Faculty of Geosciences
Department of Physical Geography
PO Box 80.115, 3508 TC Utrecht, The Netherlands

Visiting address
Vening Meinesz A, Princetonlaan 8a
3584 CB, Utrecht, The Netherlands

Website
www.uu.nl/geo/fg
Email jarno.verkaik@deltares.nl
Phone +31(0)6 4691 4636

To: Editors of Geoscientific Model Development

Date: April 18, 2023
Subject: Final Response

Dear Editor, Dear Wolfgang Kurtz,

We would like to thank you for your time for reading our manuscript "GLOBGM v1.0: a parallel implementation of a 30 arcsec PCR-GLOBWB-MODFLOW global-scale groundwater model". We would also like to thank the two anonymous reviewers for their time for thoroughly reading our manuscript and coming up with valuable comments as posted during the open discussion. We were pleased to read their positive feedbacks on our paper's main goal, which is to tackle the technical challenges for parallelizing our global groundwater model, as we state in line 67 of our manuscript: "*In this paper, the focus in on the technical challenges of implementing the GLOBGM using HPC*". We believe that from this perspective, the findings presented in our paper would appeal your readers, supported by the reviewers' comments "*The paper is well written and interesting from a computational perspective. I do not see any problems with the technical steps.*" (Reviewer 1), "*The manuscript is well-organized and successfully demonstrates a computational workflow... For this reason, it is an important contribution.*" (Reviewer 2). Besides reaching readers that are specifically interested in global groundwater modeling, we also believe that our presented methodology is interesting for readers that have large regional-scale or local-scale MODFLOW 6 groundwater models that need significantly faster computations, since our methodology is in essence not restricted to global modeling exclusively.

Although we are pleased to read these positive feedbacks, we also recognize that Reviewer 1 is critical about the usefulness and application range of the presented global model and of global models in general. Although we believe that we have been explicit about the main goal of this paper (parallel implementation of a global groundwater model needed for increased resolution) and also that many model improvements are still needed for our initial version 1.0 of our GLOBGM (e.g., in line 36 "*However, results for the transient simulation are quite similar and there is much room for improvement.*"), we welcome these comments and the valuable suggestions made by both reviewers and will take them into account for a future version. To accommodate both reviewers in their comments about the *global model performance* in this paper, we suggest updating the manuscript as follows (see in more detail our replies to the reviewers):

1. *Introduction:* Based on literature, we will add a paragraph to the beginning of Section 1 about which developments are needed to improve global groundwater models beyond the current state-of-the-art. This will also identify as one these improvements increasing the resolution and increasing the speed of

computation to make this possible. This will then immediately make clear which of these possible improvements are the focus of this paper.

2. *Methods:* In Section 2.2.1, "Description of the GLOBGM", we will add a paragraph about the assumptions and limitations of our GLOBGM model with respect the work for the 5′ global groundwater model, De Graaf et al. (2015, 2017). In this, we will again be more explicit that our work addresses *the computational challenges need to be overcome if the resolution of this model is increased.*

3. *Steady-state model evaluation:*
   3.1. We will add extra curves for our GLOBGM to Figure 12b showing the errors for sedimentary basins without karst, sedimentary basins and karst and non-sedimentary basins.
   3.2. We will mask out the karst areas in the steady-state global maps in Figure 14a.
   3.3. We will change the legend of in Figure 14a to allow for groundwater levels and heads above the surface.
   3.4. We will add a separate figure to Figure 14 showing the heads in the confined aquifers, and mask out the Karst areas.

4. *Transient model evaluation:*
   4.1. We will mask out the karst areas for the global maps of Figure 14b and Figure 14c.
   4.2. We will change the legend of Figure 14c for showing both positive and negative trends.
   4.3. We will add a supplemental document explaining the filtering of transient well locations in more detail.
   4.4. We will investigate the possibilities for comparing with GRACE data for this version of our GLOBGM, version 1.0, or an upcoming version.

Based on the comments of Reviewer 2, we suggest making the following additional changes in the manuscript regarding the technical implementation of the GLOBGM parallelization and typographical/syntactical errors (see in more detail our reply to Reviewer 2):

5. We will add text to Section 2.1 for better explaining the purpose of sorting for the global domain decomposition.
6. Furthermore, in this Section, we will add text to improve the connection of the example of Figure 2 to the actual global domain decomposition being used.
7. In Section 2.3.1, we will add text to emphasize better the relationship between the "Model Workflow" and the "Node Selection Workflow".
8. We will include the typographical/syntactical suggestions and corrections.

We hope that our suggested changes in our manuscript are adequate for you to consider our manuscript for publication. Please let me know if you have any questions or more clarification if required.

