Discussion of "A Large-scale Validation of Snowpack Simulations in Support of Avalanche Forecasting Focusing on Critical Layers"

Author response to referee comment 1

Herla et al.

June 16, 2023

1 Responses to Referee #1 (Jürg Schweizer)

1.1 General Comments

Referee Comment: *The manuscript by Herla et al. reports a validation study. It evaluates the performance of operational snowpack simulations to represent critical layers in view of avalanche triggering at the regional scale. The validation data are qualitative descriptions of potentially relevant layers by avalanche forecasters. The authors acknowledge that those do not represent the truth. Moreover, it seems that these are rather special data and not commonly used in other forecasting services. In general, the impression is that the study is much focused on the Canadian situation and the specific needs of the Canadian forecasters who seem to be skeptical about the usefulness of models. This said, I agree that the comparison presented (regional assessments by forecasters vs. snowpack simulations) is meaningful.*

The study is well designed, the methods very appropriate and overall clearly described. Still, I found some paragraphs hard to follow due to some very detailed evaluations that I occasionally had troubles seeing their relevance. Finally, I would personally prefer a slightly stronger emphasis on the research questions and the resulting answers in terms of the application of the model (and future developments) rather than, it seems, directing the discussion primarily to avalanche practitioners.

I recommend the manuscript to be accepted after minor revisions. See my detailed comments below.

Author Comment: Thank you for your positive and encouraging feedback! We value and appreciate your suggestions that have helped us improve the manuscript substantially.

Most comments were editorial, and we think that the revised presentation of the material will help readers better understand the manuscript. We are particularly thankful for your comment 1.2.32, which helped us find an oversight in the performance computation of crust layers. The revision motivated by this non-editorial comment led to a revised methodology in computing the performance curves related to crusts. Due to this revision, the corresponding results improved slightly.

We respond to each comment in a point-by-point manner below.

1.2 Detailed comments

1.2.1 Line 28: Terminology of critical layers

Referee Comment: *I argue that crusts are not critical layers. While persistent weak layers may form above or below crusts, crusts themselves are not critical layers. This misconception should definitely not be further perpetuated.*

Author Response: We thank you for this perspective, which highlights the need for a more in-depth discussion of this topic within our international community. In fact, in Canada the term critical layers that refers to persistent weak layers and crusts has been coined by Bruce Jamieson and his students a while ago. As you describe yourself, persistent weak layers may form adjacent to crusts due to processes that are enhanced by the crusts. The crust layers therefore act as the main reference for practitioners who need to (i) identify these layers in their snow pits, and (ii) extrapolate where these layers might exist and be of concern. The term critical layers (such as used in our first draft manuscript) is based on the incentive to create a concise term. We value both perspectives, but do not think that our manuscript is the right place for this discussion. We therefore accept your comment and changed the use of the term critical layers in our revised manuscript as follows.

Instead of using the term critical layer to *refer* to persistent weak layers and crusts, we use the explicit long version “persistent weak layers and crusts”, or substitute the terms “persistent layers”, “relevant layers”, “layers of concern”, and “regional layer” depending on the context. We do keep the term “critical layers” for all layers that are critical with respect to their instability. This is in line with the suggestion by the referee and with the terminology defined in the flowchart of Fig. 2.

1.2.2 Line 53-54: Limitations of validation studies with respect to the modeling setup

Referee Comment: *Please be aware that there are different setups for operational modeling. The most promising is certainly a combination of distributed modeling with continuous data assimilation. In addition, I disagree that simulations at the point scale are not useful. Obviously, almost all measurements and observations are at the point scale, and very often considered useful. As mentioned above, those measurements should then be assimilated in a distributed model.*

Author Response: Thanks for the comment. We modified the relevant paragraph to be more specific about the limitations we see in existing approaches of these kinds of validation studies. The new edits make the paragraph more clear in

- that we purely refer to validation studies, not the general model setup
- that we do not judge any model setup, and
- that the magnitude of the limitations depends on the region of interest and the setup of the operational model.

