# Author Point-By-Point Response

## Reviewer #1

The authors would like to thank Referee 1 for their helpful comments. We have substantially restructured the Results (Sections 4 and 5) and the Discussion/Conclusion (Section 6) to shorten its length and tighten the discussion. Section 5 has been modified to an investigation of variability and trends in the ensemble mean soil temperature product compared to a subset of stations with longer temperature records.

**Major Comments**

1) Reformulate and shorten the manuscript (maybe as a brief communication) with a very specific focus on soil temperature validation

We have substantially shortened the results sections (Sections 4 and 5) and have changed the section on the ensemble mean soil temperature climatology (Section 5.2) to focus on the validation of variability in soil temperature extremes (winter minimum and summer maximum soil temperatures), and soil temperature trends over a subset of grid cells with longer temperature records.

In response to the structural changes in the text, we've also made changes to a number of the figures, and their ordering has changed in some cases:

- Figure 1, Panel A has been updated substantially to account for the new grid cells added, and changes in Section 5.2 (mainly the delineation of the subset of grid cells used for soil temperature trend analysis). Panel B has been updated to show the relative importance of spatial variability and depth variation in soil temperature.
- Figure 2 now includes a new metric – a Taylor (2000) type skill score metric providing an objective estimate of the overall performance of each product
- Former Figure 3 has been renumbered as Figure 6
- Figure 4 has been revamped to only include the relationships between the station soil temperature and the product soil temperatures
- Former Figure 5 is now Figure 3
- Former Figure 6 is now Figure 5
- Former Figure 7 is now Figure 8
- Figure 7 is now a spatial map of the Ensemble Mean soil temperature biases)

- In section 5.2, former Figure 8 has been removed, and instead 3 new figures have been added (Figures 9 – 11).

2) In Sec. 4 & 5, the authors present the evaluation results together with a large part of the discussion, and additional discussions are given in Sec. 6. This makes the manuscript very unclear and difficult to follow.

The authors have separated out discussion material from the Results (Section 4 and 5). Relevant discussion material has now been incorporated into a substantially revised Section 6 that now includes a brief summary of how the results of this study relate the findings of previous work.

3) The discussion in Sec 6 is very general and superficial, and is not tightly connected to previous sections. For instance, the gap of site-scale observation and model grid (about 10–100 km), or so-called scale effects, is widely reported. P23, L392–398, this part is very confusing. Does the misclassification of permafrost affect the results? Please make sure only to present the most relevant parts here to avoid diluting your real contributions.

Section 6 discussion was reformulated to incorporate relevant discussion material originally included in Sections 4 and 5. The section begins with a summary of the key results and how these relate to previous studies' findings. The discussion now focuses around four main themes:

- Uncertainties associated with land model parameterizations and structural differences in the land models
- Impacts of discontinuities in the reanalysis timeseries
- Uncertainties associated with scale effects
- Uncertainties arising from sampling variability

The discussion is more focused as the authors clearly emphasize how these uncertainties could potentially influence the results.

4) The authors presented and discussed the reanalysis soil temperature deviation. I am wondering why this is important here and how this could be used for validation purposes? The strong variation of soil temperature in the cold season could be expected due to the presence of a snow layer, see Figure 6 from Burke et al., (2020).

The authors thank the reviewer for bringing this figure to our attention. We use the normalized standard deviation as a measure of the temporal variance of soil temperatures across the grid cells. Doing allows us to assess the range of simulated soil temperatures at each grid cell for a particular product, to see if it can capture a similar seasonal cycle of temperatures to that of the observations.

When we describe soil temperature variability in the cold season we are referring to two main features – first that the individual products themselves show a larger variance in soil temperatures than they do during the warm season. Second, we are also describing the spread in soil temperature variance between products. It is likely that differences in snow cover properties may help account for the latter, and we have added a reference to Burke et al. (2020) in Section 6.1 beginning at line 432:

> Burke et al. (2020) note that differences in snow cover properties were important in explaining soil temperature biases of several Coupled Model Intercomparison Project 6 (CMIP6) models, and it is likely that differences in snow cover properties between the land models of the reanalysis products could account for some of the observed spread - particularly in the cases of ERA-Interim, ERA5 and ERA5-Land.

5) The climatology based on the ensemble results is somehow unfocused. The purpose of this study is "validation of pan-Arctic (and Boreal) soil temperatures from eight reanalyses and land data assimilation system (LDAS) products." (see P2, L53–54), rather than analyzing the climatology. To be more focused, authors could compare and evaluate the trend of ensemble results with site-scale observations.

We agree that a focused evaluation of the trends against a subset of the stations with longer timeseries is of value and have restructured Section 5.2 to focus instead on validating decadal soil temperature trends against a subset of station estimates with longer timeseries Figure 11 and Supplemental Figure 6). We also validate the annual soil temperature minimum and maximum soil temperatures (Figure 10 and Supplemental Figure 5) in this section.

6) P8, L180: Then why not directly use the IPA map? You could also find the global permafrost zonation index map from Gruber et al., (2012).

The authors thank the reviewer for this suggestion. We have incorporated the Obu et al. (2019) permafrost zonation index map into our analysis, and the permafrost zone now refers to regions with at least 50% permafrost cover, while the zone with little to no permafrost refers to regions with less than 50% permafrost cover. The contour lines on Figure 1, Panel A encircle regions with at least 50% permafrost cover – as estimated by Obu et al. (2019). The overall conclusions of the study were not impacted by this change.

**Minor Comments**

1) P2, L24: Permafrost carbon and climate warming loop are complex, and thus ...could act as a "possibly/potentially" positive...

We have revised the sentence to read as: “Continued warming, and thawing of permafrost soils, and related decomposition of carbon could act as a **potential** positive feedback on warming, by releasing more methane ($CH_4$) and carbon dioxide ($CO_2$) into the atmosphere.”

2) P2, L31: Qinghai-Tibetan **P**lateau

Corrected.