On behalf of all authors,

With kind regards,

Jarno Verkaik

Reply to comment of Reviewer 1 om GMD-2022-226

General reply

We thank the reviewer for taking the time to evaluate our manuscript and the extensive comments. The reviewer states that the paper is well-written and does not see any problems with the technical steps of the of the paper. It is important to note this first, because the explicit goal of this paper is (line 66): *"the focus is on the technical challenges of implementing the GLOBGM using HPC"*. To our opinion, the merits of this paper should be evaluated based on that focus.

However, reviewer 1's objections against this paper are predominantly entirely based on the fact that the reviewer does not believe in the usefulness of global (we presume groundwater) models, given their current state of development and that building these and reporting on their progress should not be done. This is where we greatly disagree with the reviewer. There are two main reasons to disagree about this:

1. In any field of earth system science, modelling efforts started with grossly inaccurate first versions of models. We give three examples. the first global circulation models and numerical weather prediction models were very simple in their dynamics, model physics and parameterization (see e.g. Manabe, 1969). The first global vegetation models were simple and not at all predictive at the local scale (Prentice et al., 1992). The first global hydrological models (Alcamo et al., 1997) used very simple conceptual hydrology at low resolution, not at all capable of predicting streamflow at a specific catchment. Here is the question: If global modelling of a certain part of the earth system is deemed useful, should we wait until we are far enough to simulate all states of this system credibly everywhere, or should we start with what we know and develop from there? We firmly believe that in order to make progress, it is better to do the latter as otherwise chances are that there will be no progress at all. All progress must be done in small steps, just as climbing to the next floor needs a stair with 20 steps instead of two. In fact, if imperfect global models cannot be built and developed upon, as suggested by reviewer 1, many of the papers in GMD would not have been published.

2. Even if global models are far from perfect in quite some aspects or regions, they can still be useful. The first global climate models, especially when multiple of these were used, were all good enough to provide order of magnitude estimates of the regional impacts of CO_2 increase. The first numerical weather prediction models extended the forecast horizon several days beyond simply nowcasting pressure fields by hand. The first vegetation models were very helpful in translating local time-series of pollen-based observations to spatiotemporal maps of vegetation shifts from the last glacial maximum into the Holocene. The first hydrological models served very well in pinpointing regions of current or emerging water scarcity.

This is not different when talking about global groundwater models. There is an obvious need by the earth system science community for building these, as testified in Condon et al. (2021). And this has led to the first so-called gradient-based groundwater representations for global hydrology (Fan et al., 2013; de Graaf et al., 2015; Reinecke et al., 2019). These mentioned authors developed the models and extensively tested their accuracy with head observations, showing their inaccuracy, particularly when estimating groundwater depth. And as inaccurate as they can be, they have already proven their worth in understanding the order of magnitude of, e.g., the importance of shallow groundwater in sustaining dry season evaporation (Miguez-Macho et al., 2012), the control of groundwater on rooting depths (Fan et al., 2017) and the impact of groundwater pumping on groundwater depletion (De Graaf et al., 2017) and the violation of environmental flow limits (De Graaf et al., 2019).

When it comes to the steps needed to make a leap change in improving the current generation of models, many steps are still needed; see also the reviews by Condon et al. (2021) and Gleeson et al. (2021). Specifically these are: 1) improved hydrogeological schematization, particularly including multilayer semi-confined aquifer systems and the macroscale hydraulic properties of Karst and Fractured systems; 2) increased resolution to better resolve topography and in particular resolve smaller higher altitude groundwater bodies in mountain valleys; 3) improved knowledge on location, depth and rate over time of groundwater extractions; 4) better estimated groundwater recharge, especially in drylands and at mountain margins; 5) increased computational capabilities to be able to make simulations with the above improvements possible. Our paper specifically revolves around items 2 and 5: *If we improve spatial resolution, how should we make this computationally possible?* It is a small but important step to better global groundwater that needs to be taken to proceed further. We recognize that the above-mentioned argumentation has not been explicitly mentioned in the manuscript so far; for that we thank the reviewer 1 for putting emphasize on his concern. Therefore, in the manuscript, we will provide a statement to this angle in the introduction of the paper to show how this paper fits in the many steps needed.

The paper presented here uses the global hydrogeological schematization of De Graaf et al. (2017) and inspects *what computational challenges need to be overcome if the resolution of this model is increased*. Thus, since the objections against this paper by reviewer 1 are really about the global model itself, not about the numerical techniques displayed here, the reviewer objections are in fact about the De Graaf et al. (2017) paper. In that De Graaf et al. paper, it has been explicitly stated that 1) they simulate the top aquifer systems, so unconfined aquifers and the uppermost confined aquifers; 2) although global results are shown, the model is mostly capable for the sedimentary alluvial basins (main productive aquifers). Since we take the De Graaf et al. (2017) schematization as input, these limitations also count for the model reported here. We will make this explicit clear in the revisited manuscript.

