Dear Editor,

First of all, we apologize for taking so long for a minor revision. The remaining reviewer comments were straightforward to address (almost all about wordsmithing and clarity). We have addressed each of them in the response letter below this cover letter. The main reason it took us a while to complete the revision was that we performed a rigorous QC for the total nearly 22,000 reservoir polygons. While our previous QC emphasized the georeferenced dam points (as described in the paper), our QC for the retrieved reservoir polygons wasn't as thorough. Considering that users may rely on our provided polygons to gauge important analyses (such as assessing reservoir water level changes to infer dam operations), we do feel obligated to ensure as much as we can that each georeferenced dam point was paired to its correct reservoir polygon. The need of QC for reservoir polygons was also implied by Reviewer 1. Therefore, we decided to take more time to perform such a QC. Sorry again for taking so long.

Our reservoir polygon QC aimed at the following two goals.

First, we made sure that each dam point was paired to the correct reservoir. This could be challenging for our automated algorithm in lake dense regions (such as the case of cascade dams, where a dam may be paired to the downstream reservoir due to close vicinity). So a manual QC was necessary.

Second, once a dam and its reservoir polygon were verified to be paired correctly, we visually inspected if the reservoir boundary appeared to be reasonably complete or accurate. Some of the polygons in GRanD and HydroLAKES have very coarse boundaries. We manually replaced hundreds of them by the finer-resolution UCLA Circa 2015 Lake Dataset. Due to resolution issues, a reservoir water surface may be disintegrated to several polygons. If we saw this situation, we tried to dissolve the disintegrated portions to one multi-part feature (so the previously separate polygons were merged into one record in the shapefile).

However, we did not manually digitize any polygon. In other words, all polygons in our current version (GeoDAR v1.1) come from the three water masks (GRanD v1.3 polygons, HydroLAKES v1.0, and UCLA Circa 2015 Lake Dataset), as originally described in Methods.

Our automated reservoir retrieval method turned out to be overall accurate, but this additional round of QC (although time consuming) further improved the data quality. Accordingly, we have added a few more sentences in the main paper to explain the procedure of the reservoir polygon QC.

During our polygon QC, we also corrected some other identified issues (such as additional location errors in our georeferenced dam points and GRanD). The number of dams in GeoDAR v1.1 is now 24,783 and the number of paired polygons is 21,515. Other revised numbers and figures (all with minor changes) reflect the final results of this QC. Other text edits were also made to further improve the language readability and clarity of the paper. Unless the editor and reviewer find more issues, we have now finalized our QC and this version of the dataset.

Another change we made is that I decided to migrate our dataset from the figshare repository to Zenodo. While the figshare link was used for peer review, we do hope that once accepted, our dataset will be stored on Zenodo. This was because I have been trying to simplify the managements for both datasets and scripts by avoiding multiple accounts, and Zenodo shares the same log-in gateway with Github. This allows my lab members and me to co-manage both

platforms in a simpler and consistent way. I have contacted the figshare administrator to take off our previous GeoDAR versions and our most recently revised GeoDAR version is uploaded to Zenodo with a reserved DOI (https://doi.org/10.5281/zenodo.6163413). For consistency, we want to avoid uploading another version to this DOI, so we have not yet publish this DOI until our paper is finally accepted. For peer review purpose, our dataset can be accessed through our Github lab account:

1. https://zenodo.org/login/?next=%2F
2. Log in with Github using: Surf-Hydro (user name) and LimRivHydro@8! (password).
3. Then the dataset can be accessed through the “upload” button.

Please let us know if this is okay for the remaining peer review purpose. We will be more than happy to revise further should you find any questions or issues.

Thank you.

Sincerely,

Jida Wang

Author's response to reviewer comments

The authors have done a thorough revision of the manuscript and the additional analysis has improved the database. The methods and results sections have much improved with the restructuring and the new stand-alone sections work better. However, I still think this is an unnecessarily long manuscript (51 pages versus the original 52 pages).

We sincerely thank the reviewer for the constructive comments and an overall positive impression of our last revision. We have addressed the following comments and provided our responses below.

I only have a few minor comments:

I find Table 1 confusing. V1.0 lists a total of 22,743 dams, yet v1.1. lists GeoDAR v1.0 alone as 17,732 dams. I see that this is due to the overlap with GranD but that's not clear from the table. Perhaps add to the GeoDAR v1.0 alone (excluding GranD v1.3) or some similar note.

Thank you. As suggested, we have changed this to "GeoDAR v1.0 alone (excluding overlap with GRanD v1.3) for improved clarity.

Line 105: nominal location is described as (i.e. descriptive information) but in line 127 this is better described as (such as the dam or reservoir address..). I think it would be better to reverse this and use this latter description at earliest use.

Thank you. We tried to keep the first definition more generic and the latter more tailored to the case of dams and reservoirs. However, we do agree with the reviewer and have expanded the definition on the first occurrence as: "e.g., a descriptive address for a dam or reservoir.

Line 224. "more than one record" or use "multiple records"

Thank you for this correction, and we have revised it to: "multiple records".

Line 340. Why (if possible)? Because not all of them could be found?

That is correct. Not all of the coordinates can be corrected. For clarity, we have provided more explanation earlier in this paragraph:

"While this may be true for most cases, we identified at least 88 dams in GRanD with possible location errors. With the help of several references such as regional registers (Table 2), the recently published Dataset of Georeferenced Dams in South America (DDSA) (Paredes-Beltran et al., 2021), Google Maps, and other online documents, we were able to correct the locations of 76 of these dams and absorbed the corrected coordinates to the harmonization. The other 12 GRanD dams, including 3 duplicates with other dams and 9 we were unable to correct the locations for, were excluded from the harmonization. What was also excluded are another 5

dams in GranD that were subsumed or replaced by newer dams. For user convenience, we released these ~90 GranD dams together with the identified issues and suggested coordinates (if possible) in Supplementary Table S4 (full spreadsheet accessible at https://doi.org/10.6084/m9.figshare.13670527)”

Line 957: if he/she uses. Use the neutral “they”

Thank you. This has been revised as suggested.