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 km² north of 60°N in the CRU TS3.22 gridded air temperature product compared to 41.1 stations per million km² 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 \(T_a\) 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 \(T_a\). This should also be reflected in the conclusions about which variables are most important in driving \(T_s\).

[Response] We clarified that the contribution of LWDR on the trend of \(T_s\) 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 \(T_a\) and \(T_s\) as below.

“This indicates the importance of increasing \(T_a\) on the trend of \(T_s\), and is consistent with observations. Based on 30 climate stations observations in Canada during the period 1958-2008, \(T_s\) at 10 cm significantly and positively correlates with \(T_a\) at most sites (>90%) in spring, but at fewer sites (<30%) in winter (Qian et al., 2011). For winter \(T_s\), the winter snow depth was found to have significant and positive correlation with \(T_s\) in shallow soil layers (e.g. Zhang et al., 2001; Qian et al., 2011). Recent increases in \(T_a\) also explain the trend of \(T_s\) 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 \(T_a\) is a good indicator for the trend of deep permafrost ground temperature with some time lag (Romanovsky et al., 2007). For the modeled \(T_s\) in land surface models, the effects of \(T_a\) on \(T_s\) depend on surface energy balance and ground heat flux into soil; i.e. the extent of coupled \(T_a\) on \(T_s\) relates to the surface properties such as snow, organic soil horizons and roughness etc. in the models. The
different relative contributions of the trend of Ta to the trend of Ts in these models may
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 CO2 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.
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.

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 ?

We corrected it.

References


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 to 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]**
[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
Rawlins, M. A., Steele, M., Holland, M. M., Adam, J. C., Cherry, J. E., Francis, J. A.,


Simulated high-latitude soil thermal dynamics
during the past four decades

Shushi Peng¹,², Philippe Ciais², Gerhard Krinner¹, Tao Wang¹,², Isabelle Gouttevin¹,³, A. David McGuire⁴, David Lawrence⁵, Eleanor Burke⁶, Xiaodong Chen⁷, Christine Delire⁸, Charles Koven⁹, Andrew MacDougall¹⁰, Annette Rinke¹¹,¹², Kazuyuki Saito¹³, Wenxin Zhang¹⁴, Ramdane Alkama⁸, Theodore J. Bohn¹⁵, Bertrand Decharme⁸, Tomohiro Hajima¹³, Duoying Ji¹⁶, Dennis P. Lettenmaier², Paul A. Miller¹⁴, John C. Moore¹¹, Benjamin Smith¹⁴, Tetsuo Sueyoshi¹⁶,¹³

¹UJF – Grenoble 1/CNRS, Laboratoire de Glaciologie et Géophysique de l’Environnement (LGGE), 38041 Grenoble, France
²Laboratoire des Sciences du Climat et de l’Environnement (LSCE), CEA-CNRS-UVSQ, 91191 Gif-sur-Yvette, France
³Irstea, UR HHLY, 5 rue de la Doua, CS 70077, 69626 Villeurbanne Cedex, France
⁴U.S. Geological Survey, Alaska Cooperative Fish and Wildlife Research Unit, University of Alaska Fairbanks, Fairbanks, AK, USA
⁵National Center for Atmospheric Research, Boulder, CO, USA
⁶Met Office Hadley Centre, FitzRoy Road, Exeter, EX1 3PB, UK
⁷Department of Civil and Environmental Engineering, University of Washington, Seattle, WA, USA
⁸CNRM-GAME, Unité mixte de recherche CNRS/Meteo-France (UMR 3589), 42 avCoriolis, 31057 Toulouse cedex, France
⁹Lawrence Berkeley National Laboratory, Berkeley, CA, USA
Revised Manuscript for *The Cryosphere*

Jul 31, 2015

392 words (abstract) + 5360 words (text) + 3 table + 9 figures + 60 references

*Corresponding author:*
Shushi Peng
LGGE/LSCE, France
Tel: +0033-1 69 08 53 03
Email: Shushi.Peng@lsce.ipsl.fr
**Abstract**

Soil temperature ($T_s$) change is a key indicator of the dynamics of permafrost. On seasonal and inter-annual time scales, the variability of $T_s$ determines the active layer depth, which regulates hydrological soil properties and biogeochemical processes. On the multi-decadal scale, increasing $T_s$ not only drives permafrost thaw/retreat, but can also trigger and accelerate the decomposition of soil organic carbon. The magnitude of permafrost carbon feedbacks is thus closely linked to the rate of change of soil thermal regimes. In this study, we used nine process-based ecosystem models with permafrost processes, all forced by different observation-based climate forcing during the period 1960-2000, to characterize the warming rate of $T_s$ in permafrost regions. There is a large spread of $T_s$ trends at 20 cm depth across the models, with trend values ranging from $0.010 \pm 0.003 \degree C \text{yr}^{-1}$ to $0.031 \pm 0.005 \degree C \text{yr}^{-1}$. Most models show smaller increase in $T_s$ with increasing depth. Air temperature ($T_a$) and longwave downward radiation (LWDR) are the main drivers of $T_s$ trends, but their relative contributions differ amongst the models. Different trends of LWDR used in the forcing of models can explain 61% of their differences in $T_s$ trends, while trends of $T_a$ only explain 5% of the differences in $T_s$ trends. Uncertain climate forcing contributes a larger uncertainty in $T_s$ trends ($0.021 \pm 0.008 \degree C \text{yr}^{-1}$, mean $\pm$ standard deviation) than the uncertainty of model structure ($0.012 \pm 0.001 \degree C \text{yr}^{-1}$), diagnosed from the range of response between different models, normalized to the same forcing. In addition, the loss rate of near-surface permafrost area, defined as total area where the maximum seasonal active layer thickness (ALT) is less than 3 m loss rate is found to be significantly correlated with the magnitude of the trends of $T_s$ at 1 m depth across the models ($R=-0.85$, $P=0.003$), but not with the initial total near-surface permafrost area ($R=-0.30$, $P=0.438$). The sensitivity of the total boreal near-surface permafrost area to $T_s$ at 1 m, is estimated to be of $-2.80 \pm 0.67$ million km$^2 \degree C^{-1}$. Finally, by using two long-term LWDR datasets and relationships between trends of LWDR and $T_s$ across
models, we infer an observation-constrained total boreal near-surface permafrost area decrease comprised between $39 \pm 14 \times 10^3$ and $75 \pm 14 \times 10^3$ km$^2$ yr$^{-1}$ from 1960 to 2000. This corresponds to 9% - 18% degradation of the current permafrost area.

Key words: soil temperature, permafrost, downward longwave radiation, climate warming, land surface model
1 Introduction

Arctic permafrost regions store ~1300 Pg carbon (C) in the soil, including ~1100 Pg C in frozen soil and deposits (Hugelius et al., 2014). Decomposition of these large carbon pools in response to permafrost thawing from projected future warming is expected to be a positive feedback on climate warming through increased emissions of CO₂ and CH₄ (Khvorostyanov et al., 2008; Schuur et al., 2008; McGuire et al., 2009; Koven et al., 2011; Schaefer et al., 2011). The magnitude of permafrost soil carbon feedbacks on climate depends on the rate of soil carbon decomposition, which is related to permafrost thaw, soil water and temperature changes, the quantity and quality of soil carbon available as a substrate for decomposition, and the concentration of oxygen in the soil, which determines CH₄ vs. CO₂ production ratio (Schuur et al., 2008; Schädel et al., 2014; Elberling et al., 2013). Both the rate of permafrost thaw and the rate of soil carbon decomposition are closely related to soil thermal dynamics (Koven et al., 2011; Schädel et al., 2014; Elberling et al., 2013).