Where we do agree with reviewer 1 is that we should be open about the assumptions and limitations of the model. We presumed, since the focus was on the numerical scheme, that we did not need to extensively discuss limitations already discussed in the De Graaf et al. (2015, 2017) papers. However, as we agree with reviewer 1 that this is important, we will add a paragraph to section 2.2.1 explaining this more in detail. Particularly we will state that: The application domain of the model is as follows: 1) it is intended to simulate hydraulic heads in the top aquifer systems, so unconfined aquifers and the uppermost confined aquifers; 2) wherever there are multiple stacked aquifer systems, these are simplified in the model to one confining layer and one aquifer; 3) the model schematization is suitable for heads in large sedimentary alluvial basins (main productive aquifers) that have been mapped at a 5-arcminute resolution; 4) in as far these sedimentary basins include karst, it is questionable of a Darcy approach can be used to simulate large-scale head distributions; 5) due to the limited resolution of the hydrogeological schematizations, in mountain areas we simulate the heads in the mountain blocks but not those of groundwater bodies in hillslopes and smaller alluvial mountain valleys; 6) also, for the heads in the mountain blocks, we assume that secondary permeability of fractured hard rock can also simulated with Darcy groundwater flow; an assumption that may be questioned.

Replies to specific remarks

In the following we cite specific remarks by the reviewer in italics and provide our remarks in roman.

For example, the steady state comparison with Fan, CMG GGM in Figure 12: The comparison shows that all these results are inconsistent. The authors mention an improvement to GGM, but all of these products are highly uncertain themselves.

We do not agree that all results are inconsistent. It seems that we have not been clear enough about the model comparison, so we will improve that description in the revised paper version. We do compare between models, but we compare the differences between each model and the mean water levels in 34k well locations across the US. The models that perform best, i.e. the Fan et al. (2013) model and the Zell and Sanford (2020) USGS model, have been calibrated, while the 5-arcminute model of De Graaf et al (2019) and our 30-arcsecond model have not. This is a fair comparison of model capabilities. It also shows that especially the Zell and Sanford (2020) USGS model is the only one performing well in the mountain hard rock regions because of a) 250 m resolution that resolves many more groundwater pockets in mountain valleys; b) local calibration of transmissivities.

But the section 426-430 is very hard to read. It is a very long list how the authors worked around data gaps, with no explanation why they do so, and there is no assessment to what extent these chosen steps are reliable. To be scientifically sound, every step has to be explained and demonstrated how, where and when the assumptions hold up. It is crucial and a basic scientific principle to root any model in reality.

We regret that this has not been clear. We will add a supplemental where we will present more in detail how the selection of the locations with wells with groundwater level time series was done from the original NWIS dataset. We will also add the number of wells that remained after each filtering step. This was not a way to work around data gaps, but to retain only locations with sufficiently long time series and to ascertain that the filter of each groundwater observation well was assigned to the right model layer. We like to stress that even though actual model building is not the target of this paper - this is the introduction of an efficient scheme that makes high-resolution global groundwater modelling a reality- we have added a thorough evaluation of simulated depth to groundwater. Even though this was not global, given the huge variation of landscapes and hydrogeological settings in the U.S., we feel that this can be representative for also other world regions.

For continents where no transient head data are available a comparison with changes of water table obtained through GRACE could be informative in the context of Figure 14. Such a comparison will clearly show that the model cannot reproduce the decline of the water table on a global scale.

Thanks for this suggestion. GRACE would be an interesting means of comparing the results of the global groundwater model with. However, GRACE provides total terrestrial water storage anomalies, which includes many terms besides groundwater. So, to apply GRACE we need to correct for non-groundwater-related storage changes. We might do that with the results of the global hydrological model which is used to force our groundwater model: PCR-GLOBWB. But even then, there will be both positive as well as negative trends based on a combination of groundwater withdrawal and climate variability. We will investigate the possibilities for comparing with GRACE data for this version of the GLOBGM version 1.0 or an upcoming version.

The authors acknowledge many areas of improvement but, to highlight this point again, fall short in clearly declaring that none of the results should be used in any other context than in a software development framework.

