Reply to Referee #2 (RC2)

General comment #1:

> *Hazard concept: The article is strictly focusing on GLOF susceptibility and using this term consistently throughout the manuscript. Nevertheless, I think regarding some aspects of the study, concepts and terminologies are mixed at some places. According to international standards from UNISDR, IPCC etc., hazard is a function of probability (of occurrence) and intensity (or magnitude). Susceptibility in turn 'is a relative measure of the likelihood (or probability) that a hazard will occur or initiate from a given site, based on intrinsic properties and dynamic characteristics of that site' and 'has an inverse relationship with stability' (GAPHAZ, 2017). I.e., susceptibility can be considered as probability of occurrence and is one factor of hazard. It is determined by conditioning factors (=inherent and more or less static factors) on the one hand, and triggering factors (=factors that directly initiate an outburst) on the other. The factors (predictors) analyzed in this paper, are limited (for good reasons!) to conditioning factors. In other words, the result of an analyses based on the parameters used in the present study, is mainly a lake stability assessment.*

We thank the reviewer for this observation and these definitions. However, we would have hoped for some guidance as to where exactly we might have mixed concepts and terminologies in this regard. The specific comments below seem not to pick up this issue. Perhaps some of the confusion arises from both qualitative and quantitative uses of the term "hazard". We echo the reviewer's comments on hazard and susceptibility in principle. Yet we wish to stress that our model is far from a stability assessment of moraine dams. The reviewer may concede that such appraisals frequently hinge on hard classes such as "stable" or "unstable". Even if such an appraisal would be (geo-)technically feasible, we would have needed to collate many more parameters on the internal structure and geometry of moraine dams, including grain-size distribution, presence and size of ice cores, the volume, width, height, and slope of moraine dams, pore water pressure in the dam, armouring of the outlet channel, presence and opening of tension cracks, rates of subsidence, and many others. Such parameters have been likely prone to change during our study period and are difficult to obtain for a single lake, and so even less feasible for the size of our regional study. To avoid an elusive feeling of stability, and hence, safety, we refrain to call our approach a dam-stability assessment.
Strictly speaking, our analysis estimates the **probability of correctly detecting historic lake outbursts from a set of predictors**. The referee may acknowledge that this probability is indeed a likelihood of GLOF outburst conditioned on reporting. In this sense one could see this metric as a "relative measure of the likelihood (or probability) that a hazard will occur or initiate from a given site" like the referee suggests. Our forecasting model in particular addresses this scenario.

> *In contrast to this, most of the mentioned regional glacial lake assessment approaches with more expert based, and probably subjective, parameter weightings, follow a hazard assessment approach, rather than a stability/susceptibility assessment.*

From the literature that we compiled in this and our previous work on GLOFs, we infer that very few, if any, of these so-called hazard assessments offer probabilistic metrics that satisfy the formal quantitative definition of hazard, as the referee clearly points out. During our review, we found that most of these studies deal with hazard in a qualitative or semi-quantitative way.

The reasons why these other studies consider factors like lake area or volume, or regional glacier mass balance, is not mainly because these factors directly influence lake stability, but because they have an impact on hazard potentials. Larger lake volumes (area is often used as a proxy for volume) and lake growth imply higher potential flood volumes, and therefore increase the hazard due to higher intensities, without affecting GLOF susceptibility. For similar reasons glacier masse balance is included in such models: Negative regional mass balances lead to glacier retreat and the formation of new and growth of existing lakes. Both processes increase the GLOF hazard potential in a region, but only have minor effects on GLOF susceptibility of individual lakes.

We agree with the referee in principle here. Yet we found it difficult to trace objectively any increases (or even changes) in hazard in the literature due a distinct lack of the necessary probabilistic metrics.

