

## RESPONSE TO REVIEWER #1

[...] more evidences should be presented and discussed to assess whether wind redistribution and snowfalls are negligible during spring, when melting is the dominant process, but not the only one. This could help also to explain some models underestimations in SWE in Fig. 3(a) at the end of the melting season, which could be also due to these neglected processes, and not only to cold temperatures. This is now discussed in Section 3.4.

In particular, it could be interesting to discuss in a deeper way the implications of neglecting the full coupling between the accumulation and the melting periods, and to assess quantitatively if neglected mass fluxes are relevant or not (e.g., which is the average air temperature during winter? Is there evidence of snow events in spring? etc.).

More information on the winter and springtime conditions was added in Section 2 and possible effects on results are discussed in Section 3.4.

### Minor comments:

**1. Abstract:** I would suggest reorganizing the second part (the one dealing with the results of the study) since it seems to be a bit dispersive;

The abstract was re-written.

**2. Line 9, page 225:** I think it could help adding some quantitative references to be compared with  $3 \text{ Wm}^2$  per decade, since at this stage it is difficult to understand how much relevant this rate is, if it is reported all alone;

Done.

**3. Line 20, page 225:** please define “snow-holding capacity” since this process seems to be quite relevant in this paper;

The sentence was re-written for clarity.

**4. Sites and Data (Section 2):** maybe consider inserting a map with the location of the GB site, and improving current Figure 1 with contours (or the elevation of relevant points).

A contour map was added to Figure 1.

In the same Figure, four crosses are visible, which should refer to four instrumental sites (1 to 4, line 12 page 229). Nonetheless, at line 1, page 229, it is stated that station ‘3’ is actually composed by two different stations, so that the number of stations (5?) does not match with the number of crosses (5 stations, but only 4 crosses). Please clarify.

The station described in bullet point 4 in the previous version of the manuscript was a replacement of the one described in bullet point 2, therefore they share the same cross on Figure 1. These two stations are now combined in the same bullet point to make this clearer.

**I would also appreciate if you could provide some specifications about instrumental resolutions and accuracies;**

A table with specifications of instruments was added.

**5. Section 3.1: according to my opinion, current Appendix could be more significant if incorporated in the main text, namely in this Section;**

Agree. Done

**6. Section 3.2: since snow plays an important role in this analysis, I think more details are due about the parameterization used to model snow settling, albedo and phases dynamics in the snowpack;**

The snow scheme section was expanded.

**7. Lines 17-19, page 233: I would consider adding a quantitative comparison between the outcomes reported here and those by Bewley et al. (2010), such as maximum differences between the two predictions and the data etc. In this way, it would be possible to understand the improvements of your model without looking for the paper by Bewley et al. (2010);**

The quantitative results presented in Bewley et al. (2010) were added to Table 2 and the improvements provided by 3SOM to process representations are now more thoroughly discussed. Please note that results changed slightly between this and the previous version because 1/ the  $F_s$  parametrization was changed following a comment by Reviewer #2 2/ the reference height for wind speed was modified (snowdepth was not previously subtracted); the description of the resistances was added to the revised manuscript.

**8. Figures: as a general suggestion, I would consider reporting more labels in the Figures (such as in Fig. 9 and 10).**

Done.

## RESPONSE TO REVIEWER #2

**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 subtracted). 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  $F_s > 0$ ? 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;  $T_g$  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 km<sup>2</sup>, so what’s outside of the central domain?**

Nothing concerning this study. The sentence was rewritten.

**P243/L7: How are  $F_s$  and  $F_g$  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  $F_s + F_g = 1$ , and  $F_v$  is independent of either (if this is the case).**

$F_g$  is described as  $(1 - F_s)$ . It is now stated that  $F_v$  is independent of either.

**P258: Why is there no measured data above 700 W/m<sup>2</sup>?**

Explanations are now available in the caption of Figure 5.

## **RESPONSE TO REVIEWER #3**

**I would like to see some quantitative analyses of how big this improvement is; 3SOM sounds a conceptually exciting proposition when in a traditional 2-component situation, but this needs to be explicitly demonstrated.**

The quantitative analyses of the 2-source model in Bewley et al. (2010) were added to Table 2 for convenience and to facilitate the discussion showing model improvements. A more thorough comparison of the performance of the two models is now in Section 3.4. Please note that the reference height for wind speed was modified (snowdepth was not previously subtracted). This was not described in the previous version of the manuscript but as this was found to have an effect on model results, the description of the resistances was added to the revised manuscript.

**For example, a short additional experiment testing the impact of gridbox size to evaluate the influence of boundary line location on fluxes would be welcomed.**

Thank you for your suggestion which led to the addition of a new Section (4.1) in the revised manuscript which describes a sensitivity analysis of standard deviation of snow depth to gridbox resolution. We found that modelled standard deviation of snow depth is highly dependent on gridbox resolution and that, as a consequence, larger gridboxes would fail to represent many of the processes explicitly. A short experiment investigating the impact of gridbox size on turbulent fluxes was conducted and summarized in the first paragraph of Section 4.3.

### **Minor comments:**

**P 288, In 7 – please cite the key relevant studies (of the 100 available) rather than rely on the pers comm.**

Done.

**P 232, In 8 – can the difference to JULES albedo be stated briefly to explain why this has been changed?**

The model description has been expanded and the albedo parametrization is now described more thoroughly. As a result of this expansion, this specific reference to JULES was not deemed relevant and was removed.

**P 232, In 20 – quantify how much closer modelled SWE and depth are to measurements in 2004 than 2003.**

This was removed from the revised manuscript because model results changed following the modifications described above.

**P 233, In 27 – ‘perform well enough’. What is well enough? Can a quantitative threshold be provided for this assertion?**

This section now provides a more thorough comparison with the 2-source model and of the improvements that the addition of a 3<sup>rd</sup> source add to representation of processes. The concluding sentence of this section was changed and the terms “well enough” were removed.

**P 234, In 17 – why did you choose a 8 m grid – please justify briefly?**

An explanation is now provided in Section 4.1

**P 234, In 18 – what is the resolution of LiDAR data?**

An explanation is now provided in Section 4.1

**P 235, In 11 – was the WIA – plateau wind speed difference higher or lower?**

Corrected.

**P 236, In 3 – rewrite to say ‘there are some large errors’.**

The sentence was re-written.

**P 236, In 24 – are these ‘errors’ just enhanced uncertainty during melt as a result of increased spatial variability?**

This was clarified in the text.

**P 237, In 2 – you have not currently shown that the models have been able to capture evolution of broad spatial patterns. Need to demonstrate this or re-write.**

A figure of measured vs. modelled snow depth along the transect with the longest persisting snow cover was added to provide some information that spatial averaging in Figure 7 (previously 6) does not clearly convey. We believe that the addition of this Figure, Section 4.1 and clarification in Section 4.3 have now demonstrated this statement.

**P 241, In 28 – in relation to my first main point above, I would suggest that the known limitation (and improvement resulting from this study) need to be explicitly demonstrated here through comparison between 3SOM and the two-source model.**

Done; see answer above.

**Table 1 – although relatively intuitive, please state the units in the table.**

Done (in caption).

**Table 2 - please state units (presume meters?).**

Done (in caption).

**Fig 5 – ‘a’ and ‘b’ are not visible on the plots.**

Changed.

**Are the peaks in the measured data missing prior to April 30? Why?**

An explanation is now provided in the caption.