Utrecht University

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: September 21, 2023
Subject: second revised manuscript

Dear Editor, Dear Dr. Wolfgang Kurtz,

First, we thank the reviewers for their time and efforts to further increase the quality of our manuscript. We have taken account of all the reviewers' comments. Regarding Reviewer #2, we implemented all the (minor) changes that he / she suggested. Regarding Reviewer #3 and #4, please find below the comments of the reviewers in *italics* and our replies in roman.

Reviewer #3

GENERAL COMMENTS

I have reviewed the manuscript "GLOBGM v1.0: a parallel implementation of a 30 arcsec PCRGLOBWB-MODFLOW global-scale groundwater model" by Verkaik et al. They use a global-scale GGC (30") physics-based groundwater flow model with a parallel implementation in order to reduce model run times.

The main objective of the work is clearly stated. For evident numerical reasons, such global-scale model must have to take the most from parallel computing to reach "acceptable" simulation runtimes.

The text is mostly readable by the wider groundwater community (which I think is the intended readership for this paper), although I would suggest adding some more detail and/or vulgarising some analytical "jargon" such as the so-called "Pfafsetter level". This makes the methods section a bit arched for groundwater scientists who are not specialised in computational techniques. However, this is not a critical point for general understanding.

We thank Reviewer #3 for this brief overview of our paper. We agree to make our manuscript readable a wide groundwater community and thus have performed another round of edits in which we tried to add some additional explanations for some technical terms, such as Pfasfstetter level (e.g., line 237):

HydroBASINS catchments follow the Pfafstetter base-10 coding system for hydrologically coding river basins, where the main stem is defined as the path which drains the greatest area, and at each refinement level ten areas are defined: four major tributaries, five inter-basin regions and one closed drainage system (Verdin and Verdin, 1999).

I do not have technical comments that would tell me to reject this paper. My only major comment is rather on the over-consideration in the precision of such models to simulate arbitrary predictions at regional to local scale (for surface water capture for instance). What are the general benefits to our community of having these global scale models? To put it more clearly, what is the main objective of developing such a model?
Although it is evident for atmospheric sciences that global scale physics-based models have to deal with processes such as la Niña or el Niño; encompassing physical terrestrial boundaries; it is not so clear to me what would be the benefits of groundwater modelling at global scale. Why do we need to overcome the physical boundaries that define independent hydrogeological systems? To what extent can such a global scale model perform better than a regional scale model specifically developed for a local hydrogeological system? Finally, what is the main purpose (i.e. prediction) of such models? The paper would benefit from including such clarifications in the introduction.

This point has been made many times before and can be made for all global impact models that are out there, such as global hydrological models (used for global water resources assessments), global dynamic vegetation models and agricultural models (used for global terrestrial carbon and biodiversity and food security assessments), global economic models (to assess global economic impacts) and global inundation models (used to assess global flood risk). All of these global assessments could also be made with a patchwork of regional-scale models. Still those models exist for global scale assessments, sometimes standalone, sometimes as part of integrated assessment models or earth system models. The reason is that a complete coverage with regional-scale models is generally not available, while a global model has the benefit of uniformity of model setup such that it is more straightforward to compare different regions in the world. This is especially useful in data-scarce areas where often no models are available. With this global approach, one has at least a first order model, which incorporates hydrogeological concepts of areas where one has more information. Apart from this, there are reasons to have a global groundwater model in global scale analysis that require a global coverage. A first example is terrestrial vegetation growth and evaporation that are modulated by the presence of shallow groundwater tables with possibly global-scale impacts on the global water and carbon cycle under climate change (Anyah et al., 2008; Miguez-Macho and Fan, 2012). Another example is the importance of non-renewable groundwater use to global food production and food trade (Dalin et al., 2017) with possible global impacts on future food security. A third example is contribution of terrestrial water storage change on regional sea-level trends (Karabil et al., 2021). In all of these cases the complete global groundwater system has to be considered.

