

The manuscript “Predictive understanding of socioeconomic flood impact in data scarce regions based on channel properties and storm characteristics: Application in High Mountain Asia(HMA)” by Khanam et al. used LYI and ML methods to evaluate and predict the flood impacts and risk due to precipitation in HMA. This work is first time to evaluate socioeconomic impacts of flood hazards in data scarce region. However, it is not good writing. The structure is not reasonable. And the XGboosting tools is not clear to solve what? Thus, I would suggest it should be major revision.

We want to thank the reviewer for the insightful comments and recommendations. Following the suggestions, we will revise the manuscript. We will try to explain the methodologies clearly and improve the writing. Detailed responses to the reviewers’ comments are added below.

Note: Below is our response (italics) to each comment (regular font) from the reviewer

General comments:

- In HMA there are also GLOF which risk the human being and infrastructure. If possible, please include evaluating the socioeconomic flood impact.

Response: *We thank the reviewer for this suggestion. In general, this paper focuses on fluvial and pluvial flooding, and we will make this clearer in the introduction. GLOFs in general are triggered by glacial melt but here we focus on a climatic driver of the flooding that is related to rainfall. Addressing the damages due to GLOF is separate from the scope of this study.*

- Data-Scarce regions should be clear (which data or which type of data). In HMA, population is scarce. And Socio activity is also low.

Response: *We thank the reviewer for this comment. We will revise this part of the manuscript and rephrase our statements to avoid confusion about data scarcity. Regarding population and activity, we use population density in our work as a proxy of exposure, which varies across the region.*

Specific comments

1. Section 2.2 methods. This title is not reasonable. Is section 2.3(Machine learning model) methods? In addition, the dataset and methods in this section should be divided, for example, 2.2.4 exposure(population) is datasets.

Response: *Thank you for the suggestions. We will try to reorganize and revise the sections as per the reviewer’s comments.*

2. Line 135 Why is it classified by LYI values(<2, 2-3, and >3).

Response: *We have classified the LYI as a basis for comparison across all the watersheds and periods. The three groups correspond to <100, between 100-1000, and over 1000.*

3. Line 217 While XGBoosting is ..., this sentence is incomplete.

Line 217 this section (machine learning model) is a little difficult to understand the role that it plays.

Response: *We will revise this part of the manuscript thoroughly to clarify the methodology and explain the reasonings plainly.*