

Reviewer 1

The article demonstrates that mechanical snow profiles can be statistically evaluated and stratigraphic features reconstructed. Traditional snow profiles are limited in number due to limitation of field work. Using 921 snowmicropen profiles from different expeditions, the authors build on the work of Kaltenborn et al, who demonstrated the use of machine-learning methods to evaluate highly resolved force profiles in snow. Based on their parameterisation and algorithm, significant differences between first-year and multi-year ice floes are quantified.

The statistical quantification of snow properties is an important step towards synthetic modelling of the snowpack on Antarctic ice floes. The data are also essential to improve detailed snow pack models and develop stochastic modelling tools for these stratigraphically complex snowpacks. Concerning remote sensing, the data help to understand variation in albedo and microwave brightness temperatures. However, the article lacks standard statistical tests between observables.

On these lines, the paper is a major step towards a much more quantitative and statistically testable snowpack.

The paper is well structured. Propositions to improvements are detailed in the "Detailed comments". Figures should be adapted to the final format of TC, and the layout unified.

We thank Reviewer 1 for the careful assessment of our manuscript and for the detailed and constructive review. The comments helped us to improve the presentation and consistency of the figures, clarify methodological choices, strengthen the statistical comparison between first-year and multi-year ice, better define the reliability limits of the spatial autocorrelation analysis, and expand the discussion of methodological limitations. Below, we provide a point-by-point response. Reviewer comments are shown in black text, our responses in blue text. Manuscript text is quoted and manuscript changes are set in bold. Line numbers refer to the preprint manuscript.

We have revised the figures to better follow The Cryosphere formatting requirements and to ensure a consistent layout, typography, panel labeling, and units throughout the manuscript.

Detailed comments

Title: could be more specific, eg.

Different spatial scales and snow stratigraphy between first- and multiyear ice floes in the Weddell Sea, Antarctica

We agree that the original title could be more specific and should make clear that the study focuses on internal snow properties and their spatial variability. We have therefore revised the title.

The revised title reads: **“Variability of internal snow properties over Antarctic late summer sea ice on different spatial scales”**

Introduction: compare here the difference to the Arctic sea ice, cite the work of Sturm et al.

line 22: a short introductory paragraph on the work in the Arctic on sea ice (traditional: Sturm et al, SMP the pioneering work of King, then the MOSAiC) that forms in many ways the base to improved methods for the Antarctic.

(This is to some degree covered in lines 68 ff, but it much more in detail)

We agree that the studies by Sturm et al., King et al., and the SMP work conducted during MOSAiC have contributed substantially to our understanding of snow on Arctic Sea ice and provide an important methodological basis for subsequent sea-ice snow studies. However, because the focus of the present manuscript is on Antarctic snow rather than a direct hemispheric comparison, we have kept the Arctic background focused on the studies most relevant to our methods and objectives, which are presented from line 68 onwards.

Lines 112-116: are the data of these other measurements available? Then cite data location, references. Otherwise delete

We agree that the data sources for the accompanying sea-ice physics and biological measurements should be identified more clearly. We have therefore added the relevant references to the manuscript and clarified which observations are used in the present study:

“At all ice stations, a comprehensive sea ice physics and biological program was carried out, comprising measurements of sea ice thickness and snow depth distributions across individual ice floes, extensive ice coring for subsequent analyses of internal sea ice structure and biogeochemical composition, vertical snow characterization, and, at selected stations, the deployment of autonomous ice-tethered platforms (Arndt et al., 2018, Arndt et al., 2021, Haas et al., 2021). **The present study uses the snow pit, SnowMicroPen (SMP), snow depth and sea ice thickness observations described below.**”

Figure 1: The background sea ice concentration is only for 2018. Likely it was different for 2019 and 2021 - relevant to the interpretation of the data? The geographic separation between PS111/PS124 (southeastern Weddell) and PS118 (northwestern Weddell) confounds the FYI/MYI comparison, since regime and region covary almost completely.

We thank the reviewer for highlighting both issues. We agree that the sea ice conditions differed among the three campaigns and have therefore revised Figure 1 to show the mean sea ice extent for February 2018, February 2019 and February 2021 rather than a single sea ice concentration.

