

Author responses to reviewer comments, and edits to the manuscript “Accelerated lowland thermokarst development revealed by UAS photogrammetric surveys in the Stordalen mire, Abisko, Sweden”

(Paper submitted to ‘The Cryosphere’ | egusphere-2025-3788)

We would like to thank the reviewers and editor for the comments on our manuscript # egusphere-2025-3788. We have paid close attention to the suggested edits and made the requested changes as detailed below, with reviewer's comments in regular font, directly followed by our responses in italic (*blue font*).

Editor decision: Reconsider after major revisions (further review by editor and referees)

Anonymous Referee #1, 18 Aug 2025

(Citation: <https://doi.org/10.5194/egusphere-2025-3788-RC1>)

The authors used UAS surveys to quantify rates of palsa degradation at the Stordalen mire in Sweden from 2019 to 2021, within the context of longer-term change from almost annual UAS surveys from 2014 to 2022. EMI measurements were also collected to characterize the soil properties in stable, degrading, and degraded sections of a study transect.

While the study is interesting, I worry that the findings, as they are currently presented, are not particularly novel. The authors find that topography is important for identifying palsa degradation, that palsa degradation leads to more open water and increased soil moisture, and that palsa degradation rates have increased in recent years. These findings are useful, but they have already been demonstrated in other studies. What is potentially novel, however, is the updated estimate of emissions from Stordalen, and the integration of EMI and UAS field methods. I encourage the authors to really highlight these elements in this manuscript.

We thank the reviewer for this comment. The main idea of our study is to provide new data on the rate of lowland thermokarst development and, at the scale of the Stordalen mire, to show the acceleration of this degradation, with the potential consequences on its carbon balance. The rate of thermokarst development is a critical prerequisite for predictive models of carbon balance in such landscapes, and this rate remains very poorly constrained at present. Our review of the literature identified few local studies, with sometimes much variable figures. We therefore believe that every additional contribution of this type of assessment is very welcome to the scientific community, especially when we place the figures found in our study in perspective with those used in Arctic-scale models such as that of Turetsky et al. (2020).

We confirm the value of including slope and relative elevation as input data for land cover classifications of lowland thermokarst degradation. Such data are still rarely used, particularly slope, but the latter adds ~10% to the overall accuracy of the model. Slope further enables better detection of the early stages of degradation (i.e., it doubles the F-scores of both the areas undergoing degradation and the dynamic classes) with very short time series (i.e., on a biennial scale as opposed to at least 10 years in the previous literature). We have taken great care to emphasize these points in the new version of the manuscript, as well as the links with EMI results, carbon stock calculations, and potential GHG emissions.

GENERAL COMMENTS

This manuscript could be greatly improved by separating the Results and Discussion sections. As it currently reads, this section is quite long, and it contains a mix of reporting of results, reminders of methodology, and discussion and links to literature. I would strongly recommend separating out the Results and Discussion sections and focusing the Discussion primarily on the rate of palsa degradation compared to the literature, the implications for organic carbon

stability and greenhouse gas emissions, and the scaling up of the results. A stronger Discussion would better justify the suitability of this manuscript in a multi-disciplinary journal like The Cryosphere.

Thank you for your comment. We have taken care to minimize methodological reminders in the results and discussion section (particularly sections 3.1 and 3.3). This was facilitated by the production of a new methodological figure presenting the different model runs, as suggested by reviewer 2 (Fig. A. 5; L532).

The main result of our work concerns the acceleration of palsas degradation in Stordalen, which is based primarily on the use of historical data and recent data that we contributed to generating (see Table A. 1), which makes our results and those in the literature very tightly linked. Furthermore, thanks to the new papers you suggested, we have produced a new Fig. 5 that presents a meta-analysis of the degradation rates of palsas. We therefore consider it essential to place our study directly in the context of literature within an integrated results and discussion section.

The EMI results are interesting, but they currently feel a bit out of place with the rest of the study that is more focused on UAS-based rates of change. As above, I would suggest separating out the Results and Discussion and focusing in on the hydrological changes that were identified between stable, degrading, and degraded areas from the EMI surveys and what this means for the permafrost carbon feedback and greenhouse gas emissions. Currently, the manuscript has a separate section for the EMI results that discusses an increase in open water and soil moisture, but it would be more effective if the EMI results were properly integrated with other elements of the study, such as elaborating on what increases in ponds and soil moisture would mean for carbon stocks, emissions, etc.

We acknowledge that the connection between sections 3.2 and 3.3 was lacking. We have primarily ensured that the connection between lowland thermokarst, its consequences on hydrology, redox state, and carbon balance is clearer in the introduction (LL58-64). We have also added a few lines on the hydrological consequences of palsas degradation at the beginning of section 3.3 to further emphasize this link (LL395-397): “Lowland thermokarst development is typically associated with drastic modifications in landscape hydrology, redox conditions, vegetation species composition, organic carbon stability and GHG emissions (e.g., Palace et al., 2018; Patzner et al., 2020; Varner et al., 2022). Such is the case in Stordalen, where, using classification models from the 2014 – 2022 times series (Fig. 6), we observe [...]”.

I find it a bit difficult to follow the flow of the study, as it is currently written. I think it would be worth considering restructuring the Methods section to first describe the data processing for the 2019-2021 model, then the EMI work, then the data processing for the 2014-2022 dataset. This would help to highlight the novel data collection/work (2021 UAS flight, EMI survey, etc.) within the context of a longer study period (2014-2022).

The methodology section has been thoroughly revised (e.g., LL119-127; LL176-182; LL191-196; LL204-210; LL216-217; LL226-228; LL254-262; LL264-270; L280-282); clearer links between the multidisciplinary aspects of the paper, clarifications requested by both reviewers on specific points, syntax, etc.). However, we believe that separating the description of the classifications for the 2014-2022 time series from that of 2019-2021 would be less efficient, as the vast majority of the (pre-)processing steps, the classification algorithm, and the post-processing are identical for both models.

Overall, the figures are nice and the authors have taken care to ensure that the colour schemes are accessible. The text formatting in some of the tables may need to be reviewed, as there are some terms that are capitalized and others that are not.

We thank you for this comment. We took care to make the formatting of the tables and capital letters uniform.

SPECIFIC COMMENTS

INTRODUCTION

P2 L53, Remove “excess” from this sentence.

This has been removed (L50).

P2 L57, There is a new paper that has just come out on the use of the term “abrupt thaw” by Webb et al. 2025 that can be used to replace Turetsky et al. 2020.

Thank you very much for this reference, that we added to the manuscript (L54). Please note that we do not further mention the term ‘abrupt thaw’ in the new version of the manuscript, following comments from reviewer 2.

P2-3 L53-78, This background information on thermokarst landform types and development is interesting, but it takes away from the purpose of the study itself, which is to quantify palsa degradation in the Stordalen mire using UAS surveys. The Introduction could be improved by introducing palsas as peatland permafrost landforms, discussing the importance of peatland permafrost landscapes for permafrost carbon feedbacks, and then diving into the benefits of UAS imagery over satellite imagery and aerial photographs.

The paragraph describing the different types of thermokarst degradations has been substantially reduced and now focuses on lowland organic landscapes and peatlands (LL73-84). We have also included information on the morphology of palsa mires and the importance of peatland permafrost landscapes for permafrost carbon feedbacks (LL85-95).

P3 L90-92, I agree that reported rates of degradation are extremely variable, but it would be helpful to highlight to the reader what area or approximate time period these studies are from. In P3 L95-97, the authors state that there is accelerated degradation in more recent years, but it is difficult to understand this relative to the previous statement that does not provide a time reference/study period.

We have amended the sentence to now read “A few studies suggest an accelerated rate of degradation in more recent years (i.e., from the 1950s to the last decade; de la Barrera-Bautista et al., 2022; Borge et al., 2017; Olvmo et al., 2020), but still require at least ten years of survey data to enable quantification of palsa loss rates.” (LL101-103).

P3 L93-94, Wang et al. 2024, Verdonen et al. 2023, Zuidhoff and Kolstrup 2000, Thie 1974, Payette et al. 2004 are some other studies that also present lateral palsa degradation rates. These may be helpful for further contextualizing the results of this study on P15, L359-363.

*Thank you very much for providing these new references, which we have included in the new version of the paper, and particularly in **Fig. 5** which presents a literature-based meta-analysis of the rates of palsa degradation.*

P4 L98, What is meant here by “revisit”? This is the first mention of the Stordalen mire, and the authors do not provide examples of previous studies of degradation at the Stordalen mire, other than to say that 55% of Sweden’s largest palsa peatlands are currently subsiding in the previous paragraph. Please clarify.

We agree that the term “revisit” caused unnecessary confusion here and have replaced it with “provide” (L104). We retained the verb “update” in the discussion after presenting the rates of palsa degradation at Stordalen found in previous literature (L365).

P4 L103-105, Please provide what years the EMI surveys were conducted.

The EMI data were only collected on a single date, namely September 22, 2021. This has been added to the manuscript, in the methods section (LL282).

METHODS

P4 L109, Section 2.1 is lacking information on the climatic conditions over the study period, from 2019-2021 for the primary part of the study, and from 2014 to 2022 for the additional UAS data that was used. This would be critical for contextualizing palsa degradation.

See response to the next comment.

P4 L113, Given that the study is primarily conducted from 2019 to 2021, or even from 2014 to 2022, is there a more recent value for MAAT since 2006? Please update.

Thank you for these suggestions. We have added a figure (Fig. A. 1) showing soil temperature data (four profiles with depths of 2, 5, 10, 30, and 50 cm) and air temperature data (at 2.5 m) between 2014 and 2022 (Fig. A. 1a), as well as the 1-year moving average (Fig. A. 1b) and trends ($^{\circ}\text{C}/\text{year}$, p -values $< 10^{-3}$; Fig. A. 1c) from these measurements. This shows that soil temperature increased significantly between 2014 and 2022 in 18 out of 20 cases, up to a maximum of $0.18^{\circ}\text{C}/\text{year}$. In 2 out of 20 cases, the soil temperature decreased by $\sim 0.01^{\circ}\text{C}$. However, the air temperature decreased significantly over the same period, by 0.17°C . We have also added the following paragraph (LL119-127):

“More recently, half-hourly ICOS Sweden data shows that the mean annual air temperature for the period 2014 – 2022 at the Stordalen mire is $0.8 \pm 9.2^{\circ}\text{C}$ (mean $\pm$ standard deviation), while the mean soil temperature is $2 \pm 7^{\circ}\text{C}$ at 2 cm and $0 \pm 2^{\circ}\text{C}$ at 50 cm over the same period (Fig. A. 1a; ICOS Sweden, 2023). Annual moving averages derived from this same dataset (Fig. A. 1b) reveal a statistically significant warming trend in soil temperature at 18 of the 20 probes, with rates of up to $0.18^{\circ}\text{C}\cdot\text{a}^{-1}$ (p -values $< 10^{-3}$; Fig. A. 1c). Conversely, 2 probes exhibited slightly cooling trends of approximately $0.01^{\circ}\text{C}\cdot\text{a}^{-1}$ (p -values $< 10^{-3}$; Fig. A. 1c). In contrast to the widespread soil warming, air temperature has decreased by $0.17^{\circ}\text{C}\cdot\text{a}^{-1}$ for the period 2014 – 2022 (p -value $< 10^{-3}$; ICOS Sweden, 2023). The mean annual accumulated precipitation is 332 mm (data period 1981-2010), which is 10% more compared to the 1961-1990 normal (Abisko-Stordalen Palsa Bog | ICOS Sweden, 2025).”