3) P2, L45–49: Ensemble simulation has also been used for permafrost simulation, for instance, Cao et al., (2019), although these studies do not directly use the soil temperature

We have added a reference to Cao et al. (2019) in our introduction - in the following sentence:

“Ensemble mean datasets based on combinations of in situ, model, satellite and reanalysis data have been used to reduce biases in estimates of snow water equivalent (Mudryk et al.,2015), soil moisture (Dorigo et al., 2017; Gruber et al., 2019), precipitation (Beck et al., 2017, 2019), **as well as for local scale permafrost simulations (Cao et al., 2019)**.”

4) P4, L122: The variation of soil temperature is complex and typically depends on surface condition (i.e., snow layer, vegetation), soil properties (i.e., soil organic content), and soil depth. It could vary very large at the hourly and daily scales.

The authors thank the reviewer for this helpful comment. As we are using soil temperatures averaged between 0cm and 30cm in the near surface, and between 30cm and 300cm at depth, we had presumed that the soil temperatures should show reduced variation on daily and hourly timescales. We have revised this sentence to read “Many of the in situ (station) sites reported measurements at hourly or daily frequency, however we chose to perform the analysis at monthly time scales, in order to focus on processes controlling the seasonal cycle of soil temperatures.”

5) P6, L135: How much the difference could be? Could you please write it down?

The authors presume that the reviewer is asking by how much the soil temperatures may vary between stations within a grid cell. Panel B of Figure 1 gives an estimate of the variability of soil temperatures within a grid cell – The median spatial standard deviation is ~2°C, however soil temperatures may vary by as much as 13°C in the case of a couple of high latitude grid cells. The variability is of a similar magnitude to previous studies exploring sub-grid scale variability in cryospheric soil temperatures (e.g. Gubler et al., 2011; Morse et al., 2012; Gisnås 2014; Cao et al., 2019).

6) P6, L141: ...2 to 12..

We chose to write 2 as “two”, since style conventions in The Cryosphere specify that all single digit numbers (unless they are followed by units) should be written as a word.

7) P6, L142: The so-called "scale effects" has been widely reported, see Gubler el al., (2011) for the Alps and Cao et al., (2019) for high latitudes. Please cite relevant references.

We thank the reviewer for pointing out several relevant references on scale effects. In our revisions, we have linked our estimates of scale effects with those in the literature, and show that our results qualitatively agree with those exploring scale effects in seasonally frozen and permafrost soils (e.g. Gubler et al., 2011; Morse et al., 2012; Gisnås 2014; Cao et al., 2019).

We have added the following sentences in Section 6.3 to highlight the impacts of scale effects:

> The sub-grid scale variability in soil temperatures calculated in Figure 1, Panel B is of a similar magnitude to those calculated by previous studies exploring sub-grid scale variability in cryospheric soil temperatures (Gubler et al., 2011; Morse et al., 2012; Gisnås et al., 2014), though is smaller than those reported by Cao et al. (2019). If the strict requirements surrounding consistency in the number of stations and depths are relaxed, allowing for stations in permafrost regions to be included, spatial variability in soil temperatures is larger than 10◦ C at times in a couple of high latitude grid cells (not shown) - similar to the findings of Cao et al. (2019).

8) P8, L172: you have two "also" here

Corrected.

9) P9, L192: "more" → greater/larger

Revised.

10) P14, L250: Qinghai-Tibetan **P**lateau

Corrected.

11) P16, L256: Zero curtain period is heavily dependent on the soil moisture rather than the active layer thickness

The authors thank the reviewer for this helpful comment. We have made substantial revisions to the Results, and this paragraph is no longer present in the manuscript. We've moved our discussion of freeze-thaw parameterizations to Section 6, for example.

12) P22, L357: Remove the redundant ')'

Fixed.

13) Table 1: Could you please also add the soil discretization information here, such as depth for each layer and the total soil column depth? Please double-check the spatial resolution of all the reanalyses, ERA5 should be 0.25°, ERA-Interim is 0.75°, and MERRA-

> 2 is 0.5°×0.625°. Depending on the datasets you used, JRA-55 is 1.25° for the reanalysis level and 0.56° for the model level

We have corrected the spatial resolutions in Table 1. The information in Table 1 has also been split into 2 tables in order to include information about the depths of soil layers in each product.

> 14) Figure 4: Do you really need so many sub-plots? The inter-comparisons among different reanalyses are shown here but not discussed in the main text. Did I miss something important? Please also add the 1:1 line, so that readers could clearly see the cold/warm bias

The most important comparisons to be made here are the performances of the individual products against the station – the outer margins of Figure 4 in the paper – which is the focus of the text. We also used the histograms of the warm/cold season to look at the variability of soil temperatures in the warm and cold season. We have revised Figure 4 such that it now only shows the relationship between station soil temperatures and each of the products, for both depths.

> 15) Figure S3: Could you please improve the resolution of Figure S3?

We have recreated Figure S3 (Figure S5 in the revised manuscript) and it should be at a higher resolution now.

## Reviewer #2

The authors would like to thank Referee 2 for their helpful comments. As a part of our revisions, we have gathered substantially more data for North America, and have recalculated all metrics. In our updated database, we now have 135 validation grid cells over North America; 30 of which are located over the permafrost region. By utilizing soil temperature data from a variety of hydrometeorological and agricultural monitoring networks, our dataset now provides the most comprehensive analysis to date of soil temperatures across northern and southern Canada and the Great Lakes basin.

### Major Comments

1. The authors list potential future applications of the ensemble mean product, but I would wish to see a bit more discussion on its current usability, given that the recorded biases remain quite high and display some regional patterns. The underlying reasons for these are addressed in the manuscript but not how the biases would affect, e.g., permafrost simulations where a bias or RMSE of above 2° C can have notable implications.

We thank the reviewer for this comment. Several of the products have an RMSE of ≤4$^{o}$C – particularly over permafrost regions (as shown in Figure S1). In most products, this is expressed as a cold bias, which would suggest that reanalysis products may overestimate permafrost extent and underestimate active layer thickness. The ensemble mean biases and RMSE are generally better than (or similar to) the best performing product, especially when all seasons and depths are considered. In addition, the ensemble mean soil temperatures show a more realistic pattern of soil temperature variability in the permafrost zone compared to the individual products themselves.