The covariation between sampling region and ice regime is addressed separately as a methodological limitation in Sect. 5.3 (line 495 onwards; see our response to the corresponding Discussion comment below).

line 155: calculate a histogram of transect length and present. The calculation of autocorrelation are only statistically significant to about half of the total transect length (see Cliff & Ord 1981 on Moran's I confidence intervals).

We agree that the transect-length distribution and the decreasing statistical support at larger lag distances should be reported more explicitly. The 26 transects included in the correlation analysis range from 19 to 64 m. 17 are located on FYI and 9 on MYI, with mean lengths of 49 and 40 m, respectively. Because this relatively small distribution can be summarized compactly, we report the length ranges and means in the text rather than adding a separate histogram. We also restrict the physical interpretation of Moran's I to lag distances of approximately 35 m.

We have added a description of the present transect lengths to line 165 in Section 2.3: **“For correlation length analysis, only transects with 10 profiles or more were considered. Out of those 26 longer transects, of which 17 transects were on FYI and 9 on MYI, 10 transects between 36 and 50 m and 6 transects were 19-31 m long. Transects lengths on FYI were on average 49 m and on MYI 40 m.”**

We also revised line 178 accordingly. The final sentence of Section 2.5, now reads: **“Because the number of transects and profile pairs decrease strongly for larger lag distances, we restrict the physical interpretation to lag distances of approx. 35 m. Values at greater distances are not interpreted as robust spatial signals.”**

line 163: why was the unmodified (?) Calonne parameters used? This model is calibrated to alpine snow at rather high snow temperature (average approx -5° C). Explicitly justify this choice and discuss potential biases.

We agree that the choice of parameterization and its potential biases should be justified more explicitly. King et al. (2020) provide sea-ice-specific density coefficients but no corresponding SSA retrieval, whereas our analysis requires a consistent framework for both properties. Snow-pit temperatures in our dataset ranged mainly from 0 to -7 °C, with a mean of -3.3 °C and a mode of -1.7 °C, and therefore fall within the temperature range relevant to the Calonne et al. (2020) parameterization. In addition, this parameterization showed the best agreement with the co-located snow-pit density profiles and stratigraphic observations. Potential differences in absolute values arising from alternative parameterizations are additionally discussed in Sect. 5.3.

The revised text from line 163 onwards reads: “With appropriate assumptions and calibration, penetration resistance data **can be** converted to snow density and SSA following the approach of Calonne et al. (2020). **This parameterization was selected because it produced density profiles consistent with the co-located snow pit density and stratigraphy observations and provides both density and SSA, which are required for the subsequent analysis. Although sea-ice-specific SMP density coefficients have been proposed by King et al. (2020), no corresponding SSA retrieval is available.**”

line 214: "weighted Cross-Entropy Loss function": could you add a brief justification for the weighted version?

We agree that the purpose of the weighting should be stated more explicitly. The inverse-frequency class weights compensate for the unequal occurrence of the snow classes, preventing the more frequent classes from dominating the loss and reducing the risk that good overall performance masks poor classification of less frequent classes. We have restructured the text to connect the class weights directly to the weighted cross-entropy loss.

The revised text reads in line 210 and following:

“To address **the** imbalance between **snow class occurrence**, class weights inversely proportional to their frequency of occurrence were applied in the calculation of the **Cross-Entropy** loss function. **Without weighting, the more frequent snow classes like rounded and faceted would contribute more strongly to the loss function, so that the model could achieve good overall performance while performing poorly on the less frequent classes.**”

And in line 214:

“We therefore set up the experiments as a supervised learning, multiclass problem, and used the weighted Cross-Entropy Loss function **accounting for class imbalance** to guide the optimization process.

line 246 -248: repeated statement: "Rare ..." (cf line 209)

We thank the reviewer for pointing out the apparent repetition.

The two passages refer to different aspects: the first identifies the rare or weakly expressed classes excluded from model training, whereas the second explains the possible consequence of this exclusion when the model is applied to the full dataset. We have revised the latter sentence to make this distinction explicit.

Line 246 ff now reads: “**Because** rare or weakly expressed snow types, such as new snow, surface hoar, or thin ice lenses, were excluded from model training due to insufficient sample sizes, **their occurrence in the full dataset** may be locally classified as the most similar major snow type.”

line 268: what do you mean by "and the finite vertical support of the SMP signal"?