To make this case a bit stronger we have added the following lines to the Introduction (line 51):

Recent publications have called for a better representation of groundwater in earth system models (Bierkens et al., 2015; Clark et al., 2015; Gleeson et al., 2021). Apart from providing a globally consistent and physically plausible representation of groundwater flow using a uniform model set-up, global-scale groundwater models could serve to support global change assessments that depend on a global representation of groundwater resources. Examples of such assessments are the impact of climate change on vegetation, evaporation and atmospheric feedbacks (Anyah et al., 2008; Miguez-Macho and Fan, 2012), the role of groundwater depletion in securing global food security and trade (Dalin et al., 2017) and the contribution of terrestrial water storage change to regional sea-level trends (Karabil et al., 2021).

Despite the obvious limitations that I have discussed previously, why do you not consider other type of observations to "validate" the model, such as stream flow data and/or satellite based data?

These are good suggestions which we have partly followed in the second version of our paper (which we hope Reviewer #3 has been evaluating), see e.g. Section 2.4.3 and Section 3.3.3. We have contemplated using streamflow data, but we are only presenting the groundwater component of the terrestrial hydrological cycle, so we only simulate groundwater discharge with the groundwater model, which cannot be directly compared with total streamflow observations. We have compared total water storage anomalies based on our head simulations and other storage components from the driving hydrological model PCR-GLOBWB 2 (Sutanudjaja et al., 2018) with observations from GRACE and GRACE-FO (Figure 15, line 723) with good results.

OTHER COMMENTS

Judging from the line numbers mentioned, we are afraid that the Reviewer #3 has evaluated the original manuscript and not the revised second version that we submitted after a first round of revisions. So, to answer the questions below we have also included the original first version of the manuscript in our answers.

In lines 43 to 45 you justify the need to assess groundwater depletion, but why do we need global scale models for this? Would regional scale models be more appropriate instead?

Here, we refer to our answer above. Regional models could be used for that in principle, but there would be no global coverage and no spatially consistent and uniform way of determining groundwater depletion.

Line 48: So this is a two layer model. This is very coarse for the vertical direction. Why not consider a 2D approach? At this scale I do not see the benefit of including the 3D at this coarse resolution. Perhaps more explanation is needed here (without having to search for the information in the many papers you refer to).

To be fair, we do not mention three-dimensional or 3D anywhere in the paper. We have in fact a quasi-3D model where we simulate 2D horizontal flow in an unconfined aquifer or a confined aquifer topped with a confining layer, with vertical groundwater exchange between the confined aquifer and the overlying confining layer. However, the model could be extended to a fully 3D modelling approach if wanted and when more information about the vertical geological layering is available at the global scale. Note this is work for the future.

Line 75, more detail is needed here. The reference to Gleeson et al, 2021 is not sufficient.

We have changed this sentence to (line 96):

We provide a limited evaluation of the computed results, and we note that the current model is a first version that should be further improved in the future. We refer to Gleeson et al. (2021) for an extensive discussion on pathways to further evaluate and improve global groundwater models.

Line 99: Why is the upper model layer a confined layer? I would rather conceptualize the upper layer as an “unconfined” layer and the bottom layer as a “confined” layer. Not clear.

Sorry for being unclear here. But we talk about a confining layer, not a confined layer. A confining layer is a less permeable layer on top of a confined aquifer. Flow is mostly vertical in a confining layer. Where the confining layer is present, we have a confined aquifer below and where it is not present the second layer is an unconfined aquifer.

To make this clearer we changed the text as (line 124):

The 5′ PCR-GLOBWB-MODFLOW global-scale groundwater model (GGM) consists of two model layers: where a confining layer (having a lower permeability) is present, the upper model layer represents the confining layer and the lower layer a confined aquifer. If a confining layer is not present, both the upper and lower model layers are part of the same unconfined aquifer (de Graaf et al., 2015, 2017).

Line 103-104: Why is the water not allowed to leave the domain at the upper layer? How do the model deal with seepage faces/nodes? Is water can leave the model domain from other boundary conditions than rivers and lakes? The conceptualization of model boundary conditions is not very clear.

We acknowledge that this is indeed not clear. We have added (line 189):

Note that in the GLOBGM interaction with surface water or surface drainage is modelled by putting rivers and drains in the first active layer, seen from top to bottom.

Line 421: The average amplitude error is not so straightforward. Why not just consider the residuals, which is more often used in groundwater modelling applications?