The revised paragraph reads as follows:

> “Despite this large body of snowpack validation studies, the operational needs of (Canadian) avalanche forecasters have not been satisfied yet. While process-based validations at individual point locations based on high quality data are crucial for model development and improvement, these validation results do not provide sufficiently tangible and relevant guidance for forecasters who forecast for different locations or regions. In addition, these validation results are not necessarily representative of the real skill of operational simulations which might rely on different data sources or model configurations. [...] ”

1.2.3 Line 66: Terminology likelihood vs probability

Referee Comment: *I wonder whether the term likelihood should only be used in the corresponding statistical context rather than as a synonym for the more generic probability.*

Author Response: We agree, thanks for pointing this out! We changed all occurrences of the term "likelihood" to "probability" where applicable. Please note, that we kept the term "likelihood" unchanged where we refer to the likelihood component of an avalanche problem as defined in the Conceptual Model of Avalanche Hazard (Statham et al. 2018).

1.2.4 Line 94: Treeline elevation range

Referee Comment: *Treeline seems to be a rather straightforward to determine objective elevation (range). I wonder why to rely on forecaster consensus.*

Author Comment: Forecaster consensus was chosen for treeline elevation ranges because operational forecasters do not have consistent objective definitions for vegetation bands. We wanted to know their interpretations of elevation ranges in each region because they would be matching field observations from those ranges to vegetation bands in their assessments. This should create the best possible match in elevation between assessments and profile simulations.

As an aside, work towards an objective classification of vegetation bands in western Canada (where treeline ranges from 800 to 2400 m) has been challenging because forest density data tends to be poor quality at upper elevations and it is difficult to detect transitions in areas without alpine terrain.

The slightly modified paragraph reads as follows:

> We classified each grid point into an elevation band class, 'alpine' (ALP), 'treeline' (TL), and 'below treeline' (BTL) to match the terrain classification in the human assessments. To create the best possible match in elevation between assessments and simulations, we used forecaster consensus of TL elevation for the classification: 1600–1800 m in S2S, 1800–2100 m in GNP, and 2000–2400 m in BYK.

1.2.5 Line 102: Stability scheme

Referee Comment: *What do you mean with the "stability scheme" by Michlmayr et al. (2008).*

Author Response: We refer to the *atmospheric* stability scheme. We make that clear in the revised version.

1.2.6 Line 104: SNOWPACK's layer aggregation

Referee Comment: *Please provide a short explanation why you turned off SNOWPACK's layer aggregation feature.*

Author Response: The newly added explanation reads as follows:

> Lastly, we turned off SNOWPACK's layer aggregation feature of merging similar and adjacent layers to preserve exact knowledge of the formation dates of individual layers. The need for this decision will become apparent in Sect. 3.1.2 and 3.1.3, where we explain our approach of grouping and matching layers based on date considerations.

1.2.7 Line 107-110: Official grain type terminology

Referee Comment: *I suggest using the official terms for the grain types according to the ICSSG (Fierz et al., 2009) such as "decomposing and fragmented precipitation particles" or "rounded grains".*

Author Response: Agreed. Changed accordingly.

1.2.8 Line 116: Definition of likelihood and size

Referee Comment: *What is likelihood and size referring to?*

Author Response: In the revised version, we explain in more detail:

> Applying the Conceptual Model of Avalanche Hazard (Statham et al. 2018), forecasters partitioned the avalanche hazard into different avalanche problems and characterized each problem by its type, location, likelihood of avalanches, and destructive avalanche size.

1.2.9 Line 184: Change reference

Referee Comment: *Viallon-Galinier et al. (2022) does not seem to be the most appropriate reference to refer to the two main processes of avalanche release.*

In another context, the previous snow cover and snow instability modeling study by Reuter and Bellaire (2018) may also be of interest.

Author Response: We changed the previous reference (most recent review of stability indices in snowpack models) to more general reviews on slab avalanche formation.

> [...] which accounts for the two main processes governing slab avalanche release (Schweizer et al. 2003, 2016).

We extensively refer to Reuter et al. (2021), which is the peer-reviewed development of Reuter and Bellaire (2018). We added Reuter and Bellaire (2018) to the revised Introduction.

