Response to Reviewers

‘‘GloLakes: water storage dynamics for 27,000 lakes globally from 1984 to present derived from satellite altimetry and optical imaging’’ by Jiawei Hou et al.

We express our gratitude to the editor and two reviewers for generously dedicating their time and substantial effort to assess our manuscript. We have carefully considered all comments and suggestions provided by the editor and reviewers and have made modifications accordingly (review comments in blue, our response in black bold font).

Response to Editor’s Comments:

Many thanks for your thorough revisions of your manuscript, which has now been seen by the referees. Both referees acknowledge your improvements of the manuscript and I am happy to accept the paper for publication after some final minor adjustments. While the suggestions of referee #1, should be straight forward to address - I also want to put some of the critique of the second referee into context. Here I find that a careful adjustment of the terminology may alleviate some misunderstandings regarding your analysis. In particular I would suggest that you consider the following points:

We extend our gratitude to the editor for the encouragement and positive feedback. The insightful suggestions provided below were valuable in enhancing the quality of this manuscript.

EC1) 1. Clarify how remote sensing data are processed in the abstract (and possibly title) to avoid confusion (i.e. that lake volumes are estimated based on topographic characteristics and lake properties that can be observed by remote sensing).

We clarified this in the abstract in the revised manuscript:

> *"Measuring the spatiotemporal dynamics of lake and reservoir water storage is fundamental for assessing the influence of climate variability and anthropogenic activities on water quantity and quality. Previous studies estimated relative water volume changes for lakes where both satellite-derived extent and radar altimetry data are available. This approach is limited to only a few hundred lakes worldwide and cannot estimate absolute (i.e., total volume) water storage. We increased the number of measured lakes by a factor of 300 by using high-resolution Landsat and Sentinel-2 optical remote sensing and ICESat-2 laser altimetry, in addition to radar altimetry from the Topex/Poseidon, Jason-1, -2 and -3, and Sentinel-3 and -6 instruments. Historical time series (1984-2020) of water storage could be derived for more than 170,000 lakes globally with a surface area of at least 1 km2, representing 99% of the total volume of all water stored in lakes and reservoirs globally. Specifically, absolute lake volumes are estimated based on topographic characteristics and lake properties that can be observed by remote sensing. In addition to that, we also generated relative lake volume changes solely based on satellite-derived heights and extents if both available. Within this data set, we investigated how many lakes can be*

measured in near real-time (2020-current) in basins worldwide. We developed an automated workflow for near real-time global lake monitoring of more than 27,000 lakes. The historical and near real-time lake storage dynamics data from 1984 to current are publicly available through https://doi.org/10.25914/K8ZF-6G46 (Hou et al., 2022) and a web-based data explorer www.globalwater.online."

EC2) The term "validation" might be misunderstood by some of the readers, since it can imply comparison with an indisputable truth. I acknowledge that for many environmental variables the availability of observations or the quality of an estimate can be a hurdle. Therefore I find that any comparison to independent data is valuable. For example it may not always be possible to decide that an alternative data-product is better, but demonstrating consistency can increase confidence in new estimates. Similarly comparison using a small subset of the data to available in-situ observations at targeted locations may also be valid in light of limited data-availability. Therefore I suggest that you consider to use terms like "benchmarking", "checking consistency" or similar.

Thank you so much for your comments and suggestions. In the revised manuscript, to avoid the misunderstanding, we changed "validation" to "benchmarking" where we compared our GloLakes database against other, independent data. We only used the term "validation" where we compared our data against in situ data, which, although still not indisputable, would likely be accepted as closer to 'truth'.

Response to Reviewer #1 Comments:

I thank the authors for responding to all my comments and updating the terminology in the paper. Upon reading the revised version, I think the text does do a much better job of communicating what each dataset represents and what they can/cannot do. I just have a few remaining minor comments (see below):
Note – all lines referenced below are from the tracked changes version.