Further, unfavorable conditioning factors do not lead to a lake outburst immediately. It of course increases GLOF susceptibility, but requires still a triggering event to initiate an outburst. Clague and Evans (2000) and Emmer et al. (2020) present concepts about the timing of the causal chain of climate change, glacier retreat, glacial lake formation, and glacial lake outburst and conclude, based on empirical data from British Columbia and the Cordillera Blanca, that there is a lag between lake formation and outburst of up to several decades. The fact that a lake did not have an outburst event in the periods investigated in this study, does not automatically imply that the lake has a low GLOF susceptibility. It is indeed possible, that the lake is actually unstable (i.e. has a high susceptibility) but an outburst simply has not been triggered (yet).

We reiterate our point above and state that our method is set out to detect reported GLOFs. The reviewer's comment on dam stability is important and should be considered in geo-technical assessments, but is tangential to our objectives. Nowhere did we state that we wanted to quantify or estimate the stability of moraine dams. We also do acknowledge the concept of lag times between lake formation and outburst: every lake has a life span, but the question is whether an outburst needs to end it. This concept of lag time is thought provoking, but hinges on data similar to the models that we present here. One major advantage in our models is that we can fully capture the underlying uncertainties, something that we have so far yet to see for any lag-time model.

General comment #2:

Used data and parameters: Data availability for the entire study region is of course an important criterion for the selection of predicting parameters. But in addition to the parameters investigated in this study, there are candidates for other parameters which are often and successfully applied in other regional assessment approaches cited in the study, such as the Steep Lake front Area (SLA) developed by Fujita et al. (2013) and used by Rounce et al. (2016), or the topographic potential for rock or ice avalanches (cf. Allen et al., 2019), one of the most frequent GLOF triggers in High Mountain Asia. Considering this, I suggest to include more details about the selection of the predicting parameters.

We wish to refer the reviewer to the changes that we have made following the suggestions of referee #1. We rewrote large parts of Section 2.1 to make clear how we selected each predictor. For example, we note that:

- "Larger and growing lakes offer more area for impacts from mass flows originating from adjacent valley slopes such as avalanches, rockfalls, and landslides (Haeberli et al., 2017)."; - "A larger upstream catchment area has been associated with an increased susceptibility to GLOFs as more runoff from intense precipitation, together with glacier and snow melt, can lead to sudden increases in lake volume (Allen et al., 2019; GAPHAZ, 2017; Worni et al., 2012). ";
- "These readily available data on regional glacier-mass balances are proxies for other, less accessible, physical controls on GLOF susceptibility such as glacial meltwater input, either directly from the parent glacier or from glaciers upstream, as well as permafrost decay in slopes fringing the lake.";
- "Meteorological drivers entered previous qualitative GLOF hazard appraisals mostly as (the probability of) extreme monsoonal precipitation events: the Kedarnath GLOF disaster, for example, was triggered by intense surface runoff (Huggel et al., 2004; Prakash and Nagarajan, 2017). […] Elevated lake levels during the monsoon season also raise the hydrostatic pressure acting onto moraine dams (Richardson and Reynolds, 2000)."

Fujita et al. (2013) used the SLA approach to derive Potential Flood Volumes (PFVs) of Himalayan lakes as a proxy for GLOF susceptibility. We assume that PFVs are largely represented by our predictor lake area, given that larger lakes should produce larger floods. A major critique of the SLA concept is that lakes can have zero PFV despite large lake volumes. This issue was observed, for example, at Imja Lake in the Mt. Everest region, Nepal, that stores 78.4×10^6 m^3 of water surrounded by steep slopes (Haritashya et al., 2018). Fujita et al. (2013) also point towards an issue that the reviewer had cautioned against above (p. 1834): *"PFVs were simply calculated from the topography surrounding the moraine-dammed lakes and thus the robustness of the dam could not be evaluated. As the existence of ice within the damming moraine may alter the dam's vulnerability, understanding the distribution and degradation of permafrost will be an important factor for the further assessment of GLOF probability".* Furthermore, SLA depends on user-defined cutoffs, for example a 1-km search radius for steep slopes around lakes (Fujita et al., 2013), or a minimum slope threshold (Rounce et al., 2016). Such thresholds introduce additional subjectivity that we wished to avoid in our appraisal. Finally, errors in digital elevation models in high mountains remain unaccounted for in these slope-based metrics, though have been acknowledged for many years (Fujita et al., 2008). For example, Mukul et al., (2017) reported that *"vertical accuracy of the data decreases with increase in slope and elevation due to presence of large outliers and voids. Therefore, studies using SRTM data "as is", especially in regions like the Himalaya, are not statistically meaningful".* In summary, these findings motivated us to keep the influence of potentially error-prone model inputs at a minimum.

