Overview

We thank the reviewers for their constructive comments and suggestions. The manuscript has been appropriately revised in response to the reviewers' comments (see the point-by-point responses below). As requested by the reviewers, we compared our modeling results with some existing permafrost data sets by calculating evaluation metrics that can be compared directly against matching results reported in the literature. In addition, we also conducted several new simulations that further assess the impact on ALT of the model soil layer configuration, the soil organic carbon content, and its vertical distribution.

In summary, our modifications to the text can be categorized as follows (for reference, our response to comment "m" by reviewer "n" is labeled "R[n]C[m]"):

a) Novelty and added value:
   See R1C1 (i.e., Reviewer 1, Comment 1), R1C2, and R3C6

b) Comparison with other model-generated permafrost data sets:
   See R1C1, R1C2, R2C8b, R3C6, R3C29 and R3C30

c) Rephrasing "optimistic" discussion about ALT results:
   See R1C10, R1C11, R1C12, R3C2, R3C31

d) New sensitivity experiments and uncertainty discussion:
   See R1C3, R1C20, R2C8 and R2C12

e) Add specific evaluation metrics instead of using description words:
   See R3C6, R3C8, R3C24, R3C29 and R3C31

Throughout the discussion below, the text is colored as follows:

Black:	Reviewer comment
Blue:	Author response
Red:	Text newly inserted into the revised manuscript

Reviewer #1

This paper provides an evaluation of active layer thickness and permafrost extent as simulated by the NASA Catchmant Land Surface Model driven by MERRA-2. The model-generated dataset of permafrost conditions is evaluated again site data, global data and remotely (plane) sensed ALT. The comparison to the remotely sensed ALT is probably the most innovative part of the paper, but it also suffers from some drawbacks because the remotely sensed data conspicuously lack spatial variability. The paper is written clearly, the analysis is honest (not obviously trying to hide model shortcomings – but sometimes the assessment of the dataset quality seems a bit too optimistic), the figures are relevant and informative. The paper is a useful contribution, but some aspects detailed in the following could be improved.

We thank the reviewer for the careful reviewing. We understand the reviewer's concern about the small spatial variability with the remotely sensed ALT retrievals. The ALT retrievals were produced by the current algorithm developed by Chen et al. (2018). Somewhat larger spatial variability is presented in the original retrievals but is smoothed and reduced after aggregating to the scale of the Catchment Land Surface Model (CLSM) at 81 km^2, as also mentioned in the original manuscript. In addition, the radar penetration depth is not large enough to detect deeper thawed to frozen condition of the soil. All in all, while we eventually expect to further improve the retrieval algorithm, these are the results we have right now, and as discussed further below, their inclusion in this paper in their early form does, we feel, have value. We have tamped down our "optimistic" tone when discussing the ALT retrievals. We trust this first intercomparison of ALT among model results, remotely sensed retrievals and in-situ observations could provide useful insights to the research community. Please see specifically our response to R1C11, R1C12 and R2C7 below.

General:

- There are lots of global permafrost simulations that are driven by reanalysis-based meteorologies. What is really the added value of this one? The fact that is uses MERRA? In that case, could you say more about specific strengths and weaknesses of MERRA, please? More generally, simulations with other metorological forcing data, and comparison to other model-generated permafrost data sets (e;g. within the Permafrost Carbon Network) could be interesting.

R1C1: We thank the reviewer for encouraging us to explicitly highlight our contributions. Regarding the second, more general comment, a detailed quantitative comparison against existing permafrost data sets simulated by other land models, for example the land models participated in the Permafrost Carbon Network (PCN) (offline mode) and the Coupled Model Intercomparison Project phase 5 (CMIP5) (offline/coupled mode), is beyond the scope of this paper, which is already a bit long. (Note that aspects of such a general analysis have already been reported in literature (Peng et al., 2016;Wang et al., 2016b;Koven et al., 2013)). Inspired by the reviewer's comment, though, we added a brief discussion to our manuscript that compares our dataset with others in terms of spatiotemporal resolution and simulated results (see our response in R3C6 for details). We also summarized our dataset's particular strengths in our discussion section.

Specifically, we have added the discussion below.

a. General comparison between this work and existing model-generated permafrost data sets forced by other meteorological forcing data (both in uncoupled and coupled mode):

a.1 Regarding resolution (section 1, page 2, line 32-33 and page 3, line 1-7 in the revised manuscript)

"Most of these land models were run at coarse spatial resolutions, e.g., ranging from 0.5° × 0.5° to 1.8° × 3.6° for LSMs participating in the Permafrost Carbon Network (PCN) (Wang et al., 2016a) and from 0.188° × 0.188° to 4.10° × 5° for the models participating in the Coupled Model Intercomparison Project phase 5 (CMIP5) (Koven et al., 2013; https://portal.enes.org/data/enes-model-data/cmip5/resolution). As a result, it is difficult to compare the simulated values with in-situ observations taken at the point scale. Other types of numerical models have been run at relatively higher resolution, but not globally; such simulation domains were limited to regional scales (e.g., 2 km × 2 km in Jafarov et al. (2012) covering Alaska;1 km × 1 km in Gisnas et al. (2013) covering Norway) as necessitated by the availability of ancillary data and the heavy computational burden. As discussed further below, one of the unique contributions of the present work is a global simulation of permafrost at a somewhat higher resolution than earlier global-scale studies."

a.2 Regarding model performance in simulated permafrost extent (particularly the deficiency in western Siberia) (section 4.4, page 17, line 12-20 in the revised manuscript)

"Note that some other global models, such as CLM3 and the Community Climate System Model version 3 (CCSM3) as reported in Lawrence et al. (2012), also missed this area of permafrost and that updated versions of these models (i.e., CLM4 and CCSM4) showed improved performance in this regard (Lawrence et al., 2012). Guo et al. (2017) reported underestimated permafrost extent simulated in western Siberia using CLM4.5 driven by three different reanalysis forcings (i.e., CFSR, ERA-I and MERRA), and they showed an improved simulation of permafrost extent in this area when using another reanalysis forcing, the CRUNCEP (Climatic Research Unit - NCEP) (Guo and Wang, 2017). Guimberteau et al. (2018) found similar improvements stemming from the use of CRUNCEP forcing. We leave for further study whether the MERRA-2 forcing data is responsible for the western Siberia deficiency seen in our own results" a.3 Regarding model performance in simulated ALT (section 4.4, page 18, line 18-20 in the revised manuscript)

"Note that the existing literature on simulated ALT fields (e.g., Dankers et al. (2011), Lawrence et al. (2012) and Guo et al. (2017)) reveals a general tendency for models to overestimate ALT climatology at the global scale. Our results here suggest that the CLSM-simulated ALT fields are among the better simulation products, especially for shallow permafrost."

b. Some improvements of MERRA-2 compared to MERRA:

"Note that MERRA-2 has been found to be skilful in its simulation of near-surface atmospheric conditions (Reichle et al., 2017a;Reichle et al., 2017b;Bosilovich et al., 2015;Bosilovich et al., 2017) and to show improvements in the representation of cryospheric processes compared with its predecessor MERRA (Gelaro et al., 2017). In particular, MERRA-2 assimilates substantially more satellite observations and employs more physically reasonable hydrology representations for glaciated land surfaces compared to MERRA, and it also uses observation-based, seasonally-varying sea ice albedo as opposed to MERRA's fixed value of 0.6 (Gelaro et al., 2017). A recent study shows that permafrost and ALT simulation results obtained with forcing data from the original MERRA reanalysis are inferior to those driven by other reanalysis-based forcing data sets, particularly those from the NOAA Climate Forecast System Reanalysis (CFSR) and European Centre for Medium-Range Weather Forecasts Re-Analysis Interim (ERA-I) (Guo et al., 2017). The superiority of MERRA-2 forcing compared to MERRA forcing in the context of permafrost simulation is presumed here (given its general improvements in the cryosphere), though a side-by-side test of the two forcing datasets in this regard has not been performed." (section 1, page 4, line 1-12 in the revised manuscript)

"We note that our MERRA-2-driven permafrost simulation results, while potentially better than those we might have obtained with MERRA forcing, are still lacking (e.g., in western Siberia). " (section 5, page 20, line 10-13 in the revised manuscript)

c. Summary of novelty and added value of this work:

"The permafrost dataset presented herein can be considered unique in terms of its daily temporal resolution combined with a relatively high spatial resolution at the global scale (i.e., 81 km^2). The dataset, which is derived from a state-of-the-art reanalysis, shows reasonable skill in capturing permafrost extent and in adequately estimating ALT climatology (aside from that at the Mongolian sites). We note that our MERRA-2-driven permafrost simulation results, while potentially better than

> those we might have obtained with MERRA forcing, are still lacking (e.g., in western Siberia). Still, with its resolution and available variables (ALT, subsurface temperature at different depths), the dataset could prove valuable to many future permafrost analyses." (section 5, page 20, line 7-13 in the revised manuscript)
>
> "This work also provides a first comparison between two highly complementary approaches to estimating permafrost: model simulation and remote sensing retrieval method." (section 5, page 20, line 15-16 in the revised manuscript)
>
> "The remote sensing approach is still relatively new, and many aspects still need to be worked out. It is important, though, to begin considering the modeling and remote sensing approaches side by side, as both should play important roles in permafrost quantification in the years to come. Indeed, once the science fully develops, joint use of modeling and remote sensing (e.g., through the application of downscaling methods) should allow the generation of more accurate permafrost products at even higher resolutions." (section 5, page 20, line 23-27 in the revised manuscript)

- Some words about potential uses of this dataset could be nice.

R1C2: We thank the reviewer for this suggestion. Relevant discussion about the potential uses of this dataset has been added to the end of this manuscript (section 5, page 23, line 10-12 in the revised manuscript):

> "These data can potentially contribute, for example, to ecological studies focused on the dynamics of microbial activity and soil respiration in cold regions, on vegetation migration/adaptation in response to climate change, and so on."

Specific points:

- Page 2, line 15: "simulations… with the land surface model (Dankers et al., Guimberteau et al., Tao et al.)" -> these are different models. The sentence is misleading, and confusingly, you write "…and other numerical models" afterwards…

R1C3: We revised the relevant sentences as follows (section 1, page 2, line 25-32 in the revised manuscript).

> "Simulations and/or predictions with a variety of land surface models (LSMs) have been used to quantify large-scale permafrost patterns (i.e., distributions and thermal states) and their interactions with a warming climate. LSMs utilized for this include, for example, the Joint UK Land Environment Simulator (JULES, Dankers et al., 2011), the ORganizing Carbon and Hydrology in Dynamic EcosystEms (ORCHIDEE) - aMeliorated Interactions between Carbon and Temperature (ORCHIDEE-MICT, Guimberteau et al., 2018), the Catchment Land Surface Model (CLSM, Tao et al., 2017), and the Community Land Model (CLM,

> Lawrence and Slater, 2005;Alexeev et al., 2007;Nicolsky et al., 2007a;Yi et al., 2007;Lawrence and Slater, 2008;Lawrence et al., 2008;Lawrence et al., 2012;Koven et al., 2013;Chadburn et al., 2017;Guo and Wang, 2017)."

- Page 2, line 20-24: Strictly speaking, the fact that 2017 set records doesn't mean that permafrost conditions will change. 2017 is only one year. It's the long-term trends that matter (2017 is of course consistent with that trend)

R1C4: We agree with the Reviewer regarding this point. The message we had tried to convey is that the warming trend of our climate seems to have increased in recent years and that given the associated exacerbation in permafrost thawing, monitoring permafrost in a timely manner is critical. We have modified the relevant sentences as follows (section 1, page 2, line 17-20 in the revised manuscript). Please also see our response to next comment (R1C5).

> "In addition, given the apparent climate warming seen in recent years (exemplified by the fact that the average Arctic air temperature in 2017 (ending in September) was the second warmest on record since 1900 [Arctic Report Card; http://www.arctic.noaa.gov/Report-Card/Report-Card-2017] and that 2017 was the warmest year on record for global ocean temperatures (Cheng and Zhu, 2018)), important reductions in permafrost might be occurring as well."

- Page 2, line 23: "Some aspects of the current global permafrost thermal states are ... still unknown": can you elaborate on that, please?

R1C5: We have clarified this by expanding this sentence as shown below (section 1, page 2, line 20-23 in the revised manuscript):

> "However, current global permafrost thermal states (i.e., permafrost temperature, ice content and degradation rates across much of Northern latitudes) are still largely unknown. Monitoring permafrost degradation in a timely manner is particularly critical for ecosystem management and for various policy decisions."

- Page 3, line 10: "extensive challenges" sounds bizarre to my non-native speaker's ears

R1C6: We replaced the end of the sentence with "combined with the many challenges of physical process modelling."

- Page 3, line 16: Could you say a few words specifically about high-latitude performance? Advantages, drawbacks // other reanalyses?

