Reviewer comment on essd-2022-395

Anonymous Referee #1
Referee comment on "HMAGLOFDB v1.0 – a comprehensive and version-controlled database of glacier lake outburst floods in high mountain Asia" by Finu Shrestha et al., Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2022-395-RC1, 2023

Dear editors,
Thank you for giving me the opportunity to review the manuscript essd-2022-395 "HMAGLOFDB v1.0 - a comprehensive and version-controlled database of glacier lake outburst floods in high mountain Asia" by Shrestha and co-authors. In this study, Shrestha et al. compiled an inventory of glacier lake outburst floods (GLOFs) from sources such as scientific literature, media, and eyewitness accounts since the mid-19th century.
Of the 682 documented GLOFs, 49 cases have not been previously recorded in other databases and appear for the first time in this compilation. The authors define dozens of variables that characterize the location, size, impact, and consequences of these GLOFs and attempt to gather all available information to populate these attributes with numerical or descriptive information. An important observation is that reporting on GLOFs has not been systematic in recent decades, leaving large gaps in the database that could be filled by further research.

Shrestha et al. report on the largest GLOF database to date in High Mountain Asia. The authors validate each case using geomorphological evidence downstream of the lake (but not necessarily changes in lake size) on satellite imagery, as well as assessment by local experts. In my opinion, this assessment is among the most carefully prepared inventories in this region, and I commend the authors for this work. The authors also provide a list of previously reported GLOFs that they excluded from this database, and the reasons for excluding these cases (e.g., incorrect coordinates or little evidence of former lakes) sound largely reasonable. I also appreciate the interactive map that allows non-experts to have easy access to this data (but without a download option).

Despite these efforts to compile this database, its presentation and associated manuscript require thorough revision, not least to comply with Earth System Science Data guidelines.

I will first comment on major issues associated with the database, followed by issues in the manuscript.

Dear Reviewer,

> **We are extremely grateful for the careful reading of the manuscript and the detailed comments provided. Specifically, regarding the visual dashboard, there is indeed no direct download option, as eventually the dashboard should point to this publication. This is a bit of a chicken and egg problem and with the already two different databases we point to we have decided to remove the link to the interface for visualization. We will keep developing the dashboard further and any data that will be updated for the database will also be updated on the dashboard, but the link will only be visible in the database itself (on github as well as RDS).**
> **We will address each of the concerns raised individually below with your review in italics followed by our response in bold.**

1.) The database is currently archived on at least four different platforms (Zenodo, Github, ArcGIS and ICIMOD), and some files are available on one platform while missing on others (e.g. HMAGLOFDB_removed.csv or the list of references is not on Zenodo). I strongly encourage the authors to use ONE non-proprietary, accessible platform, i.e. without the need for a login account, to deposit their data in order to pursue their goal of presenting a version controlled database. This repository should contain all files contributing to this database (e.g. metadata, the list of references). My preferred choice would be Zenodo, but I leave that decision to the authors.

We agree with the reviewer (as with others who made similar comments) that this is not ideal. Regarding github/zenodo, these are not two repositories. Github is where the data lies and is updated, zenodo is linked to github and where any new stable release is published. RDS is the regional database system that allows us to reach stakeholders well outside of academia, and we hence believe that this is a crucial location as well. It however has the disadvantage of requiring a log in etc. We hence will now provide just one DOI for the database (for the zenodo location), which will include the link to the dashboard as well as the RDS database and will refer to the github page as the 'development version' as has been done previously (e.g. Welty et al. 2020)**. The description of the text is adjusted accordingly, and it has been added to the abstract.**

2.) The references column should contain the full name of the reference (author, title, journal, year) to make the underlying source easy to find.

We would like to refer to a full reference document provided separately in a text file (HMAGLOFDG_CIT) where the full citations are given. We now provide one column with full citations of all papers sourced ('Ref_scientific_full'). Adding separate columns for authors and titles, especially when there are multiple sources makes the table extremely clumsy (and would mean we would need to keep adding more and more columns as time goes on) and we would hence prefer to keep this information separate.

3.) This database contains dozens of new cases that have not been reported before. I believe it is important to properly document these new cases, e.g., with satellite imagery before and after the GLOF showing lake area change or downstream impacts. I would like to encourage the authors to provide supplementary material for these cases.

