Author Response for *"Impact of measured and simulated tundra snowpack properties on heat transfer"*, Victoria R. Dutch et al.

We would like to thank the editor and both reviewers for taking the time to read and comment on the original manuscript. As a consequence of these reviews, we have made three significant changes:
1. Text describing the derivation of the coefficients required to process the SMP data estimate snow density was shifted from Appendix A and is now prominent in the text of the manuscript (it was clear from the comments that Reviewer 3 did not read the Appendix).
2. The parameterization of Fourteau et al. (2021) has been added to the analysis as suggested by Reviewer 3.
3. We re-ordered the presentation and discussion so the sensitivity to different parameterizations of snow thermal conductivity is presented before the correction factor analysis.

We also provide responses with corresponding manuscript changes to all review comments, which are outlined below. For ease, reviewer comments are in black and our responses in blue. We respond to the comparatively minor revisions of Reviewer 4 before Reviewer 3.

Reviewer 4:
This is a solid study that makes use of new detailed high vertical resolution snow thermal conductivity measurements to assess the ability of the Community Land Model to represent the snow thermal conductivity. The authors find, somewhat unsurprisingly, that the lack of a process representation of depth hoar formation in CLM leads to much higher snowpack effective thermal conductivities than are observed at these sites. The paper includes a good discussion of paths forward, including introducing an adjustment factor and utilizing alternative snow thermal conductivity formulations. Overall, I think the authors did a thorough job responding to the comments of the prior reviewers. From my perspective this paper is suitable for publication with minor revisions suggested below.

1. I found typos throughout the revised manuscript. Many of them looked like they arose through accepting track changes wherein periods or other punctuation was lost. I suggest that the authors carefully review the manuscript to correct these typos.

Apologies for the typos in the revised manuscript, which resulted from some issues between the track changes and clean version. We have corrected these in the new version.

2. The discussion of the snow model in CLM5 is reasonable, but there is no mention of fresh snow density. The formulation of fresh snow density was modified substantially with a new temperature relationship and the addition of wind effects. In other assessment of CLM, we found that these changes resulted in higher and more realistic bulk snow densities (van Kampenhout et al., 2017, Lawrence et al. 2019), though the lack of widespread availability of snow density data makes these assessments more qualitative than quantitative. In any case, the fresh snow density parameterization should be introduced as well, I think, for the sake of completeness in the description of the snow model. I don't think the full set of equations need to be included, but a description would be helpful. The equations can be found here: https://escomp.github.io/ctsm-docs/versions/release-clm5.0/html/tech_note/Snow_Hydrology/CLM50_Tech_Note_Snow_Hydrology.html#ice-content

The following sentences have now been added to the end of the paragraph beginning on line 137:

"Changes implemented in CLM5 also include a new snow densification scheme, whereby fresh snow density is parameterised as a function of temperature and wind speed. The density of fresh snow can increase through the process of wind-driven compaction if wind speeds exceed 0.1 m^{-1} (Van Kampenhout et al., 2017)."

3. In a response to a reviewer, the authors note that they are exploring what could be done in terms of changes that would improve the model at the global scale. But, there is no mention of this in the text. As an interested reader, I was very curious about what approach could be taken at global scale, since this bias correction parameter would be difficult to define globally and the Sturm parameterization only really applies for tundra snowpacks. I think it would be helpful for a reader to elaborate just a bit more on the challenges here and to note explicitly that it is under investigation in an ongoing study.

We suggest the application of the Sturm parameterisation could be based on the dominant PFT for the model grid cell. The same logic could be used for the application of the correction factor α, and we had initially implied this. However, in the revised version, we no longer recommend this because the optimal values of α depend on snow depth and hence α changes at the same location between years (and would most probably also vary between locations).

4. If I am understanding correctly (I might not be, in which case the text needs to be modified to make this clearer), the same bias correction factor is applied to all the layers in the simulated snowpack. But, the main limitation of the model is that there is not representation of depth hoar that forms at the base of the snowpack. Would it be possible or advisable to apply a correction factor to just the deepest model snow layer to try to replicate this real world behavior?

Prior to applying the correction factor to all snow layers, we tested applying the correction to just snow layers 3 & 4 (bottom two layers - shown in blue on Fig. 5). However, because these layers are not initialised until later in the winter, the soil gets too cold at the start of the winter and doesn't really recover. This was particularly evident in 2017-18 due to the shallower snowpack. Because snowpack metamorphism is a continuous process and not only enacted at an arbitrary depth threshold, it is more appropriate to bias correct all layers starting from the initial snow accumulation.

