<u>Response to Reviewer #1 Comments:</u>

Hou et al present a dataset of global water storage variations. This paper is unique in that it attempts to construct absolute storage variability time series, which are challenging to produce. It also aims to fuse together multiple freely available datasets in a novel approach. However, I find it to overall be a flawed manuscript and dataset which needs many necessary improvements (see major comments below). In brief, I am concerned that the authors have not accurately described the dataset, both in regards to the long term and NRT storage dynamics and have not performed sufficient validation analyses. The paper is also poorly written in places, with numerous typos and grammatical errors as well as paragraphs that are poorly structured, and the figures are weak and do not sufficiently illustrate the dataset. Without substantial changes to the manuscript, presentation and perhaps the dataset itself, I'm not sure this paper and dataset would be of value to the broader community.

We thank the reviewer for the detailed and valuable comments and suggestions, which will enable us to greatly improve the quality of our manuscript. Below please find our response to reviewer's comments in detail.

In the revised manuscript, we will carefully and accurately distinguish between "estimate" and "measure" to describe the different approaches to derive lake water storage time series (in response to comment R1C1). We will also add the results of several more validations analyses we have undertaken to strengthen the manuscript:

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.

We apologise for any typos and grammatical errors and will carefully examine and correct them. We will also add several more figures (please see the details in the response to the comments below) to better illustrate our product. Finally, we will restructure our Results and Discussion Section to have the following sections:

(1) Lake area estimation validation

(2) Where NRT lake monitoring is feasible

(3) Validation of historical lake storage estimates

(4) Validation of NRT lake storage estimates

(5) Future opportunities

Following the review, we will also make some changes to our dataset in the revised manuscript:

(1) We will implement the ICESat-2 quick look data into our GloLakes product to improve monitoring abilities of ICESat-2 based storage estimations

(2) We will release a new product to derive lake storage changes directly using satellite-derived extents (Sentinel-2) and levels (ICESat-2)

(3) The GREALM has now included Sentinel-6A data. This will be included our product as well

Major Comments

R1C1) Estimation of long term storage dynamics. Throughout the text, the authors state that they 'measured' long term absolute storage dynamics and/or 'produced' absolute storage time series. To be clear, what the authors did was apply a geostatistical model to estimate water depth and then use this to estimate a absolute storage time series when combined with a Landsat-derived dataset of lake extent. There was thus no 'measurement of storage' here – what the authors did was 'estimate' storage based on statistical relationships. While there is nothing wrong with estimating using these geostatistical relationships, it is imperative that this is explained correctly and consistently throughout the paper so as not to cause confusion with other methods which actually calculate volume change based on water level observations.

Agreed. We will carefully revise the manuscript and make sure that we consistently use "estimate" where a geostatistical model was used while retaining "measure" where storage is calculated from water extent and level observations, to indicate the lesser uncertainty in the latter.

R1C2) ICESat-2 data is not NRT. The authors state the importance of Near Real Time (NRT) lake monitoring with a latency of ~1-10 days. They then include ICESat-2 as one of the potential datasets to use for NRT monitoring. However, this indicates a fundamental misunderstanding of ICESat-2 and how it is processed. First of all, ICESat-2 has a repeat time of 91 days, so on average you get an observation ~once every three months (though this does vary based on the size of the lake). Second of all, unlike with say MODIS, Landsat, or Sentinel-2, ICESat-2 data is not immediately released. Currently (as of Oct 9, 2022), the most recently available ICESat-2 data is through June 8th, 2022, and this has been fairly consistent over the past few years (ICESat-2 releases data about every ~6 months). While the NSIDC ICESat-2 website is perhaps a little misleading that it says data is available up to the present, so I understand some of the confusion, simple playing around with the data will quickly reveal the extremely long latency of ICESat-2 products. It is thus very much inaccurate to use ICESat-2 as a potential NRT water volume estimator in the method described here.

We agree that ICESat-2 cannot provide sufficient near real-time (NRT) information if we used the "ATLAS/ICESat-2 L3A Along Track Inland Surface Water Data, Version 5 (ATL13)" product because their data were not updated regularly when new observations obtained. In using the term NRT, we referred to the time between the ICESat-2 data becoming available and our use of it to produce storage estimates, but we appreciate this can be a misleading use of words. We also thank the reviewer for referring us to the "ATLAS/ICESat-2 L3A Along Track Inland Surface Water Data Quick Look, Version 5 (ATL13QL)" product which can provide NRT water height observations (R1C21). We will implement the ICESat-2 quick look data into our GloLakes product to improve monitoring abilities of ICESat-2 based storage estimations and make changes in the revised manuscript accordingly. We will include how we process ATL13 and ATL13QL data and combine them together to derive NRT water height observations and discuss the advantages and disadvantages of using these two products in the revised manuscript.

