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.

May 8, 2024

1 Responses to Referee #2 (anonymous)

1.1 General Comments

Referee Comment: *This is overall a well written and motivated manuscript, although heavy on jargon, technical language and acronyms in places, and assuming the reader is an 'expert' in some places without full explanations. The manuscript is likely aimed at a very specific audience, and does little to make it more accessible to those who might be at the periphery of snowpack simulations. That said, it is clearly a substantive work which will be good for those doing snowpack simulations, although they will need to read it several times to fully understand what is being done and how it is being done.*

I have made a series of suggestions below, in no order of importance, but rather notes as reading through it, and sometimes returning back to different parts of the manuscript. Because no major issues came up, I would suggest that this go through minor changes. Many of the following are stylistic and to improve the structure/content, with an occasional query about meaning and details given.

Author Comment: Thank you very much for taking the time to review our manuscript! We appreciate and value your perspective, particularly on accessibility. We respond to each comment in a point-by-point manner below.

1.2 Detailed comments

1.2.1 Abstract: quantitative summary

Referee Comment: *1. [Abstract] Consider whether to put more quantitative summaries into the abstract (currently is very narrative).*

Author Response: Thank you for the suggestion. In our revised abstract, we include more quantitative summaries. We also used your prompt to tighten up the narrative writing style a bit to keep the word count at bay. The revised abstract reads as follows (see highlighted text for added statements):

> Avalanche warning services increasingly employ snow stratigraphy simulations to improve their current understanding of critical avalanche layers, a key ingredient of dry

slab avalanche hazard. However, a lack of large-scale validation studies has limited the operational value of these simulations for regional avalanche forecasting. To address this knowledge gap, we present methods for meaningful comparisons between regional assessments of avalanche forecasters and distributed snowpack simulations. We applied these methods to operational data sets of ten winter seasons and three forecast regions with different snow climate characteristics in western Canada to quantify the Canadian weather and snowpack model chain's ability to represent persistent critical avalanche layers.

Using a recently developed statistical instability model as well as traditional process-based indices, we found that the overall probability of detecting a known critical layer can reach 75 % when accepting a probability of 40 % that any simulated layer is actually of operational concern in reality (i.e., precision) as well as a false alarm rate of 30 %. Peirce skill scores and F1 scores cap at approximately 50 %. Faceted layers were captured well but also caused most false alarms (probability of detection up to 90 %, precision between 20–40 %, false alarm rate up to 30 %), whereas surface hoar layers, though less common, were mostly of operational concern when modeled (probability of detection up to 80 %, precision between 80–100 %, false alarm rate up to 5 %). Our results also show strong patterns related to forecast regions and elevation bands and reveal more subtle trends with conditional inference trees. Explorations into daily comparisons of layer characteristics generally indicate high variability between simulations and forecaster assessments with correlations rarely exceeding 50 %. We discuss in depth how the presented results can be interpreted in light of the validation data set, which inevitably contains human biases and inconsistencies.

Overall, the simulations provide a valuable starting point for targeted field observations as well as a rich complementary information source that can help alert forecasters about the existence of critical layers and their instability. However, the existing model chain does not seem sufficiently reliable to generate assessments purely based on simulations. We conclude by presenting our vision of a real-time validation suite that can help forecasters develop a better understanding of the simulations' strengths and weaknesses by continuously comparing assessments and simulations.

1.2.2 Introduction

Referee Comment: *2. Insert somewhere in the first paragraph of the introduction "the subject of this paper" so it is clear that this will be the subject to be discussed.*

3. The introduction is strong, well cited, but would benefit (because of its length of five paragraphs, a sentence at the end of the first paragraph stating something like "In the rest of this introduction we will..." to signal to the reader where you are headed.

