Reviewer reply for Permafrost Physics

Dear Dave Lawrence,

We wish to thank our three reviewers and Executive Editor David Ham for their time and considered comments on our manuscript. Below we lay out their comments in full and include our replies in bold font. In response to reviewer comments the revised manuscript has been significantly trimmed down. Some material, which was less essential to the main discussion, has been moved to supplementary material, one figure has been removed and another moved to the appendix. We believe the paper now presents the main points in a more concise manner while still retaining the relevant details.

Executive Editor David Ham

I am writing as an executive editor of GMD to highlight an issue with the code availability section which needs to be remedied in the revised manuscript. I appreciate the lengths to which you have gone to identify the exact versions of the code corresponding to the experiments in the manuscript. There remain two concerns.

Code only on GitLab

The code reference is to a GitLab repository (in a personal account at that), and the individual tags are git tags. This doesn't provide the persistence required of the data in a journal paper. If the project moves to a different revision control system, or even
just to GitHub then the links will go dead. For this reason, the code associated with each of your tags needs to be persistently archived. Many authors find Zenodo (https://zenodo.org) a good solution for this, it basically comes down to uploading a tarball
and metadata (or you could push a copy of your repo to GitHub and use the automated GitHub-Zenodo integration). Note that you are still positively encouraged to provide a link to the preferred download location for your software (which might well be GitLab),
but this should be in addition to citing persistent archives.

We have deposited our code on Zenodo. The link is provided in the revised MS.

Insufficient documentation to reproduce

When I click through to your GitLab page, I can't find documentation which would tell me how to build the model and run the experiments in the paper. The readme file claims to contain links to documentation, but these are broken. A link to a wiki is also not very persistent: would a user coming to this after the project has ended be able to work out how to re-run the experiments in this manuscript? Please ensure that both your code and documentation are preserved in persistent, public archives.

The manual is within the code repository, not the gitlab wiki. Within the code repository it is /documentation/html/index.html. Since we structured the code and documentation this way, the manual is always part of the codebase. Our documentation is then also part of the Zenodo repository. We have made this clear in the revised manuscript.

Anonymous Referee #1

General

This paper describes improvements and tests of the CLASS-CTEM model under 18 different model configurations and input data sources. The results are compared with site observations of active-layer thickness (ALT) and ground temperatures, and estimates of permafrost extent and snow water equivalent (SWE). Such systematic tests demonstrate the effects of different configurations and data sources on the model behaviour,therefore, it provides basis for its improvements. Such information is also useful for the improvement of other land surface models in simulating permafrost regions. The topic of this paper is a good fit of the scope of the journal, and it is worthy to be published. Following are some suggestions, which may be useful to the improvement of its readability.

Major points

1. As the paper has indicated, the major issue is the sub-grid heterogeneity. The site conditions of the observations can be very different from the grid average used in the model. The input climate data for the model are also different from the climate at the observation sites. Thus, the nature of the modelled grid cells and the observation sites are quite different things. However, these multiple site observations are the data currently available, and could somewhat represent the conditions of the grid cell. Therefore, the approach used in this study is useful to check and improve the overall behavior of the model across the permafrost domain. The model is better constrained overall but not necessarily validated. Sections 3.12 and 3.13 indicated this issue. I think this is a limitation or shortcoming of the methodology rather than an excuse of the modelled bias (e.g., the authors seem to argue that the model can represent the reality better (Line 25, page 27)). Such limitation should be indicated clearly and to frame the assumptions/methods better.
   Determining how well a model performs for the physics of simulating permafrost processes is challenging. Due to the slow response times of permafrost thaw at depth there presently exist no long term observations of both meteorology (for model forcing) and soil temperatures (for evaluation) that are of suitable length to allow a proper test of the model performance. As a result it is difficult to get around the two linked problems raised by the reviewer: sub-grid heterogeneity in both the model drivers and the site-level conditions that we are hoping to simulate. We have attempted to demonstrate how sub-grid heterogeneity could impact our ability to evaluate our model (Section 3.12). Based on our analysis at the GTN-P borehole sites and the SPSMPS borehole cluster, we find that the influence of sub-grid heterogeneity makes it difficult to further decrease model bias with the observational data available. An alternative methodology could be to run the model at a single site and do an in-depth analysis of the model performance against the observations being careful to set the model up to the conditions at the site (soil textures, vegetation, permeable depths, etc.). However, as we state in our Introduction (p 3 l 25), model performance at a single site is not necessarily indicative of performance over large regions. So while we agree that our methodology does not allow a perfect validation of the model's performance, we believe it allows the most robust estimate of the model performance with the available observational information. Note also that we are not attempting to use our attempt to quantify sub-grid heterogeneity effects as an excuse for any model bias but rather we are attempting to contextualize the bias and bring that through to the interpretation of our results.
   To clarify, we don't state that our model can represent the reality better, but rather we state:

"In summary, it is likely that a slightly positive model bias, i.e. higher temperatures and greater ALT simulated than observed, would correspond to a model that best represents reality."
This statement suggests that due to the factors causing sampling bias in the ALT and borehole observations (as outlined in Section 3.13), a perfect model would simulate slightly higher temperature and greater ALT than the observations.

2. The results show that the simulated ALT is improved but no significant improvement in ground temperatures according to Figure 5. Since ground temperature is the principal state variable of the model, it would be useful to provide more information about it. Figure 4 is a nice demonstration of the improvement of the modelled ALT. Similar figures should be provided for the ground temperatures about their bias distributions. If they take too much space, you can put some of them in the Appendix.
Actually Figure 5 demonstrates that across all seasons and all depths the model simulated soil temperatures for the SoilGrids simulation are improved compared to the Base model (SoilGrids+Moss is only at certain depths/seasons). There are 105 106 borehole observations from 132 GTN-P sites used in this evaluation. An additional look at the ground temperatures is provided by Figure 6 with Gaussian KDE for the GTN-P sites. A Gaussian KDE is a similar demonstration of bias distribution to that presented in Figure 4 however the KDE plots are made up of many more data points as there are more borehole observations available than ALT. For example, the 0.05 - 0.5 m depth KDE contains 47 079 data points. Because the KDE becomes too difficult to parse with too many experiments plotted we have restricted it to only a few experiments being shown.
3. The paper is too long and some analysis is not very concise. Some parts of the analysis can be reduced, such as latent heat, sensible heat, and albedo as there are no observations for these variables anyway. Even the discharge shown in Figure 9 is not necessary. It is better just focusing on ALT, ground temperature and permafrostextent.
We agree and have streamlined the revised MS. The variables without observations for comparison have been moved to a supplement. We have retained the discharge as there are observations for comparison and several of the changes influence hydrology so this is a means to look at their impact.

Minor points

1. P.1, Lines16-18: “with seasonal . . .at the sites”. Not clear.
We have attempted to improve the clarity by rewording from: *"with seasonal wMAE values for the shallow surface layers of the revised model simulation at most 1.2 (^{\circ})C greater than those calculated for the model driving screen-level air temperature compared to observations at the sites"*, to: *"with seasonal wMAE values for the shallow surface layers of the revised model simulation at most 3.7 (^{\circ})C, which is 1.2 (^{\circ})C more than the wMAE of the screen-level air temperature used to drive the model as compared to site-level observations (2.5 (^{\circ})C"*
2. P.1 Lines 18-19: “Sub-grid heterogeneity estimates were derived from the standard deviation of ALT on the 1 km2 measurement grids at the GTN-P ALT sites”. Its sounds like you estimate the sub-grid heterogeneity for all the regions. That is not the case.

Unfortunately, we don't quite understand this comment. That sentence reads in full: "*Sub-grid heterogeneity estimates were derived from the standard deviation of ALT on the 1 km(^2) - 1 ha measurement grids at the GTN-P ALT sites, the spread in wMAE in grid cells with multiple GTN-P ALT sites, as well as from 35 boreholes measured within a 1200 km(^2) region as part of the Slave Province Surficial Materials and Permafrost Study.*". The sentence thus explicitly states where the sub-grid heterogeneity estimates were derived from. Note that now we clarify that we used both 1 ha and 1 km(^2) sampling sites.

3. P2. Line 12-13. "Since the carbon stored in frozen soils is only accessible to microbial respiration once soils thaw". Soil respiration has been observed when soil is frozen although it is low. The word "only" is too restrictive.
   Indeed, reworded to: "*Since the carbon stored in frozen soils becomes readily accessible to microbial respiration once soils thaw*"
4. P3, Line 5: "that four be considered", Four types parameterizations?
   Yes, parameterizations. Original sentence: "He investigated 15 alternative parameterizations relating to the model physics and concluded by recommending that four be considered", for better clarity rephrased to: "He investigated 15 alternative parameterizations relating to the model physics and concluded by recommending that four of those be considered"
5. P.4, Line 33: "configuration. (e.g", delete the '.'
   Thanks, removed the comma
6. Page 5: "The first seven experiments" in line 5 is too far away from "the second series of experiments" in line 32. It is better to put them closer.
   Good suggestion. We have added to line 32: *"Whereas the first series of experiments just described investigated aspects of the model setup, the second series of experiments investigates alternative parameterizations and uses the 'SoilGrids+Moss' experiment as a starting point"*
7. Page 7, Line 33: It is only for ground temperature which is converted to monthly averages? ALT is the annual maximum thaw depth. "The closest CLASS-CTEM grid cell to the ALT site's location" Why it is closest to a grid cell not within a grid cell?
   ALT is the active layer thickness. This does not imply that it is the annual maximum. Many of the sampling campaigns for ALT had one or two samples per year. It is possible they were unable to time the annual maximum of ALT in their sampling thus we compared the model to observations on a monthly basis to be most comparable. Regarding the closest/within question, we have changed the sentence for clarity to: "The closest grid cell was determined from the centre of the model grid cells to the ALT sampling location and ..."
8. Page 9, Line 24: delete the repeating "are'.
   Thanks, done.
9. Page 10, Lines 11-12: "but comparing . . .. Poor agreement", not very clear.
   Reworded to: "*Owing to the coarseness of the model grid CLASS-CTEM is not able to simulate isolated or sporadic permafrost. For regions of discontinuous and continuous permafrost, comparing the estimated distribution of \citet{Brown1997-un} to the modelled ALT indicates poor agreement.*"

10. Page 14, Fig. 3: The Y-axis is called ‘residual’. Is it the difference between simulated and observed ALT as indicated in the text? If so, it would be clearer to indicate that. The X-axis is ‘Ground depth (m)’. I think it is active-layer thickness although not sure it is modelled or observed. If so, it is better to say that.
    Another reviewer found this figure confusing. Since it was not fundamental to our paper, we have removed it.
11. Page 15, Lines 6-11 (even to line 18). It is better to put this paragraph to the section 2.2 (study design).
    We prefer our original placement as it provides a literature overview of previous work that has demonstrated the importance of increasing the ground column depth and number of ground layers.
12. Page 6, Figure 6. The top line “Depth = 0.05 – 5m |Season = JDF . . .”, 5m should be 0.5m.
    Yes, thank you. Corrected.
13. Page 21, Figure 7. The first sentence of the caption is not clear.
    We have attempted to improve clarity by rewording from: "*ALT differences (meters) for experiments that are based on the model setup of 'SoilGrids+Moss' (see Table \ref{explist}) compared to the 'SoilGrids+Moss' simulation.*" to: "*Differences (meters) between the 'SoilGrids+Moss' simulated ALT and the ALT simulated by the alternative parameterization experiments (based on the model setup of 'SoilGrids+Moss', see Table \ref{explist})*
14. Page 27, Line 25, The word ‘best’ is no appropriate.
    See our reply to Major Point #1.

Anonymous Referee #2

The manuscript presents an improved model version of the Canadian CLASS-CTEM model with respect to permafrost physics. The authors have done a great and extensive job dealing with the uncertainties of heat transfer within cold soils. Several tests were performed to see the optimised results and compared to observational datasets. The improved model version is a valuable formulation to be used in offline and coupled simulations. The analysis in the manuscript can also help identify other modelling groups for better physical formulations. The topic and the presentation fits the journal’s
scope, yet I have some minor suggestions to the authors to make the paper bit more easy to read through:

1. The extent of statistical analysis is way too long in the manuscript. I strongly suggest to move some of them to supplementary materials to make the actual paper more on point and show the most optimal formulations inside the main manuscript.
   The revised MS has been streamlined. We have moved the De Vries experiment to the supplement, moved the variables without an observational constraint to the supplement, and made the text more concise.
2. I agree with the authors to focus on the big scale improvements rather than grid point based comparisons but to actually identify the process improvements, it would be useful to show two or three selected grid points and compare the surface (∼10 - 20cm) soil temperature time series for

different experiments in addition to the borehole temperature comparisons in fig 11.
While we are glad that the reviewer appreciates our approach to focus on large scale improvements and comparisons across many different locations, we don't feel that showing surface changes at a few selected grid points would offer much additional insights. Additionally the paper, as mentioned by all reviewers, is already too long. Adding a new figure showing changes at shallow depths, with accompanying text for analysis, would add significant length to an already long manuscript.

3. To better quantify the snow pack improvement process, it would also help to show comparisons of snow depth with the observed values (if it exists). Since snow insulation plays a major role in freeze/thaw periods, the simulated snow depth should be investigated.
We agree that snow pack changes are important. The snow pack improvements in our manuscript have been suggested by earlier detailed studies and here, only their effect on permafrost is evaluated. This is why we have used what we believe to be the best available snow product (Blended5-SWE) to evaluate modelled changes in SWE. We are not aware of any large-scale snow depth products. However, snow depth and SWE are linked so looking at SWE from the model versus observation-based datasets is still valuable.

Anonymous Referee #3

In this paper the authors test and evaluate a wide range of improvements to permafrost physics in the CLASS-CTEM land surface model (which is part of the Canadian Earth System Model, CanESM). While there is nothing especially ground-breaking, this is a comprehensive and thorough assessment bringing together many different, disparate developments into a single framework and I believe is worthy of publication.

CLASS-CTEM consists of two components: broadly, CLASS does the physical calculations and CTEM performs the carbon cycle calculations. The authors improve the simulation of permafrost physics in CLASS-CTEM with a series of model developments which are successively evaluated (against multiple observations) and discussed. The default soil scheme has only three vertical layers which leads to a poor simulation of permafrost dynamics. Therefore, very reasonably, the first improvement extends the soil column and adds more layers. A moss layer is added to the surface, as this has previously been shown to improve simulation of soil temperature and freeze/thaw dynamics. Furthermore, the authors experiment with the depth to which the water can penetrate in the soil, and the impact of different driving data sets. Having established a baseline simulation based on what they consider to be these essential improvements, the authors then test a number of further developments in the representation of snow, hydrology and heat transfer. This includes, for example, allowing the presence of liquid water below zero degrees celsius ('supercooled' water), as in real life soils. Of the developments tested, this is considered to make the greatest improvement and therefore to be incorporated into the standard model version. The final simulation of permafrost by CLASS-CTEM is an improvement on the initial simulation, although the capacity for evaluation is somewhat limited by the disparity between the extremely large grid cell size in comparison to site-level observations. While this paper is relatively clear and well-written, I believe that it can and should be significantly improved prior to publication.

GENERAL COMMENTS

- In terms of experiments included, since de Vries thermal conductivity is considered to be physically unrealistic, I don’t see the need to include this experiment.
  We have moved most of this to the supplement. While we agree it is not a viable parameterization we wish to include its evaluation as it was recommended for incorporation in CLASS-CTEM by a previous study so we want to demonstrate it is not a suitable addition.
- The paper is a bit too long and the clarity could be improved. The statistics don’t need to be written out in so much detail in the text. As an example, the section 3.11 for ’modified hydrology’ could be reduced to something along the lines of: "This development generally reduces water mobility (eqs. A35/A36), resulting in wetter soils, which in turn leads to a significantly deeper ALT but minimal impact on soil temperature errors (see Figures 2,4 and 5). Overall it increases the spring runoff and reduces winter runoff, which degrades model performance compared to runoff observations (Figure 9). Therefore, this did not provide any significant improvements." [of course this could be better written, just a suggestion for what info to include -ie numbers not necessary in text, and only including the most crucial points] I would then also suggest adding a comment on the process and why it doesn’t work, for example my initial thought might be that while it may be a more realistic parameterisation, the fact that the model does not account for macropore flow / large scale cracks and defects in the ground means
  that the default parameterisation may be providing a compensation for this. In general I would like to see more reference to the processes and to the direction of the biases, rather than to absolute error, as this gives clearer information about the processes and why each development is having a particular effect. I might suggest to include a directional error in the error tables (bias in annual mean?) as well as just MAE.
  In response to the comments are length and clarity, the revised MS has been streamlined with material that is informative, but not necessary to convey the main messages moved to supplementary material. Concerning further commentary on why certain processes were successful or not in improving model performance, this is difficult to convey convincingly. To illustrate, using the same example of the modified hydrology experiment, we are not able to definitively state whether the proposed modifications are indeed more realistic based on our results presented here. While we may have a general opinion that the parameterization may be more realistic and the model could have a compensating error, we are not able to provide proof within the context of this paper. As a result it is difficult to include commentary such as that while lacking a means or simulations to definitively justify our statements. We investigated the idea of a directional error in the error tables however we found the KDE plots to give a much clearer picture of the full spectrum of model bias, i.e. calculating a directional bias tended to give a muddy picture due to cancelling positive and negative biases whereas KDE could show the spectrum of biases without that issue.
- I don’t think detailing the heat flux and LAI for each experiment is helpful to the main message. You could summarize heat fluxes and LAI as a separate section, for example noting that in warmer, wetter simulations, the LAI is generally larger.
  These variables have been moved to supplementary material to streamline the manuscript.
- In Section 3.13 I suggest adding a further comment on the potential ALT biases. I have noticed from using ALT data that the maximum that depth in the datasets is sometimes not the actual end of

season maximum because the field campaigns have not continued right until the end of the season, so this could be the cause of apparently discrepancy between soil temperatures and ALT's.
This is a good point. We were aware of this possibility thus we used the same month of ALT measurement to modelled month in our simulations for comparison purposes which minimizes this potential error (as mentioned in our response to Reviewer #1 as well).

- It appears that most of the developments are detailed in the appendix, but it doesn't appear to include details of the moss parameterisation. Please add something about the changes to thermal/hydraulic parameters for reproducibility (apologies if I missed this).
As the moss parameterization is fully detailed in Wu et al. 2016 we had originally referenced that paper. However, for clarity and to be in line with the other parameterizations we test, as suggested, we have now added a brief overview to the appendix.
- I would expect 'mean absolute error' to be higher for the air than for the soil, because variations in air temperature are typically much larger (both seasonally and on short timescales). Therefore I don't completely agree with the argument on page 27 lines 3-9.
One aspect that influences these calculations is temporal averaging. We compute the wMAE based on monthly mean values. The use of monthly means would greatly reduce the impacts of variability. If we did the same comparison at higher time resolutions, we agree that the higher variability of the air temperatures could become a factor.

FIGURES

- I find Figure 3 quite confusing and not entirely helpful. It is not clear what either of the axes are: The residual I guess is obs-model, but it's not specified in the text, and it's not at all clear what the x-axis is showing. Is this the observed ALT? If so I think it would be better, and show the same information, if you just to plot model ALT against observed ALT. I am also not entirely convinced by all the discussions that relate to that figure in terms of 'biases with depth', since there are altogether not many points on the plots and the trends do not appear to be very strong. Perhaps these discussions could be removed or modified to focus on the more significant impacts of the developments. For example, the 20 layer simulation has in general too deep ALT- would be enough information.
Since it was peripheral to the main message of the paper, and appeared to confuse roughly 66% percent of readers (Reviewer #1 found it confusing as well), we have removed it.
- Figure 7 doesn't add much. We can see basically the same thing but more clearly on Figure 4. If it is to be included, I would suggest in the appendix or supplementary.
We have moved it to the supplement.
- Figure 11: I am confused as to why the simulation with moss is warmer than the simulation with no moss. Are these labelled correctly?
Thank you for catching this, they were switched. This is now corrected.

MINOR COMMENTS

- "with seasonal wMAE values for the shallow surface layers of the revised model simulation at most 1.2C greater than those calculated for the model driving screen-level air temperature compared to observations at the sites" - This sentence is difficult to
parse. Perhaps by removing the last part ("compared to observations at the sites"), it would make

more sense.

We have taken your suggestion and reworded to: *"with seasonal wMAE values for the shallow surface layers of the revised model simulation at most 3.7 (^{\circ})C, which is 1.2 (^{\circ})C more than the wMAE of the screen-level air temperature used to drive the model as compared to site-level observations (2.5 (^{\circ})C"*

- Page 7 line 14: How is the 1851-1900 part different from just doing two more spinups? Is the CO2 varying? CO2 forcing dataset is not mentioned but should be included - please add.
  Both CO2 and land cover change during that period. We have added explanatory text around this to the revised MS. Both of which have little impact on permafrost physics which is why we neglected it in the first draft.
- Page 8 line 10: "filtered out"- not sure what that means? Does it mean removing the same values from model as are missing in observations?
  The missing values were removed from further consideration. We have removed this statement as this processing step isn't actually required since missing values aren't considered in calculating monthly means.
- Please include how you define permafrost presence in a grid cell (again, apologies if I missed this).
  This is on page 7: *'Active layer thickness in CLASS-CTEM is determined by the temperature and water content of the ground layers. If a layer's temperature is 0(^\circ)C, the frozen water fraction is used to estimate the thickness of freezing within the layer, i.e., if half of the water content in the layer is frozen, the ALT is assumed to be halfway through the layer. Permafrost area in the model domain was calculated by selecting grid cells with active layer thicknesses less than the model total ground column and multiplying by the grid cell area.'*
- Page 9 line 31. 'more grid cells' -> 'additional grid cells' (for clarity)
  Corrected
- Page 13 line 7-8. I would expect the absence of water and latent heat to make active layer too deep, because latent heat suppresses the seasonal cycle. But since latent heat works equally in both directions (applicable both in freezing and thawing periods),
  I am not convinced that its absence would lead to warmer soils. Indeed I am not seeing a major bias in soilgrids vs 20 layers on Figure 6. Perhaps reconsider this sentence.
  This section of the manuscript is comparing the 3 layer model setup to the 20 layer model setup - not yet comparing the 20 layer setup to model configurations with deeper permeable depths. The statement 'The absence of water and therefore of heat consumption by melting ice in these lower ground layers causes the model soil column to be generally too warm.' is then discussing the changes within that context. This particular comparison is complicated by the deepening of the zero flux boundary as is discussed in the same paragraph. Later in the manuscript we compare the 20 layer experiment to the SoilGrids (generally deeper permeable depths) experiment. Here we see generally colder winter and warmer summer ground temperatures in the 20 layers simulation compared to SoilGrids as one would expect due to the latent heat effects of soil water (Figure 6).
- P 13 line 25: "accurately" -> "accurately simulating"
  Corrected

- P 15 line 17-18. Soilgrids has extremely deep soils in West Siberia/Urals: Can you find some reference or ask someone who's been there whether this is at all realistic? Having claimed this is a better validated dataset it would be helpful to provide evidence
  of this.
  Unfortunately we can only draw upon the evaluation provided in the cited papers. We know of no other data sources available to us.
- P17 line 15-16. Cold bias due to too much moss is a reasonable assumption, but coupled with the fact that there doesn't seem to be a major warm bias without it (fig 6), and that you would overall expect observations to be biased cold (as written in 3.13), I am not totally convinced here. More consideration of processes and what is missing may be helpful.
  We can look at this in two ways: (a) the relative change between experiments and (b) bias. The relative change is clear, simulations with moss are colder than those without. The responsible processes are known and described as the thermal offset effect (Goodrich 1982) and as reduced warming by snow in the presence of surface materials with low thermal conductivity (Gruber and Hoelzle 2008). The difference in bias is harder to pin down and can be due to a number of things. Two examples include the inappropriate application (or not) of moss at sites that in reality don't (do) have moss and a bias in the observations available (more borehole non-moss sites than moss sites?).
- P29 line 21-22 "thus excluding the simulation of taliks". I am confused by this statement. A talik could be simulated in a 1D model. Discontinuous permafrost, on the other hand, couldn't be. Please could you either clarify or rephrase.
  Yes, this was intended to be 'discontinuous permafrost'. Thanks we have corrected this.
  Hope you find these comments helpful to improve an already good manuscript.
  Yes, thank you

References Cited:

Goodrich, L. E. 1982. "The Influence of Snow Cover on the Ground Thermal Regime." Canadian Geotechnical Journal 19 (4): 421–32. https://doi.org/10.1139/t82-047.

