Line numbers in referee comments (in blue) refer to the previously revised manuscript, whereas line numbers in our responses (in black), refer to the newly revised clean manuscript.

The authors have done a really nice job in the proposed revisions and thoroughly addressed most of my comments. The clarifications are helpful. I think the revised figures 2 and 7 are really nice. I think there are still a few points that could be improved in the manuscript text.

R1: We thank Dr. Treat for her encouraging words. Below we answer to her further comments in detail.

Line 95: How was this determined? what was the peat thickness threshold? Was it a requirement that the soil profile be described, or was anything that was called a palsa considered a palsa? Also, what about the distinction between palsa and pingo and was it applied here?

R2: In many palsa and peat plateaus studies the peat thickness is not necessarily mentioned and no soil profiles were available for all observations. Some of the utilized observations were just point data, which were accepted to our study after the satellite image verification.

In general landforms which were described as palsas in the utilized literature were considered as palsas. An exception of this was the cases where the landforms were called as mineral palsas/ palsa-like formations or it was mentioned that there was no peat at the study site. Palsas and pingos differ from each other by their formation process (Mackay, 1988; French, 2017) and thus pingos were also excluded from our observations. The same model could not realistically describe the suitable environments for palsas, peat plateaus, and pingos. Distinction between these landforms was quite easy to make based on the landform descriptions in literature and visual interpretation of the satellite images. Furthermore, pingos are usually found from much colder regions than palsas and peat plateaus (see e.g., Grosse & Jones, 2011).

To conclude, the chosen landforms were selected based on their overall description the literature (not only by a strict peat thickness threshold) or they were recognized in the comprehensive interpretation of the satellite images as mentioned in the manuscript. However, we modified the manuscript to clarify this to the readers even more. We also mentioned pingos in the revised paragraph.

Revised paragraph (starting from line 95):

Palsa observations were limited to 'true palsas' (Fig. 1b) based on the overall description of the landform provided in the used literature (see Appendix A) or a comprehensive interpretation of the satellite images (based on the size, form, and location of the landform and its surroundings for example). 'Palsa-like' formations and 'lithalsas' (so-called mineral palsas) were excluded, because of the lack of substantial peat cover on these landforms (Pissart, 2002). Pingos were also considered genetically distinctive from palsas and thus were not included (see e.g., Grosse & Jones, 2011; French 2017). Additional exclusions were made in cases where a landform was called palsa, but the source clearly stated that not peat was present at the study site. In this study, peat plateaus were considered as palsa plateaus, instead of distinguishing these two landforms from each other, (Zoltai and Tarnocai, 1975; Kershaw and Gill, 1979; Seppälä, 1988; see also Fewster et al., 2020, 2022). All compiled observations were verified by using satellite imagery in Google Earth Pro and ESRI's ArcGIS Pro (Fig. 1c). The final modelling data consisted of 961 grid cells (30 arc sec resolution, ~ 1 km) occupied by palsas or peat plateaus.

For example, sometimes geomorphologists or geologists call things a peat because it has a thick O horizon that might not technically be a peat (such as a soil with a 10cm thick organic peat horizon). (in non-permafrost terms, sometimes engineers or geochemists call things a bog because there is peat when the vegetation says it's a fen). I think this could be an issue with at least one of the studies here (Nelson et al. 1985 but maybe also Throop et al. 2012).

R3: We appreciate Dr. Treat's important notion. This aspect was not considered when we collected our landform observations. As mentioned above in reply no. R2 the peat layer thickness was not available in most of the utilized studies. Because we wanted to collect spatially as comprehensive circumpolar dataset as possible, some compromises in the required landform description had to be done. Thus, it is possible that a very small number of lithalsa-like landforms with a relatively thin peat/organic layer were included in the data.

Nelson et al. (1985) studied so-called anthropogenic palsa site, which formed in response to a road construction site nearby. Landforms in the region are described as one meter high, ice cored mounds with 10–50 cm peat layer on top of them (with small amounts of mineral soil). One of their conclusions was that even though the peat layer is relatively thin in the region, it plays a crucial role in the preservation of these palsas. Thus, we decided that these mounds, could be considered as palsas, even though there might be some uncertainties.

Throop et al. (2012) described their study sites as high elevation discontinuous permafrost mound sites (site names Wolf Creek APM and Palsa 25). The peat layer is mentioned to be thick and prominent at these sites, borehole data were available. In our opinion this description was enough to include the observations to our dataset. Fewster et al. (2020) have also utilized landform observations by Throop et al. (2012) in their palsa database, which further encouraged us to use this reference in our data compilation.

