

Response Letter to the Referees

Here we provided detailed responses to each of the comments from individual referees. For easy reading, we use blue color text for the response from us. Please note that the line numbers in the response letter refer to the revised main text (clear version). The revised China-LDRL dataset is publicly available at <https://doi.org/10.6084/m9.figshare.16964656.v3>.

Referee #2:

In this manuscript, Wang et al. produced a database entitled China-LDRL, which contains (1) thousands of detected large (>1 km²) permanent water bodies (excluding free-flowing rivers) in China, (2) an explicit separation of reservoirs from natural lakes among these large water bodies, and (3) the dam points associated with the large reservoirs, with a distinction between river dams and reservoir dams. This is yet another useful data tool in the proliferating global and regional water body datasets.

I explored the produced database, and found the layers well organized and their relationship logically associated. The manuscript is overall clear as well. However, I do have a few major concerns about the concept of dam/reservoir typology, the crosscomparison with other datasets, and some other technical issues I found in China-LDRL. I would like to see a major revision in both text and the dataset that thoroughly addresses the concerns below.

Reply: We highly appreciate your insightful comments and suggestions. In this revised version, we first double-checked our lake and reservoir maps in Google Earth one by one, and reclassified those reservoirs that were misclassified as lakes in our previous version, such as Hongze Lake in Jiangsu Province. Second, we renamed the dams as “on-stream” and “off-stream” dams following your suggestion, and rechecked each dam/reservoir and determined its classification. Third, we downloaded the newest version of GeoDAR dataset, which had much better accuracy than the previous version, and re-performed the comparison between newest GeoDAR and our China-LDRL dataset. Finally, we improved the main figures and main text, including the introduction of dam classes (Line 243-249), the comparison between our study and GeoDAR and the discussion of the causes of some of the coordinate offsets in these datasets (Line 330-395).

Dam/reservoir typology

1) The automated extraction of surface water from Landsat images follows a standard mapping pipeline and is technically sound. The separations between natural lakes and reservoirs, and then between the different types of dams, were performed by visually interpreting high-resolution Google Earth images. I don't worry too much on the way the authors identified reservoirs from natural lakes because dams and embankments are often clearly discernible from high-resolution images. My main concern is the classification of the two dam types. First, I am not in favor of the terms “river dams” and ‘reservoir dams’. They are a little confusing because many dams on rivers also form reservoirs. At the first glance, I thought a ‘river dam’ is something like a barrage which

has no evident water impoundment, whereas a ‘reservoir dam’ is the one that impounds reservoirs (either on rivers or not). But this is clearly not what the authors meant after I read the text, especially after I saw the schematic flowchart in Fig. 4. If I understand the authors’ intention correctly, I believe a better (and more intuitive) terminology can simply be “on-stream” and “off-stream” dams/reservoirs, with the former constructed on a river/stream (regardless of impoundment) and the latter formed by partial or complete embankment around an offstream lake (either manmade or originally natural).

Reply: Thank you so much for your suggestion. In this revised version, we classified the dams into two types, one is the dams constructed in a river/steam, and another one is those formed by embankment (**Fig. S1**), and we renamed them as “on-stream” and “off-stream” dams following your suggestion. We also introduced the definition of these two types of dams in detail in the main text: *“In an effort to distinguish riverine or off-stream reservoirs from lakes, we uploaded the large SWB vector layers into Google Earth Pro, and checked whether a dam existed around each polygon through the historical satellite images in 2019 within China by visual image interpretation approach. If a dam did not exist, we classified the polygon as river or lake; if a dam does exist, we classified the polygon as on-stream reservoir (constructed on a river/stream regardless of impoundment) or off-stream reservoir (formed by partial or complete embankment around an off-stream lake) (Fig. S1). Simultaneously, the corresponding dam would be classified as on-stream dam or off-stream dam. Finally, the SWB polygons were classified into lakes, reservoirs, and rivers, and the dams/reservoirs were classified into on-stream and off-stream dams/reservoirs (Fig. 3)”* (Line 240-249).