P4 L120, Zuidhoff and Kolstrup 2005 and Railton and Sparling 1973 also discuss vegetation associated with different palsa stages.

Thank you for the references. However Zuidhoff and Kolstrup. (2005). Arct. Antarct. Alp. Res. 37 (1) & Railton and Sparling. (1973). Can. J. Bot. 51: 1037–1044 do not cover the Stordalen mire. We therefore added these references to the updated description on palsa mire and peat plateaus (LL89-92; see above comment and response).

P4 L123, Is there any available information, either from this study or from previous studies, on the height of the palsas and the thickness of the permafrost at this site? It is helpful to know that the active layer thickness varies from 50 cm in stable areas to >200 cm in degraded areas, but is it possible that a talik has formed and that there is still permafrost present at depth?

The heights of the palsas are provided through the digital surface models, i.e., Fig. 2c, Fig. A. 3 (previously Fig. A. 2), Fig. A. 4g-h (previously Fig. A. 3); There are $\sim 25 - 50$ cm high (Fig. A. 4g-h). As for permafrost deeper than 200 cm, such information does not exist to our knowledge. Yet, in the degraded areas, the surface is no longer considered active layer, i.e., they show positive temperatures year-round (see [Data portal | ICOS](#))

P5 L134, The authors state here that the field campaign took place between September 14 and October 10, 2021, but that the UAS flight took place on September 17, 2021. What else occurred during this time period? When were the EMI surveys conducted?

The mission covered a wide variety of fieldwork, which is being addressed in other studies. Mentioning the total duration of the mission is indeed irrelevant, and we have removed this sentence.

P5 L137, Thanks for providing the forward overlap. What was the side overlap?

The side overlap is 63%. It has been added to the manuscript (L151).

P6 L134, Please specify that this is RGB imagery collected from UAS. While this is clear when looking at Table A 1, this should be included in the main text as well.

We believe this is already the case for this section: “To obtain data on surface properties, i.e., vegetation, elevation and ground subsidence, we generated a RGB orthomosaic and a digital surface model of the study site by UAS photogrammetry [...]” (LL149-150).

P6 L161, I think it would be best to present this information in paragraph form and to explain each of the steps and what datasets were used in each step. For example, stating that the slopes were extracted from DSMs where applicable is quite vague, and the reader is likely unsure of what is and what isn't applicable. Is this trying to convey that slopes were extracted from DSMs for 2019 and 2021, but not for the other years? And how was the area of interest extracted? Was it clipped?

The slopes were extracted from 2019 and 2021 only, as we do not have digital surface models for the other years. The areas of interest were extracted using ArcGIS ‘clip’ tool. This information has been added to the manuscript, and the bullet points have been removed.

P7 Figure 2, Remove the extra “t” in “literature” in the caption for panel a. The grey and yellow bounding boxes are very similar in colour and are a bit difficult to differentiate.

We have replaced the grey color by dark blue and corrected the spelling mistake.

P7 L183-184, Are there any historical aerial photographs or satellite images that can help to confirm that permafrost was not present in these locations for several decades?

We were cautious in stating that “[...] that may not have had permafrost conditions for several decades” (LL195-196). We cannot be certain, but past studies of the site show that part of it was already subsided/degraded in 1970 (e.g., Varner et al., 2022).

P7 L189, Are there any locations at all where permafrost aggradation and palsa expansion occurred? Having a section that describes climatic conditions from 2014 to 2022 as suggested above would be helpful for this.

We have revised the section on climatic conditions that demonstrate that soil temperature has been increasing significantly from 2014 and 2022 (LL119-127). We therefore consider it highly unlikely that there will be further aggradation of the permafrost in Stordalen. Renette et al. (2024) surveyed seasonal dynamics of palsas ~100 km northeast of Stordalen and found no evidence of aggradation from one year to the next, as did other studies in Scandinavia that we have extensively referenced (e.g., Borge et al., 2017; Olvmo et al., 2020; Verdonen et al., 2023).

P7-8 L190-222, As with P6, I think that it would be best to present much of this information in paragraph form. This could be supported by a figure or table that explains the process more visually and that possibly integrates information from Table A2 and Figure A3.

*We have rewritten this in paragraph form and added **Fig. A. 5** showing which data was used for which model run (LL200ff).*

P11 L257, Hypotheses are usually presented in the Introduction, not the Methods section. Please move this up to the Introduction and provide more information on how the authors expect the electrical properties of the soil to vary along the degradation gradient. Should the EC be higher or lower according to the factors presented (soil texture, clay content, water content, salinity, organic matter type, organic matter proportion, soil structure, soil density, soil temperature, and most importantly, permafrost presence/absence!). Instead, in this section, please focus on describing how the EMI surveys were positioned, how long they were, etc. It is helpful to know that there were 1083 points, but the reader is not informed of how far the points are from each other, whether they are all along the same line, etc.

We have moved the hypotheses up to the introduction and we added the expected variations of the soil EC along the gradient (L64-67). Soil EC should be higher with clay content, water content, salinity, organic matter proportion, soil

*density and soil temperature, and permafrost absence since there is more water content in degraded areas. Besides, the exact locations of the 1083 points are presented on **Fig. 7A** (previously Fig. 6A).*

P11 L271, Is there a reference or any more information available for this custom-made acquisition program?

As the EM38 does not provide built-in automatic acquisition system, our custom-made acquisition program simply records the GNSS positions and the corresponding EMI values (approximately one measurement every second) into text files to be processed.

RESULTS AND DISCUSSION

P13 Figure 4, This figure is very effective, particularly panel b! I would recommend changing the light blue colour of the “degraded areas” in panel a to another colour, as this looks like water at first glance.

We thank you for this comment. The color code was carefully chosen to enable easy interpretation of the plots by people suffering from color vision deficiencies. We therefore believe it inadvisable to change this color.

P14 L344-355, This is the first instance where the reader can really come to understand the authors’ “revisit” of palsa degradation rates in the Stordalen mire. These past studies should be first presented in Section 2.1 so that the reader is able to keep this information in mind as they read through the results of this study.

Please see previous response regarding this point: We now retain only the verb “update” in the discussion after presenting the rates of palsa degradation at Stordalen found in previous literature.

P15 L379, The section entitled “Palsa degradation means higher levels of humidity” does not really discuss humidity levels at all. It may be more appropriate to instead name the section something like “Palsa degradation leads to increases in soil moisture and open water”.

We have changed the section title to: “Palsa degradation leads to increases in areas with high soil moisture and open water” (L394).

P16 Figure 5, I understand that data could not be collected in 2020, so the corresponding bar has a dashed outline. But if data could not be collected in 2020, how is there a bar and a value associated with this year at all?

We agree that this caused unnecessary confusion and removed the bar with a dashed outline (L406).

P15 L385-386, The authors state here and show in Figure 2 that the processing extents for the 2014-2022 comparison and the 2019-2021 comparison are not the same. In addition to the work that has been done, is it possible to clip the results of the 2019-2021 comparison to the 2014-2022 comparison extent, so that the authors can additionally present results that are directly comparable?

*The years 2019 and 2021 are included in the 2014-2022 model, so the extent of the latter and the results you request are already included in **Fig. 6a** (previously Fig. 5a). We simply draw the reader’s attention to the fact that the rates of palsa degradation will depend on the delineation chosen, and that this varies between the 2019-2021 model and the 2014-2022 model.*

CONCLUSION

P20 L481-485, This is a helpful summary of findings that integrates the EMI and 2014-2022 work well.

We thank you for this comment.

Anonymous Referee #2, 20 Dec 2025**(Citation: <https://doi.org/10.5194/egusphere-2025-3788-RC2>)****General Comments:**

Thomas et al. looked at palsa degradation in the Stordalen mire near Abisko, Sweden. To quantify the degradation, they derived land cover changes from UAS orthomosaics and digital surface models between i) 2019 and 2021, and ii) 2014-2022. For the classification of land cover and land cover change, they combined different image processing techniques of the input data and trained a support vector machine to achieve a maximum overall accuracy of 81.3 %.

The authors identify topography information as vital model input and observe an increase of soil moisture and inundated areas throughout the mire.

Concerning the clarity and content of the study, I have the following concerns:

Methods:

- In general, this section lacks sufficient explanation and referencing regarding the choice and implementation of algorithms or techniques, eg. GLCM, RBF, the mean permutation analysis, how the refinement of initial predictions was done using the confusion matrices. These approaches should be clearly described and supported with references. Further, you mention for the first time in the results section that several model runs were performed with different inputs. It would help readers if this aspect was already introduced and described in the methods section—perhaps including a small workflow figure to illustrate the process. I included some specific comments on this further down.

The addition of texture attributes from the gray-level co-occurrence matrix (GLCM) is commonly applied in land cover classifications to provide textural metrics as a proxy of the morphological information and spatial patterns. In Stordalen's case, for example, this was used in Palace et al. (2018), which we reference extensively. In this study however, GLCM texture attributes contribute little to our model, hypothetically because morphological information is available through topography.

The radial basis function (RBF) kernel was adopted for the SVM classifier because it outperformed linear and polynomial kernels in our preliminary tests. RBF kernels can model non-linear decision boundaries without requiring prior assumptions about the form or degree of interactions between variables. Like other SVM kernels, it is known to manage a large number of dimensions.