One reason I ask is that in a study that was just published, we found that winter daily temperature variability is substantially improved in CESM2, where improved is towards less temperature variability. We inferred in that study that this reduction in variability can be largely attributed to the higher snow densities and relatedly higher thermal conductivities that acted to reduce air temperature variability due to the larger effective surface heat sink. The results of your present study imply that the thermal conductivities in CLM5 are too high, thus possibly countering the result of our study on temperature variability. However, if the uppermost snow layers are still actually relatively dense and it is just the deepest snow layer that has low thermal conductivity, it's possible that this result on temperature variability is still valid. Anyway, not sure that you need to address this comment much or at all in the manuscript, though it might be interesting as a point of reference to mention this paper (it's another argument for why snow thermal conductivity matters, in addition to the soil respiration that you already mention). The paper is Simpson et al., 2022, JAMES, https://doi.org/10.1029/2021MS002880

Thanks! This is good to know. Because we did not run the model in coupled mode, we agree that this need not be mentioned in the manuscript. However, we are taking this work further into

implications for Net Ecosystem Exchange so knowledge of this new paper will be valuable for this future work.

5. The Figure 5 figure caption needs more info about what the colors mean. I was able to infer that they indicated the snow layers (nice way to look at the evolution of the snowpack layers in CLM!), but I guess some readers may not be able to recognize that immediately.

Added the following to the caption. It now reads:

"Figure 5: Simulated snow layers and their properties for winter 2017-18; a) simulated snow layer thicknesses, b) snow layer densities, c) snow layer thermal conductivities; d) to f) as before but for winter 2018-19. Each colour represents a different computational snow layer."

Reviewer 3:

General Comments

These general comments were written after the subsequent specific comments, I apologize for the few repetitions.

This paper in fact reports two somewhat independent topics. The first one is the determination of vertical thermal conductivity profiles of snow near TVC using density profiles determined from SMP profiling. Density is derived from SMP data using an algorithm developed from snow on sea ice, if my understanding is correct.

We draw your attention to Appendix A, which clearly states new coefficients specific to tundra snow have been derived by applying the King et al (2020) methodology. The sentence pointing the reader to the appendix has been reordered slightly to make this more explicitly obvious, it now reads: "Bespoke coefficients for tundra snowpacks were calculated based on the methodology of King et al. (2020) to derive high vertical resolution snow density profiles from the SMP force profiles (see Appendix A for detailed methodology)." However, the reviewer may be interested to know that the values from snow on sea ice (King et al., 2020) and those we derived from snowpit density profiles at TVC were relatively similar (both are given in Table A1, alongside the original coefficients of Proksch (2015) derived using an earlier version of the SMP).

The second one is to compare these thermal conductivity profiles to those simulated by CLM. As expected from any snow model used in the Arctic, CLM does not perform well. It simulates much too cold temperatures and the authors discuss various adjustements to make the model better fit the data.

The experimental part is interesting and there is some novelty to it. There are however a number of weaknesses. First, the thermal conductivity is derived from density using various equations that do not agree with each other. Second, while I am willing to believe that the algorithm used to convert SMP data to density is good, it seems to have been validated for snow on sea ice, which is somewhat different from snow at TVC where vegetation impacts snow properties.

As described in Appendix A, the estimation of snow density from the SMP measurements was fully developed using only the tundra snow measurements. Coefficients were derived from Jan 2019 and March 2019 TVC field campaigns. We have now made these methods more prominent in the main body of the text by shifting some of the text from the appendix:

"Bespoke coefficients for tundra snowpacks were calculated based on the methodology of King et al. (2020) to derive high vertical resolution snow density profiles from the SMP force profiles (see Appendix A for detailed methodology). Briefly, a K-folds recalibration was used to derive new coefficients (Table A1) from 36 co-located snowpits and SMP profiles across the TVC catchment. These coefficients were then applied to all 1050 SMP force profiles from the 3 campaigns over a 2.5mm rolling window to give recalibrated density profiles. These density profiles were then used to approximate profiles of thermal conductivity using the K_{eff} relationships derived by Sturm et al. (1997), Calonne et al. (2011), Jordan (1991), and Fourteau et al. (2021), denoted $K_{eff-Sturm}$, $K_{eff-Calonne}$, $K_{eff-Jordan}$ and $K_{eff-Fourteau}$ respectively."

