**Response to Reviewer #2 Comments:**

Hou et al. presented a nice study on creating a new time-varying dataset on global lake water storage. Understanding the lake storage variablity is critical for securing freshwater supplies. The authors leveraged multiple datasets to produce a global dataset with hundreds of thousands of lakes included, which seems to be an impressive work. However, I have a few major comments on the method and data quality.

**We thank the reviewer for the thoughtful comments and constructive suggestions, which will help us to improve the quality of the manuscript. Below please find our response to reviewer's comments in detail.**

R2C1) Pekel et al. used a global model to classify water and land. The authors depended on the Pekel et al.'s data to track water area changes in each lake locally. It is not clear to me whether the global model is suitable for studying each individual water body, for example, how can lake area change be accurately captured by the global model in each lake?

**The GSWD dataset developed by Pekel et al. (2016) has been widely used to derive surface water area dynamics for individual lake in many studies (for example, Zhao and Gao (2018); Busker et al. (2019)). These previous studies have demonstrated the validity of using the GSWD dataset to study lake and reservoir changes. Pekel et al. (2016) also validated their product. The results showed that the omission errors of water mapping are less than 5%, while the commission errors are less than 1%. In the revised manuscript, we will include these uncertainties from GSWD in the storage validation analysis. Please refer to the response to R2C3. In revision, we will also include comparisons between our lake area change estimation and other published lake area datasets (i.e., Zhao and Gao (2018) and Donchyts et al. (2022)). Please see response to R2C2.**

**Most importantly, this study focuses on estimating both historical and near real-time (NRT) lake water storage dynamics. NRT lake water storage time series were only estimated if A-H-V relationships were significantly correlated, although we produced historical water storage change for more than 170,000 lakes (Fig. R5). This also means we make sure that satellite-derived extent and level observations are consistent in characterising lake change before producing storage estimates. We will include a paragraph to highlight this point in the revised manuscript:**

> *"This study mainly focuses on estimating both historical and NRT lake water storage dynamics. NRT lake water storage time series were only estimated if A-H-V relationships were significantly correlated, although we produced historical water storage change for more than 170,000 lakes. This also means we make sure that satellite-derived extent and level observations are consistent in characterising lake change before producing lake storage time series from 1984 to present.*
>
> *As Landsat, Senitnel-2 and ICESat-2 used here have comprehensive coverage of global lakes, we investigated if we can use any two of them (optical + altimetry or two optical) to monitor global lakes and for how much of the lakes in each basin storage can be measured by remote sensing. This should provide valuable information for the newly launched Surface Water and Ocean Topography (SWOT) mission that monitors storage changes in global lakes as it achieves to measure both extent and level on a single satellite platform. If remote sensing measurements from two sources (derived extent and level or*

*derived storages) are significantly correlated for a particular lake, we considered that this lake can be monitored by Earth observation. We followed three complimentary approaches: (1) geo-statistical models (Landsat + Sentinel-2), (2) H-V relationships (Landsat + ICESat-2), (3) extent and level observations (Sentinel-2 + ICESat-2). Approach (3) can be considered the most reliable as storages were directly measured by extent and level from remote sensing. However, approaches (1) and (2) make it possible to estimate absolute storage changes. Landsat and ICESat-2 together are able to measure lake water storage in nearly all (i.e., 234) river basins worldwide (Fig. R5b). Satellite-derived extents and levels are significantly correlated for over one fourth of lakes in each of the 145 basins. This feature is evenly distributed across the continents except for Antarctica and northern high latitude regions, due to the influence of frozen water surfaces. Sentinel-2 and ICESat-2 cover 122 basins globally (Fig. R5d). There are 58 basins where over half of lakes can be monitored by them, mainly located in the USA, southeastern South America, the Mediterranean, southern Africa, southern Asia, and Australia. Landsat and Sentinel-2 both measure surface water extent and they show consistency in 63 out of 124 basins. Three-quarters of lakes has a significant H-A relationship (Fig. R5c) with a distribution pattern similar to that of Sentinel-2/ICESat-2."*