Providing satellite evidence for all is not possible for different reasons. (a) Some events are based on local information from stakeholders we work with predating the satellite age. And while we did check for all cases whether a GLOF would be possible (i.e. there needs to be a lake or a depression that could have been a lake in the past) and locals did not confuse it with debris flows from heavy precipitation or snow melt not for all imagery is available (i.e. before and after documentation). (b) In some cases (especially in the Upper Indus basin) there is no evidence of the lake as we suspect them to be englacial lakes (or water pockets as we now call them, we have adapted that throughout, also considering that in some cases these pockets may be small or scattered and hence lake becomes a misnomer). Again, we checked all these cases and have discarded events where it is highly likely that it was just a debris, also based on the local information of the process chain. Following the suggestion of another reviewer, we have now also taken greater care in describing this process of data collection. We realize that a lot of information on the events we here present the first time were missing from the actual

database, which we have now added in the separate column on Ref_other. This mostly refers to information gathered locally while being present for first responder missions by co-authors of the manuscript.

For cases where it was available and where we did consult satellite imagery because local evidence was not available, we now provide an indication of the imagery where the evidence was collected in a separate column. Note that many of these events we could not associate with any particular date and have happened likely before satellite coverage started.

4.) The value of this database is that it can be used to calculate trends in GLOF occurrence, hazard, and risk. To this end, each case should be provided with at least an approximate time stamp. If the exact date is not known, the authors could at least indicate that the GLOF occurred before the first available satellite image (Landsat or Corona) was acquired or during a period embraced by two satellite images. This information should be added for the entries in lines 474 to 660 in the current database.
For these cases, it is still difficult to assess whether the occurrence of these cases could be more accurately dated, e.g., by using satellite imagery that cover the outbursts or not. In addition, it would be good to split the Year_approx column into two columns, one with the latest possible date (or year) before the GLOF and one with the first possible date (or year) after the GLOF.

Also following our response to the previous question, we would like to emphasize that the database we provide is not remote sensing based but relies primarily on field evidence, additionally to cases previously documented in literature. We believe this to be crucial as it allows us to make statements on impacts, but we also believe that this local knowledge so far has been relatively less recorded in literature. Since our co-authors are specialising on this kind of fieldwork, we are also able to discern potentially false information from actual GLOFs. We do however use satellite imagery for cases where no other way of supporting our claims is possible. We now have introduced a column called Year_Sat for those events that have likely occurred in a period covered by satellite imagery, noting the earliest year we could find that provided evidence for it. For many cases Corona or L7 are actually not adequate as the events were too small to be identified. In these cases, we rely on available Maxar imagery (Google Earth) or Sentinel imagery (see the remarks column). This still leaves us with many events that have no date, which are generally events that have been documented locally but there is no satellite imagery. We still believe this information to be valuable as a density of GLOF events, even without a time stamp (but say an occurrence during a century) tells us something about their role in forming landscapes.

5.) I have checked some GLOFs (yet not all systematically) and some cases seem to have limited evidence on satellite imagery, or the coordinates seem to be in wrong locations, e.g. at Langco, Tulaco, Phyang, Langbu Tsho or Bugyai. Oubuguoco has two entries, and I wonder if this is a duplicate? The GLOF from Pogeco in 2002 appears to be a misidentified GLOF, according to Nie et al. (2018). Looking at the removed cases, I wonder why so many GLOFs from Lake Merzbacher were discarded? Ng and Liu (2009) and Kingslake and Ng (2013) provide a thorough compilation of these cases, including flood volumes and peak discharges.

Thanks for checking through the database, we respond to the individual issues one by one. There is only one entry for Oubuguoco that we can find. We have rechecked the Lang co event and it is confirmed and matches with the reference cited (coordinate shows the lake). The Tulaco coordinates are

also confirmed with the reference cited, the coordinate is located on glacier as lake type was supraglacial. The Phyang case has been poorly documented by local sources (see the added link in the sources given) but gives some confidence on there having been a possibly small lake. The coordinates where of the impact area however and have now been adapted to the source area. For Langbu Tsho the coordinate is correct and confirmed with the reference cited. For Bugyai the coordinates were shifted, which has now been corrected and relocated to the lake. As for Poge Co, there seems to have been previous confusions over names (see (Nie et al. 2018) for a detailed discussion on the case) – the event from 2002 is the same as Dega Co. Note that however that lake's name is also used for two different cases, and they can be differentiated by their ID.