Your valuable comments and insights have made the manuscript more effective in conveying the dataset information accurately. We're pleased that you found the revised version to be an improvement. Thank you for your support and guidance.

R1C1) Line 114: "Explore all ready available, validated and frequently updated satellite data sources" does not accurately describe this dataset, as one could argue that Sentinel-2 is readily available, validated and updated with far more lakes than the version used here (BLUEDOT). Suggest rephrasing to say something like we test "some satellite data sources" or replace "satellite data sources" with something more specific about testing satellite surface water data sources.

We rephrased this sentence in the revised manuscript:

"The underlying strategy in developing this global lake monitoring system was to evaluate a selection of readily available, validated, and frequently updated satellite data sources, explore the relative advantages of each selected source and combine them to complement their respective weaknesses."

R1C2) Line 376: Replace "substantial" with significant. I assume you mean significant at $p < 0.05$? I'd also suggest clarifying 25% of which lakes you're referring to here – 25% of the 170,000 lakes in GSWD?

Apologies for the confusion. This refers to the total number of lakes that measured by both satellite sources in each basin. We rephrased this sentence in the revised manuscript:

"Extents derived from GSWD (Landsat) data and levels derived from ICESat-2 exhibit a significant correlation in over 25% of lakes measured by both in each of 145 river basins."

R1C3) Line 377: Change "this feature" to "these correlations" or similar

We changed this to "*The correlation pattern is evenly distributed across the continents…***" in the revised manuscript.**

R1C4) Line 381: Again clarify that "half of the lakes" refers to half to the 170,000 GSWD lakes.

We rephrased this sentence in the revised manuscript:

"There are 58 basins in each of which over half of the lakes measured by both satellite sources can be monitored by that method, mainly located in the USA, southeastern South America, the Mediterranean, southern Africa, southern Asia, and Australia."

R1C5) Line 505-506: "adequate to monitor dynamics in many lakes" is not specific enough – these sensors are great for measuring very large lakes, but not useful otherwise.

We changed this to " *...adequate to monitor dynamics in large lakes.***" in the revised manuscript.**

R1C6) The additional text in section 3.5 is rather poorly written/structured. The text about Sentinel-2 (lines 516-525) is a very needed addition, and I understand that the authors were trying to add additional context. However, while the information is largely here, the structure doesn't quite work, particularly the transition between the discussion of the altimeters and the Landsat-derived GSWD. I'd suggest rewriting this section and splitting it into separate discussions about the extent observations from GSWD/Landsat/BLUEDOT/Sentinel-2, followed by a discussion of the altimetry and then the nice bit at the end about the potential of radar.

Thank you for your suggestions. We restructured this section in the revised manuscript:

"Remote sensing provides an opportunity to measure water in most lakes worldwide and provide NRT information, which is impossible with the current in situ network. This study used the Landsat-derived

GSWD to estimate lake surface water extents. The main limitation of using GSWD is that it only produced monthly observations from 1984-2020 and cannot provide NRT information, despite its use of Landsat observations with a temporal frequency of 16 days. The GSWD are expected to be updated annually by the Joint Research Centre of the European Commission (https://global-surface-water.appspot.com/download), at which point lake extent estimates can be extended beyond 2020. In this study, we harnessed the capabilities of BLUEDOT (Sentinel-2) to expand lake surface water extent estimates beyond 2020, benefitting from the 5-day revisit interval of Sentinel-2 within the BLUEDOT framework. However, lake area data from BLUEDOT are presently available only for a subset of several thousand lakes across the globe. This issue could be overcome using MODIS, VIIRS or Sentinel-2 data directly, for example. The daily or 8-day composite MODIS and VIIRS products have a better chance to provide valid observations, but the hundred-meter range resolution is often not sufficient to accurately detect lake area changes. 5-day, 10-m resolution water extent derived from Sentinel-2 should be a promising candidate for NRT global lake monitoring. The sheer volume of data presented us with a challenge for data storage and processing, but there is no fundamental limitation that would prevent a similar approach from measuring all 170,957 lakes measured by GSWD (Landsat) in this study. The advantage of Sentinel-2 would be that it can provide NRT lake observations with low latency. The high computation and storage demands could potentially be met by cloud platforms like Google Earth Engine (GEE). In future research, we hope to consider such approaches to improve our data set.