Gruber, Stephan, and Martin Hoelzle. 2008. "The Cooling Effect of Coarse Blocks Revisited: A Modeling Study of a Purely Conductive Mechanism." In Proceedings of the 9th International Conference on Permafrost 2008, 557–61. Fairbanks, Alaska, USA.

Improving permafrost physics in the coupled Canadian Land Surface Scheme (v. 3.6.2) and Canadian Terrestrial Ecosystem Model (v. 2.1) (CLASS-CTEM)

Joe R. Melton[1], Diana L. Verseghy[2,*], Reinel Sospedra-Alfonso[3], and Stephan Gruber[4]

[1]Climate Research Division, Environment and Climate Change Canada, Victoria, B.C., Canada
[2]Formerly at Climate Research Division, Environment and Climate Change Canada
[3]Canadian Centre for Climate Modelling and Analysis, Climate Research Division, Environment and Climate Change Canada, Victoria, B.C., Canada
[4]Department of Geography and Environmental Studies, Carleton University, Ottawa, Canada
[*]Retired

Correspondence: Joe R. Melton (joe.melton@canada.ca)

Abstract. The Canadian Land Surface Scheme and Canadian Terrestrial Ecosystem Model (CLASS-CTEM) together form the land surface component of the Canadian Earth System model (CanESM). Here we investigate the impact of changes to CLASS-CTEM that are designed to improve the simulation of permafrost physics. Eighteen tests were performed including changing the model configuration (number and depth of ground layers, different soil permeable depth datasets, adding a surface moss layer), and investigating alternative parameterizations of soil hydrology, soil thermal conductivity and snow properties. To evaluate these changes, ~~outputs from~~ CLASS-CTEM outputs were compared to 1570 active layer thickness (ALT) measurements from 97 observation sites that are part of the Global Terrestrial Network for Permafrost (GTN-P), 105 106 monthly ground temperature observations from 132 GTN-P borehole sites, a blend of 5 observation-based snow water equivalent (SWE) datasets (Blended-5), remotely-sensed albedo, and seasonal discharge for major rivers draining permafrost regions. From the tests performed, the final revised model configuration has more ground layers (increased from 3 to 20) extending to greater depth (from 4.1m to 61.4 m) and uses a new soil permeable depths dataset with a surface layer of moss added. The most beneficial change to the model parameterizations was incorporation of unfrozen water in frozen soils. These changes to CLASS-CTEM cause a small improvement in simulated SWE with little change in surface albedo but greatly improve the model performance at the GTN-P ALT and borehole sites. Compared to the GTN-P observations, the revised CLASS-CTEM ALTs have a weighted mean absolute error (wMAE) of 0.41 - 0.47 m (depending on configuration), improved from > 2.5 m for the original model, while the borehole sites see a consistent improvement in wMAE for most seasons and depths considered, with seasonal wMAE values for the shallow surface layers of the revised model simulation at most 3.7 °C, which is 1.2 °C ~~greater than those calculated for the model driving~~ more than the wMAE of the screen-level air temperature ~~compared to observations at the sites~~used to drive the model as compared to site-level observations (2.5 °C). Sub-grid heterogeneity estimates were derived from the standard deviation of ALT on the 1 km^2 measurement grids at the GTN-P ALT sites, the spread in wMAE in grid cells with multiple GTN-P ALT sites, as well as from 35 boreholes measured within a 1200 km^2 region as part of the Slave Province Surficial

Materials and Permafrost Study. Given the size of the model grid cells (ca. 2.8°), sub-grid heterogeneity makes it likely difficult to appreciably reduce the wMAE of ALT or borehole temperatures much further.

1 Introduction

Permafrost underlies between 9 and 14 % of the exposed land surface north of 60°S (13 - 18 $\times 10^6$ km^2; Gruber, 2012). The presence of perennially frozen soil at depth has strong impacts on local hydrology, energy fluxes, plant communities, and carbon dynamics. Several factors influence ground temperature and therefore the presence of permafrost, including snow cover, vegetation structure and function, hydrology, and topography (Loranty et al., 2018). Permafrost has been warming and active layers have thickened over the last three decades (Vaughan et al., 2013). This trend is expected to continue due to climate change (Chadburn et al., 2017) making the carbon presently contained in frozen soils vulnerable to release to the atmosphere either as carbon dioxide or methane, depending on local conditions. Since the carbon stored in frozen soils ~~is only~~ becomes readily accessible to microbial respiration once soils thaw, accurately simulating the physics of the permafrost response to a changing climate is vital for reliable predictions of the permafrost carbon feedback to climate change.

The Canadian Land Surface Scheme (CLASS) is the land surface component of the Canadian Earth System Model (CanESM). CLASS has been tested for its cold regions performance in several studies previously. Tilley et al. (1997) evaluated CLASS (v. 2.5) at a site on the North Slope of Alaska. The principal conclusions of the study were that CLASS was most sensitive to ground column depth and soil composition with lesser sensitivity to variations in the radiative fluxes, specification of the overlying vegetation and the initial soil moisture. Bellisario et al. (2000) tested CLASS at a fen wetland and a willow-birch forest in the northern Hudson Bay lowlands. They found the upper soil layer temperatures to be consistently overestimated using the model's default mineral soil parameterization, whereas using the organic soil parametrization of Letts et al. (2000) improved the simulated temperatures significantly. Lafleur et al. (2000) did some tests with a subarctic open woodland site in Churchill, Manitoba using CLASS with the Letts et al. (2000) parameterization. Recommendations from their work included introducing a non-vascular plant functional type (PFT) and a sparse canopy representation, varying the minimum stomatal conductance according to PFT, and re-examination of the snow melt algorithm. The snow melt recommendations were subsequently investigated by Bartlett et al. (2006) and Brown et al. (2006). More recently, Paquin and Sushama (2014) used CLASS (v. 3.5) in the Canadian Regional Climate Model version 5 (CRCM5) to look at the impact of snow and soil parameterizations on simulated permafrost and climate. Their simulations included offline tests using the ERA-Interim meteorological forcing over the pan-Arctic region. Paquin and Sushama tested several options that have previously been made available in CLASS, but not yet implemented operationally, including, 1) increasing the number and depth of soil layers (47 levels extending to 65 m), 2) using the Letts et al. (2000) parameterization for peatlands and assuming an organic surface soil layer for most other regions, and 3) changing the snow thermal conductivity parameterization from Mellor (1977) to Sturm et al. (1997). The Sturm et al. formulation was subsequently adopted in CLASS v. 3.6 (Verseghy, 2017). Ganji et al. (2015) also used CLASS in CRCM5 to investigate cold region hydrological performance. They reported improvements by incorporating supercooled soil water, fractional permeable area, and a changed hydraulic conductivity formulation for frozen soil. MacDonald (2015) coupled CLASS v. 3.6 to the Prairie Blowing Snow Model (PBSM) to simulate the influence of chinooks (Föhn winds) over the South Saskatchewan River Basin. He investigated 15 alternative parameterizations relating to the model physics and concluded by recommending that four of those be considered for adoption in CLASS to improve the simulated snow water equivalent (SWE) and soil water. Three of the suggested parameterizations dealt with snow properties and the fourth related to soil thermal conductivity (MacDonald, 2015).

Our study evaluates the individual and combined effects of suggested enhancements to the Canadian Land Surface Scheme coupled to the Canadian Terrestrial Ecosystem Model (CLASS-CTEM) for simulating processes relevant to soils with permafrost or pronounced seasonal freezing. The model enhancements suggested above have previously been recommended in research studies but not been previously implemented into the CLASS-CTEM framework (unless otherwise noted). Here we investigate the impact of these previously proposed model enhancements as well as several model configuration changes suggested in the literature. Based on this evaluation, a revised version of CLASS-CTEM, containing several enhancements is described and also evaluated. To evaluate model behaviour we draw upon measurements of the thickness of annual thaw in perennially frozen soils (active layer thickness) and borehole temperature sites from the Global Terrestrial Network for Permafrost (GTN-P, 2016) along with other observation-based datasets for snow, surface albedo and runoff.

Numerous studies have investigated the permafrost physics performance of models (e.g. see review in Riseborough et al., 2008) including other large scale models used in ESM applications, such as JULES (Chadburn et al., 2015a, b), JS-BACH (Ekici et al., 2014), and the Community Land Model (CLM, e.g., Alexeev et al., 2007; Lawrence et al., 2008; Lee et al., 2014) allowing us to design our proposed experiments based on their conclusions. The performance of CLASS-CTEM permafrost physics will be evaluated through offline simulations where the model is forced with reanalysis meteorology to avoid biases found in the simulated climate of the coupled model as well as biases in the associated feedbacks. This study is focused on model performance at the large spatial scale of the CanESM as our principle aim is to improve the simulated permafrost physics so that the carbon cycle processes in these regions is well bounded. It is therefore not aimed at shedding light on physical processes in permafrost zones or investigating model performance at individual point locations as the model performance at a single site does not directly translate to model performance over large regions.

~~The~~ In the remainder of the paper~~consists of three sections.~~, Section 2 describes the CLASS-CTEM model, the study design as well as parameterizations tested, and the GTN-P sites used in model evaluation. Section 3 evaluates the model performance ~~over the 18 experiments~~ as well as discussing the influence of sub-grid heterogeneity while Section 4 gives overall conclusions and discusses limitations of our study and future directions for CLASS-CTEM development.

2 Experimental setup

2.1 CLASS-CTEM

CLASS (v. 3.6.2; Verseghy, 2017) coupled with CTEM (v.2.1; Melton and Arora, 2016) forms the land surface component of the CanESM. CLASS performs the land surface energy and water balance calculations on a, typically, half-hour timestep. The model uses leaf area index (LAI), rooting depth, canopy mass, and vegetation height to evaluate the energy and water balance terms of the vegetation canopy and its interactions with the atmosphere. The number of soil layers can vary depending on the application but the standard model setup uses three soil layers of 0.1, 0.25, and 3.75 m thickness. The soil texture (sand, clay, organic matter) dataset used by CLASS-CTEM is the Global Soil Dataset for use in Earth System Models (GSDE; Shangguan et al., 2014). The soil permeable depth is from Zobler (1986) (hereafter Zobler86).~~The GSDE uses a standardized data structure and processing procedures to harmonize and integrate the Soil Map of the World along with various national and regional soil databases (Shangguan et al., 2014).~~ CLASS v.3.6.2 adopts the soil albedo approach of Lawrence and Chase (2007) with the incorporation of a soil colour index geophysical field. ~~The twenty soil colour indices were derived by minimizing the difference between modelled soil albedo and MODIS satellite observed surface albedo (Lawrence and Chase, 2007).~~

CLASS prognostically determines the water content (liquid and frozen) and temperature of all soil layers at each timestep. Also calculated at each timestep, depending on ambient conditions, are the temperature, mass, albedo, and density of a single layer snow pack, interception of rain and snow on the vegetation canopy, and amount of ponded water on the soil surface. Mineral soils are parameterized using the pedotransfer functions of Cosby et al. (1984) and Clapp and Hornberger (1978). Organic soils (organic matter >30% by weight) are modelled as peat following Letts et al. (2000). In the standard CLASS-CTEM framework, lateral transfers of heat or moisture between grid cells are neglected; the treatment of processes such as streamflow and blowing snow require the inclusion of separate, specialized routines (e.g., Soulis et al., 2000; Arora et al., 2001; MacDonald, 2015). All simulations presented here have no geothermal heat flux at the bottom of the soil column.

CTEM calculates the carbon and vegetation dynamics on a daily timestep receiving from CLASS daily mean soil moisture, soil temperature, and net radiation. Photosynthesis and canopy conductance occur on the CLASS timestep. CTEM simulates the respiratory costs and carbon uptake for nine plant functional types (PFTs) which are subsets of the four CLASS PFTs. The CLASS PFTs (with corresponding CTEM PFTs in parentheses) are needleleaf tree (needleleaf deciduous and needleleaf evergreen), broadleaf tree (broadleaf cold deciduous, broadleaf drought/dry deciduous, and broadleaf evergreen), crop (photosynthetic pathway ~~C3 and C4~~C_3 and C_4), and grass (~~C3 and C4~~C_3 and C_4). CTEM carries five carbon pools representing plant leaves, roots, and stems along with two detrital pools for litter and soil C.

For global simulations, CLASS-CTEM is typically run at the CanESM atmosphere resolution which is approximately 2.8° by 2.8° corresponding to a grid cell size of approximately 49 000 km^2 at 45° latitude and about 33 500 km^2 at 70°. Various studies have used observation-based datasets to evaluate CLASS-CTEM at scales from site-level to global (e.g., Peng et al., 2014; Melton and Arora, 2014, 2016). While CLASS-CTEM is capable of running in a mosaic (multiple tiles per grid cell) configuration ~~,~~(e.g. Melton and Arora, 2014; Melton et al., 2017), the simulations presented here are run with a single tile per grid cell.

2.2 Study design

Eighteen experiments were run to assess the impact of model geophysical fields (soil texture, soil permeable depth, and meteorological forcing), model setup (number of soil layers, addition of a moss layer), and model parameterizations ~~. The experiments are listed in Table 1~~(Table 1). The physical quantities used for model evaluation are presented in the next section. The initial model version (Exp. ~~'Base model'~~*Base model*) uses 3 ground layers of thicknesses 0.1, 0.25, and 3.75 m for a total depth of 4.1 m. The first seven experiments address model configuration and input geophysical fields. To test the sensitivity of simulated permafrost to meteorological forcing, CLASS-CTEM was forced with two different meteorological datasets, the Climate Research Unit - National Centres for Environmental Prediction (CRUNCEP v. 8; Viovy, 2016) and the Climate Research Unit - Japanese 55 year Reanalysis (CRUJRA55 v. 1.0.5; Harris et al., 2014; Kobayashi et al., 2015). CRUNCEP was used as the base forcing dataset with additional runs performed for some experiments with CRUJRA55 (see Table 1). While both of these meteorological datasets use the CRU TS dataset (Harris et al., 2014) as the underlying monthly climatology, they differ in their meterological models (NCEP or JRA55). Additionally the spatial resolution of JRA55 is 0.5° while that of NCEP is 2.5°. Thus, the two datasets differ in their spatial and high frequency (sub-monthly) temporal variability. However these differences will be somewhat lessened by their regridding to the CLASS-CTEM model resolution. The meteorological inputs (surface air temperature, surface pressure, specific humidity, wind speed, precipitation, and longwave and shortwave radiation) are disaggregated from 6 hourly to half-hourly time steps while the simulation runs following the methodology in Melton and Arora (2016). Both datasets are available over the extended periods necessary for permafrost simulation (CRUNCEP v. 8: 1901 - 2016, CRUJRA55 v. 1.0.5: 1901 - 2017).

Exp. ~~'20 ground layers'~~ *20 ground layers* changes the number of ground layers from 3 to 20. The 20 layers have higher resolution near the surface with thicker layers at depth (see Table A1). If the permeable soil depth is shallower than the modelled ground column, layers below the soil permeable depth are treated like hydrologically inactive bedrock and are assigned thermal conductivity (2.5 W m^{-1} K^{-1}) and heat capacity (2.13 $\times$ 10^{6} J m^{-3} K^{-1}) values characteristic of sand particles (Verseghy, 2017). If the transition from permeable soil to impermeable bedrock occurs within a soil layer, CLASS calculates the water fluxes only in the depth of permeable soil but simulates one soil temperature for the layer.

The influence of the soil permeable depth dataset is examined by replacing the soil permeable depths of Zobler86 with either the SoilGrids dataset ~~(Exp. 'SoilGrids depth', Shangguan et al., 2017)~~ (Exp. *SoilGrids depth*, Shangguan et al., 2017) or that of Pelletier et al. (2016) (hereafter referred to as Pel16; Exp. ~~'Pel16 depth'). Both SoilGrids and Pel16 have significantly deeper soil permeable depths than Zobler86 (Figure ??).~~ *Pel16 depth*). The influence of a moss layer is examined in Exp. ~~'SoilGrids+Moss' and 'Pel16+Moss'~~*SoilGrids+Moss* and *Pel16+Moss*. In these experiments the top soil layer is ~~characterized as~~ replaced with photosynthetically-inactive moss with a higher porosity, hydraulic conductivity and heat capacity than mineral soil following Wu et al. (2016) ~~. The moss layer is applied to all locations as we are not aware of any global dataset quantifying the spatial distribution of moss.~~ (described in Appendix A1).

~~The~~ Whereas the first series of experiments just described investigated aspects of the model setup, the second series of experiments ~~used the 'SoilGrids+Moss'~~ investigates alternative parameterizations and uses the *SoilGrids+Moss* experiment as a start-

Table 1. List of experiments and the associated model theme they relate to. Experiments denoted with an asterisk were run with both the CRUNCEP and the CRUJRA55 meteorological forcing datasets.

Experiment name	Theme	Description	Starting model configuration
Base model		Original model setup	
20 ground layers	Configuration	Twenty ground layers to a maximum depth of 61.5 m (see Table A1)	Base model
SoilGrids depth	Configuration	Soil permeable depth geophysical field (SoilGrids; Shangguan et al., 2017)	20 ground layers
Pel16 depth	Configuration	Soil permeable depth geophysical field (Pel16; Pelletier et al., 2016)	20 ground layers
SoilGrids+Moss*	Configuration	SoilGrids depth setup with first soil layer treated as non-photosynthetic moss layer following Wu et al. (2016)	SoilGrids depth
Pel16+Moss*	Configuration	Setup as above but with Pel16 depths	Pel16 depth
deVries thermal cond.*	Heat transfer	Soil thermal conductivity following de Vries (1963)	SoilGrids+Moss
Tian16 thermal cond.	Heat transfer	Soil thermal conductivity following Tian et al. (2016)	SoilGrids+Moss
Snow cover:Yang97*	Snow	Snow depth to fractional snow cover relation following Yang et al. (1997)	SoilGrids+Moss
Snow cover:Brown03	Snow	Snow depth to fractional snow cover relation following Brown et al. (2003)	SoilGrids+Moss
Fresh snow density	Snow	Fresh snow density based on air temperature and wind speed following CROCUS as detailed in Essery et al. (1999) with a minimum density of 50 kg m^{-3} following MacDonald (2015)	SoilGrids+Moss
Snow albedo decay	Snow	Efficient spectral snow albedo decay (Dickinson et al., 1993)	SoilGrids+Moss
Super-cooled water	Hydrology	Unfrozen water in frozen soils (super-cooled soil water) following Niu and Yang (2006)	SoilGrids+Moss
Modif. hydrology	Hydrology	Soil matric potential and effective saturated conductivity are modified for the influence of frozen water as described in Ganji et al. (2015)	SoilGrids+Moss

ing point ~~– thus all~~ (the same geophysical fields and model ~~configurations were the same as in SoilGrids+Moss. As described in the Introduction, several recent studies have proposed recommendations for new parameterizations in CLASS-CTEM to enhance cold-regions performance. The second series of experiments examines these recommendations, which~~ configuration). The alternative parameterizations are described in detail in Appendix sections A2 to A7. Briefly, these experiments fall into three main areas related to: 1) heat transfer, 2) snow, and 3) hydrology. The heat transfer experiments replace CLASS-CTEM's default soil thermal conductivity parameterization (Côté and Konrad, 2005) with that of de Vries (1963) following the recommendations of MacDonald (2015)(Exp. ~~'deVries thermal cond.'~~ *deVries thermal cond.* results are discussed in the Supplement). As de Vries (1963) does not account for frozen water in soil, whereas ~~(Côté and Konrad, 2005)~~ Côté and Konrad (2005) does, a further experiment uses a recently published parameterization that simplifies and extends de Vries (1963) to include both frozen and unfrozen water ~~(Exp. 'Tian16 thermal cond.'; See Section A2; Tian et al., 2016)~~(Exp. *Tian16 thermal cond.*; See Section A2; Tian et . Four experiments were devoted to aspects of how snow is simulated in CLASS-CTEM. Experiments ~~'Snow cover:Yang97' and 'Snow cover: Brown03'~~ *Snow cover:Yang97* and *Snow cover: Brown03* replace CLASS-CTEM's default function to relate snow depth to grid cell fractional snow cover from a linear relationship (Verseghy, 2017) to a hyperbolic tangent ~~(following Yang et al., 1997)~~ (following Yang et al., 1997) or an exponential function (following Brown et al., 2003), respectively (~~Figure ??~~Supplement Figure 2). Another experiment (~~'Fresh snow density'~~*Fresh snow density*) changed the calculation for the density of freshly-fallen snow from one based solely on air temperature (Verseghy, 2017) to also considering wind speed following the CROCUS model (Essery et al., 1999). The final experiment concerned with aspects of the snow parameterization is ~~'Snow albedo decay'~~*Snow albedo decay*. CLASS-CTEM uses an empirical exponential decay function to simulate the decrease in snow albedo as snow ages. In ~~'Snow albedo decay'~~*Snow albedo decay*, the default parameterization is replaced by an efficient spectral method (Wiscombe and Warren, 1980; Dickinson, 1983). The last series of experiments looked at hydrology. Water in soils can be, partially or completely, unfrozen at temperatures below 0°C due to the effects of interfacial curvature, adsorption forces and solutes (Watanabe and Mizoguchi, 2002; Dall'Amico et al., 2011). Experiment ~~'Super-cooled water'~~ *Super-cooled water* incorporated the unfrozen water in frozen soils parameterizaton of Niu and Yang (2006) and Exp. ~~'Modif. hydrology'~~ *Modif. hydrology* modifies the soil matric potential and saturated hydraulic conductivity to account for the influence of frozen water following Ganji et al. (2015).

For model spinup, the meteorological forcing years of 1901 – 1925 were cycled over repeatedly until the model reached active layer thickness (ALT) equilibrium (less than 0.05 m difference between average ALT between spinup cycles across all cells with permafrost within them). To run from 1851 to 2016 while atmospheric CO2 concentration and land cover evolved, the climate was cycled over twice from 1901 to 1925 for the years 1851–1900, then the model climate was allowed to run freely from 1901 to 2016. For the simulations presented here, CLASS-CTEM was run with a prescribed, rather than prognostically determined distribution of PFTs.

Active layer thickness in CLASS-CTEM is determined by the temperature and water content of the ground layers. If a layer's temperature is 0°C, the frozen water fraction is used to estimate the thickness of freezing within the layer, i.e., if half of the water content in the layer is frozen, the ALT is assumed to be halfway through the layer. Permafrost area in the model domain was calculated by selecting grid cells with active layer thicknesses less than the model total ground column and multiplying by the grid cell area.

2.3 Datasets used for model evaluation

~~We evaluate the CLASS-CTEM simulations against measurements of the thickness of annual thaw in perennially frozen soils and borehole temperature sites from the GTN-P along with other observation-based datasets for snow, surface albedo and runoff. We also compare to published estimates of permafrost spatial coverage.~~

2.3.1 Active layer thickness sites from the Global Terrestrial Network for Permafrost (GTN-P)

To evaluate CLASS-CTEM, ~~data from~~ 97 open access GTN-P ALT sites were ~~downloaded from the GTN-P (2016)~~ chosen due to their locations in regions of continuous or discontinuous permafrost (GTN-P, 2016; Biskaborn et al., 2019) (accessed May 11^{th}, 2017; ~~Table ??~~ Supplement Table 1 and Figure 1). ~~These sites were chosen due to their locations in regions of continuous or discontinuous permafrost.~~ No sites in areas of sporadic or isolated permafrost were used due to the difficulty in representing this type of permafrost within a large model grid. While we attempted to have as broad a spatial coverage of the GTN-P sites as possible, no open access sites were available for Eastern Canada and Fennoscandia. For comparison with CLASS-CTEM, at each observation time, the average of the sampling grid was determined at each GTN-P ALT site. Then for each site, the sampling grid averages were converted to monthly mean values. The closest ~~CLASS-CTEM grid cell~~ grid cell was determined from the centre of the model grid cells to the ALT ~~site's location was selected~~ sampling location and the modelled monthly average ALTs were compared to the observed values. This resulted in the 97 GTN-P sites, with 1570 ALT observations, being placed into 37 CLASS-CTEM grid cells. As multiple GTN-P sites can be co-located in one CLASS-CTEM grid cell, the weighted mean absolute error (wMAE) for a grid cell was found by averaging the MAE calculated at each site situated within one CLASS-CTEM grid cell.