Measurements of active layer depth across circumpolar regions and borehole temperature profiles indicate that active layer thickness on top of boreal permafrost has been increasing in response to the warming that occurred during recent decades in North America, Northern Europe and Russia (e.g. Zhang et al., 2001; Qian et al., 2011; Smith et al., 2005, 2010; Romanovsky et al., 2007, 2010). For example, the borehole record of Alert in Canada (82°30′N, 62°25′W) shows that soil temperature at 9 m, 15 m and 24 m increased at rates of 0.6 °C decade⁻¹, 0.4 °C decade⁻¹ and 0.2 °C decade⁻¹ from 1978 to 2007, respectively (Smith et al., 2012). These observations provide long-term local monitoring of changes in active layer thickness and soil temperature, but the measurement sites are sparse, and their temporal sampling frequency is often low (Romanovsky et al., 2010). Because site measurements cannot document permafrost area loss on a large scale, land surface models including ‘cold
processes’, such as soil freeze-thaw and the thermal and radiative properties of snow, are important tools for quantifying the rate of permafrost degradation on a large scale, and its evolution in response to climate change scenarios.

However, there are large uncertainties in soil thermal dynamics in land surface models (e.g. Koven et al., 2013), and these uncertainties also impact predictions of carbon-cycle feedbacks on climate. To quantify and reduce the uncertainty of modeled soil temperature ($T_s$), the driving factors of $T_s$ trends need to be investigated. Besides the uncertainty in model parameterization and structure, the gridded climate forcing for offline land surface models over high latitude regions have large uncertainty (e.g. Troy and Wood, 2009; Rawlins et al., 2010). It is also important to distinguish the uncertainty caused by assigned parameter values and model structure from the uncertainty attributable to uncertain climate forcing data.

In this study, nine process-based models that participated in the Permafrost Carbon Network (PCN, www.permafrostcarbon.org) were used (1) to compare trends of simulated $T_s$ at different depths over the boreal permafrost regions during the past four decades and to assess the uncertainty of modeled $T_s$ trends across models; (2) to identify which factors drive trends of permafrost $T_s$; and (3) to quantify the sensitivity of changes in near-surface permafrost area to warming.

2 Methods

2.1 Models and simulations

The nine land surface models that were used for simulating $T_s$ in permafrost regions organized by Permafrost Carbon Network (PCN, www.permafrostcarbon.org) are listed in Table 1.
simulate $T_s$, but models have different soil depths, snow parameterizations, and soil thermal conductivity (Table 1). Three models (CLM, ISBA, UW-VIC) explicitly considered organic soil insulation and six-seven models explicitly considered the effect of water in soil on phase change. All models explicitly considered snow insulation but with different snow layers. The soil thermal conductivity depends on soil moisture in all models. More details can be found in Rawlins et al. (2015) and McGuire et al. (in preparation). We defined the Northern Hemisphere permafrost spatial domain as the definition in Figure 1, and the analysis considers three permafrost regions, Boreal North America (BONA), Boreal Europe (BOEU), Boreal Asia (BOAS) (Figure 1; Brown et al., 1998). We did not include the Tibetan plateau because not all the models covered this region. Hereafter, the term “boreal regions” is used in the following for the sum of the three sub-regions BONA, BOEU and BOAS in Figure 1.

Following the simulation protocol of the PCN project, nine land surface models performed historical simulations from 1960 to 2000, using different forcing data sets (Table 1). 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. Modelers used different forcing datasets for climate and other model boundary conditions (Table 1). Climate forcing data chosen by each group are presented in Table 1, and the differences in the trend of $T_a$, precipitation, and radiative forcing are summarized in Figure S1 and S2. How differences between these drivers are related to differences of the modeled $T_s$ is discussed in the Results and Discussion section.

To separate the contributions of the trends of four forcing variables: ($T_a$, atmospheric...
CO₂, precipitation, and LWDR) on permafrost thermal dynamics and carbon stocks, six out of the nine models conducted factorial simulations (R01-R04). The ORCHIDEE and JULES performed two additional simulations (R05-R06) to isolate the contribution of LWDR on Ts trends (Table 2 and 3). In the reference simulation R01, all drivers varied at the same time. In R02 Ts was detrended; in R03 atmospheric CO₂ was set constant to the observed 1960 level of 316 ppmv; In R04 both Ts and precipitation were detrended; in R05 Ts and LWDR were detrended; in R06 Ts, precipitation and LWDR were detrended. Differences between two simulations were used to separate the controlling effect of each driver on Ts, and some of their non-linear interactions.

2.2 Analysis

Modeled monthly Ts at 5, 20, 50, 100, 200 and 300 cm depths in every grid cell of each model were calculated by linear interpolation of Ts between the central depths of two adjacent layers. If the maximum soil depth of a model is larger than 300 cm (which is the case for seven models, except CoLM and JULES), then 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 to layers deeper than 300 cm (the maximum soil depth of each model is shown in Table 1). For each boreal sub-regions BONA, BOEU, BOAS (Figure 1) Ts was first averaged over all grid cells and the trend of regional mean Ts (denoted $\tilde{T}_s$) was calculated from a linear regression. The statistical significance of $\tilde{T}_s$ is evaluated by a t-test.

To estimate the uncertainty of $\tilde{T}_s$ caused by differences in the trend of each climate input variable, we regressed $\tilde{T}_s$ against the trends of Ts, precipitation and short-wave downward radiation (SWDR) and LWDR, respectively, using the output of R01. The
uncertainty of $\dot{J}_s$ attributed to each forcing variable was defined as the resulting range of $\dot{J}_s$ associated to different trends in each forcing variable in the models. To achieve this aim, we regressed $\dot{J}_s$ against forcing variable across the models, and the uncertainty of $\dot{J}_s$ resulting from uncertain forcing data was calculated as the range of $\dot{J}_s$ from the maximum and minimum values of forcing data in the regression equation. Then we define the uncertainty attributed to model structure, which reflects the differences in model parameterizations and parameter values, as the uncertainty of $\dot{J}_s$ assuming all models were using the same climate forcing data.

Here, we defined near-surface permafrost as in previous studies (e.g. Schneider von Deimling et al., 2012): near-surface permafrost is defined as where the maximum seasonal thaw depth (i.e., the active layer thickness, ALT) is less than 3 m. The total near-surface permafrost area (NSPA) is the sum of the areas of grid cells that fulfill this condition.