R1C7: We have added some relevant discussion about the performance of MERRA-2 in high-latitude regions (section 1, page 4, line 4-6 in the revised manuscript):

> "In particular, MERRA-2 assimilates substantially more satellite observations and employs more physically reasonable hydrology representations for glaciated land surfaces compared to MERRA, and it also uses observation-based, seasonally-

> varying sea ice albedo as opposed to MERRA's fixed value of 0.6 (Gelaro et al., 2017)."

- Page 3, line 26: Chen et al. is a paper in review. Can you reassure the reviewer that these retrievals are independent from the data produced here? One or two sentences would be nice anyway even if Chen et al. 2018 will be available to the reader soon.

R1C8: The AirMOSS radar retrievals of ALT we used here were produced by the algorithm described and analyzed in detail by Chen et al. (2018). The retrievals here are indeed identical to those produced by Chen et al. 2018, though here we examine them from a different perspective. We now added some text to distinguish the scope of this study from that of Chen et al. (2018) and to emphasize the data independence (section 1, page 4, line 18-20 in the revised manuscript):

> "In their study, Chen et al. (2018) mainly focus on the development and improvement of the ALT retrieval algorithm, whereas the present study emphasizes using the ALT retrievals to assess the (fully independent) ALT simulations."

- Page 5, line 13: because you later speak about the spinup in the trend analysis, it might be interesting to say a few words about this here. The looping through the 36 years cannot given the same soil temperatures as you would normally have if you have realistic spinup data.

R1C9: We thank the reviewer for the suggestion. We added two sentences here (section 2.1, page 6, line 19-22 in the revised manuscript).

> "The details of the spin-up procedure employed here admittedly impact our trend analysis (section 4.5); the approach makes use of the warmer conditions during the last few decades and thus should produce warmer 1980 initial conditions than would be produced with realistic historical forcing over hundreds of years (e.g., Sapriza-Azuri et al., 2018)."

- Page 8, line 15 and following: The assessment might be a bit too optimistic here: Basically one sees that the ALT is between 0.2 and 1m both in obs and simulations, not much more. Is there a significant correlation at all?

R1C10: We have modified this sentence as shown below (section 4.1, page 12, line 24-27 in the revised manuscript).

> "Figure 4b, c demonstrates that in some ways, the CLSM-simulated results roughly agree, to first order, with the in-situ observations. The overall mean bias of simulated ALT relative to the in-situ measurements is -0.05 m. Nevertheless, the scatter (blue) in Fig. 4c is large, and the corresponding correlation coefficient is quite weak (0.27)."

- Page 8, line 21 and following: The AirMOSS ALT retrievals. Bascically the retrievals are the same everywhere! Around 0.45 m. No variability. Are they actually of any use?

R1C11: As we mentioned in the original manuscript, relatively larger variability with the ALT retrievals is seen at its native resolution (Figure 3a), but this larger variability was smoothed out through aggregation to the model scale at 81 km^2, as we expected. In addition, as also mentioned in the manuscript, these retrievals cannot exceed the P-band radar sensing depth of about 60cm over this area, and thus for the shallow permafrost here, the averaged ALT retrievals appears to be around 0.45 m everywhere.

We emphasize that this is a first attempt to compare remote sensing ALT data with modeling results. An expected future direction is to take advantage of the detailed heterogeneity information in the remote sensing data to downscale model results directly or to improve modeling skill indirectly. We have added several sentences about the potential use of these ALT retrievals into the discussion section (see our response to R1C1, part c, above). Please also see our response to the next comment (R1C12).

- Page 8, line 32 and following (“Excluding: : :”): Yes, OK, but then there is still no correlation. Values are just around 0.45 m and the mean ALT of the remaining sites just happens to be around that value.

R1C12: With further analysis, we found that the correlation coefficient between the ALT retrievals and the in-situ observations, while small at the site scale, is larger than that for modeled ALT at the model scale, both for all sites and for sites with measured ALT below the AirMOSS radar sensing depth of 60cm. This is the depth below which the AirMOSS radar is expected to lose sensitivity to subsurface features, and is calculated based on the radar system noise floor and calibration accuracy. Therefore, any retrieved ALT larger than ~60 cm is expected to have large uncertainties, and the error is further expected to grow linearly as the retrieved values of ALT essentially “saturate.” We have added these discussion and included the Table R1 as a new Table 3 in the revised manuscript:

> “Note that the retrieved ALT cannot exceed the radar sensing depth of about 60 cm. This is the depth below which the AirMOSS radar is expected to lose sensitivity to subsurface features, and is calculated based on the radar system noise floor and calibration accuracy. Therefore, any retrieved ALT larger than 60 cm is expected to have large uncertainties, and the error is further expected to grow linearly as the retrieved values of ALT essentially “saturate.” This limitation may also lead to underestimates of the actual thaw depth. ” (section 2.2, page 7, line 18-22 in the revised manuscript)
>
> “For the AirMOSS retrievals, when all in-situ sites are considered, the overall ALT bias is -0.11 m at the site scale and -0.12 m at the model scale. While the correlation coefficient with the in-situ observations is only 0.05 at the site scale, it is 0.61 at the model scale. ” (section 4.1, page13, line 4-7 in the revised manuscript)
>
> “Excluding the sites with in-situ ALT measurements that exceed the AirMOSS sensing depth of ~60 cm, the overall mean bias for the AirMOSS retrievals at the model scale (site scale) drops to -0.01 m (0.02 m), and the correlation coefficient

> at the model scale (site scale) increases to 0.64 (0.20) (Table 3). In contrast, the CLSM simulation results show a bias of 0.01 m and a zero correlation coefficient at the same sites.” (section 4.1, page 13, line 12-15 in the revised manuscript)

Table R1– Evaluation metrics for model-simulated ALT and AirMOSS retrievals for 2015. (New Table 3)

Metric	All sites			Sites with ALT measurements within AirMOSS sensing depth (~ 60 cm)		
	CLSM-simulated ALT (model scale)	AirMOSS ALT retrievals (model scale)	AirMOSS ALT retrievals (site scale)	CLSM-Simulated ALT (model scale)	AirMOSS ALT retrievals (model scale)	AirMOSS ALT retrievals (site scale)
RMSE (m)	0.17	0.17	0.21	0.12	0.06	0.08
Bias (m)	-0.05	-0.12	-0.11	0.01	-0.01	0.02
R	0.27	0.61	0.05	-0.00	0.64	0.20

- Page 9, line 9-10: “Further investigation: : :”: You could nevertheless elaborate a little bit on this. Are there any common characteristics of sites with thick active layer (dry soil, highly conducive soil, southward sloping, etc.) that the model doesn’t get?

R1C13: Here we specifically meant some investigation on the zero-curtain problem with the model. Nevertheless, we have amended the text to read (section 5, page 21, line 3-5 in the revised manuscript):

> “The use of the 0ºC degree threshold in CLSM for determining the thawed or frozen layer may explain in part the model’s underestimation of ALT, as may the lack of

> an explicit treatment of local aspect, errors in assigned model parameters, and so on."

(Note that we have moved this paragraph to the discussion section, as suggested by the Reviewer #3.)

- Page 11, line 1-3 – meteorological forcing dominant control: Of course. Could anyone seriously expect something different?

R1C14: Yes. We have added "as expected" into this sentence.

- Figure 7: Good that this is quantified in such a way here. Much more synthetic &interesting than figure 6.

R1C15: Thank you. We agree.

- Page 13, line 12: Correlation might increase if time steps with snow on the ground but air temp > 0_C are not counted in Tcum - it's the soil surface temperature that counts, not the air temperature.

R1C16: The reviewer makes a reasonable suggestion, and we thus recomputed the correlation map using this modified rule. We found that the map did not differ very much from the original one. We keep the original figure as is.

- Page 14, line 5 : Problems in mountain areas: Snow forcing might be severely in error in these regions

R1C17: Yes. We have added one sentence here (section 4.3, page 17, line 1-2 in the revised manuscript).

> "In addition, MERRA-2 snow forcing might be severely erroneous in these regions."

- Page 14, line 12: "The reasons: : :" – I have probably missed the information: How deep is the model soil column?

R1C18: As mentioned in section 2.1, the depth ranges of the six soil layers are 0-0.1m, 0.1-0.3m, 0.3-0.7m, 0.7-1.4m, 1.4-3m, and 3-13m, respectively. We checked the CALM sites over western Siberia and found that the ALT observations there are basically below 2 m. Therefore, the depth of the model's soil column is not an issue, and we deleted this speculation from the sentence. The new sentence below instead was added (section 4.4, page 17, line 11-12 in the revised manuscript).

> "The reasons for this particular deficiency are unclear; perhaps the initial thermal conditions over western Siberia were too warm, or perhaps MERRA-2 overestimates current air temperatures in this region."

- Page 16, line 15, Mongolian ALT trends: How can you have a 25 cm/year trend over 17 years? That would mean that ALT increases by over 4 m over that period. That's quite improbable. These data are very suspicious.

R1C19: We share the reviewer's concerns about the quality of these data. Below in Table R2 we provide the time series of the actual ALT measurements at the three Mongolian sites M1, M1 A and M3 (https://www2.gwu.edu/~calm/data/north.html). Because M1 and M1 A are within the same CLSM modeling grid cell, we used the average of their two time series. Time series of observed and simulated ALT at the two grid cells containing these three sites are plotted in Figure R1. The observed and simulated ALT trends at the two grid cells correspond to the two dots showing the extraordinarily large observed trends in Figure 14a in the original manuscript. Note the simulated ALT trends were calculated using ALT estimates only in years when observed ALTs are available, as also mentioned in the manuscript.

Table R2 – Observed ALT (cm) at three Mongolian sites.

Site Code	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
M 1	345	350	355	345	350	340	355	-	375	365	380	380	370	350
M1 A	390	390	-	430	485	400	450	600	770	710	820	820	-	-
M3	-	610	620	600	610	660	720	760	770	750	760	770	800	-

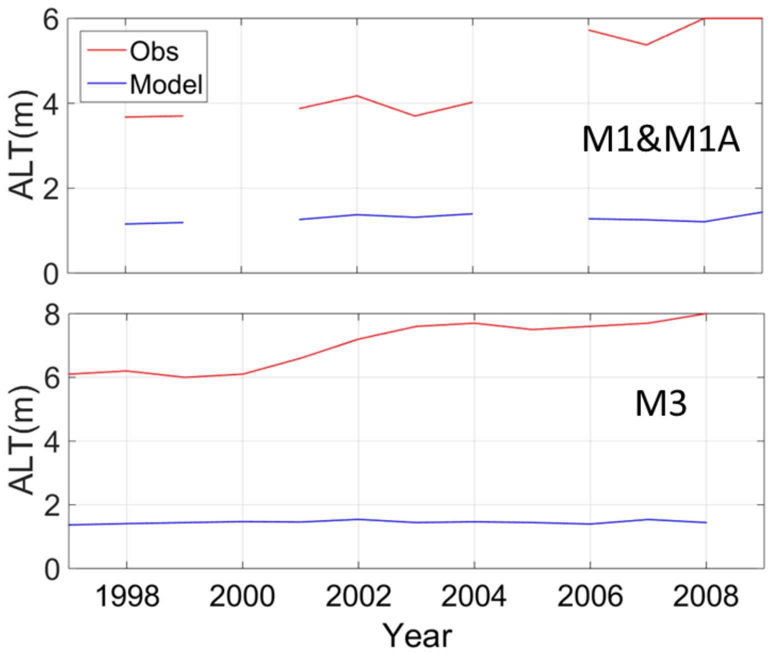


Figure R1: Time series of observed and simulated ALT at Mongolian sites collocated with two simulation grid cells, i.e., M1&M1A (upper) and M3 (bottom). The calculated ALT trends from observations and simulation are 24.38 cm/yr and 1.13 cm/yr, respectively, for the grid cell containing M1&M1A and are 19.69 cm/yr and 0.51 cm/yr, respectively, for the grid cell containing M3.

We attempted to contact the PI responsible for these data (Dr. Natsagdorj Sharkhuu from the Institute of Geography, Mongolian Academy of Sciences); however, the email address (provided here https://www2.gwu.edu/~calm/data/webforms/mg_f.html) is apparently obsolete, and the email delivery failed. We were thus unable to investigate further the data quality.

In any case, as indicated by the reviewer, this issue calls at the very least for a specific caveat about these data, which we have added (section 4.5, page 19, line 28-30 in the revised manuscript):

> "A particular caveat is required regarding the Mongolian sites, given the unusual observed trends calculated there. Attempts to contact the data providers to attain more detailed information for data evaluation were unsuccessful, and accordingly our confidence in these particular data is limited."

- Page 18, line 20 : Â´n ..addition of soil layers Â˙z : Would that be so difficult? Tests with more levels would really be interesting, but if they are really difficult, I refrain from asking for such test to be carried out.