We agree that the phrase "finite vertical support of the SMP signal" was unclear. We have therefore removed it.

The sentence now reads: **"Such effects reflect the non-discrete nature of snow stratigraphy."**

line 274: "by a single operator": The authors acknowledge single-operator labelling as a source of subjectivity but do not discuss whether this represents the dominant source of uncertainty relative to the ambiguity of the SMP signal itself. A brief discussion of this hierarchy of uncertainties would strengthen Section 3.3.

We thank the reviewer for raising this important point. We have expanded Sect. 3.3 to clarify that operator-dependent uncertainty and ambiguity in the SMP signal are closely linked and cannot be quantified independently with the available annotations. We also note that repeated labelling by independent operators would be required to assess their relative contributions.

The revised manuscript now reads in line 274 ff: " ...where expert interpretation of SMP signal characteristics is ambiguous. **Manual labelling by a single experienced operator provided a consistent interpretation across all profiles and avoided additional inter-operator variability. The remaining labelling uncertainty is closely linked to inherent ambiguity in the SMP signal and is expected to be greatest around transitions and where different snow classes produce similar responses. Since operator-dependent uncertainty arises primarily in these ambiguous intervals, the two sources cannot be considered fully independently. Repeated annotations by independent operators were not available, so their relative contributions could not be quantified, but such an inter-operator comparison would provide a useful direction for future work.** Also, while manual stratigraphic analysis may distinguish multiple adjacent layers..."

line 293-297: paragraph repeated from the methods section (cf lines 120 ff)

We agree that the original paragraph repeated information already presented in the Field Data section. We have shortened it to retain only the information needed to define the dataset used in the Results. In response to Reviewer 2, we have also clarified the thermal state of the snow during sampling.

The revised paragraph now reads:

"The results presented here are based on the full set of processed SMP data, including both manually labeled and automatically classified profiles, as described above, excluding the snow pit data. **Although sampling took place during the austral summer, observed snowpack temperatures ranged from -7°C to 0°C with a mean temperature of -3.3°C throughout all campaigns, indicating that the snowpack was not undergoing active surface melt at the time of measurement.** For the analysis, the dataset is separated into first-year ice (FYI) and multi-year ice (MYI) classes [...]"

Table 3 between FYI and MYI (density, layer thickness, SSA, snow type fractions) lacks any formal significance testing throughout the paper. A Welch's t-test or Wilcoxon rank-sum test would be straightforward additions.

We agree that the original manuscript relied primarily on descriptive comparisons between FYI and MYI. We have therefore applied Welch's t-test, which does not assume equal variances, to the comparisons summarized in Table 3, using $p < 0.05$ as the significance threshold. For snow-type-specific density and SSA, the SMP-derived values were first averaged within each individual layer so that vertically correlated measurements from the same layer were not treated as independent

observations. Table 3 now marks statistically significant differences in means in bold, and Figures 4 and 5 have been revised to show the corresponding distributions of layer-mean SSA and density. The relevant Results and Discussion sections have also been updated to distinguish significant from non-significant differences.

The following text was added to the methods in Sect. 2.4:

“For snow-type specific density and SSA, values were averaged over each contiguous classified snow layer. Probability density distributions, and statistical comparisons of snow-type density and SSA were calculated from the resulting layer-mean values. To test for differences in the mean values of the analysed parameters between FYI and MYI, Welch’s t-test was applied, which does not assume equal variances, with a significance threshold of $p < 0.05$.”

line 320 "is closely related to the snow grain size": this is only the case in theory: see the paper by Painter, Thomas H., Noah P. Molotch, Maureen Cassidy, Mark Flanner, and Konrad Steffen.

“Contact Spectroscopy for Determination of Stratigraphy of Snow Optical Grain Size.” Journal of Glaciology 53, no. 180 (2007): 121–27. <https://doi.org/10.3189/172756507781833947>.

We thank the reviewer for this clarification. We agree that a simple one-to-one relationship between SSA and grain size is not generally valid. We have therefore removed the statement and retained only the direct physical definition and relevance of SSA.

