Point-by-point replies to the question and comments by Reviewer 2

Dear Reviewer 2,

we are pleased to submit the replies to your questions and are thankful for the insightful comments and many good suggestions, as well as we are grateful for your time and effort in providing valuable feedback. We believe that addressing the issues raised by you, have now substantially improved our manuscript.

We hope our answers meet your approval. Your comments and our point-by-point responses are presented below. Please note, that we added a detailed description of a new RF modelling approach in the appendix A.

Reviewer #2 comments	**Action**	**Response**
1. The authors produce a random forest (RF) model based on lidar-derived topographic predictors and point observations of snow depth. The RF model is used to create a continuous snow depth field over the three independent palsas in Finland and Sweden. It is then evaluated against the point observations and compared to UAS-lidar-derived snow depths. Finally, the authors discuss the implications of snow depth variability on permafrost dynamics at the palsas. The manuscript has well-constructed figures, relies on a unique and interesting dataset, includes an assessment of a wide range of reasonable terrain predictors of snow depth, and the methods are on the right track. However, there are some significant concerns, including manuscript organization/framing, limited lidar validation/processing concerns, model overfitting, and generally weak analysis/discussion. These are detailed below.	Answered	We sincerely appreciate your thorough review, insightful comments, and thoughtful questions. Your valuable feedback has provided us with important guidance to improve our manuscript. We believe that addressing the issues you have raised have significantly enhanced the clarity and overall quality of our work. Below, we present our point-by-point responses to your comments and suggestions.
2. Lastly, please do not be overwhelmed by all of the comments! Addressing the major suggestions and proofreading the manuscript thoroughly should move this study much closer to publication. The specific comments are intended as suggestions/thoughts to help steer the revision process and are generally related to the below Major Suggestions/Comments.	Answered	Thank you for your encouraging words. We appreciate your detailed feedback and are confident that your suggestions have significantly enhanced the quality of our manuscript.
Major Suggestions/Comments		
Concerns with Research Objectives, Methods, and Manuscript Organization		
3. Ln 94-99: It is my opinion that the research objectives need to be refined. The first stage of the paper should be an evaluation of lidar-based snow depth, followed by an evaluation	Changed/ Answered	In response to your comments and those of Reviewer 1, we refined the focus of our manuscript during the review process. Originally, our main objective was to analyse

	of the RF modeling approach. Only then should the authors discuss the potential implications of the depth patterns, and this should be a smaller part of the manuscript focused in the discussion. Since the authors did not explicitly collect data to link snow depth to changes in the active layer (or ice loss/gain), the outcomes are more based on expectations and assumptions – which may be valid, but to verify and to be a focus of the manuscript would require more data. The points described on Ln 97-99 are underdeveloped and unsupported by observations.		snow depth patterns and their potential impact on palsas. However, based on your valuable suggestions, we now recognize the importance of a more explicit comparison between UAS-LiDAR and RF-modelled snow depth products. We have adjusted our research objectives accordingly to highlight this comparison. However, we still believe it is important to discuss the potential impact of snow depth distribution on palsa dynamics. We agree, nevertheless, that these interpretations are based on assumptions and not on direct observed data. For clarity, we explicitly point out that these ideas need to be further validated in future studies.
4.	Section 4.3 needs revision. Unless the expected errors in the lidar product are further expanded upon, these are physical observations and it is standard practice to assume that these products have uncertainties errors proportional to the sensor error (e.g., ~5 cm). This can be directly evaluated from ground observations of snow depth and was to some degree. However, the errors were much larger than expected (>20 cm, Ln 298-302), raising concerns about the processing of lidar data to produce snow depth maps. This component and the framing of the analysis are significant concerns. A section early on evaluating the lidar depth products seems necessary and considering the influence of vegetation on their accuracy explicitly (for example, examining some of the outliers in Figure 8 more closely) – addressing the concern of vegetation compression should be added here and vegetation height models produced from the summer lidar point cloud	Changed/ Answered	Thank you for highlighting these concerns. In response, we have conducted an additional model run, refining the LiDAR-based snow depth estimation by removing vegetation from the initial LiDAR products to create a more accurate DTM. Additionally, we implemented hyperparameter tuning and tested different cross-validation folds to mitigate potential overfitting. These methodological improvements are explained in detail in Appendix A. The updated results demonstrate a substantial improvement in the accuracy of the LiDAR-derived snow depth estimates. However, snow depth estimation remains less accurate for the *Thermokarst* point group, likely due to high reflectance in water-dominated areas. Furthermore, the revised RF modelling approach now has improved results in terms of representation and generalization. To ensure clarity, we have revised the relevant sections accordingly and have introduced a new section specifically detailing the processing and validation of LiDAR data. In the following, we refer to Appendix A. There you find all necessary information, which answer your questions.
5.	Comparing the lidar to an RF model trained and evaluated against <200 observations directly is not appropriate. As presented, the lidar depth analysis does not add much to the manuscript – I suggest it be redone (reprocessed data, more detailed lidar depth evaluation), and/or, the work reframed to simply build the RF model using lidar terrain and snow depth point observations, then a revised analysis on how these patterns are expected to influence the palsa stability.	Changed/ Answered	See Appendix A.
Random Forest Modeling Concerns			
6.	Ln 189-191: It seems like little consideration was given to the hyperparameters, and	Changed/ Answered	See Appendix A.