> *Then, the influence of overlapping time periods of the different data sets used should be discussed in more detail, as also mentioned in the review of A. Emmer (Emmer, 2021). In particular the fact that the lake area change period overlaps the period which is investigated for GLOF occurrence, in my view disqualifies this parameter to be considered, as actually discussed in L340-344.*

We again wish to refer the referee to our reply to referee #1:
The data on lake-area changes are not yet resolved on an annual basis for our study area, so that we had to resort to changes averaged over longer periods. However, we used our forecasting model to test whether changes in lake size between two observation periods had a credible effect on P_{GLOF}. Here we explored the weight of relative changes in lake area

between 1990 and 2005 to estimate the probability of observing GLOFs that happened in the subsequent period 2005-2018. In other words, we trained the model on GLOF data that **pre-date the testing data**, and thus offer a realistic and rigorous prediction and validation scenario. We reported in our original manuscript that "The weight of relative lake-area change in the 15 years before is ambiguous ($\beta_{A*} = -0.04^{+0.76}/_{-0.67}$) […]" (L245-246) and that "In the forecasting model, however, the influence of lake-area change remains negligible even for <50% HDIs." (L350). This indicates with 95% probability that relative lake-area change **before** the outburst is an inconclusive predictor. To further stress this result, we added the following sentences in the Discussion: "However, in the forecasting model, in which we tested whether differing data observation periods have any credible effects, the influence of lake-area change remains negligible even for <50% HDIs. We thus conclude that relative lake-area change before outburst is an inconclusive predictor. This result contradicts the assumptions made in many previous studies that assumed that rapidly growing lakes are the most prone to sudden outburst (Aggarwal et al., 2016; Bolch et al., 2011; Ives et al., 2010; Mergili and Schneider, 2011; Prakash and Nagarajan, 2017; Rounce et al., 2016; Wang et al., 2012)."

General comment #3:

> *Statistical significance: Bayesian approaches are certainly most suitable for this type of research question where a large number of lakes (3,390) had relatively few (31) GLOF events. But still this is a very limited data basis, in particular since for the Forecasting, the Glacier-mass balance, and the Monsoonality models, only 11 GLOF events were recorded in the relevant 2005 to 2018 period. Even more, these 11 events are split over four to seven groups, depending on elevation, region, or monsoonality. Over the western half of the study region, only 3 GLOFs are found. This leads to very few (often only 1 or 2) or even zero GLOF events per subgroup (cf. boxes for Hindu Kush, Karakoram and Western Himalayas in Fig. 7). I wonder, how any predictor weights can be found in these cases.*

The problem of highly imbalanced data (few GLOF reports out of several thousand lakes) was a major motivation for us to use Bayesian models. The low prior probability of detecting a reported GLOF can be compared directly with the posterior probability, a strategy that we showed in our original Fig. 9. Classical rare-events logistic regression penalises the model likelihood, and this step is done naturally via the prior distributions in the Bayesian setting. The low number of data points in some groups is even less of an issue in a hierarchical model, as this always draws strength across each group and the pooled model of all data taken together. The high posterior uncertainties tied to some model groups clearly underline the effect of fewer data points. To further underline these points, we added the following statement: "The small number of reported GLOFs introduces strong imbalance to our data, given that some regions, and hence levels, had few or no reported GLOFs. Although this would be problematic in most other modelling approaches, Bayesian multi-level models are particularly well suited for this kind of imbalanced training data (Gelman and Hill, 2007; Shor et al., 2007; Stegmueller, 2013)."