Fig. S1. Types of dams/reservoirs in this study. (a-b) On-stream dam/reservoir constructed on a river/stream regardless of impoundment; (c-d) Off-stream dam/reservoir formed by partial or complete embankment around an off-stream lake. The high-resolution images in this figure were from the Google Earth Pro.

2) Assuming I understand the authors correctly, I found many of the ‘reservoir dams’ the authors labeled are actually on rivers. Some easy examples are: Zhelin Reservoir on Xiushui River (29.257N, 115.487E), Miyun Reservoir on Chaohe River and Baihe River (40.494N, 116.851E), and Nanwan Reservoir on Shihe River with multiple inflow rivers (32.122N, 114.001E). So I am confused why they were classified as ‘reservoir dams’. Was it because some of the rivers are small tributaries that are hard to be seen from Google Earth images? If so, then I encourage that the authors take a deeper stab at the classification, and if necessary, redo some of the classification to ensure a more reliable quality. The revised text should also include a clearer description of the definitions, more detailed rationales and criteria for performing the classification, and relevant limitations of the visual interpretation method.

Reply: Yes, we did misclassify the types of some dams/reservoirs due to the classification systems in the previous version. In this revised version, we renamed the dams/reservoirs as “on-stream” and “off-stream” dams/reservoirs following your suggestion, and we have rechecked all these dams/reservoirs in Google Earth to ensure a more reliable quality. In addition, we also introduced the types of dams/reservoirs in the main text (Line 240-249). Furthermore, visual interpretation method also might bring some about some uncertainties to the classification of dam/reservoir due to the limitation of knowledge and experience of interpreters, and we updated them in the Discussion section: *“In addition, visual interpretation method for identifying dams and reservoirs in this study could also bring about some uncertainties to the classification of dams/reservoirs due to the limitations of knowledge and experience of interpreters, such as the misclassification of some reservoirs regulated by dams/gates as lakes (e.g. Hongze Lake in Jiangsu Province) and the misclassification between on-stream and off-stream dams/reservoirs.”* (Line 422-427).

Comparison with other datasets

3) I overall enjoy reading the comparison section. And I concur that China-LDRL improved the spatial documentation of cascade dams in the South and the Southwest and the reservoirs in the Northeast. Despite the merits, I would like to point out some caveats when attributing coordinate “errors” in the other datasets. Datasets were often produced using different methods and for different purposes. For example in GOODD V1.0, the original digitized dam points were purposefully snapped to the 30-arc-second HydroSHEDS river networks, which led to the offset from the actual dam locations. But on the other hand, GOODD v1.0 is directly compatible with HydroSHEDS and is therefore more convenient for modeling purposes. In GeoDAR v1.1, dam points in China were georeferenced using the Google Maps geocoding API, which led to two “issues”. First, the labels for many Chinese dams/reservoirs on Google Maps are for reservoirs rather than dams (although the names are usually the same). As a result, many “dam” points georeferenced using Google Maps API ended up falling on the reservoir surface instead of on the dams. Second, as the authors should know, Google Maps in China have substantial misalignment (500 m to 1 km or so) between the satellite images

and the map labels, because of China's GPS shift problem (which was intentional). This means the geographic coordinates returned from the Google Maps geocoding API will also carry the same offsets, even though the geocoding procedure is correct. This said, the authors may want to fully acknowledge the causes of some of the coordinate "errors" in other datasets, which will warrant a more objective and useful comparison with China-LDRL.