1.2.10 Line 186: Add references

Referee Comment: *Please provide references for the thresholds you selected (RTA, SK38 and r_c).*

Author Response: Agreed. The revised version reads as follows:

> [...] While the literature agrees on thresholds for RTA and SK38, less consensus exists for r_c. Hence, we applied two thresholds to r_c and therefore identified critical avalanche layers with poor stability if RTA ≥ 0.8 (Monti and Schweizer 2013), SK38 ≤ 1 (Reuter et al. 2021), and $r_c \leq 0.3$ (Reuter et al. 2021) or $r_c \leq 0.4$ (Reuter et al. 2015).

1.2.11 Line 208: SNOWPACK transforming DH-SH

Referee Comment: *With regard to your note that SH is transformed into DH I wonder since I think to remember that I have also seen SNOWPACK runs where SH was present for longer time periods than just a few days.*

Author Response: We have seen the full range from a few days to several weeks. In the revised manuscript, we change the wording to also account for those cases where SH is present for longer.

> [...] To address the well-known SNOWPACK behaviour of transforming most SH layers into DH layers after they have been buried for several days to weeks, [...]

1.2.12 Line 227: typo

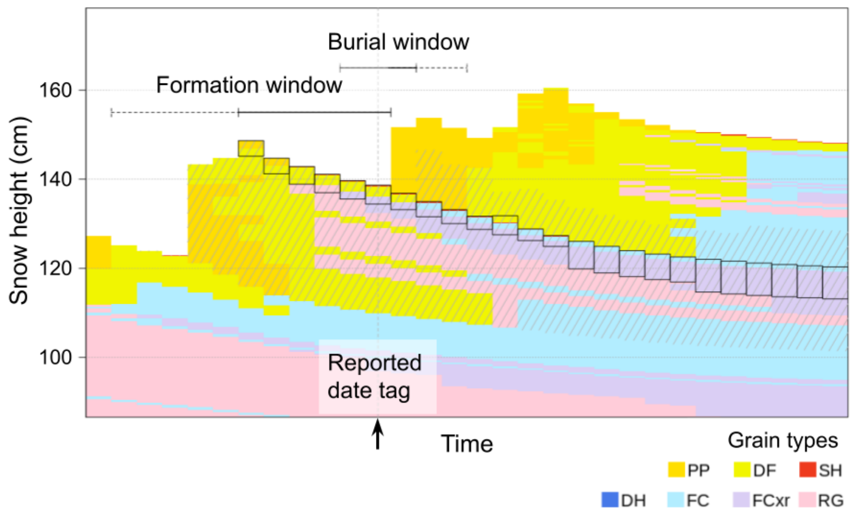
Referee Comment: *snowfall*

Author Response: Changed. Thank you!

1.2.13 Figure 3: Caption and y-axis labels

Referee Comment: *I do not follow what you describe in the second last sentence of the caption. Please clarify. Also, to improve readability, I recommend rotating labels on y-axis 90 degrees clockwise, in all figures.*

Author Response: Thanks for the heads up! We changed the caption of Figure 3 and will rotate the y-axis labels in all figures. The updated Figure 3 with its revised caption:



> An illustration of the search windows for layers of human concern around the human date tag. The panel zooms in on the near-surface layers of a time series of a simulated snow profile. The extent of the formation and burial windows are highlighted by horizontal lines, where the solid lines indicate the extent of each window based on the timing of the storms. The black boxes highlight all layers that either formed within the formation window or got buried within the burial window. To better illustrate the effect that different lengths of time windows have on the selection of layers, the gray hatched lines highlight layers that would result from fixed time windows that are not based on the timing of the storms (dashed horizontal lines). Colors refer to snow grain types defined in Sect. 2.2.

1.2.14 Line 253-254: Assigning simulated layers to datetags

Referee Comment: *Please clarify: "not assigned a simulated layer"*

Author Response: Thanks for the comment that this concept is difficult to follow. We simplified the wording and inserted a reference to the section where this concept is explained in detail. The revised sentence reads as follows:

> Analogously, all potential model-derived date tags without any simulated unstable layers (rightmost arrow in Fig. 2, Sect. 3.1.2) were counted as absent layers and directly contributed to the True Negative cell of the matrix.

1.2.15 Line 265/Figure 4: Different line colors

Referee Comment: *I suggest using colors that can be better discerned than red and black (vertical lines in Figure 4) or at least use different line style.*

Author Response: Agreed. We will change that accordingly.