Third, in the discussion, the authors mention an artefact where the penetrating SMP rod may have damaged the fragile depth hoar, causing an artefact with lower intermediate density and increased

lower density. This third issue is very important and must be mentioned in methods so that the reader has it in mind while evaluating the experimental part.

This is mentioned in detail in Appendix A as well as in the discussion. We now repeat some of this in the main body of the methods, as follows:
"Sources of uncertainty in the SMP measurements include interactions with vegetation within the snowpack and collapse of the depth hoar layer during measurement; an experienced SMP user can easily identify and remove profiles which are affected by these issues. A positive bias in derived depth hoar density occurs because of large distances between snow grain failures (see Appendix A and King et al., 2020b for more details)."

I detail this point below. So, while SMP clearly allows obtaining much more data with high vertical resolution, its limitations and potential artefacts must be stressed, with the possible conclusion that it is not (very) suitable for snowpacks with very fragile basal layers. Or that the wind slab must be removed to measure the depth hoar? In any case, I feel the SMP data deserves publication if the caveats are clearly stated.

We appreciate that uncertainties may be introduced to SMP measurements by collapse of the depth hoar layer. As mentioned above, we note this source of uncertainty in the text. With respect to mitigating this uncertainty, profiles with obvious depth hoar collapse or depression in the wind slab during measurement were removed. Experiments with removing the wind slab layer would likely result in disturbance to the remaining depth hoar, so it is not obvious how this approach would help. In addition, with the comments of this review in mind, stratigraphy and density profiles with 100 cm^3 box cutters made at TVC in late March 2022 were closely examined and the measurements showed an increase in density close to the snow-ground interface. This section of depth hoar was much more consolidated, most likely due to more moisture being available in the upper soil/veg layers at the onset of snow accumulation. In short, we now have more confidence in the increase in density close to the snow-ground interface shown by SMP profiles than when we wrote this manuscript and so the caveats do not need to be further affirmed in addition to what is already presented in this manuscript.

Another issue with the derivation of thermal conductivity (TC) profiles from SMP data is the choice of the density-TC relationship. First, the authors must remain aware that this is far from a bijective relationship and that there is enormous scatter in this relationship. For example, in (Sturm et al., 1997) there is sometimes a factor of 5 variation in TC for a given density value. The use of such an equation may therefore bear some implicit very high uncertainty and error. Second, there are large differences between the various relationships used, so which one is correct, if there is one, and what is the error induced by these relationships?

In our view, none of the density-TC relationships are correct because they all presume snow thermal conductivity to purely be a function of density, whereas we know other factors such as grain shape and temperature are also important as we acknowledge in the text. We devised an approximate snow thermal conductivity from the SMP using the density relationships to give a first-order approximation of the snowpack thermal conductivity, which is consistent with many other studies. Showing multiple relationships and the interquartile range of the values produced by each relationship gives an idea of the uncertainty across these approximations.

I would like to comment on the various relationships used. The equation of (Jordan, 1991) is based on that of (Yen, 1981) which is in fact a compilation of previous works using a variety of methods. (Sturm et al., 1997) use only the needle probe method, but there are doubts on its reliability (Riche and Schneebeli, 2013) and this is not just due to anisotropy issue as stated by the authors. (Fourteau

et al., 2022) have very recently made a detailed theoretical and experimental study of the needle probe method and demonstrated that the aproximate equation used by (Sturm et al., 1997) to derive thermal conductivity from heated needle probe data is not correct, leading to a negative artifact. Briefly, it is not valid for the short heating times used as some neglected terms in the full equations are in fact not negligible. (Fourteau et al., 2022) proposed an algoritm to correct existing needle probe data. In any case, I think the equation of (Sturm et al., 1997) is incorrect, and the fact that it produces values lower that the other parameterizations confirms this negative artefact. I very strongly recommend to stop using this equation. The equation of (Calonne et al., 2011) is based calculations from tomographic images. It is rigorous, however, it minimizes water vapor effects because is postulates slow surface kinetics. In other words the accommodation coefficient of water vapor on ice takes here a low value, so that latent heat fluxes are minimized and probably underestimated. More recently (Fourteau et al., 2021) used a similar approach but postulated fast surface kinetics, leading to slightly higher values than (Calonne et al., 2011). In conclusion, the equation of (Sturm et al., 1997) is incorrect and should just not be used. The equation of (Jordan, 1991) is based on ancient measurements using a variety of poorly documented techniques and I feel it has limited reliability. The equation of (Calonne et al., 2011) is arguably correct but makes the debatable postulate that surface kinetics is slow, while (Fourteau et al., 2021) make the more reasonable (in my opinion, but I am a coauthor of that paper) postulate of fast surface kinetics. My biased recommendation is therefore to use the equation of (Fourteau et al., 2021) and in any case not to use the equation of (Sturm et al., 1997). I recommend not to use different equations as an error compensation trick in CLM.