	several important ones (like maximum split size, and minimum node size) are not mentioned. Please clearly state the hyperparameters used, and an optimization routine should be included to select these – not just using defaults – which are likely geared towards a much larger data set. If done correctly, this will reduce overfitting (see following concerns)		
7.	Various model runs were not clear. The predictors for model 1,2, and 3 should be explicitly stated, with the appropriate reasoning within the methods section	Changed	We have described the input parameters used for each model run in lines 205–209. However, based on your (and Reviewer 1) suggestions and the revised model design, we have decided to focus on a single model run. As a result, we used only 12 parameters and removed the information related to previously unused parameters and former model runs. The section on input parameters and model runs has been updated accordingly.
8.	10-fold cross-validation is not sufficient to ensure that the model is not overfit. Each model is still trained with 90% (9/10) of all data (and the training dataset is relatively small <200 snow depths). The authors should explore the influence of fewer folds (e.g., 3-10) to assess how much model performance is degraded. For such a small dataset, around four (4) folds seems more appropriate in this case	Changed/ Answered	See Appendix A.
9.	Ln 232 – 242: This is relevant to RF model training/evaluation – I suggest moving to the random forest training section	Changed	We agree and moved this part to the description of the RF algorithm and modelling data preparation.
10.	The results shown in Table 3 (R2 > 0.99, RMSE less than the expected measurement uncertainty <3 cm) suggest substantial overfitting of the random forest model	Changed/ Answered	See Appendix A.
11.	Avoid analysis based on terms like 'potentially,' 'possibly,' and 'probably' -- you should focus your study on explaining and describing the data you have collected its likely implications with clearly stated support	Changed	We agree and changed/avoided assumptions within the abstract/discussion at lines 18, 320, 323, 347. We either supported our ideas by own observations in the field (decreasing parts, block erosion) or pointing out the uncertainties and the necessity to verify these assumptions in further studies.
12.	The terminology of cooling and warming spots was unclear – since these are not a standard term to my knowledge, these need to be explicitly defined early on and used consistently throughout the manuscript	Changed	Thank you for highlighting this important point. To ensure clarity and consistency, we have explicitly defined these terms in the introduction when explaining the role of snow dynamics in palsa environments: We define *Cooling Spots* as areas on and around a palsa where the snow cover remains relatively thin during winter. Due to the lack of insulating snow, these areas experience increased heat loss from the ground to the atmosphere, allowing frost to penetrate deeper into the subsurface. As a result, when