2.3.2 Borehole temperatures from the GTN-P

Borehole data from the GTN-P were downloaded for 132 open-access sites found in the permafrost (including continuous, discontinuous, sporadic, and isolated) or permafrost-free domains (accessed May 11^{th}, 2017). Most of the boreholes are in Eurasia with few in North America (Figure 1; Supplement Table 2). Each site has its own unique time period of observations and number and depth of observations. At each site, the depths of borehole temperatures were selected to be within 0.05 and 3.0 m of the ground surface ~~, missing values were filtered out,~~ and the observations were averaged to monthly values. For each borehole and each observation depth, the CLASS-CTEM output was selected for the nearest grid cell and the same month as the observations. Linear interpolation was then used to determine the simulated soil temperature for the same soil depth as the observation. ~~The boreholes and number of observations at each are listed in Table ??.~~ As with the ALT sites, several steps were needed to avoid biasing the comparison with CLASS-CTEM. First, borehole sites co-located in the same CLASS-CTEM grid cell were flagged. The 132 borehole sites are located in 73 unique CLASS-CTEM grid cells. Secondly, the number of observations varied by borehole site so when calculating the kernel density estimates (KDE; presented later) within a model grid cell each observation was weighted by the total number of observations per grid cell. Thus grid cells with many GTN-P borehole sites will have each observation weighted less than sites with fewer observations so each grid cell contributes equally to the KDE estimation and the calculation of wMAE.

2.3.3 Snow, albedo and runoff

Snow water equivalent (SWE) from CLASS-CTEM is compared to the Blended-5 dataset for the period from January 1981 to December 2010. Blended-5 is a multi-dataset SWE product developed by Mudryk et al. (2015) that combines five observation-based SWE datasets. Our analysis is limited to regions northward of 45°N with climatological SWE > 4 mm to avoid regions of ephemeral snow. Simulated land surface albedo is compared to the MODIS MCD43C3 white-sky albedo (MODIS Adaptive Processing System, NASA, 2016) for the period spanning February 2000 to December 2013. Similar to SWE, we limit our analysis to regions northward of 45°N. We compared our simulated seasonal runoff to measured discharge rates for seven major river basins that drain permafrost regions for the period from 1965 to 1984 (Ob, Volga, Lena, Yenisei, Yukon, Mackenzie, and Amur rivers; UNESCO Press, 1993). This comparison is limited to seasonal discharges since the CLASS-CTEM runoff is not routed, thus the timing of transport of the water from each grid cell to the river mouth is neglected. On a seasonal timescale this should not cause serious errors but the results must be interpreted with caution.

2.3.4 Permafrost distributions from the literature

Because permafrost cannot easily be observed spatially and reliable data are sparse, global or continental-scale simulation results are often compared to estimates of permafrost distributions. Most prominently, this is the "Circum-Arctic map of permafrost and ground-ice conditions" (Brown et al., 1997) that distinguishes zones of permafrost extent at a scale of 1:10 000 000. These zones are based on expert assessment and manual delineation, often following isotherms of mean annual air temperature. Here, we use *permafrost extent* to refer to the fraction (0 - 1) of the surface that is underlain by permafrost within a pixel or a polygon, *permafrost area* to the actual area (km^2) underlain by permafrost and *permafrost region* is used to denote the area (km^2) where some proportion of the ground can be expected to contain permafrost. The permafrost region is commonly taken to include areas with a permafrost extent exceeding some threshold (Zhang et al., 2000; Gruber, 2012). These definitions are relevant because CLASS-CTEM produces a binary result, i.e., permafrost is present or absent in a cell, and the classes (Zhang et al., 2000; Brown et al., 1997) and the continuous index (Gruber, 2012) of permafrost extent that are used for comparison need to be interpreted appropriately. Neglecting aggregation effects (Giorgi and Avissar, 1997), which arise when the average fine-scale behaviour of a simulated environmental variable is not equal to the simulated coarse-scale behaviour, a threshold of permafrost extent at 50% provides a first estimate of the region that should be compared with a model producing a binary result. For example, environmental conditions that give rise to a permafrost extent of 60% would likely be considered to have permafrost in the binary model and their area would be counted as having permafrost entirely (rather than only 60% of it). Similarly, conditions that produce a permafrost extent of 40% would likely result in not having permafrost in a binary model. As a consequence, we use the total area of all polygons or pixels with an expected permafrost extent larger than 50% as the appropriate area to compare with the results from CLASS-CTEM, termed 'region_50'. This includes continuous and extensive discontinuous permafrost in the Brown et al. map totalling to 15 Mkm^2 (Zhang et al. 1999) and a similar number can be interpreted from a plot of permafrost zonation index and permafrost region (Gruber, 2012).

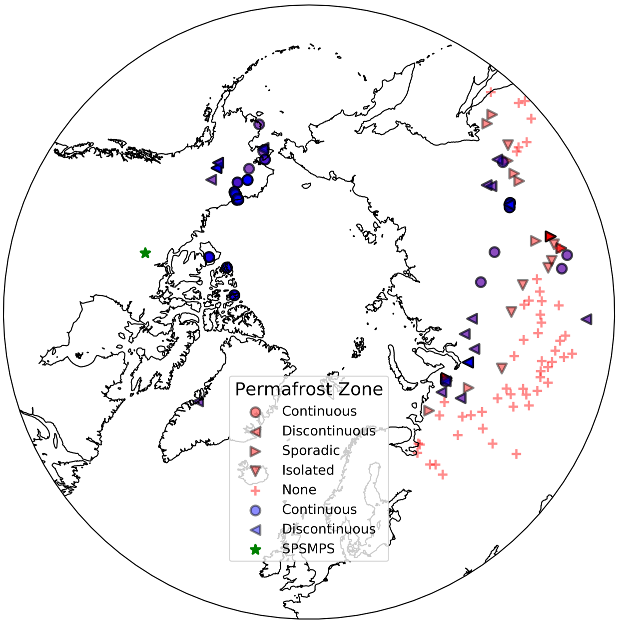

Figure 1. Locations of the 97 GTN-P ALT sites (blue; Supplement Table ~~??~~1), 132 GTN-P borehole observation sites (red; Supplement Table ~~??~~2) and the Slave Province Surficial Materials and Permafrost Study (green; SPSMPS; Lac de Gras, Northwest Territories, Canada) used for model evaluation. Each site is classified according to its permafrost zone listed in the GTN-P. The site markers are semi-transparent hence regions with many closely located GTN-P sites will cause overlap, and darkening, of the markers.

3 Results and Discussion

3.1 Comparison against GTN-P ALT sites: sites with no simulated permafrost

A first simple test of permafrost performance for CLASS-CTEM is to check whether the GTN-P ALT sites are ~~are~~ in fact simulated as containing permafrost. Given that CLASS-CTEM is being run on the CanESM grid (ca. 2.8°), it is possible that site conditions such as meteorology, orography, or vegetation at the GTN-P ALT measurement sites could be quite dissimilar to those of the nearest grid cell, which covers many thousands of km^2. In such cases, CLASS-CTEM could simulate no permafrost where some permafrost indeed exists. Per experiment, the number of sites with no permafrost simulated are listed in Table 2. These ALT sites were removed from further analysis as the ALT in sites without permafrost is not defined. Most experiments had between six and eight observation sites (corresponding to 4 to 6 grid cells) incorrectly simulated as permafrost-free (ISPF). The ~~'Base model'~~ *Base model* experiment has significantly more sites ISPF at 15, corresponding to 2 or 3 ~~more~~ additional grid cells. In general, for the same experiment, the CRUJRA55 meteorological forcing results in fewer grid cells ISPF than CRUNCEP. Small differences in the simulated presence of permafrost (or the number of sites ISPF) are to be expected given the possibility of errors in the meteorological forcing and local variations in site-level characteristics, but large differences can indicate problems with the model setup and parameterizations.

3.2 Initial model performance

The ~~'Base model'~~ *Base model* experiment simulates a permafrost area (PA) of 8.6 Mkm^2 (north of 60° S; Table 2) with permafrost confined to northern Siberia, Alaska and the northern edge of Canada (Figure 2). This low PA is in line with that simulated by CLASS-CTEM when coupled within the CanESM, although the spatial distribution is different due to the different atmospheric forcing (Koven et al., 2013). Also plotted in Figure 2 is the PE estimate of Brown et al. (1997). The Brown et al. (1997) dataset gives permafrost spatial distribution in four classifications which are not directly comparable to ALTs but may be used to give a general indication of PA from an independent estimate. Owing to the coarseness of the model grid CLASS-CTEM is not able to simulate isolated or sporadic permafrost~~, but~~. For regions of discontinuous and continuous permafrost, comparing the estimated distribution of ~~discontinuous and continuous permafrost~~ Brown et al. (1997) to the modelled ALT indicates poor agreement.

With such a small permafrost area many of the GTN-P ALT sites were ISPF as mentioned above. Of the GTN-P ALT sites where CLASS-CTEM simulated permafrost, the ~~'Base model'~~ *Base model* simulations show overly shallow ALTs with an average mean absolute error (MAE; described in Section 2.3) of 0.410 m. Thus it appears the modelled soil temperatures are too warm in the more southerly permafrost domain (PD), leading to no permafrost simulated, and too cool at the higher latitudes. ~~The 'Base model' experiment also has increasing bias with depth for the 29 sites with permafrost (Figure ??).~~ However, it should be noted that the model configuration of ~~3~~ three ground layers in this experiment makes an accurate estimation of the ALT difficult since the lowest model layer is quite thick (3.75 m).

3.3 Increasing the number of ground layers

Increasing the number of ground layers from 3 to 20 decreases the number of GTN-P ALT sites ISPF from 15 to 7 (Table 2). Figure 3 shows the difference between the simulated and observed ALT at each grid cell with GTN-P ALT sites for selected experiments. The average MAE computed against the GTN-P ALT observations for ~~'Exp 20 ground layers'~~ *Exp 20 ground layers* is over 2.5 m with simulated ALTs strongly overestimated (Figure 3). When the number and depth of ground layers is increased, but the soil permeable depth is left unchanged, CLASS-CTEM simulates the ground layers below the permeable soil depth as impermeable bedrock. The absence of water and therefore of heat consumption by melting ice in these lower ground layers causes the model soil column to be generally too warm. However, the total global PA increases from 8.6 Mkm^2 simulated by the ~~'Base model'~~ *Base model* to 16.8 Mkm^2 (Table 2) with an increase in permafrost area primarily in the southern fringes of eastern Siberia and Canada along with a general deepening of ALT across the high latitudes (Figure 2). This seeming incongruity of warmer soils with a larger permafrost area likely relates to moving the boundary of zero heat flux from 4.1 m, a depth where seasonal temperature variations can penetrate, to 61.4 m. The shallower modelled soil column in the ~~'Base model'~~ *Base model* inhibits the formation of permafrost because of the concentration of the annual heat flux oscillation in the upper few meters of the soil. ~~In the PD for the '20 layers' simulation, there is a slight general increase in sensible heat flux compared to the 'Base model' simulation (Figure ??). Changes in latent heat flux, leaf area index (LAI) and annual runoff are significant but spatially heterogeneous across the permafrost zone. Increasing the number of soil layers and depth of the zero flux boundary~~

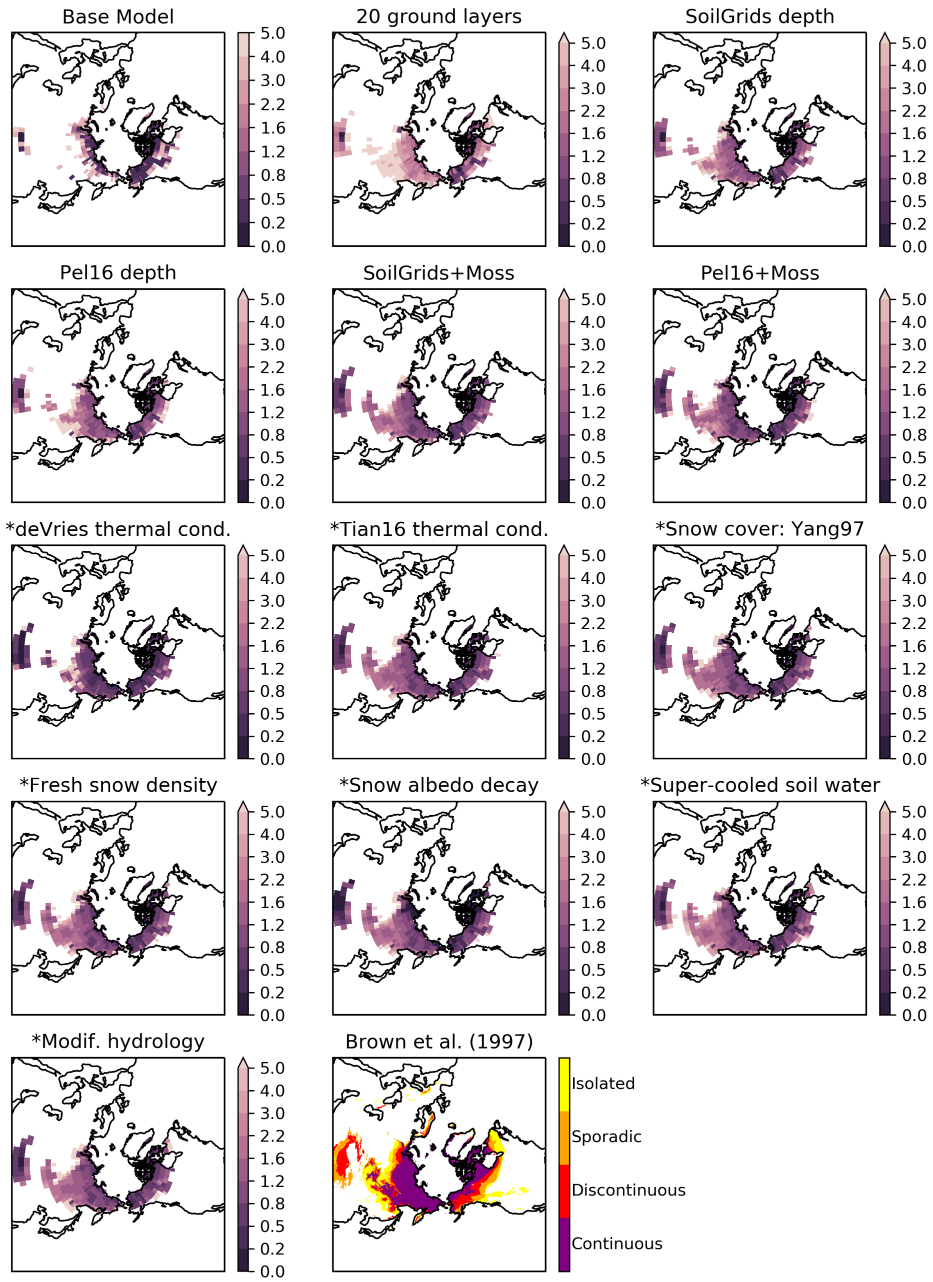

Figure 2. ALTs in meters for experiments listed in Table 1 alongside the permafrost map of Brown et al. (1997) (bottom right). Experiments with an asterisk prefixing their name use a model configuration based on the ~~'SoilGrids+Moss'~~ *SoilGrids+Moss* setup. All experiments shown here use CRUNCEP for the meteorological forcing.

Table 2. Permafrost area as simulated by CLASS-CTEM (average of 1996 - 2015) along with literature estimates for terrestrial permafrost north of 60°S. The number of GTN-P sites which CLASS-CTEM incorrectly simulated as permafrost free (ISPF) is also listed along with the number of corresponding grid cells in square brackets. These GTN-P sites were removed from further analysis since ALT is not defined in locations with no permafrost. The numbers in parentheses indicate the values when CRUJRA55 was used as the meteorological forcing instead of CRUNCEP. See Section 2.3.4 for distinction between permafrost area and permafrost region.

Experiment	Permafrost Area (10^6 km^2)	Number of sites [grid cells] ISPF
~~Base model~~ *Base model*	8.6	15 [8]
~~20 ground layers~~ *20 ground layers*	16.7	7 [5]
~~SoilGrids depth~~ *SoilGrids depth*	15.7	8 [6]
~~Pel16 depth~~ *Pel16 depth*	15.7	8 [6]
~~SoilGrids+Moss~~ *SoilGrids+Moss*	17.9 (19.8)	7 (6) [5 (4)]
~~Pel16+Moss~~ *Pel16+Moss*	18.5 (19.8)	7 (6) [5 (4)]
~~deVries thermal cond.~~ *deVries thermal cond.*	16.2 (17.8)	8 (6)[6 (4)]
~~Tian16 thermal cond.~~ *Tian16 thermal cond.*	21.2	6 [4]
~~Snow cover: Yang97~~ *Snow cover: Yang97*	19.3 (20.8)	6 (6) [4 (4)]
~~Snow cover: Brown03~~ *Snow cover: Brown03*	19.0	6 [4]
~~Fresh snow density~~ *Fresh snow density*	18.9	6 [4]
~~Snow albedo decay~~ *Snow albedo decay*	15.6	6 [4]
~~Super-cooled water~~ *Super-cooled water*	20.1	6 [4]
~~Modif. hydrology~~ *Modif. hydrology*	19.5	6 [4]
Literature estimates	**Permafrost Area (10^6 km^2)**	
Zhang et al. (2000)	12.2 - 17.0	
Gruber (2012)	12.9 - 17.7	
Literature estimates	**Permafrost Region (10^6 km^2)**	
Zhang et al. (1999)	22.8	
Gruber (2012)	21.7 with a range of 18.7 to 24.3	

~~reverses the bias trend in ALT with higher bias at shallower ground depths (Figure ??), although these differences should be interpreted with caution due to the difference in number of ALT sites included in this analysis between the 'Base model' and '20 ground layers' simulations.~~

~~The residuals of a linear regression of the CLASS-CTEM ALTs on the GTN-P ALTs for selected experiments. Each dot is the mean residual for all GTN-P site(s) that fall within a CLASS-CTEM grid cell. The number of grid cells differs between experiments due to sites being ISPF. Black curves are a LOWESS smoother applied to the residuals. For clarity all experiments are plotted with the same axis limits.~~

The wMAE calculated for each season from CLASS-CTEM's simulated ground temperatures compared to GTN-P borehole temperatures for three depth zones shows an improvement at all depths and seasons for ~~'20 layers' over 'Base model'~~ *20 layers* over *Base model* (Figure 4). Generally, across all experiments, CLASS-CTEM performs better with increasing depth. Seasonally, winter is generally simulated best with summer showing the highest wMAE values. These patterns indicate that the largest challenges to accurate ground temperature simulation are coming from the high variability in forcing at the land surface and from the difficulty in accurately simulating the summertime heat pulse into the ground column.

To look in closer detail at the model performance for the GTN-P borehole sites, Figure 5 shows the Gaussian kernel density estimate (KDE) derived from differences between the simulated and observed borehole temperatures. For shallow soils, as the seasons progress from winter to fall, the proportion of instances with a strong cold bias decreases with a warm soil bias taking over in summer, especially in the shallowest depth band. This would indicate the modelled soil heat fluxes are somewhat exaggerated. The fall period generally has the least bias, potentially due to the loss of the warm summer bias but prior to the establishment of the cold winter bias.

3.4 Increasing the soil permeable depths

Changing the soil permeable depth dataset to SoilGrids (Exp. ~~'SoilGrids depth'~~*SoilGrids depth*) from Zobler86 gives a general improvement over the ~~'20 ground layers'~~ *20 ground layers* simulations with a drop in average MAE to 1.162 m at the GTN-P ALT sites (Figure 3). ~~The residuals calculated against the GTN-P ALT sites (Figure ??) now show a general trend of increasing bias with increasing soil depth, similar to the 'Base model' simulation.~~ There is also a shift to shallower ALTs (Figure 2) with a slight decrease in PA to 15.7 Mkm^2, which is within the range literature estimates (Table 2 and discussed further in Section 2.3.4). ~~Compared to the '20 ground layers' simulation for the permafrost region, there is a decrease in latent heat flux for much of eastern Siberia and LAI shows strong decreases while runoff increases (Figure ??).~~ The greater permeable depths associated with SoilGrids lead to deeper penetration of water into the soil, resulting in more water being allocated to runoff than made available for plant transpiration or soil evaporation (Supplement Figure 3). Simulations with the alternative soil permeable depth dataset (~~'Pel16 depth'~~*Pel16 depth*) generally show similar patterns of latent heat flux, runoff and LAI (not shown) to the ~~'SoilGrids depth'~~ *SoilGrids depth* experiment. The ~~'Pel16 depth'~~ *Pel16 depth* simulations have better agreement with the GTN-P ALT observations reducing the wMAE to 0.757 m (Figure 3). ~~'SoilGrids'~~ *SoilGrids* also further improves the model's performance at all depths and seasons compared to the GTN-P borehole sites (Figure 4).

Numerous studies have pointed to the importance of increasing the simulated ground column depth and number of ground layers to better capture the decay with depth of the influence of multi-decadal variability (e.g. Smerdon and Stieglitz, 2006; Alexeev et al., 2007; Nicolsky et al., 2007; Paquin and Sushama, 2014). Of particular relevance to our study, Paquin and Sushama (2014) used CLASS in CRCM5 and found shallow soil configurations (permeable depth < 1 m throughout much of the model domain) to lead to overly strong seasonal cycles with resulting overly deep ALTs, similar to the work of Smerdon and Stieglitz (2006), and in line with our ~~'Base model'~~ *Base model* simulation with its small estimated PA.

The availability of comprehensive global soil permeable depth datasets is relatively recent. Previous studies would often assume a constant permeable soil depth, either shallow (Dankers et al., 2011) or deep (Lawrence et al., 2008) with the deeper layers hydrologically inactive. Comparing the three permeable depth datasets (Zobler86, SoilGrids, and Pel16; ~~Figure ??~~Supplement Figure 1) shows Zobler86 to be by far the shallowest while SoilGrids and Pel16 disagree on the spatial distribution of the permeable depths for the high latitude regions. Pel16 shows deep soils in the Canadian boreal forest, Finland and central southern Russia with shallower soils in the Siberian plateau. SoilGrids has more very deep soils (>50 m) especially in the West Siberian region and the Urals. These differences in permeable depth have an impact on the simulated ALT as the SoilGrids and Pel16 experiments perform quite differently at the GTN-P ALT sites (Figure 3) due to the strong impact of freezing and thawing of water in the soil column.