*We used scikit-learn's built-in function to compute permutation importance. It involves randomly shuffling the values of the input data and observing the resulting degradation of the model's score. We modified the legend of the y-axis of **Fig. C. 1a** to make this clearer (LL547-549), while the figure caption now states "(a) Mean permutation importance (scikit-learn built-in function) of each input data in the model, measured as the decrease in accuracy when the input data values are randomly shuffled. The error bars indicate the standard deviation of importance across 10 permutations".*

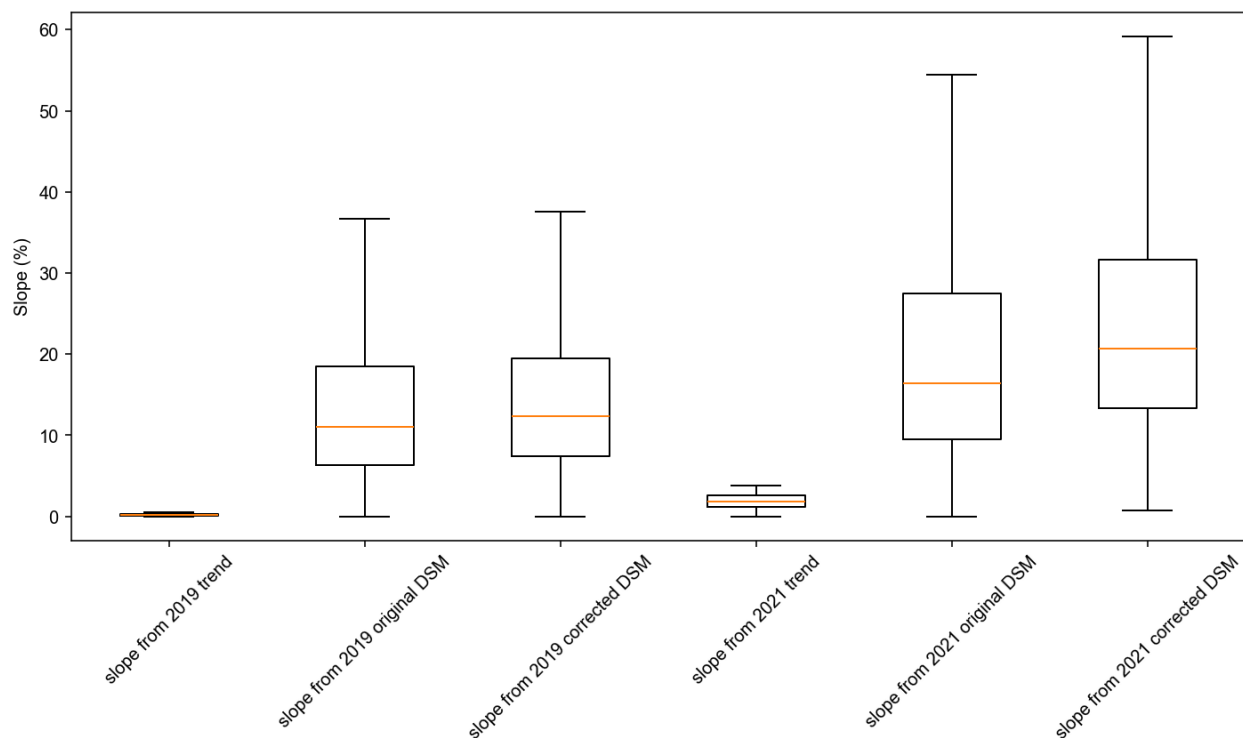
*The methodology concerning the refinement of the initial predictions is explained extensively in **section 2.4** (LL218-226): "As a first post-processing step, we refined the initial predictions with two rules involving the predicted class and the relative elevation difference between 2019 and 2021. Firstly, if a pixel was predicted as belonging to one of the dynamic classes (i.e., representing degradation between 2019 and 2021) and the corresponding relative elevation difference was positive or null (which therefore does not indicate subsidence), the model re-evaluated the prediction. In this case, the possible identification was limited to one of the static classes that remain unchanged between 2019 and 2021. Conversely, if a pixel was initially assigned to the stable palsa class, and the value of the relative elevation difference was lower than -30 cm (indicating significant subsidence), the model reclassified the pixel by selecting the likeliest identification from the dynamic classes. The reclassification process involved examining the decision scores*

produced by the model for each sample and selecting the class with the highest decision score from the set of allowed classes.” As mentioned in section 3.1 and supported by **Fig. C. 1b**, the marginal gain from this procedure is relatively limited, i.e., it increases the overall accuracy from 80.8% to 81.3%. This therefore does not affect many pixels (0,14 % of the pixels, for reference).

Finally, we have added **Fig. A. 5**, which shows the different model runs and associated input datasets, clarifying this point.

- You state that you calculated the slope using an “uncorrected” DSM, but wouldn’t the bowl effect overestimate the slope?

We analyzed the distribution of slope data and found that it is very similar between the slope calculation before and after DSM correction, and the calculated slopes are in all cases much higher than the slopes of the trends (see below)



Results:

It is somehow difficult to understand some of your results as numbers change throughout the text or are introduced without explanation, e.g.

- in Section 3, you present a value of 24 metric tons of TOC, including 6 metric tons of MAOC for your study site, but under Section 4 you refer to 12 metric tons.

We understand the possible confusion here. 12 metric tons of TOC refer to the annual mass of TOC while 24 metric tons refer to the 2-year balance assessment, i.e., 2019-2021. We have uniformized the numbers and wording to avoid any confusion (L35; L448; L501).

- The abstract reports an overall accuracy of the model of 41-77%, yet you achieved an OA of roughly 82 % after the refinement and 80 % without refinement but with the std. spatial filter. Did you intentionally select the land cover map from the less accurate (77%) model configuration, and if so, why? This ties into the incomplete description of model runs in the methods section, which makes it unclear which input configuration performed

best. Consider describing and naming the different runs clearly and then specifying which one yielded the best result.

We did all the calculations of the surface areas using the best-performing version of the model, i.e., the overall accuracy of 83%. The percentages presented in the abstract were there to indicate the increase in performance when elevation & slope were added. We changed this in the abstract, to now read "[...] that topography is an essential parameter for monitoring the evolution of palsa degradation, almost doubling the overall accuracy of the classification [...]" (L29). We have also followed your suggestion to clarify the model runs and name them in the new version of the manuscript.

- I am also somewhat unclear about how the CH₄ measurements (and their increase) under 3.4 are calculated.

*This is a weighted average of emissions reported in Łakomiec et al. (2021) by the respective surface areas of palsas and degraded areas in 2019 and 2021 (see **Tab. B. 2** for the surfaces). There was a calculation error in the previous version of the paper, and we apologize for this.*

Writing style:

- The manuscript would benefit from using the active rather than the passive voice, as this makes it easier for readers to distinguish between the work conducted by you and the data obtained from others. This distinction is especially important since both your own drone acquisitions and data from other sources are used.

*We took care to limit the use of passive voice whenever possible in the new version of the manuscript. We have significantly reduced its use in **section 2.4** of the methodology, which has been completely reworked.*

- Further, I advise to ensure consistency in the terminology and in the way processed data are referred to throughout the text (e.g. F1/ F score, thermokarst vs. abrupt thaw, DEM vs. DSM). In addition, many sentences are quite wordy and nested, which makes the text difficult to follow at times. A general English language editing would improve readability. In some cases, I included suggestions in the specific comments below.

The text has been reworked to standardize terminology, and we have included your suggestions in the specific comments.

Specific comments:

Abstract, p. 1: The abstract in its current form is too long, the results alone occupy almost 180 words. Please consider shortening it to focus on your major outcomes.

We have reduced the length of the abstract to 383 words (compared to 433 in the previous version):

"The estimation of greenhouse gas (GHG) emissions from permafrost soils is challenging, as the decomposition of organic matter depends on factors such as soil pH, temperature, and redox conditions. Over lowland permafrost soils, these conditions are directly related to the microtopography and evolve with physical degradation, i.e., lowland thermokarst development. A dynamic quantification of lowland thermokarst development – still poorly constrained – is therefore a critical prerequisite for predictive models of permafrost carbon balance in these areas. This requires high-resolution mapping, as lowland thermokarst development induces fine-scale spatial variability (~50 – 100 cm). Here we provide such a quantification, updated for the Stordalen mire in Abisko, Sweden (68°21'20"N 19°02'38"E), which displays a gradient from well-drained stable palsas to inundated fens, which have undergone ground subsidence. We produced RGB orthomosaics and digital surface models from very high resolution (10 cm) unoccupied aircraft system (UAS) photogrammetry as well as a spatially continuous map of soil electrical conductivity (EC) based on electromagnetic induction (EMI) measurements. We classified the land cover following the degradation gradient using a support vector machine algorithm and derived palsa loss rates. Our findings confirm that topography is an essential parameter for monitoring the evolution of palsa degradation, almost doubling the overall accuracy of the classification, with the addition of slope allowing the detection of the early stages of degradation. We show a clear acceleration of degradation for the period 2019 – 2021, with a decrease in palsa area of $3.3 - 3.6\% \cdot a^{-1}$ (% reduction per year relative to the initial palsa areal extent) compared to previous estimates of $\sim 0.3\% \cdot a^{-1}$ (1970 – 2000) and $\sim 0.1\% \cdot a^{-1}$ (2000 – 2014). EMI data show that this degradation leads to an increase in soil moisture, which in turn likely decreases organic carbon geochemical stability and potentially increases methane emissions. With a palsa loss of $3.3 - 3.6\% \cdot a^{-1}$, we estimate accordingly that surface degradation at Stordalen might lead to a pool of 12 metric tons of organic carbon exposed annually for the topsoil (23 cm depth), of which ~25% is mineral-interacting organic carbon. As topography changes due to lowland thermokarst are fine-scaled, circumpolar up-scaling assessments are challenging. By extending the monitoring we have conducted as part of this study to other lowland areas, it would be possible to assess the spatial variability of palsa degradation/thermokarst formation rates and thus improve estimates of net ecosystem carbon dynamics."

Please note that in the new version of the manuscript, all palsa loss rates presented are in comparison with initial palsa areal extent and calculated using eq. 5. This ensures consistency in palsa loss rates reported throughout the paper.

- L 19: "propensity", p. 1: I only know "to have a propensity to do sth." or "to have a propensity for sth." - maybe clarify with a native speaker? An alternative might be "as the decomposition of organic matter depends on ..."

We have modified the sentence based on your suggestion (LL18-19).

- L 24: "for the Stordalen mire in Abisko, Sweden for the Stordalen mire, Abisko, Sweden", p. 1: This is mentioned twice.

We have removed the repetition (L24).

- L 26: "digital elevation models", p. 1: Throughout your manuscript, you say "digital surface models". I would suggest to stay consistent in order to avoid confusion.

We have harmonized the use of "digital surface models" (L26).

- L 27: “map”, p. 1: maps.

We have only produced one map of soil electrical conductivity.

- L 30: “enhancing the overall accuracy of the classification from 41% to 77%”, p. 1: You should mention before that you classify the land cover using a SVM model in order to refer to it. Further, in the results you say that you achieved a maximum OA of 81.3 %.

We now mention the use of SVM in the abstract (L28). Furthermore, the increase from 41% to 77% only concerns the addition of elevation data to the OA. The rest of the data processing is not covered by this number. To avoid any confusion for the reader, we changed for “almost doubling” instead (L29).

L 45: “near-surface permafrost area”, p. 2: do you mean “area underlain by permafrost”?