As we argued above, we feel that just because global models are in earlier stages of development does not mean that they cannot be used for certain analyses. So, we will not make this statement. As to illustrate a possible application of this global groundwater modelling: when comparing the model runs with and without groundwater withdrawals over the last 60 years, the *large-scale* impacts of groundwater use on the global heads, groundwater storage and streamflow can be assessed, while taking into account groundwater-surface water interaction, increased capture and later groundwater flow (as cannot be done with water-balance based methods). This application of a global groundwater was shown before in De Graaf et al. (2017) and De Graaf et al (2019).

The only reliable data source is topography, but through the rough discretization a lot of information is lost, especially in steep terrains. The global geological products are speculative at best, a significant source of uncertainty. There are countless conceptual problems. Groundwater abstraction or rivers cannot be reliably simulated with these spatial resolutions in the MODFLOW conceptualization. The wells and rivers are cell-centered, making a robust simulation of drawdown cones or mounds extremely uncertain. The associated temporal dynamics of the water table decline cannot be captured with these resolutions. As the authors are presenting a transient model this is a fundamental problem.

Yes, by using a 1 km discretization lot of information is lost that would be needed to make *local inferences.* But this is not the case for *sub-regional to regional assessments* at the global extent as is the primary goal of a global extent model. We disagree that groundwater abstractions and the impacts of rivers cannot be reliably simulated. This can be done if answers are needed at the subregional to regional scale and provided that hydraulic resistances between surface waters and the groundwater systems are scale-consistent. This has been done for over four decades with regional groundwater models all over the world. These regional models had lower resolutions than ours until well into the 1990s. Does this mean, that all these models were useless? Evidently, they were not as they were used extensively as a basis for water management and policy. Of course, that is not our goal, since we do not have subregional scale hydrogeological data. But we could if these were available. For instance, on a regional scale, we want to mention the Mekong Delta model (Minderhoud et al, 2017), which reproduces well enough land subsidence caused by groundwater extractions using a grid cell size of 1x1 km^2; the same resolution as our GLOBGM version 1.0. This model has been well-received by the local community and has led to changes in groundwater extraction policies under the so-called Decree 167, which prescribes restrictions on groundwater extraction in aquifers within the territory of the Socialist Republic of Vietnam. Regarding transient simulations: these can be done if the goal is to study water table or hydraulic head variations at scales larger than the model resolution. Again, this has been done for decades with regional-scale models at similar resolutions for decades.

Also, a large part of the word is karstic, which is not reflected in the global geological model and conceptualization of MODFLOW. These areas should be fully blanked out, global maps of karst are available. Note that fractured systems are also treated the same as porous aquifers, there is no mention of the conceptual incompatibility of this approach. Moreover, all areas with permafrost or snow cannot be simulated either with the current conceptualization and should be blanked out as well.

We will mask out the karst areas that are part of the alluvial sedimentary basins in the global maps. We will also indicate in the maps the areas that do not belong to the alluvial sedimentary basins, which include fractured systems. We do not agree with masking out the permafrost, as these areas are implicitly accounted for by the global hydrological model used to force the groundwater model (with recharge, groundwater withdrawal rates and surface water levels). In this model, permafrost is included by

impermeable soils that generate no groundwater recharge. Hence, we simulate no active groundwater flow systems below permafrost areas, only the presence of groundwater.

The list could be endlessly expanded, and this assessment should have been done before the submission of the paper.

We would like to stress again that we use a previously published global groundwater and explicitly focus on the computational challenge to increase its spatial resolution with a factor 100. This is one of the many steps needed to arrive at significantly better global groundwater models (see above). For a discussion on these steps and what could be done to achieve them we refer again to Condon et al. (2021) and Gleeson et al. (2021).

The model only seems to report a groundwater decline.

This is actually not the case. We truncated the legend to focus on decline only. But of course interannual to interdecadal climate variability could be an important cause of both decline as well as increase of groundwater levels. Therefore, in an updated version of Figure 14c, we will show both negative and positive trends.

In Figure 14a there are no hydraulic heads above the surface. But this is the case for many confined aquifers, take the great artesian basin in Australia or the Nubian systems for example.

Thanks for pointing this out. Again, this is not the case and is again caused by our truncating of the legends. Also, for regions with confined aquifers, we have portrayed the groundwater levels in the confining layer, not the underlying aquifer. These groundwater levels are kept at bay by drains positioned at surface level to emulate groundwater exfiltration for groundwater in undated areas. We will the legend to allow for groundwater levels and heads above the surface. Also, we will add a separate figure showing the heads in the confined aquifers. This shows several areas with heads above the surface. Of course, there are still areas where one would expect larger heads. This can be explained by the lack of a parameterization of mountain front recharge in the model, which is an important source of recharge of confined aquifers such as the Great Artesian basin, Australia.