Author Response: We appreciate your suggestions for enhancing the clarity and structure of our introduction. In response to both of your comments, we have added a sentence to the end of the first paragraph to clearly define the subject of the paper. We believe that the suggested outline would have disrupted the flow and storyline of our introduction, which follows a conventional and widely recognized structure. The inserted sentence reads

> This paper examines the effectiveness of the Canadian operational weather and snowpack model chain to identify critical avalanche layers that are essential for regional-scale avalanche forecasting.

1.2.3 Outline Section 2

Referee Comment: *4. After the first sentence of Section 2, tell us how the Section 2 will be organized.*

Author Response: Good idea. We added the following sentence.

> This section provides the necessary background information on the study area (Sect. 2.1), the snowpack simulations employed for this study (Sect. 2.2), as well as the human hazard assessments used as validation data set (Sect. 2.3).

1.2.4 Winter season definition

Referee Comment: *5. Line 81. Indicate here or elsewhere the normal months for the winter season. In particular, if you state a winter season of 2013, is this from 2012 to 2013 or 2013 to 2014.*

14. Section 2.3 " of the winter season" Again, please define or give us an idea of how the winter season varies, or if the same months are used, which these are. If it is just the standard definition of November to March, that is fine, but state, and whether or not border line months (or other months) also would become important.

Author Response: Sure. We added the following statements:

> Sect. 2: In the manuscript, each winter season is defined to span from December of the previous year through March of the stated year (e.g., winter season 2021/22 will be referred to as 2022).
>
> Sect. 2.2: The simulations were initialized in September without any snow on the ground.

1.2.5 Figure 1

Referee Comment: *6. Figure 1. Put ALP, TL, BTL into figure caption where they appear (alpine [ALP], etc.). For all figure captions, please ensure that you acknowledge source of data explicitly in the figure caption.*

Author Response: Done.

Referee Comment: *7. Figure 1. Please include a scale for S2S, GNP, BYK insets. For overall figure map, put the line indicating 100 km slightly lower, so one can really see that it is the 'length' for 100 km. In figure caption, indicate size of grid cell (in addition to where it is mentioned in the text).*

Author Response: Done. Added statement to caption: "The grid has a 2.5 km spacing."

1.2.6 Unit: m asl

Referee Comment: *8. Line 95 and other locations, m is m asl? If so, be clear.*

Author Response: Thanks, we corrected that!

1.2.7 Study area coverage

Referee Comment: *9. Line 92. "Overall, we selected 1004 grid points (Fig. 1) covering an area of ******".*

Author Response: We included the statement. "covering an area of 6275 km^2".

Referee Comment: *10. For elevations within your classes S2S, GNP, BYK, it would be good to know the distribution of the elevation points, and some idea of which way these slopes are facing, along with any prominent wind directions.*

Author Response: Absolutely, we will include a table of grid point distributions (Table 1—please note that the table will render in the typical Copernicus style in the actual manuscript). The simulations are all run with flat field conditions and without wind transportation schemes (see next Section, first paragraph, 2.2 Snowpack simulations). Therefore we did not include a discussion of slope aspects or prominent wind directions.
We added the following sentences to Section 2.1:

> Table 1 describes the distribution of model grid points across the forecast regions and elevation bands. Due to the configuration of our snowpack simulations (flat field, no wind transport, see next section, 2.2), we do not discuss slope aspects or prominent wind directions across our study areas.

1.2.8 Table of acronyms and variables

Referee Comment: *11. Because of the large number of acronyms used in this paper, I recommend that early on you have a Table of Acronyms (Table 1) to make it easier for the reader in what is a fairly 'dense' paper.*

21. Table of variables (could be combined with table of acronyms). There are a lot of variables—consider having a table of acronyms and variables introduced early on, defining each variable, name, units, etc.

Author Response: Sure. We added the tables to an appendix (please note that the tables will render in the typical Copernicus style in the actual manuscript), but reference the tables early in the manuscript.

> A comprehensive table of acronyms (Table 2) and table of variables (Table 3) can be found in Appendix B.
>
> [...]
>
> This appendix provides a table of acronyms (Table 2) and a table of variables (Table 3) for abbreviations and symbols used throughout the manuscript.