As for the Inylchek/Merzbacher cases we were concerned that many of the recorded events were very close together and, with experience from many other GLOF cases it is hard to know whether there is not only a mix up in dates or it was the same GLOF, only with release of water over multiple dates as the channel expanded and hence events were recorded multiple times. We therefore followed the Glazirin compilation (Glazirin 2010), which based on the local insight we considered most appropriate until the date of its publication. Upon revisiting the removed cases we agree however that some of the years covered in (Kingslake and Ng 2013) do not appear at all and we have revised this accordingly and brought those into the database (e.g. the 1968 case). This has of course then also resulted in slight changes to the overall statistics (as have the changes noted above for few cases on coordinates and hence elevations) however due to the small number they do not change the overall quality of the database.

Regarding the manuscript, I have the following concerns and recommendations for improvement.

6.) The reason why the variables were selected could deserve a much stronger motivation/ justification. Why is it important to document river basin and lake volume?
Why is it important to distinguish between female and male fatalities from GLOFs? In this regard, the metadata table should also be part of the main manuscript and adequately explain the meaning and units of each variable in separate columns.

We have now added more information on our reasoning behind the variable choice in Section 2 in detail. The coordinates of the GLOF's source as well as the final impact location downstream are provided as these coordinate supports to prepare a hazard zonation map to analyse and evaluate the associated risk (Uddin and Matin 2021). The source lake location (IDs), associated glacier, lake volume before flood, flood discharge and river basin are given where known and existent. This information is crucial for planning, designing and implementation of large-scale projects like hydroelectric power plants and other types of infrastructure, in order to ensure sustainable development. Information on impacts on livelihoods and infrastructure are provided to conceptualise flood-induced coping mechanisms, enhance livelihood security, and foster self-reliance toward economic stability (Ministry of Planning Development & Special Initiatives 2022). This information is presented in quantitative as well as qualitative formats enabling the database to be read by machines, while retaining information that may not be readily quantifiable. Information on fatalities is categorized by gender and disabilities, as disasters impact women and those with disabilities differently than men (Erman et al. 2021). This data is important for addressing gender inequality, cultural beliefs, and socio-economic factors,

as well as advocating for the integration of gender and disabilities perspectives into disaster risk management efforts (UNESCAP 2022).

The name and explanation of each variable and units are already provided in Table 1.

7.) I feel that the compilation of lake data distracts from the real topic of this manuscript, which is the compilation of GLOFs. One motivation for the authors seems to be that they want to add a unique ID to each GLOF according to a previously compiled lake inventory (is my assumption correct?). However, the number of lakes may change over time, and the inventories have different minimum mapping units, different criteria for mapping lakes, or choose different distances from glaciers to consider a lake a glacial origin or influence. The methods used to map and categorize glacial lakes are subject to very different challenges and uncertainties than the creation of GLOFs and are therefore part of a very different story. It is therefore surprising that the results begin with a presentation of lake area change, given that the title and scope of this manuscript refers to GLOFs. There is also little information on how these lake inventories were mapped, particularly the multi-period inventories (2000, 2010, and 2020). Were these inventories prepared by the authors or as part of a previous study? That being said, I strongly recommend removing the lake inventories from this study and focusing on the GLOFs.

Our motivation was indeed to have this closely linked as lakes are of course the source and it would allow a quick combination of the datasets for further analysis. Following your and other's comments we appreciate however that presenting two datasets in one manuscript dilutes the focus completely and have hence decided to remove this part. Of course, the associated lakes will remain in the database, so will some of the discussion on what the value of linking them is, but the actual discussion of the databases itself has been dropped completely.

8.) Removing the lake inventory from the study would also help avoid comparing lakes with and without outbursts (e.g., in Figures 5 and 6). ESSD is a data-driven journal and discourages scientific interpretation of the data. Instead, the discussion might focus more on the completeness (or gaps) of the variables. Could these be closed in future updates of the database? If so, how? Is there a need to collect additional variables? If so, which ones?