We have changed the text to now read “[...] we are consequently anticipating the area of near-surface permafrost to decrease by [...]” (L42).

L 45: “Intergovernmental Panel on Climate Change (IPCC): scenario RCP2.6”, p. 2: I think you can leave out that the RCPs were adopted by the IPCC, this is more or less common knowledge. You should, however, once state that RCP stands for Representative Concentration Pathways, e.g. for the Representative Concentration Pathways (RCPs) 2.6 and 30 – 99% for RCP 8.5.

We have changed the text according to your suggestion (LL42-43).

L 50: “projected thaw”, p. 2: “projected thaw of permafrost soils”

We have changed the text to “[...] with a projected thaw of permafrost peatlands [...]” (L48).

L 51: “~1% of anthropogenic radiative forcing”, p. 2: Can you give a value for this?

The abstract of Hugelius et al. (2020) states “The projected thaw would cause peatland greenhouse gas emissions equal to ~1% of anthropogenic radiative forcing in this century. The main forcing is from methane emissions (0.7 to 3 Pg cumulative CH₄-C) with smaller carbon dioxide forcing (1 to 2 Pg CO₂-C) and minor nitrous oxide losses.”. We believe that providing these values in our paper is not appropriate, so as not to overload the introduction.

L 54: “arctic”, p. 2: Arctic

We have changed the text accordingly (L56).

L 56: “refer to”, p. 2: Do you mean “observe”?

We rather mean “[...] reference is often made to [...]” or “[...] there is often mention of [...]”, i.e., from a semantic point of view (L54).

L 57: “abrupt thaw”, p. 2: I assume here you mean the erosion and mass movement processes triggered by abrupt thaw, such as thaw slumps?

The terms “abrupt thaw” and “thermokarst development” are often used interchangeably in the literature (e.g., Turetsky et al., 2020), even though “abrupt thaw” can sometimes be misleading (see Webb et al., 2025). The term ‘abrupt thaw’ has been removed in the new version of the manuscript, and we rather use “lowland thermokarst degradation” for Stordalen, as it is more precise.

L 59: “cycles.”, p. 2: please add references here.

The sentence now reads “[...] which refers to physical degradations of the landscape that occur when the ground subsides or collapses with significant consequences for hydrologic and biogeochemical cycles (e.g., Turetsky et al., 2020; Varner et al., 2022; Vonk et al., 2015).” (LL54-55)

L 59: “IPCC”, p. 2: As far as I know the RCPs were developed by the research community, and only adapted by the IPCC.

We changed the sentence to now read “[...] is not included in GHG emissions models used by the IPCC, yet they [...]” (LL56-57)

L 61: “below”, p. 2: “below air temperatures of ...”

We have changed the text accordingly (L58).

L 62: “propensity”, p. 2: See comment in abstract.

We have changed the text accordingly (L59), as in the abstract (LL18-19), following your suggestion.

L 64: “the latter of which may change drastically following certain thermokarst developments”, p. 2: Could you briefly explain why that is? Do you have a reference for this?

Here we are referring to the shift in redox potential from oxidizing to reducing conditions following thermokarst development, e.g., excess ice melt leading to the creation of e.g., thaw lakes and wetlands. The new sentence is “Redox conditions may change drastically following thermokarst developments, such as from oxidizing to reducing conditions after excess ice melting and wetland formation (e.g., Turetsky et al., 2020; Varner et al., 2022; Vonk et al., 2015 and references therein).” (LL61-64)

L 66-67, p. 2: This sentence reads a bit confusing. Turetsky states that GHG emissions across 2.5 mio km² of abrupt-thaw affected land could have a comparable feedback as land affected by gradual thaw. I would not use abrupt thaw and thermokarst synonymously as thermokarst can be gradual and slow. Further, the number you are referencing for gradual thaw is for the entire permafrost region (18 mio km²), underlining the need to incorporate abrupt-thaw processes and their GHG emission in these models.

Thank you for your comment. In Turetsky's paper, “abrupt thaw” is used to refer to processes such as that observed in Stordalen, although we now use “lowland thermokarst” throughout the paper.

We have adapted the end of the sentence to mention that we are comparing this to the gradual thaw of the entire permafrost region (L70).

L 69: “increasingly negative”, p. 3: What becomes “increasingly negative”?

We refer to NECB estimates. We split the sentence in two for better readability, to now read “These assessments are based on cumulative net ecosystem carbon balance (NECB) estimates from different models of thermokarst landscape succession, which show increasingly negative balances directly influenced by the rate of permafrost degradation (Bosiö et al., 2012; Turetsky et al., 2020).”(LL70-73)

L 71: “The approach”, p. 3: Which approach? Are you still referring to Turetsky?

We do not refer specifically to Turetsky et al. here, as they did not estimate the rate of development of thermokarst landscapes. We have changed the term to ‘methodology’ (L73).

L 90-91: "Reported rates of degradation are extremely variable, i.e., range from $0.04\% \cdot a^{-1}$ to $0.7\% \cdot a^{-1}$ 90 (% reduction per year) of total land cover area", p. 3: Reported rates of degradation, quantified by changes in total land cover, range from ..."

We have changed the sentence to now read "Reported rates of degradation, quantified by comparison with initial palsa areal extent, are extremely variable, i.e., range from $0.1\% \cdot a^{-1}$ to $9\% \cdot a^{-1}$ (% reduction per year) [...]" (LL95-96).

Please note that in the new version of the manuscript, all palsa loss rates presented are in comparison with initial palsa areal extent and calculated using eq. 5. This ensures consistency in palsa loss rates reported throughout the paper.

L 98: Why revisit? Is there a study preceding this one, the reader should be aware of?

Reviewer 1 have also raised this point and we recognize that this was confusing. We have replaced the term "revisit" with "provide" here (L104), since our study is the only one to quantify palsa loss rates at Stordalen for recent years. We retained the verb "update" in the discussion after presenting the rates of palsa degradation at Stordalen found in previous literature (L365).

L 99: "obtained from color imagery/RGB orthomosaics", p. 4: "obtained from RGB orthomosaics"

We have changed the text accordingly (LL105-106).

L 103: "open water areas", p. 4: Which open waters are you referring to here (lakes, thermokarst ponds ?) and why do you do this? Is the data for 2014-19 and 2021-22 derived from satellites? Please add 1-2 explanatory sentences.

*We refer to small areas where the water table reaches the surface and represent a sort of end-member of degraded areas in the form of open water ponds (see bottom left of the photo on **Fig. 1b**). We have provided a morphological description of palsa mires following a request from reviewer 1, which clarifies this point (LL85-92). We have the opportunity to quantify the evolution of the surface area of these open water ponds for the 2014-2022 model, which is not possible in the 2019-2021 biennial model (see footnote on page 7).*

However, we have never used satellite data in the paper and only mention satellite when referring to other studies.

L 105: "classification", p. 4: This is quiet suddenly introduced. It might help to mention further up in the paragraph that you quantify palsa degradation rates from UAV imagery by using a model to predict the land cover and change thereof.

We have amended a previous sentence to now read "We combine surface properties (obtained from RGB orthomosaics) with information on micro-topography (relative elevation and slope obtained from digital surface models) to classify the land cover and thereby track palsa degradation over a two-year period." (LL105-107)

L 122: "measurements within an extent of less than 10 m^2 ", p. 4: Did you do these measurements or are they taken from another study?

Yes we did those measurements ourselves. We have added the date to the text, which clarifies this (L140).

L 124: "based on METER TEROS 12 probes measurements:", p. 4: I am unfamiliar with the term TEROS 12. I suggest referring to this device as "soil moisture sensor TEROS 12..." to clarify. Could you please also clarify, if you did these measurements yourself or if they are taken from another study? From which month and year are the measurements?

We also performed those measurements ourselves, on September 30, 2021. The new sentence now reads "[...] based on measurements from METER TEROS 12 soil moisture sensors taken on September 30, 2021" (L140).

L 137: "forward", p. 5: What was the side overlap? And where these flight parameters identical to the 2019 flight?

We have added the side overlap information of the 2021 flight (63%; L151). In 2019, side overlap was 60% and front was 80%.

L 144: “evolution of the physical degradation”, p. 6: This term appears many times throughout the text and is quite wordy. Might it be an alternative to use “land surface degradation”, or to be more specifically focused on your study “palsa degradation”?

We thank you for your suggestion. We have changed the wording accordingly (L158).

L 151: “Together, the 2019 and 2021 datasets provide complementary topographic information and adequate coverage”, p. 6: This reads like the UAS covered different areas of the mire in 2019 than in 2021.

This is indeed the case; the 2019 coverage is slightly larger than the 2021 coverage in the raw data, which is why we mentioned above that we chose the delineation representing the best compromise between the area covered by our UAS data and the polygon initially drawn by Christensen et al. (2004).

L 154 - L 160, p. 6: This paragraph reads a bit confusing. From the table A 1, I would assume that you used RGB images of the same mire originally acquired for different studies to complement your time series. However, here you refer to a dataset that delivers estimates of degradation trends. Could you please clarify what you mean?

We only have topography data for two years, namely 2019 and 2021. Topography data enables robust quantification of palsa degradation, which is why we drew the main conclusions of the study based on the evolution from 2019 to 2021. However, RGB imagery data spanning from 2014 to 2022 also exist, and can be used to run land cover classifications and thereby also track palsa degradation. However, since data on surface elevation is missing, we exercise caution and refer to “preliminary estimate of degradation trends from 2014 through 2022” instead (L171).

L 162: “The various pre-processing operations carried out on the data are detailed below.”, p. 6: I suggest to clearly state that this is the pre-processing of the UAS data to avoid confusion. Personally, I prefer full sentences, explaining processing steps as they are more reader-friendly. If you prefer bullet points, please make sure to add a little more detail on the individual steps, see comments below.

We have modified the text, which has now no bullet point and we added relevant details on the preprocessing steps (LL175ff).

L 165: “Co-registration”, p. 6: Which co-registration algorithm did you use? And was this also done for the complementary RGB images (2014-19 and 2022)

We used the georeferencing tool in ArcMap with affine transformation with 6 reference points and for all years. This has been added to the manuscript (LL178-179).

L 168: “at”, p. 6: to

We have changed the sentence accordingly (L180).

L 169- 170: “Removal of the bowl-shape effect from elevation data (DSM), to eliminate systematic distortions or artifacts caused by sensor or data processing errors, ensuring accurate inter-annual comparison (Fig. A. 21)”, p. 6: Shouldn't this step come before deriving the slope from the DSMs?

We have responded to this comment above, i.e., in the ‘general comments’ section.

L 190: “Briefly, the classification process involves”, p. 7: see comment on data pre-processing

This paragraph has been revised; we have removed the bullet points (LL202ff).

L 192: “times the standard deviation”, p. 7: Could you provide a reference for this approach or briefly explain why you did this?