The large depths to groundwater in all mountainous regions further show that the model results have nothing to do with reality. You can easily see this by consulting the measured hydraulic heads as I suggested above, or by taking a healthy hike in the mountains and appreciating countless small rivers and streams emerging at high altitudes. Again, this list could be endlessly expanded.

The mountain areas are also part of the evaluation shown in Figure 12. Here, errors are indeed large if groundwater levels are measured in mountain valleys that are too small to be resolved by the original 5 arcminute hydrogeology of De Graaf et al. (2017). Therefore, the model is only capable of simulating the mountain block hydraulic heads in the mountain areas and not the groundwater pockets in mountain valleys higher up. This limitation was already mentioned in De Graaf et al. (2015) but we agree we will repeat this here in the Introduction again. Also, we expect this to greatly improve if one re-parameterizes the hydrogeological model of De Graaf et al. (2017) at 30-arc-second resolution. Note this is work in progress but outside the scope of this paper.

Little is said about the water balances of the catchments.

The water balance is better evaluated at the aquifer level, since groundwater flows across the catchment boundaries. Water balances are by definition closed (apart from a minor numeric error) in our MODFLOW simulations since they are part of the closure criterion of the numerical solution.

Required conditions for publication

Reviewer 1 states a number of conditions that need to be met before the reviewer can support publication. We state these conditions in italics below and provide a response to what extent these conditions can be reasonably met given the scope of this paper.

- *Greatly expand the discussion on the fundamental conceptual issues and demonstrate to what extent they undermine the robustness of the model.*
- We will extent section 2.2.1 by stating the conceptual choices made in the original 5 arcminute version of the model which also pertain to this version. We will also state the limitations of this setup, i.e. the six limitations stated above. We are not sure what the reviewer means with effect on the robustness of the model, but it may be to show how well the model does in areas where it is not meant to work that well. In addition to the steady-state evaluation, we will add extra curves for the GLOBGM to Figure 12b showing the errors for sedimentary basins without karst, sedimentary basins and karst and non-sedimentary basins. *Blank out areas where karst and permafrost snow is present.*

We will provide a mask portraying the areas not belonging to the sedimentary basins and the karst areas in the sedimentary basins.

- *Demonstrate which fractures systems can be simulated with a Darcy type approach (some can but not all, depending on the properties). Blank out the ones that cannot*

With all due respect, we think that this request is not reasonable in the context of this paper. Demonstrating which can and cannot be simulated with Darcy-type flow, which also is dependent on the scale of analysis, will never be possible, even at the local scale. But since the non-sedimentary (crystalline) rocks will be masked out in a newer manuscript version, this is also not needed.

- *Refine the spatial resolution of the model that groundwater abstraction and rivers are at least conceptually implemented correction. With a finite difference approach as implemented here this is essentially impossible with the current computational capacities, so the grid should be fully redone using the unstructured features of MODFLOW which now unfortunately are only used to deal with issues in islands. Pumped aquifers or river have to be simulated with a spatial resolution adequate to the physics of the problem.*

We respectfully disagree that re-gridding is necessary. It is a misconception that one needs to refine the grid to a certain point to correctly capture the physics of the problem. The physics are resolved: Darcy's law and continuity of mass are obeyed at the scale of the numerical grid. Of course, to maintain scale-consistency, the right scaling laws of hydrogeological parameters are required and we have tried to do so as much as possible. Regular grids have been used many decades in regional-scale groundwater modelling, accepting that features smaller than the grid size are not resolved. Sub-grid parameterization or scaling laws then make sure that fluxes remain more or less constant with scale. For instance, sub-grid groundwater-surface water interaction of non-resolved streams can be taken care of by putting in additional drains (see e.g., De Graaf et al., 2017). Non-structured grids to resolve fine features of interest is something we are definitely thinking about implementing in the near future.

- *Demonstrate that the geological models and the soil maps for the remaining areas are an adequate simplification of the complexity of the subsurface. Demonstrate that populating the geological/ soil models with physical parameters is robust.*

With all due respect, we think that this is not a reasonable request. First, what is "adequate" here? Second, even if adequate could be defined, it would greatly depend on the purpose for which the model is used and at what scale (e.g. global sensitivity versus local prediction). Third, this would then mean finding alternative more detailed hydrogeological models, e.g. as used in regional groundwater models, inserting these into our model and look at the differences. This is indeed an interesting exercise, but an enormous task reserved for some future research out of the scope of this paper.

- *Test the model with real data, including streamflow, and demonstrate that the workarounds in section 2.4.2 are robust.*

The section on model evaluation 3.3 we do compare the model with real mean head data and time series over 900k locations in the US. Comparison with streamflow data does not make sense, since the groundwater model only produces baseflow. However, we could compare the streamflow data of the land surface with discharge observations (theoretically land surface baseflow and groundwater baseflow are on average the same). However, this has been done extensively in Sutanudjaja et al. (2018). We refer to that paper for these specific validation statistics.