We used monthly LWDR data from CRUNCEP v5.2 (http://dods.extra.cea.fr/data/p529viov/cruncep) and WATCH (Weedon et al., 2011) with a spatial resolution of 0.5° by 0.5° during the period 1960-2000 to derive the trend of LWDR. The CRUNCEP LWDR dataset was derived from CRU TS3.21 and NCEP reanalysis meteorology, and ancillary data sets (e.g. Wei et al., 2012, 2014). The WATCH LWDR dataset was derived from ERA-40 reanalysis (Weedon et al., 2011). Because there is no long-term large scale LWDR observation product available, we did an experiment using LWDR from CRUNCEP and WATCH data to estimate the loss of permafrost area during the period 1960-2000 by an empirical relationship found between the loss of permafrost area and LWDR.
3 Results and Discussion

3.1 Trend in upper-layer soil temperature over boreal regions

The simulated values of $\dot{\text{Ts}}$ at 20 cm depth averaged over boreal regions range from $0.010 \pm 0.003 \, ^\circ\text{C yr}^{-1}$ (CoLM) to $0.031 \pm 0.005 \, ^\circ\text{C yr}^{-1}$ (UVic) during the period 1960-2000 (Figure 2). Figure 3 shows $\dot{\text{Ts}}$ at 20 cm for BONA, BOEU and BOAS regions. Six out of the nine models show the largest $\dot{\text{Ts}}$ at 20 cm in BOAS, followed by BONA and the smallest $\dot{\text{Ts}}$ at 20 cm in BOEU. The other three models (CoLM, JULES and UW-VIC) show the smallest $\dot{\text{Ts}}$ at 20 cm in BOAS. Among the six models with smaller $\dot{\text{Ts}}$ at 20 cm in BOEU, we found that $\dot{\text{Ts}}$ at 20 cm in BOEU is significantly lower than in BOAS and in BONA ($P<0.001$, two sample t-test). This is also shown in the spatial distribution of $\dot{\text{Ts}}$ at 20 cm (Figure 4). For example, in northern Siberia, $\text{Ts}$ at 20 cm increased by more than 0.02 °C yr$^{-1}$ in five out of the nine models (ISBA, LPJ-GUESS, MICRO-ESM, ORCHIDEE and UVic) but decreased in two models (CoLM and JULES). All models show an increase of $\text{Ts}$ at 20 cm in northern BONA, but this increase is of different magnitude between models (Figure 4). Six models show significant $\dot{\text{Ts}}$ at 20 cm over northern and western Siberia, but all models show non-significant $\dot{\text{Ts}}$ at 20 cm over northern BOEU (Figure 4).

3.2 Attenuation of the trend in soil temperature with soil depth

The trend of $\text{Ts}$ at different soil depths in soil layers over 0 to 3 m and soil depth is shown in Figure 5 for each model. Based on ground soil temperature observation, annual $\text{Ts}$ at...
1.6 m increased by 0.02-0.03 °C yr⁻¹ from 1960s to 2000s in Russia (Park et al., 2014). The simulated trends of Tₛ at 1.6 m over BOAS in most models are within this range (Figure S3).

Two models (CoLM and JULES) show vertically quasi-uniform $\dot{T}_s$ over the upper 3 m of soil, probably because of too quick soil thermal equilibrium in these two models. The seven other models show decreasing values of $\dot{T}_s$ with increasing soil depth, but the vertical gradient of $\dot{T}_s$ varies among them (Figure 5a). UW-VIC has the largest negative vertical gradient of $\dot{T}_s$ (-0.0052 ± 0.0001 °C yr⁻¹ m⁻¹), followed by ISBA, MICRO-ESM, ORCHIDEE and UVic (~0.0030 ± 0.0003 °C yr⁻¹ m⁻¹) and by near-zero vertical gradient of $\dot{T}_s$ in CLM (-0.0009 ± 0.0003 °C yr⁻¹ m⁻¹) and in LPJ-GUESS (-0.0014 ± 0.0000 °C yr⁻¹ m⁻¹).

Figure 5b shows the trend of Tₛ in all soil layers over boreal regions. CLM and UVic show an increase of Tₛ even at depths deeper than 40 m, but Tₛ exhibited no changes deeper than 22 m in ORCHIDEE (Figure 5b). Tₛ increased in the deepest layer of ISBA (12 m) and MIROC-ESM (14 m), and the depth at which Tₛ exhibited no changes could not be deduced from these two models. UW-VIC shows a negative trend of Tₛ (i.e. cooling) at depths deeper than 2.5 m. The trends of Tₛ in all soil layers over BONA, BOEU and BOAS regions decrease in magnitude with increasing soil depth, but show different vertical gradients. In Figure S3, vertical gradient of $\dot{T}_s$ are is shown to be larger in BONA and BOAS than that in BOEU for most models. Figure 6 shows the spatial distribution of the difference in $\dot{T}_s$ at depths between 0.2 m and 3 m. $\dot{T}_s$ at 0.2 m is larger than that at 3 m over most regions in BONA, BOEU and BOAS in seven out of the nine models, exceptions being JULES and CoLM. Generally, boreholes records show that mean annual soil temperature at depths between 10 m and 30 m
have increased during the last three decades over the circumpolar northern permafrost regions (Osterkamp, 2003; Romanovsky et al., 2010; Smith et al., 2005, 2012; Vaughan et al., 2013). In Alaska, T at 20 m from boreholes even increased by ~1 °C between the early 1980s and 2001 (Osterkamp, 2003). The observed value of $J_z$ at one of Alert (BH3) boreholes is of ~0.04 °C yr⁻¹ at ~2.5 m depth and nearly zero at ~27 m depth during the period 1979-2004 (see Figure 9 in Smith et al., 2012). Some boreholes (BH1 and BH2) at Alert however still indicated a small warming during the period 1979-2008 (Smith et al., 2012) at 37m. 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 $T_s$ projections over the next century (e.g. Alexeev et al., 2007). Compared to the increased ground temperature observed to increase at depths deeper than 20 m in boreholes during the past three decades (Vaughan et al., 2013), most models that do not have deeper soil depth seem to underestimate the penetration of heat into deep soil layers (Figure 5b). Note that this comparison may be biased because of different periods and climate records between sites and model grid cells. It is also recommended that simulations at site level using in-situ local climate forcing can be compared with temperature profiles of boreholes (Smith et al., 2012) to evaluate why models underestimate the warming of $T_s$ at deeper depths.

3.3 Drivers of trend in soil temperature

We used the sensitivity runs (R02-R06) compared with the reference simulation with all drivers varying together (R01) to separate the effects of $T_a$, CO₂, precipitation, and LWDR on $J_z$ during 1960-2000 (Table 3). Seven of the nine models only provided results from R02,
R03 and R04. Except for JULES, all the models show a positive response of $T_s$ to increasing $T_a$, but with different sensitivities (Table 3). The fraction of the trend of $T_s$ explained by air temperature increase alone (R01 – R02) is nearly 100% in CLM, ISBA and more than 100% in UW-VIC, against only 34%, 56% and 67% in ORCHIDEE, UVic and LPJ-GUESS. This indicates the importance of increasing $T_a$ on the trend of $T_s$, and is consistent with observations. Based on 30 climate stations observations in Canada during the period 1958-2008, $T_s$ at 10 cm significantly and positively correlates with $T_s$ at most sites (>90%) in spring, but at fewer sites (<30%) in winter (Qian et al., 2011). For winter $T_s$, the winter snow depth was found to have significant and positive correlation with $T_s$ in shallow soil layers (e.g. Zhang et al., 2001; Qian et al., 2011). Recent increases in $T_s$ also explain the trend of $T_s$ 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 $T_s$ is a good indicator for the trend of deep permafrost ground temperature with some time lag (Romanovsky et al., 2007). For the modeled $T_s$ in land surface models, the effects of $T_s$ on $T_s$ depend on surface energy balance and ground heat flux into soil; i.e. the extent of coupled $T_s$ on $T_s$ relates to the surface properties such as snow, organic soil horizons and roughness etc. in the models. The different relative contributions of the trend of $T_s$ to the trend of $T_s$ in these models maybe mainly result from the different model parameterization and structures, as the trends of $T_s$ (~0.03 °C yr-1) in the climate forcing do not have a large spread (Figure 7).

