

Authors' Response to RC3

General Response

We thank the reviewer for their detailed feedback and think that their comments improved the manuscript further. We implemented all suggestions as you can see in the point-by-point response.

In response to the issues of double-checking the data and the subjectivity inherent in the manual labeling: We are glad that the reviewer found the new manuscript is more transparent on that matter. As requested, we added further information on the additional NIR and microCT measurements to make clearer how the “double-checking” during labeling worked.

Regarding the visualizations: Figure 6) shows predictions on unlabelled data and is not intended to help with comparing manual labels to ML predictions (since we need the manual labels for that). In order to visually compare ML predictions and manual labels we intended figures similar to Figure 3. Those figures are produced for every single profile and each model and are accessible on the GitHub repository alongside with further types of visualizations that have not been included in the paper. We have not fully understood which type of visualization is requested by the spaghetti/ridge plots, but would be happy to extend the GitHub repository with this type of visualization if given more detail.

We edited the manuscript to improve the paragraph structure, flow, and language use as suggested by the reviewer. We especially find the suggestion helpful to include a visualization of NIR, microCT and SMP measurements alongside and the suggestion on how to rewrite the paragraphs for the data section.

We appreciate the time and thought the reviewer has put into their feedback!

Point-by-Point Response

- *Title: Classification is a broad term. Besides grain type classification, there could also be other types of classification, e.g. by stability of the snowpack for avalanche hazard contexts. I therefore suggest including the focus on grain types. E.g., “Automatic grain type classification of Snow Micro Penetrometer profiles with machine learning models”.*
 - We agree that classification is too broad of a term and we have adapted the title accordingly. Instead of calling it “grain type classification”, we would like to call it “snow type classification” to follow the terminology used by the international classification for seasonal snow on the ground (Fierz et al, 2009).
- *L22–24: « This knowledge helps to discern fundamental snow and climate mechanisms in the Arctic and to analyze polar tipping points. Classification of snow types is essential to assess the state of our cryosphere and is thus of interest for polar, cryospheric, and climate change research. »*
 - i) *Please include references here.*
 - Thank you for pointing this out, we included additional references in the revised manuscript.
 - ii) *In your response to referee 2 you nicely explain why knowledge about grain type profiles is relevant for remote sensing in an arctic context. Please include one or two sentences that explicitly state the need for classifying grain types and possible implications thereof.*
 - We added the following sentences: “Grain type is often better reproduced in detailed snow cover models (Vionnet et al., 2012) than their effective physical properties, especially indirectly structural anisotropy (King et al., 2015). This is especially relevant for active and passive microwave sensing, essential to map the arctic snowpack during polar night (Sandells et al., 2023).”
 - iii) *I am more familiar with the terms “grain type” or “snow grain type” than “snow type”. In the newly added paragraphs starting at L42ff you also primarily use the term “grain type”. Please use the full term “snow grain type” here in the beginning to make its meaning explicit. In the following I suggest you use one term consistently, but I leave this up to you.*
 - We completely agree that it is more reader-friendly to use one term consistently. We were not sure which of the terms to use and we settled in the end with “snow type” since this is the term used by the international classification for seasonal snow on the ground (Fierz et al, 2009). We added a footnote after using the term for the first time in the introduction, listing different forms of the term.
- *L117: « The labels indicate.. ». Please state more explicitly, e.g., “The labels expressed by color in Fig. 1 . . . ”*

- Accepted.
- *Footnote 1: Fierz et al “refer” (without s)*
 - Accepted.
- *Figure 1:*
 - i) *The background color in the inset panel refers to the manual human labels. If the colors in the main panel also refer to that manual labelling, please state explicitly. Please change the wording « ground truth » in the caption.*
 - Thank you for bringing this up - accepted.
 - ii) *I understand the discussion around expert labelling, its advantages and limitations, and I fully support the presented approach in the context of this methodological contribution and demo application (one expert who did the labelling partially with additional in-situ measurements, and one person reviewing it).*
 - We also think that it is important to have this discussion and we are glad to hear that you support the presented approach in this context!

However, I have one comment about the specific labels in the inset panel that seem inconsistent to me: The surface layer is labelled as « Rounded Grains Wind Packed » with a mean force around ~2N. The spike of ~6N at 100mm depth is labelled as « Depth Hoar Wind Packed ». I am neither an expert in the interpretation of SMP signals nor an expert of Arctic snow conditions, but it seems odd to me that DHwp would be three times harder than RGwp.

- This is among the expected range for us: DHwp falls into the type of “hard depth hoar”, described in Pfeffer et al. (2002). We added some sentences about that in Appendix B to provide the necessary information to the reader for that point.