We were aware that ICESat-2 has a low temporal frequency of around 91 days. However, this kind of temporal frequency can still provide updated and useful seasonal variations of global lakes. In revising, we will clarify that NRT monitoring for ICESat-2 based storage estimations has latency to update each of lake storage but that we produce a rapid update global lake data *once* there are new observations.

Another important aspect of using ICESat-2 in this study was to investigate where lake storages can be measured by satellite-derived extent and level simultaneously, as it has very dense coverage of global lakes, thus overlapping a large number of lakes with Landsat observations. These will provide key information for the newly launched Surface Water and Ocean Topography (SWOT) to monitor storage changes in global lakes as it achieves to measure both extent and level on a single satellite platform. Please see detailed response to R1C5.

Based on all above, we will modify the paragraph on ICESat-2 data description in the revised manuscript:

> *"ATLAS/ICESat-2 L3A along-track inland surface water data, version 5 (ATL13) was used in this study (Jasinski and Ondrusek, 2021). The updated ATL13 product is not released until 30-45 days after new observations were obtained, which does not fit the purpose for the NRT monitoring. Therefore, in addition, we used ATLAS/ICESat-2 L3A Along Track Inland Surface Water Data Quick Look, Version 5 (ATL13QL), which is available within 3 days of new observations. ATL13 and ATL13QL apply the same*

algorithms to derive surface water height measurements for inland water bodies including rivers, lakes, reservoirs and coastal water. The approach is described by Jasinski and Ondrusek (2021), but in brief: 1) identifies ATLAS beams that intersect inland water body shape masks; 2) collects photons in short segments (~100 m) for calculating water height and in longer segments (1~3 km) for estimating and removing subsurface backscatter; 3) applies the physical and statistical modelling to derive inland water heights. Cooley et al. (2021) used ATLAS/ICESat-2 L3A Land and Vegetation Height (ATL08) to derive lake water height time series and found the mean absolute error between USGS gauge data and ICESat-2 height measurements is 0.14 m. Unlike ATL08, ATL13 used in this study was designed to measure water height variations on rivers and lakes as it considers physical processes of light propagation in open water bodies. An error of 6.1 cm per 100 inland water photons has been reported (Jasinski and Ondrusek, 2021). We collected both ATL13 and ATL13QL water surface height data from any of the six beams within individual lake boundaries from HydroLAKES and derived water height time series for each lake that observed by ICESat-2 from 2018 to present. ATL13QL has larger uncertainties (~100 m) in geolocation than ATL13 (~5 m). As a result, segment heights from ATL13QL are 2.7 m, with a standard deviation of ~ 7 m, lower than those from ATL13. According to the user guide (Jasinski and Ondrusek, 2021), 2.7 m can be added to ATL13QL before merging these two products and the differences between them have little impact on measuring relative heights. Each lake will be revisited by ICESat-2 around every 91 days. This means our ICESat-2 based storage product is limited to provide updated seasonal variations, rather than short-term changes, for global lakes. However, it provides updated data where there are new observations with a rapid turnaround."

R1C3) No validation of NRT data. While the authors do perform validation for 238 lakes for what appears to be the geostatistically estimated historical time series, it is not explicitly stated whether they perform any validation of the NRT time series (both the absolute and relative). More detailed information on the accuracy of the NRT time series, and how it varies between using V-H vs. V-A relationships (or how it varies by lake size, if possible), is required to be able to evaluate this dataset.

We thank the reviewer for this comment. It is a challenge to find publicly accessible in situ data especially with NRT information. Of course, this is one of the reasons we are using remote sensing in the first place. We would like to first emphasize that NRT storages were only estimated if V-H or V-A relationships were significantly correlated (please see the detailed response to R1C5). After a search, we found NRT data for nine sites and used these to validate our NRT product. In addition, we performed cross-validation between NRT estimates from our different storage products (ICESat-2, Sentinel-2, and GREALM), and summarized results for lakes with different sizes. We will include these new validation results and a new paragraph in the revised manuscript as follows:

"We validated our ICESat-2 derived and Sentinel-2 derived water storage time series from 1984 to present against in situ data that contains both historical and NRT observations (Fig. R1). We compared the performance of water storage estimates between historical and NRT periods to investigate if NRT estimation based on V-H or V-A relationships is valid. The average R and SMAPE in the historical period (1984-2020) if storage was estimated using a geostatistical model were 0.94 and 22%. The

performance for the NRT period (2020-present) when storage was estimated using a V-H or V-A relationships decreased slightly (R=0.88 and SMAPE=31%). For most lakes the agreement between predicted and observed storages in NRT period shows similar accuracy (in terms of correlation and bias, Table R1) as NRT storages in our products were only estimated if V-H or V-A relationships were significantly correlated. For example, both ICESat-2 and Sentinel-2 derived products record sharply decreasing in water storage in Big Sandy Reservoir, Vega Reservoir, and El Vado Lake in the NRT period (Fig. R1). However, the performance for the NRT period statistically decreases for Lake Summer Horsetooth Reservoir and El Vado Lake as there is one erroneous observation (Fig. R1).