Digital surface models obtained by photogrammetry can contain a few outliers/spikes for non-cooperative surfaces (e.g., reflective water, vegetation). This produces artifacts that we wanted to minimize through a statistical outlier filtering.

L 195-199, p. 8: So are these mean and standard deviation calculations applied on a stack of RGB, DSM, and slope ("original bands": or on the individual maps? Please state, that you did this to normalize the data. This will make it more accessible to readers who are not that familiar with image processing pipelines.

We did this in an attempt to reduce the sensitivity of the model to artifacts or very small local heterogeneities. This slightly improves the final classification. Because of the very high spatial resolution of the image, looking at a single pixel value is often insufficient because its color alone is not representative of the class. For example, pixels in the shadow are much darker than pixels in the light, even for the same vegetation type. Taking average values in different moving windows allows us to smooth the signal and obtain "colors" that represent the context of the pixels in a way that is less sensitive to the structure of the vegetation. On the other hand, the standard deviation highlights areas where neighboring pixels are of different colors versus area where the color of the patch is homogeneous. These two types of information are then used as pseudo-bands for the classifier, adding information about the context of the pixel. We did not do this to normalize the data.

L 201: "The calculated texture attributes include entropy, angular second momentum, contrast, homogeneity and the standard deviation of the GLCM, all applied within a 21×21 moving window.", p. 8: What are you calculating these for?

The addition of texture attributes from the gray-level co-occurrence matrix (GLCM) is quite commonly applied in land cover classifications to provide additional morphological information and spatial patterns. In Stordalen's case, for example, this was used in Palace et al. (2018), which we reference extensively. In this study however, GLCM texture attributes contribute little to our model, hypothetically because morphological information is available through topography (surface elevation and slope).

L 206: "radial basis function (RBF) as kernel and the gamma parameter set as automatic", p. 8: Why did you set them like that?

We have responded to this comment above, i.e., in the 'general comments' section.

L 208: "data distributions for the training areas are shown in Fig. A. 3.", p. 8: Here, I would be interested in some more explanation. What does that figure tell me? Is this already a result or are you simply underlining that you sampled your training points equally across classes and bands? For 2019, it looks like your training areas cover very similar pixel value ranges (ca. 130-160: for RGB bands and in 2021 this looks quite heterogeneous. Do you discuss this somewhere?

This figure is included, along with Tab. A.2, to provide complete transparency regarding the training dataset used for the 2019-2021 model. It allows us to see directly that morphological parameters (panels g to k) are more discriminating than colors (panels a to f). The differences you mention are not discussed further since the results clearly show that RGB bands are not the most crucial for the model.

L 210: "Refinement of the initial predictions.", p. 8: This is interesting. Can you recall for how many pixels this was the case?

We have responded to this comment above, i.e., in the 'general comments' section.

L 212: "the evolution of the", p. 8: I suggest "representing degradation between...".

We have changed the sentence accordingly (L220).

L 220: "increasing window sizes", p. 8: starting from?

We did this starting from 3×3 windows, to 11×11 windows. This has been added to the manuscript (L227).

L 228-229: "To identify representative classes for the evolution of the degradation between 2019 and 2021, we therefore scanned the different data for the two years of measurement and observed the changes in morphology.", p. 9: Could you please clarify this sentence? Do you mean "representative examples"?

Yes. As with the static classes, we conducted a detailed examination of the available imagery and topographic data and selected representative areas that best illustrate the observed dynamic classes, corresponding to the clearest field-based examples. We changed the sentence to "To identify representative examples for the [...]" (L233).

L 24, equation 1:, p. 9: The overall accuracy is calculated dividing the number of TPs and TNs by the total number of samples. If you only used TPs, you underestimated your model performance.

We believe that your comment applies to a model with only two classes. With a multi-class model such as the one presented here, the true negative values of one class are taken into account in the sum of the true positive values of all the other classes. Here, $\sum_i TP_i$ in eq. 1 corresponded to the total number of correct predictions across all classes. We have changed this to " $\sum_i C_i$ where C_i is the number of correctly classified samples for class i " (L261).

L 245: "For the model based on the 2014 – 2022 UAS time series (see Fig. 2a for the spatial extent:", p. 10: I suggest moving this paragraph before **2.5 Performance evaluation of classification models** as this still belongs to Data Processing and Classification. Further, you now describe your processing and classification steps in full sentences instead of bullet points. I would suggest to decide for one option and be consistent. Is the model SVM based again?

We have implemented your suggestion (LL241ff). The classifications were performed using a SVM algorithm as for the biennial model (LL247-248).

L 246: "5 cm × 5 cm", p. 10: Why did you not resample them to the same 10x10 resolution?

Since there were no topographic data available and the number of input data was lower (which therefore required less processing time), we chose to retain the best spatial resolution available.

L 250: "7 × 7 moving windows", p. 10: To normalize the data?

We have responded to this comment above.

L 250: "GLCM additional texture features", p. 10: see comment before: why are you calculating these?

We have responded to this comment above, i.e., in the 'general comments' section.

L 254: "confusion matrices to correct for misclassification bias in the output maps", p. 10: Please explain how you do this. If your CM, for example, shows that 20% of your "stable palsa" pixels get mislabeled as "other", what do you do in order to correct this for the entire study site?

Because the confusion matrix is derived from a probabilistic sample, the probability of the errors is representative of the entire area. Thus if 20% of stable palsa get mislabeled as "other", then only 80 % of the pixels labeled as palsa should be counted for the area of palsa. Of course, you also need to look at the probability of "other" pixels that is mislabeled as "palsa" (y%). Thus, the best estimate of the real area of palsa is the sum of 80% of the area classified as palsa and y% of the area classified as "other". This is done for each class. This is a completely standard procedure in remote sensing, for which we refer to Czaplewski and Catts (1992) and Hay (1988).

L 260: "To investigate these electrical properties", p. 11: This reads like you are referring to the factors you just described as electrical properties. Maybe rephrase it to "To measure the electrical conductivity..."?

We have changed the sentence accordingly (L272).

L 261: “non-invasive nature”, p. 11: Can you describe the EMI in 1-2 sentences?

We wish to point out that these are indirect measurements that do not require disturbing the soil, i.e., we pass a plastic sledge over the ground. We have added a description of the EMI (LL278-282).

L 262: “large areas”, p. 11: Can you quantify how large? Comparable sizes of 14 ha or even larger ones?

We have changed the sentence to now read “[...] efficiency in covering relatively large areas (several thousand square meters to multiple hectares) [...]” (LL273-274).

L 267: “the instrument response is linearly related to soil EC”, p. 11: high values translating to high EC? What does this mean for the factors you described above?

The measurement of soil EC, and therefore the measurement made by the EMI, does not enable discrimination between salinity, soil water content, or pH, organic matter, etc... As requested by reviewer 1, we now stand that soil EC should be higher with clay content, water content, salinity, organic matter proportion, soil density and soil temperature, and permafrost absence since there is more water content in degraded areas (Doolittle and Brevik, 2014; McNeill, 1980). Permafrost degradation should therefore lead to an increase in soil EC (LL64-67).

L 267: “For our study site, we verified that the LIN hypothesis”, p. 11: Maybe you could move this sentence to the result section.

This was already stated in the results and discussion section and we removed it from the methodology.

L 270: “GNSS”, p. 11: Is this built into the EMI or is there an additional device mounted on the sledge taking these measurements.

We used a BU-353S4 manufactured by GlobalSat as an additional device. This has been added to the manuscript (L285).

L 272: “using a combined nugget and exponential model with a maximum prediction distance of 8 m.”, p. 11: Here, an explanatory sentence would be nice - or a reference.

We refer to the covariance function that we used for kriging. The new sentence now reads “[...] using a combined nugget and exponential model as defined by the fitted covariance function, with a maximum prediction distance of 8 m.” (LL87-88).

L 279: “horizons”, p. 11: soil horizons

We have changed the sentence accordingly (LL294-295).

L 281: “consists in”, p. 11: “is described/ defined by”

We changed the sentence to “[...] between the two years (k_{gOC}) corresponds to the difference [...]” (L297).

L 282: “The results consist of a TOC or MAOC stock that is made vulnerable annually as a result of palsa degradation. The actual timing of OC loss remains unknown.”, p. 11: Is this already a result or are you pointing to the consequences? “made vulnerable” - “exposed to”?

We did not make actual measurements of OC losses, we merely suggest that this OC stock would be vulnerable to export of degradation into GHGs.

L 292-293: “Several tests have been conducted to determine which parameters provide the best classification results for palsa degradation between 2019 and 2021.”, p. 12: Which tests were those, are you explaining them somewhere? In the methods you currently only address the standard metrics (OA, Prec., Recall, F1) - were there more?

This point has been addressed through the modifications we brought to our methodology and with the addition of Fig. A. 5.

L 297: “undergoing or under recent degradation”, p. 12: Most of the time you write “undergoing degradation” instead of “undergoing or under recent degradation” in your manuscript. I would keep it this simple, to not get too wordy.

We have changed the sentence accordingly (L316).

L 300: “Fig. C. 1a”, p. 12: How is this mean permutation importance calculated? Is this one of the tests you address above? Please explain this in the method section.

We have responded to this comment above, i.e., in the ‘general comments’ section.

L 300: “relative elevation and slope to perform a land cover classification over the Stordalen catchment was also done by Siewert (2018) and resulted in an overall accuracy of 74%.”, p. 12: So, including the diff in relative elevation improved the classification result compared to Siewert et al.?

Our point is merely that Siewert et al. classified the land cover with inputs that were relatively similar to ours and achieved an overall accuracy of 74%, i.e., quite similar to what we obtained.

L 306: “79%.”, p. 12: introducing which of the three window sizes achieved this value? Is it possible to include the confusion matrices for each run - i: RGB, ii: RGB + relative elevation, iii: - in the appendix? It is still not clear to me how many model runs there were in total and if the mean spatial filters were introduced one by one or at the same time. I would suggest to include a workflow or detailed description under 2.4.

Fig. C. 1b shows the details of the increases in OA and F-scores for the different model runs, based on the input data. We believe that providing details of the confusion matrices for all runs would not add any further information and could undermine the message. Yet, we recognize that adding a Fig. A. 5 on the workflow is very welcome (see responses above).

L 307: “Adding standard deviation spatial filters improves the overall accuracy to 80%, as does the standard deviation of the GLCM,”, p. 12: See comment above. It seems this was done one by one.

You are right, we tested other inputs, which proved to be of no further relevance and were therefore not included in the rest of the study. Nevertheless, we believe it is worth mentioning them briefly. This should be clarified now with the newly added Fig. A. 5.