We absolutely agree with this assessment and have dropped these figures. We retain some discussion on the resolution of the lake inventory which goes in the direction, why recording GLOFs from local sources matters as remote sensing alone cannot capture every type.

Specific comments (line by line):

L1: It is difficult to say why this is version controlled. Have there been previous versions of that database?

This is the first GLOF database prepared for HMA. The database will be updated annually in collaboration with regional partners, and the updated version will be released at the end of each year. The advantage of the github repository (as for previous ESSD publications like (Mankoff et al. 2021)) is that any updates can be traced in future. This would also make future efforts of tracing how some events made it into records simpler. From the review we have experienced how confusing some records are because they are repeated from secondary literature and at one point it is not clear anymore where the

records have been initially documented. We also think this is crucial for the region as it helps to build trust in data and hence the sharing of the same.

L22: A number of cases seem to have occurred before 1833 according to the csv file?

Yes, those events are very limited (and no trends etc should be read into it) but it has been mentioned in line no. 189 of the original submission.

L27: It is common practice in ESSD that the abstract ends with the link and a reference to the repository. Please add.

We regret this omission; this is now stated at the end of the abstract.

L29: 'expanse': amount?

Yes – and for a spatial domain we believe that should be appropriate wording.

L31: 'on a large scale': please be more specific.

Added detail for the region.

L32: 'Many of these lakes': better use another phrase. According to your database, about 1% of all lakes had outbursts; 'results': please check the verb here.

Revised with 'numerous lakes' and 'resulted in'.

L34: 'many decades in different parts' please be more specific.

Revised with 'GLOFs have been recorded in various parts of the world for decades' – the details follow from the citations following.

L35: 'HKH': please explain.

Revised.

L39: 'n=…', just say 2916 lakes? Did these two studies use exactly the same methods, i.e. same mapping area, same buffer around glaciers?

Thanks, revised. Yes, they both apply buffer area within 10 km of glacier extent for glacial lake inventory. However, since we now remove the discussion on lake inventories, this discussion has also been removed.

L45: 'potentially dangerous lakes': I feel that this phrase adds little objectivity to the discussion of potential changes in GLOF hazard or risk and suggest to remove it.

Removed and replaced with hazardous lakes, this follows the papers cited which focus mainly on lakes that hold hazard potential.

L50: 'strain dams': if they are overtopped?

Not exactly. The pressure created by the rising water level can weaken the moraine's stability and cause the surface to erode, which could lead to the dam's failure. We see this more frequently than the actual overtopping due to an excessive water level.

L51: 'seismic events': you mean earthquakes?

Yes, predominantly but can also be other seismic shocks that may be caused e.g. my nearby mass movements. Hence 'seismic' remains inclusive rather than just focusing on earthquakes.

L56: 'in the shadow': you mean downstream?

Yes, and changed.

L77-92: As described above, I discourage from putting the compilation of lake inventory into this manuscript. If the authors would like to keep it, much more information is needed on how the lakes were mapped.

Following the response above and the advice given, we have indeed decided to remove this section and limit it only to the acknowledgment that these inventories exist, are tremendously useful and what their drawbacks may be in light of use with GLOFs.

L103: 'overreporting': maybe use another phrase such as 'misidentified' / 'confused'?

Revised and replaced with 'misidentified'.

L107: What about changes in lake areas and exposed lake beds? Is this not a criterion for identifying GLOFs?

Yes, they are significant, and we took them into account in this study, when double checking potential sites from satellite imagery. The text has been revised.

L112: This file is only on Github, as far as I can see. Please choose one repository that contains all data.

We regret the omission. As we now only refer to one DOI, all data is available at the location. Additionally, the extra files are also available on the RDS.

L126-135: The content in these sentences is largely a repetition from the introduction. Please merge this information with the introduction and delete here.

Thanks for noting this redundancy. We have shifted some of the material and integrated it in the Introduction. Section 2.3 was furthermore completely revised.

L134: 'Multiple datasets': please cite.

Cited.

L135: dh/dt: please explain.

Rephrased and citation added.