We compared NRT estimates from our different storage products (ICESat-2, Sentinel-2, and GREALM), and summarized results for lakes with different sizes (Table R2). The mean R and SMAPE between ICESat-2 and Sentinel-2 products was 0.77 and 12%, respectively. The ICESat-2 and GREALM products had a similar correlation (R=0.75) but less bias (SMAPE=6%) as they are both derived from altimetry. The evaluation results generally did not vary as a function of lake size except for the fact that ICESat-2 and Sentinel-2 products show better correlation (R=0.84) and less bias (SMAPE=5%) for lakes with an area greater than 100 km^2*."*

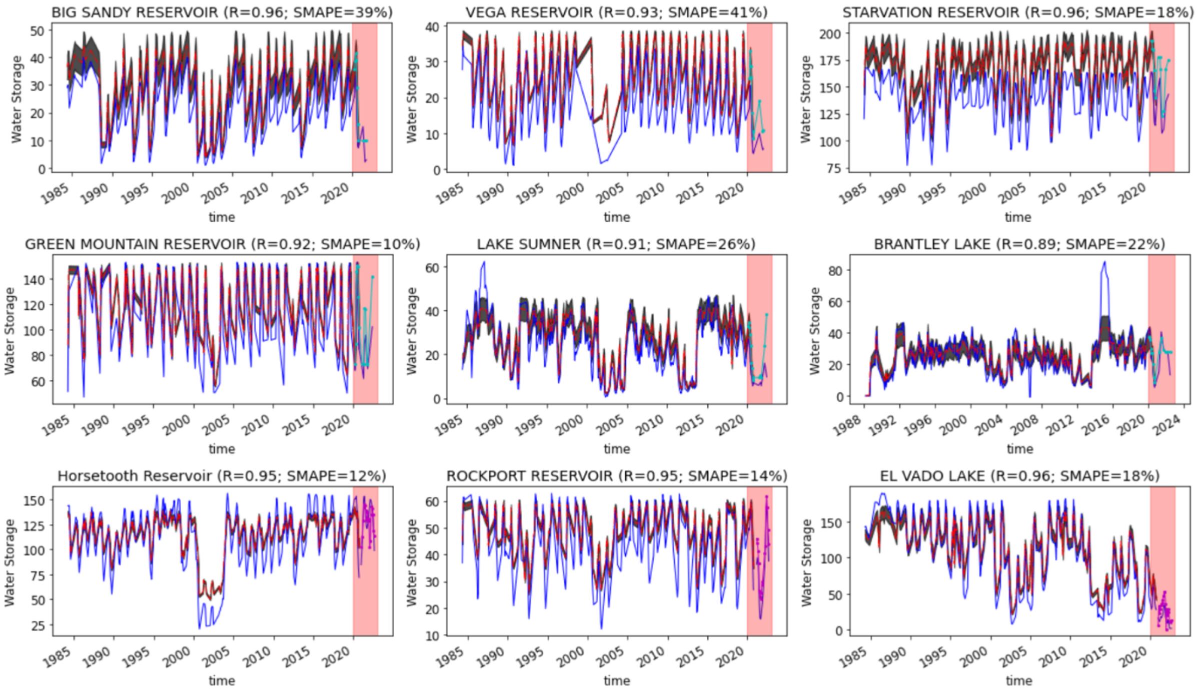

Figure R1 Comparisons of historical and NRT lake water storage time series from GloLakes against in situ (observed) data (blue line: observed data; red line: historical storage estimates; black shade: error bars of historical storage estimates; light pink shade: NRT monitoring period; cyan line with dot (first two rows): NRT storage estimates from ICESat-2; magenta line with dot (third row): NRT storage estimates from Sentinel-2).

Table R1 Comparisons of correlations and biases between historical and NRT estimates (results from Fig. R1).

Lake Name	NRT Monitoring Approach	Historical (1984-2020)		NRT (2020-present)	
		R	SMAPE	R	SMAPE
Big Sandy	L-H; ICESat-2	0.96	39%	0.94	53%
Vega		0.93	41%	0.98	52%
Starvation		0.96	18%	0.97	18%
Green Mountain		0.92	10%	0.93	13%
Sumner		0.91	26%	0.81	43%
Brantley		0.89	22%	0.89	21%
Horsetooth	L-A; Sentinel-2	0.95	12%	0.56	8%
Rockport		0.95	14%	0.99	19%
El Vado		0.96	18%	0.81	48%

Table R2 Cross-validations in NRT estimates between our different storage products for lakes with different sizes.