The increase of atmospheric CO2 concentration has almost no effect on the increase of $T_s$ in most models (-5% to +4% of increase of $T_s$, Table 3). This is expected since CO2 has no direct effect on $T_s$ apart from its impact on climate. The only indirect effect of rising CO2 on $T_s$ trends could result from feedbacks between plant productivity driven by rising CO2, soil
carbon changes and soil thermal properties. For instance, if models include heat production from microbial decomposition of soil organic carbon (Khvorostyanov et al., 2008) or if changes in soil organic carbon from the balance of NPP input and decomposition, these could impact the soil temperature directly or the profile of soil heat conductivity and capacity. In that case, the expected response is that a CO₂ driven increase of productivity will increase soil organic carbon, which will enhance the insulation effect of soil organic carbon in the soil and act to lower the trend of T_s (Lawrence et al., 2008; Lawrence and Slater, 2008; Koven et al., 2009). Further, complex changes in the surface energy balance from changes in evapotranspiration under higher CO₂ concentrations can influence soil moisture content and affect T_s trends (e.g. Field et al., 1995). Most models do not have a feedback between soil organic carbon dynamics and soil thermal properties, and the increase in soil organic carbon due to rising CO₂ is relatively small in the models compared to the initial soil organic carbon storage (< 0.1%). The changes in evapotranspiration because of increasing CO₂ are also relatively small (-3% to +1%). Therefore, the increased CO₂ concentration has a very small effect on $\dot{j}_s$ from 1960 to 2000.

Precipitation shows an increase in BONA and BOEU and a decrease since 1960 in BOAS in the climate forcing used by most models (Figure S1b). None of the regional trends of boreal precipitation are significant (P>0.05; except for the UW-VIC and JULES drivers). Changes in precipitation alone (R02 – R04) are found to cause a negative trend of T_s in CLM, JULES and UW-VIC, no effects in LPJ-GUESS and UVic, and a positive trend in ISBA and ORCHIDEE (Table 3). Increasing winter snow-fall can enhance T_s in winter through snow insulation effect (e.g. Smith et al., 2010; Koven et al., 2013). All models in this study indeed show higher winter T_s where winter snow depth became thicker and deeper, but with different magnitudes of snow insulation effects across the models. The snow insulation effects are
smaller in ISBA, LPJ-GUESS and UVic than that in the other models. A decrease in snowfall could contribute a negative trend of $T_s$ in CLM and JULES, and conversely, an increase in snowfall could enhance $T_s$ in ORCHIDEE (Figure S4; Table 3). In addition, increased rainfall in summer can cause an increase in evapotranspiration during the growing seasons, which could depress the increase of $T_s$. 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 $T_s$ should be studied in the future. Because models did not run simulations where precipitation is detrended only during a specific season, the effects of snowfall trends and growing season precipitation trends that may oppose each other, and cannot be separated in this analysis.

LWDR significantly increased since 1960 in all models yet with different trends in the forcing data used by each modeling group (0.058 – 0.200 W m$^{-2}$ yr$^{-1}$) (Figure S2a) reflecting the different climate forcing data input. LWDR forcing is mainly from two reanalysis datasets (ERA and NCEP) with corrections (e.g. Weedon et al., 2011; http://dods.extra.cea.fr/data/p529viov/cruncep). ORCHIDEE and JULES performed the simulation R05 with detrended LWDR. The results of R01 – R02 – R05 allowing to attribute $\dot{J}_s$ to trends of LWDR, indicate that the increase of LWDR explains 56% and 31% of the trend of $T_s$ since 1960 in ORCHIDEE and JULES, respectively. Increased LWDR provides additional energy to the surface, and dominates the atmosphere-to-soil energy flux in winter over boreal regions when shortwave radiation is small. Even in summer, LWDR contributes ~60% of total downward radiation (SWDR+LWDR) over boreal regions in CRUNCEP. An increase of LWDR with time thus increases the surface energy input, which
accelerates the warming of $T_s$ in case the extra energy is not dissipated by an increase of sensible and latent heat flux. The contribution of changes in LWDR, $T_a$ and other factors on all components of the surface energy budget and on $T_s$ could be further studied by testing models against observations from eddy-flux towers located in permafrost soils.

3.4 Uncertainty of modeled soil temperature trends

The uncertainty of modeled $\dot{J}_s$ at 20 cm is large, as given by the spread of model results (0.010 °C yr$^{-1}$ - 0.031 °C yr$^{-1}$). The uncertainty of $\dot{J}_s$ across the models can be conceptually decomposed into two components, a forcing uncertainty (FU) reflecting how different climate input data used by each modeling group contribute to the spread of $\dot{J}_s$ (Table 1), and a structural uncertainty (SU) related to uncertain parameter values and different equations and parameterizations of processes in models. Since $T_a$ and LWDR are the two main drivers of the increase of $T_s$ in most of the models (Section 3.3), we regressed $\dot{J}_s$ during 1960-2000 against the trends of $T_a$ and LWDR, in order to estimate the FU. We then estimated SU from the uncertainty of parameters in the regression equation for a normalized same climate forcing across all models.

We found no significant correlation between $\dot{J}_a$ and $\dot{J}_s$ over boreal regions or sub-regions across the nine models (Figure 7 and Figure S5), indicating that a bias of $\dot{J}_a$ forcing is not simply associated with the bias of $\dot{J}_s$ in a particular model compared to the others. We also found that trends of SWDR and precipitation do not significantly explain differences in $\dot{J}_s$ at 20 cm across the models (P>0.05; 21% and 19% explanation of
differences in \( \dot{I}_s \) at 20 cm for trends of SWDR and precipitation respectively; Figure S6).