We thank the reviewer for their thoughts on the relative merits of each relationship. We have added the relationship from Fourteau et al. (2021), which was not yet published when we were preparing the original manuscript. We make no judgement on the accuracy of any particular functions for this environment, rather we include them to quantify the spread between the relationships and not as any form of error compensation. It is interesting to note there are only small differences in the results produced from the Calonne, Jordan, and Fourteau parameterisations, both when applied to the SMP measurements and when used within CLM. Consequently, sensitivity to different model parameterizations is now considered in the manuscript first, with the correction factor analysis being a subsequent argument.

In fact, I am not too thrilled by the modeling part. Essentially this just says that a simple snow routine from a more complex model just does not work in the Arctic. It is interesting to confirm that the CLM snow scheme is deficient, but I do not think all the error compensation tricks used by the authors, and the lengthy discussion on parameter adjustments, have much interest. Figure 7 shows that for the first year, α =0.4 must be used while for the second year α=0.6 is better. So, the adjustment parameter changes from year to year, and is probably different for other sites, so that CLM has no predictive value when it comes to Arctic snow and ground temperature. We already know, I as well as all the authors, that even the currently most sophisticated snow model do not work in the Arctic, so no one expects the simpler CLM scheme to perform any better. The authors just need to show Figure 7, demonstrating the lack of predictive value of CLM and therefore its much reduced interest for predicting snow properties and soil temperature in the Arctic. They then could just discuss that implementing the missing process, upward vapor transfer, is too complex, and that other approaches they wish to propose must be envisaged. I Therefore think Figures 1, 2, 3, and 7 are interesting. I strongly recommend that the authors consider removing the other Figures, or at least most of them, and reduce the modeling text by at least 50%, probably more. Please also consider my comment of Figure 5 below.

Given the widespread usage of CLM, particularly for high-profile climate projections, this finding is an important one and it is critical to bring this to the attention of people outside of the snow

microphysics community. This viewpoint is echoed by Reviewer #4 who has significant experience in land surface modelling <u>and is a developer of CLM</u>. We have chosen this journal as a mechanism to bring these problems to the attention of the wider modelling community, particularly those who might be able to fix it. We were indeed not surprised that CLM performs poorly in its representation of the vertical properties of Arctic snowpacks. However, the implications of snowpack simulations on soil temperatures have broader consequences in earth system models. For example, underestimation of soil temperatures impacts simulations of soil respiration with implications for pan Arctic carbon fluxes (e.g. Natali et al. (2019)).

In addition, both reviewers from the first round of the review process also had positive things to say, with reviewer #2 describing it as "very relevant" and reviewer #1 saying that it is "only using such work presented in this paper that we could be able to better understand the soil-snow-atmosphere feedback in the Arctic." So, while reviewer #3 may not be thrilled by this modelling analysis, we are hopeful that the detailed analysis and implications will be more appreciated by the wider community.

To help better engage that community, in the revised version of the manuscript, the modelling text has been condensed and figure 8 has been removed. We have re-emphasised the variability in the ideal value of the correction factor, and use this as a tool to discuss the limitations of the model rather than advocate for the application of it as a bias correction. We also discuss the application of different parameterisations of snow thermal conductivity to show the sensitivity of snow:soil relationships in CLM.

Figures 4 and 6 show that snow thermal conductivity can be approximated from the SMP at a very high vertical resolution, which we feel is novel and exciting. Figure 5 aides in the understanding of the numerical manner in which CLM builds a snowpack and the deficiencies that result from this. Reviewer #4 also thought this was an effective way to visualise the development of the snowpack.

