Review of – **Escalating Glacial Lake Outburst Flood Risks in the Third Pole Under Future Climate Warming**

This paper presents a new continental-scale assessment of GLOF risk across the Third Pole. It integrates a long-term glacial lake inventory (TPGL), a susceptibility model (GAMS-GLOF), hydrodynamic flood simulation (HiPIMS-GLOF), and exposure analysis. The ambition of an end-to-end quantitative workflow covering the entire Third Pole is impressive, and the lake inventory and hydrodynamic modelling components are important contributions. The final output, a future GLOF risk map, is impressive. However, from a detailed reading of the methods, I am not convinced that it is robust or useful in its current state. As such, I recommend that this manuscript be rejected at this stage.

There is an overall issue throughout the manuscript of substantial assumptions being made, but the implications of these not being acknowledged in the text (and in some cases not even in the methods or supplement). In a project that strings together multiple approaches like this it is critical to be clear, transparent, and complete in the acknowledgement of uncertainties and assumptions. Without this, the results cannot be effectively evaluated. This is a particular issue here with the ambition that the outputs for a "transferable tool for early-warning design, climate-resilient land-use planning, and transboundary hazard governance" – the results here risk being rather misleading if taken at face value by decision makers. Certain uncertainties are unavoidable – this is well recognized – but this manuscript as currently written tends to gloss over substantial challenges and overrepresent confidence. Frankly, given the experience of the authorship team, these issues should have been discussed further before this manuscript was sent for review.

Specifically, some issues of concern are:

1. The GAMS-GLOF model has a very low precision with a positive predictive value of only 0.16. This suggests that for every true prediction we'd expect more than 6 false positives. Of course, this is to some degree an expected challenge in an unbalanced dataset such as this, but it still raises major questions about the next steps. There is no mention of this in the text, and the flood routing is run on all high and very high susceptibility cases without recognition of the inevitably high false positive rate. As such, this issue propagates through the full workflow. It is possible to provide useful results withing these constraints, but the presentation of AUC alone "The model achieved an Area Under Curve (AUC) of 0.92, indicating outstanding predictive capability" is misleading.
2. The reasons for selecting a 500m buffer zone are not explored. Indeed this seems to negate the point of using a sophisticated flood routing model. The exact procedure followed is somewhat unclear, but a simple 500m expansion of the flood extent would include many areas that would never be flooded (and even if considering secondary impacts, be a major overestimate). Why was the 500m value chosen? The reason "*This buffer distance was selected based on an evaluation of floodplain boundary uncertainty and the potential for secondary hazards.*" is not sufficient. Does this indicate that you believe the food routing model's lateral extent is uncertain to within 500m? This would indicate it is highly

unreliable (and is not in line with past work). An uncertainty-aware version of the model run would be best – these results end up not much better than a simple buffered flowline.

3. There are various stages in the workflow with well known and documented assumptions and uncertainties, but which are ignored or downplayed here. Cumulatively, they make it very challenging to evaluate the true quality and meaning of the final outputs. For instance, the lake detection database is subject to substantial uncertainties, particularly for smaller lakes. The workflow to identify future lakes has well-known limitations, particularly when considering individual polygons. Effects of sedimentation – both in lake creation and loss – are not well accounted for. The rates of glacier retreat are also subject to many uncertainties. The GAMS-GLOF model has challenges as mentioned above. The initial conditions for the GLOF model runs make a large difference to maximum discharge and runout distance, with certain simplifying assumptions made but not explored here. The runout polygons are buffered, with the reasoning behind this unclear. Uncertainties and temporal changes in exposure elements are not considered. Taken together, these cumulate into a large and unquantified uncertainty that makes the final risk maps difficult to evaluate or use. Overconfident scientific predictions can be as problematic for DRR planning as the absence of a robust evidence base.

Due to the issues above, this paper will require a major re-write before it can be considered for publication and I do not provide detailed in-text comments. I note a few other poorly worded sentences that should be avoided, for instance the claim that “Our simulations unequivocally confirm that global warming continues unabated” – your work does not do this.

Overall, this work presents some important novel work addressing a key question for risk in high-mountain Asia. However, poor evaluation of the uncertainties and assumptions in this complex multi-stage workflow mean that the final product does not stand up to the claims made. I thank the authors for their hard work and wish them best of luck in their future changes