	TOTAL	alpine (ALP)	treeline (TL)	below treeline (BTL)
Sea-to-sky (S2S)	476	48 (10%)	54 (11%)	374 (79%)
Glacier National Park (GNP)	233	30 (30%)	102 (44%)	101 (43%)
Banff–Yoho–Kooteney (BYK)	295	7 (2%)	163 (55%)	125 (42%)

Table 1: Number and percentage of model grid points in each region and elevation band.

Acronym	Description
ALP	Alpine elevation band
BTL	Below treeline elevation band
BYK	Banff-Yoho-Kooteney National Park
CTree	Conditional inference tree
DF	Decomposing and fragmented particles
DH	Depth hoar
FC	Faceted crystals
FCxr	rounding facteted particles
F_1	F_1 skill score
FN	False negative result
FP	False positive result
GNP	Glacier National Park
HRDPS	High resolution deterministic prediction system, a numerical weather prediction model
IFrc	Rain crust
IFsc	Sun crust
MF	Melt forms
MFcr	Melt–freeze crust
nWKL	Number of weak layers
PSS	Peirce skill score
PP	Precipitation particles
RG	Rounded grains
ROC	Receiver operating characteristics curve
RTA	Relative threshold sum approach (snow stability index)
S2S	Sea-to-Sky avalanche forecast region
SH	Surface hoar
SK38	Skier stability index
TL	Treeline elevation band
TN	True negative result
TP	True positive result

Table 2: Descriptions of acronyms used throughout the manuscript.

Variable	Name (unit)	Description
Δ_{duration}	Difference (days)	Agreement indicator of layer instability
$< \frac{\text{density}}{\text{grain size}} >_{\text{slab}}$	slab cohesion (kg m^{-3} mm^{-1})	Average ratio of density over grain size of the slab
$<$ HN24 $>_{75}$	Solid precipitation (m)	75th percentile of 24 hour new snow amounts
$<$ HN72 $>_{25}$	Solid precipitation (m)	25th percentile of 72 hour new snow amounts
Λ_{onset}	Lag (days)	Agreement indicator of layer instability
$\Lambda_{\text{turn-off}}$	Lag (days)	Agreement indicator of layer instability
p_{unstable}	Probability of layer instability (1)	A snow stability index
$<$ RAIN24 $>_{75}$	Liquid precipitation (m)	75th percentile of 24 hour rain amounts
$<$ RAIN72 $>_{25}$	Liquid precipitation (m)	25th percentile of 72 hour rain amounts
r_c	Critical crack length (m)	A mechanical snow layer property
ρ_Ψ	Spearman rank correlation (1)	Agreement indicator of layer instability

Table 3: Names, units, and description of variables used throughout the manuscript.

1.2.9 Typo

Referee Comment: *12. Line 99, Lehning et al., 2002a, b (not b, a)*

Author Response: Done.

1.2.10 Spell out acronyms

Referee Comment: *13. General: first time important acronyms are given, spell out, e.g., HRDPS (High Resolution Deterministic Prediction System)*

Author Response: Done.

1.2.11 Include extra figure of data set

Referee Comment: *15. Section 2 for data, I would have liked to have seen perhaps 1-2 other figures (e.g., photos, bulletins, maps) representing the real data that was used. Not a strong requirement on this, but it would have been helpful to bring this back to reality of what is being modelled for the reader.*

Author Response: That is a good idea, thank you! We do not have a suitable and pleasing visual yet that displays the human hazard assessment, but are trying to get a meaningful screenshot from Avalanche Canada's forecasting dashboard. We also compiled a figure that shows a snowpack simulation at a single grid point to characterize the simulations in the data section. We aim for adding these figures to the revised manuscript but can not show them here yet.