Another general comment is that there is a serious lack of attention to detail in the writing and presentation. This is surprising given that "All authors were involved in reviewing and editing prior to submission" and that the authors include a large number of high-profile esteemed and highly respected senior researchers. For example (just one, and I will not edit for typos, the authors can do it), all of these authors think it is fine to write "Snow has a low thermal conductivity, typically in the range $0.01 - 0.7$ Wm_{-2} K_{-1}". So snow can have a thermal conductivity less than half that of air? And for brevity I will refrain from commenting the 0.7 value. In any case, I will do my best to write a hopefully constructive review. Let the authors do their best to write a paper with attention to detail.

For brevity, we will comment that the range of thermal conductivities of snow was taken from Gouttevin et al. (2018): "Snow has a low thermal conductivity (K_{eff}), ranging from 0.01 to $0.7 Wm^{-1}$ K^{-1} depending on microstructure, density, and wetness, and it therefore insulates the underlying ground during the cold season." The Cryosphere, 12, 3693–3717, 2018.

Finally, it was a collective oversight of the entire author team to fail to spot some typos resulting from the acceptance of track changes during revision. However, the patronising and dismissive language accompanying this admonishment is unnecessary and undermines the scientific training of the Early Career Researcher entrusted with leading this manuscript. The open review process of The Cryosphere correctly allows for full transparency within our community. Respectful treatment of ECRs is a reasonable expectation.

Specific comments

Line 44. I am not sure what the authors mean by "indurated depth hoar". They seem to have a definition different from mine (Domine et al., 2016b) and from Sturm's (Sturm et al., 2008). Sturm and I have the same definition, since Sturm introduced me to indurated depth hoar on the Alaska north slope in 2004. I would think the authors would have a similar definition since Chris Derksen has done much Arctic snow field work with Sturm. However their stratigraphy does not show any indurated depth hoar, but faceted crystals. This is very confusing. I would also think the lower layer, with densities reaching 300 kg m-3, would often be actual indurated depth hoar, possibly formed from melt-freeze layers. In any case, I suggest the authors realize that the classification of (Fierz et al., 2009) was made by avalanche experts for avalanche motivations. It is almost exclusively based on observations in Alpine snow and is largely inadapted to Arctic snow. I discussed this with Charles Fierz and he had never seen indurated depth hoar. I think Arctic snow researchers should use symbols adapted to their problem. I have proposed symbols for indurated depth hoar and indurated faceted crystals (stage prior to indurated depth hoar) in (Domine et al., 2016b) and (Domine et al., 2018) (already cited by the authors) and in other papers. Why not popularize these symbols, adapted to Arctic snow, which by the way is much more important area-wise than alpine snow ? It would spare us these inevitable inconsistencies between text and Figures.

We use indurated hoar as identified as "wind slab to depth hoar" layer as per Sturm et al. 2008 and as described in Domine et al. 2016. To avoid unwanted confusion we have now cited Sturm et al. (2008) in our introduction:
"Between these two layers, an indurated hoar layer may also be formed (Sturm et al., 2008), where the lower part of the wind slab takes on some of the microstructural properties of depth hoar (e.g faceted grains) while maintaining the density and hardness of a wind slab (Derksen et al., 2009)."

Symbols for indurated depth hoar do not exist in the international classification of Fierz et al. (2009). We have used a symbol for facets crystals as the slab layer was turning towards depth hoar. However, if they can be made available by the reviewer in a publicly available font file (.ttf), we will add them to Figure 2 which shows symbology in an indicative manner to help the reader interpret stratigraphy. The alternative is to remove the symbol for faceted crystals from Figure 2.

Line 150. What is hsl?

h_{sl} is the height of the snow layer. This is given in the explanation sentence below eq. 1 (what was L154).

Line 185-190 are not necessary. These are very well-known considerations. In general section 3.1 can be greatly condensed.

Section 3.1 has been condensed.

Figure 2: Faceted crystals? Columnar DH? Please clarify symbols and make them consistent with earlier parts of paper.

Figure 2 shows grain type symbology (from Fierz et al. 2009) alongside changes in density, in an indicative manner to help the reader interpret stratigraphy. If a further symbol depicting indurated depth hoar (following the slab to hoar interpretation of Sturm et al. 2008) is desired, we will apply it if the reviewer can provide it in a publicly available font file (.ttf) – see comment above regarding either removal of the facets symbol or addition of a new symbol.

Plus, again, Fierz 2009 inadapted to Arctic snow. How about showing fall 2018 data? Fall data are in general scarce amnd therefore valuable.

Unfortunately, due to instrument malfunction SMP data are not available for the November 2018 field campaign. Pit data from November 2018 are shown in Figure 3.

