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.

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.

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?

Other comments

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?

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).

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

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.

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.

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?

Line 505: typo: “het”

Line 568: typo: “be left”

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.

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.

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

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.

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.