The ensemble mean product provides gridded, monthly-averaged soil temperature estimates of near surface, and deeper soil temperatures at a 1$^{o}$ resolution. Therefore, it is most suitable to regional or hemispheric-scale analyses of soil temperature climatologies, or their seasonal cycle, or to explore recent trends in soil temperatures (since 1980). The product could also be used to provide boundary conditions for hydrological models. In fact, a higher resolution version of this product (see our response to Question 12 in Minor Comments) is being used for such a purpose and will be described in a follow-up study.

The authors acknowledge that the ensemble mean soil temperature product would most likely yield an overestimation of permafrost extent, given that it is biased cold by 3-5$^{o}$C, on average, at high latitudes. That being said, over permafrost regions, the RMSE of the ensemble mean product outperforms the RMSE the best performing product by ~2$^{o}$C, on average, and hence it may still provide some added value for estimation of high latitude soil temperatures relative to the individual products.

We have added the following paragraph to the manuscript in Section 6.5 (Applications for the Ensemble Mean Product and Suggestions for Future Work):

> The ensemble mean data product provides gridded, monthly-averaged soil temperature estimates of near surface, and deeper soil temperatures at a 1° resolution. Therefore, it is most suitable to regional or hemispheric-scale analyses of soil temperature climatologies, or their seasonal cycle, or to explore recent trends in soil temperatures. The product could also be used to provide boundary conditions for models that require soil temperature inputs, such as hydrological models, and for the validation of model soil temperatures. While the ensemble mean product still exhibits substantial cold biases over permafrost regions, and therefore is likely unsuitable for permafrost modeling, the RMSE of the ensemble mean product outperforms the RMSE of the best performing product by ~2°C, on average, and hence it may still provide some added value for estimation of high latitude soil temperatures relative to the individual products.

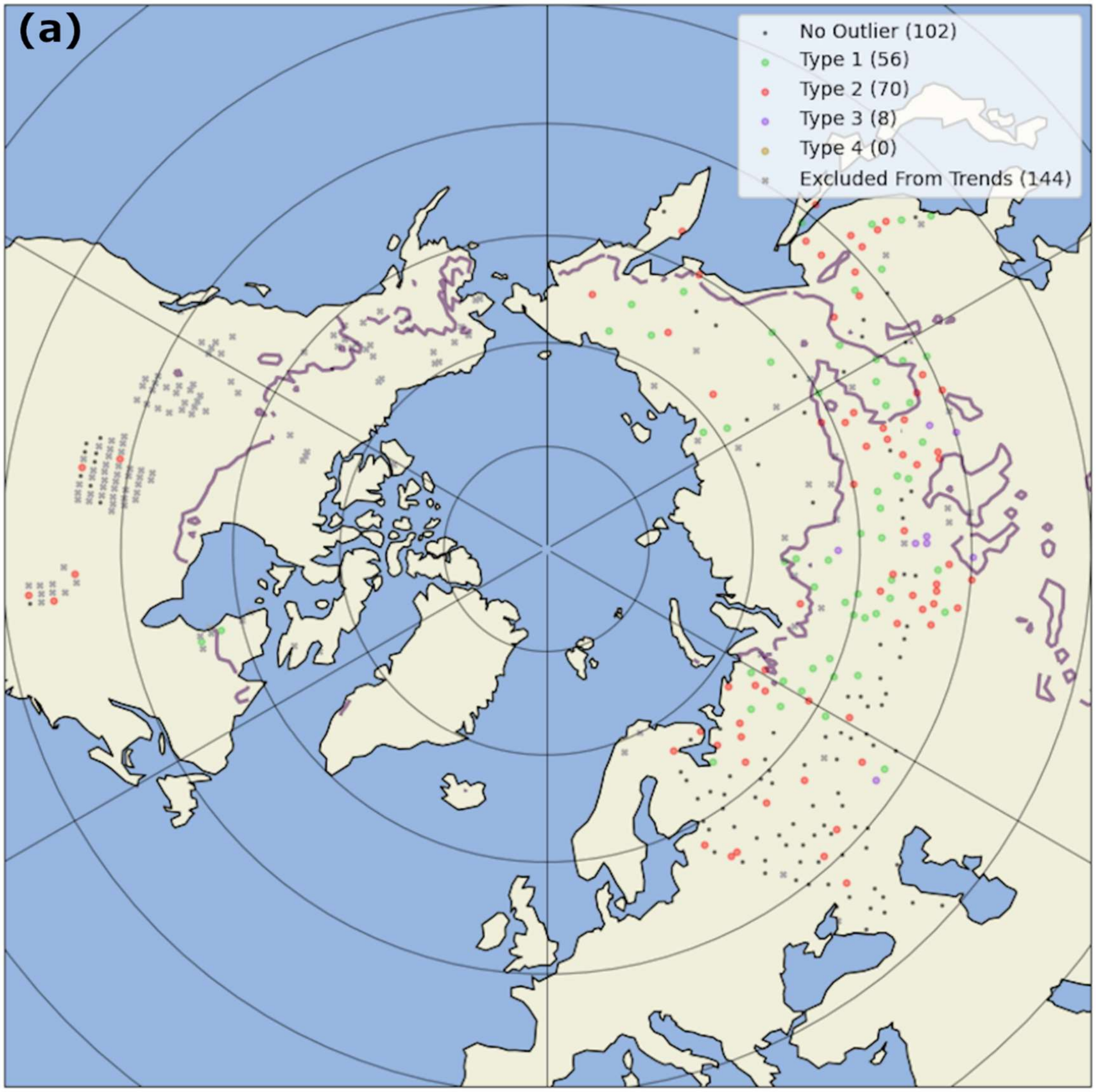
2. At places the text is hard to follow (especially Section 4.3, see detailed comments below) owing to the multiple simultaneous comparisons: near surface vs. at depth soil temperatures, cold season vs. warm season, permafrost vs. no to little permafrost, North America vs. Eurasia, and DJF vs. JJA. I suggest the authors to make sure all sections are clearly defined.