3.5 Adding an upper layer of organic matter/moss to the soil column

CLASS-CTEM ALTs with both Pel16 and SoilGrids are generally biased deeper than observed at the GTN-P sites (Figure 3) indicating that the ground surface is either overly insulated from the cold atmosphere during the winter or is absorbing too much heat during the summer months. The principal modulating influences on ground heat fluxes in cold regions are hydrology, snow cover (both of which we deal with later), vegetation structure and function, and topography (Loranty et al., 2018). Vegetation canopies shade the soil surface, attenuating radiation and reducing warming in the summer season. As well dense forests capture snow in the canopy which prevents it from reaching the ground and insulating the soil surface further cooling soils. Another aspect of vegetation influence is the insulating effect of a surface layer of moss or organic matter. Mosses are generally more abundant at high latitudes and have been shown to decrease growing season surface soil temperatures (Turetsky et al., 2012). The effect of mosses on the ground heat flux has also been demonstrated through field experiments (Gornall et al., 2007; Van Der Wal and Brooker, 2004) and modelling studies have incorporated organic layers (e.g. Lawrence et al., 2008; Paquin and Sushama, 2014) or bryophytes (Porada et al., 2016) to improve permafrost dynamics. Exps. ~~’SoilGrids+Moss’ and ’Pel16+Moss’~~ *SoilGrids+Moss* and *Pel16+Moss* both incorporate a non-photosynthetic moss layer in place of the first layer of soil (see Section 2.2) and both simulate generally shallower ALTs than their parent simulations (~~’SoilGrids depth’ and ’Pel16 depth’~~*SoilGrids depth* and *Pel16 depth*, respectively; Figure 2). The effect of moss introduction for the ~~’SoilGrids+Moss’~~ *SoilGrids+Moss* experiment is to reduce average MAE from 1.162 to 0.472 m for the GTN-P ALT sites (Figure 3) ~~with a reduced increase in bias with soil depth (Figure ??).~~ The general cooling influence is evident by comparing to the GTN-P ALT sites (Figure 3) and also through the increase in simulated PA from 15.7 to 17.9 Mkm^2. A similar improvement is seen for Exp ~~’Pel16+Moss’ where the average MAE drops from 0.757 to 0.404 m (Figure 3). Adding moss causes decreases in the mean annual latent and sensible heat fluxes from the PD with general decreases also in LAI and a large increase in annual total runoff (Figure ??).~~ *Pel16+Moss*. The high porosity of the moss layer causes less water to be available at the surface for evaporation, reducing the latent heat flux and making more water available for runoff, and its insulating effect keeps the soil surface cooler, which reduces plant growth and also the sensible heat flux (Supplement Figure 4). The reduction in plant growth due to cooler soils also reduces water uptake for transpiration further increasing runoff.

Comparing simulated ground temperatures to observations at the GTN-P borehole sites shows a slight increase in wMAE at all depth ranges and seasons compared to the ~~’SoilGrids’~~ *SoilGrids* simulation (Figure 4). Comparing the KDE plots of the bias distribution between modelled and observed borehole temperatures for the ~~’SoilGrids+Moss’ and the ’SoilGrids’~~

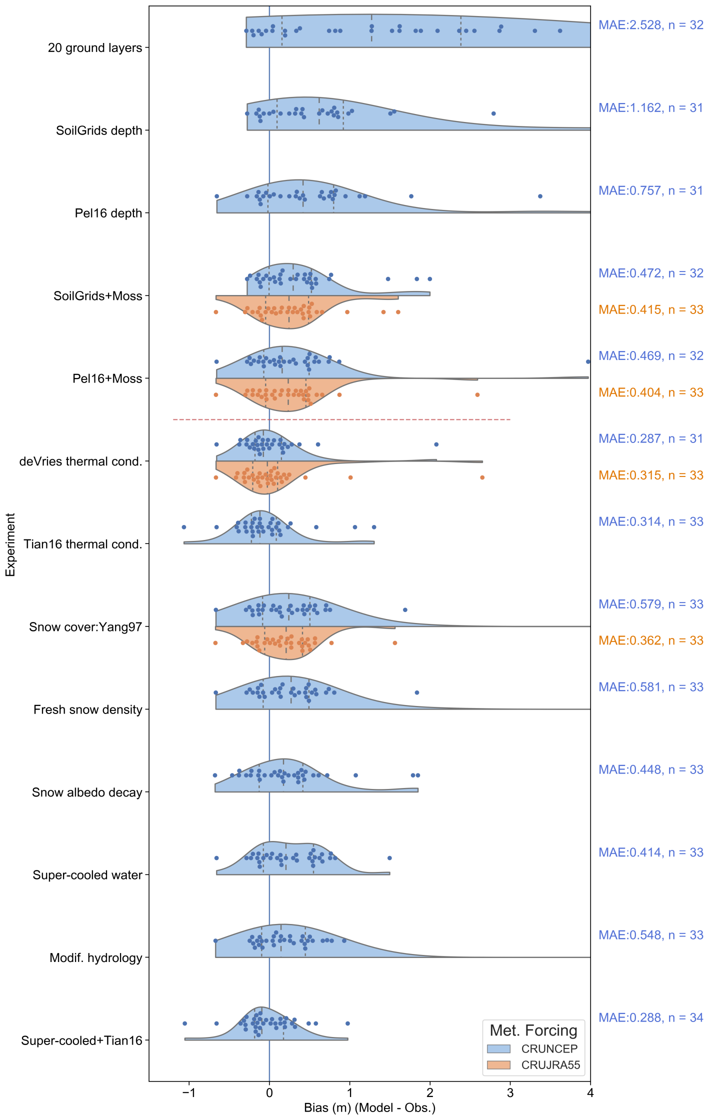

Figure 3. Differences between the ALTs from the experimental model runs and those of the Global Terrestrial Network for Permafrost ALT sites (Supplement Table ~~??~~1). Each dot represents a grid cell with one or more GTN-P sites (see Section 2.3). In this representation (a 'bee swarm'), displacement in the y-direction is only to allow each data point to be visible. The background shading is a Gaussian kernel-density estimate (KDE) with the quartiles of the distribution indicated by dashed vertical lines within the KDE plot. The mean absolute error (MAE) is produced by calculating the MAE at each grid cell and taking the average across all cells. As the number of sites ISPF differs between experiments (Table 2) the number of grid cells where CLASS-CTEM simulated permafrost is also listed. The total number of grid cells with GTN-P sites is 37. The two meteorological forcings are shown for the experiments where the CRUJRA55 forcing was also used. Experiments below the dashed red line use the model setup from Exp. ~~'SoilGrids + Moss'~~ *SoilGrids + Moss* as their starting point (Table 1).

SoilGrids+Moss and the *SoilGrids* simulations shows an increased cool bias in the shallow soil which is especially evident in summer (Figure 5). This bias extends deeper into the soil column, albeit weakening with depth. The cooling of soils due to the incorporation of a moss layer was also found by Porada et al. (2016), however their simulations included a dynamic extent for moss cover. The creation of a cold bias due to the introduction of a moss layer is reasonable considering that the moss layer was applied to all areas uniformly. While this experiment was intended to understand the impact of moss on simulated ground temperatures, future work should attempt to place moss with a more realistic distribution, similar to Porada et al. (2016).

~~When the meteorological forcing dataset is CRUJRA55 instead of CRUNCEP, the MAE of both 'SoilGrids+Moss' and 'Pel16+Moss' improves further to 0.415 m and 0.404 m, respectively, with two fewer cells ISPF (The spatial differences between the 'Pel16+Moss' and 'SoilGrids+Moss' simulations with CRUNCEP and CRUJRA55 are shown in Figure ??). The CRUNCEP meteorological forcing dataset generally produces deeper ALTs for much of Eurasia than simulations forced with CRUJRA55 with some shallower ALTs in Alaska and the Yukon. While CRUNCEP and CRUJRA55 share a common climatology (CRU), it is likely their differences in sub-monthly variability leads to significant differences in simulated PA and ALTs. This result is in line with Beer et al. (2018) who used artificially manipulated climate datasets to show that soil temperature can be 0.1 to 0.8 °C higher when climate variability is reduced in the model forcing data.~~

Comparing the model experiment outputs to the GTN-P sites in Figure 3 it is evident that increasing the number of ground layers and the soil permeable depth and incorporating a top layer of moss/organic matter improves the simulated ALTs. These changes have been suggested by other studies as mentioned above and our results are in line with them. The next experiments use the model configuration from ~~'SoilGrids+Moss'~~ *SoilGrids+Moss* as a starting point. While Pel16 generally gave better average MAE values than SoilGrids for ALT compared to the GTN-P sites (Figure 3), SoilGrids appears to be better validated (c.f. Shangguan et al., 2017, Figures 9 - 11). Both datasets, however, suffer from sparse data in high latitudes (e.g., Shangguan et al., 2017, Figure 2). Additionally, while it appears that the addition of moss can introduce a summer cool bias in ground temperatures (as discussed above), given the extensive distribution of bryophytes (c.f. the simulated distribution in Figure 4b in Porada et al., 2016), we chose to include moss in our further simulations.

3.6 Testing alternate soil thermal conductivity formulations

~~The Exp. 'deVries thermal cond.' has an smaller average MAE of 0.287 m than 'SoilGrids+Moss' (0.472 m, Figure 3). Comparing the spatial patterns of ALT for 'deVries thermal cond.' with 'SoilGrids+Moss' shows generally deeper ALTs with a large loss of permafrost in the southern zones of the PD in eastern Russia (Figures 2 and ??). Global simulated PA drops from 17.9 to 16.2 Mkm^2. In winter simulations at seven sites in the South Saskatchewan River Basin (SSRB), MacDonald (2015) found the de Vries (1963) soil thermal conductivity formulation to outperform the Côté and Konrad (2005) soil thermal conductivity parameterization in CLASS-CTEM through improvements in both simulated soil moisture and snow. However, the de Vries (1963) formulation neglects the effect of ice on soil thermal conductivity (see equations A7 - A10). As the thermal conductivity of liquid water (0.57 W m^{-1} K^{-1} at 5 °C) is much lower than that of ice (2.24 W m^{-1} K^{-1} at -4 °C), it is surprising that this formulation performs as well as it did both in our simulations against the GTN-P ALT observations, and for the SSRB runs conducted by MacDonald (2015). This neglecting of the thermal conductivity difference between liquid~~

~~water and ice could be the cause of a dampened soil temperature annual cycle as deeper soil layers are slower to cool in the 'deVries thermal cond.' simulations compared to 'SoilGrids+Moss' (Figure ??). We do see an increase in wMAE at the GTN-P borehole sites (Figure 4) demonstrating that this parameterization does indeed degrade model performance as would be anticipated.~~

~~An alternate soil thermal conductivity parameterization that explicitly accounts for the influence of ice has recently been published. Exp. 'Tian16 thermal cond.'~~ Exp. *Tian16 thermal cond.* tests the Tian et al. (2016) formulation, which is based on de Vries (1963) but explicitly accounts for the influence of ice (see Section A2 and the Supplement section 1). The new formulation simulates a much larger PE than ~~both 'SoilGrids+Moss' and 'deVries thermal cond.'~~ *SoilGrids+Moss* at 21.2 Mkm^2 with generally shallower ALTs ~~compared to 'SoilGrids+Moss'~~ in most regions except for the western edge of simulated Siberian permafrost (~~Figure ??). The 'Tian16 thermal cond.' simulation has smaller sensible and latent heat fluxes with an increase in LAI compared to 'SoilGrids+Moss' (Figure ??). The colder soils in the 'Tian16 thermal cond.' simulations retain more moisture into the summer months as less is lost to evaporation and runoff so higher LAI can be supported.~~ Supplement Figure 5). The average MAE at the GTN-P ALT sites is reduced to 0.314 m (Figure 3) ~~with no trend in bias for ALTs with depth (Figure ??). However~~however, at the GTN-P borehole sites~~there is a general increase in wMAE for~~ , the simulated ground temperatures are biased cold, primarily in summer and fall, and worsening with depth (Figure 4 ~~). The borehole KDE plots show strong distinctions between the 'SoilGrids+Moss' and 'Tian16 thermal cond.' simulations with a large cold bias developing in summer that then extends into fall (Figure~~ and 5).

~~ALT differences (meters) for experiments that are based on the model setup of 'SoilGrids+Moss' (see Table 1) compared to the 'SoilGrids+Moss' simulation. Negative values indicate that ALTs of the experiment are deeper than in 'SoilGrids+Moss' while positive values indicate shallower ALTs. Since permafrost free soils have an undefined ALT, model grid cells that have permafrost in only the experiment are green while cells that have permafrost only in the 'SoilGrids+Moss' simulation are purple. Dots indicate grid cells that are statistically significant (independent two-sample t-test p level < 0.05).~~

3.7 Changing the relationship between snow depth and snow cover

Two experiments investigated different relationships between snow depth and the grid cell snow cover in CLASS-CTEM (~~'Snow cover: Yang97' and 'Snow cover: Brown03'~~*Snow cover: Yang97* and *Snow cover: Brown03*). These modifications ~~led to increases in global PA of around 1.2~~ increased global PA (~1.2 Mkm^2) with a slightly higher PA estimated for ~~'Snow cover: Yang97'~~ *Snow cover: Yang97* (Table 2). ~~The spatial pattern of ALT for both experiments is similar with a few more grid cells simulated by 'Snow cover: Yang97' to be permafrost at the southern edge of North American and east Siberian permafrost (Figures 2 and ??).~~ For the GTN-P ALT sites, both snow cover experiments ~~show an increase in~~ increased average MAE from 0.472 m for ~~'SoilGrids+Moss'~~ *SoilGrids+Moss* to 0.579 m and 0.622 m for ~~'Snow cover: Yang97' and 'Snow cover: Brown03'~~*Snow cover: Yang97* and *Snow cover: Brown03*, respectively. ~~The two parameterization changes result in a slight increase in spring albedo (< 5%) and a small decrease in winter SWE, both not statistically significant, while annual sensible heat fluxes decrease compared to 'SoilGrids+Moss' (Figure ??). The 'Snow cover: Brown03' experiment shows similar spring albedo, and latent and sensible heat fluxes changes to 'Snow cover: Yang97' but also has a general reduction in winter SWE~~

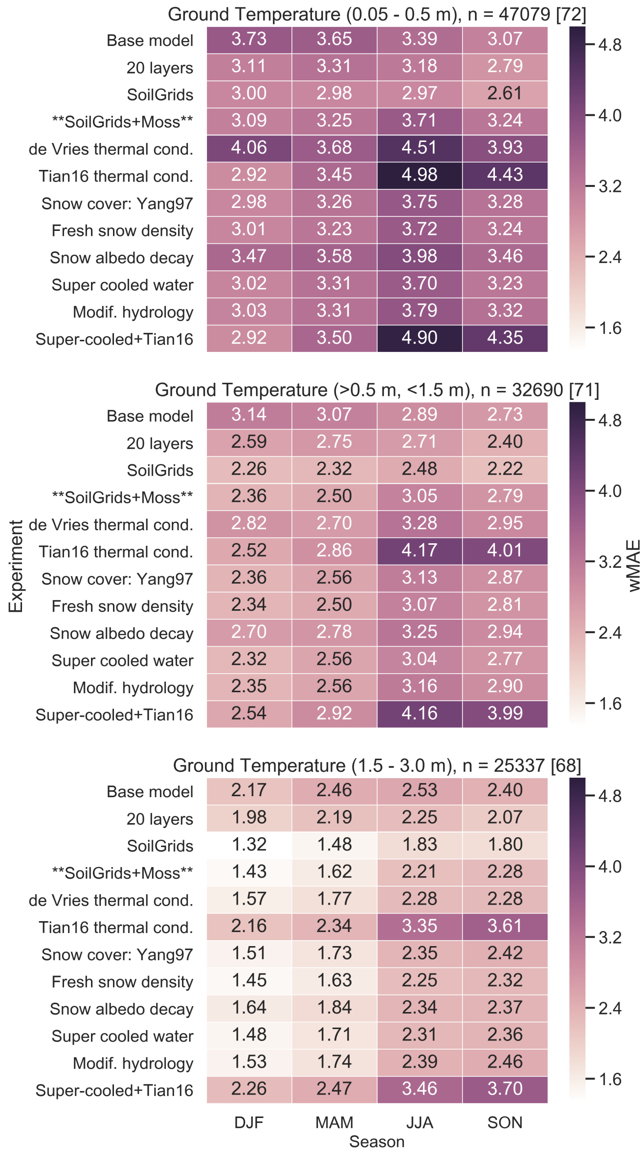

Figure 4. Weighted mean absolute error (wMAE, $^{\circ}$C) between the simulated ground temperatures and those of the GTN-P borehole temperature sites (Supplement Table ~~??~~2) for three depths: 0.05 - 0.5 m, 0.5 m - 1.5 m, and 1.5 - 3.0 m. The weighted mean absolute error (wMAE) is produced by calculating the wMAE for each depth range and season at each site within a grid cell and taking the average across all grid cells (see Section 2.3).The number of observations differs between depths and is listed along with the number of CLASS-CTEM grid cells with GTN-P borehole sites in square brackets. The colour of the text annotations is purely for clarity. The wMAE of CRUNCEP surface air temperatures compared to air temperatures measured at the GTN-P sites is 2.17 $^{\circ}$C, 2.46 $^{\circ}$C, 2.53 $^{\circ}$C, and 2.40 $^{\circ}$C for DJF, MAM, JJA, and SON, respectively over 25 337 monthly observations

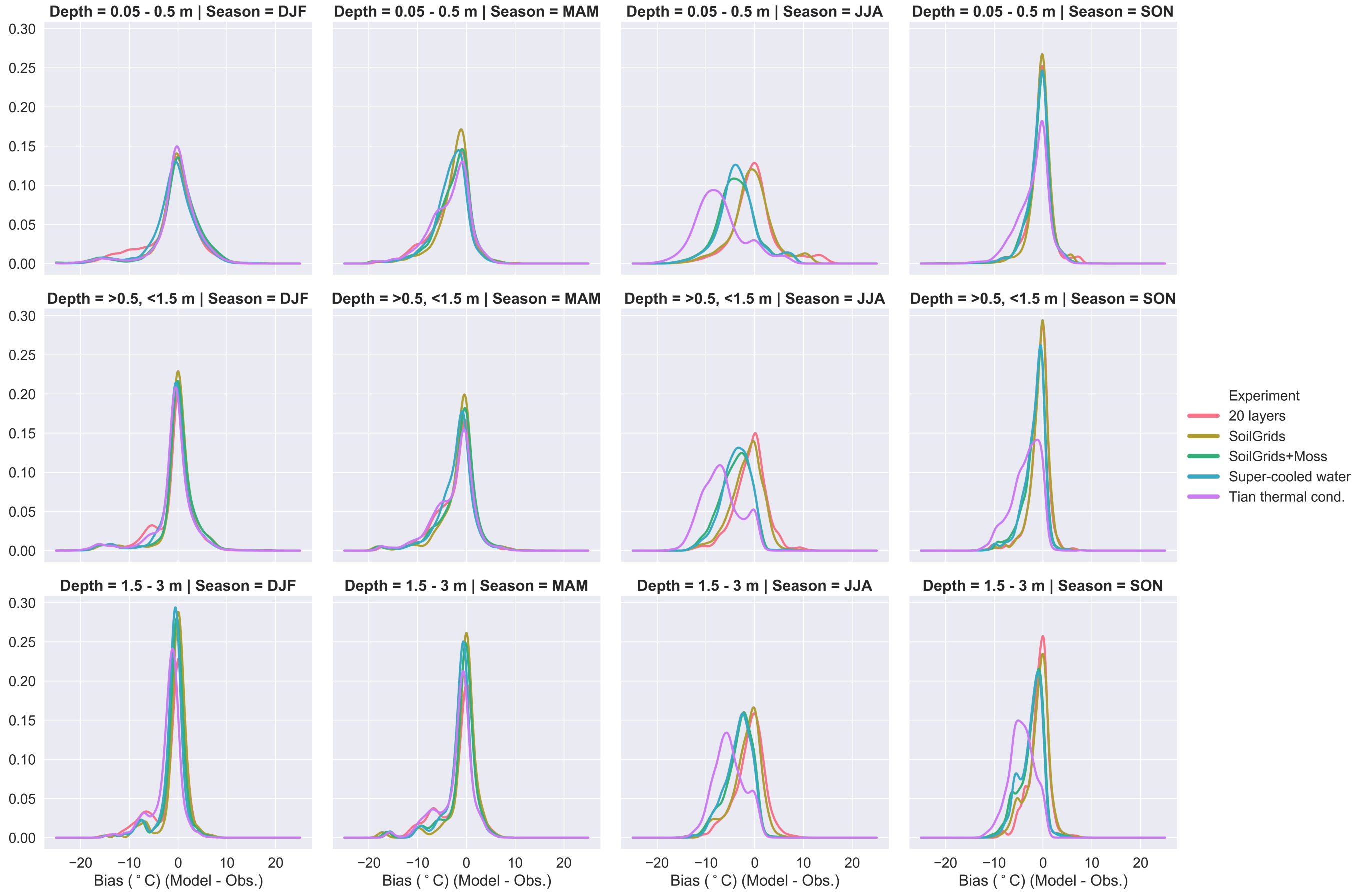

Figure 5. Gaussian kernel density estimates for the difference between the simulated ground temperatures and those of the GTN-P borehole temperature sites (Supplement Table ~~??~~2) for three depths: 0.05 - 0.5 m, 0.5 m - 1.5 m, and 1.5 - 3.0 m, for each season and for selected experiments. The bandwidth was chosen using Scott's rule of thumb (Scott, 1992).

~~and a reduction in runoff for the south central Siberian region compared to 'Snow cover: Yang97' (not shown).~~ Comparing the simulated SWE from both ~~'Snow cover: Yang97' and 'Snow cover: Brown03'~~ *Snow cover: Yang97* and *Snow cover: Brown03* to Blended-5 ~~(see Section 2.3; Mudryk et al., 2015; Kushner et al., 2018)~~ (see Section 2.3) shows a slight improvement in model performance compared to both ~~'Base model' and 'SoilGrids+Moss'~~ *Base model* and *SoilGrids+Moss* throughout the snow year, which tends to be more pronounced during fall and winter (~~Fig. ??~~Supplement Figure 6) although there is little difference between the two snow cover experiments.

Changes in snow cover can lead to large changes in albedo due to the significant brightness difference between snow and vegetation/bare ground. To investigate the impact of these experiments on albedo we evaluated seasonal averages of simulated albedo against MODIS observations over latitudes northward of 45°N for the period 2000 to ~~2013 (Figure ??).~~ 2013. We find the spring (AMJ) albedo from the various simulations is about the same (~~Fig~~Supplement Fig. 7).~~??), with anomaly~~

~~correlation coefficient (ACC)of about 0.5 and standard deviation and root mean square error of approximately 130% and 120%, respectively, relative to the standard deviation from observations.~~

3.8 Considering wind speed in the calculation of fresh snow density

In CLASS-CTEM, the density of freshly fallen snow depends on the ambient air temperature (Eqn. A19). Exp. ~~'Fresh snow density'~~ *Fresh snow density* tested a parameterization from the CROCUS model that also includes wind speed in this calculation (Eqn. A20) ~~. This experiment shows~~ which yielded an increase in PA ~~from 17.9 ('SoilGrids+Moss')~~ to 18.9 Mkm^2~~and while ALTs are generally comparable, there are more grid cells with permafrost at the southern edges of the PD similar to the experiments described above (Figures 2 and ??).~~ . Compared to the GTN-P ALT sites, the ~~'Fresh snow density'~~ *Fresh snow density* results are similar to those of the snow cover experiments with no improvement in average MAE (0.581 m; Figure 3) ~~. Compared to 'SoilGrids+Moss', the experiment 'Fresh snow density' has lower (0 - 10%) mean annual latent heat flux and generally slightly higher sensible heat flux, spring albedo, and winter SWE (Figure ??) although all are not statistically significant. There is a~~ and no discernible impact upon modelled DJF SWE compared to Blended-5 or upon spring (AMJ) albedo compared to MODIS (~~Figurem ?? and ??).~~ Supplement Figure 6 and 7).

The typical wind speed in the CRUNCEP meteorological forcing dataset when snow is falling is in the range of 1 - 5 m s^{-1} (Figure 6). With Eqn A20, the density of freshly fallen snow tends to be lower at very low wind speeds then higher as wind speed increases for the same air temperature. The generally higher density of fresh snow with the CROCUS parameterization results in a snow pack with higher thermal conductivity (Sturm et al., 1997) and thus cooler soils as evident from the expansion in PA for the ~~'Fresh snow density'~~ *Fresh snow density* experiment (Figure 2). Both the original CLASS-CTEM parameterization and that of the CROCUS model produce fresh snow densities within the range of observations. Roebber et al. (2003) evaluated 1650 snowfall events from 28 continental US sites and found the density of freshly-fallen snow to vary from 21.4 to 526.3 kg m^{-3} with a median value of 70.9 kg m^{-3} ~~. It should be noted that these data only represent~~ (for snowfall events where the wind speed was ~~less than or equal to~~ $\leq$9 m s^{-1}).