Note we implicitly do consider the residuals, but this is done by comparing: 1) the average error (bias) by evaluating the steady state model results with observed average groundwater depths; 2) looking at the timing of peaks and throughs by calculating the cross-correlation between observations and simulations. The amplitude error is additionally calculated to see if the variation is represented correctly. This has been done in previous work (De Graaf et al., 2017) and makes it comparable to that work. The three components, viz. bias, timing error and amplitude error are also the three aspects of e.g., the Kling-Gupta Efficiency (KGE) which is also often used to evaluate the fit to time series. We additionally added a comparison of the trend. We believe that this still provides a comprehensive way to evaluate the performance of a groundwater model.

Line 505: typo: "het"
Line 568: typo: "be left"

Thanks for noticing. We have corrected this.

Section 3.3: A CONUS-extent (US) is considered to validate the global scale model. Model validation is therefore conducted on a smaller scale than the global-scale. It seams to be "cherry-picked' to favour a "region" where the model is better constrained. This is where satellite based data can be useful for instance to validate over the globe.

This is exactly what we did. In the first version of the manuscript that Reviewer #3 is referring to, we restricted the validation to the U.S. because that is the continent with readily available (open) groundwater level data and with multiple other models in place. However, in the second version we extended the evaluation to the globe by comparing results with GRACE and GRACE-FO satellite-based gravity anomaly data (See Section 2.4.3 and 3.3.3 and Figure 13 / Figure 14 is second revised manuscript).

Figure 12: It seems that the GIM model of Fan et al., 2017 performs better than the current physics-based MODFLOW model with 30'' resolution. How do you explain this? Although the model of Fan et al., 2017 can be calibrated, can such a model be 'calibrated' using any of the currently available methods? If so,

calibrated for what? Heads? Flow? Model calibration must be carried out with the aim of reducing the uncertainty of a given prediction. This is where the definition of the purpose of the model is very important.

Both GIM and CGM have been calibrated by comparing simulated heads with observed head data and adjusting parameters. For GIM, this was done by iteratively changing the way hydraulic conductivity reduces with depth (thus in fact the transmissivity) till the simulated heads resemble the observed ones as good as possible. No such calibration has been performed with our model. This is left as work for the future, as the focus of this work was to introduce the numerical scheme that allows parallel computing to solve transient high-resolution multi-layer groundwater problems.

Line 632: Is this type of model really intended for the "average user"?

This is a fair point. We have removed “average” and just speak of users without access to large numbers of nodes.

Line 643: It could be dangerous to reduce the number of model iterations, as this is likely to increase numerical errors. I would not advise this, especially for large models where small numerical errors can lead to large errors in the fluxes.

We fully agree that numerical accuracy should not be lowered to prevent numerical errors. Our text was unfortunately unclear and here our intention was to refer to the parallel preconditioner in the linear solver that could be improved in the future to reduce the number of linear iterations without loss of any accuracy. Therefore, we changed the text as (line 786):

improving the parallel preconditioner for the linear solver to account for the increasing number of iterations

Line 649: So this (i.e. the memory limitation) completely precludes the use of sophisticated inverse modelling and uncertainty analysis. On the one hand you reduce the run time of the forward model, but on the other hand you increase the memory requirement. This looks like an intractable problem.

Note that we are here talking about disk storage, taken by input files and output files if all output is stored after each time step, and not about Random Access Memory (RAM) that is used by processors on nodes during computation. We don’t expect that disk storage limitations precludes any sophisticated inverse modeling and/or uncertainty analysis. All the required input data need to be read in once and, as we show here, fit the processor memory of the collective nodes. Then multiple runs can be done as part of an uncertainty analysis or as iterations in an inverse scheme while parameters are updated in processor memory, as well as comparing parts of the outputs with observations that are also read in once. The feasibility of the uncertainty analysis or an inverse scheme thus depends only on the number of model runs needed. This could be speeded up by increasing the number of nodes used and is not limited by RAM requirements.

REFERENCES

Anyah, R. O., Weaver, C.P., Miguez-Macho, G., Fan, Y. and Robock, A. (2008). Incorporating water table dynamics inclimate modeling: 3. Simulated groundwater influence on coupled land-atmosphere variability,J. Geophys. Res.,113, D07103, doi:10.1029/2007JD009087.