Table B1 states that RGwp ranges between 10–40N. I don’t think that this is a big deal, and I assume that you have good reasons for your choices. Nevertheless, potential users of your tools might be situated in Alpine conditions and get similarly confused as I did. I suggest you extend your discussion of the labelling process in Appendix B with the most important particularities encountered in your data set and environment conditions. This should also include a qualitative description of the grain types that are not included in the International Classification For Seasonal Snow On The Ground (e.g., DHwp, . . .).

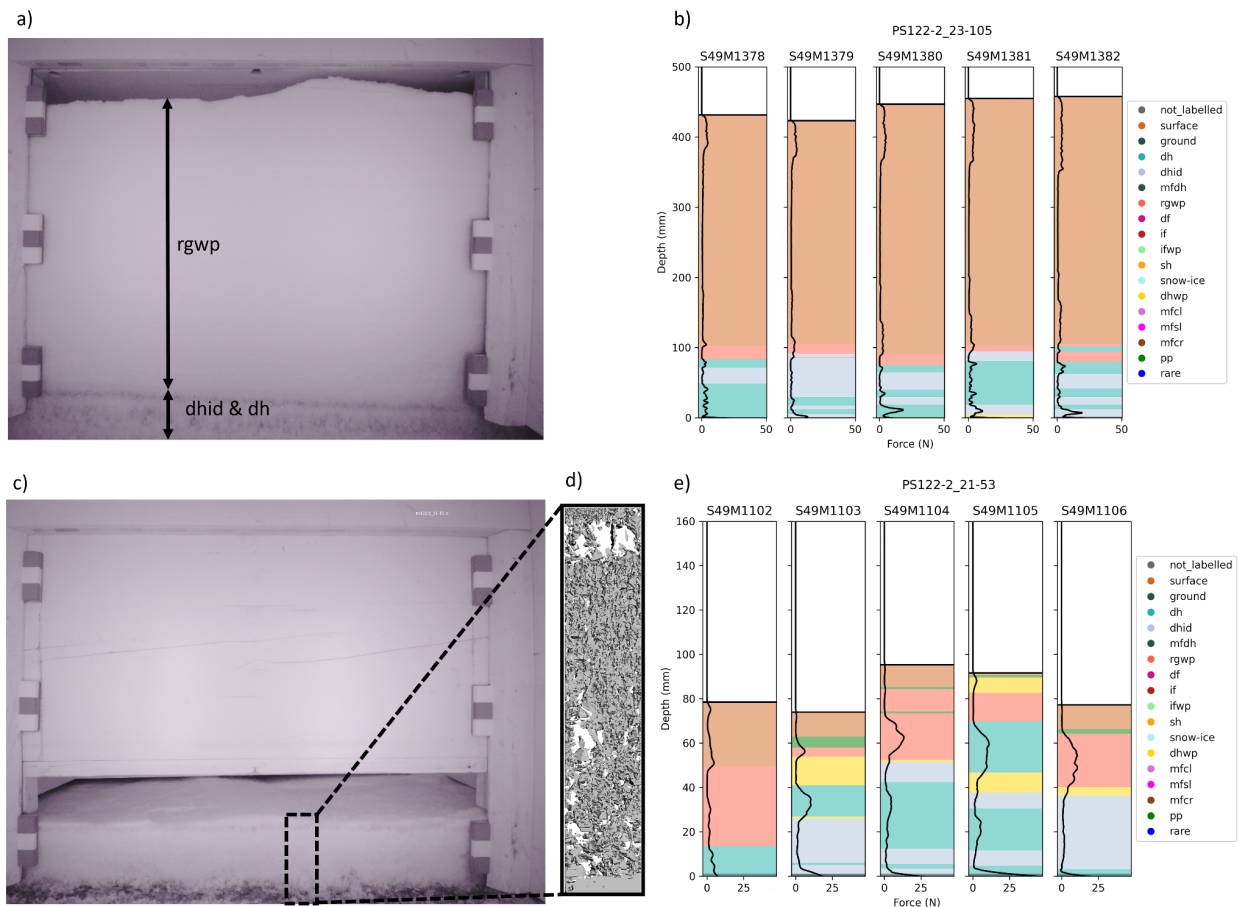
- The range is indeed too high, we corrected this to 2-20 N.
- Since an in-depth description of the environmental conditions and the particularities encountered during the data collection is beyond the scope of this paper, we refer the reader to the following publication instead: <https://online.ucpress.edu/elementa/article/9/1/00023/118092/Meteorological-conditions-during-the-MOSAiC>
- We added additional grain type descriptions, however, only dhwp and mfdh seemed to need additional explanations since all other grain types are included in the international classification. We are happy to extend this section further if

requested.