3.9 Adopting an efficient spectral method for snow albedo decay

Changing the snow albedo decay parameterization from an exponential form (Verseghy, 2017) to an efficient spectral parameterization (Dickinson, 1983)(Exp ~~'Snow albedo decay') results in a small improvement in average MAE for~~ *Snow albedo decay*) slightly improves average MAE at the GTN-P ALT sites (Figure 3) ~~. There is a drop in~~ while decreasing PA (15.6 Mkm^2)~~compared to 'SoilGrids+Moss' (17.9 Mkm^2)~~, reflecting a near uniform deepening of ALT with the exception of small areas on the western edge of the Siberian PD (~~Figures 2and ??~~Figure 2; Supplement Figure 5; Table 2). ~~The 'Snow albedo decay' experiment has higher mean annual latent and sensible heat fluxes for much of the PD with the exception of the Canadian Arctic Archipelago, the North Slope of Alaska, and the northern Urals in Russia (Figure ??). There is also reduced annual runoff with higher LAI through much of the high latitudes compared to 'SoilGrids+Moss'.~~ The efficient spectral method for albedo decay generally produces lower albedos than CLASS-CTEM's original exponential parameterization. The impact upon spring albedo and SWE leads to a notable decline in model performance compared to observation-based datasets (~~Figure~~

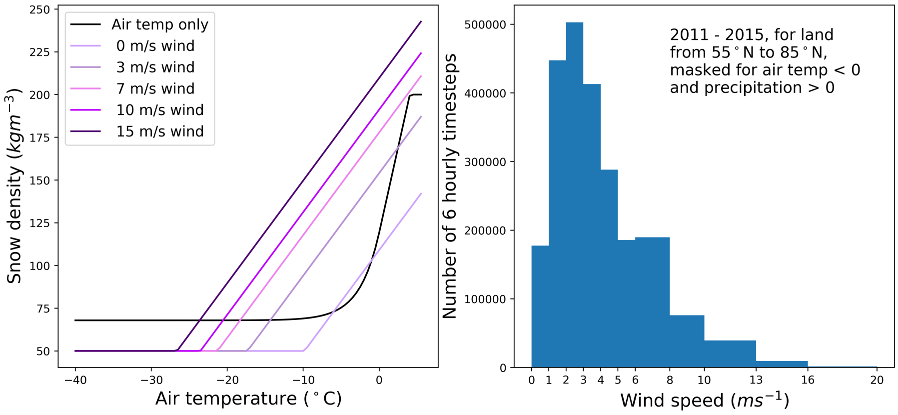

Figure 6. Left: Snow density as a function of air temperature for the original CLASS-CTEM formulation (Eqn A19) and for Exp. ~~'Fresh snow density'~~ *Fresh snow density* which includes consideration of wind speed (Eqn A20 -purple lines indicate different wind speeds). Right: Histogram of wind speeds for the period 2011 to 2015 ~~for~~ from the CRUNCEP meteorological dataset.

~~?? and ??).While the MODIS timeseries is limited, Exp. 'Snow albedo decay' is shown to under-perform having a relatively low ACC (<0.35), and slightly larger normalized standard deviation and root mean square error.~~ Supplement Figures 6 and 7).The CRUJRA55-forced experiments, on the other hand, give slightly better spring albedo for all experiments forced with that meteorological dataset. This could be due to the sub-monthly variability difference of CRUJRA55 compared to CRUNCEP as Beer et al. (2018) found one of the largest impacts of changing climate variability in model forcing to be snow depth~~, which will impact upon snow melt and surface albedo.~~. The lower albedo in the ~~'Snow albedo decay'~~ *Snow albedo decay* experiment leads to a smaller snowpack which melts earlier resulting in reduced spring runoff, a longer growing season, and a higher LAI. The warmer land surface results in larger ALTs. At the GTN-P borehole sites, the ~~'Snow albedo decay' experimentshows~~ *Snow albedo decay* experiment's warmer ground layers gives a noticeable increase in wMAE values across all seasons and most depth bands.

3.10 Allowing unfrozen water in frozen soils

The inclusion of unfrozen water in frozen soils (Exp. ~~'Super cooled water') leads to an increased PA of~~ *Super cooled water*) increased PA to 20.1 Mkm^2 ~~with no significant bias with depth compared to the GTN-P ALT sites (Figure ??). The average MAE~~ with a minor improvement at the GTN-P ALT sites ~~is slightly improved over 'SoilGrids+Moss' at 0.414 m~~ (Figure 3). ~~At the~~ The GTN-P borehole sites ~~, there is little impact evident~~ showed little change in the wMAE values (Figure 4). The larger PA for this experiment could be reflecting the thermal conductivity differences between completely frozen soil and frozen soil with some residual liquid water. The differences in bulk thermal conductivity would slow heat transfer into the deeper ground layers for the ~~'Super cooled water'~~ *Super cooled water* simulation during periods where the soil layer temperature is below 0°C. As a result spring warming would be slower to reach deeper layers. ~~The 'Super cooled water' experiment generally has~~

~~smaller latent heat fluxes, increased runoff and decreased LAI compared to 'SoilGrids+Moss' (Figure ??). The inclusion of unfrozen water in frozen soils leads to drier soils through downward movement of the liquid water while the soil temperature is below zero. The drier soils and the cooler soil temperatures lead to increased annual runoff and to lower latent heat fluxes and LAI.~~

Ganji et al. (2015) investigated streamflow for 21 watersheds in eastern Canada using CLASS and the WATROUTE routing scheme. They report their modifications (super-cooled soil water, fractional permeable area and modified hydrology due to ice; discussed in Section A7) improved streamflows particularly during the spring melt. The changes were attributed to reduced hydraulic conductivity of frozen soils causing more snow melt runoff and less infiltration. We ~~compared~~ did a rudimentary comparison of our simulated seasonal runoff for seven major river basins that drain permafrost regions (Figure 7~~). As mentioned previously (~~; Section 2.3)~~, this comparison is rudimentary and limited to seasonal discharges since the CLASS-CTEM runoff is not routed. An additional caveat is that~~ . As the CLASS-CTEM simulations did not include excess ground ice (e.g. slab ice such as ice wedges or lenses commonly found in regions affected by thermokarst processes), groundwater or interflow, all of which could increase runoff (baseflow) in the summer and fall seasons~~. As a result~~ , we limit our discussion to the spring and winter seasons. The ~~'Super cooled water'~~ *Super cooled water* experiment has lower spring runoff than both ~~'Base model' and 'SoilGrids+Moss'~~ *Base model* and *SoilGrids+Moss* but higher winter runoff, making it more in line with observed river discharges (Figure 7).

Given the ~~'Super cooled water' and 'Tian16 thermal cond.'~~ *Super cooled water* and *Tian16 thermal cond.* simulations had the lowest average MAE at the GTN-P ALT sites (Figure 3) a simulation was run with both of these parameterizations included (Exp. ~~'Super-cooled+Tian16'~~*Super-cooled+Tian16*). This experiment further reduced the average ALT MAE ~~to 0.288 m~~ but considerably worsened simulated ground temperatures at the GTN-P borehole sites (Figure 4). This incongruity between model performance at the ALT and borehole sites could be reflecting biases due to the spatial distribution of the sites (see Figure 1), the differing number of observations of ALT vs. borehole temperatures, or to biases in the observations themselves, which is discussed in Section 3.13.

3.11 Modifying hydrology due to ice

The ~~'Modif. hydrology'~~ *Modif. hydrology* experiment modified soil matric potential and saturated hydraulic conductivity to account for the impact of frozen water following the work of Farouki (1981) and Koren et al. (1999). These changes ~~to the soil hydrology led CLASS-CTEM to simulate a~~ yielded a simulated PA of 19.5 Mkm^2 (Table 2) with generally slightly deeper ALTs in much of the high latitude PD compared to ~~'SoilGrids+Moss'(Figures 2 and ??) . This experiment has~~ *SoilGrids+Moss* (Figure 2 and Supplement Figure 5) and poorer average MAE for the GTN-P ALT sites ~~(0.548 m) with a similar distribution of bias by grid cell (Figure 3) to the snow experiments above.~~ Performance at the GTN-P borehole sites is similar to Exp. ~~'SoilGrids+Moss'~~ *SoilGrids+Moss* (Figure 4). ~~The 'Modif. hydrology' experiment annual latent heat flux generally increases slightly in the Canadian PD along with a strong increase in LAI but a decrease in annual runoff compared to 'SoilGrids+Moss' (Figure ??).~~ Since the modifications to soil matric potential and saturated hydraulic conductivity (Equations A35 and A36) generally decrease water mobility in soils with ice present, the ~~'Modif. hydrology'~~ *Modif. hydrology* soils are generally wet-

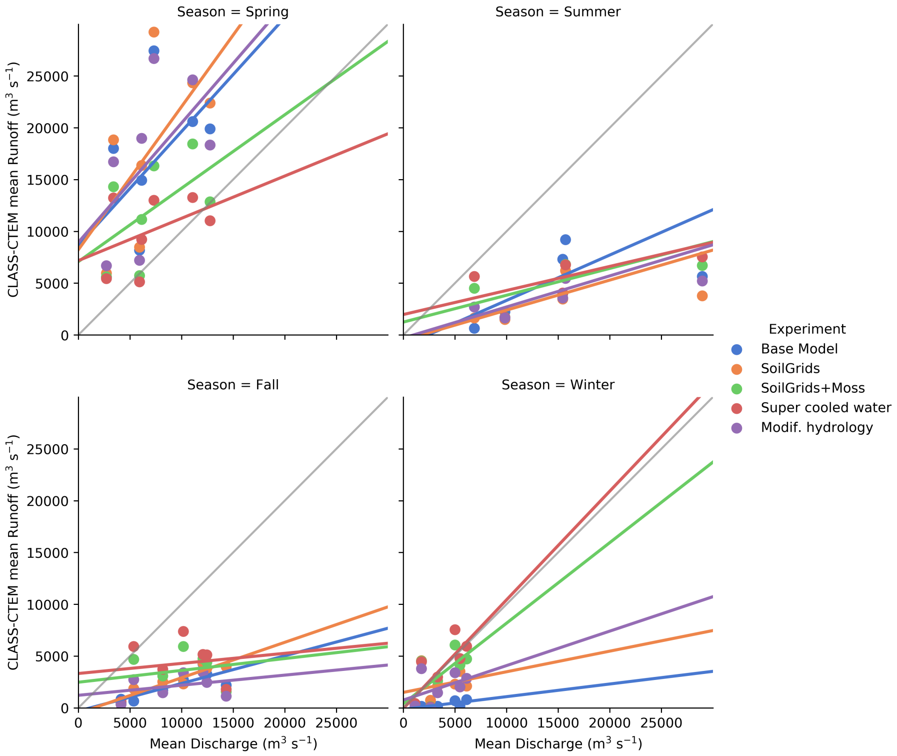

Figure 7. Mean 1965 - 1984 seasonal discharges of major rivers draining permafrost regions (Ob, Volga, Lena, Yenisei, Yukon, Mackenzie and Amur; UNESCO Press, 1993) compared to total runoff from selected model runs for the same period. Each dot represents one river basin. The CLASS-CTEM simulated runoff is not routed thus only seasonal values are compared.

ter, allowing higher annual latent heat flux and supporting higher LAI. The ~~’Modif. hydrology’~~ *Modif. hydrology* experiment has similar runoff to the ~~’Base model’~~ *Base model* experiment with higher spring runoff than observed river discharges while the winter runoff is reduced compared to ~~’SoilGrids+Moss’~~ *SoilGrids+Moss* and is also smaller than the observed river discharges (Figure 7). To investigate synergistic effects between the two modifications (~~’Modif. hydrology’ and ’Super cooled water’~~*Modif. hydrology* and *Super cooled water*), a simulation was run with both modifications applied (similar to Ganji et al.’s Exp. 3). This simulation gave slightly higher spring runoff but similar winter runoff compared to ~~’Modif. hydrology’~~ *Modif. hydrology* (not shown). Thus it appears, with respect to runoff, the modifications to hydrology have a stronger influence than super cooled soil water, in line with the conclusion of Ganji et al. (2015) that the primary effect is to reduce hydraulic conductivity which decreases infiltration and increases snow melt runoff.

3.12 Influence of sub-grid heterogeneity

The CLASS-CTEM model grid used in our study is the same as that used in the CanESM. ~~Based on the experiments in the previous section, with respect to the ALT sites of the GTN-P,~~ From the experiments conducted the lowest average MAE at

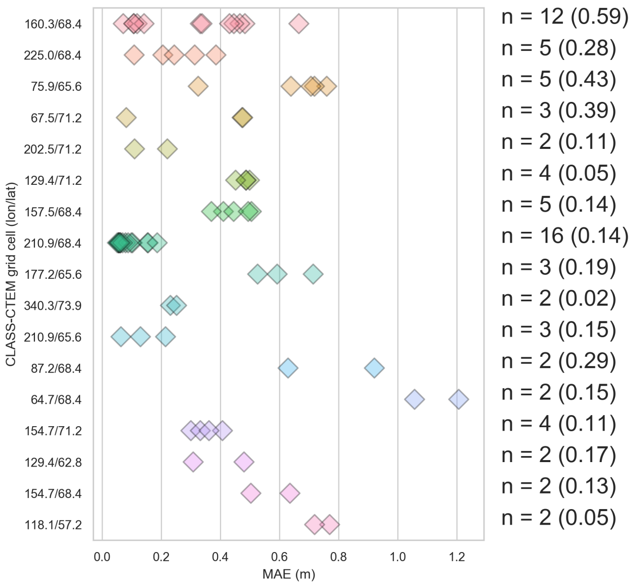

Figure 8. Mean absolute error (MAE) for CLASS-CTEM grid cells with multiple GTN-P ALT sites for the ~~'SoilGrids+Moss'~~ *SoilGrids+Moss* simulation. The number of ALT sites is listed along with the range in MAE in each grid cell in parentheses.

the GTN-P ALT sites we are able to achieve is about 0.4 m. With the size of our model grid cells, what is the best MAE we can reasonably expect given the sub-grid heterogeneity at the observation sites? Many of the GTN-P ALT measurements are performed on an 11 x 11 sampling grid covering between 1 km^2 and 1 ha giving 121 data points at one point in time per site; the mean standard deviation of measured ALT over these sampling grids varies from 0.02 m to 0.49 m (~~Table ??~~Supplement Table 1). However, one square kilometre is still small compared to model grids ranging in size from hundreds to thousands of km^2. One measure of the influence of sub-grid heterogeneity can be obtained by considering the MAE per site in the grid cells where we have more than one GTN-P ALT site (Figure 8). For these grid cells, the spread in MAE at each site ranges from 0.01 m (grid cell with 2 sites) to 0.59 m (12 sites)~~for the 'SoilGrids+Moss' simulation~~. While it is not reasonable to directly compare the sub-grid range of MAE to the model average MAE shown in Figure 3, Figure 8 demonstrates that sub-grid heterogeneity is a significant source of variability in ALT within model grid cells and that variability will impose constraints on the lower limit of MAE that is attainable by the model.

For the GTN-P borehole sites, the wMAE in temperature bias for the model varies between ca. 1.5 and 3.7 °C (Figure 4), depending on depth and season. As with ALT, what is a reasonable wMAE for ground temperatures given the size of the model grid cells and the discrete nature of a borehole? To better understand the role of sub-grid heterogeneity in borehole temperatures, we make use of the Slave Province Surficial Materials and Permafrost Study (SPSMPS; Gruber et al., 2018). The SPSMPS collected air and ground temperature measurements for 15 m x 15 m plots with hourly borehole temperatures at thirty-five boreholes all located within a ca. 1200 km^2 area. The observed screen-level temperatures are generally reasonably close to those of CRUNCEP but CRUNCEP has slightly cooler summer temperatures (~~Figure ??~~Supplement Figure 8). What is most striking about the borehole temperatures at Lac de Gras is the large spread in ground temperatures at all depths and in most seasons (Figure 9). The temperature range is smallest in fall and spring when the soils are thawing or freezing and largest in winter with differences varying from 12 to over 20 °C depending on the soil depth. This remarkable spread in temperature is due to variations in slope, aspect, soil moisture, soil texture, soil organic matter content, and vegetation type and distribution. The simulated ground temperatures from two experiments are plotted alongside the boreholes (~~'SoilGrids' and 'SoilGrids+Moss'~~*SoilGrids* and *SoilGrids+Moss*). As the model is driven by CRUNCEP and we have no precipitation information for the SPSMPS sites, it is difficult to determine the cause of any biases. Also, although the SPSMPS sampling area is considerably larger than the GTN-P sites, the same arguments apply concerning the mismatch of scales between the observational area and the model grid, and the variability introduced by sub-grid heterogeneity.

An additional measure of how reasonable the model wMAE is at the borehole sites can be obtained by comparing the CRUNCEP screen-level temperature, which is used to force the model, and the observed screen-level temperature at each GTN-P site. The MAE for screen-level temperature is between 2.17 and 2.53 °C across all seasons. Therefore the model's wMAE range for shallow soil of ca. 3 to 3.7 °C varies from ca. 0.8 to 1.2 °C above that of the MAE for CRUNCEP's screen-level temperature (for the ~~'SoilGrids+Moss'~~ *SoilGrids+Moss* simulation). Given the large spread in borehole temperatures in a relatively small area at the SPSMPS sites, and the MAE of the model's forcing air temperature, it appears the model's wMAE can be considered reasonable.

3.13 Influence of bias due to ALT or borehole sampling locations

Temperature in individual boreholes and ALT at individual sites often differ from the grid cell they are compared with because of sub-grid variability as discussed above. The underlying spatial variation of ground temperature, even at distances smaller than 1km is well documented (Smith, 1975; Morse et al., 2012; Gubler et al., 2011). If the locations of GTN-P sites were randomly sampled, sub-grid effects would be expected to cancel out and, consequentially, a mean bias (cf. Figure 3) close to zero would be indicative of good model performance. In reality, however, the choice of GTN-P measurement locations are likely biased and the nature and consequences of this bias are difficult to assess. For example, ALT sites are likely to be biased toward fine-grained and organic-rich soils and locations with small ALT where probing can be carried out. The choice of ALT and borehole sites in areas of sporadic permafrost is likely to be biased towards cold areas in the landscape. This is because ALT requires permafrost and because permafrost researchers are unlikely to drill, instrument and operate boreholes in seasonally frozen ground. Finer-scale local studies have noted that observations are strongly biased towards permafrost existence (Boeckli et al., 2012). The melt of excess ice from the top of permafrost presents an additional source of bias that may result in ALT data showing values of seasonal thaw depth that underestimate the amount of ground ice that was melted ~~. This is because ALT obtained from~~ due to frost-table probing without recording surface subsidence (Shiklomanov et al., 2013)~~omits an important part of the changes in the natural environment~~. In summary, it is likely that a slightly positive model bias, i.e. higher temperatures and greater ALT simulated than observed, would correspond to a model that best represents reality. Quantifying that effect however is beyond the present study.

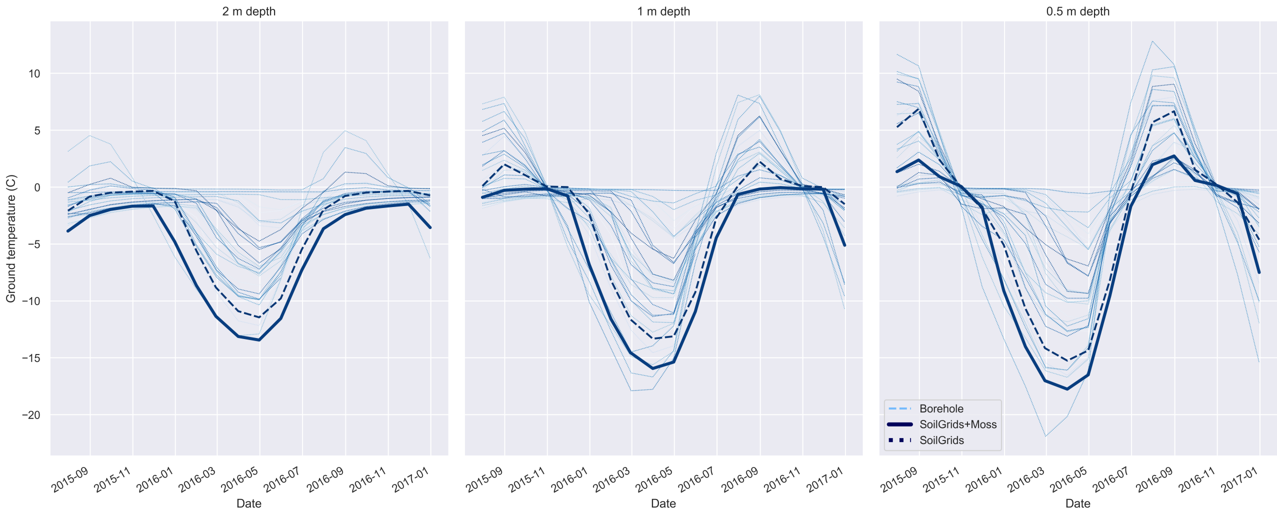

Figure 9. Borehole temperatures for 0.5, 1, and 2 m depths from the Slave Province surficial materials and permafrost study (SPSMPS, Lac de Gras region, NWT, Canada; ~~(Gruber et al., 2018)~~Gruber et al. (2018)) along with CLASS-CTEM simulated ground temperatures for Exp. ~~'SoilGrids'~~ *SoilGrids* and ~~'SoilGrids+Moss'~~*SoilGrids+Moss*. The thirty-five boreholes are each represented by a single line and are all located within a ca. 1200 km^2 area. The model output is from the grid cell corresponding to the SPSMPS study area.

4 Conclusions

The performance of CLASS-CTEM in cold regions has been investigated in the past by numerous researchers ~~(Tilley et al., 1997; Bellisario~~ ~~. As a result, several modifications have been suggested~~ who have suggested several modifications to improve the model's performance in these regions. Drawing from these recommendations and other studies, 18 experiments were carried out to investigate the influence of: 1) the number of ground layers, 2) soil permeable depth datasets, 3) the addition of a moss layer, 4) changing the soil thermal conductivity formulation, 5) altering the derivation of snow cover based on snow depth, 6) adding the effect of wind speed to the calculation of fresh snow density, 7) changing the model's snow albedo decay calculation to an efficient spectral parameterization, and 8) modifications to frozen soil hydrology including allowing unfrozen water in frozen soils and an alteration to hydraulic conductivity and soil matric potential for the presence of ice. Two soil permeable depth datasets were tested (Pelletier et al. (2016) and SoilGrids; Shangguan et al. (2017)) along with two meteorological datasets (CRUNCEP v.8; Viovy (2016) and CRUJRA55 v.1.0.5; Kobayashi et al. (2015); Harris et al. (2014)). The simulated active layer thicknesses (ALTs) were compared to 1570 observations from 97 sites from the Global Terrestrial Network for Permafrost (GTN-P; ~~Table ??~~Supplement Table 1, Figure 1), the simulated soil temperatures to 105 106 monthly observations at 132 GTN-P borehole temperature sites (~~Table ??~~Supplement Table 2), 35 borehole sites from the Slave Province Surficial Materials and Permafrost Study (SPSMPS; Gruber et al., 2018), surface albedo to a remotely-sensed dataset (MODIS MCD43C3), snow water equivalent (SWE) to a blend of five observation-based datasets (Blended5; Mudryk et al., 2015), and seasonal runoff to river discharges for major rivers draining the Arctic (UNESCO Press, 1993) as well as literature estimates of permafrost area (Table 2).