Bierkens, M. F. P. (2015). Global hydrology 2015: State, trends, and directions. Water Resources Research, 51(7), 4923–4947. https://doi.org/10.1002/2015wr017173

Clark, M. P., Fan, Y., Lawrence, D. M., Adam, J. C., Bolster, D., Gochis, D. J., et al. (2015). Improving the representation of hydrologic processes in Earth System Models. Water Resources Research, 51(8), 5929–5956. https://doi.org/10.1002/2015WR017096

Dalin, C., Wada, Y., Kastner, T. and Puma, M.J. (2017). Groundwater depletion embedded in international food trade. Nature 543, 700–704. https://doi.org/10.1038/nature21403

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. (2021) GMD perspective: The quest to improve the evaluation of groundwater representation in continental- to global-scale models, Geosci. Model Dev., 14, 7545–7571, 28. https://doi.org/10.5194/gmd-14-7545-2021, 2021

Karabil, S., Sutanudjaja, E.H., Lambert, E., Bierkens, M.F.P. and van der Wal, R. (2021). Contribution of land water storage change to regional sea-level rise over the twenty-first century. Frontiers in Earth Science, 9, 2021. https://doi.org/10.3389/feart.2021.627648

Miguez-Macho, G., and Fan, Y. (2012). The role of groundwater in the Amazon water cycle: 2. Influence on seasonalsoil moisture and evapotranspiration, J. Geophys. Res.,117, D15114, doi:10.1029/2012JD017540

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.: PCR-GLOBWB 2 (2018), A 5 arcmin global hydrological and water resources 850 model, Geosci. Model Dev., 11, 2429–2453, https://doi.org/10.5194/gmd-11-2429-2018.

Reviewer #4

GENERAL COMMENTS

Verkaik et al. present a study on the parallel setup and application of GLOBGM-MODFLOW at the global scale. They describe the parallel methodology and application including evaluation with observations over a limited area i.e CONUS. Overall, the manuscript is difficult to follow. It reads as if no iteration has been done on clarity and structure. Here, I am focusing on the (parallel) setup of the model and performance.

We are sorry to read that the reviewer found it difficult to follow the manuscript. Since the reviewer is focusing on the parallel setup of the model and performance, we presume that the reviewer specifically has difficulties with Section 2.1. Therefore, we extensively restructured this section, renamed this to Parallelization approach, and introduced four (sub)sections labeled as: General concept (2.1.1); Procedure for deriving the independent unstructured grids (2.1.2); Defining the four groundwater models of the GLOBGM (2.1.3); Groundwater model partitioning: grid cell partitioning and catchment partitioning (2.1.4); and Node selection procedure (2.1.5). Section 2.1.1 summarizes the followed parallelization concepts, referring to the successive sections for details, and the general concept is now illustrated by a new figure, Figure 1. Where possible, we reused existing text and added new text mainly for Section 2.1.1 and Section 2.2.2. To improve the readibility, we sharpened our definitions, e.g., in the revised manuscript we changed the term “straightforward METIS partitioning” to “grid cell partitioning”, which we believe is more meaningful. Hopefully our restructuring helps the reviewer, together with our replies below.

SPECIFIC COMMENTS

The description of the setup needs careful revision. The authors do not help the reader using inconsistent terminology. I am trying to reconcile what they wrote/mean: the numerical groundwater model is based on structure grid in Cartesian coordinates (they actually state the Cartesian grid represent lat/long; how does that work?).

We deliberately left the details on this and referred to previous work of Sutanudjaja et al. (2011) and de Graaf et al. (2015), since the lat/long approach is similar to the 5’ groundwater model of PCR-GLOBWB. However, to clarify this more we added the text (line 128):

Using such a grid means that we have to take into account for the fact that cell areas and volumes do vary in space, and therefore MODFLOW input for the recharge and the storage coefficient need to be corrected for this (see Sutanudjaja et al. (2011) and de Graaf et al. (2015) for details).

