Dear Pascal, Karl, and Rune,

Thank you for your valuable feedback, which we have addressed in the revised manuscript.

Please find below a detailed point-by-point response showing the changes we made. Here, we only refer to those remarks that either required some changes or that haven't been addressed in the first reply. In addition, we first respond to the remarks by editor Pascal Haegeli.

We hope that the revisions made address the comments and suggestions in an appropriate way,

Frank Techel, on behalf of all co-authors

Remarks by Editor Pascal Haegeli

Line 95+: It might be easier to just refer to local time as the time zone does not really matter that much.

We changed to **local time** *throughout the manuscript.*

Line 260: Your choice of using 1, 2, 3, and 4 for representing the danger rating levels for your calculations implies that you are interpreting the danger scale as a linear scale. In the same paragraph, you state that you tested exponential growth patterns, you discarded them because of their high correlations with the linear pattern. While I understand the simplification, and I suspect that it does not make a big difference in the results, it seems a bit unfortunate to treat the danger scale as a linear scale when we know that it is not as shown in your Figure 1 and some of the references you cite.

We are interested in the expected danger level rather than the danger potential. The latter increases (likely) in an exponential fashion, as could for instance be derived from Figure 6a and b (we actually tested this approach as well). However, as the expected value of the predicted danger level is of interest, we consider weighting of danger levels exponentially as not correct, as this would increase the relative importance of higher danger levels compared to lower danger levels. In the revised manuscript, we added a sentence to emphasize that we are interested in the expected value of a discrete random variable, and not the danger potential:

"Danger levels are rank ordered. The absolute increase in danger from one danger level to the next is unknown. To derive the expected danger rating D_{model}*, we used the respective integer values of the four danger levels 1 (low) to 4 (high) (w:= {1, 2, 3, 4}).* **This approach is in line with our interest in the expected value of the danger level, a discrete random variable, rather than the danger potential.** *However, to address the uncertainty related to w,* **and its impact on the results**, *we tested (w:= {1, f, f^2, f^3}) for various f, as for instance for f = 1.5 or f = 5. The resulting* D_{model} *vary in absolute values but are highly correlated (Pearson correlation coefficient for these two cases r = [0.91, 0.99]).*

Line 342: Should this be 1,000 km2? Everywhere else you talk about counts per 1,000 km2, and this is the only time 10,000 km2 shows up.

It is probably a matter of preference, but we used 10000 km^2 *on purpose, as this allows us to show numbers with at most one decimal place. For instance, in Figure 7b, values for avalanche size ≥ 3 per*

1000 km^2 would be 0, 0.02, 0.04, 0.07, 0.11, … rather than 0, 0.2, 0.4, 0.7, 1.1 per 10000 km^2. We believe the second notation is easier to read.

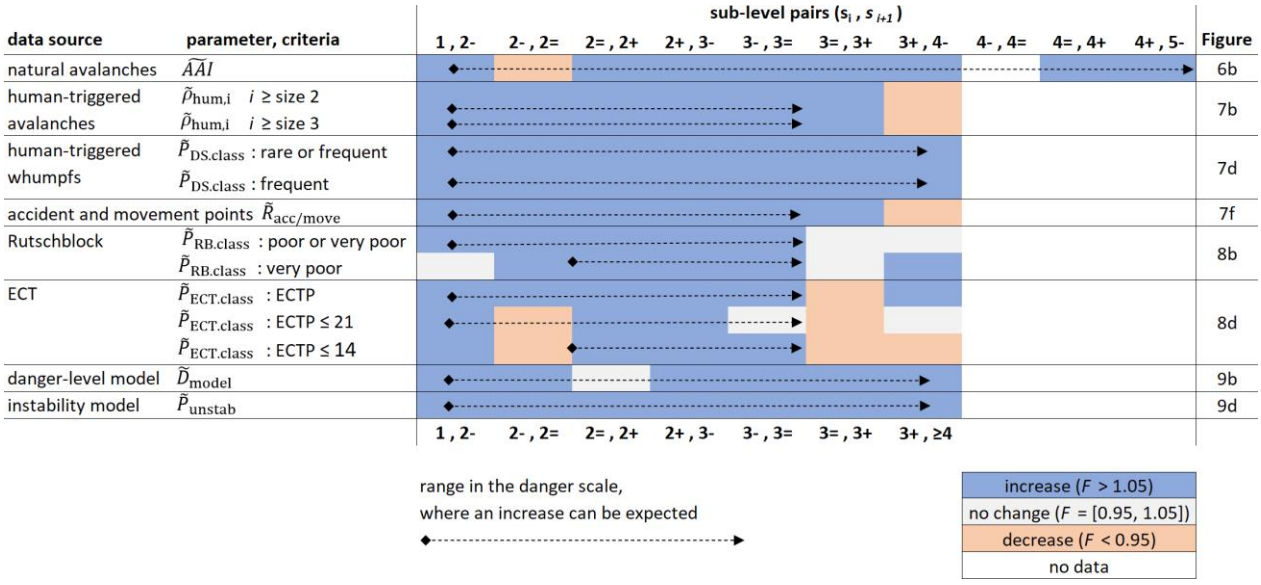
Line 355: Missing d in human triggered

Changed.