1.2.12 Figure 2

Referee Comment: *16. Figure 2. These colours do not work in the PDF downloaded, and are very difficult to read. For example, white on bright green or white on bright blue, are not recommended. Font size getting too small. Figure caption, define all acronyms, and tell us what the different colours mean. This could be overall a stronger flowchart and figure caption as currently it would need the author next to the reader to explain what they are seeing.*

Author Response: Thank you very much for the heads up on this! We changed the text color to black, which should be a lot easier to read. We also increased the font size of the smallest annotation elements. And we expanded the caption (see Fig. 1) to (1) explain the meaning of different colors, (2) explain the confusion matrix in more detail, and (3) tell the reader that all visualized processes are explained in detail in Sections 3.1.1–3.1.5.

1.2.13 Figure 3

Referee Comment: *17. Figure 3. Similar to figure 2 in terms of colours used (hard to distinguish all of these). Perhaps use https://colorbrewer2.org/ to help you pick your colour palettes. Rather than refer us to Sect. 2.2. for colours, refer us to a table or put them actually in the caption. Mostly well explained in terms of the figure caption, except I did not follow the dashed horizontal line, and why the vertical grey line on the left of the formation line goes 'before' the arrow. I did not get 'time' here—the text states that the burial window is four days, but the solid line for the burial window is 3 days and the dash line 2 days. Can time be made clearer in text caption and the figure, that each box horizontally represents one day (I think)?*

Author Response:

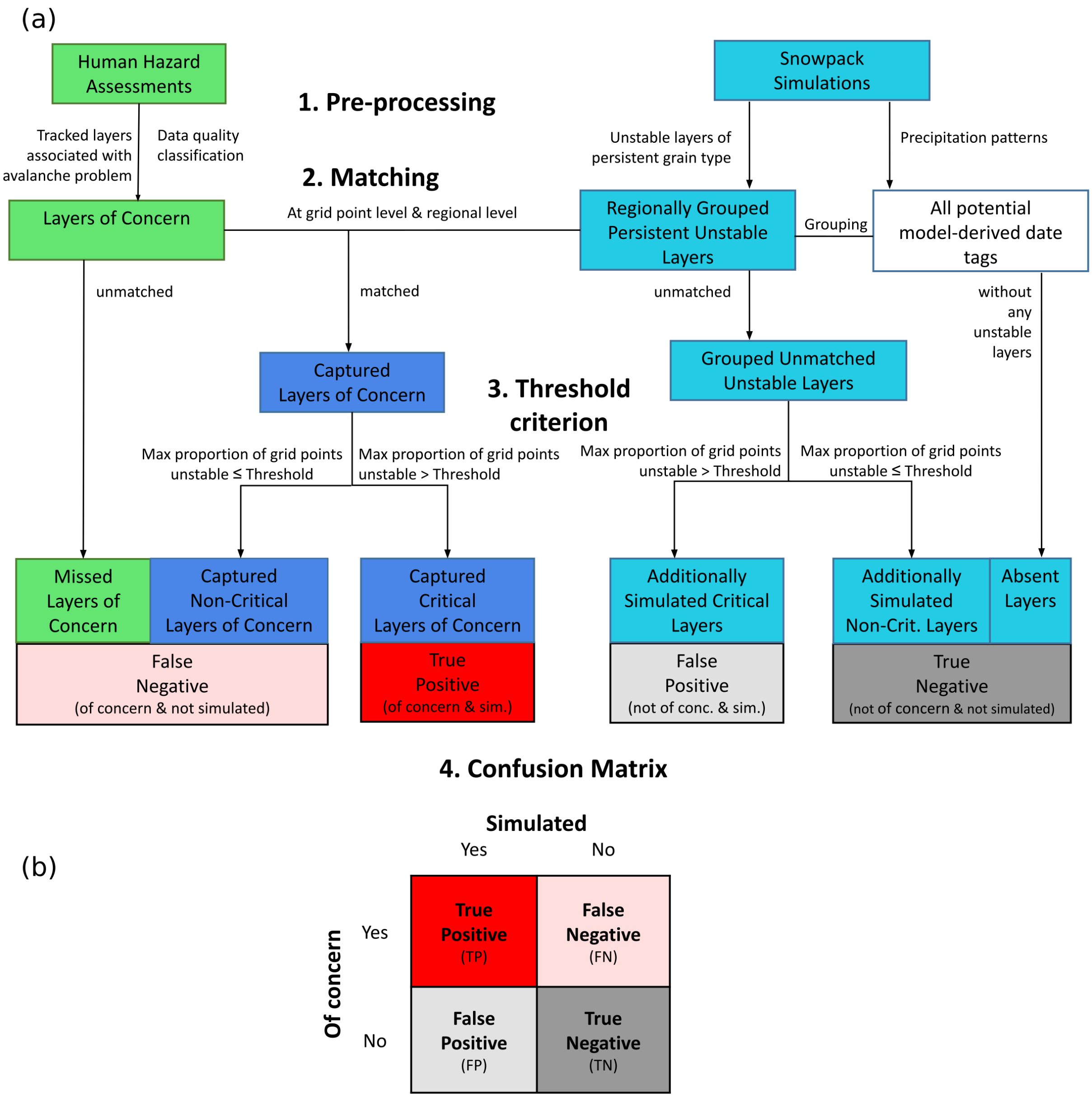

Figure 1: (a) A flowchart for assembling the four cells of the confusion matrix from the human and modeled data sets. The human data set is displayed in green boxes, snowpack simulations in turquoise, data points informed by both data sets in dark blue. (b) The resulting confusion matrix categorizes all critical layers into two dimensions: whether they are of human concern and whether they are identified by the simulation. All processes visualized by this figure are explained in detail in Sect. 3.1.1–3.1.5.

- The color palette used for grain types originates from research on designing the optimal color palette for displaying snow grain types to avalanche forecasters (Horton et al. 2020). We therefore keep the current colors.
- We reworded the reference to the acronym definitons: "Colors refer to snow grain types; acronyms for these grain types are defined in Sect. 2.2 and Table A1."
- Referee 1 already made a similar comment on the caption, so the caption was already revised to better explain the horizontal lines. After your comment, we specifically revised the sentence about the dashed horizontal line and also removed the small arrow, which seems to have rather caused confusion than helped comprehension.
- Thanks for examining this figure so closely and spotting an error on our end: The burial window is indeed limited to 5 days (like displayed in the Figure) and not 4 days (like described in the text). We changed the text accordingly.
- We included a label in the Figure that highlights that each column refers to one day and we also added it to the caption. The revised caption reads:

  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 (daily time step). 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 (potentially fixed) time windows have on the selection of layers, the gray hatched areas highlight layers that would result from time windows extended by the dashed horizontal lines. Colors refer to snow grain types; acronyms for these grain types are defined in Sect. 2.2 and Table A1.

1.2.14 Figure 4

Referee Comment: *18. Figure 4. Similar comments as above, but I was unclear about the dash vertical vs. the solid vertical line and dark grey vs. light grey in 'a'. What is the dark horizontal line in 'a'? For part 'd', can you move the 'of concern' over a bit to the 'yes' and 'no' so it is not confused as being a secondary axis for 'c'? Please make clearer the time axis, that this is 2019 to 2020 (I think). This can be signalled in the axis and in the figure caption. When I pair the figure with the text, there is a lot that feels left out in explanation (either in figure caption or in the text) and again, it almost needs the authors to be with the readers to explain each aspect. For the violoin plot in 'b', do you not need to have 0.0 to 1.0 on y-axis, and define what is meant by the white dot, and the black bars (there are MANY ways of doing violin plots, you cannot assume a give way is being shown here that everyone will automatically understand).*

Author Response:

- We made the explanation of the vertical lines (date tags) and the horizontal line in (a) clearer in both text and caption. Please consider that this paragraph is an applied example of the methodology. So a reader can not expect to understand this paragraph without having read the methods before.
- We moved the label 'of concern' over to the right as requested.
- We added labels to make the years 2018 and 2019 obvious in the figure and also state the winter season in the caption.