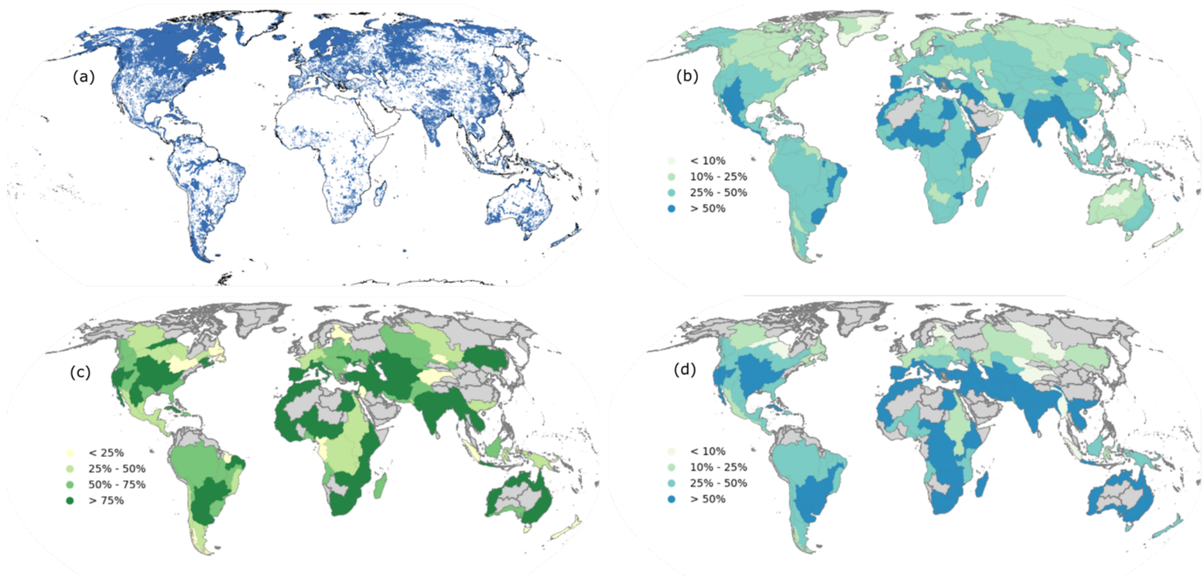


**Figure R5** The locations of 170,957 lakes whose storage dynamics for the period of 1984-2020 were estimated in this study (a), the percentages of lakes (in terms of number) whose NRT water storage dynamics can be derived using Landsat and ICESat-2 (b), Landsat and Sentinel-2 (c), and Sentinel-2 and ICESat-2 (d) in each basin.

**[1] Donchyts, G., Winsemius, H., Baart, F., Dahm, R., Schellekens, J., Gorelick, N., Iceland, C., & Schmeier, S. (2022). High-resolution surface water dynamics in Earth's small and medium-sized reservoirs. Scientific reports, 12(1), 1-13.**

R2C2) The monthly lake areas were generated from recovering water areas from contaminated images as in a previous publication by the authors (Hou et al., 2022). This is not new as quite a few recent studies have done a similar thing. As monthly lake areas are critical for generating monthly storage given monthly level data is pretty rare, the uncertainty of the recovered water areas seems to have non-negligible impact on the derived storage change. Had the authors assessed the uncertainty of areas from contaminated images? How did the generated time series compare with other existing approaches?

**We thank the reviewer for this suggestion. In revising the manuscript, we will include comparisons between our lake area change estimates and other published lake area datasets (i.e., Zhao and Gao (2018) and Donchyts et al. (2022)) (Fig. R7) as below:**

> *"Our lake area data have 5318 lakes in common with the data of Zhao and Gao (2018), and 11101 lakes with Donchyts et al. (2022). Zhao and Gao (2018) used the same Landsat data source (GSWD) and the same lake boundary delineation (GRanD included in HydroLAKES) but a different gap-filling approach to derive lake area. Nonetheless, the mean lake area between the two products scatters closely around a 1:1 relationship (Fig. R7a). In addition, we calculated correlation and bias in lake area time series for the common period 1984–2018 for each lake. The median R and SMAPE were 0.91 and 3.6%, respectively. This suggests that our gap-filling algorithm produces results that are overall similar to those of Zhao and Gao (2018). Donchyts et al. (2022) used a different Landsat data source, lake boundary delineation, and gap-filling algorithm to derive lake area time series. Despite these differences, mean lake area values still lies fairly closely around the 1:1 l relationship, especially for lakes greater than 1 km$^2$ (Fig. R7b). Larger biases exist for some lakes smaller than 1 km$^2$. This was caused mainly by different definitions of lake boundaries between HydroLAKES and Donchyts et al. (2022); we found only one-tenth of lakes had boundary area differences within 20%. The median R and SMAPE in lake area time series from 1984 to 2020 for 11101 lakes between our product and Donchyts et al. (2022) were 0.76 and 9.7%., once again lower mainly due to the differences for small lakes."*

