C. Koven (Referee)

This paper presents an exploration of a fairly theoretical question: what role does high frequency variability in meteorology have on the low-frequency dynamics of soil temperature in arctic regions. The authors show that, because snow tends to melt during brief warm period in the fall and winter, when temperature variability is reduced, less snow melts, which leads to more snow accumulation and warmer soils. This is not immediately obvious given the expectation that soil acts as a strong low band pass filter on temperature variability, but once explained, it makes perfect sense. Conversely, there is also a mechanism associated with bryophytes, which is less well explained in the text, but which leads to somewhat the opposite response.

We like to thank you for a constructive review that helped to improve the manuscript. Advancements of the previous manuscript are highlighted in red in the revised version of the manuscript.

I have a couple issues with the paper, but I first want to say that I disagree with the first reviewer's comment that this is more appropriate for GMD. This is clearly a model application rather than a model development paper, so I think this journal is an appropriate one for the work.

One issue I have is in the explanation for the bryophyte mechanism. I get the argument about reduced productivity leading to reduced bryophyte cover, but this appears minor in the actual permafrost region itself. So I don't understand why the effects on soil physical properties ought to be large. Is this just an outcome of the soils being warmer and therefore thawed longer in the spring and fall or deeper in the summer under reduced variability, which then leads to different time-averaged soil physical properties since the thawed and frozen states are so different for organic soils? If this is the mechanism, then it is really specific to the bryophyte representation per se, or ought any realistic organic soil parameterization give qualitatively the same result? In any case, the fact that soil temperatures in the annual mean appear to be almost entirely warmer in the reduced variability case argues that the snow mechanism is the dominant one, so you probably ought to state as much in the paper.

The mosses and lichens layer that has been implemented into the LSM (Porada et al., TC, 2016) is a dynamic vegetation model representing near-surface vegetation, in contrast to a soil organic layer. Its cover will change in response to CO2 and climate changes in future. Importantly, this near-surface vegetation also functions as an additional thermal and hydrological layer; it represents water and ice stocks and is part of the energy balance and hydrological scheme. The general impacts of such near-surface vegetation are described in the third paragraph of the introduction. Main functions are the vegetation cover, and its water and ice content. The latter two determine the thermal conductivity and heat capacity and hence thermal diffusivity. In the introduction section lines 54-64 and methods section lines 103-119 and also Porada et al (2016) the functionality of the lichens and bryophyte layer is described more in detail.

Interestingly, in most regions thermal diffusivity of the moss layer is higher in summer and lower in winter when climate variability is reduced (Fig 6, section 3.3). The reason is the impact of climate variability on moss water and ice contents. These effects of climate variability on moss thermal

properties have effects on soil temperature in the same direction than snow depth, it makes the soil warmer. To make this clear, we add the following sentence to results and discussion sections:
"These effects of climate variability on thermal diffusivity of lichens and bryophytes and hence soil temperature are in the same direction as effects."

Note, that in the discussion section, however, we discuss the opposite situation of increase instead of decreasing variability, which will consequently lead to cooler soils. Still, snow and moss effects go in the same direction. Hence, the important effects of climate variability on seasonal moss thermal diffusivity are discussed in lines 308-320.

Your question if such effects can be represented by a soil organic layer parametrization is interesting from a model development point of view. The wish would be to treat mosses and lichens as part of the organic soil horizon because that is sometimes already implemented in LSMs. In reality, mosses and lichens grow on top of an organic layer and their hydrological regimes differ. The vegetation functions will respond faster to environmental change, and hydrological and thermal properties of lichens and bryophytes differ from soil organic layer properties. For example mosses have a higher porosity and therefore can store more water and ice, but potentially also dry faster in summer. Most importantly, we represent moss cover at the sub-grid scale while the organic horizon usually is represented uniform for a certain grid cell. Still, specific model tests with a version including an organic layer that is fully part of water and energy balances (we do not have such representation), and a moss model like Porada et al. (2016) would be required in order to quantify the net difference and to fully answer your question. Such study would be really interesting for a model intercomparison paper about heat conduction representations but is clearly out of the scope of this paper.

