Dear Reviewer 1,

Thank you for taking the time to review our paper. We are grateful for your thoughtful comments. We trust that the revisions presented below address the issues you raise.

Yours faithfully,
Naota Hanasaki (on behalf of authors)

This is an interesting study. The authors are exploring the use of satellite remote sensing to validate reservoir operations in global hydrological models (GHMs).
I think if we take the perspective of the GHMs and their need to be improved as a modeling framework by moving away from assumptions on reservoir operations and parameterizations (which GHMs have a lot of), then this is a valuable study.
However, if this study is just exploring if we can use satellite data to track reservoirs, I don't think this study is innovative or adds much to the body of knowledge. That topic has been addressed and most of the relevant questions on that that the authors pose in this study have been answered much more vigorously and in more detail over the last 10 years. So the first question posed by the authors is really redundant here (we know the answer): Can satellite-based storage estimation data serve as a surrogate for ground truth data?

> We appreciate you sharing your fundamental standpoint and perception of our study. You raise two issues. The first is that the parameterization of reservoir operation could be removed from global hydrological models (GHMs) if the latest satellite data are used. The second is that our first research question is invalid; hence, it is redundant. We will respond to these concerns individually.
> Regarding your first point, we have a totally different view. One of the fundamental reasons for developing a GHM was to obtain long-term and continuous estimations of global hydrological variables from the past to the future by tracking hydrological processes. Satellite-based monitoring can substitute for parameterization only for past periods, when satellite data are available. This is obviously problematic for future projections, which is one of the primary interests of the many GHM developers. We firmly believe that the importance and necessity of reservoir parameterization will be unchanged regardless of how powerful satellite remote sensing becomes. This claim has now been included in the Introduction as follows: "Reservoir parameterization is indispensable in GHMs, particularly when they are

applied for future projections."

Regarding the second point, as for our first research question, we accept your point. We have used several published and validated satellite products. In the revised paper, we have withdrawn the question and recontextualized Subsection 3.1 as an introduction to the available data. Our primary claim is that satellite data should be fully utilized for the validation and intercomparison of GHMs. In this section, our intention was to explain what data are used and how they compare with ground observations. We have therefore modified the "Introduction" section to clarify further our standpoint. We have added the following text: "We first validated the satellite data against ground observations to ensure the thoroughness of our study, although this was not the primary focus of the study."

The authors have used too simplistic methods of using only altimeter from DAHITI (with limited sampling in space and time) and Landsat (optical) based reservoir area dataset from GRSADs. Today when we have a fleet of satellite sensors (optical, SAR, altimeters) that collectively provide more robust, high frequency and more accurate tracking of reservoirs in terms of surface area or elevation to generate storage change.

So I think the authors should reorganize the paper and reduce section 3.1 and focus more on 3.2 as the key focus of the paper where there is a lot to learn for the GHM community trying to improve representation of reservoirs on global models (which also fits perfectly into the theme of the special issue of HESS).

Thank you for providing details of the latest achievements in satellite data processing. We have added a paragraph to explain how we selected the satellite data in Subsection 2.4 Satellite data. "As shown in the Introduction, monitoring reservoir storage from space is a rapidly evolving study domain. Many studies have focused on one or a few reservoirs and applied the latest techniques for accurate, continuous, and frequent monitoring (e.g., Das et al. 2022). Because our final goal was to validate GHMs globally at a decadal scale, we acquired data that covered the entire globe for more than 10 years. To the best of our knowledge, there are six datasets that meet this requirement, namely, DAHITI (Schwatke et al. 2015), GRSAD (Zhao and Gao 2018), Hydroweb (Crétaux et al. 2011), G-REALM (Birket and Beckley, 2010), Glo Lakes (Hou et al., 2022), and Biswas and Hossain (2022). We excluded Hydroweb and G-REALM because the number of reservoirs monitored is one or two orders of magnitude smaller than other datasets. GloLakes is highly promising, but is still under review (as of November 17, 2023). We selected DAHITI and GRSAD because their data were easily accessible. It should be noted that new datasets are being produced by applying the latest satellite sensors. Among the other potential datasets, the Cooley et al. (2021) dataset, which utilizes the ICESat-2 altimetry, is very promising, but we excluded it because it covers only 2018 and later." Additionally, we note that DAHITI uses multiple satellite data to generate the elevation data product (https://dahiti.dgfi.tum.de/en/products/water-level-altimetry/). This was already noted in Subsection 2.4.1. Finally, we completely agree with Reviewer 1 that the latest information should be used as much as possible. This is repeatedly mentioned in the Conclusions section.