Reply: Thank you for letting us clear the reasons for the misalignment of these datasets. We agree with you that different methods and different purposes could result in the potential offsets, especially, the Google Map in China has substantial misalignment between the satellite images and the geographic data as google maps uses GCJ-02 coordinate system for the street map and labels but uses WGS-84 coordinates for satellite imagery, causing the so-called China GPS shift problem. In our study, we used WGS-84 coordinates for our polygons of surface water body, and identified the large dams within the satellite imagery in Google Earth Pro. The same coordinates in our study could reduce the misalignments between the imagery and dam locations, and improve the geographic accuracy of dams. We also added the reasons for the misalignment of these datasets in the Discussion section of the revised main text ([Line 384-395](#)).

4) In addition, the GeoDAR dataset the authors used seems to be an older version. The newest and fully peer-reviewed version of GeoDAR is available at <https://doi.org/10.5281/zenodo.6163413> (accepted paper in press). The authors of GeoDAR have manually reduced the geographic offsets of many dams in China, so I recommend the authors re-performing the comparison with GeoDAR using its newest version.

Reply: We downloaded the newest version of GeoDAR, and checked the dam locations. We found that this newest version had much better accuracy than the previous version although some dams also have offsets due to the GPS shift in China (**Fig. S2**). In addition, we further overlapped the newest GeoDAR layer with our yearlong SWB map, and we found $58.49 \pm 6.07\%$ of its dams were intersected with the yearlong SWB layer, and $82.33 \pm 3.98\%$ and $90.22 \pm 3.18\%$ intersected when the tolerance was 100-m and 500-m (**Fig. S10**). In the revised main text, we reperformed the comparison between our results and newest GeoDAR and discussed the potential reasons for the differences ([Line 368-395](#)).

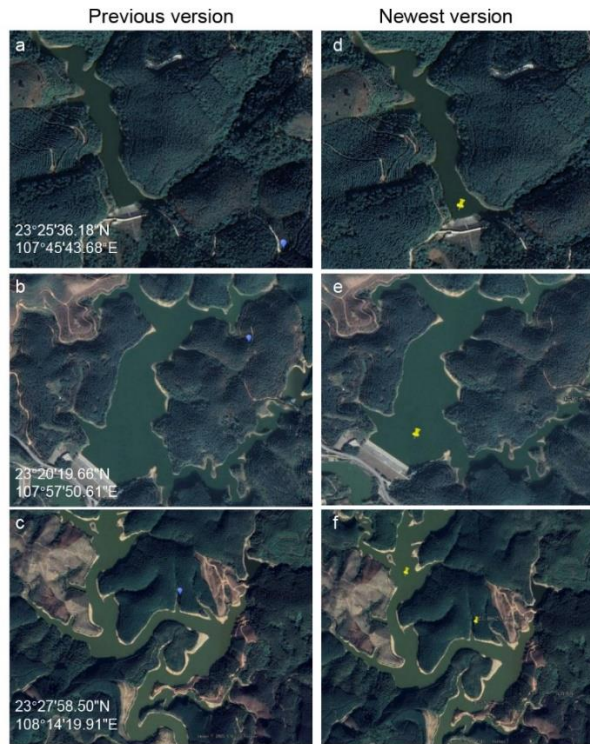


Fig. S2. Georeferenced coordinate offsets of dams from the GeoDAR within high-resolution images.