	Landsat + ICESat-2							
	R (mean)				SMAPE (mean)			
	1-10 km^2	10-100 km^2	>100 km^2	All	1-10 km^2	10-100 km^2	>100 km^2	All
Landsat + Sentinel-2	0.77	0.76	0.84	0.77	12%	12%	5%	12%
Landsat + GREALM		0.77	0.74	0.75		9%	6%	6%

R1C4) Need for large error bars and more details about the approach. The paper glosses over the specifics of the statistical method (pointing towards another paper) used to contruct the historical time series, but more details on this method should be provided to enable the reader to better understand the method and its potential accuracy. The dataset should also include significant error bars on the resulting time series, or at least clear information that these are all estimates with on average ~50% error. This is even more important for the NRT estimations, as given the use of look-up tables to approximate V-H/V-A relationships, these should have even greater error than the original volume time series (see comment about need for validation above)!

Thank you for these suggestions. We will include error bars considering the uncertainties from the geo-statistical model (Fig. R1 and R2). We will add the following sentences in Section 2.1.3 (Geo-statistical model) to better explain the method and how we assess its uncertainties:

"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."*

We extended our validation analysis from 238 lakes to 494 lakes. In the revised manuscript, we will have water storage validation results across USA, Australia, South Africa, India, and Spain. The additional in-situ water storage data were found in a recent paper by Donchyts et al. (2022). We will compare predicted mean water storage against observed data (compared to a 1:1 relationship) for 494 lakes to illustrate a ~50% error (Fig. R3). We will also display all validations results (R and SMAPE) for 496 lakes (Fig. R4), to help readers better understand the accuracy of our product.

[1] Robinson, N., Regetz, J., & Guralnick, R. P. (2014). EarthEnv-DEM90: A nearly-global, void-free, multi-scale smoothed, 90m digital elevation model from fused ASTER and SRTM data. ISPRS Journal of Photogrammetry and Remote Sensing, 87, 57-67.