R1C20: In response to this comment we have conducted such tests. Our general conclusions are consistent with other studies in terms of how soil configuration affects permafrost modeling (e.g., Alexeev et al., 2007;Lawrence et al., 2008;Sapriza-Azuri et al., 2018;Nicolsky et al., 2007b;Dankers et al., 2011).

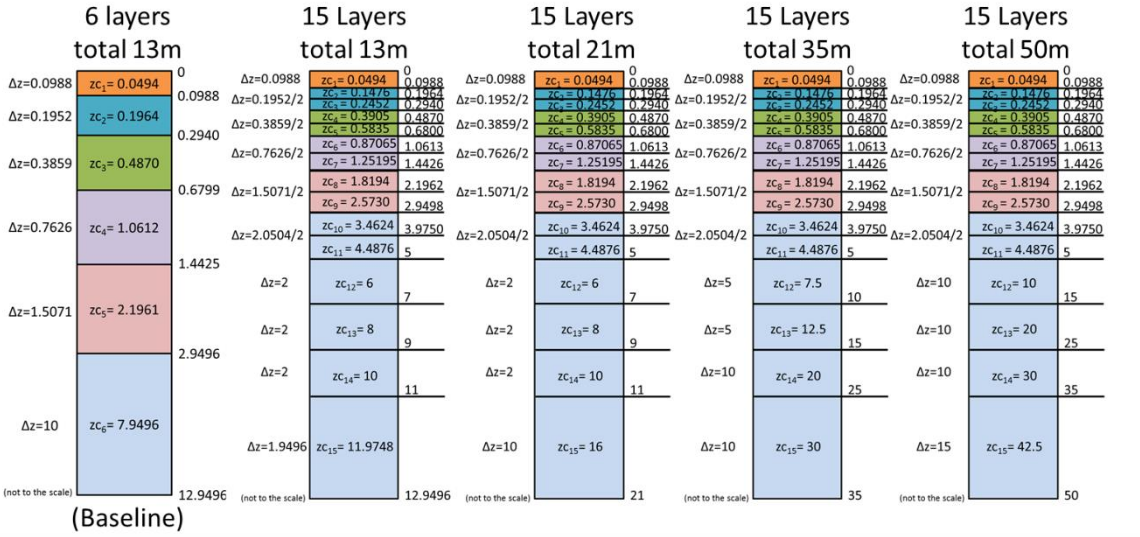


Figure R2: Soil configurations we have newly tested.

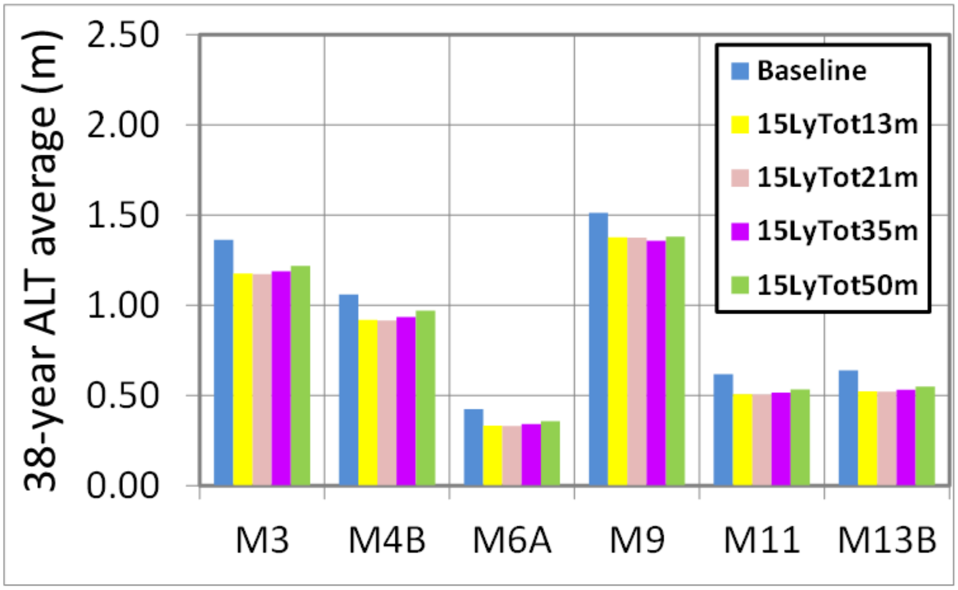


Figure R3: Simulated ALT results at six Mongolian sites with 5 different soil configurations. (Note the baseline soil configuration contains 6 soil layers and a total soil depth about 13 m.)

Specifically, we tested four new soil configurations with 15 soil layers and different soil depths (ranging from 13 m to 50 m). Figure R3 reveals that increasing the number of soil layers decreases ALT climatology at the tested sites, which is consistent with previous studies (e.g., Alexeev et al., 2007;Lawrence et al., 2008;Dankers et al., 2011). The figure also demonstrates that variations in total soil depth have only a small impact on the simulation of ALT, as also reported in the previous studies. However, the zero flux we employ at our lower boundary for all the simulations might influence heat transfer in deep soils and thus might decrease the impact of using deeper soils.

Based on these tests, we have included some appropriate discussion in the text, though without adding a figure (section 5, page 22, line 14-19 in the revised manuscript):

> "Test simulations (not shown) with alternative model configurations indicate that increasing the number of soil layers may act to decrease somewhat the simulated ALT, suggesting that our values may be a little overestimated; however, based on results from a new study by Sapriza-Azuri et al.(2018), our use of a no-heat-flux condition at the bottom boundary rather than a dynamic geothermal flux may lead to underestimates of ALT. Such uncertainties should naturally be kept in mind when interpreting our results. Our supplemental simulations also suggest that increasing the total modelled soil depth has only a small impact on simulated ALT."

We also explicitly mentioned our lower boundary condition (section 2.1, page 5, line 10-11 in the revised manuscript):

> "A no-heat-flux condition is employed at the bottom of the model's soil column."

Reviewer #2

Overall comments:

This paper used in-situ data and a remote sensing based ALT (active layer thickness) data to evaluate a model-based ALT dataset. Overall, I think it is a useful study. The analysis was done in a comprehensive way, and the results were acceptable. However, the analysis regarding the model uncertainty is somewhat general – considering we already have a good knowledge of the ability of global land models in NH permafrost simulation. I think, the study could benefit from more in-depth discussions on this aspect. The details were provided below.

We thank the reviewer for the careful reviewing. We absolutely agree with the reviewer about the importance of model uncertainty regarding ALT estimation. To further examine model uncertainty, we conducted new tests on sensitivity to organic carbon content and its vertical distribution. We also replaced the vegetation climatology at several Mongolian sites with satellite-based, time-variant leaf area index (LAI) during wintertime to investigate the impact of inter-annual variations in vegetation. Please also see our response to Reviewer #1 (R1C20) regarding the discussion about the soil layer configuration.

Major comments:

1. Page 4 Paragraph 3: I have questions regarding how the ALT was calculated. The paper indicates here it was calculated based on the simulated ice content. Does the model consider unfrozen water in frozen soils? If it does, please provide information on how the model calculates unfrozen water content. If not, this definition will be same as using a 0_C temperature threshold for thawed-to-frozen depth calculation. This information is especially important for the deep soils due to year-round low temperatures and coarse vertical resolution of the model at deeper depths.

R2C1: In our model, the ice fraction is unity if soil is fully frozen (i.e., $T < 0^oC$ and $f_{ice} = 1$), and it is zero if the soil is fully thawed (i.e., $T > 0^oC$ and $f_{ice} = 0$). The reviewer comment refers to the situation where the soil temperature is exactly at the freezing point and the soil is partially frozen (i.e., $T = 0^oC$ and $0 < f_{ice} < 1$). In the latter case, frozen and thawed soil and water coexist. This situation always occurs during freeze-to-thaw and thaw-to-freeze transitions. This is because in soil layers that are as thick as those used in the model, the phase transition does not occur instantly (relative to the model time step). More specifically, the model uses heat content as the prognostic variable, from which the temperature and ice fraction are diagnosed. Therefore, our calculation of ALT is not the same as simply using a 0^oC degree threshold to determine the thawed-to-frozen depth. Rather, we identify the deepest (fully or partially) thawed layer and then calculate the thawed-to-frozen depth based on the ice fraction within the layer. We have modified the paragraph about the ALT calculation as follows (section 2.1, page 5, line 22-29 in the revised manuscript).

> "Precisely, the thawed-to-frozen depth is calculated as:

> $$z_{\mathrm{bottom}}(l) - f_{\mathrm{ice}}(l,t) \times \Delta z(l), \qquad (1)$$
>
> where layer l is the deepest layer that is fully or partially thawed, $z_{\mathrm{bottom}}(l)$ represents the depth at the bottom of layer l, $f_{\mathrm{ice}}(l,t)$ is the fraction of ice in layer l at time t (i.e., $f_{\mathrm{ice}}(l,t) \in [0 \ \ 1]$), and $\Delta z(l)$ is the thickness of layer l. To identify layer l we use a 0^{o}C degree temperature threshold. Specifically, $T > 0^{o}$C degree indicates that a layer is fully thawed, $T = 0^{o}$C degree indicates that a layer is partially thawed, and $T < 0^{o}$C degree indicates that a layer is fully frozen. That is, layer l is the deepest layer that satisfies $T(l) \geq 0^{o}$C. Equation (1) then expresses that the thawed-to-frozen depth is equal to the bottom depth of the layer l but adjusted upward according to the ice fraction within the partially thawed layer l."

The above declaration seems contradictory to "The use of the 0_C degree threshold in CLSM for determining the thawed or frozen state of the soil may explain the model's underestimation of ALT." (Page 9, Paragraph 1). So I am confused what methods were actually used to determine the thawing depth/ALT. Please clarify.

R2C2: The 0^{o}C degree is used to determine the deepest thawed layer l, and then the ALT is calculated by $z_{\mathrm{bottom}}(l) - f_{\mathrm{ice}}(l,t) \times \Delta z(l)$ as explained above (R2C1). Here we are talking about the temperature threshold used to determine the thawed state of the soil only, not the thawing depth/ALT. We have modified the text as follows (section 5, page 21, line 3-5 in the revised manuscript):

> "The use of the 0^{o}C degree threshold in CLSM for determining the thawed or frozen layer may explain in part the model's underestimation of ALT, as may the lack of an explicit treatment of local aspect, errors in assigned model parameters, and so on."

(Note that we have moved this paragraph to the discussion section.)

2. Page 5 Paragraph 3: The spin-up scheme is questionable, though the authors themselves acknowledged this. Why do the authors using the meteorology for the entire 36-year period for spin-up? If the design is to reduce the uncertainty introduced by a single-year surface meteorology, spin-up using the first few years during the period will be more acceptable.

R2C3: As the reviewer points out, we recognized this issue and discussed it in the original manuscript. No spin-up procedure is entirely problem-free. Using a shorter period for spin-up as suggested by the reviewer would exaggerate in the resulting initial conditions any anomalies that occur during the spin-up period. The ultimate solution would be to construct a realistic historical forcing dataset over hundreds of years with a dynamic geothermal flux applied to the bottom boundary of soil column (e.g., Sapriza-Azuri et al., 2018). However, this approach is hardly feasible and would still not assure absolutely correct initial conditions. Please also see our response to Reviewer #1 (R1C9).

3. I have questions regarding the vegetation effects on permafrost simulation in Northern Alaska (Page 11, Paragraph 2). Those 4 northern flights were dominated by "dwarf trees"

as indicated by Fig. 2b (really?). Moreover, the changes in simulated maximum snow depth due to vegetation in those flights were much smaller comparing with the experiment homogenizing the forcing data (Fig. 6c). So I would expect the impact due to snow changes for the homogenizing vegetation experiment would be smaller comparing with the experiment homogenizing surface forcing, while this is not the case shown in Fig. 6a-b. Can the authors explain why?

R2C4: As we mentioned in the manuscript, the homogenization is applied cumulatively. Before we homogenized vegetation in this experiment, we already homogenized the forcing. The differences between HomF and HomF&Veg are then attributed to the changes in vegetation parameters (specifically LAI and vegetation height).

Fig.6a-b indeed demonstrates that the impact due to snow changes for the homogenizing vegetation experiment (differences between HomF and HomF&Veg) would be smaller compared to the experiment homogenizing surface forcing (differences between Baseline and HomF). Figure 7b also illustrates this in a quantitative way.

On the other hand, the Alaska North slope is dominated by tundra, while the vegetation map in CLSM indicates mostly dwarf trees or shrubs (Fig. 2b). Would this introduce uncertainties to the analysis on the contribution of different factors (i.e. forcing data, vegetation and soil) in Fig. 6?