L136: 'in .git on a rolling basis': please explain.

We have now expanded the explanation here.

L137: 'RDS / DOI': please explain

Revised the text.

L138-140: The concept of this ID is very confusion: What does GF, E, and N stand for? How do represent the precision of a given coordinate, i.e. how would you write a coordinate of 79.200239? Doesn't Z need to have two digits, if a lake bursts out more than 9 times, as reported in Table 3?

We are sorry for the confusion and this format has also changed following another reviewer comment. The acronyms have been updated in the text. The precision is measured by the three decimal places in the latitude and longitude (e.g., GF075474E36344N_3) and has been updated in the document. Z represents infinity but begins at 0. It has now been clarified.

L144: Please add the entire reference to the database, not only author and year.

We would like to point out that all details on references are provided in a separate document (HMAGLOFDG_CIT). We now provide one column with full citations of all papers sourced ('Ref_scientific_full').

L146-147: add the name of this file here.

Added.

Table 1: There is hardly any information on why these variables are part of the database. Please add more reason why you selected these variables. Please copy the information from the meta data file here. How did you extract Elevation [m a.s.l.]? How did you obtain Area? Did you map the lakes? 'Displaced_disabilities': missing 'l'; column 'Certainty' is not part of the database, as far as I can see.

We have now revised the section around this Table and the description of why and how these variables were determined. 'Certainty' refers only to removed cases and does only show up in that respective table.

L155-166: As discussed above, the presentation of a lake data base is questionable, and I would encourage the authors to remove it.

As described above, this has been amended.

L172: 'ICIMOD': copy the link here and avoid foot notes.

Revised.

L177: I couldn't see the persons. Please add a circle.

We added a marker to visualize the size of the people in the photo.

L178: 'Milad Dildar': is this the photographer? Do you have permission to use the photos?

Milad is one of our field staff and the picture was taken during a field visit of our organisation. He provided the field photograph. The photo credit has now been mentioned.

L183: 'water line': I couldn't see it. Please add an arrow

Added.

Figure 3: Please avoid red and green in figures at the same time, suggest using a viridis color scale: https://cran.r-project.org/web/packages/viridis/vignettes/intro-to-viridis.html; which study is the source of these 7 regions? Please add more space between the decades/ months. It is challenging to say when decade / month ends and the next begins. Would also be good to report the total sum per decade.

Thank you for these comments, completely revised.

L190: I would welcome very much if the authors add a supplementary file that validates the occurrence of these newly reported cases.

As discussed further up, we are not able to do the satellite image validation for all the cases, be it for time of imagery availability vs when the event happened, available resolution and for all of the events in the database the simple lack of time. However, we now do supply that for cases where we only had imagery as a source to begin with. We specify the imagery dates where it was visible in the 'remarks' column.

L193: Is this statement referring to the removed cases?

No, this number was referring to the actual database, we have now amended the text to make this clear.

L200: allowing 'for'?

Revised.

L207: 'supersaturation': unclear, please explain.

Maybe supersaturation in this context was not ideal – this is water accumulating in crevasses (similar to what you would have in surges or what is hypothesized to stand at the beginning of glacier detachments) that leads to eventual rapid drainage. It's neither a lake nor a detachment (the ice doesn't fail all the way to the bedrock) but a superficial event. We now called it 'accumulation of meltwater'.

L209: Why not simply calling it a water pocket, similar to the study of Haeberli, 1983? 'Glacier outburst' sounds like if parts of the glacier get mobilized.

We agree that the terminology may be misleading. We have now referred to this here as well as in the database and the rest of the manuscript as 'outbursts of water pockets'.

221: 22 'GLOFs'.

Revised.

L224: 'inventories': which ones?

RGI 6.0 preceding this text, we now adapted it to just say 'this inventory' as we can not proof it for others, but generally believe this is likely true for other glacier outline datasets as well.

L226: 'this data': change to 'these' (data is plural)

Revised.

L228: How likely is it that all the fatalities came from the GLOF itself? The GLOF was part of a large rainfall event with many debris flows (see also the image in Allen et al., 2016).