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

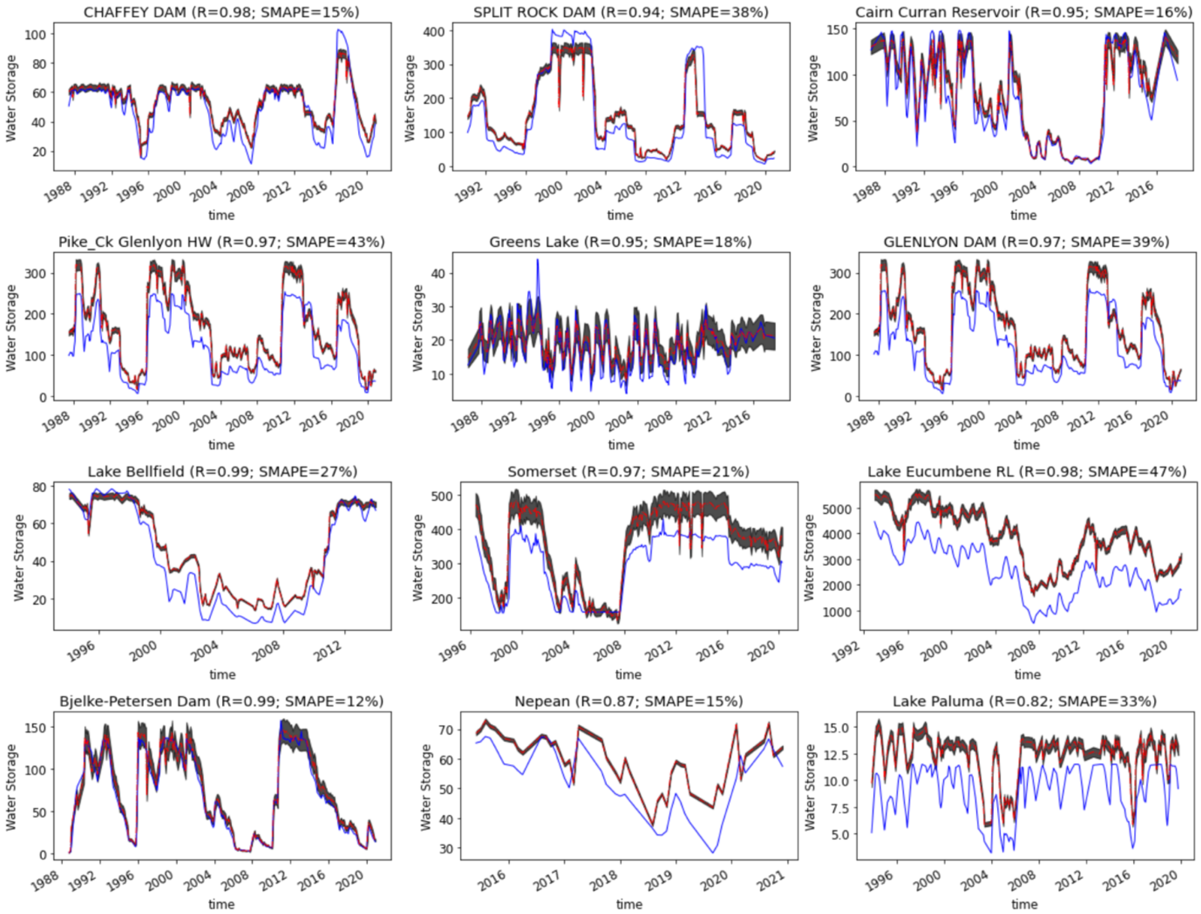

Figure R2 Some comparisons of historical lake water storage time series from GloLakes against in situ (observed) data (blue line: observed data; red line: historical storage estimates; black shade: error bars of historical storage estimates).

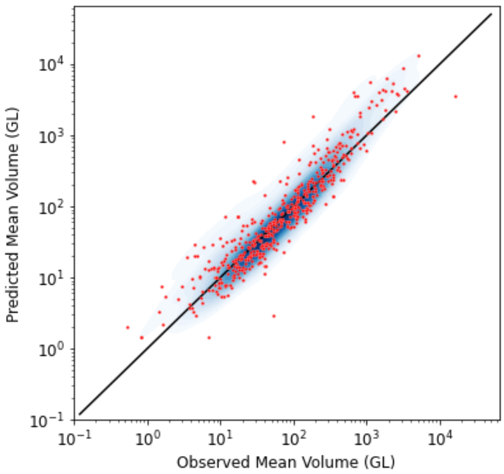

Figure R3 Scatterplots of predicted mean lake volume vs. observed mean volume (red dot: mean lake volume; black line: 1:1 relationship; background blue colours indicate data density).

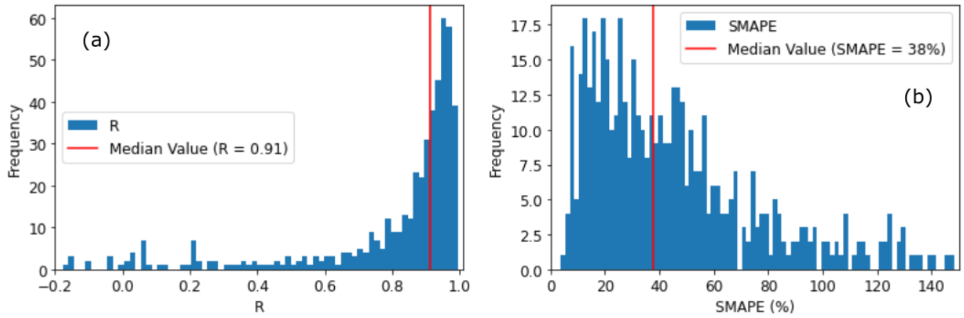

Figure R4 The distribution of R and SMAPE of water storage validation results for 494 lakes (vertical red line: the median value).

R1C5) Poor quality figures. The figures included are weak and do not provide sufficient info about the dataset. Figure 3 is very difficult to interpret (what is the difference between panel (b) and panel (c))? Figures 1 and 2 are useful but are very large and all 4 examples for each may not be needed. Figure 4 is useful and depicts the accuracy analysis well. I'd suggest adding additional figures illustrating the accuracy of the geostatistical estimation of water depth (and perhaps the resulting volume time series). It also would be useful to provide more information about where is was possible to build NRT and relative time series, and where it was not possible (perhaps in a figure? Or table?).

We apologise for the confusion. Fig. 3b and 3c showed the percentage of total basin water volume for which water storage dynamics were estimated in this study. The percentage is calculated in each basin by Equation R2:

$$p = \frac{\sum_{i=0}^{m} V_i}{\sum_{i=0}^{n} V_i} \quad \text{(R2)}$$

where *p* is the percentage of total lake water volume in each basin, *m* the number of lakes for which water storage dynamics were estimated in a basin, *n* the total number of lakes from HydroLAKES in a basin, and V_i the water volume of individual lake from HydroLAKES. The difference between these two figures is that Fig. 3b counted the lakes whose historical time series were estimated while Fig. 3c calculated those whose NRT dynamics can be monitored.

