

TTOP Sensitivity Reviewer Response

We thank the reviewers for their constructive suggestions and comments. We feel they will improve the clarity and strengthen the results of this study. As most of the recommendations were minor, we implemented the suggested changes in this version of the manuscript. If we deviated from the reviewer recommendation details are given for each individual response.

Reviewer 1 General Comment

This manuscript presents a rigorous and data-rich evaluation of the Temperature at the Top of Permafrost (TTOP) model, combining a deterministic sensitivity analysis with a random forest variable-importance approach. Using air and ground temperature data from 330 sites across northern Canada, the authors assess how key TTOP parameters (n-factors, degree days, and the thermal conductivity ratio) influence model performance and transferability. The integration of empirical sensitivity testing with machine-learning diagnostics represents a methodological advance in permafrost modelling using the TTOP approach.

The paper is well structured, methodologically sound, and clearly written. The combination of analytical and data-driven approaches is innovative. The study provides valuable empirical evidence on parameter importance and offers practical recommendations for improving permafrost model parameterization. Overall, this is a high-quality contribution deserving publication after a few corrections based on my comments below.

We thank the reviewer for their insightful feedback and constructive comments

Reviewer 1 Specific comments

Abstract

- It would help to clarify whether leave-one-out cross-validation applies to the random forest or the sensitivity analysis.

The leave-one-out cross validation applies to the sensitivity analysis. We clarified this in the abstract.

A leave - one - out cross-validation approach to determine model sensitivity was combined with random forest analysis to rank variable importance.

- Including a performance metric (for instance the RMSE reported later) could strengthen the abstract.

We included the RMSE in the abstract as suggested.

- “Changing climate” may be a more neutral term than “warming climate.”

We changed to changing climate as suggested.

- The final sentence could more clearly underline the novelty of the pan-Canadian empirical assessment.

Now reads “Overall, this study is the first empirically-based Canada-wide assessment of how different climatic and environmental factors affect the accuracy of permafrost temperature modelling and provides practical guidance for improving parameterization in regional and global permafrost models.”

Introduction

L68: Minor typo (“A of the primary challenge”).

Typo is now fixed

L68–75: The knowledge gap could be stated more explicitly, especially regarding the relative importance of nf, nt, and rk across regions.

Now reads “Few studies have examined the uncertainties arising from mischaracterization of the TTOP model parameter values or the relative importance of each parameter **which may vary substantially** in different permafrost environments (Riseborough, 2004; Way & Lewkowicz, 2018).”

L89–105: The random forest paragraph is somewhat general; a shorter description tied directly to the study aims might improve flow.

We removed the highlighted section of the paragraph to focus more on the use of random forest variable importance rankings in permafrost research.

Random forest is a supervised machine learning technique, which combines randomized decision trees with bagging, and aggregates their predictions through averaging or majority vote (Breiman 2001; Biau & Scornet, 2016). Random forest has been used in studies of air quality (Yu et al., 2016; Pendergrass et al., 2022), chemoinformatics (Mitchell, 2014), ecology (Cutler et al., 2007; Brieu et al., 2018) and remote sensing (Belgiu & Drăgu, 2016). Recently, random forest has been used in spatial mapping of permafrost presence using environmental predictors (topography, rock glaciers, vegetation, and land surface characteristics) in a variety of environments (Pastick et al., 2015; Deluigi et al., 2017; Baral & Haq, 2020). Random forest also

allows determination of variable importance rankings which can be used to either identify important variables for explanatory or interpolation purposes or to identify a small number of variables that provide a good prediction (Díaz-Uriarte & Alvarez de Andrés, 2006; Grömping, 2009; Genuer et al., 2010).

L106–111: The objectives could be phrased to highlight the combined use of sensitivity analysis and machine-learning-derived importance.

We combined the first and second objectives to the highlighted text. Now reads “The objectives of this study are: (1) to use both a sensitivity analysis and machine learning (random forest) to evaluate TTOP model parameter importance using field observations and (2) to assess the accuracy of the TTOP model using measured parameters across permafrost regions of Canada.”

Methods

L205: A brief rationale for using percentile substitution would clarify this choice.

We chose to use percentile substitution as it allowed us to increase and decrease parameter values while avoiding erroneous negative values. This method also kept the parameter change within the measured distribution, therefore keeping them more realistic than a plus/ minus change in magnitude. Additionally, percentile substitution allows for a direct comparison of sensitivity to each parameter across regions as the parameters at all sites were changed to the same value. The main disadvantage of this sensitivity testing method is that the values of the percentile are directly dependent on the range and spread of the data. As a result, the direct comparison of parameter sensitivity may not be possible, the step change for the individual observations were not the same. Therefore, the corresponding difference in the perturbed and reference TTOP may be higher for one parameter due to a larger change in value rather than an increased model sensitivity. However, this is likely not the case for this analysis as each TTOP model parameter had a large range of values since sites from a wide range of climatological and environmental conditions were sampled.