The original model version had an overly small simulated permafrost area of 8.6 Mkm^2 which was almost doubled to 16.7 Mkm^2 by increasing the number and depth of ground layers. Of the two soil permeable depth datasets, Pelletier et al. (2016) gave consistently lower average mean absolute errors (MAE) at the GTN-P ALT sites compared to SoilGrids. However, SoilsGrids was chosen for further simulations as this dataset appears to be better validated (Shangguan et al., 2017). For the two meteorological datasets used, the permafrost specific results depended on the model configuration and parameterizations tested. More consistently, spring albedo appeared to be better simulated using CRUJRA55 while winter SWE was slightly better with CRUNCEP. Changes to the model configuration by increasing soil permeable depths using the SoilGrids dataset, and adding a layer of moss reduced the average MAE at the GTN-P ALT sites from over 2.5 m (Exp. ~~'20 ground layers'~~*20 ground layers*) to 0.472 m (Exp. ~~'SoilGrids+Moss'~~*SoilGrids+Moss*). While most alternate parameterizations either degraded model performance at the GTN-P ALT and borehole sites, or degraded the performance of another model output such as albedo or SWE, incorporating unfrozen water in frozen soils following Niu and Yang (2006) is being considered for inclusion in future versions of CLASS-CTEM. A simulation with the Niu and Yang (2006) parameterization resulted in an average MAE of 0.414 m at the GTN-P ALT sites, relatively small impacts on wMAE at the GTN-P borehole sites, and a possible improvement in seasonal runoff. Further assessment of the improvements in runoff using a river routing scheme are needed before this parameterization will be fully adopted. Based on the tests performed here, the optimal model configuration will include more ground layers to a greater depth, soil permeable depths from the SoilGrids dataset, and moss in locations where it is appropriate. These changes give a simulated permafrost area of between 15.7 to 17.9 Mkm^2 (Table 2) which reasonably close to the expected 15 Mkm^2 ~~for the 'region_50'~~ based on published estimates derived from mean annual air temperature (see discussion in Section 2.3.4) .

There are six main limitations of our study. First, thermokarst processes due to melt of excess ground ice (ice wedges or lenses) are not simulated. As maps of ground ice extent improve (e.g. O'Neill et al., 2018) and become more suitable for use as a model geophysical field, parameterizations such as Lee et al. (2014) could be incorporated. Second, our treatment of mosses and their impact is simplistic. A more comprehensive approach such as the LiBry model (Porada et al., 2016) would allow for dynamic moss extents and more bryophyte subtypes including lichens. Third, the plant functional types used here are not specific to the Arctic and do not include shrubs. Shrubs, in particular, are presently expanding and have complex impacts upon Arctic regions (e.g. Figure 3 in Myers-Smith et al., 2011). Fourth, orographic influences on permafrost such as slope and aspect were not resolved. Fifth, inland water bodies and their impact upon ground thermal regimes were not considered. Finally, the influence of sub-grid heterogeneity was ignored as permafrost in the model grids is binary thus excluding the simulation of ~~taliks~~discontinuous permafrost. With regard to the influence of sub-grid heterogeneity, the standard deviation of ALT on the 1 km^2 - 1 ha measurement grids at the GTN-P ALT sites, the spread in MAE in grid cells with multiple GTN-P ALT sites, and the SPSMPS collection of 35 boreholes over a 1200 km^2 study area indicate that it is likely difficult to reduce the wMAE of ALT or borehole temperature much further, given the size of the model grid cells (ca. 2.8°). Based on the model physics performance presented here, it appears that with the modifications described above, the land surface scheme in CLASS-CTEM is well-suited to provide the physical conditions for simulating carbon fluxes in the permafrost domain.

Code availability. CLASS-CTEM is available as a tarball from Melton (2019). The following code tags correspond to experiments in this manuscript (see Table 1) with the most strongly impacted subroutines in parentheses: 1) Base model: 'archive/baseModelPermafrostPhysics', 2) deVries thermal cond.: 'archive/soilthermalcond' (TPREP), 3) Tian16 thermal cond.: 'archive/Tian16SoilThermalCond (TPREP), 4) Snow cover: Yang97/Brown03: 'archive/snowcov_changes' (CLASSA), 5) Fresh snow density: 'archive/snowdens' (CLASSI), 6) Snow albedo decay: 'archive/snowalbedorefresh' (CLASSA,SNOALBA), 7) Super-cooled water: 'archive/supercooledH2O' (CLASSB,TMCALC,TWCALC), and 8) Modif. hydrology: 'archive/arman' (GRDRAN,GRDINFL). The model manual is located within the code repository (/documentation/html/index.html)

Appendix A: Description of alternate parameterizations

A1 Moss parameterization of Wu et al. (2016)

The simple moss parameterization used here follows Wu et al. (2016) with the exception that our moss layer is non-photosynthesizing. The physical characteristics of the moss layer includes a pore volume of 0.98 m^3m^{-3}, liquid water retention capacity of 0.2 m^3m^{-3}, the residual liquid water content after freezing or evaporation of 0.01 m^3m^{-3}, the Clapp and Hornberger empirical "b" parameter set to 2.3, a soil moisture suction at saturation of 0.0103 m, a saturated hydraulic conductivity of 1.83×10^{-3} m s^{-1}, a volumetric heat capacity of 2.5×10^{-6} J $m^{-3}K^{-1}$, with the thermal conductivity of the moss set to that of organic matter (0.25 W m^{-1} K^{-1}).

A2 Soil thermal conductivity

CLASS-CTEM calculates the thermal conductivities of organic and mineral soils following Côté and Konrad (2005). The soil thermal conductivity, λ (W m^{-1} K^{-1}), is modelled via a relative thermal conductivity, λ_r, which varies between a value of 1 at saturation and 0 for dry soils:

$$\lambda = [\lambda_{sat} - \lambda_{dry}]\lambda_r + \lambda_{dry} \tag{A1}$$

Using the following generalized relationship, the relative thermal conductivity is obtained from the degree of saturation (the water content divided by the pore volume), S_r (unitless):

$$\lambda_r = \frac{\kappa S_r}{[1 + (\kappa - 1)S_r]} \tag{A2}$$

Based on the soil characteristics and state, the empirical coefficient, κ (W m^{-1} K^{-1}), takes the following values:

1. Unfrozen coarse mineral soils: $\kappa = 4.0$
2. Frozen coarse mineral soils: $\kappa = 1.2$
3. Unfrozen fine mineral soils: $\kappa = 1.9$

4. Frozen fine mineral soils: $\kappa = 0.85$

5. Unfrozen organic soils: $\kappa = 0.6$

6. Frozen organic soils: $\kappa = 0.25$

The dry thermal conductivity, λ_{dry}, is calculated via an empirical relationship using the pore volume, θ_p (m^3 m^{-3}), with different coefficients for organic and mineral soils:

$$\lambda_{dry,mineral} = 0.75e^{(-2.76\theta_p)} \tag{A3}$$

$$\lambda_{dry,organic} = 0.30e^{(-2.0\theta_p)} \tag{A4}$$

While the saturated thermal conductivity, λ_{sat}, is calculated by Côté and Konrad (2005) as a geometric mean of the conductivities of the soil components, other studies (e.g. Zhang et al., 2008) have found the linear averaging used by de Vries (1963) to be generally more accurate and this approach has been adopted by CLASS-CTEM,

$$\lambda_{sat,unfrozen} = \lambda_{liq}\theta_p + \lambda_s(1-\theta_p) \tag{A5}$$

$$\lambda_{sat,frozen} = \lambda_{ice}\theta_p + \lambda_s(1-\theta_p) \tag{A6}$$

where λ_{ice} is the thermal conductivity of ice, λ_{liq} is that of liquid water and λ_s is that of the soil solid particles.

Exp. ~~'deVries thermal cond.'~~ *deVries thermal cond.* replaces the CLASS-CTEM default soil thermal conductivity parameterization with that of de Vries (1963):

$$\lambda = \frac{\lambda_{liq}\theta_{liq} + f_a\lambda_a\theta_a + f_s\lambda_s\theta_s}{\theta_{liq} + f_a\theta_a + f_s\theta_s} \tag{A7}$$

where the a subscript denotes the air component, θ is the volumetric fraction, and f is the 'weighting' factor (unitless) which is given by:

$$f_s = \frac{1}{3}\left[\frac{2}{1+0.125(\frac{\lambda_s}{\lambda_{liq}}-1)} + \frac{1}{1+0.75(\frac{\lambda_s}{\lambda_{liq}}-1)}\right] \tag{A8}$$

$$f_a = \frac{1}{3}\left[\frac{2}{1+g_a(\frac{\lambda_a}{\lambda_{liq}}-1)} + \frac{1}{1+(1-2g_a)(\frac{\lambda_a}{\lambda_{liq}}-1)}\right] \tag{A9}$$

where g_a represents a unit-less empirical air pore-shape factor,

$$g_a = \begin{cases} 0.333 - (0.333 - 0.035)\frac{\theta_a}{\theta_p}, & \theta_{liq} > 0.09 \\ 0.013 + 0.944\theta_{liq}, & \theta_{liq} \leq 0.09 \end{cases} \tag{A10}$$

An alternate approach is tested in Exp. ~~’Tian16 thermal cond.’~~*Tian16 thermal cond.*. The Tian et al. (2016) thermal conductivity parameterization is based upon the de Vries (1963) formulation, but simplifies and extends it to both frozen and unfrozen soils. In their formulation, Tian et al. adapt equation A7 to include ice and organic matter as,

$$\lambda = \frac{\lambda_{liq}\theta_{liq} + f_{ice}\lambda_{ice}\theta_{ice} + f_a\lambda_a\theta_a + f_s\lambda_s\theta_s + f_{organic}\lambda_{organic}\theta_{organic}}{\theta_{liq} + f_{ice}\theta_{ice} + f_a\theta_a + f_s\theta_s + f_{organic}\theta_{organic}} \tag{A11}$$

for wet soil whereas the thermal conductivity of completely dry soils is calculated by,

$$\lambda = 1.25\frac{f_a\lambda_a\theta_a + f_s\lambda_s\theta_s + f_{organic}\lambda_{organic}\theta_{organic}}{f_a\theta_a + f_s\theta_s + f_{organic}\theta_{organic}} \tag{A12}$$

The Tian et al. formulation also modifies the pore-shape factor (equation A10) to be,

$$g_a = 0.333 - \left(1 - \frac{\theta_a}{\theta_p}\right) \tag{A13}$$

for air and

$$g_{ice} = 0.333 - \left(1 - \frac{\theta_{ice}}{\theta_p}\right) \tag{A14}$$

for ice. Tian et al. (2016) introduce a shape factor for ellipsoidal soil particles, g_m as,

$$g_m = g_{sand}\theta_{sand} + g_{silt}\theta_{silt} + g_{clay}\theta_{clay} \tag{A15}$$

where g_{sand} is 0.182, g_{silt} is 0.00775, and g_{clay} is 0.0534. The shape factor for organic soils, $g_{organic}$, is set to 0.5. The same ’weighting’ factor is used for ice, air, organic and mineral soil components and left unchanged from equation A9.

A3 Snow cover fraction

CLASS-CTEM relates snow depth, (d_{snow}; m), to snow cover, (f_{snow}; fraction), via a linear function (~~Figure ??~~Supplement Figure 2) (Verseghy, 2017),

$$f_{snow} = min\left[1, \left(\frac{d_{snow}}{d_0}\right)\right] \tag{A16}$$

where d_0 is a limiting snow depth assigned a value of 0.1 m. Exp. 'Snow cover:Yang97' changes the CLASS-CTEM linear function to a hyperbolic tangent function (Yang et al., 1997),

$$f_{snow} = tanh\left(\frac{d_{snow}}{d_0}\right) \tag{A17}$$

Another alternative parameterization for snow cover from snow depth was proposed by Brown et al. (2003), which was not evaluated in MacDonald (2015). This relation was developed based on analysis of a global gridded snow water equivalent product designed to evaluate GCMs. Exp 'Snow cover:Brown03' tests the impact of that parameterization by changing the snow cover function to the proposed exponential form (Brown et al., 2003),

$$f_{snow} = 1 - 0.01(15 - 100d_{snow})^{1.7}. \tag{A18}$$

A4 Fresh snow density

The density of freshly fallen snow is related to its ice-crystal structure and the volume of the ice crystal that is occupied by air. Generally, snow density is the result of 1) processes occurring in the cloud that affect the size and shape of the growing ice crystals, 2) processes that modify the crystal as it falls, and 3) compaction on the ground due to prevailing weather conditions and metamorphism in the snowpack (Roebber et al., 2003).

Fresh snow density (ϱ; kg m^{-3}) in CLASS-CTEM is calculated based on air temperature (T_a; K). For air temperatures below freezing, (T_f), a relation from Hedstrom and Pomeroy (1998) is used, while for temperatures at or above freezing CLASS-CTEM uses an equation from Pomeroy and Gray (1995),

$$\varrho = \begin{cases} 67.92 + 51.25e^{\left[\frac{(T_a - T_f)}{2.59}\right]} & T_a < T_f \\ 119.17 + 20(T_a - T_f) & T_a \geq T_f \end{cases} \tag{A19}$$

In Exp. ~~'Fresh snow density'~~*Fresh snow density*, the effect of wind speed (u, m s^{-1}) is included following the approach used in the CROCUS model as detailed in Essery et al. (1999) with a minimum density of 50 kg m^{-3} following MacDonald (2015):

$$\varrho = max[50, 109 + 6(T_a - T_f) + 26u^{1/2}] \tag{A20}$$

Wind speed may be considered important in determining fresh snow density as wind speeds greater than approximately 9 m s^{-1} can move ice crystals on the surface leading to crystal fractionation during saltation and surface compaction increasing the snow density (e.g., Gray and Male, 1981, p. 345–350).

A5 Snow albedo decay

Snow albedo (α_s; unitless) decreases as snow ages due to snow grain growth and deposition of soot and dirt (Wiscombe and Warren, 1980). In CLASS-CTEM this process is treated via empirical exponential decay functions (Verseghy, 2017). Freshly fallen snow is given a total albedo ($\alpha_{fs,total}$) value of 0.84, a visible ($\alpha_{fs,visible}$) value of 0.95 and a near-infrared (NIR; $\alpha_{fs,nir}$) value of 0.73 . It is assumed that the same decay function, calculated each timestep (Δt; 1800 s) applies to all three albedo ranges,

$$\alpha_{s,total}(t+\Delta t) = \alpha_{s,total,old} + [\alpha_{s,total}(t) - \alpha_{s,total,old}]e^{\left(-\frac{0.01\Delta t}{3600}\right)} \tag{A21}$$

If the snowpack temperature is greater than -0.01 °C or the melt rate at the top of the snowpack is not negligible, $\alpha_{s,total,old}$ is set to a value characteristic of melting snow (0.50) otherwise it is set a value representing old, dry snow (0.70). The total albedo at a given time step is converted to those of the visible and NIR ranges for dry snow via,

$$\alpha_{s,visible} = 0.7857\alpha_{s,total} + 0.29 \tag{A22}$$

$$\alpha_{s,nir} = 1.2142\alpha_{s,total} - 0.29 \tag{A23}$$

and for melting snow,

$$\alpha_{s,visible} = 0.9706\alpha_{s,total} + 0.1347 \tag{A24}$$

$$\alpha_{s,nir} = 1.0294\alpha_{s,total} - 0.1347 \tag{A25}$$

Exp. ~~'Snow albedo decay'~~ *Snow albedo decay* replaces the CLASS-CTEM exponential decay function with a spectral method based on Wiscombe and Warren (1980) and adapted for efficiency by Dickinson (1983). This efficient spectral method first calculates the diffuse radiation albedo based on the albedo of fresh snow and the transformed snow age factor (F_{age})

$$\alpha_{dif,visible} = (1 - 0.2F_{age})\alpha_{fs,visible} \tag{A26}$$

$$\alpha_{dif,nir} = (1 - 0.5F_{age})\alpha_{fs,nir} \tag{A27}$$

$$F_{age} = \frac{\tau_s}{1 + \tau_s} \tag{A28}$$

where τ_s is a non-dimensional snow age at each timestep found via

$$\tau_s(t+\Delta t) = \left[\tau_s(t) + \frac{(r_1+r_2+r_3)\Delta t}{\tau_0}\right]\left(1 - \frac{S_f \Delta t}{\Delta P}\right) \tag{A29}$$

where r_1 represents the effects of grain growth due to vapor diffusion as

$$r_1 = e^{\left[5000\left(\frac{1}{T_f} - \frac{1}{T_{g,1}}\right)\right]} \tag{A30}$$

and $r_2 = r_1^{10}$, representing the additional effects at or near the freezing of meltwater on grain growth. r_3 represents the effects of soot and dirt and is set to 0.3. $T_{g,1}$ is the temperature of the top soil layer (K), τ_0 is 10^6s, S_f is the snowfall rate for that timestep (kg m^{-2} s^{-1}), and ΔP is the snow fall amount threshold (10 kg m^{-2}). If, within a timestep, the fresh snowfall amount exceeds ΔP, the snow age is set to that of new snow ($\tau_s = F_{age} = 0$).

The direct radiation albedos are found by

$$\alpha_{dir,visible} = \alpha_{dif,visible} + 0.4 f(\mu)(1 - \alpha_{dif,visible}) \tag{A31}$$

$$\alpha_{dir,nir} = \alpha_{dif,nir} + 0.4 f(\mu)(1 - \alpha_{dif,nir}) \tag{A32}$$

where $f(\mu)$ is a factor that scales between 0 and 1 to give increased snow albedo due to solar zenith angles exceeding 60 °, calculated as

$$f(\mu) = \max\left[0, \frac{1 - 2\cos Z}{1 + b_\mu}\right] \tag{A33}$$

where Z is the solar zenith angle and b_μ is an adjustable parameter set to 2 following the BATS model (Yang et al., 1997).

A6 Super-cooled soil water

In experiment ~~'Super-cooled water'~~*Super-cooled water*, unfrozen soil water in frozen soils is introduced into CLASS-CTEM following Niu and Yang (2006). Unfrozen water can exist in frozen soils through the capillary and absorptive forces exerted by soil particles on water in close proximity. The upper limit on the residual amount of water that can remain liquid under given soil temperature and texture conditions is parameterized by Niu and Yang (2006) as,

$$\theta_{liq,max} = \theta_p \left(\frac{-L_f(T_{soil,i} - T_f)}{g\psi_{sat} T_{soil,i}}\right)^{-1/b} \tag{A34}$$

where g is gravitational acceleration (m s^{-2}), L_f is the latent heat of fusion (J kg^{-1}), and $T_{soil,i}$ is the soil layer temperature (K). According to Romanovsky and Osterkamp (2000) unfrozen water content in moss is negligible so $\theta_{liq,max}$ is set to zero for moss layers.

A7 Modified hydrology

In Ganji et al. (2015) several changes were implemented in CLASS to address how the model deals with frozen soil water. First, super-cooled soil water was added following Niu and Yang (2006) as described above. Secondly, fractional impermeable area was introduced, also following Niu and Yang (2006), but this has little impact upon our model simulations (discussed in Appendix B). Their final modification was to account for the impact of frozen water on the soil matric potential (ψ; m) after Farouki (1981) and Koren et al. (1999) by adding a new term $[(1+C_k\theta_{ice})^2]$ to the existing CLASS functional relationship,

$$\psi = \psi_{sat}\left(\frac{\theta_{liq}}{\theta_p}\right)^{-b}(1+C_k\theta_{ice})^2 \tag{A35}$$

where C_k is a constant, set to 8, that accounts for the effect of an increase in specific surface area of soil minerals and liquid water as water freezes and ice forms (Kulik, 1978). ψ_{sat} is the soil matric potential at saturation (m) and b is the Clapp and Hornberger empirical 'b' parameter (unitless) (Clapp and Hornberger, 1978). The calculation of hydraulic conductivity $k; ms^{-1}$ is also modified by multiplication with a similar term $[(1+C_k\theta_{ice})^{-4}]$,

$$k = k_{sat}\left(\frac{\theta_{liq}}{\theta_p}\right)^{2b+3}(1+C_k\theta_{ice})^{-4} \tag{A36}$$

where k_{sat} is saturated hydraulic conductivity. The effect of these changes is to generally increase soil matric potential and decrease hydraulic conductivity when ice is present in the soil. These modifications are tested in Exp. ~~'Modif. hydrology'~~ *Modif. hydrology*.

Appendix B: Fractional permeable areas in frozen soils

CLASS-CTEM accounts for the impact of frozen soil water through an empirical correction factor (f_{ice}; unitless), according to Zhao and Gray (1997).

$$f_{ice} = \left[1 - min\left(1, \frac{\theta_{ice}}{\theta_p}\right)\right]^2 \tag{B1}$$

This factor is used to correct the calculated soil hydraulic conductivity, k (m s^{-1}) which is found via the Clapp and Hornberger (1978) equation:

$$k = f_{ice}k_{sat}\left(\frac{\theta_{liq}}{\theta_p}\right)^{2b+3} \tag{B2}$$

where k_{sat} is the hydraulic conductivity at saturation and b is an empirical parameter. Soil moisture is related to soil matric potential (ψ; m) in CLASS-CTEM following Clapp and Hornberger (1978),

$$\psi = \psi_{sat}\left(\frac{\theta_{liq}}{\theta_p}\right)^{-b} \tag{B3}$$

where ψ_{sat} is the saturated soil matric potential (m).

Niu and Yang (2006) parameterize fractional permeable areas in frozen soils. Following their formulation, within a grid cell the permeable $(perm)$ and impermeable (imp) patches affect the flux of water (q; m s^{-1}) as

$$q = F_{imp}q_{imp} + (1 - F_{imp})q_{perm} \tag{B4}$$

where the impermeable grid cell fraction, F_{imp} can be estimated as

$$F_{imp} = e^{-\alpha\left(1-\frac{\theta_{ice}}{\theta_p}\right)} - e^{-\alpha} \tag{B5}$$

and α is set to 3 following Niu and Yang (2006). Assuming q_{imp} is set to zero, Niu and Yang parameterize the influence of the permeable areas on hydraulic conductivity can be parameterized as

$$k = (1 - F_{imp})k_{sat}\left(\frac{\theta_{liq} + \theta_{ice}}{\theta_p}\right)^{2b+3} \tag{B6}$$

while the soil matric potential is calculated as,

$$\psi = \psi_{sat}\left(\frac{\theta_{liq} + \theta_{ice}}{\theta_p}\right)^{-b} \tag{B7}$$

This formulation results in a soil matric potential that is insensitive to ice content within the soil (~~Figure ??~~Supplement Figure 9) which seems unreasonable (see for example Wen et al., 2012). This fact is indeed noted by Ganji et al. (2015) who state that the soil matric potential as defined by Niu and Yang (2006) is not appropriate for the case of frozen soil. The inclusion of θ_{ice} in the numerator could be a typographical error. If it is removed the hydraulic conductivity and soil matric potential behave quite similarly to the original CLASS relations which make use of the factor f_{ice} in place of $1 - F_{imp}$ (~~Figure ??~~Supplement Figure 10). Testing shows the model is relatively insensitive to the small changes visible in the plots (not shown).

Table A1. Ground layer depths and thicknesses for the 20 ground layer configuration.