		summer arrives, the snow in these areas melts earlier, exposing the ground to warm temperatures for an extended period. This prolonged exposure leads to a deeper thaw of the active layer, making these locations more susceptible to ground cracking and thermokarst formation, particularly along the palsa edges. Cooling spots are typically found in elevated or wind-exposed areas of the palsa, where snow accumulation is naturally limited. In contrast, *Warming Spots* are areas where a relatively thick snow cover accumulates during winter. The insulating properties of the snow reduce heat loss from the ground, preventing deep frost penetration and keeping the underlying soil comparatively warmer throughout the winter. In summer, the accumulated snow melts later, delaying the warming of the ground and slowing active layer thawing. Consequently, the active layer in these areas tends to be shallower compared to cooling spots. Warming spots are typically located in depressions, concave terrain, or wind-sheltered locations where snowdrifts form. These definitions have been integrated into the manuscript to provide a clearer conceptual framework for our analysis.
13. Ln 71-73: UAS-lidar-based snow depth monitoring approaches/literature should be sufficiently reviewed in the introduction (& by the authors). The approaches used to produce snow-depth products do not align with standard practice (e.g., classifying the vegetation-free ground surface). See work by Avanzi et al., 2018; Harder et al., 2020; Jacobs et al., 2021.	Changed	We acknowledge the importance of previous studies that have applied UAS-LiDAR for snow depth mapping and appreciate your suggestion. Our initial focus was primarily on demonstrating the feasibility of snow depth modelling using RF and assessing its implications for palsas. However, as the study evolved, the focus shifted more towards a comparative analysis between Random Forest modelling and LiDAR-based snow depth estimation. In response to your suggestion, we have incorporated additional references on UAV LiDAR-based snow depth mapping. Specifically, we adopted the vegetation removal approach inspired by Jacobs et al. (2021) and further reviewed relevant studies, including those by Avanzi et al. (2020) and Harder et al. (2020).
14. Grammatical and sentence structure issues limit communication effectiveness in the paper. A thorough proofreading by a third party before resubmission would benefit the paper. Some specific instances of this were noted in the comments below	Changed	Thank you for pointing this out. We will have a close look to grammar and sentence structure before resubmission and will particular improve the comments you are mentioning below.
Minor/Technical/Grammatical Suggestions		

15. Stick with snow depth or snow height throughout – be consistent with word choice	Changed	Agree, we are now using the term snow depth and changed snow height in lines 258, 264 and 266.
16. The use of the word 'precision' is questionable at times (see abstract Ln 9). Precision measures the ability for repeatable measurements. Accuracy is a better term for assessing something like a random forest model. Read through the manuscript and consider if the use of 'precision' is appropriate throughout	Changed	This is a good point. We changed the term precise/precision/precisely in lines 9, 74, 90, 97, 188, 288, 316, 365, 370 and 396.
17. Word choice should be reviewed throughout – (e.g., Ln 78: 'very strong changes' could be 'control', Ln 212: 'realism')	Changed	We scanned our manuscript for judgmental words and changed them to more neutral terms.
18. Avoid broad terminology throughout, especially before something a term is explicitly defined (e.g., Ln 150 – was not clear what 'input parameter data' referred to again at Ln 181, 182, 185). I suggest defining the types of input parameters earlier on – like was done in Section 3.3	Changed	By "input parameters", we are specifically referring to the variables listed in Table 2. These parameters are all derived from the initially generated DSM in summer and were exclusively used in the RF model. To enhance clarity and avoid ambiguities, we have explicitly defined the term "input parameters" in line 144. Additionally, we have standardized the terminology by removing the word "data" in line 150 to ensure consistency throughout the manuscript.
Abstract		
19. Ln 13-15: Machine learning is used to model the snow depth spatially and relies on observations. On its own, it does not capture snow depth patterns. Consider rewording	Changed	We agree and changed the sentence: "This considerable difference highlights the capability of machine learning to model fine-scale snow distribution based on in-situ observations."
20. The abstract should be a single cohesive paragraph, avoid splitting into two parts	Changed	We agree and changed the abstract to a single cohesive paragraph.
Introduction		
21. Ln 26, 57-58: While snow cover duration is decreasing, the suggestion that snow depth is increasing substantially in these regions is less clear. This paper suggests snowfall extremes will be reduced in the study area (https://www.nature.com/articles/s41598-021-95979-4) – can you clarify this point?	Changed/ Answered	Thank you for your suggestion. We acknowledge that while snow cover duration is generally decreasing, trends in snow depth are more regionally variable. Increased winter precipitation may lead to higher snow depths in some areas, whereas other regions might experience a decrease in snowfall extremes, as suggested by Quante et al. (2021). Since we have not specifically analysed these trends for our study region, we recognize the need to be more cautious with this statement and have revised it accordingly to reflect the regional variability and associated uncertainties.
22. Ln 32-33: Sentence structure/clarity issues – please revise	Changed	We changed the sentence: *In northern Fennoscandia, particularly in northern Finnish Lapland - the main focus of this study - specific periglacial permafrost landforms known as palsas are at risk of disappearing within this century (Leppiniemi et al., 2023).*

