We thank the reviewers for their careful review and helpful comments. Detailed responses to each comment raised by the reviewers are given in blue as follows.

On behalf of all co-authors,

Shushi PENG

Responses to Anonymous Referee #2

The manuscript presents a comparison of soil temperature trends simulated by nine numerical models. Texts are well organized and clearly written, and tables and diagrams are effective in showing the variability among simulated soil temperatures. It is an interesting exercise to demonstrate the large uncertainty in soil temperature simulation in the permafrost region. However, I am not sure what scientific advances we gain from this exercise. If I am not mistaken, the nine models are driven by different climatic forcings, and have vastly different structures and algorithms (Table 1). Therefore, the variability in model results is due to the variability in both forcing and model algorithms. As a result, the reader is left wondering what the results of this exercise really mean. For example, ColM and ISBA show completely different patterns of soil temperature trends (Figure 4). Is this caused by model algorithms or climate forcing?

As someone who has done much work on the comparison between permafrost model simulations with field observations, I am keenly aware of model sensitivity to subtle changes in surface variables (e.g. vegetation parameterization), subsurface variables (e.g. soil moisture), and boundary conditions (e.g. magnitude of geothermal flux). Permafrost models are also sensitive to the initial condition, as well as the thickness of the model domain. In order of the reader to understand the meaning and implication of model results, it would have been much more meaningful to conduct the model comparison exercises using a common set of forcing variables. For these reason, I cannot recommend publication of this manuscript in its present form. I suggest that the authors re-design and conduct new model comparison exercises, or present more meaning explanation for differences among the present simulation results and discuss how what causes these differences.

[Response] We agree that it is easier to understand the comparison exercise using the same climate forcing. However, as we shown in this manuscript, large uncertainty of climate forcing in high latitude regions also can cause the spread of modeled Ts, besides the spread of modeled Ts caused by model parameters and structure. Previous studies (e.g. Troy and Wood, 2009; Rawlins et al., 2010) shown that large spread of climate variables over Arctic regions. Troy and Wood (2009) reported 15-20 W m^{-2} of differences in radiative fluxes on seasonal timescales over northern Eurasia, between six gridded products. Between different gridded observations and reanalysis precipitation products, the magnitude of Arctic precipitation ranges from 410 mm yr^{-1} to 520 mm yr^{-1}, and the trend of Arctic precipitation also has a large spread (Rawlins et al., 2010). We do not know whether the spread of modeled trend of Ts among applications of land surface models applied to the permafrost region associated with uncertainty of climate forcing is larger than that caused by uncertainty associated different parameterization and processes in models. One protocol model inter-comparison exercise to quantify the relative uncertainty between these two sources would be to apply several climate forcing data sets with each model and then partition the variability among the climate drivers and the models. Alternatively, another approach is allow the models to conduct simulations with their preferred historical climate data, and then use the methodology in this study to partition results key aspects of variability among climate drivers (as inferred from ancillary simulations with detrended climatic components) and models. For practical reasons, the simulation protocol in the model-intercomparison project of Permafrost Carbon Network (PCN, www.permafrostcarbon.org) allow each model group to choose climate forcing and other boundary files (such as vegetation map, soil map etc.) to fully quantify the uncertainty of model results among uncertainties from climate forcing (and other forcing files) and model parameterization and structure. We indicate in the revised manuscript that future model inter-comparisons on permafrost dynamics should investigate the full uncertainty by conducting simulations for multiple climate forcing data sets.

We clarified that the main objective of this study is to distinguish the uncertainty caused by assigned parameter values and model structure from the uncertainty attributable to uncertain climate forcing data. We accomplished this by an approach of separating functional uncertainty from structural uncertainty. For the explanation of difference between models, we discussed the possible causes. In Figure 7-9, we show that the climate forcing could explain most of the differences among models. We also added further discussion about the differences in the revised version.

In addition to the fundamental comments above, the following is specific comments.

Page 2305, Line 11-16. What are the scientific objectives of this work? The objectives (1)-(3) cannot be meaningfully achieved, if model simulation results are strongly dependent on model algorithms and structure.

[Response] As mentioned in the response to fundamental comments above, we try to use the information from this model inter-comparison as much as possible to show the range of modeled Ts trend, their possible drivers (Table 3) and observation-constraint Ts trend and permafrost retreat rate (Figure 8 and 9).