- We added an explanation of the violin plot to the caption.
- The y axis is shared between panels (a) and (b), so we omitted the labeling in (b) to save some horizontal space and instead show the figure a bit larger.
- The revised paragraph and the figure:

  To summarize our methodology, we illustrate the concepts applied so far with the 2019 winter season in GNP at TL (Fig. 2). The time series of the average profile (Herla et al. 2022) for the region, shown in the bottom left panel, provides context for the date tags that mark the beginning of important snowfall periods (Sect. 2.3, 3.1.2, 3.1.3). Date tags that were reported by forecasters are visualized by solid red vertical lines while the additional model-derived date tags are indicated by dashed gray vertical lines. For each regional layer represented by its date tag, a bar in the bar chart of the top left panel represents the maximum daily proportion unstable. Using 50 % as our threshold criterion for this example, each bar is colored according to its corresponding cell in the confusion matrix in the bottom right panel. The violin plots shown in the top right panel present the distributions of the maximum daily proportion unstable in more detail, while also adhering to the same colors. In this particular case, all six layers of concern (represented by the red bars, red vertical lines, red violin, and dark red cell of the confusion matrix) were well captured, and all the other simulated layers (represented by all gray features) generally had lower proportions of unstable grid points. However, there are five layers that were considered critical by the model using the 50 % threshold but not the human forecasters.

1.2.15 Page number in citation

Referee Comment: *19. [Minor] Line 273. Why are you putting in a page number (unless this is a direct quote, in which case you should have " ").*

Author Response: It is not a direct quote. However, here we cite a book that is several hundred pages long. So putting the page number to the Section of the exact topic we refer to is only to make the interested reader's life easier.

1.2.16 Comma after equation

Referee Comment: *20. Equation 2: What is FN'? [You define FN, but not FN']. Ah, never mind, I see now that it is a comma, but looks like a '*

Author Response: This is a comma indeed. We removed it to not have anyone else fall into this trap again. Thanks for pointing it out.

1.2.17 Discussion section 5.1

Referee Comment: *22. Section 5.1 is a substantive discussion on insights.*

Can these be better broken out rather than having almost three pages of narrative text, so as to make this easier for higher level reading.

Can this be brought back a bit more to the broader literature (two citations seems really few for bringing this back to the wider community and what has been done).

