Response to comments from Mary Hill

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

Reviewer 2 Mary Hill

1. 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 study is good to excellent. The global simulation of groundwater flows is very interesting. Unfortunately, this article is Fair to Poor. The article is too confusing and important contributions are not presented in a very informative way. The article lacks a depth of consideration needed to translate model development and results to useful take away lessons. This is addressed further under Presentation Quality using some specific examples. However, only examples are presented. The same kind of analysis presented for those examples needs to be considered for every aspect of the paper.

2. 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)?

I see no real problems with the scientific method to the extent I understand it. The problem is that it is poorly explained.

3. 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)?

The article is very confusing as presently written. There are problems with both the description of methods and the presentation of results. These are solvable problems within the present length of the article.

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 her concerns at heart and improved the internal logic and reformulation of our manuscript and included a clear description of the general implication of our findings at the end of the Conclusion. Overall, we tried to improve the explanation of the methods used (see also the comments by Reviewer Zachary Subin and Nir Krakauer) and presentation of results.

Two examples of difficulties with the description of methods:

a. There are three comments about the linear store in the PCR-GLOBWB model that are confusing. I think probably the method used is ok, but the description is too confusing to tell. The three comments are as follows.
i. Page 5222, around line 20, says that the linear reservoir of the PCR-GLOBWB model

is replaced by MODFLOW.
ii. P. 5332 around line 21 says that "Because of the offline coupling and the lack of topographical detail in a 60 cell, the linear groundwater store is maintained, specifically for calculating baseflows above the drainage level to the surface water network using a cell-specific recession constant which accounts for aquifer properties and drainage density."
iii. P. 5228 describes how the MODFLOW head is used to calculate streamflow gains and losses. The role of the linear reservoir in calaculating baseflow is not described.

We understand the raised confusion, and will explain this concept clearer and in more detail to ensure any confusion is removed.

In the original version of PCR-GLOBWB the groundwater reservoir, underlying the two soil layers, is described as a linear reservoir. We replaced this with a one-layered MODFLOW model, 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. Drainage via groundwater-surface water interactions as simulated in MODFLOW forms the main contribution to the baseflow. However, at 6' resolution, the main stream is insufficient to represent truthfully where 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 floodplain elevation, explained in section 2.2), can be tapped by local springs which is also represented by a linear reservoir (but not the same one as used in the original PCR-GLOBWB model).

b. The method behind characterizing ranges and sediment basins is really confusing. I broke the confusion down into the following pieces.
i. A small thing is that the term ranges is vague – perhaps use mountain ranges or high elevation, steep terrain?

We agree with this suggestion and changed it to read mountain ranges accordingly.

ii. On p. 5224, line 9, is it correct that it could be "Next, all 6' cells with floodplain: : :"? If this is correct, then note that the text is clear on how the floodplain level of each 6' cell is determined, but is not clear on how the surface level of each 6' cell is determined.

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. (see also raised question on this point by Zachary Zubin).

iii. The z-score of eq. 2 either needs motivation and clearer description, or to be removed from the paper, described more clearly in another report or paper, and referenced from here. One example of confusion is that equation is presented in the context of the surface level and, while the figure being referred to (fig 3a) is described as "Cumulative probability of aquifer thickness". Associated with this, figure 2 lower panel is not referred to in the paper.

We concur with the reviewer that the underlying motivation for our choice may have been somewhat opaque. We have clarified this in the manuscript:

By definition basins are linked to sedimentary environments in fluvial systems and deltas. Sediments are deposited perpendicular to the main gradient (constituting the transversal axis of the basin), with grain size and volumes decreasing at greater distance away from the transversal axis.
Grain size also decreases along the transversal axis, distinguishing proximal (near the source of sediment) and distal parts. We assume gradation in grain size are somehow captured in the GLiM but differentiation in depth is not. However, since we can locate the transversal axis along the main stream, we can use relative elevation as a measure of proximity to the transversal axis and as an indicator of the associated depth. We standardize the relative elevation to obtain the standard normal ordinate Z(x) and use this to define the distribution of aquifer depths using a log-normal distribution, assuming depth is non-negative and positively skewed.

I think part of the confusion comes from it not being clearly stated that the PCRGLOBWB model is transient while the MODFLOW model is steady-state with forcings equivalent to long-term averages from the PCR-GLOBWB model.