Table 4: I like this overview of the results, but because of the shading and the graphic elements, it will be treated as a figure and not a table. I wonder whether it would be possible to come up with a design where the analysis of all data sources would be presented on a grid with the same horizontal axis and the figure column for all rows could be aligned. This might make the figure less busy and easier to read.

We had a shaded table/figure in a previous manuscript (Hutter et al., 2021). At the time, the Copernicus Editorial Team treated the table/figure as a table, although we provided the table as a figure file. We will leave it up to the editorial team whether Table 4 will be referred to as a table or a figure.

We considered your recommendation to change the layout of the figure/table. Unfortunately, we didn't find a different graphical solution that would be an improvement to the current layout. Plotting F on the same y-axis does not really work either as the F-values vary considerably between the different variables, and the figure would have become even harder to interpret. Moreover, there are 13 parameters, and results are summarized for the danger level and the sub-level. However, we made some changes to the figure/table as shown below, which hopefully make it easier to read:



Reference: Hutter, V., Techel, F., and Purves, R. S.: How is avalanche danger described in textual descriptions in avalanche forecasts in Switzerland? Consistency between forecasters and avalanche danger, Nat. Hazards Earth Syst. Sci., 21, 3879–3897, https://doi.org/10.5194/nhess-21-3879-2021, 2021.

Line 511: Avoid short forms like "can't" or "doesn't" in scientific writing.

We adjusted throughout the document.

General: Please make sure that your formatting is consistent throughout the entire manuscript. (e.g., in-text references to figures and tables are abbreviated or not, formatting of large numbers (10'000 or 10000), order of test statistics and p-values, etc.).

We checked throughout the manuscript for consistent formatting (Tab. → Table, Fig. → Figure, formatting of larger numbers → 10000, order of test statistics: F or ρ_s now always precedes the respective p-value, where we want to show this).

Remarks by Karl Birkeland

L2: I'm not sure I would say that an avalanche forecast "terminates" with assigning the avalanche danger. This is because after assigning the avalanche danger you still have the issue of how to communicate all the nuances to the public in the written avalanche forecast.

Changed to: "*In the case of public avalanche forecasts, this* **assessment** *process terminates in an expert judgment concerning summarizing avalanche conditions by using one of five danger levels.*

L4: I think this is true for some avalanche forecasts where things are simplified quite a bit. But the nuances of the conditions can be captured in other ways than the danger rating (forecast discussion, videos, etc.).

Changed to: "This strong simplification of the continuous, multi-dimensional nature of avalanche hazard allows for efficient communication but inevitably leads to a loss of information when summarizing the severity of avalanche hazard."

L13: I always treat "data" as plural. So, this would be "data are" rather than "data is". I just looked it up and I believe that used to be pretty standard, but it sounds like some people are now using "data" as singular as well. I'm not sure what NHESS prefers, but if they prefer it to be plural you may need to change this throughout the paper.

We checked and adjusted throughout the manuscript.

L31: "for each identified avalanche problem type". Note: I think it is important to clarify that this assessment is done for all avalanche problem types, as per the CMAH.

Changed to: "This requires assessing the three factors contributing to avalanche hazard **for each identified avalanche problem***…"*

L48: Here I would suggest adding a sentence saying that forecasters attempt to bridge this gap between a continuous phenomena and a discrete number using the text of their avalanche forecasts. Then, instead of starting the next sentence with "In addition", I'd substitute in "Regardless"

Changed to: "Forecasters attempt to bridge this gap between a continuous phenomenon and a discrete level using the narrative part of the avalanche forecast."

L 76: "are simply asking whether or not there is a"

*Changed to: "***Specifically, we investigate whether** *there is a rank order relationship between the data and the sub-levels?"*

L119: This is interesting. Will you discuss why you think this is the case sometime later in the paper?

See also response in initial reply. We added/changed: "In addition, an internal analysis of qualifiers assigned to danger levels describing wet-snow conditions showed that forecasters **primarily assigned sub-level plus to 2 (moderate) and 3 (considerable)**. *Hence, for wet-snow conditions, the assignment of sub-level qualifiers was halted after a test winter."*

L129-130: You said earlier you did not consider wet snow days. So, at this step did you exclude days that had only wet snow avalanche problem types?