> *A very recent study from Zheng et al. (2021) on a slightly larger study region found evidence for a total of 215 GLOF events that presumable have happened since 1900, 176 thereof so far unreported. This does not contradict any of the data used here, but offers at least a potential alternative of a database with much more GLOF evidences (in turn posing challenges on the predictor data of course).*

We checked the study by Zheng et al. (2021) but found mostly GLOFs without timestamps that are difficult to reconcile with our predictors, some of which are averaged over specified time periods. Please also see our reply to Referee #1 with respect to the validity of predictors that change over time.

Detail comment #1:

> *L11: I suggest to include something like 'regional-scale' (hazard estimations), because at the level of individual lakes, there are many quantitative assessments available, including numerical modeling, geophysical measurements etc.*

We rephrased this sentence accordingly to: "Estimating regional susceptibility of glacial lakes […]."

Detail comment #2:

> *L21: Maybe change 'with respect to' to 'compared to'?*

We changed the phrasing as requested.

Detail comment #3:

> *L81: Indicate the version number of the RGI*

We added the version number 6.0.

Detail comment #4:

> *L112: Hydrostatic pressure acting on the dam depends mainly on lake depth, not area.*

We acknowledge that this is physically more correct, though we fail to see how any change in lake area or volume could not affect hydrostatic pressure eventually. We accordingly rephrased our statement to: "Lake area scales with lake volume and depth (Huggel et al., 2002), and growing lake depths increase the hydrostatic pressure acting on moraine dams, thus raising the potential of failure (Rounce et al., 2016)."

Detail comment #5:

> *L263: The statement that upstream catchment area is well correlated with lake area is not clear to me. This needs further explanations of references. Also I do not understand why lake area is replaced by upstream catchment area in these model (Glacier-mass balance and Monsoonality), but not in others. This requires some more explanation.*

We thank the reviewer for this suggestion and added more details on our predictor catchment area. We refer to our replies to referee #1's detail comments #12 and #13:
To clarify our approach, we added more details on our predictor catchment area in the new Table 2. We explain our choice of this predictor and why we use it instead of the static lake area *A* in the glacier-mass balance and monsoonality models: "We also tested the impact of upstream **catchment area** *C* (m^2) on GLOF susceptibility. A larger upstream catchment area has been associated with an increased susceptibility to GLOFs as runoff from intense precipitation as well as glacier and snow melt can lead to sudden increases in lake volume (Allen et al., 2019; GAPHAZ, 2017)."

We find that catchment area *C* has a strong linear correlation with lake area *A* (Pearson's correlation coefficient of 0.446), such that we preferred *C* over *A* in two of our models, as *C* is constant at the scale of our study.

Detail comment #6:

> *L285: In the Forecasting and Glacier-mass balance models, A* represents lake-area change between 2005 and 2018. Is A* here also referring to this period (and not 1990 – 2018, as written)? If so, please correct, if not, another symbol should be used (ΔA?).*

Our predictor relative lake-area change A^* (not to be confused with net lake-area change *ΔA*) is calculated for three different time windows: 1990 to 2005 in the forecasting model, 2005 to 2018 in the glacier-mass balance model, and 1990 to 2018 in the monsoonality model. In order to avoid confusion and correctly refer to each respective time interval of relative lake-area change, we now assigned each with its own symbol: relative lake-area change between 1990 to 2005 is A^{*a}, relative lake-area change between 2005 and 2018 is A^{*b}, and relative lake-area change between 1990 and 2018 is A^{*c}. We explain this notation in our new Table 2 and in the respective model descriptions in the Results section 3.

Detail comment #7:

> *L321/Fig. 9: Why are the log-odds ratios negative for the first (few) lakes? Would be interesting to describe in the text.*

The negative log-odds ratios indicate lakes for which the posterior probability of a reported GLOF is lower than the prior probability. To further clarify this, we rephrased this to: "A positive log-odds ratio means that we obtain a higher posterior probability of attributing a historic GLOF to a given lake compared to a random draw. Negative log-odd ratios indicate lakes for which the posterior probability of a reported GLOF is lower than the prior probability."