1.2.16 Line 266: Typo

Referee Comment: *delete "in"*

Author Response: Done. Thank you.

1.2.17 Line 270: Typo

Referee Comment: *Four layers (considered critical by the model) in the text, five in the confusion matrix in Fig. 4d?*

Author Response: Thank you for spotting this! This is a typo in the text. Figure 4a shows that the number 5 in the confusion matrix of Fig. 4d is correct. Changed.

1.2.18 Line 330: Typo

Referee Comment: *delete "can be"*

Author Response: Done. Thank you.

1.2.19 Line 384: Typo

Referee Comment: *Wilcoxon*

Author Response: Changed. Thank you.

1.2.20 Line 384: Figure reference

Referee Comment: *Your refer to Fig. 7a,c for the process-based approach, correct?*

Author Response: Yes, the figure reference (Fig. 7a, c) is correct. We edited the sentence slightly to clarify potential ambiguity:

> The performance of the process-based approach with the higher threshold is significantly lower than the statistical approach on the 95 % confidence level. The process-based approach with the lower threshold, however, yields comparable results to the statistical approach and lies within the confidence band of the statistical approach (Fig. 7a, c).

1.2.21 Figure 9: Labels

Referee Comment: *The sequence of the subpanels (grain type, elevation, region) is different form the description in the caption. Also, I suggest labeling the curves in panels b, d, f (as in a, c, e).*

Author Response: Good catch! We corrected the description in the caption and added labels to the curves in panels b, d, f.

1.2.22 Line 500: Performance of modeled and human forecasts

Referee Comment: *I suppose not only the performance of the models, but also the forecasts are far from perfect. I guess some of the poor agreement you describe on pages 22-23 may also be related to peculiarities of forecast procedures.*

Author Response: Absolutely, we agree. To pick up readers with the same thought, we added a sentence about this fact and then refer to Sect. 5.2, where we discuss this in detail. The revised introductory paragraph to Sect. 5.1 reads as follows:

> Our results have shown that the Canadian weather and snowpack model chain is able to skillfully represent weak layers that are of operational concern. However, comparison against human assessments also demonstrates substantial differences between the two data sets. Both data sets are characterized by uncertainty, and both forecasts—modeled and human—are far from perfect. While this fact is discussed in detail in the subsequent section (Sect. 5.2 Limitations), the current section [...].

1.2.23 Line 511: Typo

Referee Comment: *I suppose you refer to Figure 13a, here, also below, in line 514.*

Author Response: Correct, thank you. Changed.

1.2.24 Line 545: Recommendation for crusts

Referee Comment: *I agree. On the other hand, you may also consider running slope simulations so that crusts will form on sunny aspects. However, as pointed out, the relevance of crusts compared to weak layers is very different.*

Author Response: Agreed. We added the recommendation.

> Given the poor model performance in capturing any type of crust, while modeling many additional crust layers that were never reported upon, we advise to use the Canadian operational weather and snowpack model chain only for assessing weak layers (such as SH, DH, FC) and not crusts, at least until the model chain includes slope simulations and further testing has been carried out.

1.2.25 Line 550: Alternative effects influencing performance in different regions

Referee Comment: *Alternatively, apart from snow climate, also forecast performance could be different, or the type of analysis you selected, following so-called layers of concern, may be better suited for GNP.*

Author Response: That is correct. We added two statements in the existing paragraph to clarify this:

> This strong effect is likely influenced by the regions' snow climates, as well as forecast practices by different agencies in different snow climates. Due to its transitional snow climate GNP is characterized by substantial snowfall amounts interspersed with frequent periods of critical layer formation (Haegeli and McClung 2007; Shandro and Haegeli 2018). Therefore many persistent avalanche problems exist each season, most of which can usually be linked to specific thin critical layers (Fig. 5c, d). In the maritime snow climate of S2S, critical layers form much less often and cause less persistent problems (Fig. 5a, b). Although continental BYK experiences the most days per season with persistent avalanche problems, less persistent problems can actually be linked to specific critical layers (Fig. 5e, f). Instead of thin critical layers, the continental snowpack is often characterized by thick bulk layers of low cohesion. Since these thick layers often get deposited by different snowfall events and facet over the course of many dry spells, it can be challenging to name these layers, let alone distinguish them in the field. Our analysis approach of focusing on specific identifiable layers of concern may therefore be most applicable to GNP. For all these reasons our data set of tracked layers of concern is skewed towards GNP and leaves BYK and particularly S2S underrepresented. Our results for S2S and BYK therefore have to be interpreted with more caution and in light of their regional peculiarities. For example, Horton and Haegeli (2022) found that BYK (and other forecast regions in the Canadian Rockies) consistently receive underestimated modeled snowfall amounts, which increases temperature gradients in the snowpack and helps explain the overestimated faceting we see in our results of BYK, which in turn leads to a low model precision in BYK. In contrast, the low model precision in S2S could be due to an underrepresentation of layers of concern. Since instabilities are often short-lived when persistent weak layers get buried by big storms, many critical layers potentially get never associated with a persistent avalanche problem.

1.2.26 Line 595: Likelihood terminology

Referee Comment: *Likelihood of problem or of avalanches?*

Author Response: Likelihood of avalanches given a specific avalanche problem. We changed the statement accordingly.

1.2.27 Section 5.3: Ideas for specific examples

Referee Comment: *I recommend you provide some specific examples in Figures 14 and 15 when, e.g. "the view will alert forecasters", or when the forecast is not supported by model result. For instance, there seems to be little variation in likelihood of avalanches but more in the proportion of unstable grid points – and how does that relate to the danger level?*

Author Response: Thanks for this recommendation. We absolutely see the added value of these tangible examples, and added an entire paragraph to the section:

> A dashboard view like this effectively summarizes the large amount of simulated observations and allows forecasters to easily put their assessments of specific layers in relation to the simulated data. For example, the reported likelihood of avalanches of persistent avalanche problems shows less variation than the corresponding proportion of simulated unstable grid points (Fig. 14c, d). Focusing on the period from Dec 03 to Jan 05, both bar charts show an almost identical progression. However, while the modeled proportion unstable spans the entire scale from 0–1, the reported likelihood of avalanches only varies between *Possible–Likely/Very Likely.* Simultaneously, the

reported danger level is *Low* at first (when the proportion unstable is close to 0), then increases to *High* within three days (Dec 13) (when the proportion unstable also increases to almost 1 at the same time). The danger rating remains constant at *Considerable/High* for roughly one week (when the proportion unstable also lingers between 0.8–1). Starting at Dec 21 the danger rating drops to *Moderate* then *Low* for a total of six days (when the proportion unstable also tapers off to 0 within the same time frame) before two snowfall events at Dec 29 and Jan 03 bring the danger rating back up to *Considerable* and *High*. The last two peaks of the danger rating are reflected in the proportion unstable at the same times, but the magnitude remains lower. This aligns with the human assessments that dropped the persistent avalanche problem at Jan 03 and only called it a storm snow problem. At the same time, the average profile (Fig. 14e) highlights substantial amounts of unstable new snow. Interestingly, it also shows a thin layer of unstable facets that got buried at Dec 28. This layer is not mentioned in the human hazard assessments that still attribute the persistent problem to the Nov 21 SH and the Dec 09 SH layers. Both of these layers are also present in the average profile, but their main activity was modeled between Dec 10 and Dec 23. Besides these nuanced comparisons at times of agreement between modeled and human data sets, the dashboard view makes any serious discrepancies easy to spot. For example, during the early season when no human assessment data is available yet, the proportion unstable highlights times of instability in early season weak layers, such as around Nov 17 and Nov 24, which are caused by the Nov 14 and Nov 21 layers (Fig. 14d, e). The opposite is possible as well. Starting with Mar 16, the assessments indicate *High* hazard and a persistent avalanche problem on the Mar 07 FC layer. However, the simulations show no signs of instability at all. Despite this dramatic discrepancy, the visualized information can still be beneficial for forecasters who get prompted to think critically about the current situation. Investigating the underlying reasons for the disagreement may help them make a more informed decision. In this specific example, additional hazard and weather information (not shown in Fig. 14) uncovered that this situation coincided with the first wet avalanche cycle of the year, a process not captured by the stability measure used in this paper.