The grid is applied over sea and land areas uniformly. Then they talk about unstructured grids. I believe they mean unstructured subdomains, because they subdivide the global model into “grids”, which are actually subdomains i.e. Afro-Eurasia, Americas, Australia, Islands?! Starting on line 179, these are then termed groundwater models which are partitioned in submodels put on one core each. Load balancing is used via METIS to improve efficiency; in this process watersheds are used in the weighing.

We note that we never used the term subdomain in the manuscript and in our terminology, we call Afro-Eurasia, Americas, Australia and Islands, *models* of the GLOBGM since they are independent models in our approach. We deliberately left out the term subdomain since we think “domain” might be nondescriptive for the readers. Instead, we use the term submodel, defined as being a part of one of the four models within the GLOBGM. We acknowledge that this could have been clearer. For that reason we introduced Section 2.1.1 (line 123) for describing the general concept, illustrating this with the new Figure 1 (e.g. line 135):

For addressing this problem, we can significantly reduce the number of grid cells by applying unstructured grids and maximize parallelism by deriving as many independent groundwater models as possible while satisfying all necessary boundary conditions. This concept is illustrated by Figure 1. Starting with the 30″ global-scale land-sea mask and boundary conditions prescribed by the GGM, we first derive independent unstructured grids and group them in a convenient way from large to small (see Section 2.1.2 for details). Then, we define the GLOBGM as a set of four independent groundwater models: three continental-scale groundwater models for the three largest unstructured grids and one remainder model called “Island model” for the remainder of the smaller unstructured grids (see Section 2.1.3 for details). The unstructured grids for these defined models are subject to parallelization: two partitioning methods (or domain decomposition methods) are considered (see Section 2.1.4 for details): one for partitioning grid cells straightforwardly (grey arrows in Figure 1) and the other for partitioning water catchments (red arrows in Figure 1). For each groundwater model, the chosen partitioning results in non-overlapping subgrids that define the computational cells for the non-overlapping groundwater submodels, where the computational work for each submodel is uniquely assigned to a processor core (MPI rank).

We note, also from other comments of the reviewer, that the GLOBGM definitely applies an unstructured grid for each of the four models of the GLOBGM (Afro-EurAsia, Americas, Australia, and Islands). This is now emphasised by Section 2.1.2 Procedure for deriving the independent unstructured grids, e.g. by line 191:

The resulting grids in the GLOBGM are clearly unstructured since the number of cell neighbors is not constant for all grid cells, and therefore the grid cell index cannot be computed directly: in the lateral direction, constant-head cells (Dirichlet; 0 m) near the coastal shore are not connected to any neighboring canceled sea cells, and in the vertical direction we cannot distinguish between upper and lower model layer anymore due to canceling of non-existing upper confining layer cells. Because of this, we apply the Unstructured Discretization (DISU) package with MODFLOW (Langevin et al., 2017).

Figure 2 is not intuitive; the caption needs to be expanded.

Sorry for being unclear here. We extensively expanded the caption (Figure 3, former Figure 2, line 277) as the reviewer suggested.

While stating on line 156 that the study is not "fixating" (non-scientific language, revise) on speedups and scaling, this is definitely required given that the parallel implementation is the focus of this study as suggested in the title. Thus strong/weak scaling results need to be added to the transient simulation section.

We apologize for our non-native English, and we changed "fixating" to "focusing". We agree with the reviewer on the importance of strong and weak scaling experiments. For that reason, a strong scaling was performed for the Americas model to estimate the submodel size for a given run-time target of 117 SYPD (Figure 9b, SYPD can interpreted as speedups). However, we did not perform a weak scaling analysis, for which grid refinement might be likely most appropriate instead of extending the computational domain. First, the focus of this research was on realizing a global groundwater model having a fixed 30" resolution, and not on models having a lower or higher spatial resolution. Second, downscaling/upscaling input data and boundary conditions are generally not straightforward (e.g. drain/river packages with different drainage networks). Third, automation of the pre-processing for accommodating grid refinement would require a significant programming effort, as well as many computation hours for running and post-processing, which is beyond the scope of this paper. Fourth, the focus was here not on code profiling our MODFLOW 6 prototyping code. From previous tests, we see comparable parallel strong-scaling performance compared to our previous parallelization efforts for MODFLOW 2005 and SEAWAT. Together with the promising results with respect to SYPD (Figure 11), we did not feel the need to conduct any weak scaling experiments for this paper.