We clarify that this is the case: "Forecasts describing exclusively wet-snow or gliding avalanches as the main avalanche problem were excluded, as no sub-level was assigned."

L 435: "For the first research question,"

Done.

L 443: "second»

Done.

L 464: «Forecasters»

Done.

L 458-483: In this "Implications for forecasters" section, I think it would be useful to discuss how the forecasters viewed this additional task when putting together the forecast. Was this difficult for the forecasters, or did it not really take much additional time?

*At the beginning of this section, we now added: "***Forecasters felt generally comfortable assigning a sub-level in dry-snow conditions. We attribute this to the fact that forecasters must characterize the severity of the avalanche conditions to accurately describe the situation in the forecast, regardless of whether a sub-level is assigned or not. However, assigning a sub-level makes this evaluation more systematic and facilitates communication with other forecasters on duty. Forecaster feedback suggests that the additional mental effort required for the assignment of a sub-level is small and that discussing the sub-level at the forecaster briefing does not take more time than discussing any of the other elements, which are communicated in the forecast products.***"**

Remarks by Rune Engeset

Line 1 I suggest adding "and manual" before "decision-making" to point out that the process is currently not done by a machine.

Changed to: "Forecasting avalanche danger at a regional scale is a largely data-driven, yet also experience-based decision-making process **by human experts**."

Line 14 Improve the “and in which part of” part of the sentence

Changed to: “Using a sequential analysis, we first establish which data are suitable, and in which part of the danger scale, by comparing their distributions at consecutive danger levels.”

Line 22 Please describe briefly what is meant by “increased predictive value”

We changed/added: “The sub-levels **show a rank-order correlation with data related to the three contributing factors of avalanche hazard. Hence,** *they increase the predictive value of the forecast, opening the discussion on how this information could be provided to forecast users.”*

Line 34 Explain the acronym CMAH when first used

Done

Figure 1 The boxes in (b) have different dimensions. I suggest adding “and boxes” between “curve” and “(b)” in the legend

Done

Line 79 Improve the part of the sentence which reads “and to data”

Done

Line 85 Replace “research” with “research questions (RQs)”

We removed the abbreviation RQ throughout the manuscript.

Line 107 / Figure 2: The white polygons referred to in line #107 are white or grey outlines rather than white polygons and are difficult to see in Figure 2b. The colours used in 2b are very difficult to separate. I recommend improving the map and text.

In Figure 2b, we changed the colours in the map. We hope that the polygon outlines are now better to see, and that the colours linked to the sub-levels are now easier to distinguish.

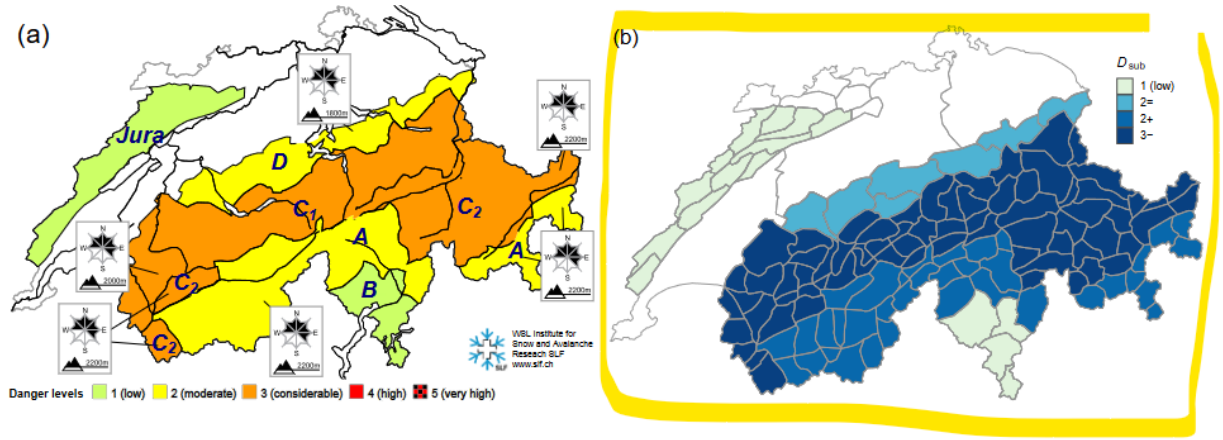


Figure 2. Maps of Switzerland showing (a) the avalanche forecast published on 10 March 2018 and (b) the (unpublished) sub-levels for this forecast. In addition, in (b) the *warning regions* (grey polygon boundaries), the smallest spatial units used in the Swiss forecast, are shown. These are aggregated to *danger regions* in the published forecast (i.e. region A in a).