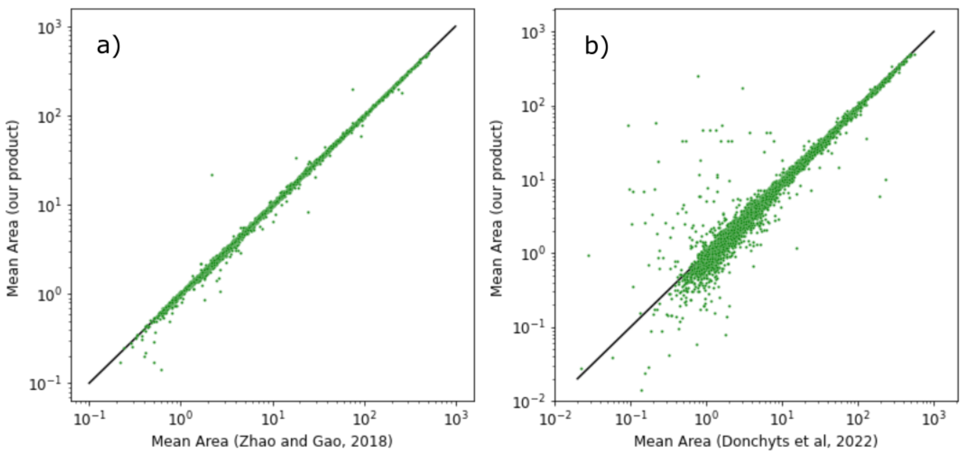


**Figure R7** Scatterplots of mean lake area from our product vs. two other published datasets (black line: 1:1 relationship).

R2C3) I do not believe the Geo-statistical model used by Messager et al., 2016 to predict the total volume of a lake can be used to derive actual lake bathymetry here. The Geo-statistical model was based on global DEM products which have an uncertainty of several meters on average. Additionally, the water level conditions at the DEM acquisition time vary. As the used DEM data in Messager et al., 2016 cannot retrieve the true land surface elevation underneath water, I think this would introduce an even larger uncertainty (e.g., dozens of meters) when the authors extrapolated water levels beneath the level at the DEM acquisition date.

**We agree that lake volume cannot be estimated very precisely unless detailed lake bathymetry data is available, but this is beyond current satellite remote sensing abilities. We emphasized this point in the manuscript. However, we respectfully disagree that the approach used in this study is affected by water**

**level condition at DEM acquisition time, as the approach is different from the traditional DEM approach. The traditional DEM approach indeed is affected by water level condition at DEM acquisition time as it used DEM within the lake to derive A-H curves. In comparison, the model used here considers DEM outside lake (i.e., extrapolating slope around the lake towards the centre of the lake to estimate lake depth) rather than inside, therefore it is not affected by water inundation at the DEM acquisition time.**

**We did compare different approaches to estimate lake depth before choosing the geo-statistical model, but did not include the full analysis in this manuscript. Khazaei et al. (2022) develop a statistical model relating lake depth with surface water area, elevation, volume, shoreline length, and watershed area. They demonstrated that the performance of this statistical model is better than the previous approaches to calculate lake depth by simply assuming lake shape in four geometries: box, cone, triangular prism, and ellipsoid (e.g., Yigzaw et al. 2018). In our study, we compared this statistical model (Khazaei et al., 2022) with the geostatistical model (Messager et al., 2016) and compared them against in situ data. The resulting SMAPE error for the GLOBahty method was 70.5%, and therefore not better than the geo-statistical method. This led our choice to use the geo-statistical method in favour of the GLOBathy dataset in this study.**

**[2] Yigzaw, W., Li, H. Y., Demissie, Y., Hejazi, M. I., Leung, L. R., Voisin, N., & Payn, R. (2018). A new global storage-area-depth data set for Modeling reservoirs in land surface and earth system models. Water Resources Research, 54(12), 10-372.**

