Response to comments on “Global mapping of lake-terminating glaciers” (essd-2025-315)

Reviewer 1 (Penny How)

Thank you for providing such a careful reading of the manuscript, including the dataset and its accessibility. We also appreciate the proactive comments directly on github to further improve the product. We address each point individually below in red, identifying the resulting changes either in manuscript or the database itself.

Steiner et al. present a community-led effort to identify and categorise lake-terminating glaciers globally, which is compatible with the Randolph Glacier Inventory version 7.0 (RGI7) and intended for future integration. The dataset has been primarily generated manually through a concerted and coordinated effort from the authors. Initial categorisations have been formed from existing global and regional glacial lake inventories, drawing upon and uniting a large portion of the glacial lake mapping and monitoring efforts. Error is estimated based on comparing classifications from two operators, which reveals low mapping conflicts (6.7 %) that is indicative of a low uncertainty in the dataset.

The dataset itself is logical and clear to a large extent, as reflected in the dataset description manuscript. The dataset description paper is well written and a thorough companion to the dataset. My feedback is mainly on the dataset itself and the accompanying Github repository, with my primary focus being to ensure that the dataset is unambiguous to users in the glaciology/cryosphere research community and beyond. Github repository pull requests and issues corresponding to some of this feedback have been included here, and I have added my comments early in the review process so that a dialogue can continue on the Github repository if needed.

In all, I would recommend acceptance after these revisions. I am looking forward to seeing this dataset integrated with RGI7. Great work!

Dataset comments

1. The naming of the lake-terminating glacier classifications

In the dataset and throughout the manuscript, the classifications to describe the relevance of lake presence are referred to as "lake level", "lake_level" and "lake level assessment". The term "lake level" is often used in reference to the water level of a lake, for example in remote sensing (e.g. Veh et al., 2025), modelling (e.g. Steffen et al., 2022), and in situ studies and monitoring efforts (e.g. Camassa et al., 2023).

I recommend that the naming convention is changed to something more suitable and less ambiguous. My suggestion would be "lake category" ("lake_catgy" in the gpkg field name), with Level 0-3 renamed to Category 0-3 (and capitalised throughout the manuscript).

This is a fair comment, and we have now implemented this (i.e. changing ‘level’ to ‘category’) throughout the manuscript, in the dataset itself (“lake_level” to “lake_cat”), the README files, and the original contributor files.

2. Level 0 to Level 3 categories are not in sequential order

The Level 0 to Level 3 relevance classifications are not in sequential order despite their numbering convention. Specifically, no lake contact (Level 0) is followed by > 50 % lake contact (Level 1), then < 50 % lake contact (Level 2), and then < 10 % and/or ambiguous lake contact (Level 3).

The classification levels should follow the magnitude of relevance sequentially, therefore my suggestion is:

- Category 0: no lake contact

- Category 1: < 10 % and/or ambiguous lake contact

- Category 2: < 50 % lake contact

- Category 3: > 50 % lake contact

Where "Level" is replaced by "Category" in accordance with the recommendation above. The dataset, processing scripts, manuscript, repository readme, and statistical analysis should also be updated accordingly.

Thank you for the suggestion, we have followed it and this has now been implemented throughout the manuscript, the ReadMe files, the dataset itself, and the original contributor files.

3. Ambiguous relevance classifications

The definitions of the relevance classifications (Level 0 to Level 3) differ between the ESSD manuscript and the Github repository readme, where the repo readme explicitly describes the relevance to glacier behaviour whereas the ESSD manuscript merely infers this. I would suggest amending the Github repository readme to align with the ESSD definitions, given that it is problematic to define an explicit connection between glacier-lake coverage and the certainty/amplitude of its impact on glacier behaviour. I have made a PR with these proposed changes: https://github.com/GLIMS-RGI/lake_terminating/pull/12.

Thank you for the suggestion. We agree that the manuscript and the repository should align and that too strong of an explicit connection may not be warranted (or necessary). We follow the PR and have merged this together with the suggested changes on naming. We have not completely aligned the text in repository and manuscript, to also preserve the nature of how this was built. Guiding documents were written at the beginning of the exercise as guidance for a large group of operators, while the manuscript reflects our collective understanding at the end of it.