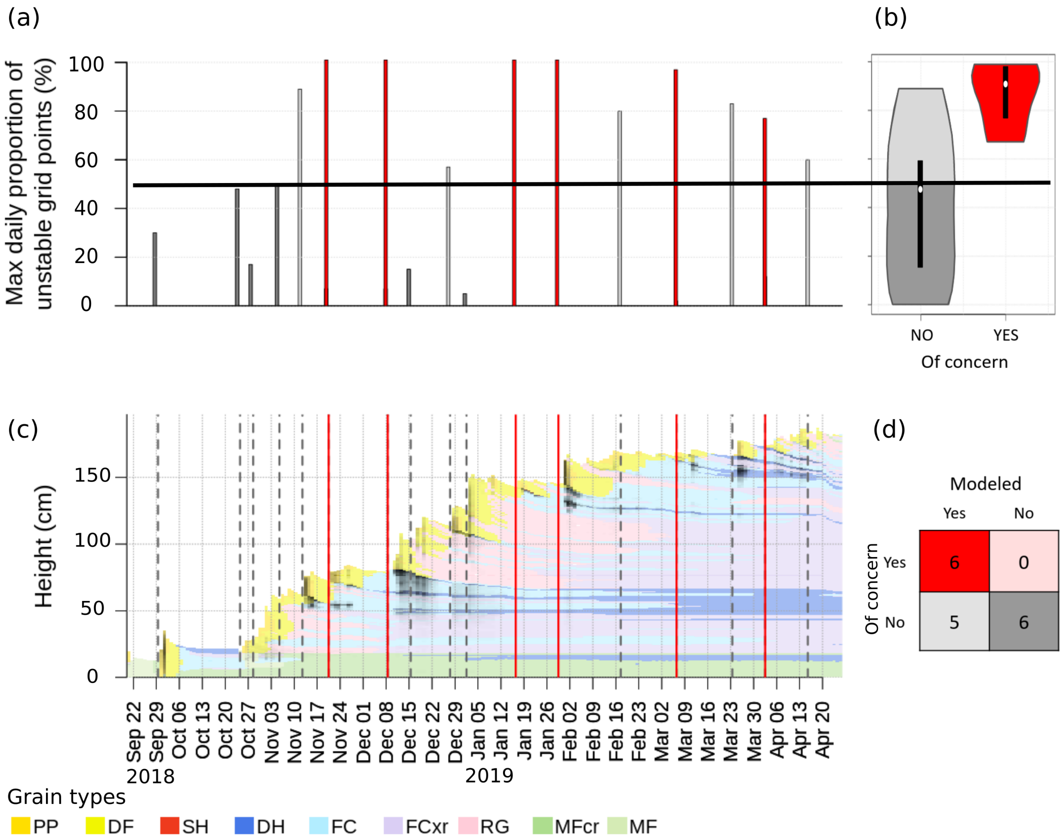

Figure 2: An illustration of the concepts applied to a case example of the 2019 season in Glacier National Park (GNP) at treeline elevation (TL). (a) Maximum daily proportion unstable for each regional layer (red: of human concern, gray: not of human concern) and an exemplary threshold criterion of 50 % (black horizontal line; shades of gray bars depend on the threshold criterion: dark gray corresponds to 'not modeled', light gray corresponds to 'modeled') and (b) the resulting distributions (colors are shared with panels a and d; boxplots within violins represent the median and interquartile range). (c) The average profile for the region and season (only shown to improve context; black shading highlights times and layers of modeled instability) with the human and model-derived date tags (red and gray vertical lines, respectively). (d) The confusion matrix evaluated for this specific example. Acronyms of snow grain types listed in the legend are defined in Sect. 2.2 and Table A1.

Author Response: In the revised manuscript, we added three subsections to brake out Section 5.1:

- A tangible interpretation of overall model performance
- Performance variations by grain types, elevation bands, and forecast regions
- Detailed comparisons of human–modeled data set

You are correct in that we do not cite many other references in this Section 5.1 (4 citations to be precise). This has the following two reasons. First, the study we present follows a validation approach that has not been carried out before (i.e., application-specific validation of critical layers on the regional scale against human hazard assessments over many seasons). Therefore there are naturally not many other references that we can meaningfully compare our results against. Due to our validation design and the simulation of snowpack stability there are indeed many interfaces to other studies. We write and cite about all these connection points in the relevant sections of the Discussion (5.2: six citations, 5.3: three citations, 5.4: seven citations), which brings us to the second reason. This Section 5.1 is specifically meant to elicit the take home points of this study, made palatable to avalanche forecasters and snowpack modelers working on applied solutions, and focusing on the Canadian context. On this intersection, there are not many other publications to bring this back to. Overall we think the Discussion as a whole finds a good balance of tying our study to the existing research.

References

Herla, F., Haegeli, P., and Mair, P.: A data exploration tool for averaging and accessing large data sets of snow stratigraphy profiles useful for avalanche forecasting, The Cryosphere, 16, 3149–3162, https://doi.org/10.5194/tc-16-3149-2022, 2022.

Horton, S., Nowak, S., and Haegeli, P.: Enhancing the operational value of snowpack models with visualization design principles, Nat Hazard Earth Sys, 20, 1557–1572, https://doi.org/10.5194/nhess-20-1557-2020, 2020.