In the revised manuscript, we will move Fig. 3b and 3c to the Supplementary Materials and add Fig. R5b-d to emphasize one of highlights in this study. To make it easier to interpreted, we calculated the percentage of total number of lakes rather than total volume explained above in the new Fig.R5b-d. Fig. R5b-d shows where and for how much of the lakes NRT storages can be measured by remote sensing in each basin around the world. These will provide valuable information for the newly launched Surface Water and Ocean Topography (SWOT) to monitor storage changes in global lakes as it achieves to measure both extent and level on a single satellite platform. We will include a paragraph to explain this analysis in the revised manuscript:

> *"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 A-H relationship (Fig. R5c) with a distribution pattern similar to that of Sentinel-2/ICESat-2."

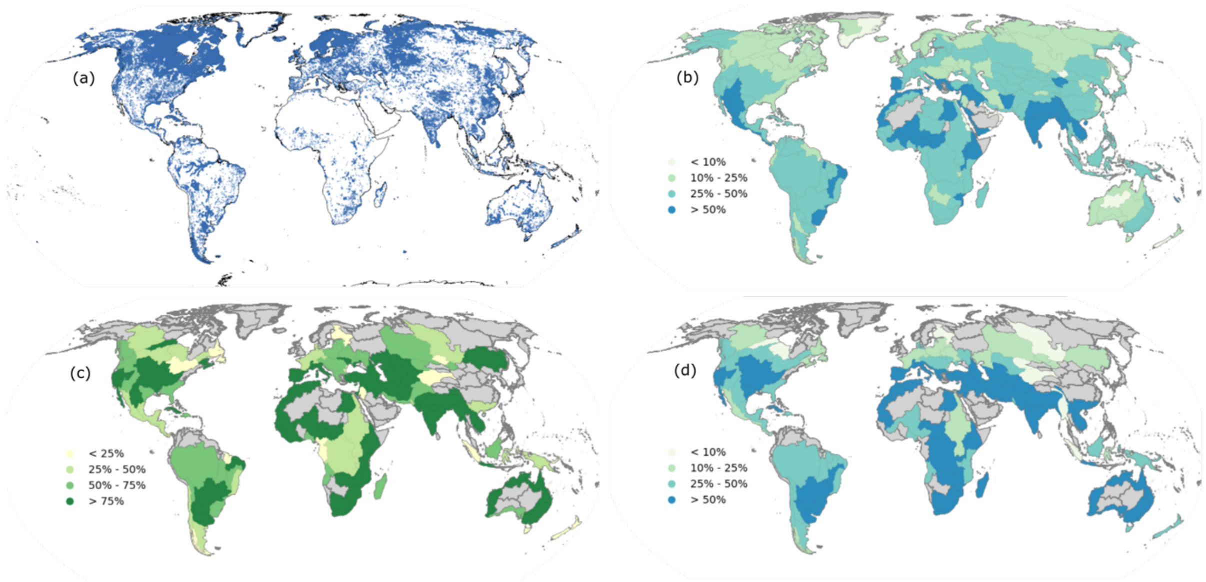

Figure R5 The locations of 170,957 lakes whose storage dynamics for the period of 1984-2020 were estimated in this study (a), the percentages of lakes (in terms of number) whose NRT water storage dynamics can be derived using Landsat and ICESat-2 (b), Landsat and Sentinel-2 (c), and Sentinel-2 and ICESat-2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

On the other hand, Fig.1 and Fig.2 showed how our gap-filling algorithm addressed missing Landsat data in different scenarios (cloud cover, SLC failure, mix issue, different contamination ratios, different shapes of lake). Following the reviewer's comment, we will clarify this and move Fig. 2 to the Supplemental Material.

R1C6) Why not include Landsat and Sentinel-2 (not BLUEDOT) as NRT? Given that the authors calculate NRT storage variations simply by building V-H or V-A relationships with the results of their Landsat and geostatistical depth model volume time series, it should be possible to construct NRT time series for thousands more lakes globally by just classifying water in Landsat-8/9 and Sentinel-2 (and this would actually be NRT, unlike ICESat-2) While I understand that part of the point of this paper is to use existing datasets, classifying water in Landsat-8/9 and Sentinel-2 is pretty darn standard and straightforward at this point, particularly with the existence of cloud platforms like GEE. I'm not suggesting the authors do this globally, more just pointing out that this approach (while likely quite inaccurate) could in theory be used for thousands more lakes. Also, why (for the Sentinel-2 data in particular) do you need to use a lookup table to estimate volume via a V-A relationship – surely you could use the same geostatistical model to calculate volume from the Sentinel-2 (or Landsat 8/9) area observation?

Thank you for this suggestion. Regards to GEE, we agree that we could directly use Sentinel-2 to derive water extent for thousands more lakes, but it was indeed well beyond the scope of this study. In the revised manuscript, we will discuss the limitation of using BLUEDOT data and how this might be addressed using Sentinel-2 and GEE in future.

