hydrological model and it is definitely justified by the increase amount of data (also geological data) that are becoming available at the global scale. We are all aware of the limitations of the global set of data, but this should not prevent us from using these data for global scale models just clearly showing all the potential drawbacks.
The paper is definitely within the scope of HESS.

Scientific Quality:
Are the scientific approach and applied methods valid? Are the results discussed in an appropriate and balanced way (consideration of related work, including appropriate references)?

The applied methods that are applied are valid, but the general description of the approach is a little confused and mainly the discussion is not complete. For example, regarding the description of the methods, I found confused the explanation of the aquifer properties presented in chapter 2.2: I would suggest some rewording there .Regarding the discussion, I think it needs to be more focused on the very important and critical point . Just as an example, many times it is noted that there are problems because of the perched water tables in the mountains but it is not clear what is the actual impact of these observations (there is a figure that should show that, Figure 6, but it is impossible to understand that). Instead of concentrating on that I would find very interesting a discussion regarding the general overestimation of the results. What the authors think as the more reasonable explanation for that? I find it more interesting potentially with a higher impact on the future work than the problem with the perched GW table.

We are thankful for the rigorous assessment of our paper and glad that the scientific quality of our work has been recognized despite the issues with the presentation and the formulation that the reviewer identifies. We have taken his/her concerns at heart and improved the description of our methods and discussion of our results (see also comments from Zachary Subin and Mary Hill).

Presentation Quality:
Are the scientific results and conclusions presented in a clear, concise, and well-structured way (number and quality of figures/tables, appropriate use of English language)?

I will provide here a more detailed list of suggestions:
p. 5220 l. 10: the sentence is not clear

Following the reviewer's suggestion, we have rewritten this to read: "Their method however does not include hydrogeological information such as aquifer depths and tranmissivities, but uses estimates from soil data. Also, the hydraulic connection between rivers and groundwater, which is the primary drainage for groundwater in humid regions, is ignored."

p. 5223 l. 25: transmissivity should be in m^2/d so I do not understand why in the parenthesis k is
correctly multiply by the thickness D but the units are m/d.

Correct, this is a typo

p. 5224 l. 20: I do not understand the difference between surface and floodplain elevation. Aren't they both obtained with Hydrosheds? And what happens when the aquifer thickness is zero?

We used the 30" data to determine the floodplain elevation at 6'as follows: Within each 6'cell and using the Hydrosheds dataset, we identified the lowest elevation at 30'' (maximum 144 values for a cell comprising only land area), and assigned this as the floodplain elevation for the entire cell.
We will clarify this into more detail in the manuscript in section 2.2. (accordingly to other reviewers Mary Hill and Zachary Subin) The aquifer thickness cannot become 0, as we assume a log-normal distribution. However, it can be very small

p. 5227 l. 10-15: this should be better explained

By naming the cells 'true' and 'apparent' we wanted to clarify the difference between the lat-lon cell area used by PCR-GLOBWB and the cell area used by MODLFOW, as the latter assumes rectangular grid cells. This means there is a difference in area, for which we should correct (as done in Eq 7). We will explain this into more detail in the manuscript in section 2.3.

p. 5229 l. 3-5: am I right that this implies that smaller rivers will not lose water?

Only rivers with a width larger than 10 m are expressed in the river-package of MODFLOW which allows groundwater to drain to or draw from the surface water, given the gradient and riverbed conductivity. Rivers with a width smaller than 10 m are included via the drain package , which will only allow groundwater to drain to the surface once intersected by a drain placed at a particular depth (in this case surface level).

p. 5229 l. 10-20: it is not clear which is the use of this in the MODFLOW model.

We understand the raised confusion, and will explain this concept clearer and in more detail to ensure any confusion is removed (see also comment of Mary Hill).