Line 115: This section still needs to be clarified. Upon both first and second reading, I interpreted "In order to reduce the effect of spatial autocorrelation on model evaluation, observations in the evaluation dataset were selected so that they were located at least 50 km from each other (Roberts et al., 2016). Because of this criterion, two absence observations had to be removed from the evaluation data (final N = 398)." To me that means the sites with cores close together would be removed from the validation, which I thought would be problematic in regions of discontinuous permafrost. The response to reviewers indicates that these observations were not excluded. Which is it? If all cores are included, what does this criteria mean? Please explain further and more clearly.

R4: We apologize for the inconsistencies between our previous responses and the manuscript. Below we try to explain better the choices made.

When we originally compiled the landform observation dataset, we controlled the distance between observations so that multiple landform observations did not occur in the same grid cell (resolution of 30 arc-second). The observations, both presences and absences, were at least approximately one kilometer from each other. Because palsas and peat plateaus are typical landforms of the sporadic and discontinuous permafrost regions, we considered that there is no reason to exclude absence observations from the grid cells neighboring presences (as the permafrost can be absent from these cells even though present in the neighboring cell in these marginal permafrost regions).

After the compilation of our landform data, we split it into separate calibration (n = 2057) and evaluation (N = 400) datasets. This was done by taking random sample separately from the presence and absence observations to retain the same prevalence (ca. 40 % presences and 60 % absences) with the original dataset. When we took the random sample we used criterion that the observations had to be at least 50 km from each other. This was done to obtain spatially as comprehensive evaluation dataset as possible. In the original dataset there are some data clusters with lots of landform observations (especially in Northern Fennoscandia) and regions with only few observations (e.g., Middle and Eastern Siberia). Without this control of the minimum distance between observations, it might have been possible to have a random sample with many observations from the data clusters, and some regions with less observations might have been unrepresented in the evaluation dataset in response to this clustering.

However, the 50-km distance was utilized also between presences and absences to do the data selection consistently. This led to an exclusion of two absence observations, which located too close (<50 km) to presence observations of the evaluation datasets, and thus the final evaluation dataset consisted of 398 observations. We revised the related paragraph in the manuscript to clarify the matter and to avoid previous inconsistencies.

Revised paragraph (starting from line 112):
Compiled data were further split into model calibration (N = 2057), and evaluation (N = 400) sets. These sample sizes are expected to be large enough to give robust predictions in geomorphological distribution modelling (see Hjort and Marmion, 2008). The separate evaluation dataset was formed by drawing a random sample from the whole data. The random sample was taken separately from presence and absence observations, to retain the same relative portion (i.e., prevalence) with the original dataset (ca. 40 % presence, and 60 % absence observations) (Hjort and Luoto, 2013). In order to avoid oversampling observations from data rich regions and underrepresenting data poor regions, the random sample for evaluation data was taken so that observations located at least 50 km from each other. Because of this criterion, two absence observations had to be removed from the evaluation data (final N = 398).

Line 320: The Olefeldt thermokarst map has some limitations, perhaps this is one of them. It is also a model, models have limitations. The map identified areas with high likelihood of thermokarst (vulnerability), not necessarily high thermokarst occurrence.
R5: Dr. Treat has a crucial point in this matter. We modified our manuscript thoroughly so that the reader understands that Olefeldt et al. (2016) data represents the likelihood of thermokarst landscapes, not necessarily the thermokarst occurrence.

We also added discussion related to the limitations that the modelling approach might have to this thermokarst comparison.

Revised paragraph (starting from line 475):
Previous studies have suggested that the development of thermokarst can be used as an indicator of the former distribution of palsas (Matthews et al., 1997; Luoto and Seppälä, 2003). Thermokarst ponds are relatively common in Fennoscandian palsa mires (e.g., Luoto and Seppälä, 2003), although Olefeldt et al. (2016) classified this region to have a low likelihood for thermokarst . This mismatch was noticed also here, as we predicted extensive degradation of the landforms in the region. Overall, our results showed that regions with higher thermokarst likelihood are in a higher risk to become unsuitable environments for palsas and peat plateaus, compared to the regions with lower thermokarst likelihood. This indicated clear spatial relationship with our results and the thermokarst likelihood. It is important to note that the thermokarst data by Olefeldt et al. (2016) is based on modelling rather than observations. However, the recognition of the regions with a high degradation risk is useful in the estimation of future greenhouse gas fluxes from permafrost wetlands (Swindles et al., 2015; Miner et al., 2022) and establishing conservation actions for these endangered ecosystems and habitats (Janssen et al., 2016).

