

Earth Syst. Sci. Data Discuss., referee comment RC2
<https://doi.org/10.5194/essd-2021-400-RC2>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on **essd-2021-400**

Jida Wang (Referee)

Referee comment on "Improved maps of surface water bodies, large dams, reservoirs, and lakes in China" by Xinxin Wang et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2021-400-RC2>, 2022

In this manuscript, Wang et al. produced a database entitled China-LDRL, which contains (1) thousands of detected large ($>1 \text{ km}^2$) 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 cross-comparison 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.

Dam/reservoir typology

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 off-stream lake (either manmade or originally natural).

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.

Comparison with other datasets

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.

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.

Other technical issues

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.

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.