R2C5: The vegetation class is only one of several model inputs. The land cover class used in the study is derived from the USGS Global Land Cover Characteristics Data Base Version 2.0 (GLCCv2). In addition to vegetation class, the model uses vegetation height, LAI, greenness fraction and albedo, which are all obtained from other satellite-based sources that reflect realistic climatologic conditions for tundra. Put differently, while the modeled vegetation class may suggest the presence of dwarf trees, the typically low (satellite) LAI values in northern Alaska will instruct the model that the tree cover is extremely sparse in this region. Please refer to Table 1 in (Tao et al., 2017) to see all the data sources. We have added one sentence to clarify this (section 4.1, page 12, line 8-11 in the revised manuscript).

> "Note that although the vegetation class (Figure 2b) suggests the presence of dwarf trees over the Alaska North Slope, the actual satellite-based LAI, vegetation height, greenness fraction and albedo will still instruct the model that the tree cover is extremely sparse in this region. The data sources for these vegetation-related boundary conditions can be found in Table 1 in Tao et al. (2017)."

Also, from Fig. 6, it does not seem to me that homogenizing soil parameters has much bigger impact on simulated ALT and surface soil temperature than homogenizing vegetation (at least for the northern flights), as the authors indicated in Fig. 7. Maybe I miss something here?

R2C6: We have added the following discussion to section 4.2 (page 15, line 21-26 in the revised manuscript).

> “Note that in Figure 6a, the soil impact on ALT (difference between HomF&Veg&Soil in black and HomF&Veg in red) appears smaller than that of the vegetation (difference between HomF in green and HomF&Veg in red) over the northern transects (ATQ, BRW and DHO). Even so, Figure 7b shows that, in terms of the integrated impact along the transects, the soil influence clearly outweighs the influence of vegetation – at several other transects, including HUS, KYK, COC, AMB, IVO and the first half of ATQ (where vegetation conditions might be similar to those used for homogenizing), the changes in vegetation parameters do not have much impact. ”

4. Most of the value for AirMOSS radar retrievals, I think, is in its ability to characterize land surface heterogeneity. Simply averaging AirMOSS data to a much coarseresolution (i.e. 9km in this study) to compare with the land model simulations is not very insightful in terms of exploring the value of this dataset. I agree with the authors that the current AirMOSS retrievals seem having large uncertainties; most notably, its ALT retrievals were in a very narrower range. However, the inconsistency of the ALT spatial pattern at some of AirMOSS flights may come from the model itself. For example, at the DHO flightâˇAˇ Tthis is the flight with most in-situ sites available, the model ALT generally increases from the north to the south, while the in-situ data show large variability, and do not show a clear increasing trend from the north to the south (Fig. 5a). There are also a number of studies pointing out that ALT is extremely variable at local scale. Therefore, analysis using a dataset like AirMOSS in this aspect would be more valuable.

R2C7: We agree with the value of AirMOSS radar retrievals in terms of (theoretically) being able to represent the spatial variability of ALT.

The differences in the spatial patterns of the AirMOSS ALT retrievals and the simulated ALT suggest that neither radar remote sensing nor modeling is perfect. As we mentioned in the manuscript, the radar sensing depth (about 60cm) strongly constrains the retrieval accuracy for ALT values larger than the sensing depth. We expanded on the analysis by adding a new Table 3, which provides several evaluation metrics for ALT restricted to less than 60cm. (Please see our response in R1C12 for details about this new Table 3.) The table suggests that the radar retrievals are in better agreement with in-situ observations especially at model scale when only using sites that have ALT less than or equal to 60cm. We have added several sentences to section 4.1 (page 13, line 12-15 in the revised manuscript):

> “Excluding the sites with in-situ ALT measurements that exceed the AirMOSS sensing depth of ~60 cm, the overall mean bias for the AirMOSS retrievals at the model scale (site scale) drops to -0.01 m (0.02 m), and the correlation coefficient at the model scale (site scale) increases to 0.64 (0.20) (Table 3). In contrast, the CLSM simulation results show a bias of 0.01 m and a zero correlation coefficient at the same sites. ”

5. It would be more interesting if the authors could provide more insightful analysis regarding the model ALT uncertainties or the correlation analysis, including:

a) Why does the model show much stronger correlation with maximum SWE in portions of NH permafrost region than with air temperature?

R2C8a: Honestly, we are not clear about the reasons yet. We have modified a relavant sentence as follows (page 16, line 29-30 in the revised manuscript).

> "Apparently, in these areas, the impacts of snow physics on ALT outweigh the impacts of lumped energy input (T_{cum}), for reasons that are not clear."

b) It would be helpful if the authors could give more explanations regarding why the model fails in western Siberia? Those areas also include continuous permafrost, so I do not think it is too challenging for global models to capture the permafrost distribution there.

R2C8b: As indicated in Figure 1b, all four types of permafrost (i.e., continuous, discontinuous, sporadic and isolated) are present in western Siberia. The literature suggests that other global models also missed this portion of permafrost, including CLM3 and CCSM3, although the updated versions of these models (i.e., CLM4 and CCSM4) demonstrated improved performance (Lawrence et al., 2012). Similarly, Guo et al. (2017) also reported underestimations in permafrost extent in western Siberia simulated by CLM4.5 when driven with three different reanalysis forcings (CFSR, ERA-I and MERRA) and improved performance extent in this region when using forcing data from a different reanalysis. We have added some discussion to compare our simulation with other existing works (section 4.4, page 17, line 12-20 in the revised manuscript):

> "Note that some other global models, such as CLM3 and the Community Climate System Model version 3 (CCSM3) as reported in Lawrence et al. (2012), also missed this area of permafrost and that updated versions of these models (i.e., CLM4 and CCSM4) showed improved performance in this regard (Lawrence et al., 2012). Guo et al. (2017) reported underestimated permafrost extent simulated in western Siberia using CLM4.5 driven by three different reanalysis forcings (i.e., CFSR, ERA-I and MERRA), and they showed an improved simulation of permafrost extent in this area when using another reanalysis forcing, the CRUNCEP (Climatic Research Unit - NCEP) (Guo and Wang, 2017). Guimberteau et al. (2018) found similar improvements stemming from the use of CRUNCEP forcing. We leave for further study whether the MERRA-2 forcing data is responsible for the western Siberia deficiency seen in our own results."

c) The ALT trends shown in Fig. 13a seem not very consistent with the trends of temperature indices shown in this figure. Have the authors explore the changes in snow cover duration? A longer snow free season generally leads to warmer soil temperature and thus deep ALT esp. in the southern area.

R2C8c: We did examine the trend in snow cover duration (see Figure R4). While some areas show a trend in snow cover duration, this trend does not seem correlated with the trend in ALT.

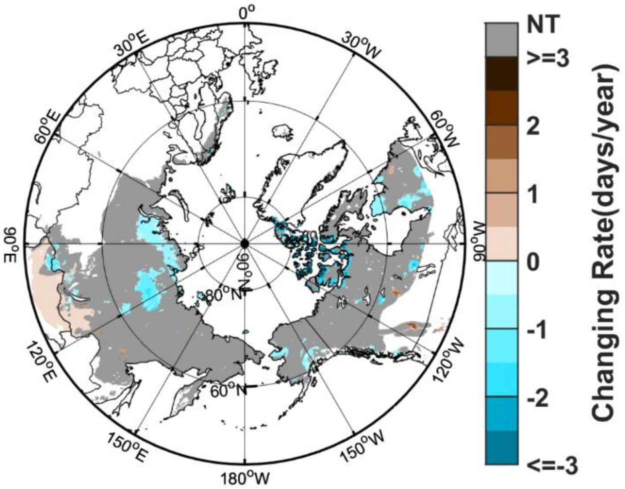


Figure R4: Spatial distribution of trend in snow persistence days when daily mean snow depth > 25cm. NT means no significant trend, i.e., the p value is larger than 0.05.

Accordingly, we modified the relevant sentence as follows (section 4.5, page 19, line 18-20 in the revised manuscript):

> "It is possible that in such cases, the computed trends are strongly affected by snowpack variability, though neither maximum SWE nor snow cover duration tends to show a significant trend in these areas (not shown)."

d) Page 18, Paragraph 3: I do not quite agree with the authors' explanation why the model fails in the Mongolian sites. For those sites, the model simulated ALT is generally less than 1.4 m, therefore, the coarse resolution at deeper soils in the model set-up (i.e. layer 5: 1.4-3m) should not be a major contributing factor there. Much drier conditions, sparse vegetation, and perhaps uncertainties in soil texture data (eps. in deep soils), I think, are more likely contributing more to the model uncertainties.

R2C8d: The reviewer is correct about the soil configuration. We conducted new tests using different soil configurations at several Mongolian sites. Please see the results and discussion in our response to the Reviewer #1 (R1C20).

The reviewer also raised a very good point regarding the influence on ALT of soil wetness, vegetation and uncertainties in soil texture for deep soils. We first examined realistic satellite-based LAI data at from Moderate Resolution Imaging Spectroradiometer (MODIS) MCD15A2H product and the Advanced very-high-resolution radiometer (AVHRR) AVH15C1 product (see Table R3). Figure R5 then shows the time series of the MODIS and AVHRR LAI, along with the LAI climatology used in the model at one CALM Mongolian site (M11). A post-processing procedure that included quality screening and gap filling was applied to the two satellite LAI products. The CLSM LAI climatology is used for the years that MODIS data is not available (1980 to 2002).

Table R3 – Information of satellite-based LAI products from MODIS and AVHRR.

Sensor	**Dataset**	**Product**	**Resolution**	**Temporal Granularity**	**Temporal Extent**
MODIS	MCD15A2H	Leaf Area Index and Fractional Photosynthetically Active Radiation	500 m	8-day Composites	July 2002 - Present
AVHRR	AVH15C1	Leaf Area Index and Fraction of Absorbed Photosynthetically Active Radiation	0.05deg	Daily	June 1981 - Present

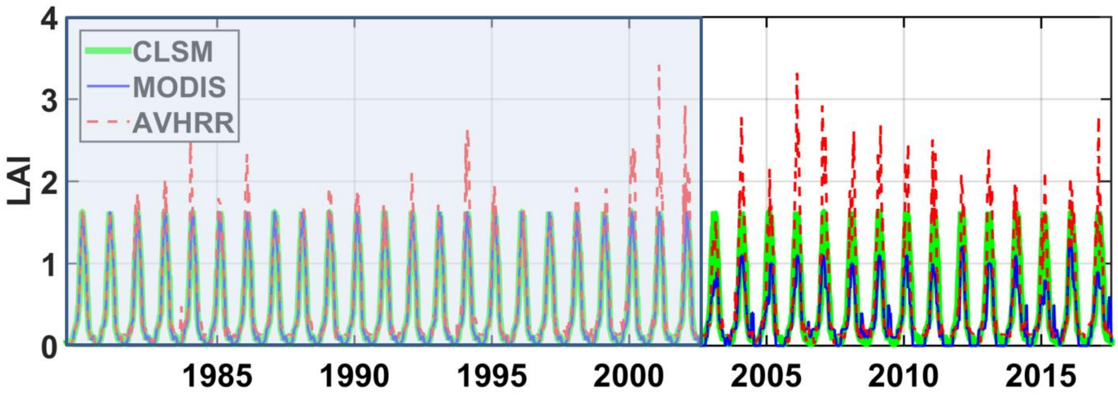


Figure R5: The LAI time series at one Mongolian site (M11). Green line represent the original LAI climatology used in the CLSM. Blue and red dash line represent the realistic (time-varying) LAI data from MODIS and AVHRR. The CLSM LAI climatology is used for the years that MODIS data is not available (1980 to 2002, shaded area).

As illustrated by Figure R5, MODIS shows smaller LAI than AVHRR over the valid period after 2002. The LAI climatology used in the model is inbetween of the two products. The differences between the LAI climatology used in the model and realistic LAI products would cause differences in energy and water partitioning at the land surface via impacting surface albedo. We conducted a simple test to examine the impact of using the more realistic, inter-annually varying vegetation inputs on the winter surface albedo and thus the snow accumulation process, which in turn would impact ALT estimates. Specifically, we replaced the LAI climatology with the satellite-based, inter-annually varying LAI products in the model, but turned off the impacts in summer, i.e., not affecting the snow-free albedo. The simulation results show only minimal differences in the estimated ALT. That is, the winter surface albedo when using realistic satellite LAI products does not differ very much from that using the original LAI climatology. However, we speculate that large differences in summer could have significant impact on ALT estimation. We leave further investigation for future work. We have added one sentence in the manuscript to bring up this issue (section 5, page 22, line 27-29 in the revised manuscript).

> "Another issue affecting our ALT comparisons is the climatological representation of vegetation parameters such as LAI used in CLSM. Additional investigation (not

> shown) revealed large differences between the LAI climatology used in CLSM and more realistic, time-varying, satellite-based LAI products at several Mongolian sites."