Additionally, there appears to be ambiguity surrounding the criteria for each relevance classification. At various points in the manuscript, the relevance classification signifies:

i) The portion of terminus in contact with lake (e.g. Line 70-72).

ii) The perceived level of influence on the adjacent glacier based on visual indicators (e.g. Line 77-78)

iii) The operator certainty of the classification/ice contact (e.g. Line 76-77, 113-115)

Therefore, the relevance classification is ambiguous as it indicates more than one criteria. In future iterations of this dataset, I propose that criteria i) and ii) should be separated from iii), with a new field denoting the operator certainty. In addition, the criteria for the relevance classification should be revised and clarified in the manuscript (Line 63-84).

Thank you for catching these inconsistencies. We have added a line before the 4 categories ('predominately resting on the relative proportion of glacier ice intersecting with lake water and the potential resulting effects of this connection'). We have also removed the ambiguous line 'Cases where contributors identified a very close lake but had doubts over its impact were classified as Category 1.', reducing the criteria to (i) and (ii). We agree that the operator uncertainty would ideally be classified in a different manner. With the difficulty of clearly framing this uncertainty across different experts, this hasn't been attempted during the compilation. We have however added this as a recommendation for potential future iterations as 'Although a consistent qualification of (un-)certainty across multiple contributing experts remains challenging, future iterations of this dataset should consider including a separate attribute that lets contributors self-assess their certainty for individual cases.'

4. Dataset directory naming/structuring conventions

It is difficult to locate the dataset itself in the Github/Zenodo repository alongside the data handling scripts and documentation. I propose renaming the directory from "tables" to "dataset" in order to make this clearer, and ensuring that only the finalised dataset is in the top level of the "dataset" directory (i.e. moving all un-collated operator definitions to a sub-directory). I have opened a pull request to the Github repository (https://github.com/GLIMS-RGI/lake_terminating/pull/11) with these proposed changes.

Implemented with PR#14.

5. The Greenland periphery glacier outlines .gpkg file is missing from dataset

https://github.com/GLIMS-RGI/lake_terminating/issues/10

Fixed with PR#14.

Line-by-line paper comments

I don't have many line-by-line comments, largely because the language and communication of the findings presented in the manuscript are to an excellent standard. Therefore, my line-by-line comments are largely remarks, questions and figure/table queries.

Line 19-20: I am unsure about the general statement that calving "remains poorly constrained with scattered observations", especially given that two of the three references to support this are over 15 years old. Can the statement be amended to better reflect the advances in calving modelling and integration into system models over recent years.

Without doubt the literature on calving (modelling and observations) is ample, but largely revolves around marine termini or even just ice shelves. We have now specified here that our statement pertains to calving into lakes (we think diving into general calving studies here would go too far) and have added a more recent study that has investigated this locally (with the following study by Minowa et al. also pointing to one of the few more comprehensive quantifications).

Line 28-31: I did not realise that these two global inventories differed so greatly, therefore it is good to see this reported here. Do you know why the difference is so vast? Is this a reflection of the difference in classification approaches and/or discrepancies in manual intervention?

The differences in these two inventories (or rather concern with the quality of each individually when looking at individual lakes) has indeed been a concern, if not further investigated. While we have not attempted to determine the source of discrepancies, the difference we believe lies in the different classification approaches.

Line 49: Great to see the processing script openly provided for this.

Thank you!

Line 58: Was there any specific reason for choosing a lake size threshold of 0.01 sq km? Was this problematic in cases where an existing inventory only contained lakes with a higher minimum size threshold (e.g. Greenland, with a minimum size of 0.05 sq km)?

The threshold was chosen based on Landsat pixels being potentially able to capture lakes of such a size but not below and numerous known lakes that interact with glaciers falling below a higher threshold like 0.05 km^2. We acknowledge that to some degree this is a bit of a subjective choice. For cases where inventories with higher thresholds were used, the manual checking that was employed was able to address this.

Line 79: Repetition of "these".