Rainfall-triggered disasters are often characterized by a variety of hazards that can lead to multiple fatalities, which may not be attributed to a single event. In the case of the 2013 India GLOF event, we have mentioned that 6000 fatalities were caused by a multitude of factors in a complex compound event. Also, mentioned in L398-399. There is no attribution study (and that is likely impossible for the event) but from the way deaths were recorded in Kedarnath (including people who died well outside the flood path but simply in the region due to the rains and ensuing floods/debris flows) as well as the intensity of the major driver (rainfall) we believe that actually only a fraction of this number can be attributed to the GLOF. And as we show here, this total number accounts for 86% of all fatalities in the region. Since such numbers matter in the public discourse on a hazard we think it is important to be cautious here since the story looks very different when we say '7000 people died in GLOFs' vs '1000 people died in GLOFs' and that eventually drives investment in hazard mitigation in one direction that maybe should not be the only focus. This is naturally true for any other events as well. This reflection on policy implications of hazard/risk data however goes beyond the scope of this data paper.

L254: 'potentially': why? Did these floods cross the borders or not?

Only a few cross-border events have been reported (see next sentence). We have now revised this part to make this argument clear.

L275: 'hundreds': suggest to add 'to thousands of meters'?

Revised.

L277: 'SRTM': please explain, and mention this part of the work also in the methods; 'between 10 and 50m': how do you know about these uncertainties?

This was based on the varying elevation values when selecting elevation from different parts of a lake via the SRTM but is indeed quite handwaving and we did not systematically assess this. We have hence rephrased this whole part together with the following sentence in the manuscript.

L278: 'errors': of what?

See response above.

L280: 'satellite imagery close before the drainage date': it's still unclear whether (or not) you mapped the lake area before the GLOF in this study. Please be more specific. 'further analysis': which analysis?

We did not systematically assess areas of lakes from satellite imagery. For many cases imagery is not available just before the event (to be sure that it didn't grow further between image acquisition and drainage) and we rather relied on the lake inventories that have been already developed systematically (and hence can be compared as well, not further increasing the uncertainty because of operator bias). The areas we report are values reported in the respective sources or if the event has been recorded on our behalf, we did try to confirm the lake area if available from within a few days before the event. We have rephrased this section now to make this clearer. We have removed the sentence with 'further analysis' as we have removed this part from the manuscript altogether.

L289: 'wilful tampering with data for political reasons': interesting thought, please elaborate.

We have now expanded on this topic, based on our experience from fieldwork. This is also true for discharge values for example, where in the subcontinent the misrepresentation of volumes (stemming from taking the originally reported cubic feet per second as cubic meters per second) is frequent and generally inflates the imagined or expected impacts.

L289: Please add more information on how you deem a source of information 'generally trustworthy'.

The events we report here that have previously not been part of literature are predominately based on fieldwork by co-authors, i.e. interaction with people impacted and visit of the respective event sites. In some cases, the sources were visited, sometimes repeatedly (e.g. the events in Afghanistan as well as most of the events in Pakistan) but during the discussion with locals as well as from local news reports often more events appeared. For those we checked whether the characteristics of a GLOF typical for the site (e.g. for subglacial drainage an initial decrease in flow followed by a steady increase, while for moraine dammed cases a rapid peak) were met and if people were aware of the lake in the upstream changing its properties (often known from herding at high elevation). The most challenging part is to make sure that it was not a debris flow without a lake in play. If satellite imagery can not confirm it. This is where satellite imagery comes into play, but it leaves cases where lakes were too small to be visible on e.g. Landsat (see our discussion on that) or simply not available yet. If people would not confirm that these were indeed lake outbursts, we disregarded the case (also because debris flows are generally more common). We have now added to the discussion in the manuscript regarding this issue.

L305: 'susceptibility': unclear how you estimated susceptibility. Please revise.

This section has been removed and only parts retained.

L305-306: Is there any hypothesis that says that lakes at higher elevations should be more prone to outburst?

As above, this section has been considerably changed. There is no such hypothesis and maybe this phrasing suggests it. In the retained text we have removed this as it doesn't add value to the statement following in the next sentence (that GLOFs are more frequent at lower elevations).

L306: 'less likely': I guess this is not a probabilistic assessment, so please avoid; 'a larger number of GLOFs happened [AT] low elevations'.

Removed completely, see point made above.

