Review of the manuscript *Avalanche danger level characteristics from field observations of snow instability* by *Jürg Schweizer, Christoph Mitterer, Benjamin Reuter, and Frank Techel.*

https://tc.copernicus.org/preprints/tc-2020-350/

General comments

The paper is well written, has a clear scope and analysis, and presents new results of scientific and practical value. It is an important contribution towards improving avalanche danger assessments and forecasting. It helps scientists and practitioners to assess and define avalanche danger, with practical implications ranging from developing machine learning tools to crowd-sourcing and interpreting signs and observations in the field. It also sheds light on the timing and combination of avalanche problems (new snow and persistent weak layer) based on observations, and provides directions for further research. The language, figures and tables are of high quality and the flow of the manuscript is sound and logical.

I recommend the editors to publish the paper after minor revisions.

Specific comments

Here are specific comments, which could be used to improve the manuscript:

- Not only stability, but also distribution/frequency of instability *and* the avalanche size affect the LN data and thus the analysis. The effect of not accounting for this on the results could be described further in the discussion and mentioned in the conclusions, if possible
- The data comes from a limited geographical area, and the study would ideally include data from other snow climates and regions of the world. I guess such data are difficult to come by, but it could be a recommendation to include more data in future studies (e.g. explain to other scientist how to collect and share data in a common data set for an analysis of more a regional or even continental/global perspective). It would also be interesting to know which other observations the authors would recommend for future studies, to improve the observational basis for quantifying avalanche danger (properties of the weak layer or the slab?)
- The introduction could explain how this study adds to – and differs from – the studies by Techel et al. (2020) and Schweizer et al. (2020)
- In chapter 3.1.2, could you include statics for quality of fracture plane (smooth, rough, irregular)?
- I would suggest reversing the x-axis in the figure in the appendix (have very poor to the left and good to the right) to align with the other figures in the manuscript, and to insert this figure (as well as its results) in the results chapter. It could also be beneficial to add a paragraph discussing, comparing or summarising the results of the three different scales used for stability (categorising stability into 3, 4 and 5 classes)

Technical corrections

Now follows specific comments, with reference to line numbers in the manuscript:

- #30-37 This is a description of a regional forecast, I suggest you add the word “regional” somewhere, and briefly describes how regional warnings are different from other forecasts (object-based / slope-scale) and, if possible, if this study may be valuable also to slope-scale assessments/forecasts
- #32 I could not easily find the EAWS 2019 in the reference list (it is on #570)

- #39 Could you improve the English of the sentence starting with “Even”?
- #59-60 Consider rewrite and simplify the first sentence, e.g. “but it was not possible to distinguish 2–Moderate from 1–Low”
- #84 Explain why you selected dry snow conditions only (dry slab selection could also be explained in #113)
- #104 Could you improve the definition of the “adjacent layer” and explain when it is above and below the failure interface?
- #108 Explain why you derived three stability classes her, while using five classes above (#95) and even 4 classes in the appendix
- #171-174 Please simplify this description
- #242 Please explain the acronym “RF” (presumably regional forecast)
- #245 Please clarify: Did the observer estimate the LN for the same region / area as the RF or is LN valid for a smaller area than RF and thus probably biased towards lower danger levels (presuming that RF is the highest danger level in the region)?
- #277 I presume “even” could be deleted (or is the sentence incomplete?)
- #367 Could you improve the sentence “the split variables and values were plausible”?
- #436 Could you improve the sentence “only in 1 out 5 cases differences in snow depth were indicative of snow instability”?
- #523 Please explain how/why this poses a challenge to forecasting?

Rune Engeset, 24 January 2021