

Comment on essd-2022-266

Anonymous Referee #1

Referee comment on "GloLakes: water storage dynamics for 27000 lakes globally from 1984 to present derived from satellite altimetry and optical imaging" by Jiawei Hou et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2022-266-RC1>, 2022

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.

Major Comments

- **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.
- **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.

- **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.
- **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 construct 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)!
- **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 it was possible to build NRT and relative time series, and where it was not possible (perhaps in a figure? Or table?).
- **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?

Specific Comments

Lines 6-7: The first sentence of the abstract is unnecessarily long and wordy and contain a

typo. Please rephrase.

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-derived 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.

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

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 absolute 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.

Line 97: Remove "Furthermore"

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

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.

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).

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.

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).

Line 291: Again, please be clear here that these results are ‘estimated’.

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.

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/>

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.