The correlations between trends in winter snowfall and trends of annual or winter \( T_s \) at 20 cm are not significant (P>0.05) across the models for boreal regions or sub-regions. However, the trend of LWDR (\( LW\dot{D}R \)) can explain 61% of the differences in \( \dot{I}_s \) at 20 cm across the models (Figure 8). This result indicates that, across the model ensemble, differences of \( \dot{I}_s \) at 20 cm between models are positively correlated (R=0.78, P=0.037) with differences of \( LW\dot{D}R \) used by the different modeling groups. \( \dot{I}_s \) at 1 m also significantly correlated with \( LW\dot{D}R \) (R=0.79, P=0.034) across the models. The values of \( LW\dot{D}R \) used by different models averaged over permafrost regions, range from 0.058 W m\(^{-2}\) yr\(^{-1}\) to 0.200 W m\(^{-2}\) yr\(^{-1}\), statistically explaining a range of simulated \( \dot{I}_s \) at 20 cm of 0.021 ± 0.005 °C yr\(^{-1}\) (solid blue arrow in Figure 8). This \( \dot{I}_s \) range defines the FU (the range of \( \dot{I}_s \) to \( LW\dot{D}R \) from 0.058 W m\(^{-2}\) yr\(^{-1}\) to 0.200 W m\(^{-2}\) yr\(^{-1}\) based on the linear regression of Figure 8). We also used multiple linear regression between \( \dot{I}_s \) at 20 cm depth and \( \dot{I}_a \), \( LW\dot{D}R \) as independent variables across the models, to derive an estimation of the FU on \( \dot{I}_s \) of 0.021 ± 0.008 °C yr\(^{-1}\) (the deviation was derived from the uncertainty of regression coefficients in the multiple linear regression). However, the uncertainty of the linear regression of \( \dot{I}_s \) at 20 cm by \( LW\dot{D}R \) shows that if all the models used the same climate forcing data, the SU would be of 0.012 ± 0.001 °C yr\(^{-1}\) (solid orange arrow in Figure 8). If all models used LWDR from CRUNCEP or WATCH, then applying the trend of annual LWDR (0.087 ± 0.023 W m\(^{-2}\) yr\(^{-1}\) from CRUNCEP and 0.187 ± 0.028 W m\(^{-2}\) yr\(^{-1}\) from WATCH) during the period 1960-2000 as an emerging observation constraint.
empirical relationship in Figure 8, the posterior range is reduced compared with the prior range (black curve in right panel of Figure 8). Overall, the total uncertainty range of \( \dot{\chi}_s \) at 20 cm (~0.02 °C yr\(^{-1}\), defined as the spread of \( \dot{\chi}_s \) at 20 cm across the models) can be broken down into FU (0.021 ± 0.008 °C yr\(^{-1}\)) and SU (0.012 ± 0.001 °C yr\(^{-1}\)). Since FU and SU are not independent, the total uncertainty of \( \dot{\chi}_s \) at 20 cm is not the sum of FU and SU.

Further, we found that correlation coefficients between trends of summer \( T_s \) at 20 cm and at 1 m and summer LWDR over boreal regions are statistically significant (P<0.05) (Figure S7). This is also found for winter (November to March) \( T_s \) at 20 cm and 1 m (Figure S8). Trends of summer and winter \( T_s \) at 20 cm or 1 m are not significantly correlated with other climate drivers than LWDR (snowfall, rainfall, \( T_s \) and SWDR) across the models (P>0.05).

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 \( \dot{\chi}_s \). 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 and evaluated in land surface models to improve the prediction of future permafrost feedbacks (e.g. van Huissteden et al., 2013; Lee et al., 2014).

3.5 Emerging constraint on how much near-surface permafrost has disappeared.

All the nine models provide estimates of the total boreal NSPA during 1960-2000 estimated by the nine models that ranges from 6.8 million km$^2$ in (CoLM) to 19.7 million km$^2$ in (ORCHIDEE), which is three times as large as the former. The average of total NSPA in the nine models ensemble (12.5 million km$^2$) is smaller than the estimate from the International Permafrost Association (IPA) map (16.2 million km$^2$; Brown et al, 1998; Slater and Lawrence and Slater et al., 2013). A statistic model based on relationships between air temperature and permafrost shows that permafrost extent over Northern Hemisphere was also estimated in the range 12.9 - 17.8 million km$^2$ (Gruber, 2012), and six out of the nine models are within this range. Eight out of the nine models show a significant decrease in NSPA with climate warming during 1960-2000 (except UW-VIC). The loss rate of NSPA is found to vary by a factor of 13 across the nine models, varying from $-4 \times 10^3$ km$^2$ yr$^{-1}$ in MIROC-ESM to $-50 \times 10^3$ km$^2$ yr$^{-1}$ in JULES (Figure 9a). The average of loss rate of NSPA across the models (-23
± 23 ×10³ km² yr⁻¹) is smaller than in the previous estimations of Burke et al. (2013) and Slater and Lawrence (2013). For example, the loss rate of NSPA was estimated at -81×10³ – -55×10³ km² yr⁻¹ during the period 1967-2000 by JULES offline simulations with different climate forcing datasets (Burke et al., 2013). The ranges of loss rate of NSPA in BONA, BOEU and BOAS across the models are -16.6×10³ – 2.2×10³ km² yr⁻¹, -4.0×10³ – 0.0×10³ km² yr⁻¹ and -34.2×10³ – -1.1×10³ km² yr⁻¹, respectively (Figure 9). This is consistent with the observed permafrost degradation (decrease in thickness) in these regions (Vaughan et al., 2013).

The retreat rate of NSPA is not correlated significantly with the initial NSPA of each model (R=-0.30, P=0.438), implying that the initial state of models is less important than their response to climate change in determining NSPA loss rates. On the contrary to the small effect of initial NSPA, the trend of summer Tₛ at 1 m is found to be strongly correlated with NSPA loss rates across the models of the ensemble. Figure 9 shows that the trend of summer Tₛ at 1 m can explain 73% of the differences in NSPA loss rates between models. The sensitivity of NSPA loss rate to summer $\dot{T}_a$ at 1 m is estimated to be -2.80 ± 0.67 million km² °C⁻¹, based on the linear regression between the loss rate of NSPA and the trend of summer Tₛ at 1 m across the nine models (Figure 9). For the BONA, BOEU and BOAS sub-regions, the sensitivities of NSPA loss rate to summer $\dot{T}_a$ at 1 m are -0.74 ± 0.10 million km² °C⁻¹, -0.09 ± 0.03 million km² °C⁻¹ and -1.74 ± 0.59 million km² °C⁻¹, respectively (Figure 9). The sensitivity of future total NSPA changes to Tₛ over Pan-Arctic regions was estimated to be -1.67 ± 0.7 million km² °C⁻¹, ranging from 0.2 million km² °C⁻¹ to 3.5 million km² °C⁻¹ in CMIP5 model ensembles (Slater and Lawrence, 2013; Koven et al., 2013). The average of trends in summer Tₛ at 1 m is only 70% (43%-100%) of $\dot{T}_a$ in the nine models, so that
the sensitivity of total NSPA to $T_s$ over boreal regions in the nine models is about $-2.00 \pm 0.47$ million km$^2$ °C$^{-1}$, which is larger than that from CMIP5 model ensemble, but comparable within the uncertainties of each estimate (Slater and Lawrence, 2013). Six out of the nine models of this study were also used as land surface schemes of the coupled CMIP5 models, but possibly for different versions.