Section 4 has been substantially restructured. We begin in Section 4.1, where we discuss the extratropical northern hemisphere mean results, before moving to discuss differences between the warm and cold seasons in Section 4.2. Regional differences are discussed in Section 4.3, and we have taken care to separate our discussions of permafrost presence from our comparisons between the continents, especially now that we have a greater number of validation grid cells over North America (making the comparisons between Eurasia and North America more meaningful).

3. L104: The authors suggest that their study is *"To the authors' knowledge, this one of the first studies to compile a comprehensive set of in situ soil temperature measurements across the Eurasian and North American Arctic, from multiple diverse sparse networks".* While it may be true that this is true for the "one of the first" part, it should be noted that the compilation is not totally novel, given that similar in situ temperature datasets have been compiled not only by Cao et al. (2020, in the references) but also, e.g., by Karjalainen et al. (2019) and Ran et al. (2022) who used mostly the same data sources, albeit computing temperatures averages for a much larger depth (several meters deep in permafrost but also in non-permafrost soils). Moreover, Lembrechts et al. (2020) have published a global soil temperature compilation of soil and near-surface temperatures. I suggest the authors to consider if their statement needs some elaboration, e.g., does the compiled dataset differ from previous datasets in some ways.

> The authors recognize the notably different sampling size for North America but retain from explaining why no more data were collected, apart from mentioning the overall data scarcity in northern Canada, to correct the imbalance between North America and Eurasia. Based on the previous data compilations (see above), there should be suitable measurement time series available from North America.

The authors thank the reviewer for making us aware of these studies. As a result, the biggest change in the revised manuscript is the inclusion of a large amount of new soil temperature data from North America. Figure R1 compares the previous and updated distribution of validation grid cells, which now contains 135 validation grid cells over North America near the surface; 30 of which are located over the permafrost region. This means that our sample of sites for North America is now more comparable to the 247 grid cells in Eurasia (45 of which span the permafrost region). The much improved coverage across North America is evident in Figure 1, Panel A:



The new data are drawn from multiple sources, and we reiterate our claim from the original manuscript that this collection of pan-Northern Hemisphere soil temperature data constitutes a novel and important contribution to the permafrost research community. Over the permafrost region, we've assembled data from the Yukon (Yukon Geological Survey, 2021) and the NWT (Cameron et al., 2019; Ensom et al., 2019; Gruber et al., 2019; GTN-P, 2018; Spence and Hedstrom, 2018a; Spence and Hedstrom, 2018b; Street, 2018).

In addition, we have incorporated data from several soil monitoring and hydrometeorological networks across Southern Canada and the Great Lakes basin of the United States, that, to our knowledge, are not included in any of the above papers. These include 85 stations from the Manitoba Mesonet network (RoTimi Ojo and Manaigre, 2021), 83 stations in Michigan and western Wisconsin (MAWN, 2022), 31 stations from the Alberta Climate Information Service network (Alberta Agriculture, Forestry and Rural Economic Development, 2022), and 150 stations from North Dakota (NDAWN, 2022). We are also including data from a peatland ecosystem in Metro Vancouver (Lee et al., 2017; Lee et al., 2021), as well as data from 11 stations in central and Northern BC (Déry, 2017; Hernández-Henríquez et al., 2018; Morris et al. 2021), and 2 stations in southern Quebec (Arsenault, 2018; Fortier, 2020).

We have also been in contact with the data providers from the Real Time In-Situ Monitoring Network (RISMA), however the data was not available to include at the time this response was submitted. We hope to include the RISMA dataset (which includes 13 stations in southern Manitoba, 6 stations in southeastern Ontario, and 4 stations in southern Saskatchewan) in follow-up studies.

While the Ran et al. (2022) study included borehole measurements from southern Canada, the data did not include information about the seasonal cycle of soil temperatures. Thus, our work presents the most comprehensive analysis to date of soil temperatures across northern and southern Canada and the Great Lakes basin.

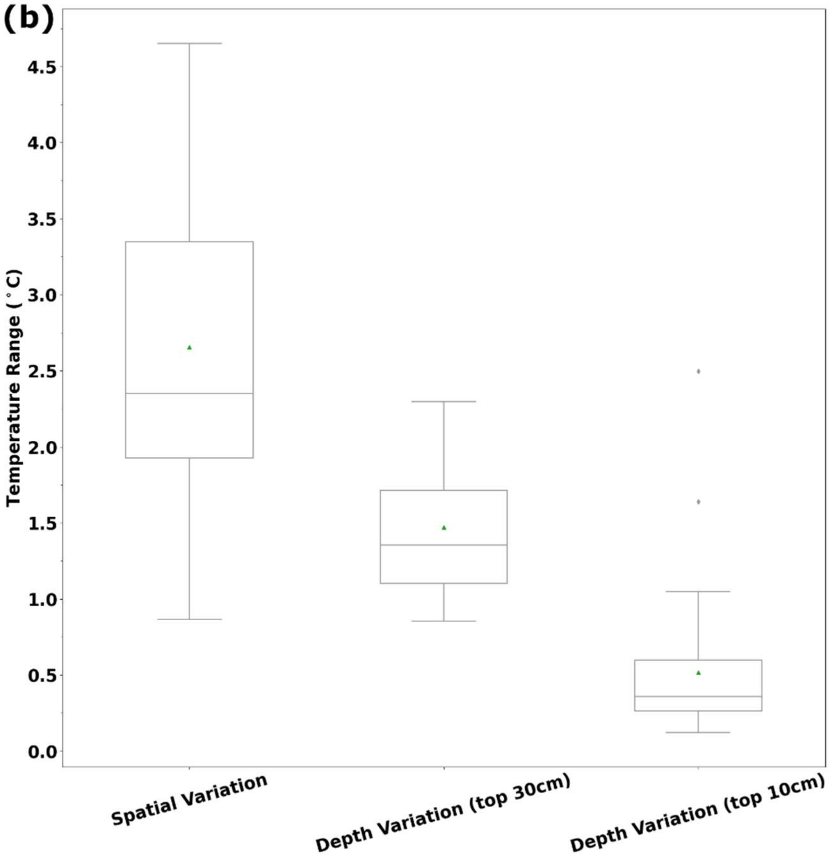
## Minor Comments

1. LL140-141: *"Panel B of Figure 1 shows the spatial standard deviation of monthly surface soil temperatures for grid cells with more than two stations included."* However, in Figure 1b, grid cells with two stations are also shown. Also, I remain unsure whether there are any grid cells with >1 stations in Eurasia?