Topex/Poseidon (1992-2002), Jason 1/2/3 (2002-present), and Sentinel-6 (a.k.a. Jason-CS; 2020-present) are all able to measure lake height every ten days, which should be adequate to monitor dynamics in large lakes. Sentinel-3A/B (2016-present) has a revisit time of 27 days, still providing valuable, quasi- monthly updates on changes on water level. The dynamic estimates of lake height from these radar altimeters were seamlessly processed and derived from G-REALM. However, radar altimeters cannot detect many smaller lakes in between the sparse ground tracks. The ICESat-2 laser altimeter covers many more lakes globally, benefiting from its dense reference tracks enhanced by the six laser beams onboard. The trade-off is its temporal resolution of ~91 days, but this is still sufficient to observe seasonal changes in many lakes worldwide, and more frequent water extent mapping can be used to interpolate between these observations.

Comparing altimetry and optical sensors to measure lake changes, the greatest challenge arises from optical remote sensing. This is primarily due to the inherent limitations associated with optical remote sensing, notably the influence of clouds, various atmospheric interferences, and, to a lesser extent, vegetation. This issue could be mitigated by using passive microwave sensors or SARs. For example, the Japanese Space Agency's AMSR2 and TRMM TMI sensors and NASA's AMSR-E and GPM instruments can provide daily observations of surface water based on differences in brightness temperature between wet and dry areas (De Groeve et al., 2015; Hou et al., 2018). Unfortunately, their resolution is generally very coarse due to the observation method. Sentinel-1 SAR could be a more practical solution to monitor lakes under cloud cover, with 12 days and 10-m resolution, provided the water detection algorithm can be automated, and vegetation cover does not interfere with the mapping. Finally, the Surface Water and

Ocean Topography (SWOT) satellite mission was launched at the end of 2022 and is expected to measure surface water height and extent simultaneously every 11 days for lakes greater than 250 m by 250 m. Its temporal resolution is intermediate to the radar and laser altimetry used here, but SWOT shows promise for monitoring lake changes given that the two basic components (i.e., A and H) needed to estimate lake volume change are measured simultaneously."

R1C7) Nice conclusion section.

We appreciate your positive feedback.

<u>Response to Reviewer #2 Comments:</u>

I appreciate the efforts that the authors made in the revisions. However, the authors should really avoid misleading information in the manuscript.

Thank you for your feedback on this manuscript. We acknowledge your concerns and would like to provide a thorough explanation of our database validation approach, as well as the limitations we encountered when validating this new global dataset. In response to the Editor's recommendations, we revised the manuscript to enhance the clarity regarding the processing of remote sensing data for our various products and the methods used for validation. Please see our detailed responses to EC1 and EC2.

R2C1) 1. In the section 3.1 Lake area estimation validation, the authors chose to compare with the Zhao and Gao (2018) while their method for estimating water areas is almost identical to Zhao and Gao. What's the purpose of this comparison and even labeling it as validation? It is almost 100% expected that these two sets of areas are (nearly) the same. In other words, the results (label as validation) here are really misleading.

In the comparison with Donchyts et al. 2022, if the authors only focus on generating one-snap of lake area (e.g., similar to the HydroLakes database), I would be likely convinced by this type of comparison, although the authors may need to provide evidence supporting that Donchyts et al. 2022 is better than theirs and why they did not adopt the methodology of Donchyts et al. 2022. However, the authors aim to focus on water area change over time, this comparison did not provide sufficient information on the applicability of their products, particularly for estimating lake volume changes. As shown in Figure 2b, the discrepancy in area become significantly larger for lakes smaller than 10 km2, although both datasets were generated from the same source (Landsat images). This is unlikely caused by the resolution limitation of Landsat given 30-m resolution should be sufficient for monitoring water bodies larger than 1 km2 (the minimum size of studied lakes here). I would like to reiterate my suggestion that the authors should consider reporting the mean absolute error on a lake-to-lake basis and compare the error to the variability in the time series. This would be more important than the results they showed in Figure 2, in my view.