Line 430: winter vs. summer controls on palsa thawing: The problem with this statement and the conclusion is that the statistical model isn't able to test these processes related to winter snowfall or summer warming because this data isn't in the statistical model. In the response to reviewers, this was discussed thoroughly but here it is not fully incorporated. Because this is a statistical model, not a process-based model where the effects of summer warming vs winter snowfall or winter processes could be explicitly tested, this also undercuts the conclusion on line 512- 515: "The pronounced importance of thawing-season climate conditions for circumpolar palsa and peat plateau occurrence suggests that the projected increases in summer temperatures and rainfall may strongly affect the stability of permafrost peatland landforms." I think the jury is still out on this but this paper isn't in a position to answer this question.

R6: We acknowledge that we would need process-based models to explicitly answer to this research question. However, our statistical modelling is congruent with previous studies (see e.g., Peng et al. 2018; Mekonnen et al., 2021) and supports the hypothesis that increased summer air temperature can strongly affect the stability of permafrost and permafrost landforms. Our conclusion is based on the result that thawing degree days (TDD) (describing the summer conditions) had greater variable importance in our models compared to freezing degree days (FDD) (describing the winter conditions) (fig. 3). In figure 3 it can also be noticed that increasing TDD values cause the occurrence probability of the suitable environments to decrease sharply after ~1200 °C-days. When the climate changes, it can be assumed that TDD values rise, leading to our conclusion.

We want to notify that our conclusion already included uncertainties (e.g., 'may strongly affect') and we are not claiming that this is necessarily the case. However, we slightly toned down the conclusion on this matter.

Revised conclusion (starting from line 520):

The pronounced importance of thawing-season climate conditions for circumpolar palsa and peat plateau occurrence is congruent with the earlier findings that the projected increases in summer temperatures and rainfall may affect the stability of permafrost peatland landforms.

Line 465: Palsa out of equilibrium, I think this misses the point somewhat. The point is that this could shift the southern edge of palsas further south than what is predicted using the 1950-2000 CE climate or do you think these points are adequately captured in your approach and this isnt' a problem? Discuss.

R7: If we understood correctly and Dr. Treat referred to the point that our models predicted suitable environments for palsas and peat plateaus further north (especially in North America) than some peat datasets such as Treat et al. (2016; see fig. 7d in the manuscript) would suggest. By better acknowledging the role of past climates the predicted suitable environments might have shifted southwards.

We decided not to model the historical distribution of suitable environments because all environmental variables used in this study were not available for historical periods. Although there are some good datasets describing the peat initiation dates and this might not even be necessary information (as Dr. Treat pointed out in her referee report) it would have been somewhat problematic to use for example the Soil Grids data for historical periods. Thus, the usage of the 1950–2000 climate baseline period (instead of e.g., 1991–2020) was the best compromise available for our research approach in our opinion. However, we recognize the value of the point that Dr. Treat brought up in her comment.

We could not take into account the state of our landform observations (in equilibrium vs. out of equilibrium with climate) when the data was compiled. This would have been a valuable addition to the data, but unfortunately impossible one. The states of the landform observations are not necessarily mentioned in the used references and thus we were not able to consider this aspect of landforms.

To acknowledge Dr. Treat's notion that there are global peat age datasets available (e.g., Treat et al., 2017) and those would not even be necessary (Treat & Jones, 2018), we removed related sentences from the revised paragraph.

Revised paragraph (starting from line 467):
It is important to note that most of palsas and peat plateaus were formed during the Little Ice Age (the 16th to mid-19th century period), in colder climates than we are currently facing (Halsey et al., 1995; Treat et al., 2016a; Vorren, 2017; Fewster et al., 2020). Because of this not all palsas and peat plateaus are in equilibrium with the current climate anymore. We did not model the historical distribution of suitable environments by using climate data from the landform initiation period, because it would have required historical data describing the other environmental factors also. However, we acknowledged the role of past climates by choosing the 1950–2000 to be our baseline period as it was considered to better resemble the conditions of the period of palsa and peat plateau formation than the current climatology (e.g., 1991–2020; Seppälä, 2003; Saemundsson et al., 2012).