Without any further information about the soil parameters in deep soils, we could not conduct further tests. Here we provided the results on the sensitivity of soil organic carbon and the vertical distribution in Figure R6. The figure reveals that a deeper ALT results from reducing the SOC content and from using a very different vertical distribution profile that arbitrarily concentrates less carbon in the top soil. Indeed, changing the vertical distribution profile for SOC content plays an almost equivalent role to reducing the SOC content. We added relevant discussion regarding this (section 5, page 22, line 20-23 in the revised manuscript).

> "Uncertainty in our description of soil organic carbon, i.e., both soil carbon content and vertical carbon distribution, leads to corresponding uncertainty in our ALT simulations. We indeed find a significant improvement in simulated ALT at several Mongolian sites when we arbitrarily impose less total soil carbon content and concentrate less soil carbon in top layers (not shown)."

This further confirms the reviewer's comment regarding the importance of vertical variation of soil properties. Thus, we added one sentence here to bring up the issue about vertical variation in soil parameters (section 5, page 22, line 23-25 in the revised manuscript).

> "Besides the vertical distribution of soil carbon, the vertical variation in other soil hydrological properties (e.g. soil texture, porosity, hydraulic conductivity, etc.) should also play a significant role since they all affect soil thermal conductivity and heat capacity."

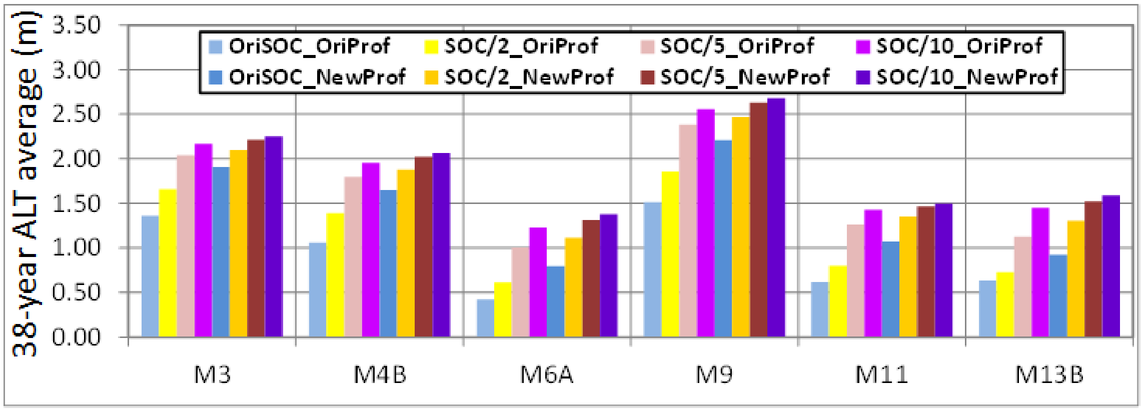


Figure R6: Simulation results at six Mongolian sites with different soil carbon contents and vertical carbon distributions. "OriSOC_OriProf" – Original soil organic carbon (SOC) content vertically distributed with the original profile as used in baseline simulation. "SOC/N_OriProf" – Reduced soil organic carbon content (by dividing the original SOC content by N) vertically distributed using the original profile. "OriSOC_NewProf" –

Original SOC content vertically distributed with a new profile which arbitrarily concentrates less carbon in top soils. "SOC/N_NewProf" – Reduced SOC content (by dividing the original SOC content by N) vertically distributed using the new profile.

Minor comments:

1. Fig. 4: The text in section 2.3 indicates most of the comparisons against AirMOSS data will be at the 4 flights in the Alaska north slope. So I wonder why COC flight was included in this figure esp. considering the AirMOSS retrievals were not included?

R2C9: We only have AirMOSS retrievals for IVO, ATQ, BRW and DHO, not for COC. However, our model provides results here, thus we included COC to add two additional measurements to compare with model results (Figure 4b).

2. Fig. 4 a-b: were the sites arranged according to the latitudinal changes?

R2C10: We arranged the sites aligning with the flight direction. We also added this into the caption of Figure 4.

> "The sites are arranged aligning with the flight direction."

3. Page 9, Line 21: Does MERRA-2 not provide air temperature at 2-m surface height?

R2C11: MERRA-2 does provide output of hourly 2-m air temperature. However, the land model within the MERRA-2 system is forced with the air temperature in the lowest (atmospheric) model layer (TLML), and the 2-m temperature is simply diagnosed from TLML and the surface temperature. For consistency, the off-line (land-only) model simulations presented here were likewise driven with TLML from MERRA-2. In any case, the sentence in question was not necessary and only caused confusion, so we deleted it.

4. section 3.2 and Figure 6: Part of the IVO flight lies in the Brooks mountain range with very low SOC content (Fig. 2c). It may be not a good representative of the average conditions in this area, at least considering SOC variability.

R2C12: The point at which we extracted the soil parameters does have a sort of intermediate SOC (greenish color in Figure 2c) which is also shown in Figure 5b. As mentioned in the original manuscript, however, we actually used an arbitrary intermediate SOC value which is 40 kg/m^2.

We also conducted two additional simulations using a very large and a very small SOC value everywhere along the AirMOSS transects. The results are shown in Figure R7 below. The "IntermC" used the same SOC that was used for homogenization (i.e., 40 kg/m^2). "LowC" and "HighC" used the lowest (10kg/m2) and highest (120kg/m2) SOC values found along the transects as shown in Figure 5b and also in Figure R7b. Figure R7a reveals that the model sensitivity to soil carbon is much larger for lower SOC than for higher SOC, and easily gets saturated for high SOC (i.e., larger than ~100 kg/m^2). However, all of this depends on the vertical soil carbon distribution profile used. Please also see R2C8d.

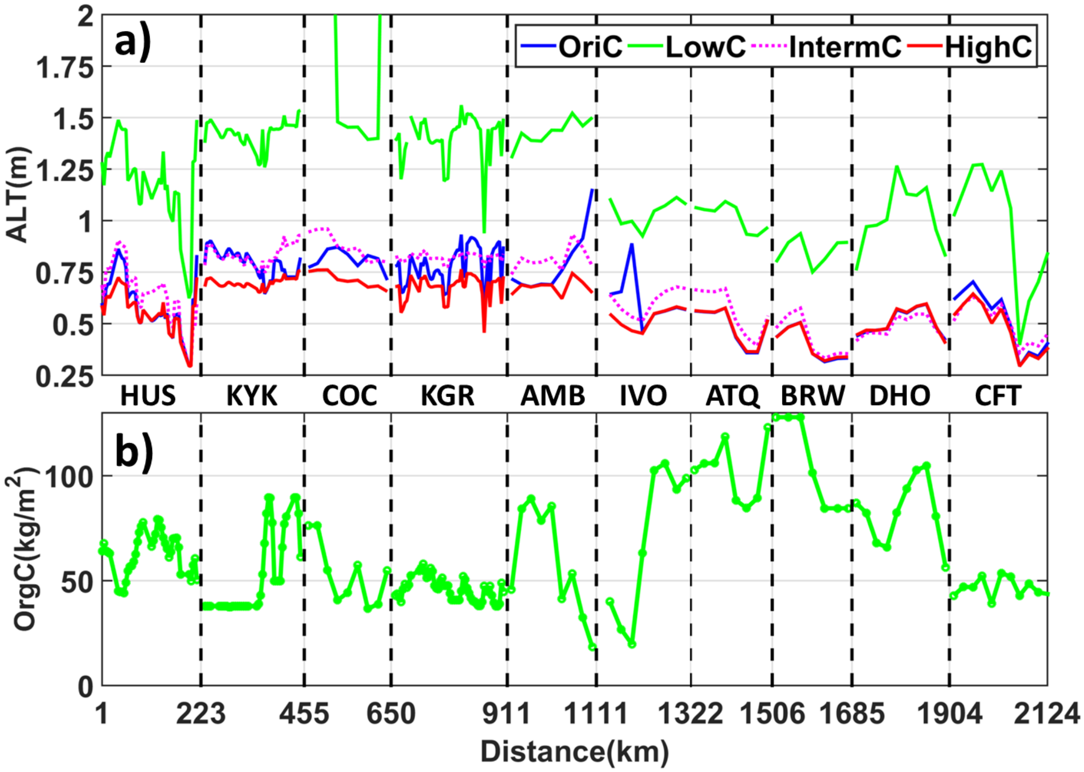


Figure R7: a) similar to the Figure 5a in the original manuscript but showing model sensitivity to organic carbon content along the AirMOSS flight transect. b) the organic carbon content along the transect.

We have added a sentence into the section 3.2 (page 10, line 8-10 in the revised manuscript).

> "Our investigation reveals that the model sensitivity to soil carbon content is much larger for lower SOC than for higher SOC, and easily gets saturated for high SOC (i.e., larger than ~100 kg/m^2) (not shown). Thus, we trust that 40 kg/m^2 is an appropriate value representing an intermediate SOC condition."

5. section 3.2: Would it be easier to follow if the description regarding the idealized experiments was included in the Methods section?

R2C13: We thank the reviewer for the helpful suggestion. We have introduced the idealize experiments in section 2 by adding a new Methods section (section 2.5).

6. Fig. 7 is not very informative. I suggest summarizing the results in a table.

R2C14: We feel that Figure 7 is informative, which is also supported by Reviewer #1 (R1C15). The figure displays the actual values and therefore also serves as a table. We therefore opt to keep the figure in its current form.

Reviewer #3

The authors compare model estimates (based on re-analyses data) of active layer thickness (ALT) with different data sources. This includes in situ measurements from an international database and an unpublished approach of airborne P-band estimation. The latter is available over Alaska. General evaluation is made for the northern hemisphere.

In addition, the model input is investigated for linear relationship with its output. The authors seek to find a simplified relationship to explain ALT change over time (driven by degree days, snow water equivalent etc).

General issues

The manuscript addresses an important topic, the modelling of active layer thickness over time. The presentation of the manuscript and setup of the experiment is however problematic.

1) The authors report 'reasonable' or 'good' results with limited and partially missing quantitative reasoning.

R3C1: We have included additional quantitative metrics to better illustrate our results. Please see our responses in R3C6, R3C8, R3C24, R3C29 and R3C31 for further details.

2) The results of the comparison to the unpublished airborne (AirMOSS) approach suggest that it actually does not work (AirMOSS values are all at the same level, not representing the in situ range, Figure 4). The comparison to this unpublished approach should be removed from the manuscript.

R3C2: We thank the reviewer for pointing out this. However, we do not agree with the reviewer regarding this point. We believe the AirMOSS data has value, which is also the opinion of Reviewer #2. We have expanded the relevant discussion about the potential use of the AirMOSS retrievals in the manuscript. The most important thing to note here is that the performance metrics for AirMOSS retrievals are quite good when evaluated within the instrument's sensing depth of approximately 60 cm (as shown in the new Table 3 in the revised manuscript), but are known to deteriorate for larger values of ALT.

Please see specifically our response to R1C11, R1C12 and R2C7, with particular attention to the Table R1 (new Table 3 added in the revised manuscript).

3) The stated objectives comprise issues regarding active layer thickness. The results section however does also cover permafrost extent.

R3C3: We thank the reviewer for pointing out this. We do not solely focus on ALT. We now modified the statement of our objectives in section 1 (page 4, line 24-27 in the revised manuscript):

> "Overall we pursue three scientific objectives: 1) evaluate the relative importance of the factors that determine the spatial variability of ALT, 2) evaluate CLSM-simulated ALT climatology and permafrost extent against observations, and 3)

quantify and assess the large-scale characteristics of ALT (in terms of means, interannual variability and trend) in Northern Hemisphere permafrost regions from 1980 through 2017. ”

4) In addition, common structure of methods, results and discussion is lacking. Methods are included in the introduction and mostly in the results section. The discussion is largely included in the results section. The ‘approach’ section mostly only includes the dataset description.

R3C4: We have added a new section (“section 3 Methods”) and moved all the method description to this new section. We also moved some discussion about our results to the last section to supplement other ‘discussion’ of the research findings.

5) Some of the used datasets are not appropriately cited. See detailed comments below

R3C5: We thank the reviewer for pointing out this. We have checked our reference list and correctly cited all of the datasets that we used. Please also see our response in R3C10 and R3C11 for additional details.

6) A main weakness of the manuscript is that the results are not discussed with respect to already published material on permafrost related parameter modelling and on the parameters which are investigated for explanation as drivers of ALT change over time. The role of temperatures and snow water equivalent for active layer thickness is known (what is presented as one of the main results in the abstract). The novelty of the paper is not clear.

R3C6: We have added extensive text that further highlights the novelty of our study. Please see our responses in R1C1 and R1C2. In the discussion below we focus specifically on the comparison between our results and other existing studies, for the reviewer’s interest.

