Reply on RC1

R1: This paper describes recent updates to the WaterGAP global water resources model (v2.2e) and provides benchmarks against observations.

As with previous versions, the model is impressive and the analysis is comprehensive. However, the improvements described in this paper, and their impacts on performance, appear fairly minor. There should be space in GMD for efforts that focus primarily on software and data updates rather than scientific ones, but revisions are needed.

Answer: Thank you for your time to review the manuscript and for providing your comments and suggestions. We will reply to the referee comment, indicated by **R1** by our answer, indicated by **Answer**, and corresponding actions, indicated by **Action** and textual changes in *italic font*:

R1: 1. The improvements include the treatment of reservoirs, updates to datasets for reservoirs, non-irrigation water use, and streamflow stations for calibration. Updates to model capabilities include PET representations, glaciers, and water temperature, though the latter has already been described in a previous publication. While all of these changes are justified, and no doubt an intensive effort, their impact on the modeled global water balance and distribution of NSE presented in the results does not seem to be a large change from previous versions of WaterGAP. Section 7.5.1 reports a nearly identical performance to the previous version. The scientific contribution of the new model capabilities should be more strongly justified.

Answer: Thank you for your suggestion. As mentioned by the referee in the statement of the beginning, GMD has space for such rather technical descriptions. In particular, the manuscript is submitted as a model description paper type which (among other purposes) "… should be detailed, complete, rigorous, and accessible to a wide community of geoscientists." (https://www.geoscientific-model-development.net/about/manuscript_types.html#item1). Our approach is to describe this model version, provide evaluation details and give examples of model output. Scientific contributions are not necessarily the focus of this manuscript type and not necessarily intended by this manuscript. It is rather a documentation for the model outputs e.g. used within the ISIMIP phase 3 and to bring together all the features of this model version in a description of the model version. So, we do not aim for more scientific justification as this is not the focus of this manuscript type.

Action: none

R1: 2. A main focus of the updates is the reservoir model. From the previous paper (2021), the release policy is assumed to follow Hanasaki (2006) and Döll (2009), which distinguishes between irrigation and non-irrigation reservoirs. This is a simplified rule that can be applied globally, but is often inaccurate at the level of individual reservoirs. The current paper does not investigate changes to this assumption, but removes a previous limitation about the maximum storage capacity for flood prevention. The accuracy of reservoir storage shown in Supplemental Figures S1, S2 from version 2.2c leaves much room for improvement, and it is not clear that removing the storage capacity threshold will fix this. The updated results after the change are not shown.

Answer: We have to apologize. The storage capacity threshold was already removed for the model version 2.2d (noted in Müller Schmied et al., 2021, Section A2.4 (last bullet point) and the inclusion in this manuscript was a result of an internal communication problem. We agree that the generic reservoir algorithm has several limitations and efforts are ongoing to improve the representation of reservoirs in large-scale hydrological models (Dong et al, 2023, Otta et al, 2023, Shrestha et al, 2024, Steyard & Condon, 2024).**Action**: We have removed the update of the reservoir algorithm completely from this manuscript.

R1: 3. More information should be included about the potential scale mismatch between reservoir outflow (a point) mapped to a larger grid cell. The same goes for the stream gage data used for calibration (Step 3 in Section 2.5.2). It is possible this information is included in previous papers, but it would help to discuss here the potential impacts of this scale mismatch.

Answer: Indeed, there is a scale mismatch of a point information that is located somewhere in a grid cell and a drainage direction map at 0.5 x 0.5 degree spatial resolution. Basically, the model assumes that there is one river per grid cell, and generally, the whole grid cell(s) contribute to the runoff for the basin. With regards to both, the reservoir outflow location and the location of streamflow station for calibration, the given coordinate does not always fit to the hydrological situation and the (not always provided) upstream drainage area from the station/reservoir location and the drainage network. We have manually checked the location of the coordinate of the station/reservoir and its hydrological situation, esp. if the station or a reservoir outflow is located at a tributary or the main stream. Hence, we have co-registered stations to the best-located grid cell (in order of good match of observed and DDM upstream basin area) but decided not to use correction factors in case basin sizes differ (given all the other uncertainties and as we anyhow use a threshold of 9000 km2 ~ 4 grid cells as minimum for calibration). In the shapefiles of the calibration data (Müller Schmied & Schiebener 2022), the basin area from the data provider as well as from the drainage direction map are provided. We will not add this information to the manuscript. However, we have added some text to reflect this discussion.