L 313-314: “Then, initial predictions were refined using elevation differences and model scores to ensure alignment with domain knowledge (see section 2.4:.”, p. 12: In the methods you refer to “highest decision score from the set of allowed classes”. Could you shortly elaborate on that? There is no information on how many pixels in the study area were refined like this. Do you have a table of these decision scores? It is hard to understand this without any numbers.

We have responded to this comment above, i.e., in the ‘general comments’ section.

L 316: “evolution”, p. 13: improvement?

We have changed the sentence accordingly (L333).

L 326: “Tab. C. 1; Tab. C. 2:”, p. 13: This is a little confusing: In C1 and C2 you present the CMs for the model run with all (?) 55 bands as: (i) 11 bands with original data (3 spectral bands, relative elevation and slope for the years 2019 and 2021 as well as the difference in relative elevation between 2019 and 2021: along with (ii) 3×11 bands with spatial filters (mean & standard deviation: over windows of increasing size, i.e. 3×3 , 5×5 and 7×7 , and finally (iii) the 11 bands from the texture attribute ‘homogeneity’. But in C 1b where you show the performance, the 55 bands also include the refinement and noise filter. At the same time, all of the latter model runs have an input of 55 bands. So, some bands were substituted by others? I suggest, when you add these clarifications to the method section to name the model runs more clearly and establish which bands were used for which run.

We have added a figure that clarifies the different runs of the model (Fig. A. 5), as a function of the input data and added references to the models runs in the text.

L 330-333, p. 13: See comment for Tables C1 and C2.

See above responses.

L 348: “2014 WorldView 2 satellite”, p. 14: Where was the 2000 dataset from?

The 2000 dataset comes from aerial photos. Yet, we understand that mentioning that the 2014 dataset comes from satellite added unnecessary confusion and we removed this information, which is besides well described in Varner et al. (2022).

L 351: “0.9%·a-1 to 1.1%·a-1”, p. 14: Is this range due to the different model outputs? Could you briefly state, which values result from which model run?

This range comes from the calculation of run 8. It consists of (i) the degradation rate calculated from the dynamic classes (confusion matrix Table 2) and (ii) the difference between the 2019 and 2021 after aggregation of the classes (confusion matrixes Table C.1 & C.2).

L 352: “previous periods”, p. 14: “the 1970-2000 period”

We have changed the sentence accordingly (L368).

L 371: “Recent modeling studies besides”, p. 15: “Besides, recent modeling studies ...”

We have changed the sentence accordingly (L380).

L 378: “a large number of study sites”, p. 15: “a large number of study sites of comparable sizes”

We have changed the sentence accordingly (L388).

L 379: “humidity”, p. 15: This is misleading. You don't measure humidity, but quantify the proportion of open water and inundated areas to total land cover and - in the case of the EC data - higher soil moisture.

We have changed the title of this section to “3.3 Palsa degradation leads to increases in soil moisture” (L395)

L 380: “model from the broader temporal view”, p. 15: This reads like the model used for the 2014-2022 data was a different one.

We have changed the sentence to “[...] using the classification models from the 2014 – 2022 times series (Fig. 6)” (L398).

L 380: “we observe a trend towards an increase in the area”, p. 15: “we observe an increase”

We have changed the sentence accordingly (L399).

L 386-387: “Furthermore, the model from the 2014 – 2022 UAS time series does not use terrain morphology data (relative elevation and slope: for classification.”, p. 15: Okay, here it is clearer, that the model seems to be the same (SVM), but the input varied. Please, clarify this in the method section.

Thank you. We have clarified this (L247).

L 388: “quality indicators (Fig. 5a) are weaker than for the biennial model (Tab. 2)”, p. 15: Please be consistent and precise with your wording, you are comparing the F scores here, not all quality indicators.

We have changed the sentence to now read “As a result, the overall accuracies and F-scores (Fig. 5a) are weaker than for the biennial model” (LL403-404).

L 401-402: “enabling the estimation of soil EC from the quadphase component of the measured field, as the LIN assumption is met (McNeill, 1980:.”, p. 17: Could you explain this in 1-2 sentences?

In the methods section, we now clarify what we mean by the LIN hypothesis (LL277-282): “[...] The EC is estimated using the McNeill model (McNeill, 1980), which is widely applied in EMI surveys. EMI sensors such as the EM38 generate a primary electromagnetic field that induces electrical currents in the ground. These currents generate a secondary electromagnetic field that is measured by the sensor’s receiver. Under conditions known as “operating under low induction numbers (LIN)” the secondary field is proportional to the ground current and is used to calculate soil EC (Doolittle and Brevik, 2014); meaning that the instrument response is linearly related to soil EC. [...]”

L 407-408: “vegetation types, as shown in the RGB orthophoto”, p. 17: Can you repeat them here?

The new sentence now reads: “Electrical conductivity patterns also align with the distribution of vegetation types, as shown in the RGB orthophoto, with stable palsa associated with low-growing ericaceous and woody plants and degraded areas covered by green sedges (Fig. 7d).” (LL423-424).

L 408: “more developed vegetation”, p. 17: What do you mean by “more developed”? Are the plants in high EC areas considerably higher or lushier?

High EC areas correspond to more degraded areas, associated with the presence of sedges such as Eriophorum Spp which are therefore more developed, while stable palsa areas are dominated by low-growing ericaceous and woody plants.

L 410: “ $R^2 = 51\%$ ”, p. 17: Could you explain this? In my experience 51 % is a moderate correlation, not a clear linear one.

The new sentence now states “A relationship is observed between EC and relative elevation ($R^2 = 51\%$; Fig. 7e), further emphasizing the influence of micro-topography on soil hydrology and conductivity.” (L427)

L 412: “p-values from Kruskal-Wallis test $< 10^{-3}$ ”, p. 17: could you list these p-values here or add them to the figure?

The new sentence now stands as “Significant differences in EC are also observed between the different degradation classes (p-value from Kruskal-Wallis test $< 10^{-3}$ & p-values from Wilcoxon rank-sum tests $< 10^{-3}$; Fig. 7f).” (LL428-429). We do not believe it is relevant to include all p-values in the text or figure, for the sake of brevity. For reference, the results of Wilcoxon rank-sum tests:

- other vs intact: p-value = 2.8×10^{-7}
- other vs undergoing: p-value = 6.1×10^{-4}
- other vs degraded: p-value = 2.2×10^{-26}
- intact vs undergoing: p-value = 2.2×10^{-71}
- intact vs degraded: p-value = 0
- undergoing vs degraded: p-value = 1.3×10^{-73}

L 420: "Figure 6:", p. 18: What is the white space on the left of b,c,d ? The data points for class "other" in subfigure e) are hard to see. Could you use a different color for this class?

The white space is solely due to the shape of the 2019-2021 model's delineation. The color code was carefully chosen to enable easy interpretation of the plots by people suffering from color vision deficiencies. We therefore believe it inadvisable to change this color.

L 423-424: "Boxplots of the evolution of electrical conductivity as a function of the class.", p. 18: "Boxplots of the EC per land cover class"

We have changed the sentence accordingly (LL440-441).

L 427: "land-cover", p. 19: land cover

We have changed the sentence accordingly (L444).

L 432: "23 cm", p. 19: Where does this number come from? Is this derived from your DEM?

*This is the depth down to which we calculated carbon stocks (see **Tab. B. 1** & Patzner et al., 2020).*

L 434-435: "They consist of TOC or MAOC stocks that are made vulnerable annually as a result of palsa degradation and represent first order estimates.", p. 19: Could you please check the grammar of this sentence again? What is "they"?

We changed the sentence to "These are of TOC or MAOC stocks [...]" (L451).

L 445: "~ 7.1 g·C·m⁻²·a⁻¹ in 2019 to ~ 7.3 g·C·m⁻²·a⁻¹", p. 19: Could you briefly say where these numbers are coming from?

*This is a weighted average of emissions reported in Łakomiec et al. (2021) by the respective surface areas of palsas and degraded areas in 2019 and 2021 (see **Tab. B. 2** for the surfaces). There was a calculation error in the previous version of the paper, and we apologize for this.*

L 450: "show much lower or higher results than", p. 19: L450 : "deviate from"

We have changed the sentence accordingly (L467).

L 456: "evolution", p. 19: change?

We have changed the sentence accordingly (L473).

L 460: "arctic", p. 20: "Arctic"

We have changed the sentence accordingly (L477).

L 473-474: "We have conducted an evaluation of palsa degradation through time based on photogrammetric surveys providing access to RGB imagery and topography", p. 20: e.g. "We quantified palsa degradation using RGB imagery and topography data from UAS surveys, ..."

We have changed the sentence accordingly (L489).

L 487-488: "pool of 12 metric tons of organic carbon", p. 20: under 3.4 you state "24 metric tons of TOC, including 6 metric tons of MAOC". How did you get to this number?

The response to this comment is presented above.

L 488: “~25% is mineral-interacting organic carbon”, p. 20: could you please explain briefly, how you arrived at this number?

It is essentially the ratio of MAOC to TOC: 3 metric tons of MAOC versus 12 metric tons of TOC ; $3/12 = 25\%$

L 489: “~ 7.1 g·C·m⁻²·a⁻¹ in 2019 to ~ 7.3 g·C·m⁻²·a⁻¹ in 2021”, p. 21: see comment L 445.

The response to this comment is presented above.

Figure A. 1: “evolution of active layer depth along the gradient; (c) evolution of the volumetric water content along the gradient”, p. 22: See comment under 2.1: Did you do these measurements? Could you briefly describe?

The response to this comment is presented above.

“Table A. 1:”, p. 23: Do Robota Triton XL and Sensefly Ebee also use RTK?

Robota Triton XL did not use RTK, and the 2019 dataset (Sensefly Ebee) was georeferenced with > 10 ground control points, yet all datasets were co-registered (see above response).

“Figure A. 3”, p. 26: Could you please also label the x-axes of subfigures a-f and normalize them to a uniform range? This would make interpretation a little easier for the reader.

*The figure has been updated accordingly. Note that it is now **Fig. A. 4**.*

New section for the revised version of the manuscript.

2.6 Quantitative analysis of palsa loss rates

To quantify changes in palsa extent, we calculated the average annual rate of reduction of the areas of intact palsa from one year to the next (e.g., Olvmo et al., 2020; Verdonen et al., 2023; **eq. 5**). This method assumes that the palsa area decreases annually by a constant percentage applied to the remaining area, analogous to an exponential decay model.

$$\text{average annual rate of reduction (\% a}^{-1}\text{)} = \left(\left(\frac{A_{\text{start}}}{A_{\text{end}}} \right)^{\frac{1}{Y_{\text{start}} - Y_{\text{end}}}} - 1 \right) \quad (5)$$

where A_{start} and A_{end} are the total area of palsas at the start year Y_{start} of the respective period and at the end of it (Y_{end}). This formulation allows comparison across sites and time intervals of varying length, since the value is normalized to an annual percentage rate.