- *The outcome of such an assessment will show that the model is far away from being used in any real-world context. The paper needs a clear statement in the abstract and the conclusions on this.*

As already explained under point 2 in the beginning of this reply: if models are inaccurate in certain areas, they can still be useful for (sub-)regional sensitivity studies of global extent as some of the references given show. We therefore feel that this model is far away from being used *in any context* is too strong. However, we will provide a caveat in the text related to the limitations of global groundwater models, when we describe the model in section 2.2.1.

References

Alcamo, J., P. Döll, F. Kaspar, and S. Siebert (1997), Global change and global scenarios of water use and availability: An application of WaterGAP 1.0., Rep. A9701, Cent. for Environ. Syst. Res., Univ. of Kassel, Kassel, Germany.

Condon, L. E., Kollet, S., Bierkens, M. F. P., Fogg, G. E., Maxwell, R. M., Hill, M. C., et al. (2021). Global groundwater modeling and monitoring: Opportunities and challenges. Water Resources Research, 57, e2020WR029500.

De Graaf, I. E. M., Sutanudjaja, E. H., van Beek, L. P. H., & Bierkens, M. F. P. (2015). A high-resolution global-scale groundwater model. Hy-drology and Earth System Sciences, 19(2), 823–837.

De Graaf, I. E. M., van Beek, R. L. P. H., Gleeson, T., Moosdorf, N., Schmitz, O., Sutanudjaja, E. H., & Bierkens, M. F. P. (2017). A global-scale two-layer transient groundwater model: Development and application to groundwater depletion. Advances in Water Resources, 102, 53–67.

De Graaf, I. E. M., Gleeson, T., van Beek, L. P. H., Sutanudjaja, E. H., & Bierkens, M. F. P. (2019). Environmental flow limits to global ground-water pumping. Nature, 574(7776), 90–94.

Fan, Y., Li, H. & Miguez-Macho, G. Global patterns of groundwater table depth. Science 339, 940–943 (2013).

Fan, Y., Miguez-Macho, G., Jobbágy, E. G., Jackson, R. B., & Otero-Casal, C. (2017). Hydrologic regulation of plant rooting depth. Proceedings of the National Academy of Sciences of the United States of America, 114(40), 10572– 10577.

Gleeson, T., Wagener, T., Döll, P., Zipper, S. C., West, C., Wada, Y., Taylor, R., Scanlon, B., Rosolem, R., Rahman, S., Oshinlaja, N., Maxwell, R., Lo, M.-H., Kim, H., Hill, M., Hartmann, A., Fogg, G., Famiglietti, J.S., Ducharne, A., de Graaf, I., Cuthbert, M., Condon, L., Bresciani, E., and Bierkens, M.F.P. (2018), GMD perspective: The quest to improve the evaluation of groundwater representation in continental- to global-scale models, Geosci. Model Dev., 14, 7545–7571.

Manabe, S. (1969), Climate and the ocean circulation: The atmospheric circulation and the hydrology of the earth's surface, Mon. Weather Rev., 97, 739–774.

Miguez-Macho, G., and Y. Fan (2012), The role of groundwater in the Amazon water cycle: 2. Influence on seasonalsoil moisture and evapotranspiration,J. Geophys. Res.,117, D15114.

Minderhoud, P.S.J., Erkens, G., Pham, V.H., Bui, V.T., Erban, L.E., Kooi, H., Stouthamer, E., 2017. Impacts of 25 years of groundwater extraction on subsidence in the Mekong delta, Vietnam. Environ. Res. Lett. 12, 13.

Prentice, I. C., W. Cramer, S. P. Harrison, R. Leemans, R. A. Monserud, and A. M. Solomon (1992), A global biome model based on plant physiology and dominance, soil properties and climate, J. Biogeogr., 19, 117–134.

Reinecke, R., Foglia, L., Mehl, S., Trautmann, T., Cáceres, D., & Döll, P. (2019). Challenges in developing a global gradient-based groundwater model (G3M v1.0) for the integration into a global hydrological model. Geoscientific Model Development, 12(6), 2401–2418.

Sutanudjaja, E.H., van Beek, R., Wanders, N., Wada, Y., Bosmans, J.H.C., Drost, N., van der Ent, R J., de Graaf, I.E.M., Hoch, J.M., de Jong, K., Karssenberg, D., López López, P., Peßenteiner, S., Schmitz, O., Straatsma, M.W., Vannametee, E., Wisser, D. and Bierkens, M. F. P. (2018), PCR-GLOBWB 2: a 5 arcmin global hydrological and water resources model, Geosci. Model Dev., 11, 2429–2453.

