

## Reviewer comment on essd-2022-395

### Anonymous Referee #2

Dear Reviewer,

We are very grateful for the close reading of the manuscript, the appreciation for making the data accessible beyond academia and also your concerns regarding the quality of impact data and information through local knowledge. We appreciate that these are issues that need to be taken seriously and respond to them point by point in bold below, with your original review kept in cursive.

*Dear editor and authors*

*In this manuscript, Shrestha et al., presented a comprehensive database of Glacier Lake Outburst Floods (GLOFs) in High Mountain Asia (HMA). The authors combined three databases of glacier lakes in this work. The compilation of the GLOF database is largely based on literature review of articles from different sources including peer-reviewed papers, book chapters, technical reports as well as online news articles. The authors also take local knowledge into consideration, which is believed to be necessary considering the issue with under or over reporting of hazards in rural areas in HMA. However, this also brings challenges to the reliability of the sources. The novelty of this database lies in its inclusion of impacts of these GLOFs downstream, which could be difficult to quantify, fact-check and describe in a single .csv file. The authors have also created an interactive map and dashboard for visualization and quick check for non-academic users. I praise the efforts that have been put on compiling such a database, which has great value in creating vulnerability assessments and hazard adaptation plans for mountain communities.*

*However, I also have a few major comments on how the article is written regarding ESSD guideline and how the data is archived, and quality controlled.*

- 1) First, the authors spent a lot of efforts on analyzing and interpreting the data in Sect. 3 and 4. Of course, an overall statistic of the data could be included. However, I believe, since the article is about presenting a dataset, the emphasis should be put on elaborating the methods used to produce the data, the choices of the variables, analyzing the quality, uncertainties, and limitations in the data and how it could be useful in other studies.*

**We agree that the manuscript so far is quite heavy on going in further depth rather than only pointing out what further potential would have been. This is especially the case for the discussion of lakes (which we have now completely removed). As you point out we think that highlighting the general nature (i.e. Figure 2), statistics (i.e. Figure 3 and 4) as well as potential avenues of investigation (Figure 7/8) is warranted within a data journal, while the analysis vs lake types for example goes too far. We have therefore removed these parts (while the baseline data as well as the connection of the database to lake databases of course remains). We furthermore now add a more critical reflection on data quality, uncertainty (as far as quantifiable) as well as general limitations under section 3.1.2.**

- 2) The methodology about how the lake dataset and GLOFs dataset were derived are not detailed enough. I understand that the lake data was compiled from three earlier datasets from different years. But it is not clear how they are different in terms of coverage and quality, if the three datasets are merged or used for GLOFs happened in different years separately, and if they are merged what the rule for merging is. For*

*the GLOF dataset, the authors stated briefly that the data was derived by reviewing articles and interviews from different sources and verified by satellite imageries, and that false reported events are removed. But I think more details are need in describing and discussing this process as from this description the readers have no idea how reliable these derived data are. And the removed cases need more vigorous justification. Since this article is about the dataset not the interpretation of the data more discussion could be put in these aspects.*

**Following also on the response to the previous questions, we agree that there needs to be some shift in focus. The first part of the question is solved by removing the section on lakes to the minimum necessary as their source (but not describing the datasets), the second is now addressed by having a more detailed discussion of the process in section 3.1.1.**

- 3) *In terms of the datasets, the authors indicate that the dataset is publicly available on ICIMOD data portal (<https://doi.org/10.26066/RDS.1973283>). I assume this will be the main platform for downloading the data. However, the HMAGLOFDB\_v1.0.csv file downloaded from there is not accompanied by either a metadata file, a description file or the HMAGLOFDB\_removed.csv file. Thus, the user who downloads the data from there has no idea what each column in the data file means. I later realize that those files are included on the Github repository. But in the ICIMOD data port there is no mentioning of the Github repository. It would be nice to centralize these different bits of data or at least link them together to be more user friendly, especially for non-academic users since they are considered as important stakeholders in the article.*

**Thank you for making this point. Our decision to have two locations for our database, the RDS as well as Github, was indeed a bit of a challenge that stems from our motivation to make this data as much accessible as possible to both non-academics (RDS, established in the region as a tool that aims to ease the challenge of data sharing between countries) as well as academics (github, making it better traceable and integrateable into future analysis). We understand that this was a bit confusing and now make clear in both repositories the presence of the other. We have now also added the Metadata and the removed file to the RDS. However, we would also like to note that on github the database will be continuously updated in future (i.e. as soon as new events are recorded and checked individually), while on RDS this is only done annually after a revision of the data.**

- 4) *My biggest concern is the criteria for choosing the mechanism involved in lake breach or drainage. It is not clear to me how the authors decided to adopt a definite mechanism or mark it as unknown. I did not check all the GLOF events in the data files. But for the lakes I checked (next to Kyagar glacier and Khurdopin glacier) there seem to be some issues with this.*
- The 34 GLOFs of the lake next to Kyagar glacier are well documented in different articles, which are cited in the HMAGLOFDB\_v1.0.csv file. The GLOFs before 2018 were believed to be triggered by ice-dam failure linked to subglacial drainage. And the one in 2018 was more likely to be linked to overtopping. However, the entry for the mechanism is unknown. I don't know what the authors' reasoning behind that.*
- The GLOF next to Khurdopin glacier are marked as caused by 'englacial tunnels. However, Bazai et al. (2022) used a subglacial hydrology model to simulate the sudden drainage. Clearly, Bazai et al. (2022) thought the drainage was likely to be linked to subglacial drainage system. The authors of this manuscript have cited Bazai et al. (2022) but decided to adopt englacial tunnel as the mechanism.*