*Based on the grid cells that met our criteria for validation, there were no grid cells in Eurasia with two or more stations included. A clarification has been added to the text:*

> *Over Eurasia, grid cells contained a single in situ measurement location.*

*Figure 1 has been revised as part of the changes that were made to Section 5.2 (Trends and Variability in Seasonal Extremes in the Ensemble Mean Product). Panel B now compares the average temperature variation between stations within a grid cell (spatial variability) with the average variation in soil temperatures across the top 30cm at a* *particular station, for a subset of grid cells that have a consistent number of stations and depths in their timeseries.*



2. L236: Reference should be to Fig. S1, right?

L236 mentions that “several factors may explain the increased variability in soil temperatures over permafrost regions.” We presume that you may have meant L226, which describes the difference in the mean bias/RMSE over North America versus Eurasia?

In the original manuscript, Figure S1 displays the mean bias and RMSE over the combined Pan-Arctic permafrost zone. Here we meant to refer to Figure S2, which showed the difference in bias between Eurasia and North America.

We have since made a change so that Figure S1 now refers to the bias and skill score of products in permafrost regions versus those outside the permafrost zone:

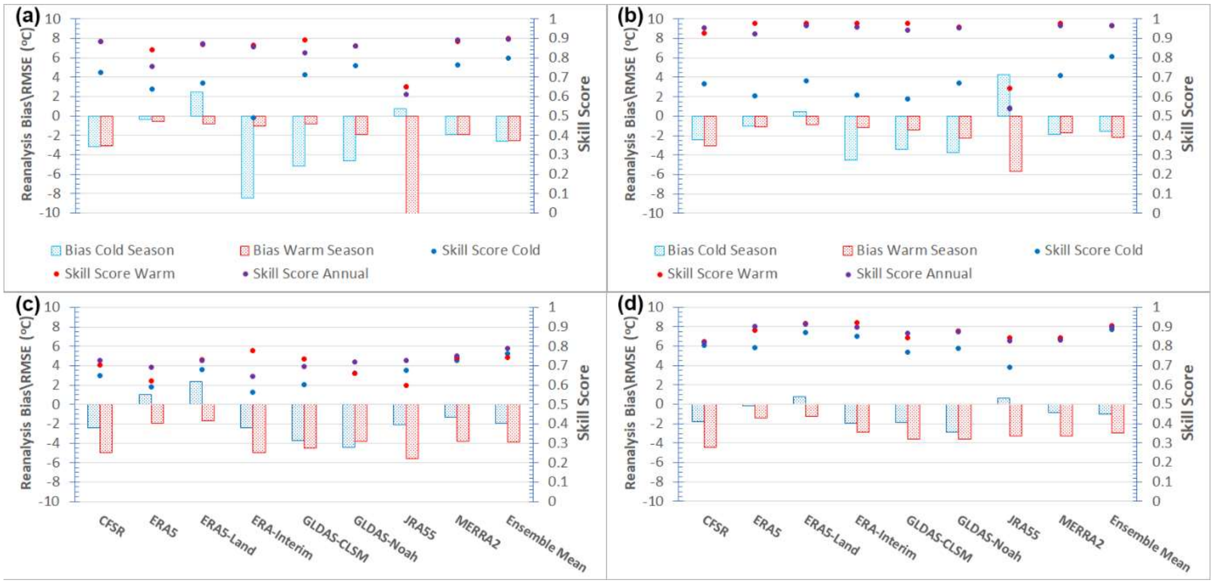


*Figure S1. Bias (bar plot) skill scores (scatter) for the cold season (≤-2°C) and the warm season (> -2°C) over the permafrost zone (Panels A and C) and the zone with little-to-no permafrost (Panels B and D) . The top panels display the bias and RMSE for the near surface (0cm - 30cm) layer, while the bottom panel displays the bias and RMSE at depth (30cm - 300cm). Models are ordered based on cold season RMSE (from the smallest to largest). Products are listed in alphabetical order, with the ensemble mean listed at the end for comparison.*

Figure S2 now shows the same metrics, but for Eurasia and North America:

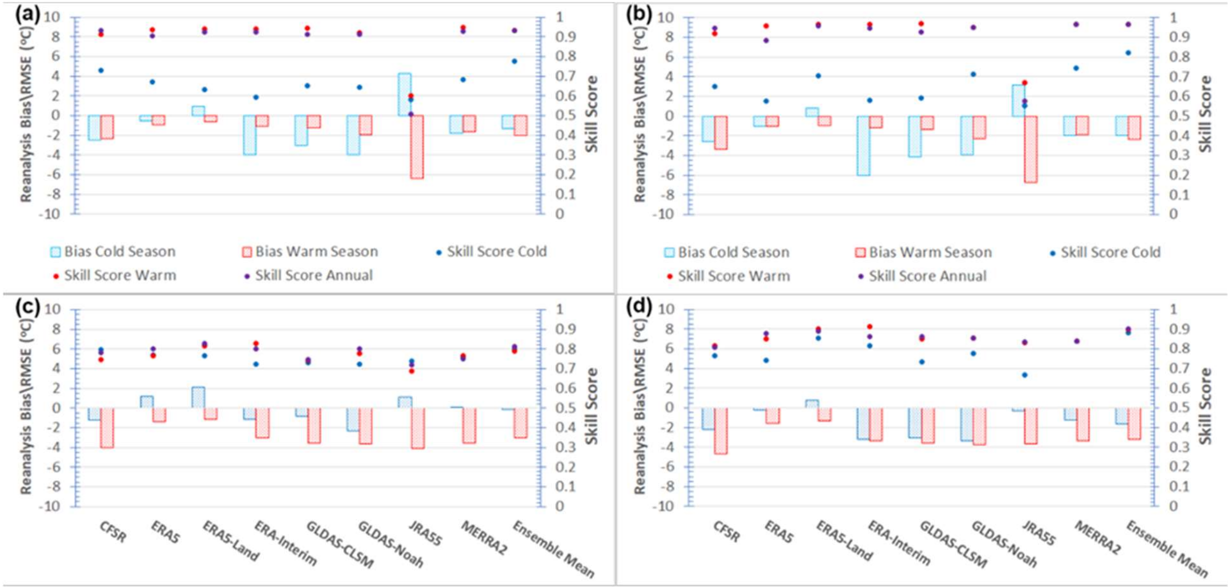


*Figure S2. Bias (bar plot) skill scores (scatter) for the cold season (≤-2°C) and the warm season (> -2°C) over North America (Panels A and C) and Eurasia (Panels B and D) . The top panels display the bias and RMSE for the near surface (0cm - 30cm) layer, while the bottom panel displays the bias and RMSE at depth (30cm - 300cm). Products are listed in alphabetical order, with the ensemble mean listed at the end for comparison.*

3. L239: What correlations, the ones between measurements and reanalysis temperatures? A slight elaboration would help the reader to see that what are compared in the sentence.

Yes, we were referring to correlations between the observed soil temperatures and the reanalysis temperatures. Section 4 has been substantially revised, and we have taken care to clarify that correlations are referring to those between reanalysis soil temperatures and station soil temperatures.

4. LL240-241: I also struggled with this sentence. What is the opposite situation here? It is hard to follow the comparisons between permafrost and little to no permafrost, as well as near-surface and at depth temperatures at the same time, especially since the results are not shown.

Here we were referring to the fact that the reanalysis products are more likely to overestimate the observed variance over the permafrost region at depth. The results section has been substantially altered in a way that we believe is easier to follow (see our response to comment #3 for further details).

5. LL243-246: Are these results related to the permafrost binning? It's fine if they are not, but overall Section 4.3 is at times hard to follow because it deals with both permafrost binning and regional comparisons.

We have revised Section 4.3 such that differences between permafrost/no permafrost and regional differences are more clearly separated. We begin with a comparison of the differences between regions with permafrost, and regions outside the permafrost zone, and then include a short paragraph at the end of Section 4.3 that explains differences over North America and Eurasia.

6. L405: Instead of the cold season standard deviations, should you not refer here to cold stations/observations? That is, figure 6 does not distinguish between warm and cold season.

What we were describing here is that when soil temperatures are frozen (and particularly for soil temperatures below –20°C), soil temperature standard deviations increase to near 10°C in several products. The sentence has been altered as follows:

> For individual products, the variability in reanalysis soil temperature for a given observed soil temperature (as measured by their standard deviation) is generally greatest over frozen soil conditions (particularly temperatures below -20°C) - further evidence of the reduced agreement between product soil temperatures and observations.

7. L261: The ensemble mean product is not properly addressed until deep into the results (validation) section. I suggest presenting the ensemble mean product and its calculation procedure already in the early stages (possibly inside section 2.1.).

As suggested, we have added a subsection in the Methods – Section 3.3 that explains how the ensemble mean soil temperature product was created, and the soil layers included from each product for the two depths.

8. L303: I find “coastal regions” not the ideal term here because the regions with the highest variability are far more than that. In winter, greatest variation associates with the coldest regions, yet not exclusively either. Could the variation here be related to snow cover duration / snow thickness as mentioned elsewhere in the text?

We agree that “coastal regions” does not adequately describe the spatial pattern of variability – particularly in winter – a more appropriate description would likely be “coldest regions”. Figure R1 shows a scatterplot of the relationship between mean annual air temperature (MAAT) and soil temperature standard deviation, when soil temperature variance is largest. The figure shows that soil temperature standard deviation and MAAT have a moderately strong negative correlation of -0.69. Moreover, it appears that regions with extreme continentality (such as eastern Siberia) show the largest standard deviations. While it is possible that snow cover characteristics may be important in certain regions, a detailed snow cover analysis is beyond the scope of this paper – and will be the focus of a follow-up paper.

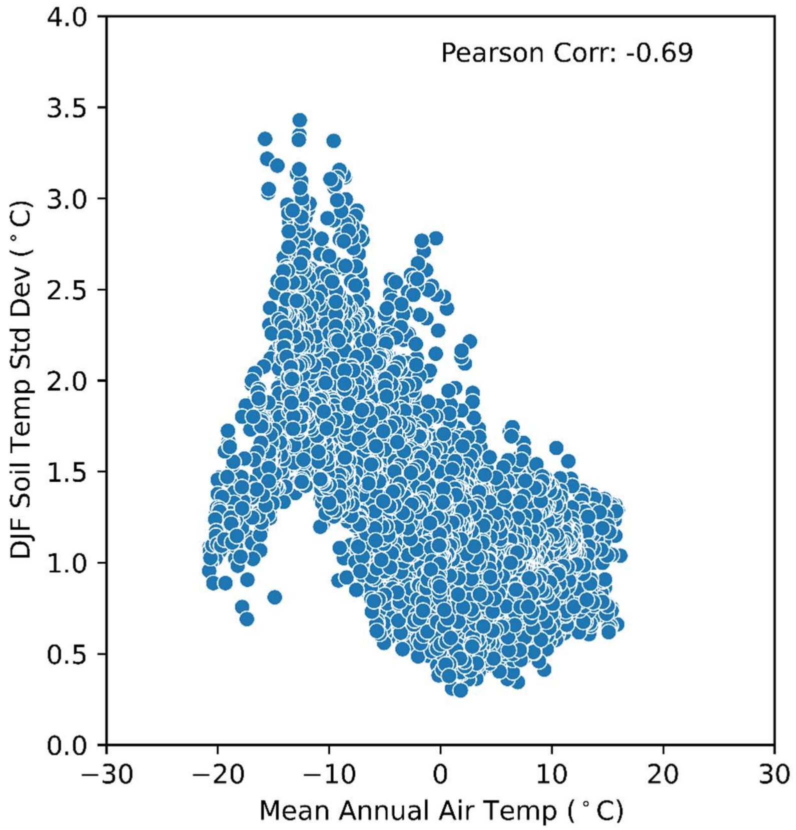


*Figure R1. Relationship between soil temperature standard deviation and mean annual air temperature (MAAT) for DJF over the 1991-2010 climatological period.*

Section 5.2 has been substantially altered and now focuses on validating soil temperature trends. When describing the ensemble mean temperature trends, we now say the following:

> The ensemble mean soil temperature dataset shows that most regions see small positive annual mean soil temperature trends of ≤ 1°C decade$^{-1}$, with slightly larger trends in the Canadian Arctic Archipelago and in Siberia, for example.