1.2.28 Figure 14: Figure legend

Referee Comment: *I suggest adding a legend for the danger rating. Also, please add year and region to the seasonal overview.*

Author Response: Agreed. We will add the legend and add the year/season to the figure.

1.2.29 Figure 15: Unit error

Referee Comment: *There seems to be an error with regard to the units of the slab cohesion. The values you indicate are in the range of 100 to 500, the parameter is supposed to be density divided by grain size, and the units indicated are kg m -4 ?*

Author Response: Thank you for pointing this out. Indeed, the units need to read kg m^{-3} mm^{-1}. This choice of units is consistent with Mayer et al. (2022, Fig. B2f). We changed the unit error in the manuscript.

1.2.30 Lines 674-682: Paragraph needs rewrite

Referee Comment: *I do not really follow the argumentation here, for instance, why there should be a model bias and what do you refer to by Fig. 12d, 11, and 10? What is the reasoning for* *density and sphericity for the sensitivity to the length of the dry spell?*

Author Response: That comment calls for a thorough revision of the entire paragraph. We added explanations to improve the logic of the idea, stated explicitly that this idea is a hypothesis, and we separated the discussion of grain size from the discussion of the potential bias. The revised version reads as follows:

> A further side observation outside of the scope of this paper but nevertheless interesting for developers of snowpack models was brought up by the length of the dry period before the burial of the potential weak layer, which emerged as a variable with strong explanatory power in many of our CTree analyses (Sect. 3.2 and 3.3). Since longer dry periods conceptually increase the chance of weak layer formation and growth on the snow surface, these findings seem plausible at first. However, in sum all of the following observations appear slightly odd: (1) If the dry period before the burial of a SH layer was shorter than 7 days, the proportion of grid points that structurally contained the layer was significantly lower than for longer dry periods (Fig. 10, Nodes 6–8). (2) Barely any SH layer was modeled unstable when buried after a dry period shorter than 7 days, whereas significantly more SH layers were modeled unstable when the dry period was longer (Fig. 11, Nodes 7–9). A less strong but similar effect was observed for FC and combinations of SH/FC (Fig. 1, Nodes 2, 3, 6). (3) Weak layers that were buried after a dry period longer than 12 days were simulated unstable for significantly longer duration than human assessments suggest than weak layers that were buried after a shorter dry period (Fig. 12d). Based on this strong influence of the variable in several different CTree analyses we hypothesize that the model is biased towards overestimating the structural prevalence and instability of weak layers that were buried at the end of long dry periods. Since both structural existence and instability show the same patterns, we further believe that this effect originates in the SNOWPACK model itself and propagates through the stability module by Mayer et al. (2022). To dig even deeper, the variables in the stability module that are potentially affected by the length of the dry period are grain size, density, and sphericity of the weak layer (Mayer et al. 2022). Interestingly, the median simulated grain size at the time of burial was substantially less impactful in our CTree analyses. Hence, we hypothesize that the suggested bias is caused by density or sphericity.
>
> Our data set of human assessments reported multiple SH layers with grain sizes beyond 10 mm, while the median simulated grain sizes at the time of burial rarely exceeded 2.5 mm. Although grain sizes were simulated substantially smaller than reported in many cases, our findings still suggest that the simulated grain sizes were consistent within the simulations. Our CTree analysis has shown that median simulated grain sizes at the time of burial that exceeded 0.9 mm were associated with a higher proportion unstable than smaller grain sizes. This is in line with Mayer et al. (2022), who report an increased influence of the simulated grain size on layer instability for grain sizes above 1–1.5 mm (Fig. B2d in Mayer et al. 2022).

1.2.31 Lines 691: Language edit

Referee Comment: *I suggest replacing human.*

Author Response: Agreed. Done.