A second issue I have is in what these idealized results actually mean in terms of more policy-relevant questions such as climate projections. Under warming, snow cover is expected to be reduced over most of the warmer part of the permafrost region. So that ought to attenuate the importance of the snow mechanism here. Furthermore, models which are driven directly by GCM future scenarios ought to explicitly capture this effect, so in principle this may already be built into those permafrost projections. But perhaps an issue is that for some studies, e.g. those which are used in the permafrost carbon network future scenario MIP, and many papers that are using that output, (and which I've been involved in, hence my waiving of anonymity in this review), the protocol forcing data for the future period is a hybrid that uses high frequency data that comes from reanalysis data, but with climatological anomalies relative to the present from a GCM imposed onto the reanalysis output for future. This design was chosen to avoid the mean-state biases present in almost all GCMs, but perhaps it leads to new problems that are exposed here if the variance changes. I.e., such an experimental design implicitly assumes that only the long-term means are changing and not the high frequency variability about the means, which is still coming from the reanalysis data. So, how big an effect is there from ignoring the projected changes in the variability? I suggest this paper would be most useful to the community if it actually tried to answer this question. Doing so properly would require a few more runs: i.e. something along the lines of taking a GCM scenario meteorology output, and then either include changes to the variance over time or not, while in both cases leaving the changes in the mean state due to the scenario, and then use those to force JSBACH offline to assess how large the bias in projected soil temperature, permafrost extent, etc, changes are in the current PCN-MIP way things are done. I recognize this would require substantially more work on the authors' part, but I'd suggest the authors consider doing such an experiment if they really feel like this is an important result that the community needs to take seriously.

Thank you for this very important point. Actually, since raw ESM output data will be biased we suggest avoiding JSBACH experiments forced by this original climate model outputs. The specific numbers of equilibrium state variables (from model spin-up), such as soil temperature, are important for quantifying any effects due to artificial model experiments. Still, we see a way of directly quantifying this impact of continuously changing short-term variability of meteorological measures into the future. Forcing data for the CNTL run would be from a state-of-the-art statistical bias correction of ESM output data, and for the additional artificial experiment the short-term variability should be dynamically reduced in time until 2100 (Fig 8a,b).

For the revised version of the paper, we extend the study substantially by
i) creating such climate forcing data and
ii) run such additional experiment (REDVARfut) for a single representative region at 70N/100E
in order to demonstrate that by using such slightly different method we will also arrive at the same conclusions: Increasing variability of meteorological measures in future will lead to less pronounced warming of the permafrost soil in response to climate change. For this, methods, results and discussion sections have been substantially advanced (red color). In particular, the reviewer comment and the new experiment also advanced the discussion section: Our findings also show that climate data preparation should continuously be improved also w.r.t. higher statistical moments and a change in short-term variability. This has been also added into the discussion section lines 360-362.

Conclusions and the last sentence of the abstract have been also revised in order to highlight this discussion about future forcing data preparation and the new results.

A last minor issue, on page 9, line 268. This statement needs a lot more support if it is to remain. Several of the models in that study have some version of the effects described here. While every model that does will include effects like organic soil and snow physics differently, it is not at all clear that most of the models in that MIP will not include the essence of these effects. I don't think it is the authors' intention to assert that no other models have snow which melts during brief warm spells, or that the models that include organic soils (but not bryophytes) necessarily miss out on all of those effects described here, so I'd suggest not making the statement.

We fully agree with you and deleted that sentence. It is not the aim of this study to compare different process-oriented models in the way they represent heat conduction.

Minor point: The symbol (-) as a unit in several figures is unclear. I see that you are using it to mean relative fractional differences, but please state that somewhere in each figure

Hope it is more clear now.