We would also like to clarify that we did in fact use the geostatistical model to estimate water storage for both of Landsat and Sentinel-2 and then merged them together to derive historical and NRT water storage time series. We will modify the relevant sentences to clarify this in the revised manuscript.

Considering that Sentinel-2 and ICESat-2 can measure surface water extent and level, respectively, simultaneously, we produced an additional data set as part of this collection to derive lake storage changes directly rather than using V-H or V-A relationships. Below are some examples of H-A (Sentinel-2 + ICESat-2) relationships for the overlapping period (2016-present).

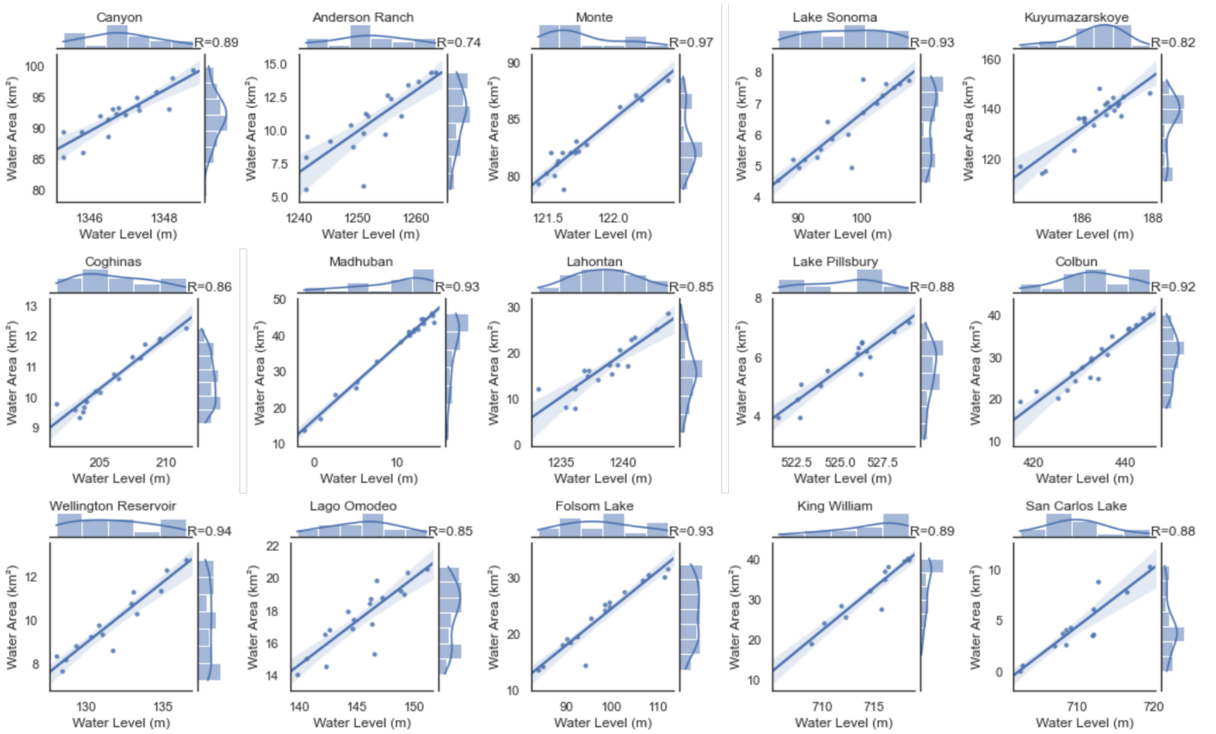

Figure R6 Some examples of lake H-A (Sentinel-2 + ICESat-2) relationships from 2016-present (blue line: fitting relationships; blue shade: 95% confidence interval; top axis: the distribution of water level; right axis: the distribution of water area).

Specific Comments

R1C7) Lines 6-7: The first sentence of the abstract is unnecessarily long and wordy and contain a typo. Please rephrase.

Thank you. We will rephrase this sentence in the revised manuscript.

R1C8) Lines 56-74: Nice review of all of the different global water datasets, but this paragraph could use some restructuring as in its current form, this transition from the statement about ICESat-2 laser altimetry towards stating that "the spatial resolution of global satellite-dervied surface water dynamics projects…" doesn't make much sense, as the paragraph then moves to talking about measurements of surface water extent, not water level. I would suggest reorganizing this paragraph and being clearer about developments in observations of water level vs. water extent.

Thank you. We will rephrase this paragraph in the revised manuscript.

R1C9) Line 85: Add "cannot measure lake depth and therefore are unable to measure absolute water volume without the use of bathymetric data"

Thank you. We will add this in the revised manuscript.