Layer number	Thickness (m)	Depth (m)
1	0.1	0.1
2	0.1	0.2
3	0.1	0.3
4	0.1	0.4
5	0.1	0.5
6	0.1	0.6
7	0.1	0.7
8	0.1	0.8
9	0.1	0.9
10	0.1	1.0
11	0.2	1.2
12	0.3	1.5
13	0.4	1.9
14	0.5	2.4
15	1.0	3.4
16	3.0	6.4
17	5.0	11.4
18	15.0	26.4
19	30.0	56.4
20	5.0	61.4

~~GTN-P ALT sites used in the CLASS-CTEM evaluation. The mean ALT is calculated by first taking the mean of the sampling grid for each observation in time and then taking the mean across all observation times at each site. The standard deviation (SD) at each site is calculated across the sampling grid. The mean SD is then the average SD of the sampling grid across all observation times. Name Latitude Longitude Obs. period Mean obs./yr Mean ALT Mean SD (°N) (°W) (m) over ALT grid (m) 1 Cape Chukochii R13a 70.08 159.92 2000 - 2015 0.94 0.41 0.09 2 Taglu Grid 69.37 -134.95 1998 - 2008 1.27 0.99 0.22 3 Mt Rodinka PLOT 68.75 161.50 2003 - 2015 1.00 0.99 0.14 4 Mt Rodinka Control Site 68.73 161.40 2003 - 2015 1.00 1.35 0.11 5 Urengoy GAS FIELD GP5 66.32 76.91 2008 - 2014 1.14 0.77 0.41 6 Vaskiny Dachi 2 70.30 68.88 2007 - 2015 1.00 0.70 0.12 7 Mt Rodinka Burn Site 68.72 161.53 2003 - 2015 0.77 1.53 0.09 8 Barrow 71.32 -156.60 1995 - 2015 1.52 0.36 0.08 9 Panteleekha River 68.42 161.22 1996 - 1996 1.00 0.45 0.12 10 Tiksi 71.58 128.78 1997 - 2000 1.25 0.42 0.11 11 Chukochya River 69.49 156.99 1996 - 2015 0.65 0.44 0.06 12 Betty Pingo 70.28 -148.87 1995 - 2015 1.43 0.52 0.17 13 Cape Rogozny 64.78 176.97 1994 - 2015 1.05 0.50 0.05 14 Zackenberg ZEROCALM 2 74.47 -20.50 1996 - 2010 9.27 0.63 0.13 15 Chandalar Shelf 68.07 -149.58 1996 - 2015 0.90 0.36 0.09 16 Old Man 66.45 -150.62 1996 - 2015 0.85 0.40 0.04 17 Franklin Bluff 69.68 -148.72 1996 - 2015 1.00 0.62 0.15 18 Alexandria Fiord 78.88 -75.92 1996 - 2001 0.67 0.57 0.07 19 North Head Grid 69.72 -134.46 1998 - 2008 1.00 0.47 0.08 20 Lousy Point Grid 69.22 -134.29 1998 - 2008 1.00 0.56 0.13 21 Wickersham 65.27 -148.05 1972 - 2015 0.93 0.47 0.08 22 Allaiha 70.56 147.43 2004 - 2015 1.00 0.47 0.08 23 Talnakh 69.43 88.47 2008 - 2015 1.25 0.95 0.21 24 Igarka 67.48 86.44 2008 - 2017 1.10 0.76 0.32 25 Happy Valley 1km 69.10 -148.50 1995 - 2015 1.52 0.44 0.09 continued 26 West Dock 1ha 70.37 -148.55 1996 - 2015 1.00 0.31 0.07 27 Marre Sale 69.72 66.75 1995 - 2015 1.00 1.10 0.33 28 Segodnya Pingo 69.09 158.90 1996 - 2015 0.70 0.51 0.12 29 Talnik 67.33 63.73 1998 - 2015 2.50 1.29 0.27 30 Zackenberg ZEROCALM 1 74.47 -20.50 1996 - 2010 7.87 0.71 0.07 31 Malchikovskaya Channel 68.52 161.43 1996 - 2015 1.45 0.54 0.10 32 Ivotuk 68.48 -155.74 2000 - 2014 0.87 0.52 0.11 33 Konkovaya River R15a 69.41 158.45 1996 - 2015 0.70 0.35 0.07 34 Kuropatochya River R12a 70.92 156.63 1996 - 1996 1.00 0.37 0.09 35 Bykovsky Cape Plakor 71.79 129.42 2001 - 2014 0.86 0.33 0.06 36 Ayach 67.58 64.18 1996 - 2015 1.55 0.80 0.11 37 Vaskiny Dachi 1 70.28 68.89 2007 - 2015 1.00 0.70 0.09 38 Khomus2 69.98 153.58 2005 - 2005 1.00 0.54 0.09 39 Plosky Tolbachik 1 55.75 160.29 2003 - 2012 1.10 0.69 0.27 40 Betty Pingo WET 70.28 -148.92 1995 - 2015 1.48 0.41 0.05 41 Kougarok 65.46 -164.63 1999 - 2015 0.71 0.57 0.11 42 Norman Wells Grid 65.19 -126.47 1998 - 2008 1.18 0.46 0.12 43 Bolvansky 68.29 54.51 1999 - 2015 1.65 1.09 0.17 44 Kruglaya 64.63 176.97 2010 - 2015 1.00 0.45 0.07 45 Happy Valley 1ha 69.17 -148.83 1996 - 2015 0.80 0.40 0.08 46 Tuymada 62.01 129.66 2008 - 2015 1.12 2.02 0.08 47 Imnavait Creek MAT 68.61 -149.31 1995 - 2015 1.52 0.45 0.09 48 Lavrentiya 65.60 171.05 2000 - 2012 3.46 0.65 0.10 49 Bykovsky Cape Alas 71.78 129.40 2004 - 2015 0.92 0.31 0.07 50 Sagwon Hills MNT 69.44 -148.67 1995 - 2015 1.62 0.58 0.12 51 Mt Rodinka Station 68.70 161.55 2003 - 2015 1.00 0.77 0.14 52 Pearl Creek 64.90 -147.80 1969 - 2015 1.00 0.64 0.08 53 Atqasuk 70.45 -157.40 1995 - 2015 1.33 0.48 0.19 54 Andryushkino 69.17 154.43 2005 - 2015 1.55 0.38 0.11 55 Toolik 1km 68.62 -149.60 1995 - 2015 1.43 0.48 0.12 continued 56 Deadhorse 70.17 -148.47 1996 - 2015 0.95 0.65 0.08 57 Yubileynoe 3 DRY 65.95 75.87 2007 - 2007 1.00 0.23 0.04 58 Betty Pingo MNT 70.28 -148.89 1995 - 2014 1.55 0.38 0.08 59 Samoylov 72.37 126.48 2002 - 2015 7.50 0.48 0.06 60 Toolik MAT 68.62 -149.62 1995 - 2015 1.52 0.45 0.10 61 Yubileynoe 2 DRY 66.01 75.78 2007 - 2007 1.00 0.27 0.08 62 Most 56.91 118.28 2013 - 2014 1.00 0.49 0.08 63 Belenkiy 56.76 118.19 2013 - 2014 1.00 0.54 0.14 64 Fort Simpson Grid 61.89 -121.60 1999 - 2008 1.00~~

~~0.90 0.24 65 Lorino 65.54 -171.63 2010 - 2012 1.00 0.47 0.11 66 Bykovsky Cape 71.78 129.42 2015 - 2015 1.00 0.34 0.05 67 Plosky Tolbachik 2 55.76 160.32 2004 - 2012 1.00 0.56 0.04 68 Kuropatochya River R12b 70.92 156.63 1996 - 1996 1.00 0.27 0.08 69 Kashin Island 68.23 53.85 2010 - 2015 1.50 0.74 0.19 70 Rengleng River Grid 67.80 -134.13 1998 - 2008 1.18 0.78 0.13 71 Khomus1 69.98 153.58 2005 - 2005 1.00 0.51 0.09 72 Talnik 67.33 63.73 1999 - 2015 2.24 0.58 0.14 73 Konkovaya River R15b 69.41 158.45 1999 - 2015 0.71 0.45 0.06 74 Akhmelo Channel 68.81 160.99 1996 - 2015 0.95 0.52 0.08 75 Cape Chukochii R13b 70.08 159.92 1999 - 2015 1.00 0.43 0.06 76 Willowlake River Grid 62.70 -123.06 2001 - 2008 1.00 0.81 0.19 77 Alazeya River 69.32 154.97 1998 - 2015 0.78 0.51 0.09 78 West Dock 1km 70.37 -148.56 1995 - 2015 1.43 0.50 0.12 79 Yubileynoe 2 WET 66.01 75.78 2007 - 2007 1.00 0.28 0.04 80 Mountain Dionisiya 64.57 177.20 1996 - 2015 0.95 0.55 0.10 81 Yubileynoe 3 WET 65.95 75.87 2007 - 2007 1.00 0.35 0.03 82 UNISCALM 78.20 15.75 2011 - 2015 1.00 1.03 0.06 83 Yakutskoe Lake 69.85 159.49 1999 - 2015 0.82 0.46 0.06 continued 84 Neleger 62.32 129.50 2008 - 2015 1.00 1.24 0.10 85 Lake Glukhoe 68.80 160.96 1996 - 2015 0.95 0.85 0.16 86 Svyatoy Nos Cape 72.86 141.01 2001 - 2001 1.00 0.38 0.04 87 Parisento 70.12 75.58 1992 - 1995 0.75 0.91 0.31 88 Bonanza Creek 64.75 -148.00 1990 - 2015 0.96 0.71 0.11 89 Vaskiny Dachi 3 70.30 68.84 2007 - 2015 1.00 1.13 0.13 90 Imnavait 1km 68.50 -149.50 1995 - 2015 1.48 0.52 0.10 91 Nadym Grid 65.33 72.92 1997 - 2016 1.25 1.35 0.48 92 Urengoy GAS FIELD GP15 67.48 76.70 2008 - 2015 1.00 0.84 0.15 93 Reindeer Depot Grid 68.68 -134.15 2000 - 2008 1.11 1.10 0.10 94 Council 64.84 -163.71 1999 - 2015 0.71 0.57 0.33 95 Lake Akhmelo 68.83 161.03 1996 - 2015 0.95 0.97 0.16 96 Galbraith Lake 68.48 -149.50 1996 - 2015 0.90 0.53 0.09 97 Tolbachinsky Pass 55.89 160.54 2006 - 2015 0.90 0.47 0.02~~

~~GTN-P permafrost temperature (borehole) sites used in the CLASS-CTEM evaluation. Name Latitude Longitude GTN-P site number Permafrost zone Number of observations (°N) (°W) 1 Chevak 61.54 -165.60 746 Continuous 45 2 Smith Lake 4 64.87 -147.86 619 Discontinuous 964 3 Circle 65.82 -144.07 752 Discontinuous 105 4 Tobolsk aerologicheskaya 58.15 68.18 1670 None 1079 5 Belenkiy 56.76 118.19 1835 Continuous 34 6 Kerak 57.98 125.50 686 Discontinuous 99 7 Bayandai 53.10 105.53 1674 Isolated 1536 8 Rubtsovsk 51.50 81.22 1666 None 1403 9 Ust 65.45 52.17 1699 None 954 10 Karam 55.33 107.50 1676 Sporadic 990 11 Olkhon 53.23 107.44 1135 Sporadic 82 12 Nadym Pingo 65.30 72.89 178 Discontinuous 169 13 Franklin Bluffs dry b 69.67 -148.72 103 Continuous 691 14 Mould Bay 76.23 -119.30 1108 Continuous 24 15 Ishim 56.13 69.52 1605 None 1550 16 ILU2007 69.22 -51.10 535 Discontinuous 213 17 Taiga 56.07 87.62 1663 None 565 18 Vologda Molochnoe 59.28 39.87 1648 None 579 19 Svobodnyi 51.45 128.12 1623 None 1546 20 Russkaya Polyana 53.83 73.83 1610 None 1089 21 Onega 63.90 38.12 1658 No 1225 22 Berezovo 63.93 65.05 1639 Sporadic 1090 23 Komsomolsk 51.08 137.03 1693 Sporadic 1248 24 Nozhovka 57.08 54.75 1637 None 808 25 Deadhorse 2 new instrumentation 70.16 -148.47 88 Continuous 699 continued 26 Turukhansk 65.78 87.95 1652 Discontinuous 1369 27 Boguchany 58.42 97.40 1704 Isolated 1567 28 Saranpaul 64.28 60.88 1638 Discontinuous 1070 29 Smith Lake 3 64.87 -147.86 618 Discontinuous 1134 30 West Dock 1 surface 70.37 -148.55 118 Continuous 436 31 Sagwon MNT 69.43 -148.67 116 Continuous 1021 32 Kupino 54.37 77.28 1620 None 1253 33 Vyazemskaya 47.55 134.82 1689 None 1627 34 Rodino 52.50 80.20 1705 None 1396 35 West Dock 1 surface 70.37 -148.55 118 Continuous 1061 36 Irkutsk obs grass 52.28 104.30 1653 Sporadic 1589 37 Tyumen 57.15 65.50 1647 None 1642 38 Eniseisk 58.45 92.15 1657 None 1356 39 Smith Lake 2 64.87 -147.86 620 Discontinuous 991 40 Bikin 46.80 134.27 1694 None 1061 41 Ivotuk 3 68.48 -155.74 65 Continuous 39 42 Anderson 64.35 -149.20 845 Discontinuous~~

~~90 43 Makushino 55.25 67.30 1697 None 1350 44 Barnaul agriest 53.33 83.70 1673 None 1358 45 Tura 64.17 100.07 1641 Continuous 1092 46 Yartsevo 60.25 90.23 1642 Isolated 1500 47 Banks Island 73.22 -119.56 1107 Continuous 478 48 Salmon Lake 64.91 -165.05 1191 Discontinuous 185 49 Azarova 56.90 117.58 54 Continuous 52 50 Franklin Bluffs dry be 69.67 -148.72 104 Continuous 280 continued 51 Bomnak 54.72 128.93 1696 Discontinuous 1064 52 Irkutsk obs bare soil 52.28 104.30 1695 Sporadic 244 53 Tolstovka Amurskaya agrexpst 50.17 127.92 1687 None 1629 54 Kamen 53.80 81.33 1622 None 1391 55 Isil 54.90 71.27 1649 None 1627 56 Ivdel 60.68 60.43 1631 None 839 57 Eletskaya 67.17 64.17 1645 Sporadic 619 58 Erbogachen 61.27 108.02 1654 Continuous 1331 59 Anaktuvuk Pass 68.15 -151.72 833 Continuous 44 60 Tashtyp 52.80 99.88 1671 Continuous 1521 61 Solyanka 56.17 95.27 1682 None 1340 62 Ivotuk 4 68.48 -155.74 74 Continuous 316 63 Howe Island 1 b 70.32 -147.99 93 Continuous 568 64 Tatarsk 55.20 75.97 1661 None 842 65 Khanovey 67.29 63.65 1184 Continuous 29 66 Slavgorod 53.97 78.65 1665 None 1476 67 Troitsko 62.70 56.20 1698 None 983 68 Bogotol 56.23 89.58 1617 None 1176 69 Last Bridge 65.39 -164.66 1188 Continuous 60 70 Olkhon 53.23 107.44 1118 Sporadic 138 71 Sidorovsk 66.67 82.33 1678 Discontinuous 1380 72 Rhonda Basin 66.57 -164.48 1185 Continuous 49 73 Shimanovskaya 51.98 127.65 1624 None 857 74 Franklin Bluffs surface 69.66 -148.72 106 Continuous 1131 75 Ambler 67.08 -157.87 780 Continuous 57~~

~~continued 76 Olkhon 53.23 107.44 1136 Sporadic 82 77 Nadym ND3 4 65.31 72.86 176 Discontinuous 116 78 Belenkiy 56.76 118.19 224 Discontinuous 192 79 Leushi 59.62 65.78 1643 None 1061 80 Sterlitamak 53.62 55.98 1616 None 1135 81 Mould Bay 76.23 -119.30 1108 Continuous 1304 82 Franklin Bluffs dry ib 69.67 -148.72 105 Continuous 1103 83 Krasnoyarsk expfield 56.00 92.88 1655 None 1447 84 Barabinsk 55.37 78.40 1662 None 1046 85 Minusinsk expfield 53.70 91.70 1656 None 1119 86 Tulun agro 54.60 100.63 1686 Isolated 1591 87 Kolpashev 58.30 82.90 1659 None 1187 88 Syktyvkar 1 61.67 50.85 1672 None 969 89 Rhonda Upland 66.56 -164.46 1186 Continuous 79 90 Howe Island 1 ib 70.32 -147.99 94 Continuous 104 91 Ust 65.97 56.92 1614 None 1387 92 Zima railst 53.93 102.05 1677 Isolated 1109 93 Kosh 50.02 88.68 1707 Discontinuous 938 94 Aldan D 57.53 124.53 692 Discontinuous 80 95 Khomutovo 52.50 104.33 1685 None 1052 96 Kotkino 67.02 51.20 1611 Sporadic 496 97 Happy Valley 1 b 69.15 -148.85 95 Continuous 380 98 Nadym ND3 65.31 72.86 175 Discontinuous 246 99 Franklin Bluffs dry b 69.67 -148.72 103 Continuous 447 100 Biisk zonalnaya 52.68 84.95 1621 None 1349~~

~~continued 101 Ivotuk 4 68.48 -155.74 74 Continuous 51 102 Ivotuk 3 68.48 -155.74 65 Continuous 245 103 Happy Valley 1 ib 69.15 -148.85 110 Continuous 441 104 Ust 61.80 57.92 1675 None 603 105 Poliny Osipenko 52.42 136.50 1690 Sporadic 1535 106 Tunka 51.73 102.53 1668 Continuous 1588 107 Olkhon 53.23 107.44 1840 Sporadic 36 108 Ushelistiy 56.54 118.48 227 Continuous 94 109 Ivotuk 3 68.48 -155.74 65 Continuous 140 110 Skovorodino 54.00 123.97 1612 Sporadic 854 111 Tomsk 56.43 84.97 1660 None 697 112 Nadym PiCla 65.31 72.89 177 Discontinuous 49 113 Franklin Bluffs wet b 69.66 -148.72 107 Continuous 310 114 Mary s Igloo East 65.11 -164.70 1190 Discontinuous 160 115 Kazachinskoe expfield 57.75 93.18 1681 None 569 116 Kargopol 61.50 38.95 1627 None 1216 117 Deadhorse 1 surface 70.16 -148.47 87 Continuous 972 118 Shitkino 56.37 98.37 1669 None 1038 119 Tarko 64.92 77.82 1606 Discontinuous 1003 120 Howe Island 1 ib 70.32 -147.99 94 Continuous 512 121 Kalachinsk 55.03 74.58 1644 None 1163 122 Nadym ND2 65.31 72.89 174 Discontinuous 160 123 Banks Island 73.22 -119.56 1107 Continuous 385 124 Olkhon 53.22 107.45 1137 Sporadic 70 125 Belenkiy 56.76 118.19 1116 Continuous 231~~

~~continued 126 Vikulovo 56.82 70.62 1646 None 1295 127 Mould Bay 76.23 -119.30 1108 Continuous 1092 128 Konosha 61.00 40.17 1700 None 659 129 Srednii Vasyugan Vasyuganskoe 59.22 78.23 1684 None 1169 130 Khoseda 67.08 59.38 1701 Discontinuous 933 131 Ivotuk 4 68.48 -155.74 74 Continuous 508 132 Erofei Pavlovich 53.97 121.93 1651 Sporadic 814~~

~~Comparison of soil permeable depth datasets from Zobler86 (Zobler, 1986), Pel16 (Pelletier et al., 2016), and SoilGrids (Shangguan et al., 2017).~~

~~Snow cover as a function of snow depth for the CLASS-CTEM linear relation (Verseghy, 2017), the hyperbolic tangent form of Yang et al. (1997), and the exponential relationship proposed by Brown et al. (2003)~~

~~Comparison of mean annual latent and sensible heat flxues, spring (March, April, May; MAM) total surface albedo, winter (December, January, February; DJF) snow water equivalent (SWE), mean annual leaf area index (LAI) and total runoff. Positive percent change values indicate that the '20 ground layer' experiment has higher values of a quantity than the 'Base model' while negative values indicate the opposite. The green polygon indicates regions of permafrost simulated by that experiment. The green polygon on the percent change plots is the permafrost region from the '20 ground layers' experiment. Dots indicate grid cells that are statistically significant (independent two-sample t-test p level < 0.05). The left column shows the results from the original model version ('Base model') while the middle column shows the '20 ground layer' experiment.~~

~~Same as Figure **??** but for experiments '20 ground layers' and 'SoilGrids depth' Same as Figure **??** but for experiments 'SoilGrids depth' and 'SoilGrids+Moss' ALTs for experiments 'SoilGrids+Moss', 'Pel16+Moss', 'deVries thermal cond.', and 'Snow cover: Yang97' using the CRUNCEP meterological forcing dataset (left column), CRUJRA55 meteorological forcing dataset (middle column), and their difference (right column). Dots indicate grid cells that are statistically significant (independent two-sample t-test p level < 0.05). Same as Figure **??** but for experiments 'SoilGrids+Moss' and 'Tian16 thermal cond.' Mean monthly soil temperature over a permafrost region in eastern Siberia for the top 5 ground layers for the 'SoilGrids+Moss' and 'deVries thermal cond.' experiments. The first layer is moss in both simulations. Same as Figure **??** but for experiments 'SoilGrids+Moss' and 'Snow cover: Yang97' Taylor diagrams for winter (DJF) SWE compared to Blended-5 for the period spanning January 1981 to December 2010. Blended-5 is a multi-dataset SWE product developed by Mudryk et al. (2015) that combines five observation-based SWE datasets. Exp. 'Snow albedo decay' is outside the plot boundaries. This plot shows anomaly correlation coefficient as well as ratio of standard deviations and root mean square error normalized by the standard deviation from observations. Values shown correspond to the centroid over the values obtained for every grid cell northward of 45°N, with climatological SWE > 4 mm to avoid regions of ephemeral snow. Taylor plot of total spring albedo (AMJ) compared to MODIS MCD43C3 white-sky albedo (MODIS Adaptive Processing System, NASA, 2016) for the period spanning February 2000 to December 2013. The Taylor plot shows the anomaly correlation coefficient (polar coordinates), ratio of standard deviations (y axis) and root mean square error (RMSE) normalized by the standard deviation from observations (x axis). Values shown correspond to the centroid over the values obtained for every grid cell northward of 45°N. Same as Figure **??** but for experiments 'SoilGrids+Moss' and 'Fresh snow density'. Same as Figure **??** but for experiments 'SoilGrids+Moss' and 'Soil albedo decay'. Same as Figure **??** but for experiments 'SoilGrids+Moss' and 'Super-cooled water'.~~

~~Same as Figure ?? but for experiments 'SoilGrids+Moss' and 'Modif. hydrology'.~~ ~~Mean monthly 2 m air temperature at Lac de Gras sites from meteorological stations as part of the Slave Province Surficial Materials and Permafrost Study (SPSMPS) (Gruber et al., 2018) and reanalysis meterological datasets CRUNCEP (Viovy, 2016) and CRUJRA55 (Harris et al., 2014; Kobayashi et al.~~

~~Soil hydraulic conductivity (a,c) and matric potential (b,d) for a soil with sand and clay content by weight of 40% and 20%, respectively, with a maximum saturation level of 95%. The Zhao and Gray (1997) f_{ice} and the Niu and Yang (2006) $(1-F_{imp})$ parameters are shown in plots e and f, respectively.~~

~~Same as Figure ?? but the θ_{ice} term is removed from the numerator of equations B6 and B7.~~

Author contributions. JRM initiated the study, performed the model simulations and analysis and wrote the paper. DV led the development of the CLASS model, conducted initial research into the recommendations of MacDonald (2015) and was liaison to the Sushama group for the work of Arman Ganji. RS-A performed the statistical analysis and plotting for SWE and MODIS albedo. SG provided the Lac de Gras data and participated in discussions around model evaluation. All authors contributed to the final version of the paper.

Competing interests. None

Acknowledgements. We thank the Global Terrestrial Network for Permafrost for generously sharing their data and for making it easily accessible on-line. We thank Dr. Vivek Arora for processing the CRUJRA55 meteorological data, Mr. Ed Chan for processing the MODIS data, and Drs. Christian Seiler and Paul Bartlett for providing comments on a pre-submission version of our MS. Sampling at the Khanovey site is supported by the RuNoCORE CPRU-2017/10015 https://www.siu.no/eng/content/view/full/81242 ; the SAMCoT WP6 https://www.ntnu.edu/web/samcot/home and Lomonosov Moscow state university, Geology faculty, permafrost department.

References

Alexeev, V. A., Nicolsky, D. J., Romanovsky, V. E., and Lawrence, D. M.: An evaluation of deep soil configurations in the CLM3 for improved representation of permafrost, Geophys. Res. Lett., 34, 2007.

Arora, V., Seglenieks, F., Kouwen, N., and Soulis, E.: Scaling aspects of river flow routing, Hydrol. Process., 15, 461–477, https://doi.org/10.1002/hyp.161, 2001.