**The geo-statistical model used in this study has uncertainties sources from DEM and observed lake area as demonstrated in Equation R1. The GSWD data used in this study to derive lake area has the omission and commission errors of 5% and 1%, respectively. The DEM used in geo-statistical model to calculate slope has an error of around 10 m. We have considered all these uncertainties in our storage validation results. In the revised manuscript, we will include a new paragraph (please see below) to better describe the geo-statistical model and include uncertainties analysis (Fig. R2).**

*"The geo-statistical model (Messager et al., 2016) used in this study is:*

$$Log_{10}(D) = C_1 + C_2 \times Log_{10}(A) + C_3 \times Log_{10}(S_{100}) + s^2 \qquad (R1)$$

*where D is the predicted mean depth (m), A the observed surface area of the lake (km$^2$), $S_{100}$ the average slope (derived from DEM) within a 100-m buffer around the lake, $C_1$, $C_2$ and $C_3$ constant parameters estimated from best fitting the model using global lake data and present for different sizes of lakes (i.e., 0.1-1 km$^2$, 1-10 km$^2$, 10-100 km$^2$, 100-500 km$^2$), and $s^2$ the residual variance. The fundamental assumption is that one can extrapolate slope around the lake towards the centre of the lake to estimate lake depth. The traditional approach uses a DEM within the lake to derive A-H curves, but cannot retrieve true land surface elevation below water. In comparison, the model used here considers the DEM outside lake rather than inside, and hence is not affected by water inundation at the time of DEM acquisition. Messager et al. (2016) reported that the symmetric mean absolute percent error between predicted and reference volume is 48.8% without significant bias in volumes for the majority of lakes around the world, with the exception of Finland, Sweden and northwestern Russia, the European Alps, and the Andes. The uncertainties from the geo-statistical model are mainly from observed surface area* *and DEM used to derive slope. The omission and commission errors of surface water mapping from GSWD used in this study are 5% and 1%, respectively. EarthEnv-DEM90 was used in the geo-statistical model to derive slope, and its vertical accuracy is around 10 m (Robinson et al., 2014). We assessed how these errors propagate into water storage estimation in the validation analysis."*