L320-321: interesting thought regarding the monitoring of debris flows. Is there any reference for that?

Unfortunately, we do not have hard evidence for this hypothesis, but this rather stems from the visible amount of literature (in Russian often) from the time before Independence and an eventual decrease afterwards. However (Medeu et al. 2019) argue that at least for Kazakhstan this can be attributed to successful mitigation efforts and we have added this line of thought here now as well (which remains impossible to prove).

L326: could be backed up with more references.

We have now elaborated with a focus on the KKH and the reference that discusses this development.

L327: 'increase of [reported] events'

Revised

L327-328: not exactly sure how the trends in the reported GLOFs and the research activity fit together. Please elaborate.

The increased interest has raised the bar to find and document more events, which we believe is linked – as there was more interest/funding available especially regional researchers wrote more detailed studies on individually events which eventually landed on the radar of the global community. Many of the events we document here for the first time for example were well known in the communities but had simply not been picked up by scientists to model for example and hence resulted in them not appearing in any previous records compiling events. We have rephrased the sentence.

Paragraph 4.3: Suggest to avoid this discussion. ESSD is a data-driven journal, and too much science can be a reason for rejection.

We agree with the reviewer and have completely removed this section, including the Figure.

Figure 6, second panel: shouldn't lakes with outburst have negative lake area change?

This figure has been removed. Note however that these are the overall changes over decades. Hence this also incudes area changes that may go beyond a GLOF, i.e. the lake could be 10 m2 at the beginning of the period, drain in between completely but refill again to 20m2 due to increased glacier melt or other contributing factors.

L365: 'risk': you mean hazard (i.e. probability of failure)?

Yes, revised.

Figure 7: Honestly speaking, I cannot see the added value of this figure. 'PZI/ RGI': please explain.

We believe that the figure is helpful in visualizing how this database can be combined with other datasets. We have however expanded on this now in the Discussion to make it clear. PZI is also explained.

Figure 8: Please elaborate how you obtained the Fahrböschung. What is the overall reason behind this analysis? I could not find any motivation in the introduction for this analysis. Why do you plot the data from Kääb et al. (2021) here? Please add axis labels to the inset and describe in the figure caption.

As for Figure 7 we believe this Figure is useful to show the potential of having both source as well as impact data available for the whole region for many events, also since this was not possible for previous inventories and is crucial for risk estimation. We agree however that the discussion of this was lacking and have now elaborated on the concept and why it plays a role here. It follows the same argument as in other studies (like (Kääb et al. 2021)) to evaluate a pattern in reach of these events.

L388: 'mean reach angle': all previous statements refer to the median?

Yes, but the mean is also given in brackets. We reported the median for the length and elevation drop, as the mean can be biased to the few outliers.

L400: Do you consider the same study period of Carrivick and Tweed (2016)? Their study has been published 6 years ago, so the database is shorter.

This is correct. However, the large difference stands. We have now clarified that in the text (i.e. over the same period we have 854 vs 300 deaths).

L423: Did your appraisal account for road disruptions? What is a 'ripple effect'?

Unfortunately, there is no appropriate documentation of impacts on roads, but we do mention that this is the biggest concern based on the reports collected in this database (see infrastructure impacts). Together with the change of wording of 'ripple' to 'knock on' effects, a common term when talking about climate risks we have now expanded on this explanation. As we note we also believe that this indirect effect of GLOFs could however be further investigated.

L426: are thee minimum and maximum values behind the mean?

Yes, we have clarified that now.

L447: It is not expected in ESSD that authors investigate these trends or provide deeper mechanistic insights. This phrase therefore can be deleted (also in L315 and L336).

Thanks, we have now removed these statements.

L454-456: I could not follow this statement, please revise.

Our argument here is that GLOFs in the region have received too much attention if you will. This is based on our experience looking at hazards more holistically. Generally, projects in high mountain and cryosphere hazards are always associated to GLOFs, hence also any large-scale financing of projects

is made available when it addresses GLOFs. This has caused local reports to immediately call every mass flow a GLOF, even if no lake whatsoever was involved. The high number of studies is a by product of this focus. Other hazards, like avalanches, debris flows, or other mass movements associated to permafrost change for example have hardly received any attention, even though they often had higher impacts. We have now revised this sentence considerably to make our argument clear and why this is important when critically reflecting on GLOF data.