Given that different models run at different spatiotemporal resolutions, and that the simulated results were evaluated using different observation data sets (or using different time periods of the same data set), a direct and credible comparison with existing model-simulated permafrost data sets requires extensive analysis and is not conducted here as it is considered beyond the project scope. Instead, we did a rough comparison with existing permafrost products simulated by land models with reanalysis-based forcing as shown in Table R4 below. (Note that ALT products simulated with climate forcing that is not observation-based are not included here.) We also want to emphasize that many existing studies define permafrost as areas with an ALT less than 3 m (Dankers et al., 2011) or 3.8 m (Guo et al., 2017;Guo and Wang, 2017;Lawrence et al., 2012), i.e., focused on the “near-surface permafrost”. We do not specifically limit ALT within a certain depth, although our simulated results generally are less than or equal to the bottom depth of the 5th soil layer (2.95m). Also, many of these studies evaluate the simulated results for “present-day” (i.e., before 2000). To make a relatively fair comparison with the existing studies, we used the same validating time period.

Table R4 – Existing available ALT products and evaluation metrics over Northern Hemisphere simulated by land surface models with reanalysis-based forcing.

Model and Soil Configuration	Forcing	Spatiotemporal Domain and Resolution	Findings and Conclusions Regarding ALT and Permafrost Extent	Reference
JULES (Joint UK Land Environment Simulator) (total soil depth 3m)	GSWP2 (Global Soil Wetness Project 2) WATCH (WATer and Global Change)	1983–1995 at 1°×1° resolution with GSWP2 covering areas north of 25°N (includes the Tibetan Plateau). 1959–2000 at 0.5°×0.5° resolution with WATCH covering areas north of 45°N (does not include the Tibetan Plateau).	- Captures 97% of the continuous and discontinuous permafrost areas. - Overestimates the total extent; simulates permafrost where it only occurs sporadically or only in isolated patches (25%); overestimates an additional 14% in areas permafrost free. ALT: JULES-simulated ALT is generally too deep compared with CALM observations: mean bias in the GSWP2 run (1990–1995) is 0.81±0.48 m, and 0.53±0.50m in WATCH-GPCC (1990–2000). The Root Mean Square Error (RMSE) is 0.94 and 0.73m, respectively.	Dankers et al. (2011)
ORCHIDEE-MICT land surface model (total soil depth is 38 m)	GSWP3 (Soil Wetness Project Phase 3) CRUNCEP (Climatic Research Unit - NCEP)	Northern Hemisphere (>30° N) at 1°×1° spatial resolution (does not include the Tibetan Plateau). 1901–2007 with GSWP3; 1901–2015 with CRUNCEP.	- Both simulations underestimate permafrost extent when using ALT<3m as the definition of permafrost; CRUNCEP-forced simulation shows better permafrost extent using an alternative definition of permafrost. (No actual bias value provided.) - GSWP3-forced model generally overestimates ALT by more than 1m; CRUNCEP-forced output shows relatively better agreement with the observations. (No actual bias value provided.)	Guimberteau et al. (2018)
CLM 4.5 (50m soil depth)	CRUNCEP (Climatic Research Unit - NCEP)	Northern Hemisphere (includes the Tibetan Plateau). 1901 – 2010	(Defines near-surface permafrost as that within the upper 3.8 m of soil.) - Simulated present-day (mean from 1981 to 2000) permafrost distribution; shows good agreement – only discrepancy is on the Tibetan Plateau. Bias in	Guo and Wang (2017)

<table>
<tr><td></td><td></td><td>Spatial Resolution: 0.5°×0.5°</td><td>permafrost extent is 2.02×10^6 km^2.

- Global ALT trend with a correlation coefficient and an The Nash-Sutcliffe Efficiency (NSE) of 0.73 and 0.21, respectively. The simulated trend was smaller than observed trend. (No evaluation for ALT climatology provided.)</td><td></td></tr>
<tr><td>CLM 4.5

(50m soil depth)</td><td>CFSR: NOAA Climate Forecast System Reanalysis

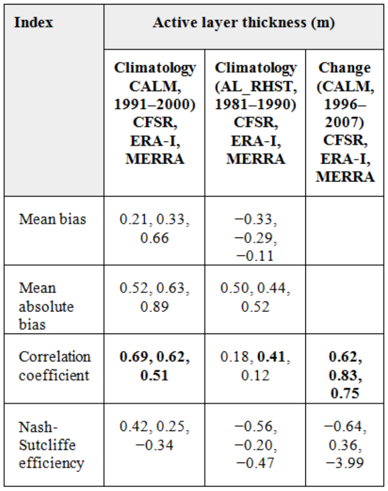
ERA-I: European Centre for Medium-Range Weather Forecasts Re-Analysis Interim

MERRA: NASA Modern Era Retrospective-Analysis for Research and Applications</td><td>Northern Hemisphere (includes the Tibetan Plateau).

Spatial Resolution: 0.5°×0.5°

1979 -2009</td><td>(Permafrost is defined as ground where monthly soil temperature is less than 0°C for 24 consecutive months in at least one layer of the upper 10 soil layers (3.8 m) (Lawrence et al., 2012).)

- The model underestimates the permafrost extent in southern Alaska, northern Western Siberian Plain, and over the Tibetan Plateau.

Specific evaluation metrics:

<table>
<tr><th>Index</th><th colspan="3">Active layer thickness (m)</th></tr>
<tr><th></th><th>Climatology CALM, 1991–2000) CFSR, ERA-I, MERRA</th><th>Climatology (AL_RHST, 1981–1990) CFSR, ERA-I, MERRA</th><th>Change (CALM, 1996–2007) CFSR, ERA-I, MERRA</th></tr>
<tr><td>Mean bias</td><td>0.21, 0.33, 0.66</td><td>−0.33, −0.29, −0.11</td><td></td></tr>
<tr><td>Mean absolute bias</td><td>0.52, 0.63, 0.89</td><td>0.50, 0.44, 0.52</td><td></td></tr>
<tr><td>Correlation coefficient</td><td>0.69, 0.62, 0.51</td><td>0.18, 0.41, 0.12</td><td>0.62, 0.83, 0.75</td></tr>
<tr><td>Nash-Sutcliffe efficiency</td><td>0.42, 0.25, −0.34</td><td>−0.56, −0.20, −0.47</td><td>−0.64, 0.36, −3.99</td></tr>
</table></td><td>Guo et al. (2017)</td></tr>
<tr><td>CLM3 and CLM4 (offline)

3.5m depth with CLM3 and 3.8m depth with CLM4.</td><td>Forced with observed meteorological data based on NCEP-DOE.</td><td>Northern Hemisphere (includes Tibetan Plateau);

Simulation period: 1980–1999;</td><td>(Focused on shallow-permafrsot with ALT < 3.8 m.)

- CLM3 simulation underestimates permafrost in southern and western Siberia.
- CLM4 simulations show improved permafrost extent.</td><td>Lawrence et al. (2012)</td></tr>
</table>

		0.9375° latitude × 1.25° longitude	- In general, ALT is overestimated; CLM4 underestimated ALT in regions of shallow permafrost. (No actual evaluation metrics provided.) - CLM3-simulated continuous and discontinous permafrost extent is 11.1 × 10^6 km^2 for the period 1970–89, which is below 11.8–14.6 × 10^6 km^2 estimate from Zhang et al., 2000. - CLM4-simulated continuous and discontinous permafrost extent is 13.7 × 10^6 km^2 for the period 1970–89.	

Existing studies that evaluated ALT that can be compared with ours include Dankers et al. (2011) and Guo et al. (2017). Other studies did not report specific evaluation metrics within a common simulation domain as ours, hence, they cannot be used for comparison. Comparing with reported evaluation metrics as shown in Table R5, Figure R8 and Figure R9 reveal that: 1) for the early 1990s (i.e., 1990 – 1995 period), our simulated results show a much better agreement with the observations than when using JULES with GSWP2 forcing as reported in Dankers et al. (2011); 2) for the whole 1990s (i.e., 1990 – 2000) period, our results are better than that using JULES with WATCH-GPCC forcing or when using CLM4.5 with both ERA-I and MERRA forcing as reported in Guo et al. (2017); and 3) our results show a smaller mean absolute bias and a larger correlation coefficient than that using CLM4.5 with CFSR, but show worse performance regarding mean bias and NSE. In addition, all of these existing studies overestimated ALT at the global scale and reveal a positive bias while our results underestimate deep ALT, and hence, reveal a negative bias. However, it is worth noting that our results demonstrate much better agreement against observations for regions of shallow permafrost as shown in Figure R8 and R9. We now added a short discussion along the lines of the following (section 4.4, page 18, line 18-20 in the revised manuscript):

> “Note that the existing literature on simulated ALT fields (e.g., Dankers et al. (2011), Lawrence et al. (2012) and Guo et al. (2017)) reveals a general tendency for models to overestimate ALT climatology at the global scale. Our results here suggest that the CLSM-simulated ALT fields are among the better simulation products, especially for shallow permafrost.”

Table R5 – Summary of evaluation metrics for ALT estimates reported in literature. The same metrics calculated with the simulation results in this study for an identical evaluation period are provided.

Evaluation metrics	(Dankers et al., 2011)* with two different sets of climate forcings				(Guo et al., 2017)[#] with three different sets of climate forcings			
	GSWP2 (1990–1995)	**This Study (1990–1995)**	WATCH-GPCC (1990–2000)	**This Study (1990–2000)**	CFSR (1991-2000)	ERA-I (1991-2000)	MERRA (1991-2000)	**This Study (1991-2000)**
RMSE (m)	0.94	0.18	0.73	0.72	None			
Mean Bias (m)	0.81	0.04	0.53	-0.17	0.21	0.33	0.66	-0.24
Mean Absolute Bias (m)	None				0.52	0.63	0.89	0.36
Correlation Coefficient					0.69	0.62	0.51	0.70
Nash-Sutcliffe Efficiency (NSE)					0.42	0.25	-0.34	0.25

*Evaluation was conducted with annual ALT.

[#] Evaluation was based on ALT climatology.

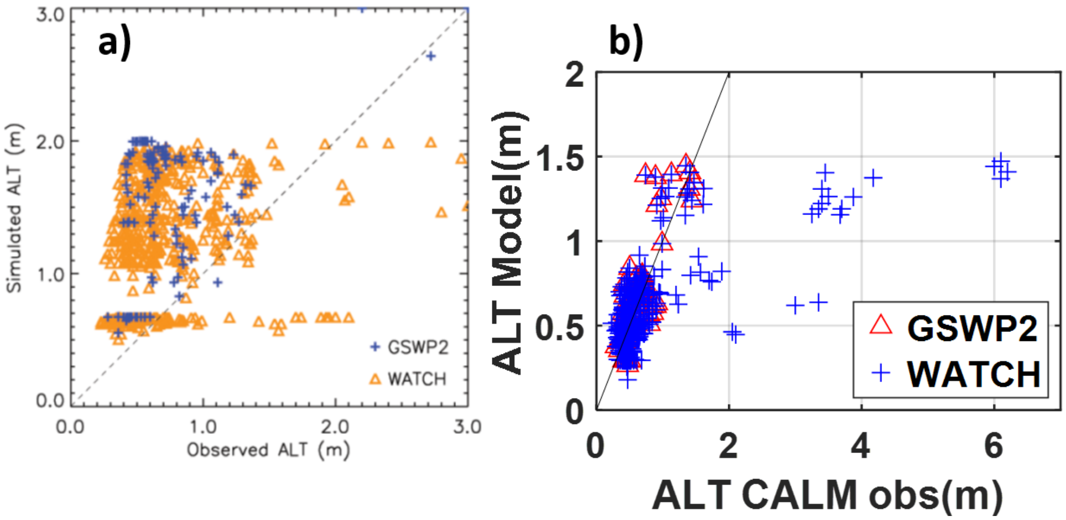


Figure R8: Observed and simulated ALT at CALM observations sites. a) copied from Dankers et al. (2011). The comparison was made for the period 1990–1995 using GSWP2 forcing, and for the period 1990–2000 using WATCH. b) Results from this study.