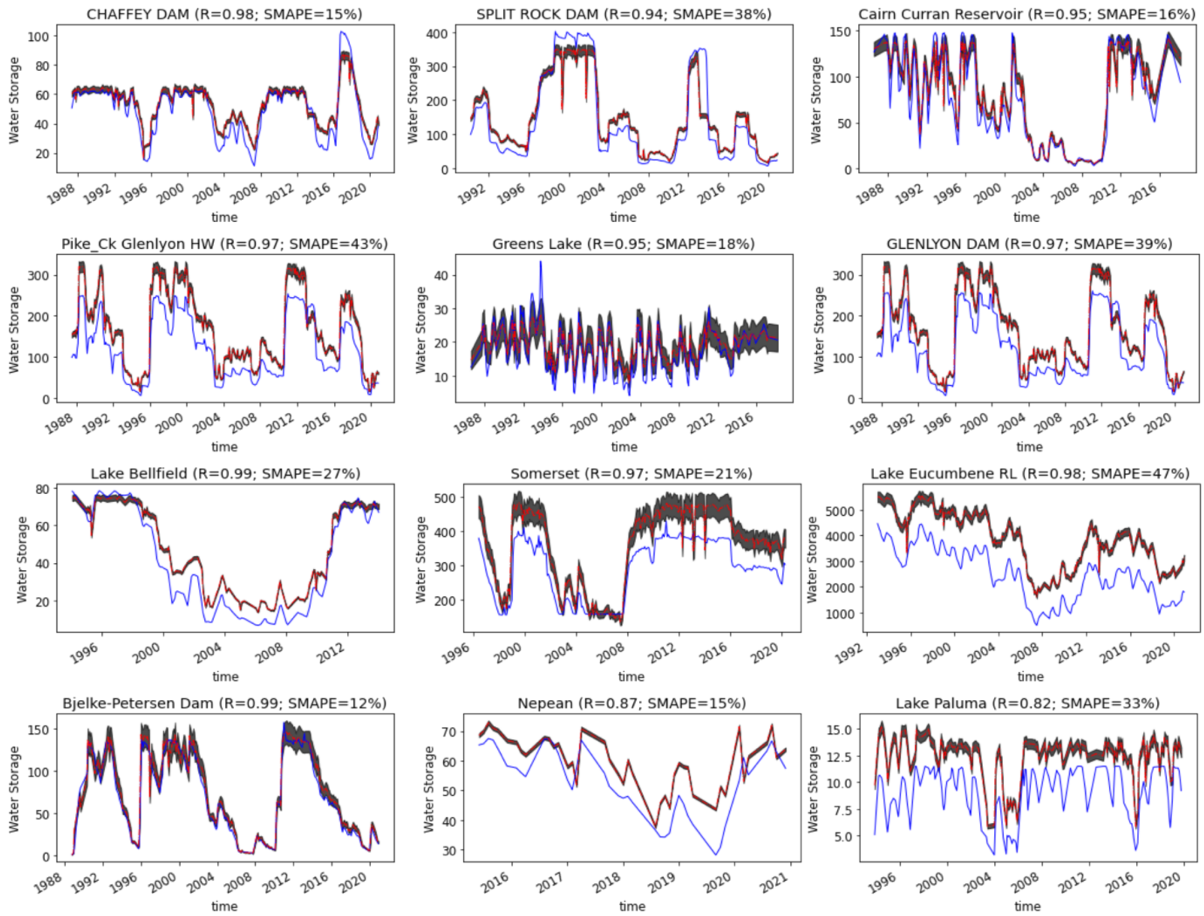


**Figure R2** Some comparisons of historical lake water storage time series from GloLakes against in situ (observed) data (blue line: observed data; red line: historical storage estimates; black shade: error bars of historical storage estimates).

R2C4) The validation appears to be insufficient. How did the authors select the 238 lakes for the validation and why this is a comprehensive evaluation? Did the authors consider the performance of the method on cold-region lakes (e.g., in Canada). What the accuracy for smaller lakes given this method is only significant on small lakes as existing studies did a fairly good estimate on large lakes. Why the authors use relative metric R given R only gives a correlation estimate? For example, if the storage was scaled from area, no matter how large the level error is, the R value remains the same. Without a comprehensive validation, it is hard to foresee that the produced datasets would be useful for scientific inquiries.

**Thank you for this comment. In the revised manuscript, we will add several more validations analyses:**

**(1) We extended our water storage validation analysis from 238 lakes to 494 lakes (now including in situ data from USA, Australia, South Africa, India, and Spain), and show correlation and bias results for all 494 lakes.**

**(2) We performed uncertainty analysis of the geo-statistical model in the water storage validation analysis for 21 lakes as examples. Uncertainties sources include observed lake area (the omission and commission errors of surface water mapping from GSWD used in this study are 5% and 1%, respectively; Pekel et al. (2016)) and DEM (±10 m; Robinson et al. (2014)).**

**(3) We compared lake mean volume from our product against in situ data for 494 lakes to examine any systematic bias in our data.**

**(4) We validated near real-time (NRT) storage estimates for 9 lakes where in situ data include NRT information and analysed the change in the performance of storage estimation between NRT and historical estimates.**

**(5) We cross-validated NRT storage estimates between our different approaches and show the performance of storage estimation for lakes with different sizes.**

**(6) We compared lake water area time series from our product against Zhao and Gao (2018) and Donchyts et al. (2022) for 5318 and 11101 lakes, respectively, in terms of correlation and bias, and showed the comparisons of lake mean area between our product and these two published studies in the 1:1 relationship to examine the uncertainties of our derived lake area time series.**

**(7) We provided a summary as to where and for how many lakes NRT storage can be estimated by remote sensing in each basin around the world. This will provide key information for the newly launched Surface Water and Ocean Topography (SWOT) to monitor storage changes in global lakes as it will be able to measure both extent and level on a single satellite platform.**

**(8) We show examples of comparisons between our relative storage product and Tortini et al. (2020) and correlation results for all evaluated lakes.**

**Unfortunately, we could not find any in situ data in Canada or other cold regions, although we have extended our storage validation results to cover more regions around the world. However, we assessed where and how much of lakes in each basin whose storage can be measured by remote sensing. The results showed that only a modest percentage of lakes can be monitored by our approach in northern high latitude regions due to the influence of frozen water surfaces in the cold season (see response to R2C1).**