Overall, the results with respect to SYPD appear to be promising. This raises the question, why a one/two-layer groundwater model has been implemented and applied.

The main reason for this is the present lack of borehole and geophysical data at the global scale (see e.g. Section 4.1, bullet 2 in Condon et al. 2021). Note that we are working on collecting and interpreting more borehole/geophysical data for a future version of the GLOBGM to add more model layers to represent the geology better.

One/two-layer groundwater flow models have been used 40 years ago. Thus, the results obtained with this type of model are at a level of sophistication that is not state-of-the-art.

That depends on the data scarcity. For data-rich countries where many borehole/geophysical data are publicly available, like in the Netherlands (Van der Meulen et al., 2013), (regional-scale) groundwater models having one/two model layers are indeed not state-of-the-art, and we agree with the reviewer in that respect. However, when little to no subsurface data is available, which is the case for global-scale groundwater models like the GLOBGM (see our reply above), models are inherently bound to having one/two model layers. Although from a physical/modeling point of view this is undesirable, from a data-availably point of view those models can still deemed to be considered state-of-the-art.

Especially, when figure 10 suggests that a more complex flow model is computationally possible.

We agree that computationally there seems much more possible since we only used 12 nodes (Table 4) of the Snellius supercomputer instead of all the total ~1300 nodes, but this revelation came to the surface as our research progressed. Even with the current version of the GLOBGM, we believe performance can be further improved by adding more nodes. However, our focus was rather on usability of the current version of the GLOBGM than on computability of future versions. For future versions of the GLOBGM, we believe that the updated flow model complexity should be re-tuned with the available nodes and scenario runtimes.

This is perhaps also the reason for the low and negative correlations in figure 14.

Thank you for your suggestion. When looking at Figure 15 (former Figure 14), i.e. a comparison with GRACE, we only observe negative correlation for the annual correlation (Figure 15b), mainly in the Amazon and Sahara. As mentioned in the text, we think these negative correlations are more likely to be caused by large water storages in floodplains for the Amazon (line 716-717: “Furthermore, the low… floodplains during,”) and higher noise-to-signal ratio in the Sahara (line 719-720: “Additionally, it is… …like the Sahara”).

This brings up the purpose of the paper that is to show that 30’’ global transient groundwater simulations are possible. Coming from the mathematical/computational perspective, the authors solve a linear PDE on a structured grid over a domain with complex geometry.

We unfortunately disagree with the reviewer on the comment “*the authors solve a linear PDE on a structured grid*”. First, the groundwater flow equation subject to solving in our manuscript is non-linear, since river and drain (Robin/Neumann) boundary conditions are used that require (here Picard) linearization and therefore non-linear (outer) iterations. Second, the grid is unstructured and not structured (see above explanation; Section 2.1.2 in the revised manuscript). From a groundwater modeling perspective, this makes our GLOBGM one of the very first peer-reviewed presented models using the new MODFLOW 6 unstructured grid functionality.

In total ~9.3 million cells (line 120) corresponding to DOFs are implemented.

First, we apologize for making a mistake with this number the reviewer is referring to on line 120 of the first revised manuscript. This should be ~18.7 million cells, and we corrected this in the revised manuscript (line 131):

Each of the two GGM layers has 9.3 million 5′ cells (4,320 columns times 2,160 rows), and therefore the GGM has a total of ~18.7 million 5′ cells. A straightforward refinement of this grid to 30″ resolution would result in ~100 times more cells, hence 1.87 billion cells (two model layers of 43,200 columns times 21,600 rows).

Second, we disagree with the reviewer that only ~9,3 million cells / DOFs (correct: ~18.7 million) are implemented for the GLOBGM. We emphasize that the GLOBGM is not a straight-forward (nearest-neighbor) interpolated version of the 5' GGM. Although some model input data are used at a coarser resolution using interpolation, e.g. recharge and well abstractions (see Table 1), the GLOBGM really uses detailed 30" model input data: e.g. a 30" model top layer derived from upscaling the 3" MERIT DEM (see Table 1), and 30" surface water levels derived from 30" HydroSHEDS data (see Table 2). This makes the GLOBGM v1.0, having 278 million DOFs in total (168 million DOFs for the largest Afro-Eurasia model), sufficiently numerically challenging for HPC, stressing both the linear and non-linear solver, as well as the pre-processing for transient simulation.