## Technical Corrections

1. L61: Please, open the abbreviation GLDAS-CLSM already here.

We have expanded GLDAS-CLSM to read “Global Land Data Assimilation System – Catchment Land Surface Model” here:

> Products were remapped onto the Global Land Data Assimilation System – Catchment Land Surface Model (GLDAS-CLSM) grid for comparison, using three different methods: nearest neighbour, bilinear interpolation, and first-order conservative remapping.

2. LL80-83: Check grammar of the sentence. Maybe delete the word "that" at line 81?

This sentence should say “In ERA5, a weak coupling exists between the land surface and atmosphere. It includes an advanced LDAS that incorporates information regarding the near-surface air temperature, relative humidity, as well as snow cover (de Rosnay et al., 2014), along with satellite estimates of soil moisture and soil temperature from the top 1m of soil (de Rosnay et al., 2013).”

3. L191: Figure 2 does not have panels C and D.

This sentence should read “Warm season biases tend to be slightly larger at depth (Figure 2, Panel B) for most products (by $1^{o}$C – $2^{o}$C).”

4. Figure 3: This is a nice figure with lots of information in it. The letters in “Correlation coefficient” are clumped together and could be corrected.

We agree – Figure 3 (now Figure 6 in the revised manuscript) has been updated to correct the issues with “Correlation Coefficient” in the text.

5. Figure 4: Stratification of the values in histograms is not explained. Please add it to the caption.

This figure has been altered in the revised manuscript in response to other revisions, and the histogram no longer appears in the updated version.

6. Figure 5: Y-axis is a bit messy. Consider adjusting the interval at which temperatures are denoted.

We have altered Figure 5 (now Figure 3 in the revised manuscript) so that major ticks at every 2◦C.

7. Figure 8: DJF missing from Panel A label.

Figure 8 has changed in response to other revisions and has been replaced.

8. L286: NH → northern hemisphere

Instances of NH have been changed to “northern hemisphere”.

9. L290: Why are ensemble mean at depth temperatures not shown? Could be part of the supplement. Figure 9 also shows at depth results, so it would be interesting to

> see how the models reconstruct frozen ground in JJA, although it is acknowledged that this is not explicitly representative of permafrost.

Our decision to not include the results at depth was because the pattern correlations were quite similar to those near the surface (with a pattern correlation of ~0.95 over the study area). The overall features were generally quite similar, however showing a smaller annual range of temperatures.

Section 5.2 has been substantially altered to focus on validation of annual mean soil temperature trends and performance in the winter minimum and summer maximum soil temperatures (Figures 9 - 11). While the section focuses on near surface performance, we have included a brief comparison with the performance at depth and have included equivalent figures as supplementary figures (Figure S5 and A6).

10. L366: Please put Gruber et al. 2018 inside parentheses.

This sentence has been removed as the Discussion section has undergone substantial revisions. We have ensured that citations are correctly formatted.

11. L369-370: *"Moreover, the impact of snow cover on soil temperature is generally more pronounced over permafrost regions (regions of seasonal frost).“* Is something missing here? Should it be "compared to regions of seasonal frost" or what is the idea?

We agree that this sentence is confusing. It should have read “Moreover, the impact of snow cover on soil temperature is generally more pronounced over permafrost regions relative to regions of seasonal frost.”

Section 6 (Discussion and Conclusion) has undergone substantiative revisions in response to comments from Referee 1, so the sentence no longer appears in the updated manuscript.

12. LL418-419: Could you elaborate, what does it mean *"is being explored"?*

Using a similar blending methodology, we have been exploring the impacts of spatial resolution on soil temperature performance, using a subset of the products examined in this study. A 0.31-degree product based on CFSR, ERA5, ERA5-Land and GLDAS-Noah was explored, along with a 0.05-degree soil temperature product, using interpolated soil temperatures from the Arctic System Reanalysis version 2 (ASR), ERA5-Land, and the Famine Early Warning Systems Network (FLDAS).

We have included the following information in our revised manuscript:

> Using a similar blending methodology, we have been investigating the performance of a 0.31-degree product (using a smaller subset of products that provide data at higher spatial resolution). We have also performed similar analyses with a 0.05-degree soil temperature product, using interpolated soil temperatures from the Arctic

> System Reanalysis version 2 (ASR), ERA5-Land, and the Famine Early Warning Systems Network (FLDAS). The goal has been to assess the impact of spatial resolution on performance of the ensemble mean product. We are hoping to include these results in a follow-up paper.

13. L428: Please provide a url for the ensemble mean dataset on the ADC.

The original version we submitted had all URLs as hyperlinks. We see that the hyperlinks are not present in the version available online, so we have added a hyperlink to the ensemble mean dataset:

> The ensemble mean soil temperature dataset has been made available on the Arctic Data Center (https://doi.org/10.18739/A2RN3085P).

14. L583: Database title and url missing.

We have added a database title and URL for Heather Kropp's dataset.

> while the Kropp et al. (2020) dataset is available from https://doi.org/10.18739/A2736M31X.