**We assessed the accuracy of our products for lakes of a wide range of sizes. The evaluations not only include validation of lake area estimates (see response to R2C2), but also lake storage (see Fig. R3 below). The lakes for which area estimates were validated had areas from 0.1 km$^2$ to 1000 km$^2$ while those for which storage was validated had volumes from 1 GL to 10000 GL.**

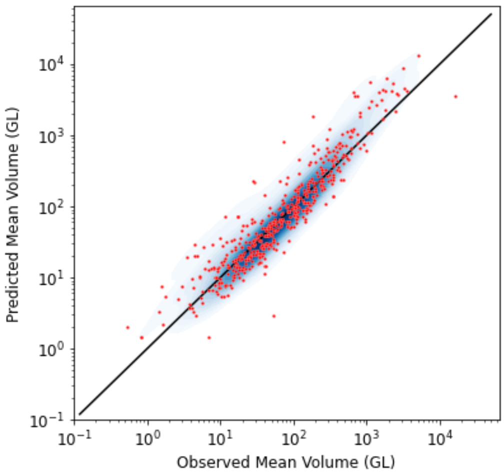


**Figure R3** Scatterplots of predicted mean lake volume vs. observed mean volume (red dot: mean lake volume; black line: 1:1 relationship; blue density: background blue colours indicate data density).

**We calculated not only R (correlation) but also SMAPE (bias) to assess the accuracy of our product. Storage was not just scaled from area, in fact, monitoring approaches include three types: (1) geo-statistical model (e.g., Landsat + Sentinel-2); (2) H-V relationships (e.g., Landsat + ICESat-2); (3) direct extent and level (e.g., Sentinel-2 + ICESat-2). We will clarify this in the revised manuscript.**

Specific comments:

R2C5) Line 64: I would suggest replacing "a larger number" with the actual number.

**Thank you. The actual number was 227,386. We will change this in the revised manuscript.**

R2C6) Line 86: It seems that Avisse et al. is not the only study on estimating lake storage change based on DEM data. Maybe also highlight other relevant studies here.

**Thank you. We will include the additional studies below in the revised manuscript.**

**[2] Bonnema, M., Sikder, S., Miao, Y., Chen, X., Hossain, F., Ara Pervin, I., Mahbubur Rahman S.M., & Lee, H. (2016). Understanding satellite-based monthly-to-seasonal reservoir outflow estimation as a function of hydrologic controls. Water Resources Research, 52(5), 4095-4115.**

**[3] Vu, D. T., Dang, T. D., Galelli, S., & Hossain, F. (2022). Satellite observations reveal 13 years of reservoir filling strategies, operating rules, and hydrological alterations in the Upper Mekong River basin. Hydrology and Earth System Sciences, 26(9), 2345-2364.**

R2C7) Line 280: the authors only show a few case studies for the validation. It would be better to have a figure or table to show all data used for the validation.

**Thank you, agreed. We will show R and SMAPE validation results for all 494 lakes (Fig. R4) in the revised manuscript.**

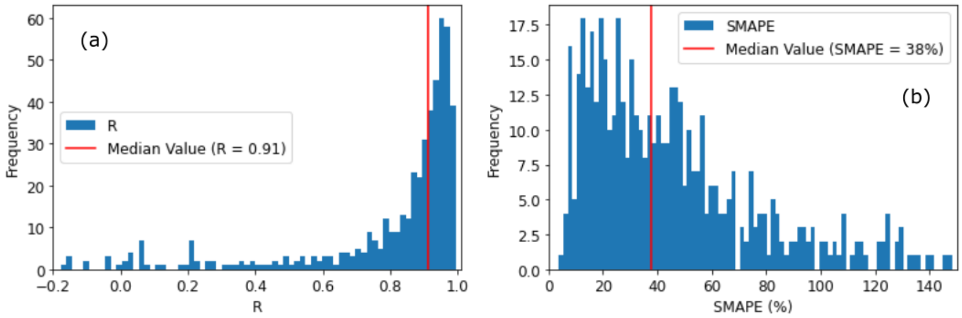


**Figure R4** The distribution of R and SMAPE of water storage validation results for 494 lakes (vertical red line: the median value).

.