Page 2305, Line 19-23. In addition to surface forcing, the forcing from the bottom boundary of models needs to be explained clearly. Energy input in the form of geothermal flux has strong effects on soil temperature.

[Response] We agree that bottom geothermal flux has effects on simulated soil temperature. We added the number of geothermal flux for each model in Table 1.

Page 2305, Line 23-24. If I have understood correctly, three out of nine models do not consider the "effects of water in soil on phase change". Does that mean these three models do not simulate the freezing and thawing of soil water? Since permafrost is the phenomenon of pore water freezing and thawing, I am not sure if these three models are even suitable for the purpose of this exercise. Clear justification is needed for the inclusion of these models.

[Response] Most models used in this study are the land surface component of Earth System Models participating CMIP5 project. The original idea was to compare these models and to show possible guidelines for further permafrost simulations in CMIP6 or other future projects. Actually, two of the nine models used in this study (CoLM and MIROC-ESM) do not represent the effects of water phase change processes (freezing and thawing) in frozen soil. But these two models represent other important processes/issues such as maximum 3 and 5 snow layers in MIROC-ESM and CoLM, respectively. We think that it is important to include all the nine models to show their results for comparison.

Page 2306, Line 8-10. Why were different forcing data sets used for different models? Clear justification is needed in relation to the scientific objectives of the study.

[Response] Please see the response to general comments. We added two sentences in Introduction and Methods section as below.

"both uncertainties from climate forcing (and other forcing files) and from model parameterization and structure"

"The different modeling groups in this study used different forcing datasets for climate and other model boundary conditions (Table 1), which collectively represent both uncertainty from climate forcing (and other forcing files) and from model parameterization and structure in simulating soil thermal dynamics across the permafrost region."

Page 2307, Line 18-22. The thermal condition of top 3m is strongly dependent on the presence of absence of permafrost in the underlying zone. In some regions permafrost is more than 10-20m thick. Table 1 indicates that some of the models are not sufficiently deep to represent the effects of underlying permafrost.

[Response] Yes, we pointed this out in the text and suggested that models should have deeper soil depth to simulate permafrost as below.

"…This suggests that much deeper maximum soil depth than the currently prescribed maximum soil depths (Table 1) are needed for some models to calculate the heat flux into the entire soil profile (Stevens et al., 2007). CoLM, JULES and LPJ-GUESS have too shallow maximum soil depth for the calculation of permafrost soil temperature trends over the last four decades, which makes these models even less realistic for deeper Ts projections over the next century (e.g. Alexeev et al., 2007)."

References

Alexeev, V. A., Nicolsky, D. J., Romanovsky, V. E., and Lawrence, D. M.: An evaluation of deep soil configurations in the CLM3 for improved representation of permafrost, Geophysical Research Letters, 34, L09502, doi:10.1029/2007gl029536, 2007.

Rawlins, M. A., Steele, M., Holland, M. M., Adam, J. C., Cherry, J. E., Francis, J. A., Groisman, P. Y., Hinzman, L. D., Huntington, T. G., Kane, D. L., Kimball, J. S., Kwok, R., Lammers, R. B., Lee, C. M., Lettenmaier, D. P., McDonald, K. C., Podest, E., Pundsack, J. W., Rudels, B., Serreze, M. C., Shiklomanov, A., Skagseth, O., Troy, T. J., Vorosmarty, C. J., Wensnahan, M., Wood, E. F., Woodgate, R., Yang, D. Q., Zhang, K., and Zhang, T. J.: Analysis of the Arctic System for Freshwater Cycle Intensification: Observations and Expectations, Journal of Climate, 23, 5715-5737, doi:10.1175/2010jcli3421.1, 2010.

Stevens, M. B., Smerdon, J. E., González-Rouco, J. F., Stieglitz, M., and Beltrami, H.: Effects of bottom boundary placement on subsurface heat storage: Implications for climate model simulations, Geophysical Research Letters, 34, L02702, doi:10.1029/2006gl028546, 2007.

Troy, T. J., and Wood, E. F.: Comparison and evaluation of gridded radiation products across northern Eurasia, Environmental Research Letters, 4, 045008, doi:10.1088/1748-9326/4/4/045008, 2009.