1.2.32 Lines 701: Performance of crusts

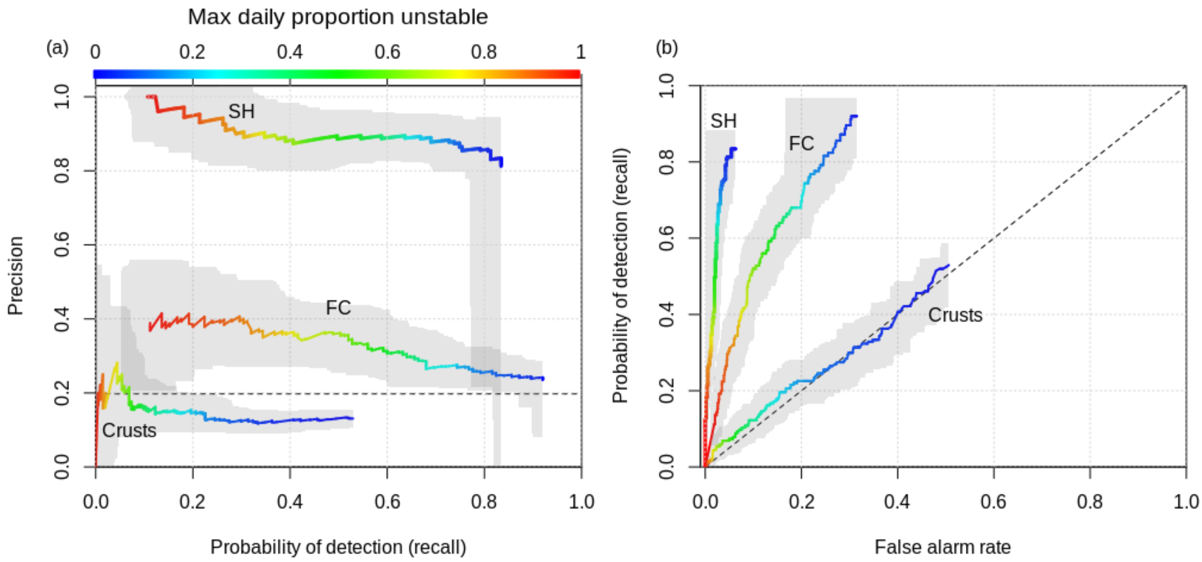
Referee Comment: *Is the lack of skill with regard to crusts due to the model setup (flat) or the misconception about the role of crusts?*

Author Response: At the outset of this project, we actually planned to solely look into structural presence or absence of weak layers and crusts. As the project matured, we decided to take stability into account as well. While implementing that additional step, we indeed incorporated this oversight. Upon verifying our calculations due to this comment, we found that we only looked at the stability of the crust grains instead of the stability of adjacent facets. That is an error that impacts the results presented in Fig. 9a, b, curve labeled 'MFcr'. In the meantime, we ran the relevant calculations, and we have made the following edits to the manuscript.

We added the appropriate routines for selecting crust layers with adjacent unstable faceted layers:

> To identify relevant layers in the snowpack simulations, we searched all simulated snow profiles for layers that were characterized by persistent grain types and poor stability. Persistent grain types included all faceted layers (FC, DH), surface hoar layers (SH), and crust layers (MFcr, IF). Since crust layers are strong layers, we counted crust layers as unstable if unstable weak layers (FC) were present in the vicinity of the crusts.
>
> [...]
>
> While the search windows identified all simulated layers that align with a specific human date tag, the layers still needed to be filtered for the reported grain types to produce a meaningful match. To accomplish this, we employed up to four different grain type searches for each layer of concern: a) a strict grain type search that only accepted the specific *reported* weak grain types (e.g., SH, SH/FC, etc.) in the simulated layers; b) a relaxed grain type search that accepted *all* persistent weak grain types (SH/DH/FC) in the simulated layers; and a grain type search for c) any crust layer (IF, MFcr) or d) any crust layer with an adjacent unstable faceted layer if crusts were reported to play a role in the given layer of concern.

We re-computed Fig. 9a, b where the curves that represent crusts changed:



We changed the description of the figure in the text:

> Crust layers are poorly represented in the simulations with probabilities of detection below 50 % at a low precision of 10-20%. All performance measures for crusts closely follow the no-skill base line (Fig. 9a, b).

We added a statement about the poor precision in the discussion, which was not an issue in the old version:

> Given the poor model performance in capturing any type of crust, while modeling many additional crust layers that were never reported upon, we advise to use the Canadian operational weather and snowpack model chain only for assessing weak layers (such as SH, DH, FC) and not crusts, at least until the model chain includes slope simulations and further testing has been carried out.

