Review #1

What is the major message of this study, and for which audience is this message? Sometimes, the manuscript reads like the authors want to inform the atmospheric community that the LSMs have deficits. This impression arises as almost all of the findings in the study are reasonably well known for the cryosphere community, but maybe the findings are not that clear for the atmosphere modelling community. Reading the abstract and the conclusions, the main message is that snow is of high importance for modelling the ground thermal regime close to the surface, which is not a new message. Therefore, before eventual publication, the authors should after my opinion address the following points:

1. The selection of models. You should give reasons why you choose these six models, as there are obviously more models capable to represent ground thermal conditions, both stand-alone models used in permafrost science (e.g. the UaF model GIPL), but also other LSM schemes. Why use Hybrid8? If snow is lacking in the scheme it is clear that the model will fail.

We agree that there are several other models that could be used in such a comparison, and theoretically one should include as many models as possible. The idea was to compare models that act as land-surface schemes in Earth System Models (JSBACH, ORCHIDEE, JULES) and models that run offline at site level (COUP) or as vegetation models (LPJ, HYBRID8). The selection of models is also a result of the existing collaboration within the European projects Greencycles2 and PAGE21, or otherwise between the first author and co-authors. This practical issue was important for managing the task as the site-level model experiments needed to be designed (including preparation of forcing data), performed, and analyzed from scratch. This was a larger project than e.g. downloading model results from a database and analyzing them. The already high organizational workload and required discussions could be managed by running the experiments as part of the EU projects, e.g. using project meetings for discussion etc.
The particular usage of the HYBRID8 model is utterly important to emphasize the segregated effect of having snow or not in a model, and to give a clear quantification of it. We believe this is representative for many other models that are not addressed to cryospheric questions and so our results can be useful to a bigger audience than just the cryosphere community. Hence, we feel that the model results should not be removed even if the effect of no snow on winter soil temperature is qualitatively clear to the audience of this journal.

2. The selection of sites. What is the background motivation here? To show diversity? I am certain there are many sites in the world where the data bases is as good or even better than of some of your sites, so why those sites?

Similar to the motivation of model selection, our study sites are primarily chosen as a result of international networks described above. We are aware of other databases with more sites; however, many Arctic sites lack a complete set of observational variables that are vital for site level model analysis as presented in our manuscript. It is very important to have reliable data that are provided from our coauthors themselves in order to provide forcing data for the models. Having year-long atmospheric forcing data (radiation, temperature, precipitation) and evaluation data (snow depth, soil temperatures) from exact same observational site is not always possible from other databases and using them could lead to more uncertainty in model results and less

confidence in our conclusions. Then, we tried to select sites, which cover a great range of climatic conditions and landscape types.

3. The model domain and discretization. Normally, for permafrost models a domain of a few m depth is not good enough. So please, discuss this issue. There is also a different discretization with depth, also for the topsoil layer. This has an influence and should be quantified somehow.

We thank the reviewer for bringing up this point. Soil depth and layering scheme are indeed important factors for permafrost specific thermal balance modeling. The model depths and discretizations are explained in Appendix A. Most of our models have less than 10 m depths; however, the aim of this manuscript is not to compare models that are particularly developed for permafrost studies but instead using commonly chosen Earth System models and vegetation models to asses their performance. The scope of the paper is to evaluate the current versions of common models in order to guide further improvements. We now mention this in our discussion and clear the importance of soil depth and layering with the following sentences:

"In general, permafrost specific model experiments require deeper soil representation than 5-10 meters. As discussed in Alexeev et al. (2007), at least 30 m soil depth is needed for capturing decadal temperature variations in permafrost soils. The improvements from having such extended soil depth are shown in Lawrence et al. (2012) when compared to their older model version with shallow soil depth (Lawrence and Slater, 2005). Additionally, soil layer discretization plays an important role for the accuracy of heat/water transfer within the soil, and hence can effect the ALT estimations. Most of the models in our intercomparison have less than 10 m depths, so they lack some effects of deep soil factors. However, most of the models used in global climate simulations have similar soil depth representations and the scope here is to compare models that are not directly aimed to simulate permafrost and to show general guidelines for future model developments."

4. The modelled deviations in relation to observations: Most of the deviations described in the manuscript are caused by different snow schemes of the models. This is fine, but it is certainly not new or surprising that near surface or ground temperatures then deviate, especially during winter. And of course, if snow is wrong, ground temperatures become wrong independently how good the soil physics is represented in the model

This is absolutely true. The importance of snow is known in the cryospheric community for long time. However, this is the first time many different parts of cryospheric research (modeller-field scientists, high latitude-high altitude conditions, permafrost-nonpermafrost sites, global modelers-site scale modellers) are compared in a single study, so the overall conclusions include something new for one party or another. Most importantly, Earth System models are being used for future climate change scenarios and our results can emphasize the need for better model representations for such projections.

5. The paper structure: The paper is very hard to read, you should really consider distinguishing between pure results and discussion. This is intermixed, and therefore you must write very often sentences in the sense of "as mentioned above, snow matters. . ..". It is enough to do this once in a paper, and that you can achieve if you discuss snow in a discussion paragraph. Now, you do this for all result paragraphs. This would also considerably shorten the paper.

We have chosen to combine results and discussion because the paper structure includes analyzing results from surface to deeper soil layers in a sequential order. The discussion points of the first part are needed to analyze the results of the following sections. As an example, the discussion of topsoil comparisons highlight the importance of snow insulation process for soil temperatures, however in section 3.2 and 3.3, the subsoil thermal dynamics/ALT comparisons reveal additional important factors (soil heat transfer/hydrology) for modeling soil temperatures and needs to mention the previous discussion points. We agree that this choice makes the paper a bit harder to follow, so we have revised the manuscript in order to clear the message by reducing some general parts (see below point). If the reviewers/editor wishes, we can change the structure, but we would require more time for that.