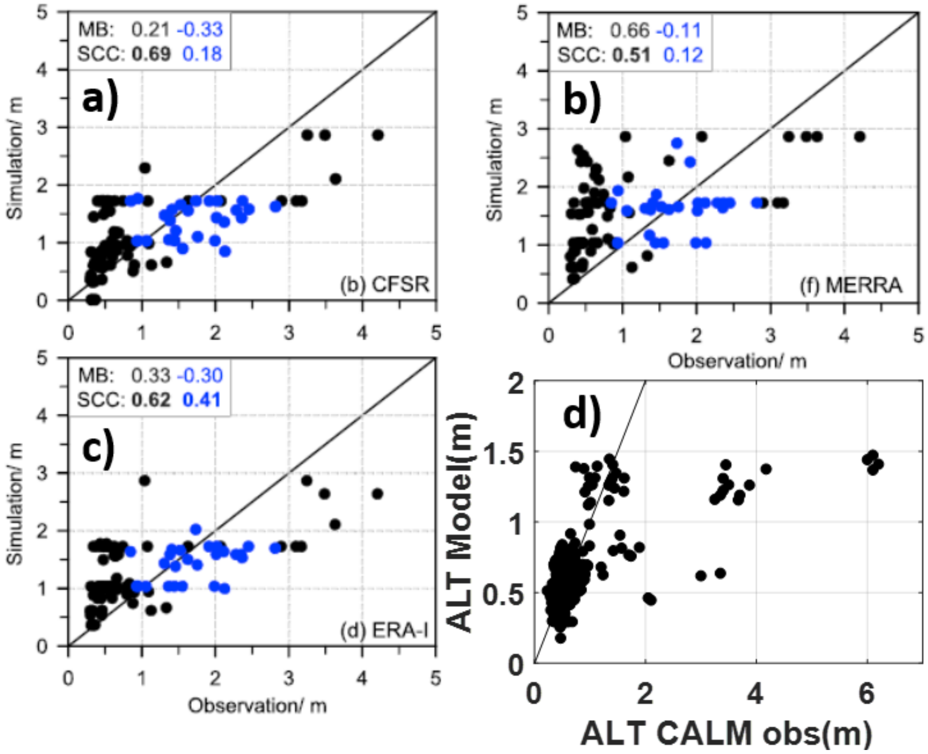


Figure R9: a), b) and c) are copied from Guo et al. (2017). Black dots represent the validating pair of data using CALM observations whereas the blue dots represent the validating pair of data using AL_RHST observations. d) shows comparison results with ALT estimates in this study. All subplots are for the validating period of 1990 to 2000.

In terms of the evaluation of simulated permafrost extent, although it is not fair to compare our results with these existing studies due to differences in simulation domain, we did conduct similar calculations to quantitatively evaluate our simulated permafrost extent against other studies. In addition, most of these existing studies compare the model simulated permafrost area with that of the total area of continuous and discontinuous permafrost from Brown's map due to their coarse spatial resolution (i.e., at least 0.5°). Given the high resolution (i.e., roughly 9km × 9km) of our simulation results, we compare and discuss our simulated permafrost area with that of the total area of continuous, discontinuous, **and** sporadic permafrost together as shown in Figure 10 in our manuscript. For the revised manuscript, we have incorporated these evaluation results into a new table, as shown below (Table 4 in the revised manuscript). We also compared the simulated permafrost extent with that of the total area of continuous and discontinuous permafrost area from Brown's map in order to better compare our results with those of existing studies, which is shown in the parentheses in Table R6 below (new Table 4 in the revised manuscript).

Table R6– Evaluation results for simulated permafrost extent. (New Table 4)

Case	CLSM	Obs.	Simulated Area ($\times10^6$ km^2)	Percentage Relative to Observed
4	No	No	48.8	-

3	Yes	No	1.9	-
2	No	Yes	3.2 (1.7)	18.7% (12.3%)
1	Yes	Yes	13.8 (12.3)	81.3 % (87.7%)

We also added the following short discussion into the revised manuscript (section 4.4, page 17, line 27-31 in the revised manuscript):

> "The specific areas of each type shown in Figure 10a are listed in Table 4. The simulated permafrost extent covers 81.3% of the observation-based area (i.e., the total area of continuous, discontinuous and sporadic permafrost regions), and misses 18.7% of the observed permafrost area. When comparing simulated permafrost extent with only continuous and discontinuous types, these metrics change to 87.7% and 12.3%, respectively. Meanwhile, the permafrost extent is overestimated by 3.2×10^6 km^2."

Regarding the role of temperature and SWE, although it is well known that these two factors affect permafrost, the *relative* contribution of the two is still unclear. This study attempts to quantitatively estimate this relative impact, which, to our knowledge, has heretofore not been done. This, we feel, is indeed one of our useful, novel contributions to the permafrost literature.

Detailed comments

Title: the title is misleading; the paper focuses on active layer thickness what should be reflected in the title

R3C7: We do not solely focus on ALT. Simulated permafrost extent is also evaluated and discussed in the manuscript. We believe our title regarding "variability" implicitly includes both ALT and permafrost extent. In addition, ALT might be too technical to be included in a title since many people know about permafrost but may not know what ALT is. We then retain our title as is.

Abstract:

Line 19: 'measurements demonstrates reasonable skill' – what is 'reasonable'? add numbers

R3C8: We have added the following sentence in order to clarify what we mean by "reasonable" skill (Abstract, page 1, line 18-19 in the revised manuscript).

> "Specifically, the RMSE (and bias) of climatological ALT is 1.22 m (and -0.48 m) across all sites and 0.33 m (and -0.04 m) without the Mongolia sites."

Line 27: 'significant degradation, with ALT increasing up to 0.5 cm/year' – it should be noted that this cannot be confirmed with in situ observations

R3C9: We thank the reviewer for pointing out this. This is very true. Below we provide the same figure as our Figure 14a in the original manuscript but only showing sites at which both the observed and estimated ALT trend is found to be statistically significant. Only four and ten sites are shown below with p value less than 0.05 and 0.10, respectively. Although the estimated and observed trends are closely clustered around the 1:1 line, the limited number of sites shown below cannot assure the accuracy of the estimated trends.

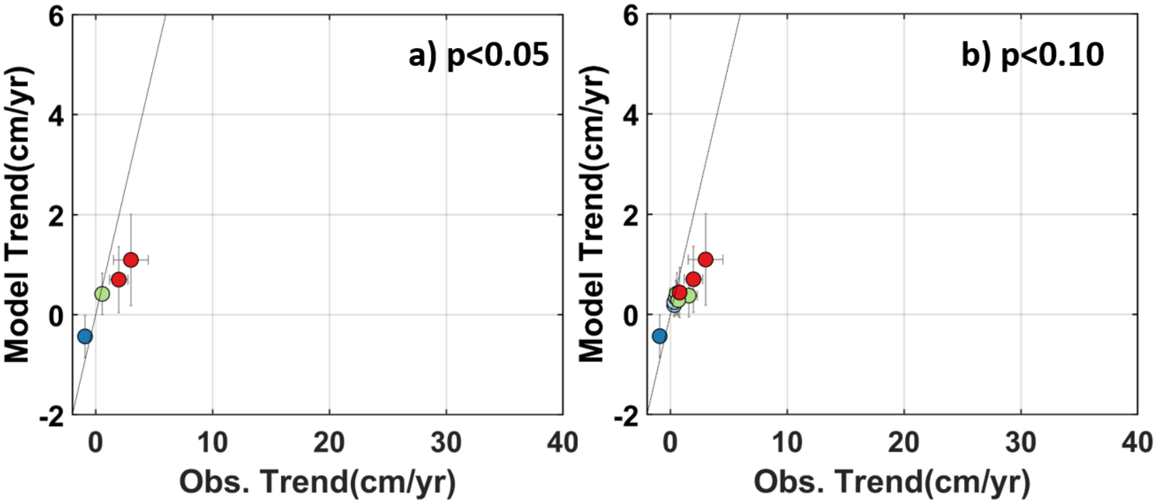


Figure R10: ALT trend from CLSM estimates vs. CALM observations, based on years common to both datasets within the period of 1990 to 2017. The horizontal and vertical error bars represent 95% confidence intervals of observed ALT trend (regression slope) and CLSM-simulated ALT trend, respectively. This current figure is similar to Figure 14 in the original manuscript, but only showing sites with both observed and estimated trends that are statistically significant at a) 0.05 and 0.10 level.

For the manuscript, we have added an additional description of the significance of estimated trends into the caption of Figure 14, and we also added the following to the discussion in the Abstract (page 1, line 28-29 in the revised manuscript):

> "It is difficult, however, to adequately assess the accuracy of the simulated ALT trends given the limited availability and relatively short records of in-situ measurements"

We also added a sentence into section 4.5 as shown below (page 19, line 30 to page 20, line 2 in the revised manuscript):

> "The overall comparison in Figure 14 is, in any case, highly uncertain, given the limited number of data points available to compute the trends. Note that only four (ten) points remain when screening out sites at which either the observed or estimated ALT trend is not statistically significant at the 0.05 (0.10) level. Simply put, the limited number of sites with meaningful trends cannot assure an accurate trend assessment."

Page 5, line 9: NCSCDv2 – citation missing, follow instructions of https://bolin.su.se/data/ncscd/

R3C10: We sincerely apologize for missing the important references. We thank the reviewer for pointing out this. We have added the references for NCSCDv2 (section 2.1, page 6, line 3-4 in the revised manuscript):

> "… from the Northern Circumpolar Soil Carbon Database version 2 (NCSCDv2, https://bolin.su.se/data/ncscd/) (Hugelius et al., 2013a;Hugelius et al., 2013b)."

Page 6, line 24: Brown et al. 2002: the reference is missing in the list

R3C11: We are sorry for failing to include this reference. We have added it.

Page 7, line 30 following: this paragraph belongs to methods

R3C12: Done as suggested.

Page 8, line 9: remove 'relatively'

R3C13: Done as suggested.

Page 9, first paragraph: this belongs to discussion

R3C14: Done as suggested.

Page 9, line 17 following: introduce this comparison in the methods section

R3C15: Done as suggested.

Page 9, line 22: provide correlation analyses results as table

R3C16: These results only consist of three correlation coefficients. Therefore, instead of adding these values into a table, we have added these correlation coefficients to Figure 5 for each variable for greater clarity.

Page 9, line 23: which SOC value did you use? Is it representative for the upper soil layer?

R3C17: We used spatially varying SOC provided by NCSCDv2, and not a constant SOC everywhere. Please see section 2.1 and Tao et al. (2017) for more details.

Page 10, line 4-28: this all belongs to the methods section

R3C18: We have moved these paragraphs to the new method section (section 3).

Page 11, line 7: spell out LAI

R3C19: We have spelled out LAI at the place it first appears (page 9, line 24 in the revised manuscript) and thus keep LAI as used here.

Page 12, line 6-12: move to methods

R3C20: Done as suggested.

Page 12, line 19: ‘permafrost areas shown in Fig. 1b are well confined within the cold side ..’ this cannot be seen. Add outlines to map

R3C21: The 273.15K isotherm actually is shown in the figure (i.e., the edge of dark green block). We have added the boundary containing continuous and discontinuous permafrost regions into Figure 8d. The original Fig. 8d showed mean Tair all over the Northern Hemisphere. We also modified it and only show mean Tair over the simulation domain in order to better illustrate the results.

Page 13, up to line 7: to methods; what is the reasoning for the static time period approach?

R3C22: We have moved the method description to the new section 3.3. We are not sure what the reviewer means by “static time period”. We evaluated the full 37-year period in order to best leverage all of the available observations. We also modified equation (1) and (2) in the original manuscript (equation (2) and (3) in the revised manuscript) as follows (section 3.3, page 10, line 28 to page 11, line 2 in the revised manuscript):

$$\text{“}T_{cum}(\mathrm{N}) = \sum_{t=1}^{t=M} T_{pos}(t) \;\; , \qquad (2)$$

where

$$T_{pos}(\mathrm{t}) = \begin{cases} T_{air}(t) - T_f & if\ T_{air}(t) > T_f \\ 0 & if\ T_{air}(t) \le T_f \text{’} \end{cases} \qquad (3)$$

The index t in equation (2) for year N starts with a value of 1 on 1 September of year (N-1) and ends with a value of M on 31 August of year N. M could be 365 or 366 depending on the presence of a leap year over the preceding annual period.”

Page 13 – move explanation of linear regression analyses to methods

R3C23: Done as suggested.

Page 14, line 14: ‘geographically thin disagreements’ – quantify this

R3C24: It is quite hard to quantify this, but we have attempted to do so with the following modification (section 4.4, page 17, line 22-25 in the revised manuscript):

> “Aside from western Siberia, the geographically thin disagreements (i.e., about a few degrees latitude) between the simulated and observed permafrost extents toward the south in Figure 10a (green and blue areas at the southern edge of permafrost regions) are not as much a concern, since the comparison in such areas is muddied by the interpretation of “isolated” permafrost in the observational map (Figure 1b).”

Page 14, line 28 following: this is methods

R3C25: We have modified and moved these sentences to the new method section (section 3.4).

Page 14, line 18: this error may seem small in absolute numbers, but ALT is much thinner than for the Mongolian sites. The error is still relatively large.

R3C26: We recognized this behavior, and hence we discussed the model performance in high-latitude regions separately from that in mid-latitude regions. We also explicitly emphasized the large error for the Mongolian sites and discussed the possible reasons for the underestimation. We conducted some new sensitivity tests and have added relevant discussion into the manuscript. Please see our response in R2C8d and R3C6 for details.