The revised manuscript now reads in line 320:

“The specific surface area (SSA) of snow quantifies the interfacial area between the ice crystals and the pore air enclosed within the snowpack and influences key physical snow properties such as radiative transfer and albedo.”

line 347: replace "in the methods in Section 2.5" by "(see Section 2.5, page xx)"

We have revised the cross-reference.

The sentence now reads:

” The spatial autocorrelation of fractional snowpack composition is computed using Moran’s I (see **Section 2.5**).”

We have omitted the page number because it may change during typesetting.

Figure 7: very interesting data. I guess however that all values and interpretation beyond 30 m are spurious (see comment to line 155).

We agree that Moran’s I values at the largest lag distances are not sufficiently supported for physical interpretation. Based on the decline in the number of contributing transects and profile pairs, we restrict interpretation to lag distances of approximately 35 m and have stated this explicitly in line 535 in Sect. 4.4.

The revised sentence now reads: **“As the number of profile pairs and transects decrease for larger lag distances, Moran’s I values at lag distances larger than 35 m cannot be interpreted as robust signals.”**

We have also added this restriction to the length scale discussion in Sect. 5.2

The revised text from line 453 ff now reads: “... whereas FYI snowpack correlations decayed rapidly over the first tens of meters from 0.4 - 0.6 down to below 0.2 and stayed low **thereafter**. **However, it needs to be kept in mind, that signals past 35 m cannot be considered statistically robust. Nevertheless, this contrast indicates ...**”

line 355 ff: Are these differences statistically significant? Perform a test.

We agree. This point is addressed by the Welch's t-tests described in our response to the Table 3 comment above. The corresponding Results text has been revised to state explicitly which FYI-MYI differences are statistically significant.

Discussion

I wonder why you did not apply standard statistical tests to all your comparisons between FYI and MYI. Are the differences significant? Which differences are significant?

We agree that formal tests are needed to distinguish robust FYI-MYI differences from descriptive contrasts. As described above, we have applied Welch's t-tests to the comparisons summarized in Table 3 and revised the Results and Discussion to distinguish significant from non-significant differences.

The dataset has 19 FYI floes versus only 9 MYI floes, and the MYI floes are almost entirely from a single expedition (PS118). This is a notable limitation, as it could bias the spatial autocorrelation comparisons.

We agree that the unequal number and regional distribution of FYI and MYI floes limit the attribution of the observed differences solely to ice regime. Most MYI floes were sampled during PS118 in the northwestern Weddell Sea, whereas most FYI floes were sampled during PS111 and PS124 in the southeastern Weddell Sea. We have therefore added the coupling between ice regime and sampling region to the methodological limitations in Sect. 5.3 and now make clear that the FYI-MYI contrasts may also reflect regional differences.

The revised discussion from line 495 on now reads:

“The interpretation of Antarctic snow variability derived from SMP measurements and machine-learning-based snow-type classification is subject to several methodological limitations. These primarily affect the absolute quantification of snow properties and the resolution of basal and fine-scale stratigraphy, with implications for interpreting variability across spatial scales.

The sampling design limits the attribution of differences specifically to a single ice-regime because ice-type is partly coupled to the sampling region. Most MYI floes were sampled during PS118 in the northwestern Weddell Sea, whereas most FYI floes were sampled during PS111 and PS124 in the southeastern Weddell Sea. The FYI-MYI contrasts therefore also reflect regional differences. However, because the southeastern Weddell Sea is typically dominated by FYI and the northwestern Weddell Sea by MYI (Gupta et al., 2025), the coupling between ice regime and sampling region reflects, at least in part, the natural regional distribution of these ice types. The observed contrasts should therefore be interpreted as differences between the sampled FYI- and MYI-dominated environments rather than as region-independent effects of ice regime alone.”

Practical recommendation for sampling in future campaigns?

We agree that the practical implications for future sampling should be stated clearly. The discussion therefore emphasizes two recommendations: sampling multiple floes rather than concentrating measurements on a single floe (line 465 onwards), and using a higher within-floe sampling density for layer-resolved properties than for bulk properties (line 480 onwards).

Code availability: TC data policy request that the code must be made public. Without code results are irreproducible, as the machine-learning part depends on subjective choices. This is strong request for the next round of reviews.

