General comments

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.

Detailed comments

Line 28:	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.
Lines 53-54:	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.
Line 66:	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.
Line 94:	Treeline seems to be a rather straightforward to determine objective elevation (range). I wonder why to rely on forecaster consensus.
Line 102:	What do you mean with the “stability scheme” by Michlmayr et al. (2008).
Line 104:	Please provide a short explanation why you turned off SNOWPACK’s layer aggregation feature.
Lines 107-110:	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”.
Line 116:	What is likelihood and size referring to?
Line 184:	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.

Line 186: Please provide references for the thresholds you selected (RTA, SK38 and r_c).

Line 208: 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.

Line 227: snowfall

Figure 3: 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.

Lines 253-254: Please clarify: “not assigned a simulated layer”

Line 265: 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.

Line 266: delete “in”

Line 270: Four layers (considered critical by the model) in the text, five in the confusion matrix in Fig. 4d?

Line 330: delete “can be”

Line 384: Wilcoxon

Line 399: Your refer to Fig. 7a,c for the process-based approach, correct?

Figure 9: 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).

Line 500: 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.

Line 511: I suppose you refer to *Figure* 13a, here, also below, in line 514.

Line 545: 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.

Line 550: 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.

Line 595: Likelihood of problem or of avalanches?

Section 5.3: 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?

Figure 14: I suggest adding a legend for the danger rating. Also, please add year and region to the seasonal overview.

Figure 15: 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} ?

Lines 674-682: 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?

Line 691: I suggest replacing human.

Line 701: Is the lack of skill with regard to crusts due to the model setup (flat) or the misconception about the role of crusts?

Line 707: 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.

Davos, 5 May 2022
Jürg Schweizer

References

Reuter, B., and Bellaire, S.: On combining snow cover and snow instability modelling, Proceedings ISSW 2018. International Snow Science Workshop, Innsbruck, Austria, 7-12 October 2018, 949-953, 2018.

Viallon-Galinier, L., Hagenmuller, P., Reuter, B., and Eckert, N.: Modelling snowpack stability from simulated snow stratigraphy: Summary and implementation examples, Cold Reg. Sci. Technol., 201, 103596, https://doi.org/10.1016/j.coldregions.2022.103596, 2022.