We will clarify this more in section 2.2 where we describe the land-surface model PCR-GLOBWB. Long term averages for recharge and river discharges (outcomes of PCR-GLOBWB run for 1960-2010 (Wada et al. 2014)) are used to force the steady state groundwater model with.

The presentation of results is difficult because the global scale figures are hard to evaluate and there are too many of them. Some ideas are as follows.
Following the suggestion of the reviewer, we have rigorously evaluated the figures in our manuscript and reduced and focused the presented material and improved the overall visual presentation.

However, in this study we want to give global pictures and overviews of the different parameters and simulated groundwater heads, so that larger scale patterns are shown and can be understood. To focus the discussion at some points on specific processes we will use insets, as the reviewer suggests.

We will shortly discuss what we will change in the figures, following the reviewer's comments.

a. Figure 3. put the top panel in auxiliary material or another pub with a better description of how it is obtained. This allows the thickness and T maps to be larger and easier to read.

As answered above (answer to iii), we will clarify the use of the "frequency distribution of aquifer thickness" better to resolve the confusion. We agree with this suggestion to put this figure to the auxiliary material.

b. Figure 4. Choose a location and show detailed inset for each figure. This might require another figure. Choose the location to that a them can be followed and a transferable lesson told.

For the top figure “groundwater recharge”, we want to show the spatial differences over the globe. At the regional scale groundwater recharge is one of the main controls for groundwater levels, we want to show the differences at the global scale.

For the bottom figure “rivers and drains” we agree with the reviewer that insets for specific regions will be useful. We will use Europe and Africa, as this are also the regions of Fig. 9. The main message of this figure is to show that larger rivers (>10 m width) are modelled in MODFLOW with the ‘River Package’ and the smaller rivers (<10 m width) are modelled in MODFLOW with the ‘Drain Package’. All remaining groundwater-surface water interactions (springs and streams higher up in the valley on mountainous areas) is modelled with the linear reservoir concept (all now better explained in section 2.3)

c. Figure 5. Just show panel B and in the caption say that this is the dominant parameter. Say the other two parameters had very little sensitivities and the figure with all parameters varied looks a lot like the one shown. In the new figure, include an inset for the same area in the new figure.

We agree with the reviewer at this point. The improved figure will show: “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.”

d. Figure 6. Make the print on the axes much bigger. The caption says A and B are pretty similar, so it is not clear that the distinction shown is very important. Omit one panel? I think the red points are in the sedimentary basins and the blue are in the ranges, but this is not in the caption. I may have missed item but I do not recall that the obs consistently being larger than the simulated values is discussed. Any thoughts on why this is so?

The spatial difference between observed and simulated groundwater depths (Fig. 7) show that in general groundwater depths are underestimated compared with observations for steeper and higher elevated terrains. This is explained in p. 5233 line 27 by the grid resolution, which is to coarse to capture small local valleys. As a result, local drainage is underestimated for these terrains, and groundwater levels are lower. In fact, the simulated result shows the regional scale groundwater depth pattern, where the local groundwater is (most often) sampled. We assume that for steeper and higher elevated terrains local and perched water tables are likely to occur. The shallow groundwater depths observed for these areas confirm this assumption (see also reaction on comment Nir Krakauer). We will discuss this in more detail in our manuscript.

The difference between the two scatter plots is the parameter set that is used (explained at p. 5233 l. 10-15). But we agree with the reviewer that this figure can be reduced to one scatter only. This will be the left panel, as this result of average steady state run that is also presented in Figure 8.

e. Figure 7 is very hard to read. Decide what message is most important and focus the figure accordingly.

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.

f. Figure 8 is not presented at a scale that allows the dotted areas to be distinguished, as far as I can tell. I am not sure what "lighter/washed-out colors" refer to – the lighter shades?
Perhaps take the same area for which detailed insets are provided and do the same for this figure?

In this figure we want to show the global scale patterns or groundwater depths (as discussed in section 3.3). We improved the figure captain: "…" washed-out colours indicate deep groundwater regions, where most likely shallow perched and local water tables occur, which are not captured by the model. The hyper arid regions area distinguished at the grid resolution.

g. Figure 09. This is the inset idea, but I think at too large a scale to make the figure very meaningful. Use the same location used for the other figures?

In these figures we want to show long inter-basin flow paths as well shorter ones that exist. We chose Europe and Africa as there is a clear difference in groundwater recharge en discharge for these regions.

Again, I commend the authors for their fine study of an important topic.

We thank the reviewer for her kind words and invaluable comments.