Reply to comment of Reviewer 2 om GMD-2022-226

General reply

We thank the reviewer for taking the time to evaluate our manuscript and the extensive comments. Below we present a point-by-point reply to the points raised by Reviewer 2. The Comments of the Reviewer are represented in Italics, our replies in Roman.

GENERAL COMMENTS:

The manuscript is well-organized and successfully demonstrates a computational workflow towards the efficient simulation of long term monthly GW flow and discharge at ~ 1 km scales at a global extent. For this reason it is an important contribution.

We thank Reviewer 2 for this positive evaluation of the contribution.

My general concern with the manuscript is that the (i) benefit of increasing the model resolution and (ii) the resulting utility (or persisting limitations) of a 30″ GW simulation are assumed rather than clearly motivated. While the introduction (55ff) names the need for better GW estimates and describes the opportunity afforded to GW modelers by datasets generated at increasing resolution, the manuscript would be strengthened by persuading the reader that (a) 'better GW estimates' need to be made by a global model; (b) a 30″ model makes better estimates than a 5′ model for some set of GW questions; and/or (c) the development of a 30″ model is a provisional but critical step towards the type of better, global GW estimates that are at some point in the future. Some discussion of these factors would also better situate the discussion of model performance in Section 3. While it is true that the paper is about the technical dimensions of model construction and execution, and a full interrogation of model performance may belong in an altogether separate paper, models must be suitable for some purpose(s), and it is not discussed whether this model is or is not, or is on the way to being so.

These are valid points by the reviewer and these have also been indirectly raised by Reviewer 1.

When it comes to the critical steps that are needed to obtain better global GW estimates, or better global groundwater models in general, we will now add a section to the Introduction that states (based on other literature) which developments are needed to improve global groundwater models beyond the current state-of-the-art. This will be a statement that reads something like (see also the reply to Reviewer 1):

"When it comes to the steps needed to make a leap change in improving the current generation of models, many steps are still needed; see also the reviews by Condon et al. (2021) and Gleeson et al. (2021). Specifically these are: 1) improved hydrogeological schematization, particularly including multilayer semi-confined aquifer systems and the macroscale hydraulic properties of Karst and Fractured systems; 2) increased resolution to better resolve topography and in particular resolve smaller higher altitude groundwater bodies in mountain valleys; 3) improved knowledge on location, depth and rate over time of groundwater extractions; 4) better estimated groundwater recharge, especially in drylands and at mountain margins; 5) increased computational capabilities to be able to make simulations with the above improvements possible. Our paper specifically revolves around items 2 and 5: *If we improve spatial resolution, how should we make this computationally possible?* It is a small but important step to better global groundwater that needs to be taken to proceed further."

Regarding the (persisting) limitations that even a 30 arcsecond model still suffers, we intend to add a paragraph to the beginning of section 2.2.1 highlighting the limitations the global groundwater model (as it derives from de model of De Graaf et al. 2015; 2017) has in simulating global groundwater levels and heads (see also the reply to Reviewer 1):

Particularly we will state that: "The application domain of the model is as follows: 1) it is intended to simulate hydraulic heads in the top aquifer systems, so unconfined aquifers and the uppermost confined aquifers; 2) wherever there are multiple stacked aquifer systems, these are simplified in the model to one confining layer and one aquifer; 3) the model schematization is suitable for heads in large sedimentary alluvial basins (main productive aquifers) that have been mapped at a 5-arcminute resolution; 4) in as far

these sedimentary basins include karst, it is questionable of a Darcy approach can be used to simulate large-scale head distributions; 5) due to the limited resolution of the hydrogeological schematizations, in mountain areas we simulate the heads in the mountain blocks but not those of groundwater bodies in hillslopes and smaller alluvial mountain valleys; 6) also, for the heads in the mountain blocks, we assume that secondary permeability of fractured hard rock can also simulated with Darcy groundwater flow; an assumption that may be questioned."

Section 3 should briefly help the reader understand that. For example, the steady state Fan (2017) model, which is at the same 30" resolution as GLOBGM, performs better with respect to DTW observations; hypotheses about why this is so (conceptual model? recharge formulation? subsurface parameterization using newly available datasets?) would help guide subsequent GLOBGM development and analysis, and would help the reader understand the significance of the model comparisons included in this manuscript.