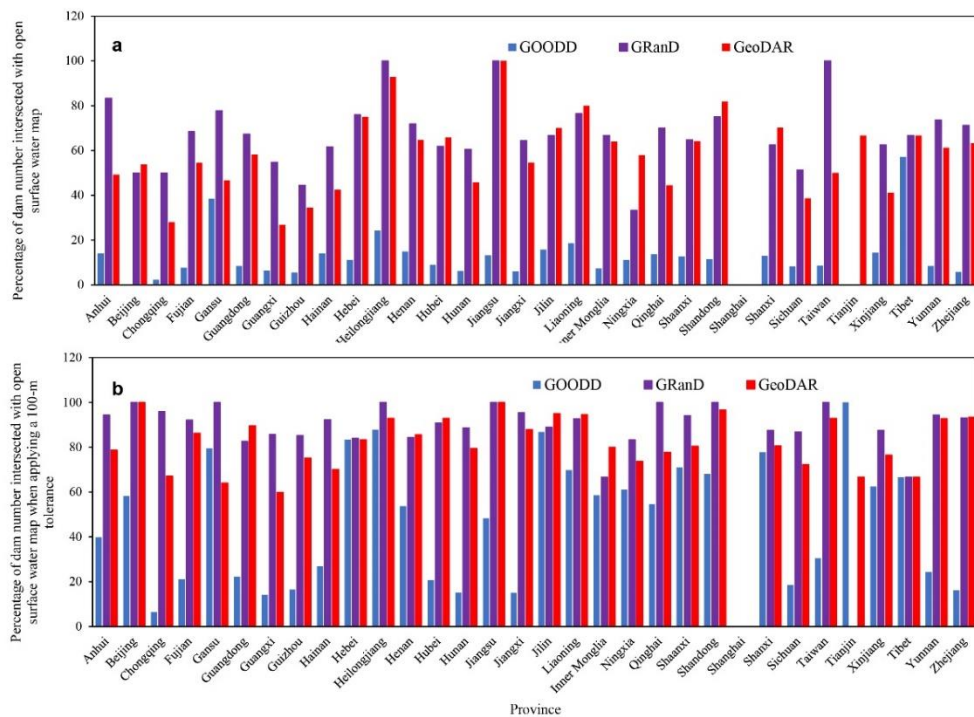


Fig. S10. Numbers of dams intersected with surface water body (SWB) map in China for 2019 by province. (a) Percentage of numbers of dams intersected with SWB map; (b) Percentage of numbers of dams intersected with SWB map when applying a 100-m tolerance.

Other technical issues

5) I would like to point out a few other technical issues I found in China-LDRL. Some of the natural lakes the authors classified are actually reservoirs or regulated (dammed/gated) lakes. Examples are lake ID 18395, which is part of the Danjinkou Reservoir, and lake IDs 1904, 2176, 1428, and 1483, which are arguably part of the Three Gorges Reservoir stretching to the lateral tributaries (some of the polygons were cut off by bridges), and lake ID 76265, which duplicates (conflicts) with reservoir ID 1261. The authors mentioned both GRanD and GeoDAR misidentified Hongze Lake as a reservoir. After a careful examination, I believe Hongze Lake should be a regulated lake (controlled by Sanhezha Gate), thus compliant with the reservoir category (by the way, lake ID 90753 seems to be an editing glitch). In general, I encourage the authors to perform another round of quality control on the classified natural lakes and reservoirs to ensure the accuracy as much as they can.

Reply: You raised very good problems in our dataset. We did classify some reservoirs as natural lakes because of their names, such as the Hongze Lake in Jiangsu Province and Dongchang Lake and Dongping Lake in Shandong Province. In this revised version, we double-checked these lake polygons one by one, and tried to make sure these lakes were classified correctly. In addition, as we used the yearlong surface water body to generate the lake/reservoir maps and some polygons were cut off into several polygons by bridges or provincial boundaries, in this revised version, we tried to recheck each polygon and combined potential polygons as one polygon with one same ID when they were regarded as one lake or reservoir. After the checking of each polygon, we recalculated the areas of lake and reservoir, and reassigned an ID for each polygon to make sure there is no conflicts.

6) I agree with Lines 105 to 111 that water body classes have not been adequately considered in water body dynamics studies. This echoes the PNAS letter from Song et al. (<https://doi.org/10.1073/pnas.2005584117>), which demonstrates the importance of water body types in sorting out the recent surface water dynamics in China.

Reply: Song et al. (2020) did a nice work of mapping lakes in China and demonstrating the importance of lakes and reservoirs. According to Song's study, we introduced the importance and necessary to distinguish lake and reservoirs from surface water body in the Introduction section (Line 37-54), and discussed the potential of our study for exploring the dynamics of different water body types and the conversions between lakes and reservoirs (Line 434-444).