

# egusphere-2022-938 referee report 3

## Detailed comments

Please note that all line numbers refer to the ATC2 document where changes from the previous manuscript version are highlighted.

- *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”.
- 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.
  - 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.
  - 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.
- L117: « The labels indicate.. ». Please state more explicitly., e.g., “The labels expressed by color in Fig. 1 ...”
- Footnote 1: Fierz et al “refer” (without s)
- 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.
  - 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). 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. 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, ...). 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.
- L113–146: The flow of the paragraphs and their storyline jumps back and forth a little bit. I suggest to reorder 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

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.

- L153: Poisson shot noise *model*
- Figure 2: Please change the label « Ground Truth » in panel (a).
- L536: Can you cite these data sets?
- 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.
- 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)
- 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.
- 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.
- 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.
- One last question out of curiosity: What inspired the name « snowdragon » ;-)?