



Comment on **essd-2021-58**

Anonymous Referee #2

Referee comment on "GeoDAR: Georeferenced global dam and reservoir dataset for bridging attributes and geolocations" by Jida Wang et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2021-58-RC2>, 2021

Wang et al., describe a new global georeferenced database of dams based on geo matching attributes from the proprietary ICOLD database with publicly available sources to match attributes to spatial locations of dams.

This database complements existing geo-referenced databases as it a) expands on the number of dams for which more attributes such as reservoir storage are available and b) allows for connecting spatial locations of dams with attributes from the ICOLD database. As such I believe this database is a valuable addition to the growing number of global dam datasets (e.g. see www.globaldamwatch.org).

Whilst this database certainly has merits I think it promises more than it delivers. The authors frame the paper as a significant improvement over other dam databases in that it includes more attributes as there are already other global datasets (e.g. GOODD and GROD) that include more dams but lack such attributes. However, the majority of the paper is focused on identifying the spatial location of dams and improvements in quantity and spatial locations of dams. E.g. line 812 "GeoDAR's major improvement lies on the quantity or spatial details of the dams". This framing is understandable given the challenge of linking to a guarded proprietary database. However, this limits the use of the database as only people who have purchased access from ICOLD *may* be able to connect the attributes with the spatial location of the dams. Whilst this point is made clear in the conclusions it could be made more clear in abstract and introduction and the overall framing of the paper. It might be worth focusing on potential applications of the dataset.

Overall, it is also not entirely clear to me why there are two versions of the dataset released simultaneously. It seems to me that V1.1 supersedes v1.0 in that it includes more dams and associated reservoirs and the harmonising with Grand is just part of the method. The authors in line 929 also refer to V1.1 as "our end product".

As noted by the authors, (line 967) connecting the dam locations with a hydrographic network would enable research into hydrological implications and ecological connectivity. This would greatly enhance the utility of the dataset.

As also noted by an earlier reviewer, the paper is very detailed and quite repetitive. I think it could be significantly shortened to make it more readable. In particular, the methods section is very detailed. Whilst this may be useful for some readers, the majority

of readers will not require such extensive detail and could perhaps be referred to supplementary material if more detail is required. The methods section already includes a lot of the numbers later presented in results and discussion while some validation methods get introduced in the results section so some re-organisation would be required.

Specific comments:

I'm surprised that only about 60% of dams from GRanD were found in GeoDAR considering GRanD dams tend to be the largest and usually well documented dams and as such I would expect their attributes to be easily found.

I was wondering if there could be a potential bias in WRD data since this is a volunteered database? Are there any countries not included because they don't contribute to ICOLD?

Line 54. "inaccessible" in what sense? I believe many WRD coordinates can be made available at cost. Suggest change to e.g. not freely or publicly available. In particular as the point about public availability is made in line 58 for regional registers.

Line 93. We may decrypt? Perhaps link to more detail provided in sections 4 and 5.

Line 98. How is it possible that about 1/3 of the WRD dams (v1.1) capture a similar total storage capacity as the full WRD inventory of ~60,000 dams. Is this because the remaining ~40k dams in WRD are non-reservoir dams? Please explain this in this section. Also would be good to provide the total storage in WRD here (which is only provided in line 138)

Line 118: "We acknowledge..." I suggest rephrasing this sentence to something like: "Whilst we have made every endeavour to remove duplicates, we acknowledge that some duplicates may remain in the dataset"

Line 138. 7388 km³ in original WRD. Is this the figure for all WRD dams or for the cleaned version of 56,783 dams?

Line 139: I don't think the Venn diagrams are very clear. Not sure if they are even needed as the text explains the process. A simple flowchart might be easier to understand.

Line 239: "rest parts of the world" is a strange phrase

Line 322: ICOLD storage capacity erroneous. See earlier comment (line 98) on ICOLD reported storage. This can also explain the discrepancy. Note that Mulligan et al (2020) also note erroneous reporting of catchments in ICOLD.

Line 525-562, section 3.2 in results and discussion seems to introduce more methods on validation. This should be moved to methods.

Line 797: I find the term (global) capacity improvement a bit confusing. I guess what is meant is a higher reporting of total dam storage capacity by country or globally which is hardly surprising given that more dams and reservoirs are included.

Technical corrections:

Numbers in some cases use thousand separator (e.g. 7,163 line 238) but not in others. Please be consistent throughout

Line 163. "NID records were accessed"

Line 223 "This led to a conservative success rates"

Line 258: "this process *was* repeated"

Line 570 "We believed"

Line 700: Mulligan et al (2020)

Line 744: GeoDAR

References: line 39 Doll should be Döll, line 42 Vorosmarty should be: Vörösmarty and there may be others.