Regarding the latter part of the comment, we have revised our paper by taking on board the Reviewer's insightful suggestions. First, we have changed the title of Subsection 3.1 from "Validation of satellite data" to "Satellite data used in this study". It now starts with the following paragraph. "Satellite data have been cross-referenced with the growth of observations from various studies (e.g., Hou et al. 2022). The following section presents the critical details of the data and methods used in this study, to ensure the comprehensiveness of our approach." Please note that is it important to retain Subsection 3.1 because it explains to readers why we needed to normalize the data. We have presented a more straightforward explanation of this and shortened it as much as possible.

Let me share some key references the authors miss in capturing the state of the art of satellite-based reservoir tracking:

Bonnema, M., C.H. David, R.P. d. M. Frasson, C. Oaida, & S. -H. Yun. (2022). The global surface area variations of lakes and reservoirs as seen from satellite remote sensing. Geophysical Research Letters, vol. 49, e2022GL098987. https://doi.org/10.1029/2022GL098987

Cooley, S.W., J.C. Ryan and L.C. Smith (2021), Human alteration of global water storage variability, Nature, vol. 591, pages 78–81 https://doi.org/10.1038/s41586-021-03262-3

Das, P., F. Hossain, S. Khan, N. K. Biswas, H. Lee, T. Piman, C. Meechaiya, U. Ghimire, K. Hosen (2022) Reservoir Assessment Tool 2.0: Stakeholder driven Improvements to Satellite Remote Sensing based Reservoir Monitoring, Environmental Modeling and Software, Vol. 157. https://doi.org/10.1016/j.envsoft.2022.105533

Biswas, N. and F. Hossain (2022) A Multi-decadal Analysis of Reservoir Storage Change in Developing Regions, Journal of Hydrometeorology, Vol. 21(1), pp 71-85. https://doi.org/10.1175/JHM-D-21-0053.1

Biswas, N., F. Hossain, M. Bonnema, H. Lee, F. Chishtie (2021). Towards a Global

Reservoir Assessment Tool for Predicting Hydrologic Impacts and Operating Patterns of Existing and Planned Reservoirs, Environmental Modeling and Software, Vol. 140. https://doi.org/10.1016/j.envsoft.2021.105043

Zhou, T., Nijssen, B., Gao, H., & Lettenmaier, D. P. (2016). The Contribution of Reservoirs to Global Land Surface Water Storage Variations. Journal of Hydrometeorology, vol 17(1), pp. 309–325. https://doi.org/10.1175/JHM-D-15-0002.1

I urge the authors to explore the above papers very carefully to reframe their introduction and research questions. It will become clear from these papers that
bjust relying on altimeter data on reservoir levels from DAHITI and a corrected Landsat only reservoir area dataset really do not reflect the true state of the art on what we can do today in capturing reservoir dynamics at sub weekly time scales using multiple sensors, wavelengths, and innovative methods. The authors will realize that they are answering that question with the most primitive tools and hat too by not really defining what the end goal is to be a useful 'surrogate' for GHM model development.
Monthly reservoir tracking is déjà vu and not hard to do with or without satellites as at that scale, most reservoirs lag the prevailing hydrology (unless they store multiple years of annual runoff). At monthly or longer time scales, GHMs can work reasonably well with parameterizations based on capacity, embedded rule curves and objective functions.
However, today, we really do not need to do that as we have 40 years of satellite data, with the last 10 years of that being a very high frequency, multi-sensor to help us track at a granular level how reservoirs have been operated (to help us model them in GHMs).