We agree that public availability of the classification code is necessary to ensure reproducibility, particularly because the machine-learning workflow includes methodological choices that should be transparent.

We are currently finalizing the repository and will make the code available on GitHub. The repository link will be added to the Code Availability statement.

Reviewer 2

The study investigates the spatial variability and characteristic length scales of snow properties on Antarctic sea ice during late summer. This is important because snow significantly affects sea ice thermodynamics, mass balance, and remote sensing. However, knowledge about its variability is limited beyond local scales. Researchers combined manual snow pit observations with more than 900 high-resolution SnowMicroPen (SMP) profiles collected during three expeditions in the Weddell Sea from 2018 to 2021. Using a supervised one-dimensional convolutional neural network (1D-CNN), they classified snow microstructural layers and types from SMP data. The paper focuses on characterizing snow properties for FYI and MYI to improve representation of snow on Antarctic sea ice. I only have specific comments.

We thank Reviewer 2 for the positive and careful review of our manuscript. The comments helped us to improve the clarity of the terminology, better justify the SMP-derived density and SSA retrievals, clarify the snow state during the late-summer measurements, and more carefully frame the implications for remote sensing and upscaling. Below, we provide a point-by-point response. Reviewer comments are shown in black text, our responses in blue text. Manuscript text is quoted and manuscript changes are set in bold. Line numbers refer to the preprint manuscript.

Line 8: which regimes?

We agree that the wording was ambiguous. We have clarified that the statement refers to the two sea-ice regimes considered in this study, first-year ice and multi-year ice.

Line 8 now reads: “Across both **sea-ice** regimes, intrinsic properties of individual snow types, including density and specific surface area, were remarkably similar.”

Line 13: 50% of the variability of what? Snow density...

We agree that the original sentence did not specify which property the variance estimate referred to. We have clarified that this statement refers to snow bulk-density variability. This is consistent with the analysis in Sect. 4.5, where floe-wise and pooled variances are calculated for bulk snow density.

Line 13 now reads: “Individual ice floes capture only about 50 % of the **snow density** variability characteristic of their respective ice regime, underscoring fundamental limits to the representativeness of point measurements.”

Line 60: I would mention SSA here if you are going use it.

We agree and have added SSA to the description of the SMP-derived snow properties, as SSA is an important parameter throughout the manuscript and is later discussed in relation to radiative transfer and remote-sensing applications.

The sentence now reads: “The SMP offers an efficient approach to rapidly characterize snow structural properties, including density **and SSA**, through mechanical resistance profiles in sub-millimeter-scale vertical resolution and independent of the observer.”

Line 73: microstructural

We agree and have changed the wording.

The sentence now reads: “However, snow **microstructural** properties such as grain or snow types are not directly measured by the SMP and must be inferred from characteristic features of the penetration signal.”

Line 107: What is the mean temperature? does that mean that the snow is wet? Can you describe what the effect is on the snowpack if temps are above zero.

We agree that the snow thermal state could be described more clearly, especially because the measurements were conducted during late austral summer. We have added a clarification that the measured snow temperatures were below 0 °C during the campaigns. This indicates that the snowpack was not undergoing active surface melt at the time of measurement. The manuscript already discusses that SMP force signals can be sensitive to temperature, liquid water content, and salinity, which can increase uncertainty in inferred snow properties during melt or flooding events.

We therefore clarify the state of the snow in Section 4, line 295 (see specific comment below).

Line 163: maybe a word on coefficients? King et al. (2020) provided coefficient for density on sea ice but no ssa...

We agree that the choice of SMP parameterization requires further clarification. We selected the Calonne et al. (2020) approach because it provides both density and SSA, which are required for the subsequent analysis, and because it showed the best agreement with the co-located snow-pit density and stratigraphy observations in our dataset. We now also acknowledge that King et al. (2020) provide sea-ice-specific density coefficients, but that no corresponding SSA retrieval is available in that framework. This distinction is important because our study benefits from a consistent retrieval framework for both density and SSA. Your notes already identify this as one of the key points.