New manuscript text “We selected percentile-based substitution to test TTOP Model sensitivity as it allowed us to increase and decrease parameter values within observed ranges while avoiding introducing negative values. Additionally, percentile substitution allows for a direct comparison of sensitivity to each parameter across regions as the parameters at all sites were changed to the same value.”

L215–225: The target variable used in the RF models is worth specifying.

The target variable was the mean annual ground temperature at the top of permafrost (where present) or the bottom of seasonal frost where near surface permafrost was absent. This is now included in the methods section as suggested

L226–235: Commenting on RF repeatability and possible predictor correlations would strengthen this section.

In the uncertainties section in the discussion, we included this text “Finally, it is important to note that due to the nature of random forest, the variable importance rankings are not perfectly repeatable. However, in several random forest runs the most important and least important parameters were consistent even if they were not in the exact same order each time.” Some correlation values are given in the discussion as well and the correlation matrices are provided in the Supplemental data.

L236–241: Listing the performance metrics used (RMSE, bias, R^2) and whether they are per site or per site-year would aid transparency.

Now in the methods “TTOP model performance was based on model root mean square error (RMSE), r^2 , and bias compared to measured temperatures for individual years and sites”.

Results

L242–252: Adding mean \pm SD to Table 4 would contextualize the percentile ranges.

This has been added to Table 4

L296–307: Clarify whether the importance values are averaged across sites or regions and consider noting the correlation between TO and rk.

Now included in manuscript “Variable importance plots were created for each random forest model both for the entire dataset (averaged across all sites) and for each region individually (averaged for all sites within each region) (e.g. Colyn et al. 2025).

As in previous comment we include in the correlation between TO and rk and correlation matrices are now in the supplemental data.

L322–331: A simple correlation (e.g., Spearman) between sensitivity and RF rankings could help compare approaches.

We ran a Spearman correlation for the parameter importance rankings between the two methods as suggested and the results are in Table 6

L332–341: The validation could be complemented with R^2 or NSE, and a brief note on the permafrost classification criterion.

Bias and r^2 are now given in this section in addition to the permafrost classification criterion. Now reads “The model also has a slight warm bias (0.02 °C) and a high r^2 (0.99).”... “Additionally, the

model only misclassified permafrost presence ($TTOP < 0^{\circ}\text{C}$) or absence ($TTOP > 0^{\circ}\text{C}$) at 3 of the 612 observations.”

Discussion and Conclusions

L351–366: A short explanation of why n_f dominates (snow–air coupling) could enhance clarity.

Now included “Additionally, n_f represents the impact of freezing season air temperature and snow depth which has an important influence on the ground thermal regime and therefore permafrost occurrence in the discontinuous zone (Riseborough & Smith, 1998; Smith & Riseborough, 2002, Gislén et al., 2014; Way & Lapalme, 2021, Tutton et al., 2021; von Oppen et al., 2022). Therefore, it is unsurprising that it was consistently ranked as an important parameter.”

L387–406: A bit more context on why TO and r_k rank highly in RF, including their correlation, would aid interpretation.

We believe that the importance of these parameters in the RF with all parameters is skewed based on the highly correlated nature of the other parameters to each other since in the RF with only TTOP model parameters r_k is not important. In the random forest r_k and TO are only highly correlated with each other and not the other parameters. Potentially, their elevated importance may also come from the smaller number of High Arctic sites compared to the sensitivity analysis, though we think the correlation is the more likely explanation.

As with other comments correlation data is now included in the text and the supplemental

L437–463: The uncertainty section is strong; providing the correlation matrix and distinguishing sensitivity from statistical importance could add clarity.

As in previous responses we included correlation matrices in supplemental. Now included in text “ Lastly, parameters sensitivity rankings do not inherently relate to statistical importance. TTOP model sensitivity to changes in a parameter value may not be statistically different from the sensitivity to changes in another (Table 5). This is especially true for certain regions (N NWT & NU and the more southern regions), where there are few statistically different sensitivities between parameters.”

Reviewer 2 Comments

It was a pleasure to read this manuscript.

The manuscript includes a well-crafted introduction and provides an appropriate overview, incorporating both key studies and recent advancements in the field.

