

This paper “Snow Avalanche Frequency Estimation (SAFE): 32 years of remote hazard monitoring in Afghanistan” attempts to produce inventories of avalanche debris using Landsat optical satellite imagery in late spring when snow, bare ground and water are easily distinguishable. The concept of using a long time series of remote sensing data to identify hotspots of avalanche deposition zones and trends in their spatial occurrence is good, but there are many pitfalls with the overall implementation and communication of the work which reduces the impact.

1. The paper requires some major restructuring of the content, starting with the introduction. Throughout the paper I found that information was in the wrong order and/or wrong section. Results were presented already in section 2 (eg. Table 2) and discussions were being made in the results section. This makes the work difficult to follow, even with the flow chart provided. Moreover figures are wrongly labeled (Fig. 10) and have unsatisfactory captions or text to explain what is being shown or how they were produced, color scales are not constant making figures hard to compare (Fig. 6-8).

*Dear reviewer, the authors are very thankful for your comments on our paper. For now we address your comments here, but we are already working on the new version of the manuscript based on your points. Regarding the order of the information, you must be referring to the validation that we indeed presented in the methodology section. We will move it to the results section based on your comment. Regarding Figure 10, we will improve the caption and ensure that the pixels on this map are the pixels where significant temperature trends occurred. As for the colors of figures 6-8, it is unclear exactly what you are referring to here. If you mean that we used different discretizations for each of the 4 categories in Fig. 7 (i.e., different binning of total avalanches per village), as well as for number of avalanches per km of road (Fig. 8), then if we normalized each of these it would not clearly show where the major impacts were – a key point of our paper. Thus, we are reluctant to change this as it would obscure the importance of the avalanche impacts.*

2. It seems to me that the authors are basically identifying late season snow patches in valley bottoms close to rivers which they are assuming to be avalanche deposits. This is made quite straightforward by the fact that the regions of interest are snow-free and snow is easily distinguishable by higher NDSI in the Landsat images compared with bare ground or water. This just reduces the problem to a simple thresholding and classification of image pixels into 3 classes, and I fail to see what is state-of-the-art in this approach. Moreover the authors have employed MODIS data to identify the snowline in order to select the dates and regions which are snow-free. MODIS has poorer spatial resolution than Landsat, so why not just use the Landsat data to identify the snowline? I can't see any value in using MODIS vs. Landsat for this purpose.

*Indeed, the NDSI reclassification approach of SAFE is straightforward but the date as well as the region of interest are the key parameters in this model, and to our knowledge, no other studies have adopted this approach before. In the introduction, we reviewed the literature related to avalanche detection using optical, radar, and Lidar data, but none of these studies used the NDSI as we did in SAFE.*

*Regarding the snowline extraction, we used MODIS because of its coverage and its ease of application. Landsat archives can certainly provide snowline maps with higher resolution but the amount of data to be extracted for this purpose is much greater. Moreover, the cloud coverage on Landsat images presents a greater challenge compared to MODIS because there are more tiles to merge from different times, and the coverage is smaller*

than MODIS. MODIS was used to separate highlands from lowlands across the entire study area and a coarser resolution was acceptable for this purpose.

3. Throughout the paper the authors emphasise that the approach is based on Landsat data and the use of the google earth engine because it should be used in areas where internet connection is poor. However, they also highlight that the main end-users of such a dataset are stakeholders and decision makers. Are these stakeholders and decision makers likely to be located in remote mountain villages or the main cities (where internet connection is presumably good)? Are local villagers in these mountain environments really likely to be making use of this dataset? I find it hard to believe that knowing where a large avalanche deposit has occurred several months prior to its detection is likely to be of interest to these people.

*Thank you for those comments. It is true that decision makers are potential users of SAFE, but academics and local research institutes can be potential users as well. There is no restriction of use of this model, and this is why SAFE was implemented in Google Engine, to be used in Europe, Central Asia, Andes, in big and well-connected cities, in small cities and villages in remote areas or wherever someone wants to and can use it. That is the point of open-access scripts such as SAFE. Stakeholders and decisions makers in remote mountain villages must not be excluded from this process and, in the mountains of Central Asia and nearby areas, in particular, these stakeholders are very interested and play an active role in hazard monitoring, mainly because of road blockages in winter. The road connections (and therefore the supplies in food, energy, medical supplies, and other items) to these villages are highly affected by snow avalanches and landslides. Hence, there is a critical need for local decision makers to have hazard frequency maps based on long-term data. Moreover, the remoteness of these villages where local stakeholders reside does not mean that they do not have internet access or competent personal to run SAFE, and we know by experience that such models can be easily run in small towns of high Badakhshan in Afghanistan or Tajikistan where this study was conducted. We could have decided to design SAFE as a toolbox (potentially for income) using expensive software, but this would not be useful in the local context where most of the agencies and stakeholders cannot afford expensive software or programs (and Google is a widely used browser that makes Google Engine a user friendly-interface for any stakeholder). Please remember that the high mountains of Central and East Asia are among the poorest regions of the world and residents are very vulnerable to these hazards. This is our primary target area for SAFE, although it certainly can be used in other mountain regions.*

4. As pointed out by reviewer 1 the classification of avalanche size seems quite arbitrary and does not have much meaning when it is being detected late in the season after it has already partly melted out. It would be more meaningful to show for example a histogram of the avalanche size to show what is being detected rather than applying some random size classification to the detected deposits.

*The size classification was actually based on the distribution of our “avalanche surface areas” as shown in Figure 5a. We thought about classifying our deposit zones based on EAWS, 2018 classification, but our surface areas did not match this universal classification. Additionally, it was stated in review 1 that snow avalanches sizes must be classified by volume; however, the two datasets published on EnviDat (<https://www.envidat.ch/#/metadata/satellite-avalanche-mapping-validation>) were actually classified by area (m<sup>2</sup>), therefore we retained our size classification based on surface area.*

5. Inconsistent terminology. Avalanche debris/deposits are referred to as "snow packages", "snow patches", "avalanche depositional" in the paper. The authors should use the correct term and use it throughout.

*Thank you for this comment. In the paper we had used the term "depositional zone", but we now made it more consistent in the manuscript.*

6. Poor validation. In section 3 the authors state that over the 32 years of data analysed they identified around 810,000 avalanche deposits using their dataset. However for the calculation of POD and PPV as shown in Table 2 they have only used 158 deposits observed using Google Earth images. Moreover they do not describe how the validation data were identified (was this done visually or was there some other algorithm used to detect them in these images?). Overall this does not come across as a satisfactory validation dataset with which to evaluate their detections.

*Thank you for your comment. We used 158 snow avalanches because those were visible in some regions during specific years of Google Earth images as explained on line 243: "A total of 158 snow avalanche depositional zones were easily identified in the riparian buffer zones on Google Earth images in 2001, 2003, 2015, 2017, and 2019". And Google Earth images were indeed used to verify the locations of the avalanches predicted by SAFE. The lack of Google Earth imagery over this 32 year period restricted the number of avalanche deposits we could assess.*

*Moreover, in order to delve into the validation, we are currently conducting the comparison between outlined snow avalanches using SPOT-6 images and SAFE results in Switzerland, as recommended by reviewer one.*

---

**NOTE: We now incorporate Landsat-9 images to SAFE model, which improves the coverage of Landsat archives worldwide. This will be added into the manuscript (Methodology section).**