23. The relevance of palsas is not addressed clearly in the introduction. Please add some sentences on their general significance, e.g., Do they stabilize permafrost? Provide habitat? Have societal relevance?	Changed	We have added that palsas serve as indicators of climate warming, as their degradation and disappearance reflect rising temperatures (Leppiniemi et al., 2023). Additionally, they provide important habitats for various animal species (Luoto et al., 2004) and hold significant cultural and societal relevance for the Sámi people, particularly in the context of traditional reindeer herding (Markkula et al., 2019).
24. Ln 53: remove 'exemplarily'	Changed	Removed.
25. Ln 56-57: These points seem essential for understanding the relevance of palsas – I suggest this is moved earlier in the introduction when palsas are defined	Changed	We moved the sentence "*Microtopography affects snow depth and creates an environment, in which the palsas usually receive enough penetrating cold air to remain stable and to last year after year due to a thin snow cover.*" to line 41.
26. Ln 60-61: 'in-situ measured data' or 'observations' need to be clarified. Is this temperature data? Snow depth? Other?	Changed	We are referring to "snow depth data" an inserted this term for clarification.
27. Ln 78-79: Wording is unclear – '…limits information value of satellite data…'	Changed	We changed the sentence: *Small-scale structures, such as palsas, exhibit significant variations in snow depth at fine spatial scales, which reduces the usefulness of satellite data for analysing small-scale processes in these structures.*
28. Ln 81:'Another' should start a new paragraph – this section is also very short relative to the prominent role that machine learning plays in the paper. I suggest adding more detail.	Changed	We would like to point out kindly, that there is a paragraph between line 80 and 81. As mentioned previously, we initially focused more on the impacts of the snow depth distribution to palsas in this paper. However, after your useful comments, we agree and inserted more details about machine learning and specifically RF.
29. Ln 90: '…test methods for generating detailed snow distribution maps..'' should lead this section. The objectives need to be clearly stated up front	Changed	Yes, we agree and have adjusted the focus of the paper accordingly. Specifically, we have ensured that the section begins with a clear statement of our objectives, emphasizing the evaluation of methods for generating detailed snow distribution maps.
Data and Methods		
30. Ln 141: a comprehensive dataset of what? Specify	Changed	We have clarified this statement by specifying that we collected a comprehensive dataset consisting of UAS-LiDAR data and in-situ snow depth measurements for modelling purposes.
31. Figure 3: Should clearly state the actual observations that were collected	Changed	We have clarified that the collected data include UAS-LiDAR measurements, which were used to generate DTMs for both winter and summer, as well as in-situ snow depth measurements, which served as training data for the modelling.
32. Ln 151: This is the first time LiDAR is mentioned. Needs to be introduced within the introduction	Changed	We agree and introduced LiDAR in lines 73 – 74 in the context of the studies by Rauhala et al. (2023) and Meriö et al. (2023).