First, thank you for introducing these new references, which have now been cited in the Introduction section of the revised manuscript.

Second, we learned from these papers that the state-of-the-art satellite technology you refer to is basically applicable from 2016 onwards, following the launch of Sentinel 3. Because the ISIMIP historical simulation covers the period between 1901 and 2019, the period of overlap is only four years. This point has been clarified in the aforementioned data selection paragraph.

Third, there are two issues regarding the temporal resolution. We have now added the following text to the Methods Section. "We analyzed the data at a monthly interval for two reasons. First, although we understand that the latest satellite product can estimate the sub-monthly dynamics of reservoir storage. Even though GHM simulations can be run at a daily scale, ISIMIP simulation protocol foresees only monthly resolution of nearly all variables to avoid excessive storage and transfer demand. However, for the future simulations, we would propose daily scale to considered as a part of the ISIMIP protocol. Second, as Masaki et al. (2017) clearly show, global hydrological modelers have still observed substantial differences in monthly simulations among models. These differences stand out clearly when the catchment area is relatively small or multiple reservoirs are cascading. For a global simulation at a spatial resolution of half a degree, we considered a monthly resolution to be largely valid."

Fourth, as we understand, and as stated before, highly accurate multi-sensor satellite data covering more than a decade will potentially be available in the near future (i.e., satellite data typically started from 2016 once the Sentinel 3 and ICESat-2 data became available). For the immediate validation for ISIMIP Phase 3, we believe our method is still valid. We have now clearly acknowledged Biswas and Hossain (2022) who developed a 35 year sub-monthly scale time series of global reservoir storage data. This is clearly stated in the aforementioned data selection paragraph.

My request to the authors therefore is to get rid of question 1, tone down or revise substantially section 3.1 and focus most of the paper on section 3.2 where the key contribution lies. Authors should explore using additional datasets -there is now Hydroweb (multiple altimeters), use Sentinel series from 2016 and even MODIS (for larger surface area reservoirs). For the selected reservoirs, the authors have one advantage that cloud cover is minimum. They also do not need to rely on SRTM DEM as that can be very erroneous and over the US, SRTM captured only the 'free bathymetry' above the water level that existed during February 2000 when the Shuttle flew. Perhaps authors could test the validity of SRTM DEM with topo maps and published bathymetry and area-elevation curves (consider checking the RESSED database of USGS). The authors use DAHITI only for elevation but as they point out, multiple sensors can be used to generate surface area which apparently DAHITI has done but is not available in the authors' reservoirs of interest. So why not generate those reservoir areas from multiple sensors and indices by the authors themselves?

Thank you for sharing this idea. First, we agree with your recommendation regarding the restructuring of Subsections 3.1 and 3.2 and this has been adopted in the revised manuscript.

Second, we considered the recommendation regarding the use of multiple sensor satellite products to be largely unworkable, primarily because, as mentioned above, the simulation and data periods do not overlap. It could be possible to implement the use of Hydroweb. We found that Hydroweb offers data for four of the seven studied reservoirs. Hydroweb provides water level data slightly more frequently than do GRSAD and DAHITI (12 records per year) for two reservoirs (the Glen Canyon Dam and the Fort Peck Dam; see Table R1), but far less information regarding changes in water storage (See Table R2). Because of the large amount of missing data for our analysis period (1980–2017, see Figure R1), we ultimately decided against the use of Hydroweb data in our revised draft.

Table R1 Water level data from Hydroweb

Name	Records	Period (in decimal year)	Records per year
Hoover	197	1995.45-2020.57 (25.12 yr)	7.8
Glen Canyon	347	1992.79-2020.31 (17.52 yr)	19.8
Structure 193	250	1995.41-2020.82 (25.41 yr)	9.8
Fort Peck	479	1992.79-2022.68 (29.89 yr)	16.0

Table R2 Changes in water storage from Hydroweb

Name	Records	Period (in decimal year)	Records per year
Hoover	108	1995.45-2020.57 (25.12 yr)	4.3
Glen Canyon	176	1992.79-2020.31 (17.52 yr)	10.0
Structure 193	227	1995.41-2020.82 (25.41 yr)	8.9
Fort Peck	59	1992.79-2022.68 (29.89 yr)	2.0