6. You write several times long textbook-like paragraphs that snow insulates, vegetation insulates, the soil water content is important etc, this is information well-known for the cryosphere community and can mainly be omitted.

We agree with the reviewer. This knowledge is basic knowledge in the cryosphere community. However, as such processes are being implemented into Earth system models or vegetation models, our paper also addresses a bit wider community and we feel that some basics could be repeated. At the same time, we have improved the text according to your points by removing some of the general descriptions like:

p4973,l15-20
"Modeling soil heat transfer requires accurate coupling of atmosphere and soil. Vegetation cover, snow pack, litter layers, organic layer or a combination of these can form a buffer zone at this interface. This buffer zone provides thermal insulation for soil. Especially at high latitudes/altitudes, where soil is snow-covered during most of the year, insulation effects are critical for estimating the soil thermal regime as well as the active layer thickness and permafrost conditions."
p4976,l8-10
"After snowmelt, surface vegetation cover and organic layer form the buffer zone at the atmosphere/soil interface. Insulating properties of vegetation cover depend on the type of species, its thickness and wetness of the layer. In our study sites (except for the bare Schilthorn site), vegetation comprise of moss cover with ca. 5–15 cm thickness."
p4978,l5-12
"Being in contact with the atmosphere and surface cover, topsoil conditions show strong relations to seasonal surface processes and atmospheric changes. Despite their importance for the surface, above soil conditions are not enough to explain all the changes in subsoil temperature dynamics. Soil water and ice content, soil freezing and thawing, the characteristics and depth of the permafrost layer and its specific heat and water transfer mechanisms are especially important for cold regions' soil thermal regime. Therefore it is essential for models to capture deeper soil temperature dynamics together with the surface conditions."

7. For the ALT you compare a transient approach with an equilibrium approach, of course there are deviations as the simple semi-empirical relations between GST and ALT normally relate to an equilibrium climate conditions, e.g. a climate normal.

It is not clear to us what the reviewer wants us to do here. We refer to the review by Riseborough et al., which clearly discusses the drawbacks of semi-empirical approaches. We have also added the following sentences to the end of section 3.3 in order to clear the message: "*The equilibrium models are not capable to estimate ALT in long-term*

simulations. So, using transient models and considering internal soil physical factors are critical to properly assess ALT within climate change context."

8. Conclusions: 1-4 and 5-7 are conclusions from your intercomparison, but not new for the cryosphere community. Conclusion 5 I do not understand. Please explain.

For the first 4 conclusion points, please see comments above to point 4. For the conclusion point 5, we have rephrased it as: "*Model heat transfer rates differ due to coupled heat transfer and hydrological processes. This leads to discrepancies in subsoil thermal dynamics.*"

9. The model spin-up: Why different spin-up times for the different models, why not the same spin up, or you spin up one model, and start from this for all the models? Maybe this is better for intercomparison? Probably, I miss something here, but anyway.

Model spin-up is performed in order to reach steady state of state variables (e.g. temperature) to mean climate conditions before starting transient simulations under changing conditions, for which models are using the exact same climate files as described in our manuscript. The procedure and time period required to perform this spin-up is dependent on model formulations and particular set-up. For example, the deeper the soil layering is considered, the longer time is required for bringing the last layers temperature in steady state with atmospheric forcing. That's why model spin-up is performed separately.

10. Introduction: Really, give a good scientific question; objectives etc. (see comments above)

The scientific questions and objectives are summed up in the last paragraph of the introduction: "*We have compared the performances of six different land models in simulating soil thermal dynamics at four contrasting sites. In comparison to previous works, we used advanced model versions specifically improved for cold regions and our model simulations are driven by (and evaluated with) site observations. To represent a wider range of assessment and model structures, we used both land components of ESMs (JSBACH, ORCHIDEE, JULES) and stand-alone models (COUP, HYBRID8, LPJ-GUESS), and compared them at Arctic permafrost, Alpine permafrost and Arctic non-permafrost sites. By doing so, we aimed to quantify the importance of different processes, to determine the general shortcomings of current model versions and finally to highlight the key processes for future model developments.*"
See also our replies to the previous comments.

Minor editorial points: - a location map is missing, I have no idea where all these sites are located - many of the figures are hard to read

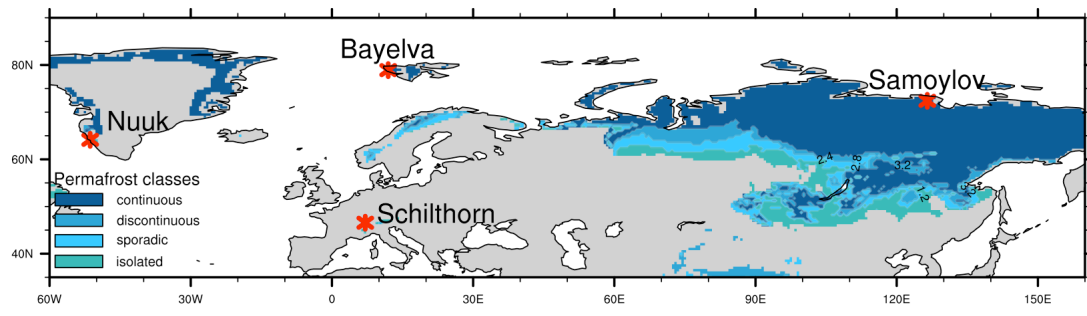In addition to the already provided site information table, we have now included a location map (Figure 1).

Figures

Figure 1: Site location map.