A mean positive trend of summer LWDR of $0.073 \pm 0.030$ W m$^{-2}$ yr$^{-1}$ and $0.210 \pm 0.027$ W m$^{-2}$ yr$^{-1}$ over boreal regions from 1960 to 2000 are derived from the CRUNCEP and WATCH datasets respectively described in Section 2. We applied this trend of LWDR to an emerging constraint on summer $T_s$ trends from the relationship between the trend of summer LWDR and the trend of summer $T_s$ at 1 m (Figure S7). This approach constrains the trend of summer $T_s$ to $0.014 \pm 0.004$ °C yr$^{-1}$ with CRUNCEP and to $0.027 \pm 0.004$ °C yr$^{-1}$ with WATCH. The uncertainty is reduced by 50% from the prior range including different models and different forcings. A total NSPA loss rate of $39 \pm 14 \times 10^3$ km$^2$ yr$^{-1}$ can be constrained by multiplying the sensitivity of total NSPA loss rate to summer $j_s$ at 1 m ($-2.80 \pm 0.67$ million km$^2$ °C$^{-1}$) by the trend of $T_s$ at 1 m itself empirically estimated by $LWDR$ during 1960-2000 from CRUNCEP ($0.014 \pm 0.004$ °C yr$^{-1}$). The constrained loss rate of NSPA over BONA, BOEU and BOAS based upon the CRUNCEP $LWDR$ from 1960 to 2000 are $11 \pm 5 \times 10^3$ km$^2$ yr$^{-1}$, $1 \pm 1 \times 10^3$ km$^2$ yr$^{-1}$ and $25 \pm 11 \times 10^3$ km$^2$ yr$^{-1}$, respectively. Similarly, if WATCH $LWDR$ is used to constrain NSPA loss rate, the total NSPA loss rate is $75 \pm 14 \times 10^3$ km$^2$ yr$^{-1}$, and loss rate of NSPA over BONA, BOEU and BOAS are estimated to be $28 \pm 10 \times 10^3$ km$^2$ yr$^{-1}$, $2 \pm 1 \times 10^3$ km$^2$ yr$^{-1}$ and $39 \pm 19 \times 10^3$ km$^2$ yr$^{-1}$, respectively. The southern boundary of the discontinuous permafrost zone has been observed to shift northward during the recent past decades is also observed (Vaughan et al., 2013), which is generally consistent
with the simulations reported in this study. The larger warming rate and higher sensitivity of NSPA loss to $T_s$ over BOAS could explain the reason for significant degradation of permafrost over BOAS than the other boreal regions (Vaughan et al., 2013). The larger permafrost degradation rate in BOAS than that in BONA may have larger effects on changes in vegetation distribution and growth, and permafrost carbon in these two regions, and can be quantified in future studies. Obviously, there is a large difference in constrained NSPA between CRUNCEP and WATCH. In the future, long-term climate reanalysis including radiation evaluated against sites with long-term radiation measurements (http://www.geba.ethz.ch) would be extremely useful for land surface models to provide improved estimate of NSPA.

4 Conclusions

In this study, trends of $T_s$ over boreal regions from nine process-based models were analyzed for the past 40 years. All models produce a warming of $T_s$, but the trends of $T_s$ at 20 cm depth range from $0.010 \pm 0.003$ °C yr$^{-1}$ (CoLM) to $0.031 \pm 0.005$ °C yr$^{-1}$ (UVic) during 1960-2000. Most models show a smaller increase of $T_s$ with deeper depth. $T_a$ and LWDR are found to be the predominant drivers of the increase in $T_s$ averaged across large spatial scales. The relative contribution of $T_a$ and LWDR trends to the increase of $T_s$ is however different across the models. Note that the relative contribution of LWDR is based on only two models in this study, and this needs further investigation. The total uncertainty of the trend of $T_s$ at 20 cm is decomposed into the uncertainty contributed by uncertain climate forcing datasets ($0.021 \pm 0.008$ °C yr$^{-1}$) and the uncertainty reflecting model structure ($0.012 \pm 0.001$ °C yr$^{-1}$). The NSPA loss rate is significantly correlated among the model results with the simulated trend of $T_s$ at 1 m, with a linear sensitivity of total NSPA loss rate to summer $\dot{J}_s$ at 1 m of $-2.80 \pm 0.67$ million km$^2$ °C$^{-1}$. Based on LWDR from CRUNCEP and WATCH data, the total
NSPA decrease is estimated to be $39 \pm 14 \times 10^3 \text{ km}^2 \text{ yr}^{-1} - 75 \pm 14 \times 10^3 \text{ km}^2 \text{ yr}^{-1}$ from 1960 to 2000. The constraint method used in this study could be applied to estimate historical and future permafrost degradation rate, and further to quantify the permafrost carbon loss by permafrost carbon distribution map.

Given that meteorological stations are sparse in the cold permafrost regions, especially in Siberia and other unpopulated land in the north, the gridded climate products over high-latitude regions have a large uncertainty as well (Mitchell and Jones, 2005; Rawlins et al., 2010; Weedon et al., 2011). This large uncertainty could propagate into simulated permafrost dynamics and feedbacks. More sites are needed in high-latitude regions for reducing the climate uncertainty. Future model inter-comparisons on permafrost dynamics should investigate the full uncertainty by conducting simulations for multiple climate forcing data sets. Since the beginning of the satellite era, microwave emissivity data related to land surface temperature has become increasingly available (e.g. Smith et al., 2004). These images could be used to independently evaluate soil surface temperature in models on a large scale, although they have their own uncertainties. In addition, many complex processes affect permafrost thermal dynamics in the models, such as soil organic insulation effects, snow insulation effects, soil freeze-thaw etc., it is valuable to evaluate the uncertainty of each process effects on soil thermal dynamics simulations based on site measurements. This could be helpful for reducing permafrost simulation uncertainty.

Acknowledgments
This study has been supported by the PAGE21 project, funded by the European Commission FP7-ENV-2011 (grant agreement NO. 282700) and has been developed as part
of the modeling integration team of the Permafrost Carbon Network (PCN, www.permafrostcarbon.org) funded by the National Science Foundation. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. Bertrand Decharme and Christine Delire were supported by the French Agence Nationale de la Recherche under agreement ANR-10-CEPL-012-03.

References
Bonfils, C. J. W., Phillips, T. J., Lawrence, D. M., Cameron-Smith, P., Riley, W. J., and Subin,


Decharme, B., Martin, E., and Faroux, S.: Reconciling soil thermal and hydrological lower


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.,
and Zhang, T. J.: Analysis of the Arctic System for Freshwater Cycle Intensification:

Rawlins, M. A., McGuire, A. D., Kimball, J. K., Dass, P., Lawrence, D., Burke, E., Chen, X.,
Delire, C., Koven, C., MacDougall, A., Peng, S., Rinke, A., Saito, K., Zhang, W.,
Alkama, R., J. Bohn, T., Ciais, P., Decharm, B., Gouttevin, I., Hajima, T., Ji, D., Krinner,
G., Lettenmaier, D. P., Miller, P., Moore, J. C., Smith, B., and Sueyoshi, T.: Assessment
of model estimates of land-atmosphere CO2 exchange across Northern Eurasia,

and recent changes in air and permafrost temperatures in eastern Siberia, Global and

Romanovsky, V. E., Smith, S. L., and Christiansen, H. H.: Permafrost thermal state in the
polar Northern Hemisphere during the international polar year 2007–2009: a synthesis,

Schädel, C., Schuur, E. A. G., Bracho, R., Elberling, B., Knoblauch, C., Lee, H., Luo, Y.,
Shaver, G. R., and Turetsky, M. R.: Circumpolar assessment of permafrost C quality and
its vulnerability over time using long-term incubation data, Global Change Biology, 20,

Schaefer, K., Zhang, T., Bruhwiler, L., and Barrett, A. P.: Amount and timing of permafrost
carbon release in response to climate warming, Tellus B, 63, 165-180,

30


Table 1. Soil depth for soil thermal dynamics and climate forcing used in each model.