I acknowledge that the authors did additional validation on 20 lakes (two snapshots per lake) using the higher-resolution sentinel-2. If this is a local or regional scale study, this may work. For a global study of 170k lakes, this value of this evaluation on 20 lakes is really limited.

We think the comparison against Zhao and Gao (2018) is necessary. Although we utilized the same data source (GSWD), our approach to gap-filling contaminated images differs. If these two datasets exhibit consistency, it would indicate the effectiveness of our gap-filling algorithm, at least akin to that of Zhao and Gao (2018).

In our opinion, the relatively large discrepancy observed between our database and Donchyts et al. 2022 does not dominate the comparison results. We have also conducted investigations into these discrepancies and determined that they are primarily attributable to differences in defining lake boundaries, at times making it challenging to conceptually distinguish lakes from interconnected water bodies (L317). Additionally, in both comparison analyses, we reported the mean absolute error as SMAPE on a lake-to-lake time series basis (L313 and L319).

We acknowledge the limitations of evaluating this global database. We conducted a manual comparison with high-resolution Sentinel-2 data. Although we incorporated additional benchmarking results for 20 lakes, we tried to choose a diverse set of lakes (i.e., in different continents and climate zones and with varying surface water color, lake shape and size and surrounding land cover.) for this analysis. Therefore, our comparison results could be considered representative for global studies.

R2C2) 2. The added Figure 9 is also misleading. The authors chose to validate the water levels derived from ICESat-2 (2018 to now) using in-situ level, while their study focuses on generating lake volume time series from 1984 to the present. The authors seem to use the validated accuracy of ICESat-2 levels to report the accuracy for the derived water levels over the historical period (1984 -). In the last review, I commented that the geospatial model may not work for estimating volume changes and suggested that additional validation on the levels estimated from the geospatial model to be validated. What a surprise to see such a misleading response? Additionally, the authors selected 10 lakes for this validation without a careful thought of whether the number is sufficient and whether the samples are representative.

By adding Figure 9 with satellite-derived lake height validation results for 96 lakes, we believe our study now has a thorough validation/benchmarking analysis of all the lake water resources variables, i.e., height, extent, and storage. However, these validation/benchmarking analyses indeed have been limited by the availability of in situ data or published datasets.

We think the geo-statistical model mainly determines the lake bathymetry, rather than volume changes. The accuracy of volume changes mainly relies on the precision of lake height and extent measurements, both of which have undergone validation and benchmarking within this study. In the manuscript, we reported both correlation and bias error for lake height and extent measurements and the geo-statistical model used in this study (L458-459, L313-319 and L399). As there are no available in situ lake extent data, we compared satellite extent measurements against published datasets. While satellite-derived lake heights and the geo-statistical model have been validated by in situ data where available.

R2C3) 3. The authors should carefully go through the entire manuscript to eliminate possible confusions. Many places are misleading. For example, the title implies that the lake volume derived using satellite altimetry and optical imaging. But the generated data largely came from the geostatistical model that seems to be empirical (definitely not based on satellites). In the Methods, “relative storage products are unaffected by any bias from the geostatistical model”. I think that volume changes estimated from empirical models are subject to large uncertainties. The authors should clarify that what this statement means..

In this study, we not only generated absolute lake storage (which is based on the geo-statistical model), but also relative lake storage (relying solely on satellite-derived lake height and extent measurements). The relative lake storage time series were not estimated by the geo-statistical model, therefore we claim that relative storage products are unaffected by any bias from the geostatistical model. We labeled our different products with corresponding methods (L549-553). In the revised manuscript, we clarified this in the abstract.