Line 120 It would be of value if you explain how this was (internally) analysed, and why there is a difference between wet and dry avalanche situations.

See also previous reply. We changed to / added: "In addition, an internal analysis of qualifiers assigned to danger levels describing wet-snow conditions showed that forecasters **primarily assigned sub-level plus to 2 (moderate) and 3 (considerable)**."

Table 1 Add "and" before "(x)" in the third line of the caption

Done

Line 145-147 Could you improve readability of this sentence, and add a short explanation of why this was done?

*We changed to / added: "***We considered a high number of days with reported avalanches as an indicator for regular observations. Consequently, we expected that the number of days with no avalanches due to missing observations or wrong dating of avalanches is reduced, and hence the quality of the avalanche observations is increased.***"**

Figure 3 Why not reduce the number of rows for full ECTP to one? Why have an "*" after "only" in the caption?

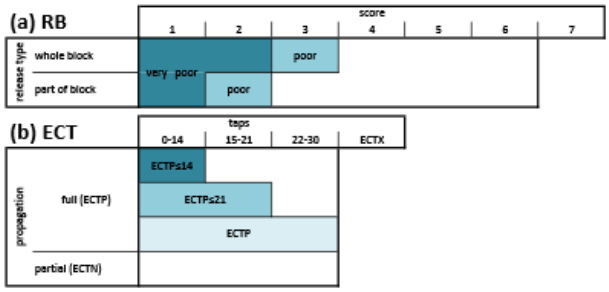


Figure 3. Classification of stability tests (a) Rutschblock (RB) and (b) Extended Column Test (ECT). The RB classification ($RB.class$) considers the score (7 loading steps) and the release type (whole block, part of block includes release type edge only). Similarly, ECT are classified combining the number of taps to initiate a fracture (30 loading steps) and the propagation propensity (full propagation: ECTP, partial or no propagation: ECTN). RB score 7 and ECTX indicate that no failure could be initiated following loading.

See also previous reply. - We removed the asterisk in the figure and changed the caption to: "Classification of stability tests (a) Rutschblock (RB) and (b) Extended Column Test (ECT). The RB classification considers the score (7 loading steps) and the release type (whole block, **part of block includes release type edge only***). Similarly, ECT are classified combining the number of taps to initiate a fracture (30 loading steps) and the propagation propensity (full propagation: ECTP, partial or no propagation: ECTN). RB score 7 and ECTX indicate that no failure could be initiated following loading."*

Line 196 Please explain what you mean by avalanche terrain in this context. Release areas only or release areas and runout zones?

*We added: "***This approach to classifying avalanche terrain considers a relevant slope area for each point in the terrain. Therefore, points lying in avalanche release areas but also in slopes below, may be considered avalanche terrain (for details refer to Schmudlach and Köhler, 2018).***"*

Line 226-228 I recommend improving this very long sentence

We changed to a list:

"For this analysis, we distinguished between data sources,

- *which mostly included only a single data point, or which mostly included only a single data point or even no data at all per forecast danger region (Figure 4a - step 1) and data,*
- *which allowed the calculation of a proportion or a mean for each forecast danger region (Figure 4b)."*

Line 258 I recommend improving this sentence.

We split the sentence into three shorter ones: "Danger levels are rank ordered. The absolute increase in danger from one danger level to the next is unknown. To derive the expected danger rating D_{model}*, we used the respective integer values of the four danger levels 1 (low) to 4 (high) (w:= {1, 2, 3, 4})."*

Line 355 You would like to write "human-triggered"

Done

Figure 7 In the first line in the caption, consider if “surface area” should probably be replaced by e.g. “area of PRA”. Consider using x103 (by thousand) rather than x104 on the vertical axis of the top row diagrams

Changed to “area of PRA”.

Figure 8 The shaded areas in the diagrams c and d are difficult to see, consider improving legibility of the shading

We adjusted the color scale and shading of these figures, but also checked the other figures.

Table 4 This is probably a matter of preference, but I suggest using “,” instead of “/” in the table as these are pairs of sublevels

Done

Line 443 I suggest replacing “main” with “second”

Done

Line 453 I suggest using another word than “tendency”, as tendency is often used to describe a temporal trend

We replaced tendency with severity to: “This highlights that assigning a sub-level can provide an important indication about the **severity** *within a danger level, and thus has the potential to reduce the magnitude of the forecast error.”*

Line 472 I presume you would like to say that there is a need for enough (not only relevant) data. I suggest improving the sentence. If possible, could you describe what criteria should be met (in terms of amount and relevance)?

We added “enough”.

Line 528 I suggest adding “enough” before “relevant” and removing “if” after “(2)

Done.