<table>
<thead>
<tr>
<th>Model</th>
<th>Soil depth (m)</th>
<th>Soil discretization</th>
<th>Bottom boundary geothermal heat flux (mW m^{-2})</th>
<th>Climate forcing (Reference)</th>
<th>Model reference</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLM</td>
<td>45.1</td>
<td>30</td>
<td>0</td>
<td>CRUNCEP v4 (<a href="http://dods.extra.cea.fr/">http://dods.extra.cea.fr/</a>)</td>
<td>Oleson et al., 2013</td>
<td></td>
</tr>
<tr>
<td>CoLM</td>
<td>3.4</td>
<td>10</td>
<td>0</td>
<td>Princeton (Sheffield et al., 2006)</td>
<td>Dai et al., 2003, 2004</td>
<td></td>
</tr>
<tr>
<td>ISBA</td>
<td>12.0</td>
<td>14</td>
<td>0</td>
<td>WATCH (1901-1978)</td>
<td>Decharme et al., 2011, 2013</td>
<td></td>
</tr>
<tr>
<td>JULES</td>
<td>20.8</td>
<td>16</td>
<td>0</td>
<td>WATCH (1901-2001)</td>
<td>Best et al., 2011; Clark et al., 2011</td>
<td></td>
</tr>
<tr>
<td>LPJ-GUESS</td>
<td>3.0</td>
<td>8</td>
<td>0</td>
<td>CRU TS 3.1 (Harris et al., 2014)</td>
<td>Smith et al., 2001; McGuire et al., 2012</td>
<td></td>
</tr>
<tr>
<td>MIROC-ESM</td>
<td>14.0</td>
<td>6</td>
<td>0</td>
<td>CMIP5 Drivers (Watanabe et al., 2011)</td>
<td>Watanabe et al., 2011</td>
<td></td>
</tr>
<tr>
<td>ORCHIDEE</td>
<td>47.4</td>
<td>32</td>
<td>28</td>
<td>WFDEI (1978-2009)</td>
<td>Koven et al., 2011; Gouttevin et al., 2012</td>
<td></td>
</tr>
<tr>
<td>UVic</td>
<td>250.3</td>
<td>14</td>
<td>0</td>
<td>CRUNCEP v4 (<a href="http://dods.extra.cea.fr/">http://dods.extra.cea.fr/</a>)</td>
<td>Avis et al., 2011; MacDougall et al., 2012</td>
<td></td>
</tr>
<tr>
<td>UW-VIC</td>
<td>25.0</td>
<td>25</td>
<td>0</td>
<td></td>
<td>Bohn et al., 2013</td>
<td></td>
</tr>
</tbody>
</table>

Soil temperature in the top 3 meter is based on another 6 parameter bottom layer condition. Surface shortwave downward and cloudiness data set; No longwave downward radiation and surface shortwave and longwave downward radiation is applied.
Table 2. Description of simulations used in this study.

<table>
<thead>
<tr>
<th>Simulation ID</th>
<th>Climate</th>
<th>CO₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>R01</td>
<td>variable</td>
<td>variable</td>
</tr>
<tr>
<td>R02</td>
<td>variable, but with detrended $T_a$</td>
<td>variable</td>
</tr>
<tr>
<td>R03</td>
<td>variable</td>
<td>constant in the year of 1960</td>
</tr>
<tr>
<td>R04</td>
<td>variable, but with detrended $T_a$ and precipitation</td>
<td>variable</td>
</tr>
<tr>
<td>R05</td>
<td>variable, but with detrended $T_a$ and LWDR</td>
<td>variable</td>
</tr>
<tr>
<td>R06</td>
<td>variable, but with detrended $T_a$, precipitation and LWDR</td>
<td>variable</td>
</tr>
</tbody>
</table>
Table 3. The trends of annual air temperature (T<sub>a</sub>), precipitation and longwave downward radiation (LWDR) in the second to fourth columns. The fifth column shows the trends of annual T<sub>s</sub> at 20 cm in the reference simulation (R01). The last four columns show the contributions of drivers (T<sub>a</sub>, precipitation, CO<sub>2</sub> and LWDR) on the trend of T<sub>s</sub> as mentioned in Methods section. The relative contributions (divided by the trend of T<sub>s</sub> in Ref) are shown in the parentheses. The bold font indicates statistically significant (P<0.05).

<table>
<thead>
<tr>
<th>Model</th>
<th>Trend of T&lt;sub&gt;a&lt;/sub&gt; (°C yr&lt;sup&gt;-1&lt;/sup&gt;)</th>
<th>Trend of precipitation (mm yr&lt;sup&gt;-1&lt;/sup&gt;)</th>
<th>Trend of LWDR (W m&lt;sup&gt;-2&lt;/sup&gt; yr&lt;sup&gt;-1&lt;/sup&gt;)</th>
<th>Simulated Trend of T&lt;sub&gt;s&lt;/sub&gt; (R01) (°C yr&lt;sup&gt;-1&lt;/sup&gt;)</th>
<th>Contribution from T&lt;sub&gt;a&lt;/sub&gt; (R01-R02) (°C yr&lt;sup&gt;-1&lt;/sup&gt;)</th>
<th>Contribution from precipitation (R02-R04) (°C yr&lt;sup&gt;-1&lt;/sup&gt;)</th>
<th>Contribution from CO&lt;sub&gt;2&lt;/sub&gt; (R01-R03) (°C yr&lt;sup&gt;-1&lt;/sup&gt;)</th>
<th>Contribution from LWDR (R02-R05) (°C yr&lt;sup&gt;-1&lt;/sup&gt;)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLM</td>
<td>0.031</td>
<td>0.13</td>
<td>0.114</td>
<td>0.015(92%)</td>
<td>-0.002(-12%)</td>
<td>0.001(4%)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CoLM</td>
<td>0.031</td>
<td>-0.05</td>
<td>0.058</td>
<td>0.010(100%)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ISBA</td>
<td>0.033</td>
<td>-0.17</td>
<td>0.183</td>
<td>0.030(100%)</td>
<td>-0.002(-12%)</td>
<td>0.001(2%)</td>
<td>0.000(-1%)</td>
<td>-</td>
</tr>
<tr>
<td>JULES</td>
<td>0.034</td>
<td>0.31</td>
<td>0.189</td>
<td>0.017(100%)</td>
<td>-0.001(-6%)</td>
<td>-0.005(-28%)</td>
<td>0.000(0%)</td>
<td>0.005(31%)</td>
</tr>
<tr>
<td>LPJ-GUESS</td>
<td>0.033</td>
<td>0.11</td>
<td>0.140</td>
<td>0.026(100%)</td>
<td>-</td>
<td>0.000(-1%)</td>
<td>-0.001(-5%)</td>
<td>-</td>
</tr>
<tr>
<td>MIROC-ESM</td>
<td>0.025</td>
<td>0.44</td>
<td>0.140</td>
<td>0.024(100%)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ORCHIDEE</td>
<td>0.045</td>
<td>0.00</td>
<td>0.201</td>
<td>0.030(100%)</td>
<td>0.010(34%)</td>
<td>-0.002(7%)</td>
<td>0.001(2%)</td>
<td>0.017(56%)</td>
</tr>
<tr>
<td>UVic</td>
<td>0.031</td>
<td>0.11</td>
<td>0.125</td>
<td>0.031(100%)</td>
<td>0.017(56%)</td>
<td>0.000(0%)</td>
<td>0.000(-1%)</td>
<td>-</td>
</tr>
<tr>
<td>UW-VIC</td>
<td>0.031</td>
<td>2.01</td>
<td>0.125</td>
<td>0.011(100%)</td>
<td>0.029(266%)</td>
<td>-0.005(-47%)</td>
<td>0.000(0%)</td>
<td>-</td>
</tr>
</tbody>
</table>
Figure legends

Figure 1. The spatial extent of regions defined in this study. Red, green, blue and magenta indicate the regions of boreal North America (BONA), boreal Europe (BOEU) and boreal Asia (BOAS), other permafrost areas (Other), respectively. We only selected BONA, BOEU and BOAS sub-regions for analysis in this study.