In PCR-GLOBWB the groundwater reservoir is described as a linear reservoir. This we replaced with one layer of MODFLOW, which makes simulated of lateral groundwater flow possible. In the MODFLOW model groundwater-surface water interactions are considered via river drainage along the main stream as represented in PCR-GLOBWB and recharge. This is the main component of the baseflow, especially for sediment areas where groundwater flow is relatively slow. However, at 6‘ resolution, the main stream is insufficient to represent how groundwater levels intersect the terrain and additional drainage is needed to represent local sags, springs and streams higher up in valleys in mountainous areas. To resolve this issue, groundwater above the drainage level (taken equivalent to the river plain elevation), can be tapped by local springs also presented by a linear reservoir (but not the same one as used in PCR-GLOBWB)

p. 5230 l. 3-10: I am not sure that changing together conductivity and recharge can provide useful
results: usually they are strongly correlated

In the uncertainty analysis the groundwater recharge and conductivity are indeed independent parameters here, while in reality they may indeed be positively correlated. By doing this, we may somewhat overestimate the uncertainty (allowing for combinations of larger groundwater recharge and low conductivity and vice versa). However, we draw conductivities within permeability classes that depend on the lithological map used both in the groundwater model as in the land surface model that computes the recharge. So for low conductivity lithologies (and also low transmissivities) the groundwater recharge will be small and for high conductivity lithologies in wet climates large. So this positive correlation is partly accounted for at the scales transcending the lithological units.

p. 5230 paragraph 2.5: how have transient data been used

The dataset used only contains steady-state data.

p. 5230 l. 26: put the figure number where these results are presented
p. 5231 l. 14: figure number is missing
We refer to the Figures were results are discussed (p. 5232 l. 17 and p. 5234 l. 5 respectively)

p. 5232 l. 8-9: not clear

We rewrote this section to make it clearer: “A change in aquifer thickness leads to a change in calculated transmissivities. However, the spatial variability of aquifer thickness is fixed across all realizations (F’(x), eq 1), and the impact on calculated groundwater depths is therefore small”.

p. 5232 paragraph 3.2: did I understand correctly that the observations have been used as they are?

The observed data was used as they are, but only data was used where surface level was reported as well (added on p. 5230 l. 19). The average of the reported data was used when more than one measurement was available in one 6’ grid cell (discussed on p. 5230 l. 19-20)

p. 5233 l. 13: I do not understand why the blue dots are still in the figure and also regarding the same
figure (figure 6) the difference between the two versions is not clear and I would suggest to present only one of them

The blue dots present the groundwater depths for mountain regions. The difference between the two plots is the parameter set used. For A) the best performing run is selected. For B) the best performing run based on validation of groundwater heads in sediment areas is selected.

However, we decided to change this figure after the review of reviewer 2 to only the scatter of A. The main message of this figure should be to 1) show the correlation of the best performing run, 2) show the difference between sediment areas and mountain ranges.

p. 5233-5234: the last paragraph of 3.2 is very confusing

We understand the confusion and will rewrite the paragraph. We also decided to rearrange this figure (accordingly to the suggestions of Mary Hill) to make the presented results easier to understand.

p. 5235: the conclusion needs to be restructured: the problem with the perched groundwater tables is
repeated again, but I think that a discussion on the potential reason for the residuals being always
negative should be included

We will discuss this issue in more detail in the discussion of the residuals (section 3.2).

Figure 5: it is really hard to distinguish and I suggest a better explanation in the caption

We will rearrange this figure (according to the suggestions of Mary Hill).

The parameter that effects groundwater heads most is the saturated conductivity. The other two parameters, aquifer thickness and groundwater recharge, are of minor importance (as explained on p.5232 l. 7-14). We will focus in this figure on the dominant parameter.
We will rewrite the figure caption to: "A) coefficient of variation in groundwater depth of 100 runs with different parameter settings for aquifer thickness, groundwater recharge, and saturated conductivity. B) coefficient of variation of 100 runs with different parameter settings for saturated conductivity."

Figure 7: in this figure, what is presented in the three maps is hardly readable and I suggest to keep just the histograms

We will rearrange this figure to make the main message clearer. With the maps of spatial distribution of residuals we want to show were the absolute differences are. The corresponding relative residuals show that for deep groundwater the absolute error can be large but relative is small, and vise-versa for shallow depths. The histogram shows the distribution of the absolute errors linked to the observed groundwater depths.