Table 2 lacks detail and does not correspond with stratigraphy of Figure 2. There seems to be a dense basal layer, perhaps indurated depth hoar, in the lower 10-20%. Then the next 20-60% seem to have a homogeneous low density typical of columnar depth hoar. Therefore, these 2 layers should be separated in Table 2. Just having one DH layer is not consistent with data.

We feel there is a balance to strike here. An exciting reason to use SMP measurements is that traditional binary thresholds for wind slab / depth hoar or wind slab / indurated depth hoar / depth hoar do not capture the transitional nature of snowpack properties in a vertical profile. This is arguably a significant advance on traditional snowpit stratigraphy. However, pragmatic classification of snow layers will always have a role in model evaluation, hence we feel that the three-part (surface snow, wind slab, depth hoar) classification in Table 2 provides useful information to readers, especially those to whom land surface modelling is a focus.

Line 214. Please descrive indurated DH. Written as faceted crystals in Figure 2.

See previous responses to comments on Figure 2 which directly address this issue.

Line 221. Ice lenses not shown in stratigraphy.

Ice lenses were only observed in a small number of pits, whereas figure 2 shows an amalgamation (median and interquartile range) of ~400 SMP profiles. Adding ice lens symbols to the indicative symbology of Figure 2 would not aid in conveying the properties of the total distribution of density measurements.

Line 235. Please check grammar
Done.

Line 257. Please be consistent with snow layers. Figure 2c and d mention 5 layers. Here just 3.
Our response here follows on from the comment on Table 2. The transitional element of SMP derived density profiles rather than traditional binary layer identification or computationally modelled layers, means that measurements made at the sub-millimetre scale may not directly translate. Simplification of layers in the way the arguments are presented, especially when discussing modelled layering, means this perceived discrepancy in consistency of the number of snow layer consistency is valid.

Line 265. 0.344 is 4 times as large as 0.08, not 3.

The original sentence was to state that snow thermal conductivities from CLM (0.344) are ~3 times larger than those from the needleprobe (0.08) *or the SMP (0.11)*. This has now been changed to "at least 3 times" for clarity.

Sturm is not an approximation, but an equation, or a parameterization.

The word approximation has been changed to parameterization.

Line 268. Most of Sturm's measurements were in fact in the Boreal forest of interior Alaska, where the mostly depth hoar snowpack layers have very low thermal conductivities. Many measurements were also from tundra snow but what the authors say is incorrect.

Changed to refer to "the Alaskan Arctic".

Line 281. Probably unnecessary explanation.

Removed.

Figure 5. The graphs may be easier to visualize if the grey and black colors were swapped.

Contrast between the colours of the different layers alternates in order to make the individual layers easier to distinguish, particularly for readers who are colour blind. Putting black and navy blue layers next to each other would make the figure harder to read.

In any case, I am not sure about the utility of panels b-d and in fact I am surprised by how CLM seems to work. The origin seems to be the top rather than the base of the snowpack, even though the snowpack forms from the bottom up, as the authors know. Following density from the top then does not really correspond to anything physical, as a given snow layer is not monitored. I do not have time to get into the methods and architecture of CLM, which is totally unknown to me, but it seems very strange. Since it does not seem to take into account actual processes, it seems that having such a model fit data will result in mandatory adjustments on a case by case basis, with no hope of ever having any predictive value. From what little I understand, I therefore wonder whether there is actually any hope of ever getting any reliable snow simulations from CLM. I'll be more than happy to be proven wrong.

The process by which CLM builds a snowpack is entirely computational and does not reflect the physical processes which create an Arctic snowpack. However, this does not mean that any adjustment can only be made on a case-by-case basis or that the model has "no hope of ever having any predictive value". Improvements to the properties of this simplified snowpack structure can still provide an improvement to the simulation of snow insulation and soil temperatures for example, such as when reducing the simulated thermal conductivities. Changes to this simple snow model, such as through the substitution of snow thermal conductivity parameterisation provide further improvement to the simulated subnivean conditions.

Line 293. Why just mention 2018-2019? Is not it interesting to realize that in 2017-2018, $\alpha = 0.4$ works best?

The following sentence has been added:
"However, a smaller value of α was required for best model performance in 2017-18, with an α of 0.4 giving the lowest RMSE of 1.6°C and highest SHTM of 0.986"

A single value cannot simulate both years. And therefore, expectedly, different thermal conductivity parameterizations have to be used for each year. This may be stressed.