Bartlett, P. A., MacKay, M. D., and Verseghy, D. L.: Modified snow algorithms in the Canadian land surface scheme: Model runs and sensitivity analysis at three boreal forest stands, Atmosphere-Ocean, 44, 207–222, https://doi.org/10.3137/ao.440301, 2006.

Beer, C., Porada, P., Ekici, A., and Brakebusch, M.: Effects of short-term variability of meteorological variables on soil temperature in permafrost regions, The Cryosphere, 12, 741–757, 2018.

Bellisario, L. M., Boudreau, L. D., Verseghy, D. L., Rouse, W. R., and Blanken, P. D.: Comparing the performance of the Canadian land surface scheme (CLASS) for two subarctic terrain types, Atmosphere-Ocean, 38, 181–204, https://doi.org/10.1080/07055900.2000.9649645, 2000.

Biskaborn, B. K., Smith, S. L., Noetzli, J., Matthes, H., Vieira, G., Streletskiy, D. A., Schoeneich, P., Romanovsky, V. E., Lewkowicz, A. G., Abramov, A., Allard, M., Boike, J., Cable, W. L., Christiansen, H. H., Delaloye, R., Diekmann, B., Drozdov, D., Etzelmüller, B., Grosse, G., Guglielmin, M., Ingeman-Nielsen, T., Isaksen, K., Ishikawa, M., Johansson, M., Johannsson, H., Joo, A., Kaverin, D., Kholodov, A., Konstantinov, P., Kröger, T., Lambiel, C., Lanckman, J.-P., Luo, D., Malkova, G., Meiklejohn, I., Moskalenko, N., Oliva, M., Phillips, M., Ramos, M., Sannel, A. B. K., Sergeev, D., Seybold, C., Skryabin, P., Vasiliev, A., Wu, Q., Yoshikawa, K., Zheleznyak, M., and Lantuit, H.: Permafrost is warming at a global scale, Nat. Commun., 10, 264, https://doi.org/10.1038/s41467-018-08240-4, 2019.

Boeckli, L., Brenning, A., Gruber, S., and Noetzli, J.: Permafrost distribution in the European Alps: calculation and evaluation of an index map and summary statistics, The Cryosphere, 6, 807, 2012.

Brown, J., Ferrians, Jr, O. J., Heginbottom, J. A., and Melnikov, E. S.: Circum-Arctic map of permafrost and ground-ice conditions, US Geological Survey Reston, 1997.

Brown, R., Bartlett, P., MacKay, M., and Verseghy, D.: Evaluation of snow cover in CLASS for SnowMIP, Atmosphere-Ocean, 44, 223–238, https://doi.org/10.3137/ao.440302, 2006.

Brown, R. D., Brasnett, B., and Robinson, D.: Gridded North American monthly snow depth and snow water equivalent for GCM evaluation, Atmosphere-Ocean, 41, 1–14, https://doi.org/10.3137/ao.410101, 2003.

Chadburn, S., Burke, E., Essery, R., Boike, J., Langer, M., Heikenfeld, M., Cox, P., and Friedlingstein, P.: An improved representation of physical permafrost dynamics in the JULES land-surface model, Geoscientific Model Development, 8, 1493–1508, https://doi.org/10.5194/gmd-8-1493-2015, 2015a.

Chadburn, S. E., Burke, E. J., Essery, R. L. H., Boike, J., Langer, M., Heikenfeld, M., Cox, P. M., and Friedlingstein, P.: Impact of model developments on present and future simulations of permafrost in a global land-surface model, The Cryosphere Discuss., 9, 1965–2012, https://doi.org/10.5194/tcd-9-1965-2015, 2015b.

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. Chang., 7, 340, https://doi.org/10.1038/nclimate3262, 2017.

Clapp, R. B. and Hornberger, G. M.: Empirical equations for some soil hydraulic properties, Water Resour. Res., 14, 601–604, https://doi.org/10.1029/WR014i004p00601, 1978.

Cosby, B. J., Hornberger, G. M., Clapp, R. B., and Ginn, T. R.: A Statistical Exploration of the Relationships of Soil Moisture Characteristics to the Physical Properties of Soils, Water Resour. Res., 20, 682–690, https://doi.org/10.1029/WR020i006p00682, 1984.

Côté, J. and Konrad, J.-M.: A generalized thermal conductivity model for soils and construction materials, Can. Geotech. J., 42, 443–458, https://doi.org/10.1139/t04-106, 2005.

Dall'Amico, M., Endrizzi, S., Gruber, S., and Rigon, R.: A robust and energy-conserving model of freezing variably-saturated soil, The Cryosphere, 5, 469, 2011.

Dankers, R., Burke, E. J., and Price, J.: Simulation of permafrost and seasonal thaw depth in the JULES land surface scheme, The Cryosphere, 5, 773–790, https://doi.org/10.5194/tc-5-773-2011, 2011.

de Vries, D.: Thermal properties of soils, Physics of Plant Environment, 1963.

Dickinson, R. E.: Land Surface Processes and Climate—Surface Albedos and Energy Balance, in: Advances in Geophysics, edited by Saltzman, B., vol. 25, pp. 305–353, Elsevier, https://doi.org/10.1016/S0065-2687(08)60176-4, 1983.

Dickinson, R. E., Henderson-Sellers, A., and Kennedy, P.: Biosphere/Atmosphere Transfer Scheme (BATS) Version 1e as coupled to the NCAR Community Climate Model, Tech. rep., Climate and Global Dynamics Division, National Center for Atmospheric Research, Boulder, Colorado, 1993.

Ekici, A., Beer, C., Hagemann, S., and Hauck, C.: Simulating high-latitude permafrost regions by the JSBACH terrestrial ecosystem model, Geoscientific Model Development, 7, 631–647, 2014.

Essery, R., Martin, E., Douville, H., Fernández, A., and Brun, E.: A comparison of four snow models using observations from an alpine site, Clim. Dyn., 15, 583–593, https://doi.org/10.1007/s003820050302, 1999.

Farouki, O. T.: The thermal properties of soils in cold regions, Cold Reg. Sci. Technol., 5, 67–75, https://doi.org/10.1016/0165-232X(81)90041-0, 1981.

Ganji, A., Sushama, L., Verseghy, D., and Harvey, R.: On improving cold region hydrological processes in the Canadian Land Surface Scheme, Theor. Appl. Climatol., 127, 45–59, https://doi.org/10.1007/s00704-015-1618-4, 2015.

Giorgi, F. and Avissar, R.: Representation of heterogeneity effects in Earth system modeling: Experience from land surface modeling, Rev. Geophys., 35, 413–437, https://doi.org/10.1029/97RG01754, 1997.

Gornall, J. L., Jónsdóttir, I. S., Woodin, S. J., and Van der Wal, R.: Arctic mosses govern below-ground environment and ecosystem processes, Oecologia, 153, 931–941, https://doi.org/10.1007/s00442-007-0785-0, 2007.

Gray, D. M. and Male, D. H.: Handbook of Snow: Principles, Processes, Management and Use, Pergamon Press, 1981.

Gruber, S.: Derivation and analysis of a high-resolution estimate of global permafrost zonation, Cryosphere, 6, 221–233, 2012.

Gruber, S., Brown, N., Stewart-Jones, E., Karunaratne, K., Riddick, J., Peart, C., Subedi, R., and Kokelj, S.: Air and ground temperature, air humidity and site characterization data from the Canadian Shield tundra near Lac de Gras, Northwest Territories, Canada, v. 1.0 (2015-2017), http://www.cen.ulaval.ca/nordicanad/dpage.aspx?doi=45561XD-2C7AB3DCF3D24AD8, https://doi.org/10.5885/45561XD-2C7AB3DCF3D24AD8, accessed: 2018-11-13, 2018.

GTN-P: Global Terrestrial Network for Permafrost Database: Active Layer Thickness Data (CALM-Circumpolar Active Layer Monitoring), 2016.

Gubler, S., Fiddes, J., Keller, M., and Gruber, S.: Scale-dependent measurement and analysis of ground surface temperature variability in alpine terrain, The Cryosphere, 5, 431–443, https://doi.org/10.5194/tc-5-431-2011, 2011.

Harris, I., Jones, P. D., Osborn, T. J., and Lister, D. H.: Updated high-resolution grids of monthly climatic observations–the CRU TS3. 10 Dataset, Int. J. Climatol., 34, 623–642, 2014.

Hedstrom, N. R. and Pomeroy, J. W.: Measurements and modelling of snow interception in the boreal forest, Hydrol. Process., 12, 1611–1625, 1998.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., and Takahashi, K.: The JRA-55 Reanalysis: General Specifications and Basic Characteristics, Journal of the Meteorological Society of Japan, 93, 5–48, https://doi.org/10.2151/jmsj.2015-001, 2015.

Koren, V., Schaake, J., Mitchell, K., Duan, Q.-Y., Chen, F., and Baker, J. M.: A parameterization of snowpack and frozen ground intended for NCEP weather and climate models, J. Geophys. Res., 104, 19 569–19 585, https://doi.org/10.1029/1999JD900232, 1999.

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, J. Clim., 26, 1877–1900, https://doi.org/10.1175/JCLI-D-12-00228.1, 2013.

Kulik, V. Y.: Water infiltration into soil (in Russian), Gidrometeoizdat, Moscow, 19878, 1978.

Kushner, P. J., Mudryk, L. R., Merryfield, W., Ambadan, J. T., Berg, A., Bichet, A., Brown, R., Derksen, C. P., Déry, S. J., Dirkson, A., Flato, G., Fletcher, C. G., Fyfe, J. C., Gillett, N., Haas, C., Howell, S., Laliberté, F., McCusker, K., Sigmond, M., Sospedra-Alfonso, R., Tandon, N. F., Thackeray, C., Tremblay, B., and Zwiers, F. W.: Assessment of snow, sea ice, and related climate process in Canada's Earth-System model and climate prediction system, The Cryosphere, 12, 1137–1156, 2018.

Lafleur, P. M., Skarupa, M. R., and Verseghy, D. L.: Validation of the Canadian land surface scheme (class) for a subarctic open woodland, Atmosphere-Ocean, 38, 205–225, https://doi.org/10.1080/07055900.2000.9649646, 2000.

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, J. Geophys. Res., 113, F02 011, https://doi.org/10.1029/2007JF000883, 2008.

Lawrence, P. J. and Chase, T. N.: Representing a new MODIS consistent land surface in the Community Land Model (CLM 3.0), J. Geophys. Res., 112, G01 023, https://doi.org/10.1029/2006JG000168, 2007.

Lee, H., Swenson, S. C., Slater, A. G., and Lawrence, D. M.: Effects of excess ground ice on projections of permafrost in a warming climate, Environ. Res. Lett., 9, 124 006, https://doi.org/10.1088/1748-9326/9/12/124006, 2014.

Letts, M. G., Roulet, N. T., Comer, N. T., Skarupa, M. R., and Verseghy, D. L.: Parametrization of peatland hydraulic properties for the Canadian land surface scheme, Atmosphere-Ocean, 38, 141–160, https://doi.org/10.1080/07055900.2000.9649643, 2000.

Loranty, M. M., Abbott, B. W., Blok, D., Douglas, T. A., Epstein, H. E., Forbes, B. C., Jones, B. M., Kholodov, A. L., Kropp, H., Malhotra, A., Mamet, S. D., Myers-Smith, I. H., Natali, S. M., O'Donnell, J. A., Phoenix, G. K., Rocha, A. V., Sonnentag, O., Tape, K. D., and Walker, D. A.: Reviews and syntheses: Changing ecosystem influences on soil thermal regimes in northern high-latitude permafrost regions, Biogeosciences, 15, 5287–5313, https://doi.org/10.5194/bg-15-5287-2018, 2018.

MacDonald, M. K.: The Hydrometeorological Response to Chinook Winds in the South Saskatchewan River Basin, Ph.D. thesis, University of Edinburgh, 2015.

Mellor, M.: Engineering Properties of Snow, J. Glaciol., 19, 15–66, https://doi.org/10.3189/S002214300002921X, 1977.

Melton, J. R.: CLASS-CTEM code for "Improving permafrost physics in the coupled Canadian Land Surface Scheme (v. 3.6.2) and Canadian Terrestrial Ecosystem Model (v. 2.1) (CLASS-CTEM)", https://doi.org/10.5281/zenodo.3369396, https://doi.org/10.5281/zenodo.3369396, 2019.

Melton, J. R. and Arora, V. K.: Sub-grid scale representation of vegetation in global land surface schemes: implications for estimation of the terrestrial carbon sink, Biogeosciences, 11, 1021–1036, https://doi.org/10.5194/bgd-10-16003-2013, 2014.

Melton, J. R. and Arora, V. K.: Competition between plant functional types in the Canadian Terrestrial Ecosystem Model (CTEM) v. 2.0, Geoscientific Model Development, 9, 323–361, https://doi.org/10.5194/gmd-9-323-2016, 2016.

Melton, J. R., Sospedra-Alfonso, R., and McCusker, K. E.: Tiling soil textures for terrestrial ecosystem modelling via clustering analysis: a case study with CLASS-CTEM (version 2.1), Geoscientific Model Development, 10, 2761–2783, https://doi.org/10.5194/gmd-10-2761-2017, 2017.

MODIS Adaptive Processing System, NASA: MODIS/Terra+Aqua Albedo 16-Day L3 Global 0.05Deg CMG V005, title of the publication associated with this dataset: MODIS/Terra+Aqua Albedo 16-Day L3 Global 0.05Deg CMG V005, 2016.

Morse, P. D., Burn, C. R., and Kokelj, S. V.: Influence of snow on near-surface ground temperatures in upland and alluvial environments of the outer Mackenzie Delta, Northwest Territories, Can. J. Earth Sci., 49, 895–913, 2012.

Mudryk, L. R., Derksen, C., Kushner, P. J., and Brown, R.: Characterization of Northern Hemisphere snow water equivalent datasets, 1981-2010, Journal of Climate, 28, 8037–8051, 2015.

Myers-Smith, I. H., Forbes, B. C., Wilmking, M., Hallinger, M., Lantz, T., Blok, D., Tape, K. D., Macias-Fauria, M., Sass-Klaassen, U., Lévesque, E., Boudreau, S., Ropars, P., Hermanutz, L., Trant, A., Collier, L. S., Weijers, S., Rozema, J., Rayback, S. A., Schmidt, N. M., Schaepman-Strub, G., Wipf, S., Rixen, C., Ménard, C. B., Venn, S., Goetz, S., Andreu-Hayles, L., Elmendorf, S., Ravolainen, V., Welker, J., Grogan, P., Epstein, H. E., and Hik, D. S.: Shrub expansion in tundra ecosystems: dynamics, impacts and research priorities, Environ. Res. Lett., 6, 045 509, 2011.

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, https://doi.org/10.1029/2007GL029525, 2007.

Niu, G.-Y. and Yang, Z.-L.: Effects of Frozen Soil on Snowmelt Runoff and Soil Water Storage at a Continental Scale, J. Hydrometeorol., 7, 937–952, https://doi.org/10.1175/JHM538.1, 2006.

O'Neill, H. B., Wolfe, S. A., and Duchesne, C.: New ground ice maps for Canada using a paleogeographic modelling approach, The Cryosphere Discuss., pp. 1–37, https://doi.org/10.5194/tc-2018-200, 2018.

Paquin, J.-P. and Sushama, L.: On the Arctic near-surface permafrost and climate sensitivities to soil and snow model formulations in climate models, Clim. Dyn., 44, 203–228, https://doi.org/10.1007/s00382-014-2185-6, 2014.

Pelletier, J. D., Broxton, P. D., Hazenberg, P., Zeng, X., Troch, P. A., Niu, G., Williams, Z. C., Brunke, M. A., and Gochis, D.: Global 1-km Gridded Thickness of Soil, Regolith, and Sedimentary Deposit Layers, https://doi.org/10.3334/ORNLDAAC/1304, 2016.

Peng, Y., Arora, V. K., Kurz, W. A., Hember, R. A., Hawkins, B. J., Fyfe, J. C., and Werner, A. T.: Climate and atmospheric drivers of historical terrestrial carbon uptake in the province of British Columbia, Canada, Biogeosciences, 11, 635–649, https://doi.org/10.5194/bg-11-635-2014, 2014.

Pomeroy, J. W. and Gray, D. M.: Snowcover accumulation, relocation and management, Bulletin of the International Society of Soil Science no, 88, 1995.

Porada, P., Ekici, A., and Beer, C.: Effects of bryophyte and lichen cover on permafrost soil temperature at large scale, The Cryosphere, 10, 2291–2315, https://doi.org/10.5194/tc-10-2291-2016, 2016.

Riseborough, D., Shiklomanov, N., Etzelmüller, B., Gruber, S., and Marchenko, S.: Recent advances in permafrost modelling, Permafrost Periglacial Processes, 19, 137–156, https://doi.org/10.1002/ppp.615, 2008.

Roebber, P. J., Bruening, S. L., Schultz, D. M., and Cortinas, J. V.: Improving Snowfall Forecasting by Diagnosing Snow Density, Weather Forecast., 18, 264–287, https://doi.org/10.1175/1520-0434(2003)018<0264:ISFBDS>2.0.CO;2, 2003.

Romanovsky, V. E. and Osterkamp, T. E.: Effects of unfrozen water on heat and mass transport processes in the active layer and permafrost, Permafrost Periglacial Processes, 11, 219–239, 2000.

Scott, D. W.: Multivariate Density Estimation: Theory, Practice, and Visualization, John Wiley & Sons, 1992.

Shangguan, W., Dai, Y., Duan, Q., Liu, B., and Yuan, H.: A global soil data set for earth system modeling, Journal of Advances in Modeling Earth Systems, 6, 249–263, https://doi.org/10.1002/2013MS000293, 2014.

Shangguan, W., Hengl, T., Mendes de Jesus, J., Yuan, H., and Dai, Y.: Mapping the global depth to bedrock for land surface modeling, J. Adv. Model. Earth Syst., 9, 65–88, https://doi.org/10.1002/2016MS000686, 2017.

Shiklomanov, N. I., Streletskiy, D. A., Little, J. D., and Nelson, F. E.: Isotropic thaw subsidence in undisturbed permafrost landscapes, Geophys. Res. Lett., 40, 6356–6361, https://doi.org/10.1002/2013GL058295, 2013.

Smerdon, J. E. and Stieglitz, M.: Simulating heat transport of harmonic temperature signals in the Earth's shallow subsurface: Lower-boundary sensitivities, Geophys. Res. Lett., 33, 2006.

Smith, M. W.: Microclimatic Influences on Ground Temperatures and Permafrost Distribution, Mackenzie Delta, Northwest Territories, Can. J. Earth Sci., 12, 1421–1438, https://doi.org/10.1139/e75-129, 1975.

Soulis, E. D., Snelgrove, K. R., Kouwen, N., Seglenieks, F., and Verseghy, D. L.: Towards closing the vertical water balance in Canadian atmospheric models: Coupling of the land surface scheme class with the distributed hydrological model watflood, Atmosphere-Ocean, 38, 251–269, https://doi.org/10.1080/07055900.2000.9649648, 2000.

Sturm, M., Holmgren, J., König, M., and Morris, K.: The thermal conductivity of seasonal snow, J. Glaciol., 43, 26–41, https://doi.org/10.1017/S0022143000002781, 1997.

Tian, Z., Lu, Y., Horton, R., and Ren, T.: A simplified de Vries-based model to estimate thermal conductivity of unfrozen and frozen soil, Eur. J. Soil Sci., 67, 564–572, https://doi.org/10.1111/ejss.12366, 2016.

Tilley, J. S., Chapman, W. L., and Wu, W.: Sensitivity tests of the Canadian Land Surface Scheme (CLASS) for Arctic tundra, Ann. Glaciol., 25, 46–50, https://doi.org/10.1017/s0260305500013781, 1997.

Turetsky, M. R., Bond-Lamberty, B., Euskirchen, E., Talbot, J., Frolking, S., McGuire, A. D., and Tuittila, E.-S.: The resilience and functional role of moss in boreal and arctic ecosystems, New Phytol., 196, 49–67, https://doi.org/10.1111/j.1469-8137.2012.04254.x, 2012.

UNESCO Press: Discharge of selected rivers of the world, vol. 2, part 2. Mean Monthly and Extreme Discharges, 1965-1984,, Tech. rep., UNESCO, Paris, 1993.

Van Der Wal, R. and Brooker, R. W.: Mosses mediate grazer impacts on grass abundance in arctic ecosystems, Funct. Ecol., 18, 77–86, https://doi.org/10.1111/j.1365-2435.2004.00820.x, 2004.

Vaughan, D. G., Comiso, J. C., Allison, I., Carrasco, J., Kaser, G., Kwok, R., Mote, P., Murray, T., Paul, F., Ren, J., Rignot, E., Solomina, O., Steffen, K., and Zhang, T.: Observations: Cryosphere, in: Climate Change 2013 - The Physical Science Basis, edited by Stocker, T. F., Qin, D., K. Plattner, G., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, pp. 317–382, Cambridge University Press, Cambridge, https://doi.org/10.1017/CBO9781107415324.012, 2013.

Verseghy, D.: CLASS – The Canadian land surface scheme (v. 3.6.2), Climate Research Division, Science and Technology Branch, Environment Canada, 2017.

Viovy, N.: CRU-NCEP Version 8, title of the publication associated with this dataset: CRU-NCEP version 8, 2016.

Watanabe, K. and Mizoguchi, M.: Amount of unfrozen water in frozen porous media saturated with solution, Cold Reg. Sci. Technol., 34, 103–110, https://doi.org/10.1016/S0165-232X(01)00063-5, 2002.

Wen, Z., Ma, W., Feng, W., Deng, Y., Wang, D., Fan, Z., and Zhou, C.: Experimental study on unfrozen water content and soil matric potential of Qinghai-Tibetan silty clay, Environ. Earth Sci., 66, 1467–1476, https://doi.org/10.1007/s12665-011-1386-0, 2012.

Wiscombe, W. J. and Warren, S. G.: A Model for the Spectral Albedo of Snow. I: Pure Snow, J. Atmos. Sci., 37, 2712–2733, https://doi.org/10.1175/1520-0469(1980)037<2712:AMFTSA>2.0.CO;2, 1980.

Wu, Y., Verseghy, D. L., and Melton, J. R.: Integrating peatlands into the coupled Canadian Land Surface Scheme (CLASS) v3.6 and the Canadian Terrestrial Ecosystem Model (CTEM) v2.0, Geoscientific Model Development, 9, 2639–2663, https://doi.org/10.5194/gmd-9-2639-2016, 2016.

Yang, Z.-L., Dickinson, R. E., Robock, A., and Vinnikov, K. Y.: Validation of the snow submodel of the Biosphere–Atmosphere Transfer Scheme with Russian snow cover and meteorological observational data, J. Clim., 10, 353–373, 1997.

Zhang, T., Barry, R. G., Knowles, K., Heginbottom, J. A., and Brown, J.: Statistics and characteristics of permafrost and ground-ice distribution in the Northern Hemisphere, Polar Geogr., 23, 132–154, https://doi.org/10.1080/10889379909377670, 1999.

Zhang, T., Heginbottom, J. A., Barry, R. G., and Brown, J.: Further statistics on the distribution of permafrost and ground ice in the Northern Hemisphere, Polar Geogr., 24, 126–131, https://doi.org/10.1080/10889370009377692, 2000.

Zhang, Y., Carey, S. K., and Quinton, W. L.: Evaluation of the algorithms and parameterizations for ground thawing and freezing simulation in permafrost regions, J. Geophys. Res., 113, D17 116, https://doi.org/10.1029/2007JD009343, 2008.

Zhao, L. and Gray, D. M.: A parametric expression for estimating infiltration into frozen soils, Hydrological Processes, 11, 1761–1775, https://doi.org/10.1002/(SICI)1099-1085(19971030)11:13<1761::AID-HYP604>3.0.CO;2-O, https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291099-1085%2819971030%2911%3A13%3C1761%3A%3AAID-HYP604%3E3.0.CO%3B2-O, 1997.

Zobler, L.: A world soil file for global climate modelling, title of the publication associated with this dataset: NASA Technical Memorandum 87802, 1986.