We are thankful for the constructive criticism and suggestions which we have considered and incorporated in the revised version of the manuscript, and which have contributed to significant improvements. We provide responses to all points in the following (review in bold font, responses in normal font):

This paper introduces a new land surface scheme, CRYOGRID3 which is a flexible platform for exploring new parameterizations of permafrost processes. In particular, novel parameterizations of thaw processes that account for important lateral and vertical movement of water are developed. The scheme is described in detail and then evaluated against in-situ measurements in NE Siberia. The authors find that the model is able to reproduce long term evolution of the ground thermals regime and also the coupling of atmosphere-surface energy processes. A parameterization is introduced that extends excess ice melt routine of Lee et al. 2014 to include energy transfer in shallow ponds. Finally, future climate simulations are conducted that demonstrate that the hydrological regime can both accelerate and delay permafrost thawing. Overall the paper is well written and structured (with one caveat below), addresses an important question and aims to permit the inclusion of poorly/non-represented sub-grid surface processes in climate model simulations. This is an important and emerging area of research and has the potential to greatly improve the representation of surface-atmosphere interactions with respect to permafrost thaw, particularly in high latitude regions.

MAIN COMMENTS

1. **p.6935 l. 7. I don't think all aims are included in the "in this study..." section. E.g. climate simulations are not mentioned. I think this section could set the scene and introduce the structure of the paper better so the reader can follow more easily where you're going.**

   We have extended the last paragraph of the Introduction: "In this study, we describe a new permafrost modeling scheme, CryoGrid 3, which aims to narrow the gap between traditional transient permafrost models and land-surface schemes employed in Earth System modeling. CryoGrid 3 adds land-atmosphere coupling to the subsurface heat transfer model CryoGrid 2 (Westermann et al., 2013), which has been employed to simulate ground temperatures in Norway (Westermann et al., 2013) and NE Greenland (Westermann et al., 2015). In this way, it is possible to run CryoGrid 3 with the same forcing data as ESM land-surface schemes and compare results of present and future simulations of the ground thermal state. Moreover, the numerical structure is designed to be as simple as possible which eases the requirements for adding and investigating parameterizations for permafrost processes. Following a similar philosophy as Essery (2015), CryoGrid 3 is dedicated to "active users" with the programming skills to modify and extend the code, rather than being a "black box tool" with a predefined and thus limited set of options. To illustrate the potential of the model scheme, we implement simple parameterizations for ground subsidence and initial thermokarst formation and demonstrate the effect on future simulations of the ground thermal regime for a permafrost site in NE Siberia. "

2. **A table summarizing differences between CRYOGRID2 + 3 would be useful to quickly see main processes accounted for and where (which paper) to look for which description.**

   We have inserted a table summarizing the differences between the models CryoGrid 2, 3 and 3 Xice.

3. **The aim to be a modular model is really interesting and goes in the direction of the Factorial Snow Model (Essery 2015) philosophy. These are great and very useful developments. However, (a) this needs to be stressed earlier in the paper I.e. in the Intro/Aims and not left just to p.6960 where it is then really clear what the ambition of this model is. (b) would it be possible to include a figure that schematically illustrates how such a modular model works – which components are targeted, options etc? I think this is quite a strength of CRYOGRID3 and would be great for this potential to be more explicitly presented. I think such a figure would also help in tuning the reader into exactly what this model focuses on.**

a) To make the goals of the model more clear we have inserted the following sentence in the last paragraph of the Introduction: "Following a similar philosophy as Essery (2015), CryoGrid 3 is dedicated to "active users" with the programming skills to modify and extend the code, rather than being a "black box tool" with a predefined and thus limited set of options."

b) CryoGrid 3 constitutes work in progress, so we and other identified users aim to continuously add options to the basis module. However, this process very much depends on the research interests of different users, so we cannot provide a concise overview of the next steps yet. There is work in progress to further explore and improve the representation of water bodies in CryoGrid 3. Another direction of research is the representations of small-scale variability and hereof resulting lateral fluxes of energy, water and snow through interacting ensembles of 1D-realizations of CryoGrid 3. Furthermore, data assimilation schemes for satellite observations are targeted. However, this is ongoing work which may or may not succeed, so it is not adequate to already announce it as a future option in this manuscript. For all future modifications and model options, we aim for dedicated publications including a thorough validation to in-situ data.

4. It would also be nice through this figure (+ short text) to get an idea of what foreseen limits to this modularity are e.g. could you plugin routines from other models which are currently superior in a non-target area of CRYOGRID3 such as the snowpack accumulation/ablation processes?