You can also consider picking an SMP signal for Figure 1 that demonstrates unambiguous (textbook-like) labels. Please see this comment also in light of my general comment earlier. The fact that all three referees spent considerable amount of thinking around the subjectivity of the manual labels and the resulting impact on the predictions highlights the potential benefit of a more in-depth approach (/visualization) to look at the data set.

- We hope that the extension of Appendix B addresses your comments. The SMP signal in the inset panel serves only the purpose of showing how the SMP signal looks like in general and should inform the reader about the general task at hand. The inset panel is not supposed to demonstrate how the signal of specific snow types look like (textbook-like). For that, we prefer to forward the reader to the Appendix B where the signals are visualized for each snow type. There, the signals are more visible than in the inset panel of Figure 1, plus we can show the whole range of signals and not just a subset.
- *L113–146: The flow of the paragraphs and their storyline jumps back and forth a little bit. I suggest reordering the sequence of the paragraphs. e.g., (1) Intro to MOSAiC (incl. measurements and conditions), (2) explicit statement about the data used for this study, (2a) SMP, (2b) Micro-CT: please explain and cite, mention 1 full term before using acronym, state how many Micro-CTs were available, (2c) NIR: please explain and cite, mention full term before using acronym, state how many photographs were available.*
 - Thank you, we agree that the flow can be improved. We restructured the content as suggested and added the information you requested. We added further information, especially on the numbers, citations, and explanations on NIR and micro-CT.
- *L153: Poisson shot noise model*
 - Accepted.
- *Figure 2: Please change the label « Ground Truth » in panel (a).*
 - Accepted.
- *L536: Can you cite these data sets?*
 - We added three examples of publicly available SMP datasets.
- *Appendix A: I appreciate this detailed user guide very much. Consider referring readers to that appendix when introducing the high-level process of labelling in the main body.*
 - Thank you - we have added a reference at the end of the introduction when outlining the rest of the paper. We thought that might be the best place (grouped together with data & code availability) since the user guide includes so many different aspects (collection, labeling, model selection, etc.).

- *Appendix B: Please include at least a brief explanation of how the Micro-CT and NIR measurements were used to inform/nudge/validate the grain type labelling of the SMP signal. I personally would appreciate one example figure that shows how an additional in-situ measurement made a challenging SMP signal more interpretable. A good example could be the signal in Fig. 3 that raised a concern with referee 1. – See related comment about Appendix B that I raised earlier (when writing about Fig 1).*
 - Thank you for that suggestion. Unfortunately, the requested SMP profile does not have any corresponding microCT or NIR images. We choose a set of SMP profiles instead for which we retrieved microCT and NIR images. Please take a look at Figure B2 that we included upon your suggestion. The new subsection of the appendix intends to give the reader a general intuition of how the different measurements are used in parallel, however, an in-depth explanation is beyond the scope of this paper. For further information, we refer the reader to the metadata section in the corresponding dataset publication (<https://doi.pangaea.de/10.1594/PANGAEA.935934>).



- *Figures Appendix B: I appreciate the figures that demonstrate typical SMP signals for different grain types very much. It would help a lot to see those signals in subplots in one figure on one page, rather than spread out. To make it even easier for the reader, you could clip the x-axis to the window that shows the desired signal and not show the*

remainder of the measurement. At least a background shading that highlights the window of interest would be nice.

- We summarized the figures into one figure with 8 panels spread across two pages. We hope that this gives a better overview. We highlighted the windows of interest with a thin red box. When there is no red box it means that the complete signal represent the mentioned snow type.
- *Appendix D: I appreciate the definitions of variable names, features, and metrics. Please reference these definitions in the main body of the manuscript when new terms come up.*
 - The appendix about the metrics is referenced when the metrics are mentioned in the Section “Evaluation.”
 - The appendix about the features and variable names is referenced in the data section. We added the reference at several other locations where applicable.
- *Figure G1: Please cross-reference the list in Appendix D that explains the variable names. When I jumped to the decision tree figure from the main body of the text I had skipped Appendix D, and was missing exactly this piece of information.*
 - We included this reference now.
- One last question out of curiosity: What inspired the name « snowdragon » ;-)?
 - First author responding on her own: I am just a very big fan of dragons (they are my favorite fantasy animal). The snow-layer structure of a snowpack reminded me of the scales of a dragon - for some reason. And dragons are hoarding - since we implemented so many different ML models, I had the feeling we are building a dragon hoard of ML models at some point. As you can see it was purely association-driven :)

Other Revisions

- “Labeling” instead of “labelling”, “labeled” instead of “labelled”
- Requested by copernicus:
 - We adjusted the list of corresponding authors
 - C2 and H1 are both inserted as figures and called / referred to as “Tables”, hence the request should already be satisfied?