New figures for the revised version of the manuscript.

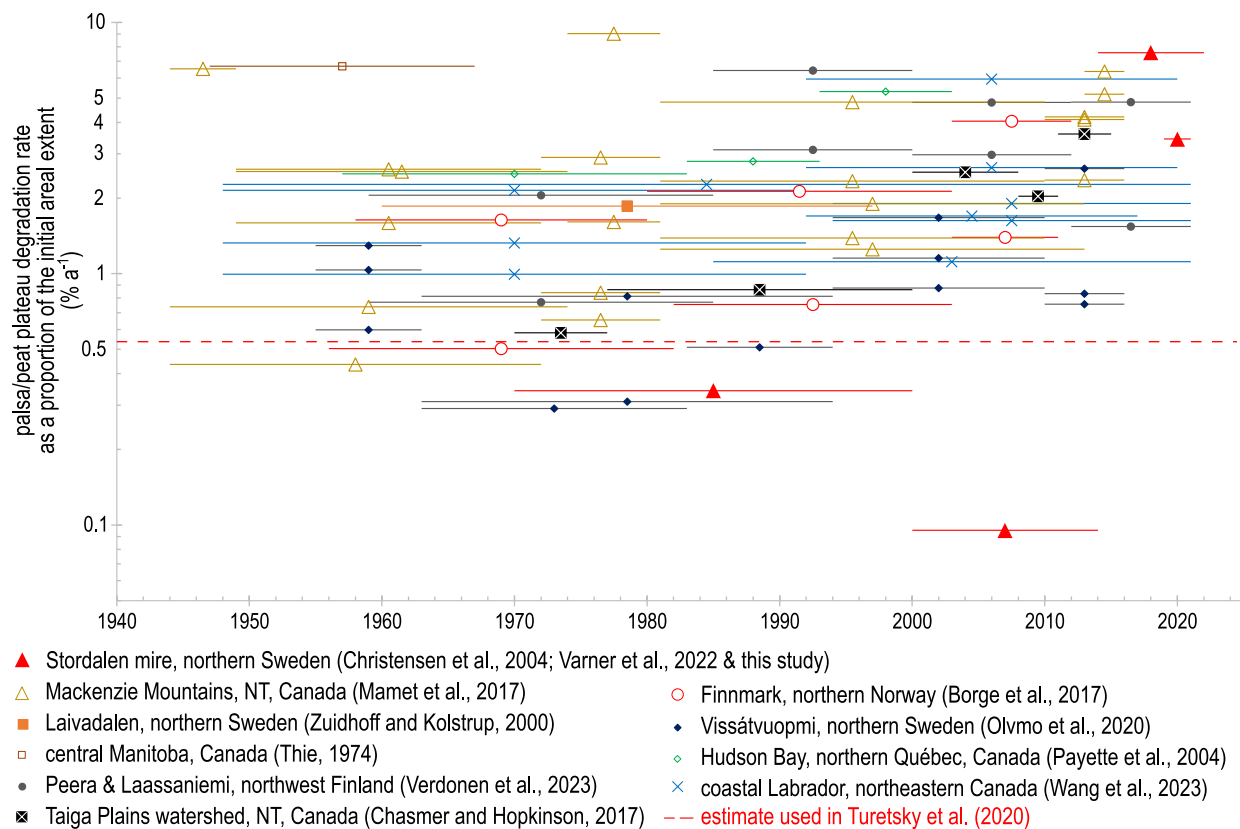


Figure 5: Meta-analysis showing the rates of degradation of palsas/peat plateaus relative to the initial palsa areal extent. Where required, rates reported here were recalculated using eq. 5 for inter-study comparability.

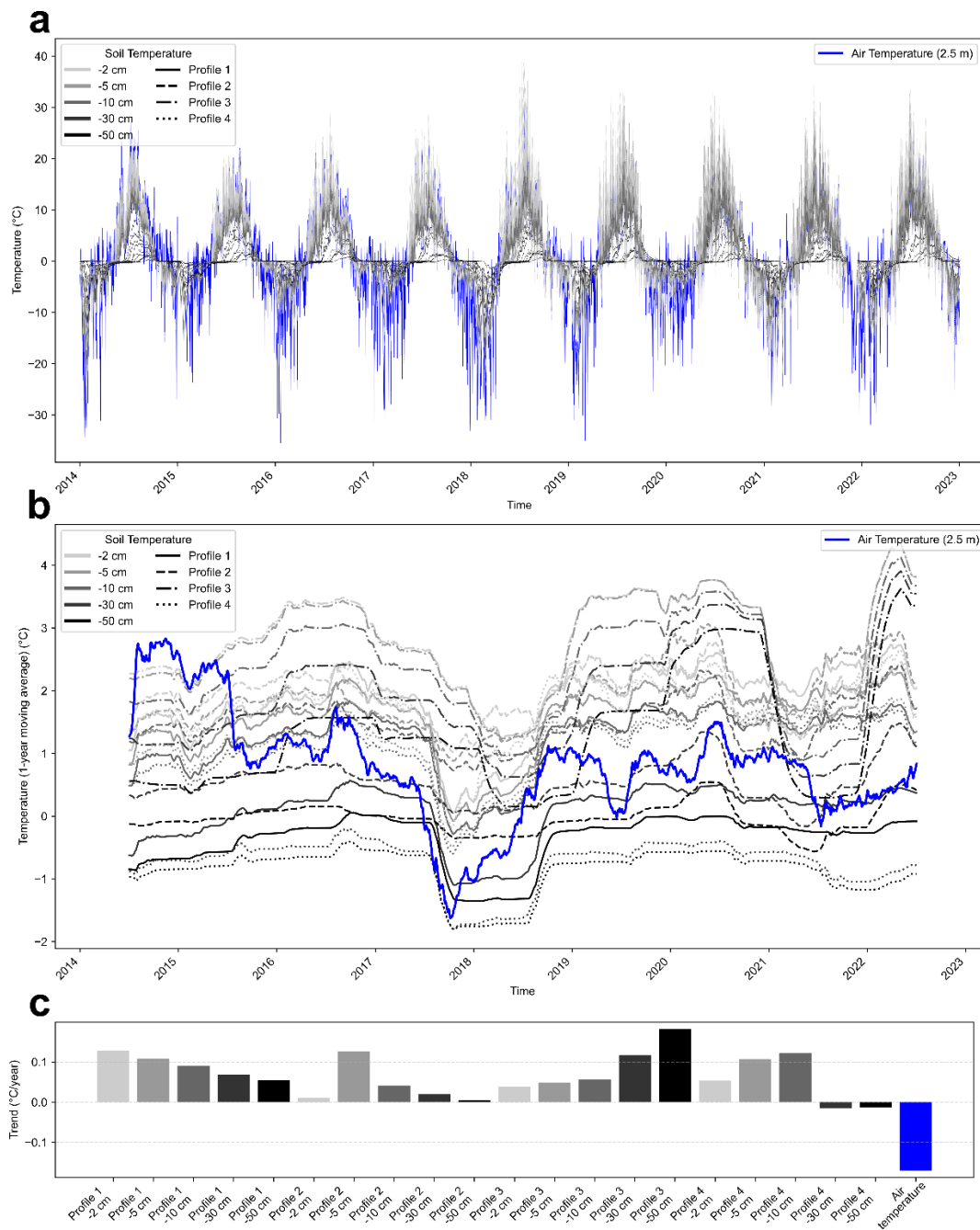


Figure A. 1: Air and ground temperatures at Stordalen between 2014 and 2022. (a) half-hourly air temperature at 2.5 m and soil temperatures from 4 profiles at 2, 5, 10, 30, and 50 cm depth (ICOS Sweden, 2023); (b) annual moving averages calculated on these same data; (c) significant trends in temperature changes calculated from moving averages (p-values < 10⁻³).

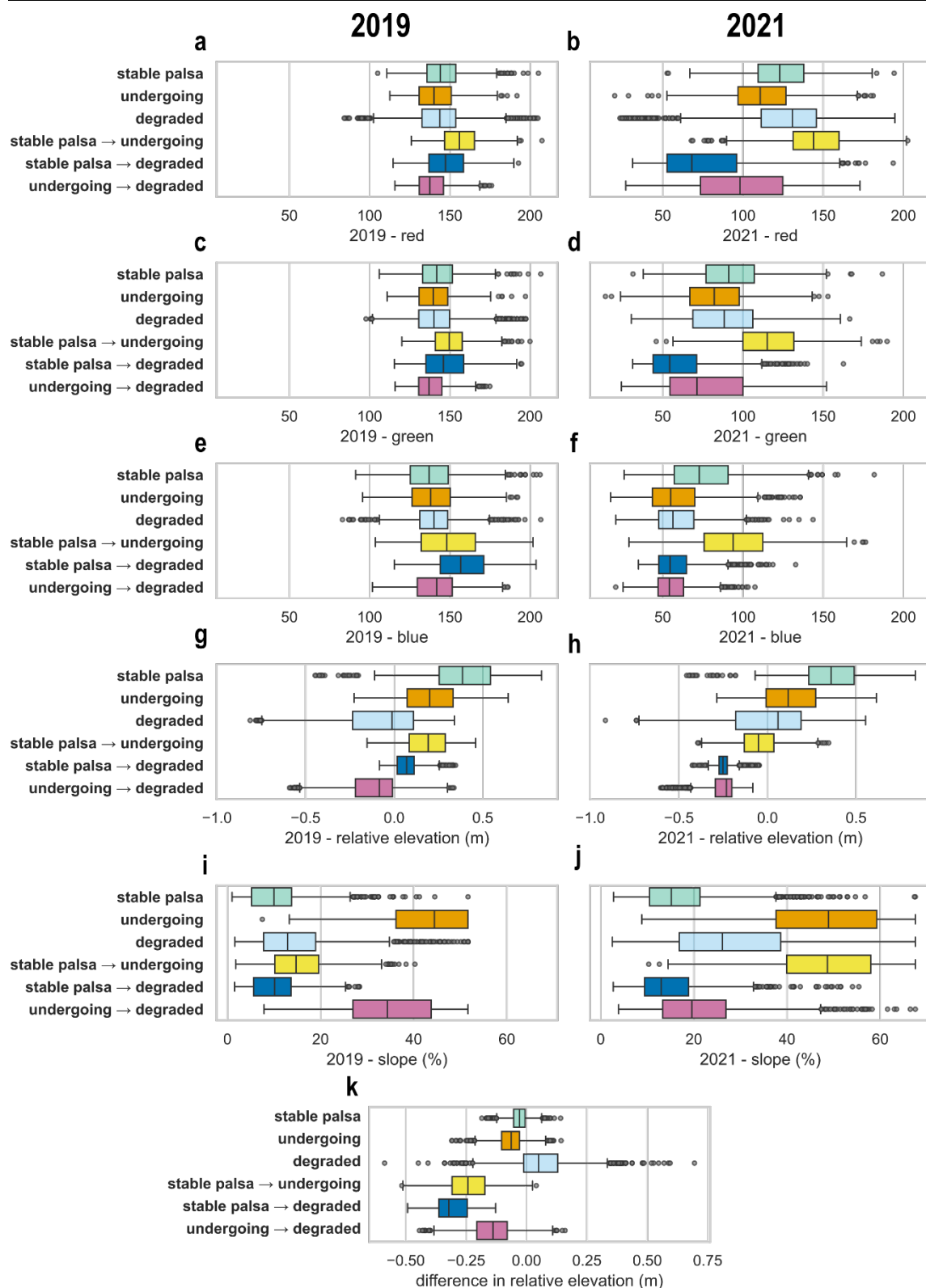


Figure A. 4: Distribution of the selected input data values across the training areas of the classification model, namely (a-b) red color, (c-d) green color, (e-f) blue color, (g-h) relative elevation data corrected by removal of the bowl-shape effect, (i-j) slope values derived from the original DSM and (k) the difference in relative elevation between 2019 and 2021. DSM = digital surface model. Color code as in Fig. 4.