This is a crucial point for the future development of CryoGrid 3, but frankly said, we cannot provide a definite answer at this point. It would be quite possible to couple a compiled version of e.g. a sophisticated snow model (SNOWPACK, CROCUS) to the ground part of CryoGrid 3 simply by exchanging fluxes of energy after each time step. However, doing so will make it much harder to detect hidden errors and erroneous feedback loops between the two models that could lead to unphysical behavior of the model, since the "other model" would in first place behave as a black box for developers. Conversely, hard-coding and maintaining such a sophisticated scheme as part of CryoGrid 3 will be challenging due to the limited manpower of the CryoGrid 3 developer team. Our feeling is that we would rather aim for the first option if we can expect strong improvements or modifications to a dedicated "submodel" in the near future, while the second option would be preferable if results are satisfactory with an existing scheme (since the considerable effort would have to be done only once).

5. p. 6952 l.2 Any idea how good the ERA snow data is in your area? Limiting accumulation to height of polygon suggests you had too much snow from ERA, which is interesting as often it is found that ERA is negatively biased for precipitation.

When comparing the measurements of late-winter snow water equivalent on Samoylov Island with accumulated winter precipitation form ERA reanalysis, one finds that the latter is considerably larger (more than factor of three in some years, so the bias is clearly positive). However, it is completely unclear if this is due to a considerable overestimation of the true winter precipitation in the reanalysis, or if wind drift removes a large part of the snow (basically everything exceeding the height of the polygons) from the island which is several meters higher than the surrounding arms of the Lena River. Lacking measurements of the "true" winter precipitation (which is a general problem in Arctic conditions), we cannot decide this question and do not wish to go into detail in the present study. However, we note that our procedure is most error-prone during the build-up-phase of the snow cover until it reaches the predefined maximum (but cannot decide if the potential bias would be negative or positive). This was not clearly stated in the original version of the manuscript, but it is possible that it is part of the reason for the model bias in ground temperatures during the freeze-back in fall (Fig. 4), in addition to the already mentioned bias of the onset of the snow cover. We have therefore inserted "The largest uncertainties can be expected during the build-up phase of the snow cover, before the threshold height is reached" in Sect. 4.2. Furthermore, we have inserted a sentence in Sect. 5.1: "…while the length of the zero curtain (…) is in some years underestimated by up to one week. While the reason for this is not entirely clear, we note that the uncertainty of the snow forcing data is largest during the onset and build-up of the annual snow cover (Sect. 4.2)."

6. What is the elevation difference between the ERA grid and surface elevation? Its probably not too much in your relatively flat domain – but still could have an influence on downscaling success so could be interesting to know. This is also relevant to p.6954 l.27: the elevation difference would have an effect on rain/snow threshold if using snow data directly and if CRYOGRID doesn't compute rain/snow based on the downscaled air temperature. Is this likely to have a significant effect on received phase of precip?

In the rather flat landscape of the Lena River Delta, differences between real and model topography most likely do not play a large role. The employed grid cell of the ERA reanalysis features a model elevation of around 40 m a.s.l. and is thus very close to the true altitude of Samoylov Island of around 20 m a.s.l. We presume that the elevation of the model grid cell does not play a strong role for the downscaling and will be overridden by larger-scale effects, e.g. due to an errors in the representation of the sea ice cover in the nearby Laptev Sea. However, we can only speculate on this issue, and therefore do not wish to comment on it in the manuscript.

TECHNICAL COMMENTS

1. **p. 6935 l.18 "up on" → upon**

   done

2. **p.6946 l.1 mains → maintains**

   changed to "remains"

3. **p.6947 l.18 if → is**

   done

4. **p.6949 l.27 in → from**

   done

5. **p.6951 l.28 months → month**

done

6. **p.6953 l.21 snowfall data is also capped to height of polygon as in the vaqlidation run?**

   Yes, it is also capped, thanks for this comment! We have inserted “but as in the validation run, the maximum height of the snow cover is limited to the approximate height of the polygon rims” to make this issue clear.

7. **p.6954 l.9 on → of**

   done

8. **p.6954 l.15 despite of → despite**

   done

9. **p.6954 l.17 remove second use of “snow”**

   done

10. **p.6958 l.14 despite → in despite**

    done

11. **p.6959 l.12 an → until**

    done

12. **p.6965 l12 use of ’hereby’ sounds a bit awkward here. Perhaps just: “...excess ground ice and resulting hydrological processes...”**

    We agree, thanks!

REFERENCES

Essery, R.: A Factorial Snowpack Model (FSM 1.0), Geosci. Model Dev. Discuss., 8, 6583-6609, doi:10.5194/gmdd-8-6583-2015, 2015.

On behalf of all authors,

Sebastian Westermann