Page 14, line 20-26: this belongs to discussion

R3C27: We thank the reviewer for this suggestion, but we believe this paragraph should stay here since it describes the results.

Page 15, first two paragraphs of section 3.5 – this all belongs to methods

R3C28: We have modified and moved these paragraphs to the new methods section (section 3.4).

Page 17, 21: 'shows good general agreement' – this is not really the case, quantify the agreement and compare with other published results

R3C29: We calculated the bias in percentage (please see R3C6 for details), and we have modified as follows (section 5, page 21, line 21-23 in the revised manuscript):

> "The spatial distribution of CLSM-simulated permafrost shows general agreement with the observation-based permafrost map of Brown et al. (2002), capturing 81.3% of total areas of continuous, discontinuous and sporadic types while capturing 87.7% of the total area of continuous and discontinuous types."

Page 17, line 29: 'The simulated ALTs agree well with the in-situ observations' – not really, see my comment for page14, line18; how do your results compare to other published results?

R3C30: Our results are among the best existing permafrost products. In particular, our results demonstrate much better agreement against observations for regions of shallow permafrost (as shown in Figure R8 and R9) compared to other existing results. Please see details in R3C6.

Page 18, lines 1-2: 'retrievals from airborne remote sensing for 2015 and the corresponding simulated ALT exhibit reasonable accuracy vs. in situ measurements' – this is not clear from the material presented.

R3C31: Point taken. We have modified this sentence as follows (section 5, page 20, line 16-27 in the revised manuscript):

> "In northern Alaska, ALT retrievals from airborne remote sensing for 2015 and the corresponding simulated ALT exhibit limited skill versus the in-situ measurements.

> At the model scale, the mean bias for the simulated results is better (-0.05 m) than that for the retrievals (-0.12 m), but the opposite is true for the correlation coefficient against observations (0.27 for the model vs. 0.61 for the retrievals). At the in-situ site scale, however, the ALT retrievals show a very weak correlation coefficient with the observations (0.05). Excluding sites that have ALT measurements exceeding the radar sensing depth (~ 60cm), the evaluation metrics for ALT retrievals become better than that for simulated ALT at the model scale. However, the remotely sensed ALT estimates generally show lower levels of spatial variability than the simulated ALT estimates, and their spatial patterns differ considerably. The remote sensing approach is still relatively new, and many aspects still need to be worked out. It is important, though, to begin considering the modeling and remote sensing approaches side by side, as both should play important roles in permafrost quantification in the years to come. Indeed, once the science fully develops, joint use of modeling and remote sensing (e.g., through the application of downscaling methods) should allow the generation of more accurate permafrost products at even higher resolutions."

Figure 5: what is the red rectangle?

R3C32: The red rectangle highlights an example to illustrate the anti-correlated relationship between ALT and organic carbon content. We discussed this point in section 4.2. We also added the following sentence to the caption of this figure (page 37, line 6-7 in the revised manuscript):

> "The red rectangle crossing a) and b) highlights a portion of the domain that shows an anti-correlated relationship between organic carbon content and modelled ALT (see section 4.2)."

Figure 7: convert to table

R3C33: Since the values are already provided on the figure, we feel that converting it to a table would not help much. We opt to keep the figure as is.

Reference

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, Geophys Res Lett, 34, 10.1029/2007gl029536, 2007.

Bosilovich, M. G., Akella, S., Coy, L., Cullather, R., Draper, C., Gelaro, R., Kovach, R., Liu, Q., Molod, A., Norris, P., Wargan, K., Chao, W., Reichle, R., Takacs, L., Vikhliaev, Y., Bloom, S., Collow, A., Firth, S., Labow, G., Partyka, G., Pawson, S., Reale, O., Schubert, S. D., and Suarez, M.: MERRA-2: Initial Evaluation of the Climate, NASA Technical Report Series on Global Modeling and Data Assimilation, NASA/TM-2015-104606, Vol. 43, National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland, USA, 2015.

Bosilovich, M. G., Robertson, F. R., Takacs, L., Molod, A., and Mocko, D.: Atmospheric Water Balance and Variability in the MERRA-2 Reanalysis, Journal of Climate, 30, 1177-1196, 2017.

Brown, J., Ferrians, O., Heginbottom, J. A., and Melnikov, E.: Circum-Arctic Map of Permafrost and Ground-Ice Conditions, Version 2. [Permafrost Extent]. NSIDC: National Snow and Ice Data Center, Boulder, Colorado USA., 2002.

Chadburn, S. E., Burke, E. J., Cox, P. M., Friedlingstein, P., Hugelius, G., and Westermann, S.: An observation-based constraint on permafrost loss as a function of global warming, Nat Clim Change, 7, 340-+, 2017.

Chen, R. H., Tabatabaeenejad, A., and Moghaddam, M.: Retrieval of permafrost active layer properties using time-series P-band radar observations, IEEE Transactions on Geoscience and Remote Sensing. (In Review), 2018.

Cheng, L., and Zhu, J.: 2017 was the warmest year on record for the global ocean, Adv Atmos Sci, 35, 261-263, 2018.

Dankers, R., Burke, E. J., and Price, J.: Simulation of permafrost and seasonal thaw depth in the JULES land surface scheme, Cryosphere, 5, 773-790, 2011.

Gelaro, R., McCarty, W., Suarez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G. K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), Journal of Climate, 30, 5419-5454, 2017.

Gisnas, K., Etzelmuller, B., Farbrot, H., Schuler, T. V., and Westermann, S.: CryoGRID 1.0: Permafrost Distribution in Norway estimated by a Spatial Numerical Model, Permafrost Periglac, 24, 2-19, 2013.

Guimberteau, M., Zhu, D., Maignan, F., Huang, Y., Yue, C., Dantec-Nedelec, S., Ottle, C., Jornet-Puig, A., Bastos, A., Laurent, P., Goll, D., Bowring, S., Chang, J. F., Guenet, B., Tifafi, M., Peng, S. S., Krinner, G., Ducharne, A., Wang, F. X., Wang, T., Wang, X. H., Wang, Y. L., Yin, Z., Lauerwald, R., Joetzjer, E., Qiu, C. J., Kim, H., and Ciais, P.: ORCHIDEE-MICT (v8.4.1), a land surface model for the high latitudes: model description and validation, Geoscientific Model Development, 11, 121-163, 2018.

Guo, D. L., and Wang, H. J.: Simulated Historical (1901-2010) Changes in the Permafrost Extent and Active Layer Thickness in the Northern Hemisphere, Journal of Geophysical Research-Atmospheres, 122, 12285-12295, 10.1002/2017jd027691, 2017.

Guo, D. L., Wang, H. J., and Wang, A. H.: Sensitivity of Historical Simulation of the Permafrost to Different Atmospheric Forcing Data Sets from 1979 to 2009, Journal of Geophysical Research-Atmospheres, 122, 12269-12284, 2017.

Hugelius, G., Bockheim, J. G., Camill, P., Elberling, B., Grosse, G., Harden, J. W., Johnson, K., Jorgenson, T., Koven, C. D., Kuhry, P., Michaelson, G., Mishra, U., Palmtag, J., Ping, C. L., O'Donnell, J., Schirrmeister, L., Schuur, E. A. G., Sheng, Y., Smith, L. C., Strauss, J., and Yu, Z.: A new data set for estimating organic carbon storage to 3m depth in soils of the northern circumpolar permafrost region, Earth Syst Sci Data, 5, 393-402, 10.5194/essd-5-393-2013, 2013a.

Hugelius, G., Tarnocai, C., Broll, G., Canadell, J. G., Kuhry, P., and Swanson, D. K.: The Northern Circumpolar Soil Carbon Database: spatially distributed datasets of soil coverage and soil carbon storage in the northern permafrost regions, Earth Syst Sci Data, 5, 3-13, 10.5194/essd-5-3-2013, 2013b.

Jafarov, E. E., Marchenko, S. S., and Romanovsky, V. E.: Numerical modeling of permafrost dynamics in Alaska using a high spatial resolution dataset, Cryosphere, 6, 613-624, 10.5194/tc-6-613-2012, 2012.

Koven, C. D., Riley, W. J., and Stern, A.: Analysis of Permafrost Thermal Dynamics and Response to Climate Change in the CMIP5 Earth System Models, Journal of Climate, 26, 1877-1900, 10.1175/Jcli-D-12-00228.1, 2013.

Lawrence, D. M., and Slater, A. G.: A projection of severe near-surface permafrost degradation during the 21st century, Geophys Res Lett, 32, 2005.

Lawrence, D. M., and Slater, A. G.: Incorporating organic soil into a global climate model, Clim Dynam, 30, 145-160, 2008.

Lawrence, D. M., Slater, A. G., Romanovsky, V. E., and Nicolsky, D. J.: Sensitivity of a model projection of near-surface permafrost degradation to soil column depth and representation of soil organic matter, Journal of Geophysical Research-Earth Surface, 113, 10.1029/2007jf000883, 2008.

Lawrence, D. M., Slater, A. G., and Swenson, S. C.: Simulation of Present-Day and Future Permafrost and Seasonally Frozen Ground Conditions in CCSM4, Journal of Climate, 25, 2207-2225, 10.1175/Jcli-D-11-00334.1, 2012.

Nicolsky, D., Romanovsky, V., Alexeev, V., and Lawrence, D.: Improved modeling of permafrost dynamics in a GCM land-surface scheme, Geophys Res Lett, 34, 2007a.

Nicolsky, D. J., Romanovsky, V. E., Alexeev, V. A., and Lawrence, D. M.: Improved modeling of permafrost dynamics in a GCM land-surface scheme, Geophys Res Lett, 34, 2007b.

Peng, S., Ciais, P., Krinner, G., Wang, T., Gouttevin, I., McGuire, A. D., Lawrence, D., Burke, E., Chen, X., Decharme, B., Koven, C., MacDougall, A., Rinke, A., Saito, K., Zhang, W., Alkama, R., Bohn, T. J., Delire, C., Hajima, T., Ji, D., Lettenmaier, D. P., Miller, P. A., Moore, J. C., Smith, B., and Sueyoshi, T.: Simulated high-latitude soil thermal dynamics during the past 4 decades, Cryosphere, 10, 179-192, 10.5194/tc-10-179-2016, 2016.

Reichle, R. H., Draper, C. S., Liu, Q., Girotto, M., Mahanama, S. P. P., Koster, R. D., and De Lannoy, G. J. M.: Assessment of MERRA-2 land surface hydrology estimates, Journal of Climate, 0, null, 10.1175/jcli-d-16-0720.1, 2017a.

Reichle, R. H., Liu, Q., Koster, R. D., Draper, C. S., Mahanama, S. P. P., and Partyka, G. S.: Land Surface Precipitation in MERRA-2, Journal of Climate, 30, 1643-1664, 10.1175/jcli-d-16-0570.1, 2017b.

Sapriza-Azuri, G., Gamazo, P., Razavi, S., and Wheater, H. S.: On the appropriate definition of soil profile configuration and initial conditions for land surface-hydrology models in cold regions, Hydrol Earth Syst Sc, 22, 3295-3309, 10.5194/hess-22-3295-2018, 2018.

Tao, J., Reichle, R. H., Koster, R. D., Forman, B. A., and Xue, Y.: Evaluation and Enhancement of Permafrost Modeling With the NASA Catchment Land Surface Model, Journal of Advances in Modeling Earth Systems, 9, 2771-2795, 10.1002/2017MS001019, 2017.

Wang, W., Rinke, A., Moore, J. C., Cui, X., Ji, D., Li, Q., Zhang, N., Wang, C., Zhang, S., Lawrence, D. M., McGuire, A. D., Zhang, W., Delire, C., Koven, C., Saito, K., MacDougall, A., Burke, E., and Decharme, B.: Diagnostic and model dependent uncertainty of simulated Tibetan permafrost area, Cryosphere, 10, 287-306, 10.5194/tc-10-287-2016, 2016a.

Wang, W. L., Rinke, A., Moore, J. C., Ji, D. Y., Cui, X. F., Peng, S. S., Lawrence, D. M., McGuire, A. D., Burke, E. J., Chen, X. D., Decharme, B., Koven, C., MacDougall, A., Saito, K., Zhang, W. X., Alkama, R., Bohn, T. J., Ciais, P., Delire, C., Gouttevin, I., Hajima, T., Krinner, G., Lettenmaier, D. P., Miller, P. A., Smith, B., Sueyoshi, T., and Sherstiukov, A. B.: Evaluation of air-soil temperature relationships simulated by land surface models during winter across the permafrost region, Cryosphere, 10, 1721-1737, 10.5194/tc-10-1721-2016, 2016b.

Yi, S. H., Woo, M. K., and Arain, M. A.: Impacts of peat and vegetation on permafrost degradation under climate warming, Geophys Res Lett, 34, 2007.