Detail comment #8:

> *L323/324 (Caption Fig. 9): '...in the past four decades' applies only to the lakes in the x-axes of (a) and (c), for the other panels it's 2005-2018. I suggest to replace this with 'in the period 1981 – 2018 (a and c) and 2005 – 2018 (b and d-h). (Or change panel letters, see suggestion below).*

We changed panel labelling and added information on used time periods for lake subsets on the x-axis to the figure caption accordingly.

Detail comment #9:

> *Table 1: This is a pretty large table for only presenting the 6 predictor parameter selected for this study. I suggest to present the 6 parameters used here in separate table, giving some more details as well. (By the way, I think dam type could be ticked as well, at least a tick in brackets. As only moraine-dammed lakes are investigated here, this criteria is inherently considered). If the authors wish to keep having a table with other potentially relevant parameters for GLOF hazard assessment, this could be done in a more compressed format. But in this case, further geotechnical and geomorphic parameters would need to be included, such as permafrost conditions, lithology, seismicity, etc. The annex tables of the GAPHAZ guidelines (GAPHAZ, 2017) might give some indications for this.*

To meet the requests from both referees, we split our former Table 1 into two separate tables with an overview table listing the predictors for HKKHN lakes described in the literature (now labelled as Table 1) and a comprehensive table listing our predictor choices (now labelled as Table 2). We added a number of additional parameters to Table 1 and more details (used notation and selection reasoning) to Table 2. Now the indicator "dam type" is also ticked in Table 1.

Detail comment #10:

> *Fig. 1: According to the caption, white triangles represent GLOFs since 1935. But as the study only deals with GLOFs that have occurred on the periods 1981-2018 and 2005-2018, respectively, only these should be shown here. Preferably with two colors, one for 1981-2005 and another for 2005-2018 to discriminate these to reference data sets. Please also indicate the spacing of the lake bubbles.*

We thank the reviewer for this suggestion and changed the figure accordingly.

Detail comment #11:

> *Fig. 3: Blue-green combinations are hardly readable in the bubbles. I can see it in the text, slightly see it in the middle ('mulit-level') bubble, but do not see any green in the right ('many models'). Colors to be adjusted.*

We modified Fig. 3 to make the principles of multi-level modelling more clear. We also chose a green-purple-yellow colour combination to improve contrast.

Detail comment #12:

> *Figure 4: I suggest to sort sub-groups from highest (top) to lowest (bottom) in (a) and (b), West to East (or East-West) in (c) and highest monsoonality on top to lowest monsoonality in (d). (Same as ordering in Figs. 5-8).*

We appreciate this suggestion. However, the point of the figure is to highlight better the ranked deviations of the group-level coefficients from the pooled mean (at the bottom of the stack). We believe that this ranking allows a better visual assessment of which groups deviate most from the pooled means.

Detail comment #13:

> *Figs. 5-8 (general): In none of the figures I can see the middle (blueish) line. I only see the purple and orange lines. Similar for the color shades, I guess I only see the purple and the orange and the overlap of the two. Is this middle line represented in the Figures? If so, please adjust coloring, if not please add (or remove from the legend). For all figures it would be very nice to also have a panel for the pooled data, similar to Fig. 4.*

The middle line is grey in the original Figs. 5-8 and lies between the orange and purple lines. We acknowledge that this may be hard to decipher and changed the colour scheme to provide more contrast. Adding a panel for pooled data is a good suggestion and we are to amend the figures as requested.

Detail comment #14:

> *Figs. 5 and 6: It would be helpful to indicated elevation bands in m a.s.l.*

This is a good suggestion and we are to amend the figures as requested.

Detail comment #15:

Fig. 9: To me it would make more sense to number the TP a-d and the TN e-h.