**References:**

Agriculture and Agrifoods Canada. Real Time In-Situ Monitoring for Agriculture. (Retrieved: July 2022 from: https://agriculture.canada.ca/SoilMonitoringStations/historical-data-en.html)

Alberta Agriculture, Forestry and Rural Economic Development, Alberta Climate Information Service (ACIS). Current and Historical Alberta Weather Station Data. (Retrieved: July 2022 from: https://acis.alberta.ca).

Cao, B., Quan, X., Brown, N., Stewart-Jones, E., and Gruber, S.: GlobSim (v1.0): deriving meteorological time series for point locations from multiple global reanalyses, Geosci. Model Dev., 12, 4661–4679, https://doi.org/10.5194/gmd-12-4661-2019, 2019.

Cameron, E. A., Lantz, T. C., O'Neill, H. B., Gill, H. K., Kokelj, S. V., and Burn, C. R.: Permafrost Ground Temperature Report: Ground temperature variability among terrain types in the Peel Plateau region of the Northwest Territories (2011-2015), Northwest Territories Geological Survey, Northwest Territories, Canada, 2019.

Déry, S.: Cariboo Alpine Mesonet (CAMnet) Database,. Zenodo. https://doi.org/10.5281/zenodo.1195043 , 2017.

Ensom, T., Kokelj, S. V., and McHugh, K. K.: Permafrost Ground Temperature Report: Inuvik to Tuktoyaktuk Highway stream crossing and alignment sites, Northwest Territories, Northwest Territories Geological Survey, Northwest Territories, Canada, 2020.

Gisnås, K., Westermann, S., Schuler, T. V., Litherland, T., Isaksen, K., Boike, J., and Etzelmüller, B.: A statistical approach to represent small-scale variability of permafrost temperatures due to snow cover, The Cryosphere, 8, 2063–2074, https://doi.org/10.5194/tc-8-2063-2014, 2014.

Gruber, S. Derivation and analysis of a high-resolution estimate of global permafrost zonation. *The Cryosphere*, *6*(1), 221–233. https://doi.org/10.5194/tc-6-221-2012. 2019.

Gruber, S., Brown, N., Stewart-Jones, E., Karunaratne, K., Riddick, J., Peart, C., Subedi, R., and Kokelj, S. V.: Permafrost Ground Temperature Report: Ground temperature and site characterisation data from the Canadian Shield tundra near Lac de Gras, Northwest Territories, Canada, Northwest Territories Geological Survey, Northwest Territories, Canada, 2019.

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.

Lee, S., Black, T. A., Nyberg, M., Merkens, M., Nesic, Z., Ng, D., and Knox, S. H.: Biophysical Impacts of Historical Disturbances, Restoration Strategies, and Vegetation Types in a Peatland Ecosystem, J Geophys Res Biogeosci, 126, https://doi.org/10.1029/2021JG006532, 2021.

Lee, S.-C., Christen, A., Black, A. T., Johnson, M. S., Jassal, R. S., Ketler, R., Nesic, Z., and Merkens, M.: Annual greenhouse gas budget for a bog ecosystem undergoing restoration by rewetting, Biogeosciences, 14, 2799–2814, https://doi.org/10.5194/bg-14-2799-2017, 2017.

Hernández-Henríquez, M. A., Sharma, A. R., Taylor, M., Thompson, H. D., and Déry, S. J.: The Cariboo Alpine Mesonet: sub-hourly hydrometeorological observations of British Columbia's Cariboo Mountains and surrounding area since 2006, 18, 2018.

MAWN.: Michigan Automated Weather Network. (Retrieved: July 2022 from: https://mawn.geo.msu.edu/)

Morris, J., Hernández-Henríquez, M., and Déry, S.: Cariboo Alpine Mesonet meteorological data, 2017-2021. Zenodo. https://doi.org/10.5281/zenodo.6518969, 2021.

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, https://doi.org/10.1139/e2012-012, 2012.

NDAWN.: North Dakota Mesonet Network. (Retrieved: July 2022 from: https://ndawn.ndsu.nodak.edu/).

Obu, J., Westermann, S., Bartsch, A., Berdnikov, N., Christiansen, H. H., Dashtseren, A., et al. Northern Hemisphere permafrost map based on TTOP modelling for 2000–2016 at 1 km2 scale. *Earth-Science Reviews*, *193*, 299–316. https://doi.org/10.1016/j.earscirev.2019.04.023. 2019.

RISMA.: Real-Time In-Situ Soil Monitoring for Agriculture. (Retrieved: July 2022 from: https://agriculture.canada.ca/SoilMonitoringStations/)

RoTimi Ojo, E. and Manaigre, L.: The Manitoba Agriculture Mesonet: Technical Overview, Bulletin of the American Meteorological Society, 102, E1786–E1804, https://doi.org/10.1175/BAMS-D-20-0306.1, 2021.

Spence, C. and Hedstrom, N.: Baker Creek Research Catchment Hydrometeorological and Hydrological Data, 2018a.

Spence, C. and Hedstrom, N.: Hydrometeorological data from Baker Creek Research Watershed, Northwest Territories, Canada, Earth Syst. Sci. Data, 10, 1753–1767, 2018b.

Street, L. E., Mielke, N., and Woodin, S. J.: Phosphorus Availability Determines the Response of Tundra Ecosystem Carbon Stocks to Nitrogen Enrichment, Ecosystems, 21, 1155–1167, https://doi.org/10.1007/s10021-017-0209-x, 2018.

Streletskiy, D. A., Shiklomanov, N. I., & Nelson, F. E. Spatial variability of permafrost active-layer thickness under contemporary and projected climate in Northern Alaska. *Polar Geography*, *35*(2), 95–116. https://doi.org/10.1080/1088937X.2012.680204. 2019.

Tao, J., Koster, R. D., Reichle, R. H., Forman, B. A., Xue, Y., Chen, R. H., and Moghaddam, M.: Permafrost variability over the Northern Hemisphere based on the MERRA-2 reanalysis, The Cryosphere, 13, 2087–2110, https://doi.org/10.5194/tc-13-2087-2019, 2019.

Yukon Geological Survey: Yukon permafrost reports data. In: Yukon Permafrost Database. Government of Yukon, 2021.