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 #1

General comments: Peng et al. compares simulated change in soil temperature from 1960 to 2000 with nine process-based land surface models which are forced with historical data. Through a set of additional factorial simulations the effect of different forcing variables on soil temperature is explored as well as the uncertainty contribution from forcing versus model structure (and parameters). Finally, they study present estimates of how much the near surface permafrost area has decreased in this period and its sensitivity to soil temperature. Altogether it is my understanding that this study presents new results which fits with the scope of The Cryosphere. The aim is clearly stated and the manuscript is with few exceptions well written and logically structured. However, the discussion should in my opinion be deeper on some points before the conclusions can be drawn. Also, it seems that the spread in model results is not always reflected in the text and not shown on certain figures. I therefore recommend the paper to be published in The Cryosphere after major revisions.

[Response] Overall, we revised the manuscript following your suggestions. We expanded the discussion section, especially on the uncertainty of Ts trend between models. The details are provided below.

Specific comments:

I would have liked to see a deeper discussion about the differences between the models and to what extent this can be assumed to represent the uncertainties in our understanding of the actual system. This issue is briefly touched upon in the last part of the conclusion, but should in my opinion be expanded, and be discussed before conclusions are drawn.

[Response] We expanded the discussion about uncertainty of modeling Ts including climate forcing uncertainty and 'cold processes' in models as below.

"Meteorological stations are sparse in the cold permafrost regions. For example, there are only 8.8 stations per million km^2 north of 60°N in the CRU TS3.22 gridded air temperature product compared to 41.1 stations per million km^2 between 25°N and 60°N. This results in uncertainty in gridded climate products over Arctic regions, especially for trends of Arctic climate variables (Mitchell and Jones, 2005; Troy and Wood, 2009; Rawlins et al., 2010; Weedon et al., 2011). 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). These large uncertainties in climate forcing in Arctic undoubtedly can cause large spread of modeled T$_s$. We found that the FU dominates the total uncertainty of $\dot{T}_s$. This suggests that modelers not only need to improve their models, but also need better climate forcing data (or need to test the effects of different climate input data) when modeling long term changes of T$_s$ in permafrost regions. However, to quantify the SU, simulations using the same agreed upon climate forcing data are highly recommended to further attribute the contribution of each process in the soil thermal dynamics of models such as organic carbon insulation effects, snow insulation effects, latent heat formation and emission, soil conductivity and surface properties (see Lawrence and Slater, 2008; Koven et al., 2009; Bonfils et al., 2012; Gouttevin et al., 2012). In addition, important processes in permafrost regions such as ice content (e.g. ice wedge) in permafrost and thermokarst lakes etc. should be developed in land surface models to improve the prediction of future permafrost feedbacks (e.g. van Huissteden et al., 2013; Lee et al., 2014).

The discussion of the different drivers of trend in soil temperature (section 3.3) does not reflect the importance of the different variables. In particular the importance of Ta and the spread in sensitivity to this variable should be discussed more. For LWDR it should be more clearly stated that this is based on only two models, which does not have a representative sensitivity to for instance Ta. This should also be reflected in the conclusions about which variables are most important in driving Ts.

[Response] We clarified that the contribution of LWDR on the trend of Ts is based on only two models (JULES and ORCHIDEE) in the Conclusions section. "Note that the relative contribution of LWDR is based on only two models in this study, and this needs further investigation." was added into the revised version.

We also added the discussion about Ta and Ts as below.

"This indicates the importance of increasing Ta on the trend of Ts, and is consistent with observations. Based on 30 climate stations observations in Canada during the period 1958-2008, Ts at 10 cm significantly and positively correlates with Ta at most sites (>90%) in spring, but at fewer sites (<30%) in winter (Qian et al., 2011). For winter Ts, the winter snow depth was found to have significant and positive correlation with Ts in shallow soil layers (e.g. Zhang et al., 2001; Qian et al., 2011). Recent increases in Ta also explain the trend of Ts at 1.6m measured at Churapcha metrological station (N62.02, E132.36), and at 5 m measured in a borehole at Iqaluit (N63.47, W68.48) in Canada (Smith et al., 2005; Romanovsky et al., 2007). To some extent, the trend of Ta is a good indicator for the trend of deep permafrost ground temperature with some time lag (Romanovsky et al., 2007). For the modeled Ts in land surface models, the effects of Ta on Ts depend on surface energy balance and ground heat flux into soil; i.e. the extent of coupled Ta on Ts relates to the surface properties such as snow, organic soil horizons and rougness etc. in the models. The

different relative contributions of the trend of Ta to the trend of Ts in these models maybe mainly result from the different model parameterization and structures, as the trends of Ta (~0.03 °C yr-1) in the climate forcing do not have a large spread (Figure 7)."

The statement that in UW-VIC "nearly 100 %" of the trend in soil temperature can be explained with Ta (P2310 line 17-19) does not reflect the results presented in table 3.

[Response] We corrected it.

Figure 4 and 6 seems to show a too narrow range of values, excluding the extremes. This is clearly also the case in figure 5, but here I think it is sufficient to include a note about this in the figure caption.

[Response] We clarified this in the figure captions for these two figures.

P2306 line 25: The term "some of their non-linear interactions" should be explained better.

[Response] We used the difference between two simulations with and without the trend of one driver to identify the effects of such driver on Ts. For example, R01 includes the effects of all drivers, and R02 includes the effects of all drivers except Ta. In R02 with detrended Ta, the interactions between increasing Ta and CO_2 or other drivers were also excluded. Thus, the difference between R01 and R02 shows the effects of Ta and some interactions between Ta and other drivers. We removed "non-linear" here.