We added a statement to the concluding comment:

> While the performance was substantially better for surface hoar (SH) layers than facets (FC), the model had no skill in representing any type of crust layer (IFrc, IFsc, MFcr), neither structurally nor when taking adjacent unstable facets into account.

Overall, the revised methodology is now in line with the referee's comments, and the crust performance improved slightly from extremely poor to poor. We believe that the reason for the poor performance is related to (i) the configuration of the simulations (flat field instead of slopes) for IFsc layers, and (ii) the phase determination of precipitation based on a temperature threshold implemented in SNOWPACK (instead of using phase information from the weather model) and (iii) other meteorological inputs from the weather model for IFrc and MFcr layers.

1.2.33 Lines 707: Conclusion

Referee Comment: *I agree that it is essential to find the critical weak layers and assess their degree of instability. In addition, we may also ask whether the temporal evolution of the instability is properly modeled. In other words, are the parameterizations implemented capable to adequately simulate the temporal evolution of strength and toughness.*

Author Response: Thanks for this encouraging comment—this is what we're working on right now ;-). Although again on the regional scale and constrained by our data set of human assessments, so our results will likely not contribute to evaluating the parametrizations you mention. I'm curious whether Switzerland has the data to run these validations on the regional scale (around Davos?) and for many winter seasons?

We added one more sentence to the end of the paragraph:

> Future research may benefit from a more in-depth analysis of the temporal agreement between modeled and reported instability.

References

Haegeli, P. and McClung, D. M.: Expanding the snow-climate classification with avalanche-relevant information: Initial description of avalanche winter regimes for southwestern Canada, J Glaciol, 53, 266–276, https://doi.org/10.3189/172756507782202801, 2007.

Horton, S. and Haegeli, P.: Using snow depth observations to provide insight into the quality of snowpack simulations for regional-scale avalanche forecasting, The Cryosphere, 16, 3393–3411, https://doi.org/10.5194/tc-16-3393-2022, 2022.

Mayer, S., van Herwijnen, A., Techel, F., and Schweizer, J.: A random forest model to assess snow instability from simulated snow stratigraphy, The Cryosphere, 16, 4593–4615, https://doi.org/10.5194/tc-16-4593-2022, 2022.

Monti, F. and Schweizer, J.: A relative difference approach to detect potential weak layers within a snow profile, in: Proceedings of the 2013 International Snow Science Workshop, Grenoble, France, pp. 339–343, URL `https://arc.lib.montana.edu/snow-science/item.php?id=1861`, 2013.

Reuter, B. and Bellaire, S.: On combining snow cover and snow instability modelling, in: Proceedings of the 2018 international snow science workshop, Innsbruck, AUT, pp. 949—-953, URL `https://arc.lib.montana.edu/snow-science/item/2684`, 2018.

Reuter, B., Schweizer, J., and van Herwijnen, A.: A process-based approach to estimate point snow instability, The Cryosphere, 9, 837–847, 2015.

Reuter, B., Viallon-Galinier, L., Horton, S., van Herwijnen, A., Mayer, S., Hagenmuller, P., and Morin, S.: Characterizing snow instability with avalanche problem types derived from snow cover simulations, Cold Reg Sci Technol, 194, 103 462, https://doi.org/10.1016/j.coldregions.2021.103462, 2021.

Schweizer, J., Jamieson, J. B., and Schneebeli, M.: Snow avalanche formation, Rev Geophys, 41, https://doi.org/10.1029/2002rg000123, 2003.

Schweizer, J., Reuter, B., Van Herwijnen, A., and Gaume, J.: Avalanche Release 101, in: Proceedings of the 2016 International Snow Science Workshop, Breckenridge, CO, USA, URL `https://arc.lib.montana.edu/snow-science/item/2235`, 2016.

Shandro, B. and Haegeli, P.: Characterizing the nature and variability of avalanche hazard in western Canada, Nat Hazard Earth Sys, 18, 1141–1158, https://doi.org/10.5194/nhess-18-1141-2018, 2018.

Statham, G., Haegeli, P., Greene, E., Birkeland, K. W., Israelson, C., Tremper, B., Stethem, C., McMahon, B., White, B., and Kelly, J.: A conceptual model of avalanche hazard, Nat Hazards, 90, 663–691, https://doi.org/10.1007/s11069-017-3070-5, 2018.