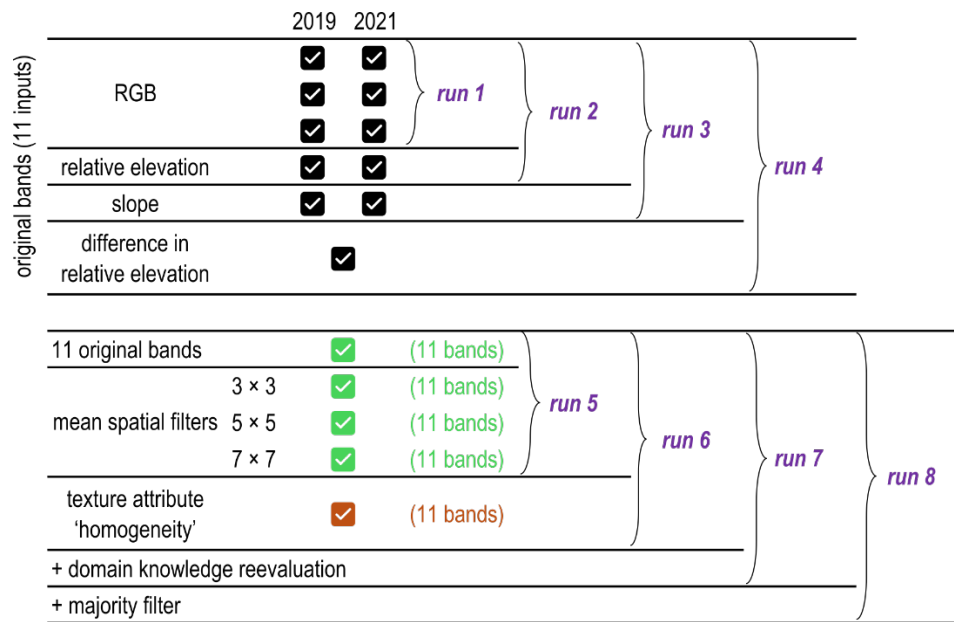


Figure A. 5: Workflow of the different runs of the model with successive additional data inputs and post-processing steps.

References

- de la Barrera-Bautista, B., Boyd, D. S., Ledger, M., Siewert, M. B., Chandler, C., Bradley, A. V., Gee, D., Large, D. J., Olofsson, J., Sowter, A., and Sjögersten, S.: Towards a Monitoring Approach for Understanding Permafrost Degradation and Linked Subsidence in Arctic Peatlands, *Remote Sensing*, 14, 444, <https://doi.org/10.3390/rs14030444>, 2022.
- Borge, A. F., Westermann, S., Solheim, I., and Etzelmüller, B.: Strong degradation of palsas and peat plateaus in northern Norway during the last 60 years, *The Cryosphere*, 11, 1–16, <https://doi.org/10.5194/tc-11-1-2017>, 2017.
- Bosiö, J., Johansson, M., Callaghan, T. V., Johansen, B., and Christensen, T. R.: Future vegetation changes in thawing subarctic mires and implications for greenhouse gas exchange—a regional assessment, *Climatic Change*, 115, 379–398, <https://doi.org/10.1007/s10584-012-0445-1>, 2012.
- Christensen, T. R., Johansson, T., Åkerman, H. J., Mastepanov, M., Malmer, N., Friborg, T., Crill, P., and Svensson, B. H.: Thawing sub-arctic permafrost: Effects on vegetation and methane emissions, *Geophysical Research Letters*, 31, <https://doi.org/10.1029/2003GL018680>, 2004.
- Czaplewski, R. L. and Catts, G. P.: Calibration of remotely sensed proportion or area estimates for misclassification error, *Remote Sensing of Environment*, 39, 29–43, [https://doi.org/10.1016/0034-4257\(92\)90138-A](https://doi.org/10.1016/0034-4257(92)90138-A), 1992.
- Doolittle, J. A. and Brevik, E. C.: The use of electromagnetic induction techniques in soils studies, *Geoderma*, 223–225, 33–45, <https://doi.org/10.1016/j.geoderma.2014.01.027>, 2014.
- Hay, A. M.: The derivation of global estimates from a confusion matrix, *International Journal of Remote Sensing*, 9, 1395–1398, <https://doi.org/10.1080/01431168808954945>, 1988.
- Hugelius, G., Loisel, J., Chadburn, S., Jackson, R. B., Jones, M., MacDonald, G., Marushchak, M., Olefeldt, D., Packalen, M., Siewert, M. B., Treat, C., Turetsky, M., Voigt, C., and Yu, Z.: Large stocks of peatland carbon and nitrogen are vulnerable to permafrost thaw, *Proceedings of the National Academy of Sciences*, 117, 20438–20446, <https://doi.org/10.1073/pnas.1916387117>, 2020.
- ICOS Sweden: Collection of Abisko Stordalen Palsa Bog Swedish network data, <https://doi.org/10.18160/Q6H6-B94B>, 2023.
- Abisko-Stordalen Palsa Bog | ICOS Sweden: <https://www.icos-sweden.se/abisko-stordalen>, last access: 29 August 2025.
- Łakomiec, P., Holst, J., Friborg, T., Crill, P., Rakos, N., Kljun, N., Olsson, P.-O., Eklundh, L., Persson, A., and Rinne, J.: Field-scale CH₄ emission at a subarctic mire with heterogeneous permafrost thaw status, *Biogeosciences*, 18, 5811–5830, <https://doi.org/10.5194/bg-18-5811-2021>, 2021.
- McNeill, J. D.: Electromagnetic terrain conductivity measurement at low induction numbers, Geonics Limited, Tech. Note TN-6, 8, 1980.
- Olvmo, M., Holmer, B., Thorsson, S., Reese, H., and Lindberg, F.: Sub-arctic palsa degradation and the role of climatic drivers in the largest coherent palsa mire complex in Sweden (Vissátvuopmi), 1955–2016, *Sci Rep*, 10, 8937, <https://doi.org/10.1038/s41598-020-65719-1>, 2020.
- Palace, M., Herrick, C., DelGreco, J., Finnell, D., Garnello, A. J., McCalley, C., McArthur, K., Sullivan, F., and Varner, R. K.: Determining Subarctic Peatland Vegetation Using an Unmanned Aerial System (UAS), *Remote Sensing*, 10, 1498, <https://doi.org/10.3390/rs10091498>, 2018.

Patzner, M. S., Mueller, C. W., Malusova, M., Baur, M., Nikeleit, V., Scholten, T., Hoeschen, C., Byrne, J. M., Borch, T., Kappler, A., and Bryce, C.: Iron mineral dissolution releases iron and associated organic carbon during permafrost thaw, *Nat Commun*, 11, 6329, <https://doi.org/10.1038/s41467-020-20102-6>, 2020.

Renette, C., Olvmo, M., Thorsson, S., Holmer, B., and Reese, H.: Multitemporal UAV lidar detects seasonal heave and subsidence on palsas, *The Cryosphere*, 18, 5465–5480, <https://doi.org/10.5194/tc-18-5465-2024>, 2024.

Turetsky, M. R., Abbott, B. W., Jones, M. C., Anthony, K. W., Olefeldt, D., Schuur, E. A. G., Grosse, G., Kuhry, P., Hugelius, G., Koven, C., Lawrence, D. M., Gibson, C., Sannel, A. B. K., and McGuire, A. D.: Carbon release through abrupt permafrost thaw, *Nature Geoscience*, 13, 138–143, <https://doi.org/10.1038/s41561-019-0526-0>, 2020.

Varner, R. K., Crill, P. M., Frolking, S., McCalley, C. K., Burke, S. A., Chanton, J. P., Holmes, M. E., Isogenie Project Coordinators, Saleska, S., and Palace, M. W.: Permafrost thaw driven changes in hydrology and vegetation cover increase trace gas emissions and climate forcing in Stordalen Mire from 1970 to 2014, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 380, 20210022, <https://doi.org/10.1098/rsta.2021.0022>, 2022.

Verdonen, M., Störmer, A., Lotsari, E., Korpelainen, P., Burkhard, B., Colpaert, A., and Kumpula, T.: Permafrost degradation at two monitored palsa mires in north-west Finland, *The Cryosphere*, 17, 1803–1819, <https://doi.org/10.5194/tc-17-1803-2023>, 2023.

Vonk, J. E., Tank, S. E., Bowden, W. B., Laurion, I., Vincent, W. F., Alekseychik, P., Amyot, M., Billet, M. F., Canário, J., Cory, R. M., Deshpande, B. N., Helbig, M., Jammet, M., Karlsson, J., Larouche, J., MacMillan, G., Rautio, M., Walter Anthony, K. M., and Wickland, K. P.: Reviews and syntheses: Effects of permafrost thaw on Arctic aquatic ecosystems, *Biogeosciences*, 12, 7129–7167, <https://doi.org/10.5194/bg-12-7129-2015>, 2015.

Webb, H., Fuchs, M., Abbott, B. W., Douglas, T. A., Elder, C. D., Ernakovich, J. G., Euskirchen, E. S., Göckede, M., Grosse, G., Hugelius, G., Jones, M. C., Koven, C., Kropp, H., Lathrop, E., Li, W., Loranty, M. M., Natali, S. M., Olefeldt, D., Schädel, C., Schuur, E. A. G., Sonnentag, O., Strauss, J., Virkkala, A.-M., and Turetsky, M. R.: A Review of Abrupt Permafrost Thaw: Definitions, Usage, and a Proposed Conceptual Framework, *Curr Clim Change Rep*, 11, 7, <https://doi.org/10.1007/s40641-025-00204-3>, 2025.