Section 3.4 Statistical and spatial evalution of the models. I would like to see this section immediately after the model description. I still have some of the problem that I had before the revisions: these model statistics don't mean anything to me ("For AUC the standard deviation range was 0.003–0.006, and for TSS 0.014–0.019"). This could be more informative as fractions or percentages, in relation to a mean, as a fraction of mis-prediction. The numbers need a lot more explanation to put them into context. The paragraph starting at line 355 does a better job at this, so open this more.

R8: We acknowledge that it would be useful to have these results immediately after the model description. However, we decided to place model evaluation results at the end of the results section, because these are not the most interesting and important results for the majority of the readers. We preferred to present the results which answered our research questions, and which are probably appreciated more by the wider audience first, and then provide these evaluation results as background knowledge to justify the choices we made (e.g., focusing on results from random forest).

To make the evaluation scores more informative we added text related to the interpretation of these evaluation metrics to the section '2.3 Statistical modelling' and also modified the related paragraph in the Results section.

Revised paragraph (starting from line 226):
Calibrated models were evaluated in two ways (Araújo et al., 2005). First, 100-fold cross-validation was performed with the calibration data. At each validation run a random sample of 70 % (N = 1440) of observations were used to calibrate models and the remaining 30 % (N = 617) was used to validate them. Secondly, semi-independent evaluation was conducted by using the separate evaluation data (N = 398), which was set aside before model calibration (as described in the section 2.1). Performance of the models was evaluated with two prevalence-independent statistical measures of classification accuracy, area under the receiving operating characteristic curve (AUC) and TSS (Allouche et al., 2006). AUC and TSS take into account both sensitivity and specificity the of the model (Allouche et al., 2006). AUC values range between 0–1, and models with AUC values > 0.5 are considered to predict the distribution better than random. AUC values 0.7–0.8 are usually interpreted as acceptable, 0.8–0.9 as excellent, and values > 0.9 as outstanding (Hosmer & Lemeshow, 2000; Mandrekar, 2010). TSS values range from -1 to 1, where value 1 indicates perfect agreement. If TSS value is ≤ 0, model performance is no better than random (Alloche et al., 2006). TSS is interpreted in the same manner than Cohen's kappa, and for example 0.6–0.79 can be considered as moderate model performance, 0.8–0.9 as strong model and >0.9 as almost perfect model performance (Landis & Koch, 1977; Allouche et al., 2006). Model reliability was also evaluated spatially by exploring the modelling agreement between four independent modelling methods (Luoto et al., 2010), and by the comparison of our predictions and the four peatland datasets (Treat et al., 2016b; Xu et al., 2018, Hugelius et al., 2020; Olefeldt et al., 2021).

Revised paragraph (starting from line 350):
The 100-fold cross-validation based on the split-sample approach yielded high AUC and TSS values (Fig. 6), indicating very good predictive performance of the models. RF and GBM best predicted the distribution of suitable environments, and the ensemble model performed slightly weaker. Between-model variability within the 100 modelling runs was estimated by the standard deviations. For AUC the standard deviation range was 0.003–0.006, and for TSS 0.014–0.019, indicating that the performance of the generated 100 models varied only a little. Evaluation with the separate dataset resulted in lower AUC and TSS values than the cross-validation, but still indicated outstanding predictive performance based on AUC values (AUC > 0.9), and moderate predictive performance based on TSS values (TSS around 0.7; Fig. 6). With the separate evaluation data, RF and GBM had the highest TSS values, whereas the ensemble and RF models had the best AUC score. Compared to the cross-validation, larger between-model variability was observed in TSS values (sd. 0.017–0.04), whereas variability in AUC values (sd. 0.001–0.004) was smaller. GLM and GAM were the weakest models in both evaluations. Based on the evaluation metrics, response curves, and predictive maps RF was considered the best method for predicting suitable environments for palsas and peat plateaus.

Data availability statement: I also think this is missing the mark. This study has clearly benefitted from other publicly available data as listed in the data availability statement. How are these study results going to be disseminated in a similarly open way? Point coordinates are available in Appendix A, but the map availability is not specified. This needs to be addressed for publication in an EGU journal and would increase the impact and re-usability of this study.