Thanks for noting these challenges. Specifically, for Kyagar as well as Khurdopin, we have made an error in the database – these are definitely subglacial drainage mechanisms at play and we have corrected this error now and have redone Figure 4 and adapted the text. We regret this error. As for the many other cases, we generally do only add a known mechanism when this is either documented from the source publication, the news item or without doubt visible from satellite imagery. In most cases this is simply not given, and while we could make an 'educated guess' we believe this would not be beneficial for further analysis. We hope to gather more evidence in future to potentially complete this part, but as of now are not able to do so considering the limited documentation regarding individual events.

- 5) *There are many events that only recorded in 'this study'. We have no idea how they are identified and quality controlled.*

**We apologize for the omission in this case. While for some of the events we have provided documentation from other sources (i.e. non-academic, news reports etc) there were indeed many cases where this was missing. As the co-authors are responsible for monitoring mountain hazards in the region, either in the field or applying remote sensing, there are a number of cases we are aware of from our daily work. Some of these have been detailed in technical reports (e.g. Ashraf et al. 2015), which are however not publicly accessible. In all cases where this was still feasible we checked all sites with satellite imagery as we did for other cases, following (Zheng et al. 2021). For some of the cases (e.g. the ones in Afghanistan) co-authors were responsible for the rapid response missions, visiting the field sites including the sources. We have not included cases where local populations reported GLOFs where all evidence points to simply debris flows (i.e. no lake source available).**

*Besides, the major issues I also have a few minor comments:*

*L90: (Chen, et al., 2021) -> Chen et al. (2021). There are also some other places that the citations are written not according to the convention.*

**Thank you, revised throughout the manuscript.**

*L115: This kind of statement or practice does not seem to be very rigorous. The authors have excluded many cases that might have been caused by debris flows but include cases that have happened far away from any glacier just because they appear in a landscape that was most likely glaciated at one point?*

**We agree that this is confusing but argue this is due to our formulation. The events we included are indeed all from glaciated terrain, where ice is still present even if not necessarily part of the current inventories anymore. The Lang Co event is actually the only case where no more inventorized glaciers are close, but very recent glacier cover is likely given the geomorphology and rock glacier presence is likely. For other cases where we report 'no glacier' it is simply not possible from the source or the satellite imagery to define which glacier would be the potential source of melt water from the existing inventories. We have reformulated the text in this paragraph as below:**

*We also record GLOFs that cannot be directly associated to a glacier, either because from the source or satellite imagery it is not clear which glacier upstream feeds into the lake or because there is no adjacent glacier in any of the available inventories.*

*L120-124: The information in this part is not included in the dataset but only describes how the data is analyzed for the discussion in Sect.4. Following my major comment no.1 I don't know if this should be put here or be included at all.*

**The aggregation is also important for Figure 3, which we believe is important to show the overall distribution of the data we have compiled, and hence we would prefer to keep it. We have however considerably shortened the discussion in section 4, following the suggestions above.**

*L193-194: In which place did the other 0.8% happen?*

**In China (29.4%) and Kyrgyzstan (22.4%), we had rounded the numbers. Revised.**

*L 339-342: Need some references here.*

**Following the suggestions to considerably shorten the discussion as well as the data on lakes itself, this section has been removed completely.**

*Fig. 2: It is not clear why this figure should be included and why the pictures of these glaciers are selected. These pictures are not a part of the database; or should they be included as a part of the database? Something could be a reference is the Norwegian Water Directorate GLOF map (<http://glacier.nve.no/Glacier/viewer/GLOF/en/>).*

**The figure is included to visualize the type of GLOFs that are included in this dataset and also visualizes to the reader what these types would look like and how we assess them in the field. All the shown examples are from the database. To include the photos into the database is the plan and will eventually be realized on the dashboard (<https://experience.arcgis.com/experience/20a0ef1d86ec4a77b2744df9e495214e>) where the photos available of a GLOF will appear as you click on the event. However, the development of this dashboard is lagging behind the publication and is subject to staff availability at our employer and hence could not be brought to the final status of the manuscript.**

*Fig. 3 Maybe it is better to use another colormap to be color-blind friendly.*

**Thanks for the suggestion, the colormap has been changed to a color-blind friendly scale.**

*Fig. 4 A Should the cause of lake appearing be 'glacier melt' or 'glacier retreat'? Or it means something else?*

**This is again based on the information from studies. While some of the lakes where also formed of course as ice retreats, the information generally pertains to melt water provision, irrespective of retreat or stability. We have now explained that better in the caption.**

*Fig. 6 Not quite sure what the x-axes represent.*

**We have decided to remove the Figure as it comprises analysis that goes too far in depth.**