33. Ln 157-160: very confusing. No SfM, but then orthophotos were created? That relies on photogrammetry -- but then you state point cloud densities. Are these associated with lidar or RGB orthophotos? If lidar, need to put it right after the lidar. Also, should report density per square meter as it is the standard.	Changed	We acknowledge the potential for misunderstanding and have clarified our statement in lines 157/158. Specifically, we used SfM techniques solely for the creation of orthophotos. The RGB flights were conducted using an Autel EVO II Pro V2 UAV at a flight altitude of 80 m, with a 75% overlap for each flight. Initially, we had stated that the orthophotos were acquired with the integrated RGB sensor of the LiDAR mapper. However, this was a misunderstanding, and we have now corrected this statement. The orthophotos do not contribute specific data to the analysis but were solely used for figure creation.
34. Ln 162-164: Revise sentence structure for clarity	Changed	We changed the sentence: "*By substracting the winter by the summer DSM in Geographic Information Systems (GIS) – ArcGIS Pro by Esri was used – snow depth distribution datasets were calculated, allowing the comparison of UAS-LiDAR snow depth (SD_{LiDAR}) and RF modelled (SD_{RF})*."
35. Ln 166: should be 'by an RTK GPS system'	Changed	We agree and added the term.
36. Ln 170: word choice - 'optimal'	Changed	We changed "optimal" to "diverse".
37. Ln 172: The sampling strategy is claimed to be randomized, though it appears observations were collected along transects with some random points. Some of these could be biased, so it would be useful to add a bit more description. There are also areas with clear gaps	Answered	We acknowledge that the sampling strategy may appear structured, potentially suggesting a bias. However, no strict transect approach was followed when measuring snow depth in Pousu. Instead, the sampling locations were selected based on terrain features, as illustrated in the figure in Appendix E, which we will include in the manuscript appendix. Measuring snow depth under these environmental conditions is challenging, and our data collection was constrained by a limited time frame. Therefore, we prioritized a well-distributed dataset that captures the variability within our palsa sites as effectively as possible.
38. Related to the previous point, the distribution of snow depth observations included in the appendix should be split by site (in my opinion)	Changed	See Appendix D.
39. Ln 176: It isn't easy to make out any snow-free areas on the palsas in the imagery – can these be indicated?	Answered	The snow-free points represent extreme locations in highly exposed areas. These points were specifically captured to ensure that the model is trained with the full range of snow depth variations observed in the field. However, due to the limited resolution of the orthophotos, it is difficult to clearly visualize these areas as the images appear too blurry to clearly highlight them. However, these snow-free areas are mainly located on steep slopes where wind-induced redistribution of snow and downslope movement have either