The revised paragraph from line 163ff now reads:

“With appropriate assumptions and calibration, penetration resistance data **can be** converted to snow density and SSA following the approach of Calonne et al. (2020). **This parameterization was selected because it yields density profiles consistent with the co-located snow pit density and stratigraphy observations and provides both density and SSA, which are required for the subsequent analysis. Although Arctic sea-ice-specific SMP density coefficients have been proposed by King et al. (2020), no corresponding SSA retrieval is available.**”

Line 189: Is this how it was done in Kaltenborn et al. (2023)?

Yes. We have clarified this in the manuscript by explicitly stating that the initial identification of characteristic SMP signal patterns from co-located snow pit observations follows the approach suggested by Kaltenborn et al. (2023).

The sentence now reads: “In an initial step, all SMP profiles co-located with the 50 snow pit sites were analysed to identify characteristic signal patterns associated with snow types determined from the corresponding snow pit observations, **as suggested in Kaltenborn et al. (2023).**”

Line 202: Maybe mention these parameters are also used to derive density and SSA.

We agree that this connection can be made explicit. We have added a sentence explaining that the Poisson shot-noise parameters are also the parameters used to derive density and SSA from the raw SMP force signal.

The revised text reads:

“Furthermore, the parameters deflection at rupture in mm (δ), mean rupture force (f_0), element size (L), intensity of point process (λ), as well as the median force (F_m), calculated within a 4 mm context window, were derived using the Poisson shot-noise model of Löwe and Van Herwijnen (2012). **These are the same parameters used to derive density and SSA from the raw force signal.**”

Line 210: Switch the order to match "Rare or weak". New snow is not rare...

We agree and have revised the wording so that the order of examples matches the description of "rare or weakly expressed layers."

The sentence has been changed to:

"Rare or weakly expressed layers, such as **surface hoar or new snow** were excluded."

Line 295: everything is still frozen? Clarify the snow state with respect to melting.

We agree that the snow state during the late-summer measurements could be stated explicitly. We have added a sentence clarifying that observed snowpack temperatures were below 0 °C throughout the campaigns, indicating that the snowpack was not undergoing active surface melt at the time of measurement.

Line 295 now reads:

"Although sampling took place during the austral summer, observed snowpack temperatures ranged from -7°C to 0°C with a mean temperature of -3.3°C throughout all campaigns, indicating that the snowpack was not undergoing active surface melt at the time of measurement."

Table 3: SSA values are key for remote sensing. I think it should be emphasized more in the conclusion or abstract.

We agree that SSA can be emphasized more strongly, because it is a key microstructural parameter for radiative transfer and microwave remote-sensing applications. We have therefore added a sentence to the abstract highlighting that the consistency of SSA within snow types across sea-ice regimes may help constrain snow-type-specific inputs to microwave emission and radiative transfer models. This addition is consistent with the manuscript's finding that snow-type-specific SSA differences between FYI and MYI are generally small, while snowpack structure and snow-type fractions differ more strongly.

Line 19 in the abstract now includes: **"In particular, the relative consistency of specific surface area within snow types across ice regimes provides observational constraints that can be evaluated for future microwave-emission and radiative-transfer modeling studies."**

Line 550 in the conclusion now reads:

"... particularly on multi-year ice. This consistency suggests that snow-type-specific SSA values may be transferable across ice regimes, offering a basis for parameterizing snow-type-dependent inputs to microwave emission and radiative transfer models, even as bulk snowpack properties diverge structurally between FYI and MYI. "

Line 324: Probably because of more developed facet on MYI or melted Facet which will have a small SSA.

We thank the reviewer for this helpful interpretation. We have revised the discussion of the bimodal SSA and density distributions in rounded and faceted layers on MYI to explicitly mention more developed or melted facets. This provides a more physically specific explanation for the observed lower-SSA and density mode in MYI snow.

Line 425 in Section 5.1 now reads:

"For rounded and faceted layers in this study, both density and SSA distributions transitioned from unimodal on FYI to bimodal on MYI, likely reflecting the coexistence of decomposing precipitation and older, more developed or melted facets."

Line 344: 35% of MF... is this even with accounting for the basal layer that couldn't be measured by the SMP on MYI?

