We would like to thank the referee for valuable comments and suggestions our manuscript. Please find our responses and relevant changes (*italic*) to comments (**bold**) below.

Anonymous Referee #2

In this paper, the authors take steps towards an ability to represent in a large-scale model the important lateral snow redistribution, water, and heat processes that impact the trajectory of permafrost thaw and related processes in different permafrost landscapes. The approach is parsimonious, which I like. The authors propose to represent these systems with just two 'tiles' (rim and center for polygonal tundra), rather than explicitly modeling the full complexity of the heterogeneous landscape. I like this approach as it does lend itself to potential inclusion across the pan-Arctic. A significant limitation is that the model is not explicitly modeling the formation of these permafrost landscape features. Instead, the goal is simply to be able to simulate the transition from a low-centered to a high-centered polygon. This is a reasonable first step and the authors acknowledge this limitation. Clearly, to have 'full' confidence in the model, one would want it to be able to simulate the full set of physical processes that drive both the formation and the decay of low-centered polygons. Nonetheless, this is a practical first step that is clearly an improvement over the current 1-tile assumption that cannot at all account for the real spatial heterogeneity of the system.

As noted also in the reply to referee #1, we completely agree that simulating the formation of excess ice would be desirable, although we do not see this as feasible within the current study, both due to the long time scales, and the complexity and lack of well-developed parameterizations for the buildup processes.

Overall, I enjoyed reading this paper and I find it suitable for publication with a few relatively minor revisions and clarifications.

Specific comments

1. When the Noah-MP model is introduced, it would be good to explain why Noah-MP is being used instead of any other model. I believe that it is because of the lateral flow capabilities in WRF-hydro, but that capability isn't introduced until section 2.2.4.

We now include a short justification for the use of this model section 2.2, when the NoahMP model is first introduced.

Furthermore, lateral subsurface water fluxes are already implemented in this model as part of the WRF-Hydro modelling system (see sec. 2.2.4). With some modifications it is therefore a suitable base model for studying the geophysical aspects of permafrost thaw, including the importance of lateral fluxes.

2. P. 8, line 2 typo: "only elevated only"

Corrected. Thank you!

3. I wonder if the "coupled" is the best way to reference the multiple tile simulations. Coupled can mean a lot of things in different contexts. Perhaps you could rename as Reference and Tiled or Single column and Two column or something else that is more descriptive.

We agree that only referring to the two-tiled simulation as the "coupled" simulation is ambiguous. We have now carefully gone through the manuscript to make sure that whenever we refer to the "coupled" simulation, it is clear that we are referring to lateral coupling between tiles.

4. Figure 5: Why is the ref simulation at depth so much warmer than either the RIM or CENTER simulation?

We attribute this to the non-linear effect of snow. Maintaining an almost snow-free rim throughout the winter season increases the heat loss more on the RIM than it is reduced from the CENTER. The tiled system is therefore colder than the REF which receives the average snow accumulation.

5. P.9, Line 16: "The simulated maximum snow depths in 2008 compares quite well with observations for both RIM (0.23 m compared to 0.16 m), and centers (0.39 m compared to 0.46) although the observations show considerable spread (see Nitzbon et al., 2018)." Statements like this are a bit misleading. Should make it clear that the simulated snow depths matching observations is probably mostly good fortune. You are using large-scale forcing from CRU-NCEP. It would be completely unsurprising if the snow depths didn't match up with the observations at the local site when using large-scale forcing. It would be more appropriate to note that due to this good fortune, it is easier to make direct comparisons to observations.

We agree that the raw CRU-NCEP data cannot be expected to reproduce local snow depths accurately, and the agreement is partly due to the scaling factor for precipitation. This is now noted in the text:

This was partly achieved by applying a scaling factor for precipitation (Pscale) of 0.6 (Table 2).

6. P. 10, line 1: Similar to above, the discrepancy in temperature between model and obs is likely substantially a result of using the large-scale CRUNCEP data to force the model. You wouldn't really expect the soil temperatures to match the observed site level soil temperatures in this circumstance.

We again agree that one cannot expect to match soil temperatures exactly when forced with a large-scale reanalysis like CRU-NCEP. This is now pointed out in the discussion section.

However, given the relatively coarse resolution of the forcing data, a certain disagreement must be expected

7. P. 13, line 4: Same again as above. The stability of the peat plateau is at least partly related to what you are getting from the large-scale forcing. You can't go as far as to make the argument that you have to have certain couplings to maintain the peat plateau permafrost, which is what is implied. What you are finding, which is interesting and important, is just that soil conditions are colder on the peat plateau when snow and water coupling is included.

We agree that permafrost could be maintained without these couplings in colder conditions. However, the snow and soil water conditions are recognized also by others as key factors for maintaining these marginal permafrost features in this region, which we now also include a reference for.

This is in agreement with previous studies of palsas and peat plateaus in this region, pointing to low snow accumulation and dry peat during summer as the most important factors for their stability (see Seppäla, 2011).

8. The Discussion section brings up a lot of good points. One thing that isn't clear in the discussion of how one could potentially employ this method at pan-arctic scale is the question of how one would specify the tile structure for each grid cell (is it a polygonal system or a peat plateau, something else, or a mixture of several permafrost landscapes within each large-scale grid cell). Along same lines, how would you know how to initialize the amount and depth of excess ice across the pan-Arctic domain? Based on the information provided in the paper, it seems like this took some trial and error to get it 'right'.

This is a good point. We have expanded the discussion with some more details on this:

Ground ice data from Brown et al. (1998) could provide a starting point here, similar to the study by Lee et al. (2016).Assigning excess ground ice to the first soil layers below the simulated ALT has been a reasonable first-order choice for the two test sites, but this procedure is likely not adequate for areas with excess ice well below the current active layer, e.g. due to burial or melting of excess ground ice in the past (e.g. truncated ice wedges, Brown, 1967). Ultimately, new global data sets for ground ice depth, excess ice density and geometries of the two tiles must be compiled, for example building on approaches as in Hugelius et al, (2014) and Strauss et al., (2017).