R9: We appreciate this insight and decided to publish continuous predictions of the suitable environment probability in raster format for all used modelling methods (GLM, GAM, GBM, RF and an ensemble of former methods) and climate change scenarios (RCP2.6, 4.5 and 8.5) and the baseline period (1950-2000). These files are now publicly available in Zenodo (https://doi.org/doi:10.5281/zenodo.7745085). This information was added also to the Data availability statement. True Skills Statistics cut-off values used in the binary classification are provided in the supplementary materials (Table S1).

Revised Data availability statement (starting from line 740):
Used climate data were obtained from WorldClim – Global Climatic Data, available at https://worldclim.org/ (Hijmans et al., 2005). Variables describing the soil properties were calculated from data in SoilGrids – global gridded soil information database https://files.isric.org/soilgrids (Poggio et al., 2021). Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010; Danielson and Gesch, 2011) https://doi.org/10.5066/F7J38R2N were used to calculate TWI. Thermokarst dataset (Olefeldt et al., 2016) can be obtained from https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1332. BAWLD -database (Olefeldt et al., 2021) is available at https://arcticdata.io/catalog/view/doi:10.18739/A2C824F9X, Hugelius et al. (2020) data at Bolin Centre for Climate (https://doi.org/10.17043/hugelius-2020-peatland-2), PEATMAP (Xu et al., 2018) at https://doi.org/10.5518/252, and Treat et al. (2016b) data at https://doi.org/10.1594/PANGAEA.863689. The coordinates of the landform observations compiled in this study are available in the supplementary materials. List of references used in the compilation is provided in Appendix A. Our predictions of the suitable environments for palsas and peat plateaus are publicly available in raster format in Zenodo: https://doi.org/doi:10.5281/zenodo.7745085.

Additional sites not included in current synthesis:

Turetsky, M. R., R. K. Wieder, D. H. Vitt, R. J. Evans and K. D. Scott (2007). "The disappearance of relict permafrost in boreal north America: Effects on peatland carbon storage and fluxes." Global Change Biology 13(9): 1922-1934.

Quinton, W. L. and J. L. Baltzer "The active-layer hydrology of a peat plateau with thawing permafrost (Scotty Creek, Canada) Hydrologie de la couche productive d'un plateau tourbeux avec dégel du permafrost (Scotty Creek, Canada) La capa hidrológica activa de un plateau de turba con deshielo del permafrost (Scotty Creek, Canada) 加拿大Scotty Creek泥炭高原融化的永久冻土活性层的水文研究 Hidrologia da camada ativa de um planalto turfoso com permafrost em fusão (Scotty Creek, Canadá)." Hydrogeology journal 21(1): 201-220.

Bellisario, L. M., J. L. Bubier, T. R. Moore and J. P. Chanton (1999). "Controls on CH4 emissions from a northern peatland." Global Biogeochemical Cycles 13(1): 81-91.

Wickland, K. P., R. G. Striegl, J. C. Neff and T. Sachs (2006). "Effects of permafrost melting on CO2 and CH4 exchange of a poorly drained black spruce lowland." Journal of Geophysical Research-Biogeosciences 111(G2): G02011,.

Blyakharchuk, T. A. and L. D. Sulerzhitsky (1999). "Holocene vegetational and climatic changes in the forest zone of Western Siberia according to pollen records from the extrazonal palsa bog Bugristoye." The Holocene 9(5): 621-628.

Andreev, A., V. A. Klimanov and L. D. Sulerzhitsky (2001). "Vegetation and climate history of the Yana River lowland, Russia, during the last 6400 yr." Quaternary Science Reviews 20: 259-266.

Hugelius, G. (2012). "Spatial upscaling using thematic maps: An analysis of uncertainties in permafrost soil carbon estimates." Global Biogeochem. Cycles 26(2): GB2026.

R10: We thank Dr. Treat for these useful references. At this point we are not able to utilize them in this study, but we will check these out and possibly use them in our future studies.

References used in this response letter:

Fewster, R. E., Morris, P. J., Swindles, G. T., Gregoire, L. J., Ivanovic, R. F., Valdes, P. J., and Mullan, D.: Drivers of Holocene palsa distribution in North America, Quaternary Sci. Rev., 240, 106337, https://doi.org/10.1016/j.quascirev.2020.106337, 2020.

French, H. M. The Periglacial Environment. 4th ed., Wiley-Blackwell, Hoboken. ISBN: 978-1-119-13278-3, 2017.

Grosse, G., and Jones, B. M.: Spatial distribution of pingos in northern Asia. Cryosphere, 5, 13–33, https://doi.org/10.5194/tc-5-13-2011, 2011.