Even without disaggregation into independent groundwater models, this is not considered a large computational problem.

First, from our estimates the serial GLOBGM would require at least 3 months of runtime (line 608: "…or 87 days runtime") to simulate 1958-2015 considering the largest Afro-Eurasia model of the GLOBGM, which is a large computational problem. Second, since the reviewer seems to refer to the transient 5' GGM performance, even for this model runtimes are significant. Although this was not a subject of research for our paper, the GGM requires about ~50 minutes runtime in serial. Hence, for 78 years of simulation (1958-2015, 38 years + 20 years spin-up) about 78 x 50 min = 65 hours = 2.7 days is required, which is still significant.

Thus, in terms of modeling and HPC, the study is of limited scope. The authors need to discuss why the study is relevant given the limitations outlined in the discussion above.

We hope that our above answers have convinced the reviewer that the GLOBGM model is challenging with respect to HPC, i.e. as summarized in the conclusions of our manuscript:

- We believe that the GLOBGM is a numerically challenging, transient, groundwater model having 278 million DOFs in total (168 million DOFs for the largest Afro-Eurasia model), that we have sucessfully implemented with limited and reachable parallel hardware requirement (line 753: "successfully implemented…").
- To our knowledge, the GLOBGM is the first global-scale transient 30″ groundwater model made possible by our parallelization (line 755: "To our knowledge, …").
- The GLOBGM uses parallel pre-processing, to our knowledge new for MODFLOW-based groundwater modeling, and we showed its necessity (line 759: "using parallel pre-processing').
- The GLOBGM uses unstructured grids, making this model one of the first MODFLOW 6 models presented for peer-reviewing using this functionality (line 756: "Our implementation uses…").

- The GLOBGM is the first parallel application using our new parallellized version of MODFLOW 6 (line 760: “a new parallel distributed memory prototype version”).
- Besides straight-forward cell partitioning, the GLOBGM can apply sub-optimal partitioning using catchments that turned out to be a promising method (769: “For catchment partitioning, …”’).

Although we do believe that in terms of modeling and HPC we have significant scope and made advances, we note that these HPC advances mainly serve the primary goal of the paper (Figure 11 and 12), i.e. realizing a first transient 30” global-scale groundwater model meeting user requirements.

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. https://doi.org/10.1029/2020WR029500

de Graaf, I. E. M., Sutanudjaja, E. H., Van Beek, L. P. H., and Bierkens, M. F. P.: A high-resolution global-scale groundwater model, Hydrol. Earth Syst. Sci., 19, 823–837, https://doi.org/10.5194/hess-19-823-2015, 2015.

Sutanudjaja, E. H., Van Beek, L. P. H., De Jong, S. M., Van Geer, F. C., and Bierkens, M. F. P.: Large-scale groundwater modeling using global datasets: A test case for the Rhine-Meuse basin, Hydrol. Earth Syst. Sci., 15, 2913–2935, https://doi.org/10.5194/hess-15-2913-2011, 2011.

Van der Meulen, M.J., Doornenbal, J.C., Gunnink, J.L., Stafleu, J., Schokker, J., Vernes, R.W., Van Geer, F.C., Van Gessel, S.F., Van Heteren, S., Van Leeuwen, R.J.W., Bakker, M.A.J., Bogaard, P.J.F., Busschers, F.S., Griffioen, J., Gruijters, S.H.L.L., Kiden, P., Schroot, B.M., Simmelink, H.J., Van Berkel, W.O., Van der Krogt, R.A.A., Westerhoff, W.E., Van Daalen, T.M., 2013. 3D geology in a 2D country: Perspectives for geological surveying in the Netherlands. Netherlands J. Geosci. / Geol. en Mijnb. 92, 217–241. https://doi.org/10.1017/S0016774600000184

Please let us know if you have any questions or more clarification is required. We hope that after this second round of revisions the paper has improved sufficiently to be considered for publication in GMD.

On behalf of all authors,

With kind regards,

Jarno Verkaik