We thank the reviewer for this suggestion and changed the figure accordingly.

Detail comment #16:

Fig. 10: In the legend (e) (the letter e is not needed in my view) swap ordering, that a is on top and d at the bottom, as in the main panels. Ad '%' to the numbers at the bottom. In the panels it would be helpful to include the locations with a recorded GLOF (for 1990-2018 in (a) and 2005-2018 in (b), (c) and (d)).

We thank the reviewer for this suggestion and changed the figure accordingly.

References cited in this Reply

Allen, S. K., Zhang, G., Wang, W., Yao, T. and Bolch, T.: Potentially dangerous glacial lakes across the Tibetan Plateau revealed using a large-scale automated assessment approach, Sci. Bull., (April), doi:10.1016/j.scib.2019.03.011, 2019.

Fujita, K., Suzuki, R., Nuimura, T. and Sakai, A.: Performance of ASTER and SRTM DEMs, and their potential for assessing glacial lakes in the Lunana region, Bhutan Himalaya, J. Glaciol., 54(185), 220–228, doi:10.3189/002214308784886162, 2008.

Fujita, K., Sakai, A., Takenaka, S., Nuimura, T., Surazakov, A. B., Sawagaki, T. and Yamanokuchi, T.: Potential flood volume of Himalayan glacial lakes, Nat. Hazards Earth Syst. Sci., 13(7), 1827–1839, doi:10.5194/nhess-13-1827-2013, 2013.

GAPHAZ: Assessment of Glacier and Permafrost Hazards in Mountain Regions: technical Guidance Document. Standing Group on Glacier and Permafrost Hazards in Mountains (GAPHAZ) of the International Association of Cryospheric Sciences (IACS) and the International Per, Zurich, Lima., 2017.

Gelman, A. and Hill, J.: Data Analysis using Regression and Multilevel/Hierarchical Models, Cambridge University Press, New York., 2007.

Haritashya, U. K., Kargel, J. S., Shugar, D. H., Leonard, G. J., Strattman, K., Watson, C. S., Shean, D., Harrison, S., Mandli, K. T. and Regmi, D.: Evolution and Controls of Large Glacial Lakes in the Nepal Himalaya, Remote Sens. , 10(5), doi:10.3390/rs10050798, 2018.

Huggel, C., Kääb, A., Haeberli, W., Teysseire, P. and Paul, F.: Remote sensing based assessment of hazards from glacier lake outbursts: a case study in the Swiss Alps, Can. Geotech. J., 39(2), 316–330, doi:10.1139/t01-099, 2002.

Mukul, M., Srivastava, V., Jade, S. and Mukul, M.: Uncertainties in the Shuttle Radar Topography Mission (SRTM) Heights: Insights from the Indian Himalaya and Peninsula, Sci. Rep., 7(February), 1–10, doi:10.1038/srep41672, 2017.

Rounce, D. R., McKinney, D. C., Lala, J. M., Byers, A. C. and Watson, C. S.: A new remote hazard and risk assessment framework for glacial lakes in the Nepal Himalaya, Hydrol. Earth Syst. Sci., 20(9), 3455–3475, doi:10.5194/hess-20-3455-2016, 2016.

Shor, B., Bafumi, J., Keele, L. and Park, D.: A Bayesian multilevel modeling approach to time-series cross-sectional data, Polit. Anal., 15(2), 165–181, doi:10.1093/pan/mpm006, 2007.

Stegmueller, D.: How many countries for multilevel modeling? A comparison of frequentist and bayesian approaches, Am. J. Pol. Sci., 57(3), 748–761, doi:10.1111/ajps.12001, 2013.

Zheng, G., Bao, A., Allen, S., Ballesteros-Cánovas, J. A., Yuan, Y., Jiapaer, G. and Stoffel, M.: Numerous unreported glacial lake outburst floods in the Third Pole revealed by high-resolution satellite data and geomorphological evidence, Sci. Bull., 2021.