Yes, the reported melt-freeze fraction is based only on the part of the snowpack that was penetrable by the SMP and could therefore be classified. We agree that this should be stated more clearly, especially because hard basal melt-freeze or ice layers can prevent full SMP penetration, particularly on MYI. We have clarified this in the results and expanded the methodological limitations to state that melt-freeze fractions and density estimates are biased toward the penetrable part of the snowpack. The manuscript already notes that the SMP reached its force threshold primarily on MYI floes with extensive basal melt-freeze layers.

In line 339, Sect. 4.3 we added:

“The relative fraction of each snow type within individual snow profiles is calculated, **based on the classification of the penetrable snowpack.**”

In line 498 in Sect. 5.3 we revised the limitations: “As a result, density **and snowtype fraction** estimates for melt-freeze layers are biased toward the penetrable fraction, and bulk snow density may be underestimated where thick, hard layers occur, predominantly on MYI (Arndt, 2022a).

Line 372: “indicating that bulk density is the most spatially homogeneous property.” That is good for remote sensing retrieval.

We agree that the comparatively low variability of bulk density is relevant for large-scale applications. We have added a short pointer in the results and expanded the discussion in Sect. 5.2 to distinguish between applications that rely mainly on bulk-integrated properties, such as first-order snow-load or SWE estimates, and applications that are sensitive to internal stratigraphy.

Line 372 now reads: "In contrast, lowest RSD values are found for bulk snow density, indicating that bulk density is the most spatially homogeneous property, **a favourable characteristic for large-scale snow-load estimates, as discussed further in Sect. 5.2.** However, individual layer densities exhibit RSD values approximately twice as high as those of bulk density."

Line 430: King et al. (2020) measured data in winter.... this study in the summer. How do you think this affect the comparison.

We agree that the comparison with King et al. (2020) is affected not only by hemisphere but also by season. King et al. (2020) measured Arctic snow under winter conditions, whereas our dataset represents late austral summer conditions in the Antarctic. Therefore, the observed contrast may reflect both hemispheric differences in sea-ice surface processes and seasonal differences in snow metamorphism. In particular, late-summer Antarctic snow is more likely to preserve signatures of melt-refreeze modification, flooding, snow-ice formation, and superimposed-ice processes, whereas winter Arctic snow is more strongly shaped by cold-season accumulation and temperature-gradient metamorphism. We have therefore revised the discussion.

The paragraph now reads:

"Finally, we note a hemispheric contrast, though a direct quantitative comparison is limited by the fact that King et al. (2020) sampled Arctic winter snow, whereas our dataset represents Antarctic late summer. In their study, Arctic winter FYI showed higher bulk density than MYI, linked to a larger fraction of low-density depth hoar on MYI and pronounced surface heterogeneity. Our Antarctic snowpacks, by contrast, were sampled at the opposite end of the seasonal cycle: by late summer, depth hoar is comparatively minor and MYI bulk density instead exceeds FYI, driven by the greater prevalence of dense melt-freeze layers near the base of the snowpack. Because the two datasets capture different stages of seasonal snowpack evolution, the contrasting patterns reflect a combination of seasonal and hemispheric

differences in snowpack processes rather than ice-age effects alone. We therefore highlight that snowpack evolution and bulk properties are strongly shaped by both regional (climate) regime and season, underscoring that density-stratigraphy relationships established for one hemisphere and season cannot be directly transferred to another without accounting for these differences."

Line 431: Do you also see the same thing for SSA?

As King et al do not report any SSA values, we cannot directly compare to their dataset. However, within our dataset we observe a related but slightly weaker pattern: snow-type-specific SSA values are generally similar between FYI and MYI, with slightly lower mean SSA values on MYI for most snow classes. This is consistent with more metamorphosed snow grains in snowpacks affected by repeated melt-refreeze cycling.

The revised sentence in line 419 in Section 5.1 now reads:

"In particular, **lower mean SSA and** higher snow bulk density on MYI is not caused by denser individual layers, but by the nearly three-times greater prevalence and two-times higher thickness of **low-SSA**, dense melt-freeze layers (for details see Table 3)"

Line 486: What if you want to retrieve bulk properties like SWE?