We will add a description to Section 3.3.1 about the presumed reasons for the improvement performance of the 30 arcsecond model compared to the 5-arcminute model. We think that the main reason is that the higher resolutions are better in following topography and relief, in particular to resolve smaller higher altitude groundwater bodies in mountain valleys. Also, higher resolution models have a smaller scale gap with the in-situ head observations in wells. The reason that the Fan et al. (2013) model does better is that their model has been calibrated, while ours has not. They do this by optimizing a two-parameter relationship between the e-folding depth value that reduces conductivity with depth and the topographic slope. So, they effectively calibrate the transmissivity of their model. For the US, many of the groundwater level observations they use for this calibration are also part of the validation dataset. We will include this explanation in Section 3.3.1 as well.

SPECIFIC COMMENTS:

121ff (and Figure 2). I do not understand how 'sorting by cell count' constrains or informs the decomposition of the global domain into 9050 separate 'grids'.

The sorting of the grids using the cell count is important to determine the largest models subject to parallelization. After sorting from large to small, the first three largest directly correspond to the grids for the Afro-Eurasia, Americas and Australia models, respectively. We will add some text explaining the purpose of sorting.

Figure 2 should note that, while helpful for illustrative purposes, this submodel distribution does not actually occur in the global domain decomposition because in GLOBGM all of the islands belong to a single model.

Although Figure 2 is indeed for illustrative purposes, the submodel distribution illustrated in this figure actually does occur in the global domain decomposition: the largest island in model 1 on the left corresponds to the Afro-Eurasia, Americas or Australia model; the three islands in model 2 on the right correspond to the Islands model. Although the Islands model contains 9047 islands, the larger "islands" may be split up to run on multiple cores. This is actually the case for the largest "island" Great Britain in the model when using 32 cores as in Table 4. We will add text to the paragraph describing Figure 2, line 189, to improve the connection with the global domain decomposition.

Figure 3: is the Node Selection Workflow before the Model Workflow? I'm not clear on the logical or temporal relationship between the 2 workflows.

As shown in Figure 3b, the Node Selection Workflow makes usage of the Model Workflow, see line 293 "Then, the Model Workflow generates the model input files for this number of cores". We will emphasize this relation, directly at the beginning of the Node Selection workflow description, line 288, as well as in the caption of Figure 3b.

Section 3.1.2 - How does tiling reduce storage requirements?

This is described in Section 2.3.2, line 306, "using tiles cancels a significant number of redundant sea cells (missing values)". Although in section 3.1.2 we refer to this in line 469 "Using data tiles therefore saves storing ~8 TB of redundant data (43% reduction; see Section 2.3.2)", we will clarify this more in the text of 2.3.2. Basically, for the global extent, we use 163 square tiles instead of 288 tiles (24 x 12; including all oceans, Greenland and Antarctica that are not used as input for our model), and therefore for each global map we only need to store 100 x 163/288 = 57% of the data corresponding to 43% reduction. Hence, instead of storing one 30-arcsec global raster file having 43200x21600 cells, we store 163 smaller raster files, each having 1800x1800 cells.

TYPOGRAPHIC/SYNTACTICAL SUGGESTIONS AND CORRECTIONS

Thank you for the close reading. We will include these suggestions and corrections in our revised manuscript.

References

Condon, L. E., Kollet, S., Bierkens, M. F. P., Fogg, G. E., Maxwell, R. M., Hill, M. C., et al. (2021). Global groundwater modeling and monitoring: Opportunities and challenges. Water Resources Research, 57, e2020WR029500.

De Graaf, I. E. M., Sutanudjaja, E. H., van Beek, L. P. H., & Bierkens, M. F. P. (2015). A high-resolution global-scale groundwater model. Hy-drology and Earth System Sciences, 19(2), 823–837.

De Graaf, I. E. M., van Beek, R. L. P. H., Gleeson, T., Moosdorf, N., Schmitz, O., Sutanudjaja, E. H., & Bierkens, M. F. P. (2017). A global-scale two-layer transient groundwater model: Development and application to groundwater depletion. Advances in Water Resources, 102, 53–67.

Fan, Y., Li, H. & Miguez-Macho, G. Global patterns of groundwater table depth. Science 339, 940–943 (2013).

Gleeson, T., Wagener, T., Döll, P., Zipper, S. C., West, C., Wada, Y., Taylor, R., Scanlon, B., Rosolem, R., Rahman, S., Oshinlaja, N., Maxwell, R., Lo, M.-H., Kim, H., Hill, M., Hartmann, A., Fogg, G., Famiglietti, J.S., Ducharne, A., de Graaf, I., Cuthbert, M., Condon, L., Bresciani, E., and Bierkens, M.F.P. (2018), GMD perspective: The quest to improve the evaluation of groundwater representation in continental- to global-scale models, Geosci. Model Dev., 14, 7545–7571.