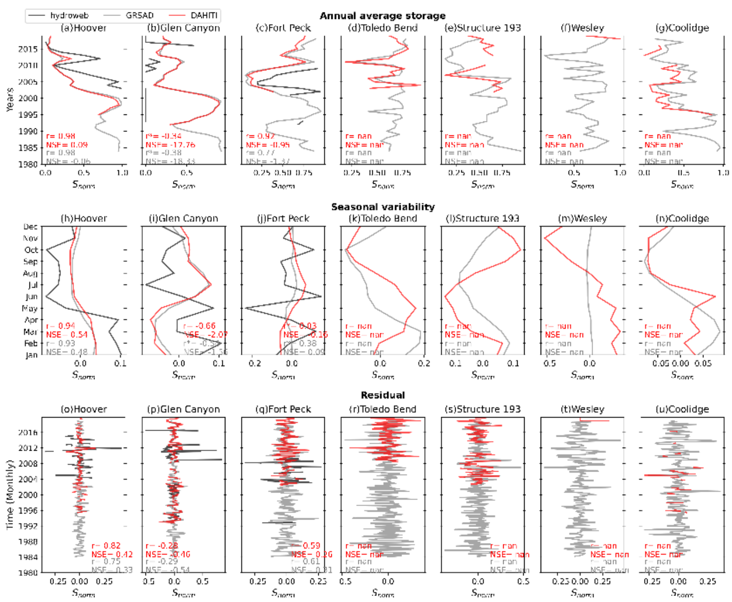

Figure R1 Figure 4 including Hydroweb data for the Hoover, Glen Canyon, Fort Peck, and Toledo Bend dams. Due to the large amount of missing data, the time series cannot be meaningfully compared with the other two datasets.

Third, we only mentioned that the GRSAD data utilized the shapefile of GRanD data, which was based on the SRTM. We have removed this text because it is not fundamentally important to the study.

Fourth, we believe that the generation of new data products would be impractical for the reasons shown above. Also, generating a new satellite data product was obviously beyond the scope of this study. This has been clearly explained in the Introduction as follows: "The following topics are beyond the scope of this study: 1) the improvement of the generic reservoir operation algorithms embedded in GHMs; 2) the generation of new satellite-based reservoir data; and 3) the production of case studies at a handful of reservoirs (this is, rather, a global study)."

There is also the SWOT mission the authors can talk about in a few sentences in the conclusion section (see http://swot.jpl.nasa.gov). The whole premise of SWOT is to generate simultaneously area and elevation so that we don't have to jury rig the observation system to derive storage change. There is now plenty of 'help' resources to help the community build literacy on SWOT (just click on 'applications' of the SWOT website).

Thank you. We have now mentioned the SWOT mission in the Conclusion section. "Advanced data are being produced by developers (Hou et al., 2022 and Cooley et al., 2022) and, with the launch of the SWOT mission, the potential for satellite-based surface water tracking is poised to expand significantly (Biancamaria et al. 2016)."

It is also no surprise that DAHITI (elevation) is generating better results for GHMs in section 3.2 and the GRSADs - this is something we keep seeing all the time as all reservoirs experience storage change via elevation change at levels detectable by altimeters. The same can't be said about surface area changes unless the reservoir is very large and does not have a complex dendritic shape. A lot depends on the shape, shoreline, climate, surrounding terrain of the reservoirs in how well or poorly a specific satellite data will work in tracking surface area or volume change. That is why I reiterate that the authors should focus more on the GHM model validation part rather than the satellite data assessment part in Section 3.1

Thank you. We have given your comment serious consideration and have added additional detail to the Discussion section: “This is a recurring observation because all reservoirs undergo storage changes reflected in elevation alterations detectable by altimeters (Busker et al., 2019; Verma et al., 2021). However, the same is not true for surface area changes unless the reservoir is particularly large and does not have a complex dendritic shape. The effectiveness of specific satellite data in tracking changes in surface area or volume is greatly influenced by factors such as the reservoir shape, shoreline characteristics, climate, and the surrounding terrain.” As advised, the focus on data assessment in Section 3.1 has now been toned down by moving section *3.1.2 (Decomposed monthly reservoir storage from satellite and ground observation)* to supplementary information.