We agree that bulk applications such as snow load or first-order SWE estimates should be distinguished from stratigraphy-sensitive retrievals. We have revised the discussion accordingly. Our results suggest that bulk density may be more robustly upscaled than layer thickness, snow-type fractions, or SSA because it is the most spatially homogeneous property in our dataset. However, SWE also depends directly on snow depth, which remains spatially variable, and therefore robust SWE estimates still require spatially representative snow-depth information. Moreover, the comparatively low variability of bulk density does not imply that bulk snow representations are sufficient for retrievals that depend on vertical snow structure, such as microwave emission, microwave scattering, or snow-ice interface detection.

We revised the discussion from line 482 on:

"... snow-type composition and stratigraphy requires substantially higher spatial sampling density. **This distinction is also relevant for remote-sensing applications: for first-order estimates of snow load or snow water equivalent, bulk density may be more robustly upscaled than layer thickness, snow-type fractions, or SSA. However, the comparatively low variability of bulk density does not imply that bulk representations are sufficient for all retrieval applications. Microwave emission and scattering are sensitive to the vertical distribution of density, SSA, and melt-freeze layers, so retrieval approaches that depend on radiative transfer or snow-ice interface detection still require information on stratigraphic structure (Durand et al., 2011; Royer et al., 2017; Tonboe et al., 2022).** More fundamentally, representing Antarctic snow processes ..."

and in line 486:

"Consequently, Antarctic snow variability is governed by meter-scale structural heterogeneity and ice-regime dependence rather than by bulk properties alone. This implies that observational strategies, remote-sensing retrievals, and model parameterizations **that depend on internal stratigraphy, as opposed to bulk-integrated quantities such as snow load**, should move beyond uniform bulk representations toward approaches that explicitly account for ice-regime-dependent snow stratigraphy and its scale-dependent variability."

Line 507: Different parametrizations also exist...

We agree and have expanded the methodological limitations to acknowledge that different SMP parameterizations exist and may yield different absolute values. We selected the Calonne et al. (2020) approach because it provides both density and SSA and showed the best agreement with the co-located snow-pit stratigraphy in our dataset. We now state explicitly that alternative parameterizations, including Proksch et al. (2015) and King et al. (2020), were calibrated on different snow populations and could shift absolute density or SSA magnitudes. However, because the same retrieval framework is applied consistently across all profiles, we expect the relative contrasts between ice regimes and snow classes to be more robust than the absolute values. This is consistent with the methodological limitations already discussed in Sect. 5.3.

Revised text from line 509 on: **"Alternative parameterizations for retrieving density and/or SSA from SMP signals also exist (e.g., Proksch et al., 2015; King et al., 2020), calibrated on different snow populations, and these can yield systematically different absolute values. We selected the Calonne et al. (2020) approach based on its agreement with our co-located snow pit stratigraphy. Using an alternative parameterization would likely shift absolute density and SSA magnitudes, though relative contrasts between ice regimes and snow types, as well as spatial variability patterns, are expected to be less sensitive than the absolute values."**

Line 550: Yes but is there an impact of a bulk representation for remote sensing retrieval? This wasn't shown here.

We agree that our study does not directly quantify the impact of bulk versus stratigraphic snow representations on individual remote-sensing retrieval algorithms. We have therefore revised the conclusion to distinguish between findings demonstrated by our observations and implications that require dedicated forward-modeling studies, which go beyond the scope of this manuscript. The revised wording now states that uniform bulk representations are insufficient for applications that depend on internal stratigraphy, while acknowledging that direct retrieval impacts remain to be tested.

Together with the changes pointed out in the comment above on Table 3 and SSA values as a key variable for remote sensing, the revised conclusion from line 550 onward now reads:

"This consistency suggests that snow-type-specific SSA values may be transferable across ice regimes, offering a basis for parameterizing snow-type-dependent inputs to microwave emission and radiative transfer models, even as bulk snowpack properties diverge structurally between FYI and MYI. At the same time, this transferability applies at the level of snow types rather than snowpacks as a whole. This highlights that Antarctic snow cannot be adequately represented by uniform bulk properties or sparse point observations. Instead, snow on Antarctic sea ice should be treated as a stratigraphically structured, ice-regime-dependent medium, with implications for sampling strategies **and, pending dedicated forward-modeling studies for remote-sensing retrievals,** and the representation of snow processes in sea ice models."