The study's objectives, methods, and results are clearly defined and effectively presented. It is well-written, accompanied by well-designed figures, and easy to follow.

The study highlights that the TTOP model's performance is most sensitive to freezing-season parameters, particularly the freezing n-factor and freezing degree days and clarifies the relative importance of climatic and environmental factors. It offers practical recommendations for improving model parameterization at different scales and locations. It also emphasizes the significance of in situ data in validating parameters for accurate predictions of permafrost and ground temperatures.

With the incorporation of the suggestions mentioned below in a revised version, I recommend it for publication in TC.

We would like to thank the reviewer for taking the time to read our manuscript and provide helpful and constructive comments.

General comment:

The manuscript highlights that parameters derived from landcover classes are not transferable between sites, emphasizing the need for locally calibrated inputs. Expanding on how calibration methods can be improved across regions could strengthen the practical applicability of the findings.

The lack of transferability likely stems from differences in snow depth and morphology between landcover classes across Canada, which is why using external n_f values always resulted in higher errors than n_t (although this could also just be a product of the relative importance of the parameters as well). Now included in the discussion “The lack of transferability of n_f likely stems from differences in snow depth and density across Canada (Bormann et al., 2013; Way & Lewkowicz, 2016, Simpson et al., 2022). As a result, utilizing snow redistribution algorithms are likely a more viable way to accurately capture).permafrost presence and temperature on national and circumpolar scales (Gisnås et al., 2014; L’Hérault et al., 2017; Obu et al., 2019). However, using landcover as a proxy even when parameter values were outside of the region still resulted in a smaller error range than when values from the entire dataset (regardless of landcover type) were used.”

Specific comments:

P10, L148-150: It would be useful to include (here or in the discussion) a short description on the sensitivity of the variation in measuring depth (within the 2 to 5 cm range) for calculation of

the ground surface temperature variables (MAGST, FDDs, and TDDs) and how it may influence the results.

Difference in the ground surface sampling depth from 2 to 5 cm likely did not impact the resulting analysis. The average change is -0.003 °/cm which is orders of magnitude less than the accuracy and precision of the loggers used. Even over 5 cm the temperature change would still only be -0.02 °C still outside of the logger accuracy and precision as noted in the manuscript. In a previous paper, Garibaldi et al., 2021 we were also asked this by a reviewer and showed in our response that the temperature change over 3 cm in the Canadian High Arctic had no impact on the resulting analysis (though this was not included in publication). Added to methods “The slight difference in surface sampling depth (~ 3 cm) did not have an impact on the results as the temperature difference is outside the logger accuracy.”

P10-11: Regarding the different data loggers used, you should include the specific manufacturer name, city, and country of origin etc.

We included the manufacturer name in the text. In previous publications using similar loggers the city and country typically are not included within the text. As we use several different logger types (with different manufacturers) we feel this information may unnecessarily hurt the flow of the paper. However, we can include this information in the supplemental data if required or perhaps this could be included as a footnote within the paper if the reviewer deems it necessary.

P27, L424-436: The Southern NWT and Southern Yukon-Northern BC regions contained several outliers with significant errors. Expanding the discussion on this result would be beneficial.

The Southern NWT outlier point was one year at a site with relatively warm ground conditions during the freezing season leading to a low n_f (0.1) compared to the other 12 years (0.46 on average). However, this warm winter did not impact the ground temperature as the annual mean ground temperature increased slightly during this year but remained comparable to the other years. As a result, the TTOP model produced a larger error for this site during this year. This has been added to the discussion.

The Southern Yukon-Northern BC main outliers were a small calculation mistake (and are no longer outliers) and will be corrected in the next version of the figure.

P37, L753-L759: This reference includes some misspelled names and an incomplete co-author list. Please verify it against the name list from the original report.

Thank you for pointing out the omission of one of the lead authors and this will be corrected. For these reports normally only the lead author team is listed in the citation (contributing authors are not included as indicated in the suggested citation provided in the report). The citation included in the revised reference list is:

Romanovsky, V., Isaksen, K., Drozdov, D., Anisimov, O., Instanes, A., Leibman, M., McGuire, A.D., Shiklomanov, N., Smith, S., & Walker, D. Chapter 4, Changing permafrost and its impacts. In Snow, Water, Ice and Permafrost in the Arctic (SWIPA) 2017. Arctic Monitoring and Assessment Program (AMAP) Oslo, Norway. 65-102. ISBN 978-82-7971-101-8, 2017

<https://www.amap.no/documents/doc/snow-water-ice-and-permafrost-in-the-arctic-swipa-2017/1610>