R1C10) Lines 93-105: This paragraph is confusingly worded. The authors state that "Relative storage changes were estimated … while absolute storage changes were…" which is immediately followed by a statement that this was possible for more than 27,000 lakes worldwide. According to my understanding of the paper, the absolutely storage changes estimated using a geostatistical model were possible for all ~170,000 lakes, whereas the relative changes were possible for ~23,000 and NRT volumes (using the statistical model plus V-H or V-A relationships) were estimated for ~27,000. As currently written, this paragraph thus inaccurately describes the results.

Thank you. We will rephrase this paragraph in the revised manuscript.

R1C11) Line 97: Remove "Furthermore"

Thank you. We will remove "Furthermore" in the revised manuscript.

R1C12) Line 153: What is meant by "future GSWD water bodies"?

Apologies for the confusion. We can simply remove "future" in the revised manuscript

R1C13) Line 195: While NSIDC is where the ICESat-2 data is hosted, it is incorrect to call it a monitoring platform. Just call it ICESat-2 data.

Thank you. We will call it ICESat-2 data consistently in the revised manuscript

R1C14) Table 1: It is incorrect to call NSIDC the platform for ICESat-2. I'm not sure exactly what term would be useful here, but I'd suggest simply calling the datasets something like ICESat-2, USDA G-REALM and BLUEDOT (Sentinel-2).

Agreed. We will use ICESat-2, USDA G-REALM and BLUEDOT (Sentinel-2) as their names in the revised manuscript.

R1C15) Figure 3 caption (and throughout the manuscript). It is incorrect to state that "storage dynamics for the period of 1984-2020 were measured in this study". Given that storage was estimated based purely on statistical relationships that are likely to be highly inaccurate in many places (and with an overall error of ~50%), 'estimated' is the only appropriate word to describe this approach.

Thank you. We will carefully use either "estimate" or "measure" to describe different approaches (see response to R1C1).

R1C16) Line 277: What is meant by "Overall, the relative volume dynamics are generally more reliable, as indicated by the correlation values"? Does this refer to the NRT time series? Or specifically the NRT time series calculated from A-H relationships? And correlation with what? The results on this accuracy are not reported in the paper (see major comment above).

Apologies. By this we mean that the validation results (predicted volume vs. observed volume) showed strong correlation so our product should be more reliable to indicate lake volume changes than absolute depth changes. We also produced relative storage products that avoid using the geo-statistical model and focus on providing accurate relative changes in lake storage. In addition, we will add validation of the NRT estimates in the revised manuscript (see response to R1C3).

R1C17) Line 291: Again, please be clear here that these results are 'estimated'.

Agreed, see R1C15.

R1C18) Line 299: It is not necessary to state that Busker's dataset was not publicly available (you can always just email an author and ask for it, as technically every dataset should be publicly available on some level). I suggest removing this part of the sentence.

Agreed. We will remove this.

R1C19) Line 343: SWOT will measure water height every 11 days (its orbit has a return period of 21 days, but since it is an off-nadir satellite, it will measure every water body every 11 days – see https://swot.jpl.nasa.gov/

Thank you. We will change it to "11 days" in the revised manuscript.

R1C20) Line 356: I explored the online Global Water Monitor linked in the paper. This is a cool way of displaying and communicating the data, and I commend the authors for putting it together, though it needs some cleaning up a bit (for example, what is the unit "GL" on the y axis for the annual time series of water volume)? If possible, it also would benefit from including error bars.

Thank you for this suggestion. The Global Water Monitor is still under development and we only included the link at this stage for review purposes. The official version will be released in January 2023. The data quality and visualization will keep improved along the new version of this web application.

Additional Comments:

R1C21) I wanted to lightly addend my review, specifically about the latency of ICESat-2, as I just learned about the existence of ICESat-2 Quick Look, low-latency products (https://nsidc.org/data/user-resources/help-center/faqs-icesat-2-quick-looks). My comments about the limited value of ICESat-2 for NRT data due to its 91 day repeat cycle still apply, but the existence of these quick look products does mean that it is technically possible to use the data within a few days after collection. I apologize for not acknowledging this before.

If the authors are to continue to use ICESat-2 data for their NRT dataset, I would advise explicitly discussing the quick look data as well as its advantages/disadvantages relative to the final product. Similarly, while the other two datasets mentioned in the NRT section (G-REALM and BLUEDOT) are relatively easy to download and process into lake height/area (since they already come processed to individual lake height/area) doing so with ICESat-2 is significantly more complicated as it requires choices around how to aggregate different tracks, filter out poor quality or outlier water level observations, etc. This additional difficulty should be described in the manuscript with additional details on how the authors process the ICESat-2 data automatically.

Thank you for these addition comments, see response to R1C2.