Mackay, J. R.: Pingo collapse and paleoclimatic reconstruction, Can. J. Earth Sci., 25, 495–511, https://doi.org/10.1139/e88-050, 1988.

Mekonnen, Z. A., Riley, W. J., Grant, R. F., and Romanovsky, V. E.: Changes in precipitation and air temperature contribute comparably to permafrost degradation in a warmer climate, Environ. Res. Lett., 16, 024008, https://doi.org/10.1088/1748-9326/ABC444, 2021.

Nelson, F., Outcalt, S., Goodwin, C., and Hinkel, K.: Diurnal Thermal Regime in a Peat-Covered Palsa, Toolik Lake, Alaska, Arctic, 38, 310–315, 1985.

Olefeldt, D., Goswami, S., Grosse, G., Hayes, D., Hugelius, G., Kuhry, P., Mcguire, A. D., Romanovsky, V. E., Sannel, A. B. K., Schuur, E. A. G., and Turetsky, M. R.: Circumpolar distribution and carbon storage of thermokarst landscapes, Nat. Commun., 7, 1–11, https://doi.org/10.1038/ncomms13043, 2016.

Peng, X., Zhang, T., Frauenfeld, O. W., Wang, K., Luo, D., Cao, B., Su, H., Jin, H., and Wu, Q.: Spatiotemporal Changes in Active Layer Thickness under Contemporary and Projected Climate in the Northern Hemisphere, J. Climate, 31, 251–266, https://doi.org/10.1175/JCLI-D-16-0721.1, 2018.

Throop, J., Lewkowicz, A. G., and Smith, S. L.: Climate and ground temperature relations at sites across the continuous and discontinuous permafrost zones, northern Canada, Can. J. Earth Sci., 49, 865–876, https://doi.org/10.1139/e11-075, 2012.

Treat, C., Jones, M. C.: Near-surface permafrost aggradation in Northern Hemisphere peatlands shows regional and global trends during the past 6000 years. Holocene, 28, 998–1010, https://doi.org/10.1177/0959683617752858, 2018.

Treat, C. C., Jones, M. C., Camill, A., Gallego-Sala, A., Garneau, M., Harden, J. W., Hugelius, G., Klein, E. S., Kokfelt, U., Kuhry, P., Loisel, J., Mathijissen, P. J. H., O'Donnell, J. A., Oksanen, P. O., Ronkainen, T. M., Sannel, A. B. K., Talbot, J., Tarnocai, C., and Väliranta, M.: Synthesis dataset of physical ad ecosystem properties from pan-arctic wetland sitesusing peat core analysis, PANGEA, https://doi.org/10.1594/PANGAEA.863697, 2016.

Treat, C; Broothaerts, N., Dalton, A. S., Dommain, R., Douglas, T., Drexler, J., Finkelstein, S. A., Grosse, G., Hope, G., Hutchings, J. A., Jones, M. C., Kleinen, T., Kuhry, P., Lacourse, T., Lähteenoja, O., Loisel, J., Notebaert, B., Payne, R. J., Peteet, D. M., Sannel, A. B. K., Stelling, J., Strauss, J., Swindles, G. T., Talbot, J., Tarnocai, C., Verstraeten, G., Williams, C., J; Xia, Z., Yu, Z., Brovkin, V.: Global dataset of peatland basal ages. PANGAEA, https://doi.org/10.1594/PANGAEA.873065, 2017.

List of relevant changes:

- We revised the manuscript so that our criteria for accepting landform observations to our data is clearer (lines 95–104).
- We clarified the section related to the division of evaluation and calibration datasets (lines 112–119).
- We modified the manuscript thoroughly so that it is clear for the reader that the Olefeldt at el. (2016) dataset represents modelled thermokarst likelihoods instead of observed occurrence of thermokarst landscapes.
- To help the interpretation of our evaluation scores, we revised the paragraph in the section 2.3 Statistical modelling, by adding a couple of sentences explaining the interpretation of AUC and TSS values (lines 231–238). We also revised the related paragraph in the results section (lines 350–361).
- We slightly toned down our conclusion related to the role of thawing season to the degradation of the studied landforms (lines 520–522).
- Our results are now publicly available in raster format in Zenodo (https://doi.org/doi:10.5281/zenodo.7745085). The data availability statement was revised accordingly (lines 749–750).