The ideal value of the correction factor changes between years, and more emphasis has now been placed on this in the text.

Line 296. I do not understand (0.3 ≥3≥ α ≥ 0.5555)

Thanks for flagging this. This is a typo as a result of merging different sets of tracked changes and has now been fixed to read (0.3 ≥ α ≥ 0.55).

Line 316. Sturm's parameterisation works best, but that parameterization is wrong. This is just an error compensation game. The model is wrong, and this is compensated by a wrong thermal conductivity parameterization.

All models are imperfect. The value of these simulations comes from what they tell us about the dynamic relationship between snow and soil, and how this is influenced by both the application of alternative snow thermal conductivity parameterisations and the use of the correction factor α. All parameterisations are by necessity a simplification, and as you previously mentioned, none of them are perfect. We feel using the largest reduction in RMSE, common in snow modelling evaluation, to guide parameterization choice is a valid approach.

And the compensation is different for each year, meaning that the model has no predictive value.

The agreement between simulated and observed soil temperatures changes within and between winters. This is the case regardless of the thermal conductivity parameterisation used. However, we dispute the assertion that this means the model has "no predictive value" because it does provide insight into the potential for bias correction. The ideal value of α does change between years, and more attention is now brought to this in the text. As detailed in figure 7c and Table 1, CLM considerably underestimates the depth of the snowpack in 2017 – 18. This is most likely due to uncertainties in observed precipitation values used to force the model (as mentioned in the methods), and we believe this error in snowpack depth is the most likely reason that the model gives better soil temperature simulations (and requires a smaller correction) in 2018 – 19. Consequently, in the overall narrative of manuscript we changed the sequence of the results to first show the sensitivity of CLM to the various thermal conductivity parameterizations, and then applied the correction factor as an illustration of bias correction.

Lines 326-328. This comparison is very misleading. The authors compare their indirect estimation of thermal conductivity based on an indirect estimation of density to actual measurements of snow thermal conductivity. Furthermore, the measurements of (Domine et al., 2015; Domine et al., 2016b) and of (Morin et al., 2010) are continuous season-long time series, so that the focus of those papers are on time-variations, while the authors'work is on height variations.

We have no intent to mislead. Rather, as stated in the text, it is a novel approach to use very high vertical resolution estimates of density from SMP measurements to try and understand changes in the snowpack thermal conductivity. From our perspective, this leverages and promotes knowledge from the classical thermal conductivity studies we cite, a larger mistake would be to ignore this and not cite it.

By the way, vertical profiles of thermal conductivity measured by (Gouttevin et al., 2018) and by (Domine et al., 2012; Domine et al., 2016a) is closer to 5 cm resolution than to 10 cm.

Changed to read "~ 5 – 10 cm".

Lines 330-344. This discussion is interesting but I think it should already be stated in the methods section. I have been wondering about this high density basal layer, thinking it may reflect rain-on-snow in the fall but only now do I realise it is probably just an artifact! I have been misled in my understanding of the data all along! So please shift this up. And by the way, why did not the authors

perform SMP measurements with the wind slab remove to test for the actual impact of this artefact, since they must have been aware of it while making the measurements.

This was stated in the appendix as well as being re-iterated in the discussion. The following text has also been added to the methods section:
"Sources of uncertainty in the SMP measurements include interactions with vegetation within the snowpack and collapse of the depth hoar layer during measurement; an experienced SMP user can easily identify and remove profiles which are affected by these issues. A positive bias in derived depth hoar density occurs because of large distances between snow grain failures (see Appendix A and King et al., 2020b for more details)."

Please also see our response above to a similar comment where very recent measurements at TVC in late March 2022 suggest this higher density depth hoar near the soil-snow interface is not an artefact.

Lines 345-352. Continuous vertical profiling is indeed very nice and is clearly a significant improvement over discrete layer sampling. However, is SMP suited to Arctic snowpacks, with a basal depth hoar layer that can be extremely fragile and collapse at the slightest touch? (see details in (Domine et al., 2016b). The authors seem aware of this problem, and this clearly limits the interest of SMP for some Arctic snowpacks. Not all, I agree, since very windy areas such as Barrow and polar deserts seldom have very fragile depth hoar. This should also be discussed. Furthermore, SMP is blind sampling, since a snowpit is not dug in most cases as this would cancel the benefit of the technique. Therefore, artefacts due to soft layers would be undetected. In conclusion, while I do see the benefit of SMP in some cases, the authors may wonder whether a low resolution reliable manual density profile is better than a high resolution profile with potential and unverifiable artefacts.