Thanks, edited.

Table 2: I would like to see the entry type (i.e. string, integer, float) for each of these fields, mainly to guide users who are importing these using R or Python. Also, a short description for each field should be added, similar to those described in Lines 124-128.

Thanks for the suggestion, we have added this to the caption now.

Line 123-127: The AutoTerm field is not defined here, in Table 2, or in the Github repository. I am guessing this is a categorisation of the level to which an external glacial lake inventory dataset was used?

This field is the modified *term_type* field, updated after running **scripts/assign_lake_flag.py**. We have now added the definition/explanation of this field at Line XX of the manuscript, and to the README file in the repository.

Line 149: "...(Table 3,Figure 5 and 6)." >> "...(Table 3, Figure 5 and 6)."

Corrected.

Table 3: I think the region name should be included here, if possible, rather than having to refer to Table 1.

While in principle we agree that this would ease the readability, we are already maxing the page here quite far and adding the full names would in our view use unnecessary space, which we need here for the numbers. We would therefore prefer to simply stick to the IDs.

Line 194-216: An additional table would neatly summarise and compliment these findings (i.e. "Table 4. Statistics from independent flag submissions of glacier classifications")

Thank you for this suggestion. We have now added a confusion matrix of the categories assigned by multiple reviewers (Table 4).

Line 218-220: Are these discussions openly available, for instance through issue postings on the Github repository? I think this could be a great approach to open, transparent discussion and resolutions in future iterations of this dataset. If you would like to use the repository as a user contribution portal then I would recommend: 1) adding a section to the readme on how to contribute; 2) adding an issue template to guide users in writing their contributions; and 3) adding a repo action to check the compatibility of user contributions (e.g. ensuring the submission is a .csv with all essential fields included).

The discussions are not openly available, simply because they largely occurred in an unstructured manner (e.g. by email or whatever form of communication regional teams preferred). One example of such a discussion is visible as issue #4 on the GitHub repository (https://github.com/GLIMS-RGI/lake_terminating/issues/4). Considering that many contributors now (and possibly also in future) are not familiar with how github works and git-literacy shouldn't necessarily be a requirement to contribute on this topic, we also believe that requirements for exchange on these issues or future contributions should not be too stringent. However, we agree that easing the submission process for potential future contributors should be attempted at this stage. We have therefore included a section to the ReadMe on how to contribute, including an example template for users. We have now included a GitHub Action to check contributions (PR#15), and have included additional instructions in the Contributing section of the README.

Line 228: The .gpkg information should be included when describing the format and contents of the .csv tables (Line 123-127), rather than at the end of the manuscript. In addition, the geographic projection (OGC:CRS84) and field descriptions (i.e. fid, IDs, aut_trm, lak_lvl, image_d, imag_dt, invntr_, cntrbtr, notes) of the gpkg files should also be included. Perhaps the field description names in the .gpkg files could be incorporated into Table 2.

Thanks for pointing this out - the attributes in the gpkgs are actually the same as described in Table 2, but the names were corrupted in the process of preparing the packages. This is now updated.

Line 254-256: Normally ESSD publications require a section, or some comment, and how this dataset could be used in future work. I think a couple of comments could easily be added to the Conclusions here, tying back to the relevant literature highlighted in the introduction.

We have expanded on the very last sentence of the Conclusions where we noted the relevance of the dataset for future studies, noting its suitability for scaling in-situ insights.

References

Camassa, R. et al. (2023) Extreme seasonal water-level changes and hydraulic modeling of deep, high-altitude, glacial-carved, Himalayan lakes. Sci Rep 13, 11705. https://doi.org/10.1038/s41598-023-37667-z

Steffen, T. et al. (2022) Volume, evolution, and sedimentation of future glacier lakes in Switzerland over the 21st century, Earth Surf. Dynam., 10, 723–741, https://doi.org/10.5194/esurf-10-723-2022

Veh, G. et al. (2025) Progressively smaller glacier lake outburst floods despite worldwide growth in lake area. Nat Water 3, 271–283. https://doi.org/10.1038/s44221-025-00388-w