L478: The uncertainty and completeness of this variable (and many others) is not assessed or discussed in greater detail. Suggest to extend the associated paragraphs.

We agree that this has so far been weakly discussed. We now provide an in-depth discussion of these issues in the section on 'Data structure and variables' (New section 2.2) and pick this up here in the Discussion again.

L496-499: Please add to the discussion how you would achieve that.

We have now expanded on this end of the discussion with more detail, proposing a way forward.

L500: The data availability statement usually comes before the conclusions.

Revised.

L575: Please use the published version of that paper.

Revised.

References

Erman, Alvina, Sophie Anne De Vries Robbe, Stephan Fabian Thies, Kayenat Kabir, and Mirai Maruo. 2021. "Gender Dimensions of Disaster Risk and Resilience." Washington D.C.: World Bank. http://hdl.handle.net/10986/35202.

Glazirin, Gleb. 2010. "A Century of Investigations on Outbursts of the Ice-Dammed Lake Merzbacher (Central Tien Shan)." *Austrian Journal of Earth Sciences* 103: 171–79.

Kääb, Andreas, Myléne Jacquemart, Adrien Gilbert, Silvan Leinss, Luc Girod, Christian Huggel, Daniel Falaschi, et al. 2021. "Sudden Large-Volume Detachments of Low-Angle Mountain Glaciers - More Frequent than Thought." *The Cryosphere* 15: 1751–85. https://doi.org/10.5194/tc-2020-243.

Kingslake, Jonathan, and Felix Ng. 2013. "Quantifying the Predictability of the Timing of Jökulhlaups from Merzbacher Lake, Kyrgyzstan." *Journal of Glaciology* 59 (217): 805–18. https://doi.org/10.3189/2013JoG12J156.

Mankoff, Kenneth D., Xavier Fettweis, Peter L. Langen, Martin Stendel, Kristian K. Kjeldsen, Nanna B. Karlsson, Brice Noël, et al. 2021. "Greenland Ice Sheet Mass Balance from 1840 through next Week." *Earth System Science Data* 13 (10): 5001–25. https://doi.org/10.5194/essd-13-5001-2021.

Medeu, A. R., V. P. Blagoveshchenskiy, T. S. Gulyayeva, and S. U. Ranova. 2019. "Debris Flow Activity in Trans-Ili Alatau in the 20th — Early 21st Centuries." *Geography and Natural Resources* 40 (3): 292–98. https://doi.org/10.1134/S1875372819030120.

Ministry of Planning Development & Special Initiatives. 2022. "Pakistan Floods 2022: Post-Disaster Needs Assessment (PDNA)." Islamabad, Pakistan.

https://www.undp.org/pakistan/publications/pakistan-floods-2022-post-disaster-needs-assessment-pdna.

Nie, Yong, Qiao Liu, Jida Wang, Yili Zhang, Yongwei Sheng, and Shiyin Liu. 2018. "An Inventory of Historical Glacial Lake Outburst Floods in the Himalayas Based on Remote Sensing Observations and Geomorphological Analysis." *Geomorphology* 308: 91–106. https://doi.org/10.1016/j.geomorph.2018.02.002.

Uddin, Kabir, and Mir A. Matin. 2021. "Potential Flood Hazard Zonation and Flood Shelter Suitability Mapping for Disaster Risk Mitigation in Bangladesh Using Geospatial Technology." *Progress in Disaster Science* 11: 100185. https://doi.org/10.1016/j.pdisas.2021.100185.

UNESCAP. 2022. "Background Paper for Regional Consultation on Facilitating Innovative Action on Disability-Inclusive and Gender-Responsive DRR." Bangkok, Thailand: UNESCAP.

Welty, Ethan, Michael Zemp, Francisco Navarro, Matthias Huss, Johannes J. Fürst, Isabelle Gärtner-Roer, Johannes Landmann, et al. 2020. "Worldwide Version-Controlled Database of Glacier Thickness Observations." *Earth System Science Data* 12 (4): 3039–55. https://doi.org/10.5194/essd-12-3039-2020.