Technical comments:

P2306 line 6: Either "hereafter" or "in the following" should be sufficient.

[Response] We corrected it.

P2306: The sentence starting with "To separate . . ." is not very fluent. Dropping the ":" and writing the variables in a parenthesis would make it clearer.

[Response] We corrected it.

P2307: I found the sentence staring with "if the maximum soil depth. . ." hard to understand. Please clarify.

[Response] We revised this sentence as "Modeled Ts at depths deeper than 300 cm (six models modeled Ts deeper than 300 cm, except CoLM, JULES and LPJ-GUESS) was not extrapolated (the maximum soil depth of each model is shown in Table 1).".

P2307 line 7: differenced => differences

[Response] We corrected it.

P2307 line 27: Wei et al. 2013 not found in reference list. Is it Wei et al. 2014?

[Response] We corrected it.

P2308 line 3: which "seven models"?

[Response] We clarified here "seven models" out of the nine models except LPJ-GUESS and UVic because LWDR was not used by these two models.

P2308 line 22: Unclear sentence.

[Response] We revised this sentence as "The trend of Ts at different soil depths is shown in Figure 5 for each model.".

P2309 line 8-9/Figure 5: why is CLM only showed down to 35 m?

[Response] In the previous version, we only plotted the depth of soil node for each soil layer. In CLM, the node of deepest soil layer in CLM is 38.2 m, and the bottom of deepest soil layer is 45.1 m. We re-plotted Figure 5, and extended the soil depth to the bottom for each soil layer.

P2309 line 12-13: Please elaborate on the negative trend seen in UW-VIC below 2.5m.

[Response] The negative trend in UW-VIC below 2.5 m is about -0.035 °C yr^{-1}. We clarified this in the revised version.

P2311 line 12: Please rephrase. Does not all show trend since 1960?

[Response] We corrected it.

P2311 line 14: is => are

[Response] We corrected it.

P2311: Please rephrase last sentence.

[Response] We revised the last sentence as "The effects of snowfall trends and growing season precipitation trends may oppose each other as mentioned above. These two contrasting effects cannot be separated in this analysis, because models did not run simulations with seasonally detrended precipitation. But the different effects of seasonal precipitation on Ts should be studied in the future.".

P2315: Please rephrase first sentence.

[Response] We revised this sentence as "The total boreal NSPA during 1960-2000 estimated by the nine models ranges from 6.8 million km2 (CoLM) to 19.7 million km2 (ORCHIDEE).".

P2315 line 7: Lawrence and Slanter et al. 2013 => Slanter and Lawrence 2013 ?

[Response] We corrected it.

References

Bonfils, C. J. W., Phillips, T. J., Lawrence, D. M., Cameron-Smith, P., Riley, W. J., and Subin, Z. M.: On the influence of shrub height and expansion on northern high latitude climate, Environmental Research Letters, 7, 015503, doi:10.1088/1748-9326/7/1/015503, 2012.

Gouttevin, I., Krinner, G., Ciais, P., Polcher, J., and Legout, C.: Multi-scale validation of a new soil freezing scheme for a land-surface model with physically-based hydrology, The Cryosphere, 6, 407-430, doi:10.5194/tc-6-407-2012, 2012.

Koven, C., Friedlingstein, P., Ciais, P., Khvorostyanov, D., Krinner, G., and Tarnocai, C.: On the formation of high-latitude soil carbon stocks: Effects of cryoturbation and insulation by organic matter in a land surface model, Geophysical Research Letters, 36, L21501, doi:10.1029/2009gl040150, 2009.

Lawrence, D., and Slater, A.: Incorporating organic soil into a global climate model, Clim Dyn, 30, 145-160, doi:10.1007/s00382-007-0278-1, 2008.

Lee, H., Swenson, S. C., Slater, A. G., and Lawrence, D. M.: Effects of excess ground ice on projections of permafrost in a warming climate, Environmental Research Letters, 9, 124006, doi:10.1088/1748-9326/9/12/124006, 2014.

Mitchell, T. D., and Jones, P. D.: An improved method of constructing a database of monthly climate observations and associated high-resolution grids, International Journal of Climatology, 25, 693-712, doi:10.1002/joc.1181, 2005.

Qian, B., Gregorich, E. G., Gameda, S., Hopkins, D. W., and Wang, X. L.: Observed soil temperature trends associated with climate change in Canada, Journal of Geophysical Research: Atmospheres, 116, D02106, doi:10.1029/2010jd015012, 2011.

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.

Romanovsky, V. E., Sazonova, T. S., Balobaev, V. T., Shender, N. I., and Sergueev, D. O.: Past and recent changes in air and permafrost temperatures in eastern Siberia, Global and Planetary Change, 56, 399-413, doi:10.1016/j.gloplacha.2006.07.022, 2007.

Smith, S. L., Burgess, M. M., Riseborough, D., and Mark Nixon, F.: Recent trends from Canadian permafrost thermal monitoring network sites, Permafrost and Periglacial Processes, 16, 19-30, doi:10.1002/ppp.511, 2005.

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.

van Huissteden, J., Berrittella, C., Parmentier, F. J. W., Mi, Y., Maximov, T. C., and Dolman, A. J.: Methane emissions from permafrost thaw lakes limited by lake drainage, Nature Clim. Change, 1, 119-123, doi:dx.doi.org/10.1038/nclimate1101, 2011.

Weedon, G. P., Balsamo, G., Bellouin, N., Gomes, S., Best, M. J., and Viterbo, P.: The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data, Water Resources Research, 50, 7505-7514, doi:10.1002/2014wr015638, 2014.