Action: We have added the following text to Line 66 (after the citations Döll and Lehner, 2002 and Schewe and Müller Schmied, 2022): *The location of the new reservoirs was manually co-registered in the drainage network with the help of web-based map information in order to match the given hydrological situation, in particular if a reservoir is located on the mainstream or its tributary.*

We have added the following text to Line 143 (after “re-map the station to a grid cell that fits with the drainage network”): *Re-mapping of the position focused on accurately relating the station either to the mainstream of the river or the tributary. A correcting factor for mismatches of drainage areas between the values provided by the station data producers and those calculated from the drainage direction map was not implemented but both areas can be found in the shapefiles of Müller Schmied & Schiebener (2022).*

R1: 4. The calibration process attempts to find an optimal value of gamma, the runoff coefficient, to align the modeled mean annual streamflow within either 1% or 10% of the observed. Failing this, additional correction factors are applied to the runoff. While I can appreciate the difficulty of calibrating a global model, this calibration setup would have several problems for a basin-scale study. The gamma parameter can compensate for any mass balance error without a physical relationship to the runoff curve shown in Figure 3 of the 2021 paper. The additional correction factors only worsen this problem, and many regions of the model rely on these (Fig 4 of the current paper). By calibrating to mean annual data, monthly dynamics could be lost, though the efficiency metrics reported in Figure 7 seem to be doing well at many stations.

This calibration approach may be standard for global models. But at the basin scale, we could expect to see more diagnostics applied to investigate whether the results are physically based, or to analyze how much of the calibration uncertainty comes from each component of the mass balance. The calibration is more of a bias correction that is not able to distinguish between the many degrees of freedom in the model.

Answer: We fully agree that this bias adjustment it is a rather simple approach and it is good to add the term bias adjustment into the text for clarification. More extended calibration approaches are of course available and tested also with WaterGAP (Döll et al., 2024, Hasan et al., 2023), but not yet applicable on global scale, and also not included in a standard version of the model. Indeed, the original idea was to reduce biases by calibrating the HBV beta (our gamma) to match mean observed annual streamflow for water resources assessment and in many cases the uncertainties are large enough that gamma alone is not enough to achieve the aim. With regards to the last sentence of the first paragraph (the effect on monthly dynamics when only the mean value is calibrated), we here refer to the abstract of Hunger and Döll (2008) where, it was stated that “other flow characteristics like low flow, inter-annual variability and seasonality, the deviation between simulated and observed values also decreases significantly, which, however, is mainly due to the better representation fo average discharge but not of variability”. This is also reflected in relatively weak performance of the KGE variability parameter.

We also agree that further diagnostics are needed to elaborate on reasons for different model performance but this is outside of focus of this manuscript.

Action: We introduce the term bias adjustment in the paper for clarity. In the beginning of Section 5.2 (Line 314) we add the sentence: *“The calibration as implemented in the standard version of WaterGAP focuses on adjusting biases in a rather simple method. More comprehensive approaches are currently in development (Döll et al., 2024, Hasan et al., 2023) and might be used in future model versions.”*

References

Dong, N., Yang, M., Wei, J., Arnault, J., Laux, P., Xu, S., et al. (2023). Toward improved parameterizations of reservoir operation in ungauged basins: A synergistic framework coupling satellite remote sensing, hydrologic modeling, and conceptual operation schemes. Water Resources Research, 59, e2022WR033026. https://doi. org/10.1029/2022WR033026

Döll, P., Hasan, H. M. M., Schulze, K., Gerdener, H., Börger, L., Shadkam, S., Ackermann, S., Hosseini-Moghari, S.-M., Müller Schmied, H., Güntner, A., and Kusche, J.: Leveraging multi-variable observations to reduce and quantify the output uncertainty of a global hydrological model: evaluation of three ensemble-based approaches for the Mississippi River basin, Hydrol. Earth Syst. Sci., 28, 2259–2295, https://doi.org/10.5194/hess-28-2259-2024, 2024.

Hasan, H. M. M., Döll, P., Hosseini-Moghari, S.-M., Papa, F., and Güntner, A.: The benefits and trade-offs of multi-variable calibration of WGHM in the Ganges and Brahmaputra basins, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2023-2324, 2023.

Müller Schmied, H., & Schiebener, L. (2022). The global water resources and use model WaterGAP v2.2e: streamflow calibration and evaluation data basis (1.1) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.7255968

Otta, K., Müller Schmied, H., Gosling, S. N., and Hanasaki, N.: Use of satellite remote sensing to validate reservoir operations in global hydrological models: a case study from the CONUS, Hydrol. Earth Syst. Sci. Discuss. [preprint], https://doi.org/10.5194/hess-2023-215, in review, 2023.

Shrestha, P. K., Samaniego, L., Rakovec, O., Kumar, R., Mi, C., Rinke, K., & Thober, S. (2024). Toward improved simulations of disruptive reservoirs in global hydrological modeling. Water Resources Research, 60, e2023WR035433. https://doi.org/10.1029/2023WR035433

Steyaert, J. C. and Condon, L. E.: Synthesis of historical reservoir operations from 1980 to 2020 for the evaluation of reservoir representation in large-scale hydrologic models, Hydrol. Earth Syst. Sci., 28, 1071–1088, https://doi.org/10.5194/hess-28-1071-2024, 2024.