Figure 2. Simulated anomaly of annual $T_s$ at 20 cm averaged over boreal regions of each model, during the period of 1960-2000.

Figure 3. Simulated trends of annual $T_s$ at 20 cm averaged over boreal regions and sub-regions of each model, from 1960 to 2000. * indicates significant trend of $T_s$ (P<0.05).

Figure 4. Spatial distributions of trends of annual $T_s$ at 20 cm over boreal regions from 1960 to 2000 in (a) CLM, (b) CoLM, (c) ISBA, (d) JULES, (e) LPJ-GUESS, (f) MICRO-ESM, (g) ORCHIDEE, (h) UVic and (i) UW-VIC models. The black dots indicate regions with significant trends of $T_s$ (P<0.05). Note that extreme values outside of the range of -0.05 °C yr$^{-1}$ - 0.05 °C yr$^{-1}$ are shown in deepest blue and red in the color bar.

Figure 5. Simulated trends of annual $T_s$ over boreal regions as a function of soil depths (a) 0 - 3 m and (b) 0 - 40 m for the nine models. Note the different total soil depths of the models and negative trends for UW-VIC (~ -0.01 - -0.03 °C yr$^{-1}$) below 2.3 m are not shown in the plots.

Figure 6. Spatial distributions of difference in trends of annual $T_s$ at 0.2 m and 3 m over boreal regions from 1960 to 2000 in (a) CLM, (b) CoLM, (c) ISBA, (d) JULES, (e) LPJ-GUESS, (f) MICRO-ESM, (g) ORCHIDEE, (h) UVic and (i) UW-VIC models. The black dots indicate statistically significant difference by t-test (P<0.05). Note that extreme values outside of the range of -0.005 °C yr$^{-1}$ - 0.005 °C yr$^{-1}$ are shown in deepest
blue and red in the color bar.

Figure 7. Simulated trends of annual $T_s$ at 20 cm and $T_a$ in the climate forcing data across the nine models.

Figure 8. Simulated trends of annual $T_s$ at 20 cm and annual LWDR in the climate forcing data over boreal regions across the seven models which used and provided LWDR in their climate forcing. The black dotted lines indicate the linear regression and 95% confidence interval. The gray dashed line indicates the uncertainty of trend of LWDR in the climate forcing data. The solid blue and orange lines with double arrows indicate FU and SU, respectively. The red solid line with shade area shows the trend of LWDR (0.087 ± 0.023 W m$^{-2}$ yr$^{-1}$) during the period 1960-2000 from CRUNCEP v5.2 dataset. The purple solid line with shade area shows the trend of LWDR (0.187 ± 0.028 W m$^{-2}$ yr$^{-1}$) during the period 1960-2000 from WATCH dataset. The right panel shows the prior normal probability density function (PDF) with modeled mean and standard deviation (black solid line) and posterior normal PDF (red and purple dotted line) with given trend of LWDR from CRUNCEP and WATCH respectively.

Figure 9. Simulated trends of summer $T_s$ at 1 m and loss rate of NSPA over (a) boreal regions, (b) BONA, (c) BOEU and (d) BOAS across the nine models.
Figure 1. The spatial extent of regions defined in this study. Red, green, blue and magenta indicate the regions of boreal North America (BONA), boreal Europe (BOEU) and boreal Asia (BOAS), other permafrost areas (Other), respectively. We only selected BONA, BOEU and BOAS sub-regions for analysis in this study.
Figure 2. Simulated anomaly of annual $T_s$ at 20 cm averaged over boreal regions of each model, during the period of 1960-2000.
Figure 3. Simulated trends of annual $T_s$ at 20 cm averaged over boreal regions and sub-regions of each model, from 1960 to 2000. * indicates significant trend of $T_s$ ($P<0.05$).
Figure 4. Spatial distributions of trends of annual $T_s$ at 20 cm over boreal regions from 1960 to 2000 in (a) CLM, (b) CoLM, (c) ISBA, (d) JULES, (e) LPJ-GUESS, (f) MICRO-ESM, (g) ORCHIDEE, (h) UVic and (i) UW-VIC models. The black dots indicate regions with significant trends of $T_s$ ($P<0.05$). Note that extreme values outside of the range of $-0.05 \, ^{\circ}C \, yr^{-1}$ - $0.05 \, ^{\circ}C \, yr^{-1}$ are shown in deepest blue and red in the color bar.
Figure 5. Simulated trends of annual $T_s$ over boreal regions as a function of soil depths (a) 0 - 3 m and (b) 0 - 40 m for the nine models. Note the different total soil depths of the models and negative trends for UW-VIC ($\sim -0.01$ - -0.03 °C yr$^{-1}$) below 2.3 m are not shown in the plots.
Figure 6. Spatial distributions of difference in trends of annual T, at 0.2 m and 3 m over boreal regions from 1960 to 2000 in (a) CLM, (b) CoLM, (c) ISBA, (d) JULES, (e) LPJ-GUESS, (f) MICRO-ESM, (g) ORCHIDEE, (h) UVic and (i) UW-VIC models. The black dots indicate statistically significant difference by t-test (P<0.05). Note that extreme values outside of the range of -0.005 °C yr⁻¹ - 0.005 °C yr⁻¹ are shown in deepest blue and red in the color bar.
Figure 7. Simulated trends of annual $T_s$ at 20 cm and $T_a$ in the climate forcing data across the nine models.
Figure 8. Simulated trends of annual $T_s$ at 20 cm and annual LWDR in the climate forcing data over boreal regions across the seven models which used and provided LWDR in their climate forcing. The black dotted lines indicate the linear regression and 95% confidence interval. The gray dashed line indicates the uncertainty of trend of LWDR in the climate forcing data. The solid blue and orange lines with double arrows indicate FU and SU, respectively. The red solid line with shade area shows the trend of LWDR ($0.087 \pm 0.023$ W m$^{-2}$ yr$^{-1}$) during the period 1960-2000 from CRUNCEP v5.2 dataset. The purple solid line with shade area shows the trend of LWDR ($0.187 \pm 0.028$ W m$^{-2}$ yr$^{-1}$) during the period 1960-2000 from WATCH dataset. The right panel shows the prior normal probability density function (PDF) with modeled mean and standard deviation (black solid line) and posterior normal PDF (red and purple dotted line) with given trend of LWDR from CRUNCEP and WATCH respectively.
Figure 9. Simulated trends of summer $T_s$ at 1 m and loss rate of NSPA over (a) boreal regions, (b) BONA, (c) BOEU and (d) BOAS across the nine models.

(a) All

$$y = -2802.8x + 45.65$$
$$R^2 = 0.73, P = 0.003$$

(b) BONA

$$y = -740.5x + 9.08$$
$$R^2 = 0.85, P < 0.001$$

(c) BOEU

$$y = -88.0x + 0.78$$
$$R^2 = 0.46, P = 0.046$$

(d) BOAS

$$y = -1742.5x + 31.76$$
$$R^2 = 0.62, P = 0.012$$