Thank you for accepting our paper for publication in The Cryosphere! We thank the reviewers for their decision and the time they have taken to review this paper. In light of the comments by Reviewer 4 (in black – only the relevant parts of the reviewer comments are shown), we have made some minor alternations (in blue) to the text in the discussion section:

However, one thing that occurred to me is that the authors note that the simulated snow depth is biased low in the CLM simulations. This low bias makes it difficult to know what emphasis should be placed on the simulated soil temperature biases in the control and modified versions of the model. It may be beyond the scope of this study, but did the authors consider using this low snow depth bias to try to correct the snowfall rates at these sites so as to drive a more realistic seasonal and end of season snowpack depth.

After the following text in the original,

"We note that issues in simulating the initial accumulation of the snowpack are likely linked to uncertainties in the forcing data caused by measurement limitations surrounding the use of precipitation gauges in tundra environments (Smith, 2008; Watson et al., 2008; Pan et al., 2016)."

We have added this text to the discussion (on line 425):

"However, attempting to correct for snow depth errors through adjustment to the precipitation forcing beyond the corrections outlined in Pan et al. (2016) is not advisable due to high variability of snow depth (Fig 2a) over short spatial scales (metres to tens of metres). Additionally, Fig 7c suggests that the timing of the snow onset is more important in determining the soil temperature than the absolute snow depth error, as in 2018-19 soil temperatures simulated using the Sturm parameterisation are closer to observations than in the previous year, despite an absolute snow depth error of up to 0.2m."

If the snowpack depth was represented more realistically, then the comparison of modeled and observed soil temperatures would be more meaningful. I wonder if the relative impact of the different tested parameterizations would be different under a low snow depth bias. If such a simulation is not easy to do, then I would at least suggest a better discussion of the potential interpretation error that the low snow depth bias could impart.

Additional simulations at this point are beyond the scope of this study. However, we added the following text to the discussion (on line 410):

"A similar bias compensation effect could apply for the use of alternative parameterisations of snow thermal conductivity. If snow depth bias was consistently positive, we suspect that the Calonne, Fourteau and Jordan parameterisations would likely compensate for an overthickened snowpack through increased thermal conductivity. However, under a negative snow depth bias the Sturm parameterisation remains more suitable; although the absolute magnitude of the improvement in soil temperatures using the Sturm parameterisation was lower when the snow depth bias was greater in 2017-18, the relative order of impact of the different parameterisations remained the same."