With the main focus on the spatial heterogeneity of snow, I'm irritated by the fact that the authors do not offer any spatially explicit validation
A figure of measured vs. modelled snow depth along the transect with the longest persisting snow cover was added.

Just to clarify, spatially explicit representation was not offered in the previous version of the manuscript because of differences in scale between point measurements and model gridboxes. As is now more explicitly explained in the last paragraph in Section 4.3. and further emphasized with an added section on the effect of grid box resolution on process representation, issues of scale occur when comparing single point measurements with a snow probe against model output from a 8 m resolution gridbox and, thus, will inevitably lead to potentially large errors. Other scale-related potential sources of error were described in Section 4.3 (4.2 in previous version). Spatially averaged comparisons were therefore provided, on the basis that the broad measured and modelled snow patterns would emerge more clearly.

Figure 7 would allow spatial validation, so would the snow data measured along the transects.
Pictures of each slope taken from the opposite slope, were taken daily in view to georeference them and use them to evaluate the spatial evolution of the modelled snow distribution. A single camera, which was screwed daily to a tripod installed, but not fixed, on the ground, was used to take the pictures. However, while initial processing of the images showed promising results and were presented at some conferences, further processing showed that even small changes in the position of the camera, which were occurring because neither the tripod nor the camera were fixed, the poor resolution of the handheld GPS used to define the coordinates of the ground control points and the barrel and radial distortion of a 6mm focal length caused errors in georeferencing which were potentially larger than model errors. As a consequence, the photos were not used.

In addition, to give one example of problems I see related to the transferability of the model: Given the structure of 3SOM, the calculation of snow cover fraction (Equation 7) must be considered a key component of the model. The function was calibrated using local survey data. However, this calibration is representative of the existing shrub distribution and terrain. How can this parameterization be transferable to scenarios like no-terrain or a vegetation fraction projected to be 6.5 times higher than the actual fraction?"
This is an excellent point. The snow cover fraction was modified and the calibrated parameter replaced by the pre-melt standard deviation of snow depth following Essery and Pomeroy (2004). The changes incurred by this new parametrization on the snow depletion curve can clearly be seen when comparing Figure 11 in the new version of the manuscript and the Figure 9 in the previous version.

3) Figure 6 presents data that allow identification of severe mismatches between modeled and observed CV data. Modeled CV data for south facing slopes seem to be underestimated by a factor of up to three, whereas CV data for north facing slopes seem to be about right.
We conducted a sensitivity analysis on standard deviation as a function of gridbox resolution, which led to the addition of a new Section (4.1). We found that modelled standard deviation of snow depth is highly dependent on gridbox resolution and that it obeyed a power function. The modelled standard deviation with 8 m resolution gridboxes was extrapolated to a 4 m resolution gridbox using the power function; considering issues of scale considerably reduced errors between modelled and measured standard deviation. (Also both are closely related, please note that Figure 6 (now 7) shows standard deviation, not CV).

Maybe the model is not yet up for a sensitivity study to challenge previous studies about the feedback of tundra shrub expansion on land surface albedo?
Please note that the reference height for wind speed was modified (snowdepth was not previously substracted).This was not described in the previous version of the manuscript but as this was

found to improve modelled turbulent fluxes, the description of the resistances was added to the revised manuscript. We believe that this, added to the findings described in Section 4.1 and to the change in the snow cover parametrization have consolidated the results and, by extension, the points addressed in the Discussion and Conclusions section. The manuscript stays cautious about "challenging" previous studies but instead points out that the proposed high resolution study identified that processes which are not represented in large scale studies may affect model results. We do not refute previous findings but instead propose areas in large scale modelling in need of future research. We also acknowledge that the "study was conducted at a single location and [that] further studies are required to confirm the relevance of these findings in other sub-arctic and arctic environments".

a) upgrade 2SOM / 3SOM to include heat transfer between grid cells. If this is impossible the authors should consider to run DBSM at 8 m resolution but then decrease the grid resolution when switching to 3SOM.
Thank you, this was a very valuable suggestion which led to the section of the sensitivity of the standard deviation of snow depth to gridbox resolution. As is now explained in Section 4.1, we found that the standard deviation of snow is dependent upon gridbox resolution. As a result, a short experiment investigating the impact of gridbox size on turbulent fluxes found that larger gridbox sizes do not solve the between-cells advection; as the snow is more homogeneous with increasing gridbox size, latent heat increases but sensible heat decreases (Section 4.2).

b) provide explicit evidence that the models can replicate the evolution of spatial snow patterns as observed;
See answers paragraphs 4 and 5 above.

c) allow more space for the introduction of 3SOM.
The appendix was moved to the main text and the description of the model was expanded.

What would Figure 3 a and b look like if modeled with 2SOM, or 3SOM at various grid resolutions?
Figure 3 a and b show results from the model being run at a single point and initialized with field measurements in order to allow a direct comparison with the 2-source model, which is a single point model. This was made clearer in the text.

Would any improvements relative to 2SOM show up in Figure 6 and 7?
No, the 2-source model is a single point model.

d) reduce the weight currently put on the model sensitivity exercise
More weight was added to the model description and evaluation.

and discuss the findings more cautiously.
See answer above to "Maybe the model is not yet up for a sensitivity study to challenge …"

Given Figure 6 it may not be justifiable to dedicate more than 50 % of the abstract to findings of the modeled scenarios.
Although Figure 6 has changed and shows improved results compared to the previous version of the manuscript, modeled scenarios now cover less than 50% of the abstract.

SPECIFIC COMMENTS

P232/L11: If the three surface sources share a single soil column, does this mean that the surface temperature of snow-free patches cannot exceed 0 _C if Fs > 0? I'm probably misunderstanding something here, otherwise this approach would severely compromise

the benefit of having a separate energy balance equation for bare ground. This needs further context.
No, each source has a separate temperature; Tg is the ground (snow-free) source surface temperature (Section 3.1 after Equation 13). A sentence was added in the penultimate paragraph of Section 3.1 to clarify the relationship between the soil temperature and the temperatures of the snow and ground sources.

P235/L16-26: I suggest to move the content to the previous page (P234).
Done.

P237/L2-3: There is no evidence that the models can replicate the evolution of spatial snow patterns.
The addition of Section 4.1, Figure 6 and Figure 8, the discussion about scale effects and potential sources that may affect them, and a clarification of what we define as “the spatial snow pattern” should now provide sufficient evidence that the model can replicate them.

P237/L20: I thought the domain was 1 km2, so what’s outside of the central domain?
Nothing concerning this study. The sentence was rewritten.

P243/L7: How are Fs and Fg factored in?
There was a mistake in the manuscript (but not in the code). Equation corrected.

P254: Why do the transects not extend into zones with variable shrub density?
They do, although it is not clear from Figure 1. Some details were added in the last paragraph of Section 2.

P255: Somewhere in the paper it should be mentioned that Fs+Fg = 1, and Fv is independent of either (if this is the case).
Fg is decribed as (1-Fs). It is now stated that Fv is independent of either.

P258: Why is there no measured data above _700 W/m2?
Explanations are now available in the caption of Figure 5.