I noticed many typos that I tried to note down. One is:

Line 144. I don’t think it’s year 286 to 2020 (although we do have many reservoirs built that early still functional in many places)

Thank you for allowing us to revisit this. GRanD includes Lake Qattinah (also called Lake Homs), which was a reservoir constructed by the ancient Romans. We have therefore not changed this text.

References

Biancamaria, S., Lettenmaier, D. P., and Pavelsky, T. M.: The SWOT Mission and Its Capabilities for Land Hydrology, Surveys in Geophysics, 37, 307-337, 10.1007/s10712-015-9346-y, 2016.

Birkett, C. M., and Beckley, B.: Investigating the Performance of the Jason-2/OSTM Radar Altimeter over Lakes and Reservoirs, Marine Geodesy, 33, 204-238, 10.1080/01490419.2010.488983, 2010.

Biswas, N. K., and Hossain, F.: A Multidecadal Analysis of Reservoir Storage Change in Developing Regions, Journal of Hydrometeorology, 23, 71-85, https://doi.org/10.1175/JHM-D-21-0053.1, 2022.

Busker, T., de Roo, A., Gelati, E., Schwatke, C., Adamovic, M., Bisselink, B., Pekel, J. F., and Cottam, A.: A global lake and reservoir volume analysis using a surface water dataset and satellite altimetry, Hydrol. Earth Syst. Sci., 23, 669-690, 10.5194/hess-23-669-2019, 2019.

Cooley, S. W., Ryan, J. C., and Smith, L. C.: Human alteration of global surface water storage variability, Nature, 591, 78-81, 10.1038/s41586-021-03262-3, 2021.

Crétaux, J. F., Arsen, A., Calmant, S., Kouraev, A., Vuglinski, V., Bergé-Nguyen, M., Gennero, M. C., Nino, F., Abarca Del Rio, R., Cazenave, A., and Maisongrande, P.: SOLS: A lake database to monitor in the Near Real Time water level and storage variations from remote sensing data, Advances in Space Research, 47, 1497-1507, https://doi.org/10.1016/j.asr.2011.01.004, 2011.

Das, P., Hossain, F., Khan, S., Biswas, N. K., Lee, H., Piman, T., Meechaiya, C., Ghimire, U., and Hosen, K.: Reservoir Assessment Tool 2.0: Stakeholder driven improvements to satellite remote sensing based reservoir monitoring, Environmental Modelling & Software, 157, 105533, https://doi.org/10.1016/j.envsoft.2022.105533, 2022.

Hou, J., Van Dijk, A. I. J. M., Renzullo, L. J., and Larraondo, P. R.: GloLakes: a database of global lake water storage dynamics from 1984 to present derived using laser and radar altimetry and optical remote sensing, Earth Syst. Sci. Data Discuss., 2022, 1-20, 10.5194/essd-2022-266, 2022.

Schwatke, C., Dettmering, D., Bosch, W., and Seitz, F.: DAHITI – an innovative approach for estimating water level time series over inland waters using multi-mission satellite altimetry, Hydrol. Earth Syst. Sci., 19, 4345-4364, 10.5194/hess-19-4345-2015, 2015.

Verma, K., Nair, A. S., Jayaluxmi, I., Karmakar, S., and Calmant, S.: Satellite altimetry for Indian reservoirs, Water Science and Engineering, 14, 277-285, https://doi.org/10.1016/j.wse.2021.09.001, 2021.

Zhao, G., and Gao, H.: Automatic Correction of Contaminated Images for Assessment of Reservoir Surface Area Dynamics, Geophys. Res. Lett., 45, 6092-6099, https://doi.org/10.1029/2018GL078343, 2018.