We have never advocated for not making manual snowpit measurements. In fact, some manual density profiles from snow pits are required to derive coefficients to convert SMP force measurements to density. Snowpits are also used to give an idea of what features might be likely within the SMP profiles, and which are more likely to be artefacts. Our sampling protocol required taking SMP profiles at pit locations as well as across the landscape, with the nearest neighbour SMP to each pit used to derive the Force: Density relationship. The necessity of both snowpit and SMP measurements to derive recalibration coefficients is detailed in Appendix A and is now re-iterated within the methods section as follows:
"Use of the SMP allows for a large increase in both the number of sites and the vertical resolution at each site compared to traditional snowpits, but some coincident snowpit measurements are still required to derive the coefficients to estimate snow density."

Line 355. Alpine snow does have a density profile as simulated by CLM, but not the taiga snow, which by the way is not discussed by the references cited. In the boreal forest (for some reason, I refrain from using Russian words these days…?) the profile may be as in CLM at the beginning of the season but by the end of winter it is either flat or with lower basal density because of the upward vapor flux. See e.g. (Taillandier et al., 2006).

The reference to taiga snow has been removed.

Line 361-362. "such as the snowpack vapour kinetics necessary to form depth hoar." The term "vapor kinetics" is unclear. Use flux, and more accurately upward flux.

The word "kinetics" has been replaced with "flux".

References:

Derksen, C., Sturm, M., Holmgren, J., Liston, G. E., Huntington, H., Silis, A., and Solie, D.: Northwest Territories and Nunavut Snow Characteristics from a Subarctic Traverse: Implications for Passive Microwave Remote Sensing, Journal of Hydrometeorology, 10, 448-463, 10.1175/2008jhm1074.1, 2009.

Fourteau, K., Domine, F., and Hagenmuller, P.: Impact of water vapor diffusion and latent heat on the effective thermal conductivity of snow, The Cryosphere, 15, 2739-2755, 10.5194/tc-15-2739-2021, 2021.

Gouttevin, I., Langer, M., Löwe, H., Boike, J., Proksch, M., and Schneebeli, M.: Observation and modelling of snow at a polygonal tundra permafrost site: spatial variability and thermal implications, The Cryosphere, 12, 3693-3717, 10.5194/tc-12-3693-2018, 2018.

Natali, S. M., Watts, J. D., Rogers, B. M., Potter, S., Ludwig, S. M., Selbmann, A.-K., Sullivan, P. F., Abbott, B. W., Arndt, K. A., Birch, L., Björkman, M. P., Bloom, A. A., Celis, G., Christensen, T. R., Christiansen, C. T., Commane, R., Cooper, E. J., Crill, P., Czimczik, C., Davydov, S., Du, J., Egan, J. E., Elberling, B., Euskirchen, E. S., Friborg, T., Genet, H., Göckede, M., Goodrich, J. P., Grogan, P., Helbig, M., Jafarov, E. E., Jastrow, J. D., Kalhori, A. A. M., Kim, Y., Kimball, J. S., Kutzbach, L., Lara, M. J., Larsen, K. S., Lee, B.-Y., Liu, Z., Loranty, M. M., Lund, M., Lupascu, M., Madani, N., Malhotra, A., Matamala, R., McFarland, J., McGuire, A. D., Michelsen, A., Minions, C., Oechel, W. C., Olefeldt, D., Parmentier, F.-J. W., Pirk, N., Poulter, B., Quinton, W., Rezanezhad, F., Risk, D., Sachs, T., Schaefer, K., Schmidt, N. M., Schuur, E. A. G., Semenchuk, P. R., Shaver, G., Sonnentag, O., Starr, G., Treat, C. C., Waldrop, M. P., Wang, Y., Welker, J., Wille, C., Xu, X., Zhang, Z., Zhuang, Q., and Zona, D.: Large loss of CO_2 in winter observed across the northern permafrost region, Nature Climate Change, 9, 852-857, 10.1038/s41558-019-0592-8, 2019.

Sturm, M., Derksen, C., Liston, G., Silis, A., Solie, D., Holmgren, J., and Huntington, H.: A reconnaissance snow survey across northwest territories and Nunavut, Canada, April 2007, Cold Regions Research and Engineering laboratory, Hanover, New Hampshire, 1–80, 2008.