		removed or significantly reduced the snow cover.
40. Ln 184-188: To clarify, the full training set is based on only 185 observations – but increased due to the buffering? Please indicate how many unique features were actually used to train the model after the buffering. Will help the reader understand the robustness of the model	Changed/ Answered	See Appendix A.
41. Ln 193-194: Just state the metrics were normalized 0-1, with the highest output importance set as 1	Changed/ Answered	See Appendix A.
42. Ln 205: The removal of elevation as a predictor needs more explanation – the logic that is will 'reduce possible overfitting' is not apparent	Changed/ Answered	Please refer to our previous responses, especially comment #7, regarding the removal of the initial model runs. As a result, it is no longer necessary to elaborate on the exclusion and reintroduction of the *Elevation* parameter. For clarity regarding our initial approach: *Elevation* was excluded in the second model run because all other input parameters were derived from it. This step was taken to assess whether *Elevation* might introduce bias into the modelling results. After analysing the outcomes, we found no indication of such bias and subsequently decided to retain *Elevation* as an input parameter in the final model.
43. Ln 206-207: wordy, what is 'initial minimal impact'?	Changed	We changed the term to "low impact".
44. Ln 211: Unclear how this offers a balanced representation. The idea of taking an area is usually to remove noise, reduce the influence of sampling or geolocation errors, and to grow the training set size (taking groupings of nearby points vs. a single one - which should improve the robustness of the model). Please explain further.	Changed/ Answered	Thank you for your comment. We agree that the phrase "balanced representation" was not the most precise wording. To clarify, the buffering strategy was implemented to reduce noise, minimize the influence of geolocation and sampling errors, and enhance the robustness of the model by increasing the number of training points. By incorporating groupings of nearby points rather than relying on single-point measurements, this approach helps improve the model's stability and realism, as demonstrated in Bergamo et al. (2023). We have revised the manuscript accordingly to reflect this explanation more clearly.
45. Table 2: Nice table! For features like TPI (which are determined to be very important), you should be more detailed in their definition. More than 'it combines several topographic features.' TPI is generally just the relative elevation of a point to surrounding points within some radius (or adjacent pixels)	Changed	Thank you for your positive feedback! We agree with your suggestion and have added more detailed explanations for *TPI, Wind Effect, Valley Depth, Channel Network Base Level,* and *Wind Exposition* to ensure clarity and precision in their definitions.
46. Ln 238-239: Be careful with wording. Correlation (strength of linear relationship) and significance (based on statistical testing) are not the same thing	Changed	We recognize the difference between correlation and statistical significance and have adjusted the wording accordingly to ensure accuracy. In particular, we clarify that

		the analysis aimed to assess the strength of the relationships between the input parameters and the SD_{RF} predictions, rather than implying statistical significance unless explicitly tested.
Results		
47. Section 4.1: Nice job describing results clearly and sequentially by site.		Thank you very much!
48. Figure 5: Nice figure, it would be useful to add annotations for areas of interest referred to in Section 4.1 on the figure (e.g., the collapsed areas)	Answered	Thank you for your valuable suggestion! While we acknowledge that adding annotations could enhance interpretability, we aim to maintain clarity and avoid overloading the maps with excessive information. Additionally, we want to prevent cross-referencing multiple figures within a single visualization. For these reasons, we have decided not to modify Figure 5 further but will ensure that the areas of interest are clearly described and referenced within the text.
49. Ln 252: When stating things like 'slightly higher,' specify the magnitude (is this 10cm, 20cm, 5cm?). Same as Ln 257, how much lower?	Changed	We agree with your suggestion and have included exact numerical differences in centimetres to provide a more precise comparison in lines 252 and 257.
50. Ln 272-273: sentence clarity issue	Changed	We changed the sentence: Notably, deviations in the areas surrounding the palsas are primarily characterized by higher snow depths predicted by the RF model.
51. Figure 7: Nice figure! Be sure to add more specifications on the model runs in the methods section	Answered	See comment #42.
52. Ln 294: How were they separated into 'point groups' used to produce Table 4 – how were the different areas delineated and can these be added to the maps?	Changed/ Answered	See Appendix A and E.
53. Ln 310-313: Correlation analysis results should be included as a table – this could be added to the appendix if the authors do not want to include it in the body of the paper	Changed/ Answered	See Appendix F.
Discussion		
54. Ln 318-319: Revise based on previous comments	Answered	Revision have been done.
55. Ln 329: warming and cooling spots need to be defined more before this point. What is a good technical definition? For example, are warming spots where the net heat flux into the ground during the winter is highest - making these areas warmer? vs. Cooling spots, where the net heat flux into the ground is lowest? We need to have a clear and more scientific definition	Changed/ Answered	See answer to comment #12.
56. Ln 345: 'Cooling spots inhibit a greater active layer thickness in summer' – is this the technical definition? It comes across as difficult to interpret. An alternative version of this: 'Cooling spots result in shallower active	Changed/ Answered	We appreciate your suggestion and have revised the sentence accordingly to improve clarity:

layers in summer compared to warming spots.’		“Cooling spots result in shallower active layers in summer compared to warming spots.” Additionally, we have provided a detailed definition of cooling and warming spots in response to comment #12 to ensure consistency throughout the manuscript.
57. Figure 8 - Nice figure. The delineations are helpful. Similar delineations would help the interpretation of results in prior figures	Answered	See answer to comment #48.
58. Section 5.2: This section should be revised thoroughly – see previous comments on lidar snow depth and RF model comparison	Answered	We have revised this section based on our new results.
59. Ln 354: Luo and Panda studies were based on satellite remotely sensed snow cover – not sure I understand the link to UAS-lidar observations. Also, not clear what ‘not in depth post-processed data’ is. I did not understand the transition of the discussion from snow depth to snow cover	Changed	Thank you for pointing this out. We acknowledge that the studies by Luo et al. (2022) and Panda et al. (2022) focus on satellite-based snow cover observations rather than snow depth. To avoid confusion, we have removed these references in this context and ensured that our discussion remains focused on snow depth mapping. Additionally, since our revised model approach now utilizes a DTM instead of a DSM, we have removed the statement regarding “not in-depth post-processed remote sensing data” to accurately reflect the improved data processing methodology.
60. Ln 363-364: How did manual probing address the issue of vegetation? The uncertainty in these observations was never discussed	Answered	Thank you for highlighting this important point. We acknowledge that the impact of vegetation on manual snow depth probing was not explicitly discussed. In our study, manual probing was conducted with a heavy yardstick, which allowed us to reach the ground despite the presence of vegetation. However, we recognize that vegetation, particularly tall grasses and shrubs, can introduce uncertainties in snow depth measurements. In areas with denser vegetation, there is a possibility that the probe may not always reach the exact ground surface, leading to slight overestimations of snow depth. To address this, we have expanded our discussion on potential uncertainties in manual snow depth measurements and their implications for model accuracy.
61. Ln 366-368: This doesn't seem like only a lidar limitation - but a measurement challenge in general. Measuring snow over dense vegetation with air voids, compression, etc.. is always challenging. New approaches to correct the lidar based on the underlying vegetation type/density/height may improve lidar snow depth products.	Answered	We agree that the challenges of measuring snow depth over dense vegetation are not solely a limitation of LiDAR but rather a general measurement issue. We acknowledge that new approaches, such as correcting LiDAR-based snow depth estimates based on vegetation type, density, and height, could improve the accuracy of these products. We briefly addressed this in the discussion and highlight it as a potential avenue for future research.

62. Ln 375-387: The discussion in this paragraph was strong, and it was easier to follow the logic. This could be an example to use when revising the discussion.	Answered	Thank you! We have now used this paragraph as example for revising the discussion.
63. Ln 392, 407-408: Why was vegetation not removed from the summer point cloud? I do not understand why this was done in this manner. This step is critical for snow depth mapping with lidar.	Changed/ Answered	See Appendix A.
64. Ln 395-397: There is a growing body of literature on this that would be useful to review. See Buhler 2016, 2017; Adams et al., 2018; Avanzi et al., 2018, Cho et al., 2024 (Preprint), Eker et al. 2019; Harder et al. 2020 (compares lidar and RGB)	Answered	We have reviewed the recommended literature and incorporate relevant findings or ideas if they make a meaningful contribution to the context of our study.
65. Much of the discussion relies on findings from other studies and assumed links to snow depth observed in this study to conclude – not clear to me what value the work presented here has to understanding palsa permafrost dynamics more than point observations on a transect across one of these features would. Related to previous comments on reframing and refocusing the research objectives	Answered	Thank you for your comment. We recognize that directly linking our modelled snow distribution to permafrost dynamics remains a complex challenge. However, we believe that our study offers significant value beyond point transect measurements by providing the first spatially continuous snow depth maps over palsas using validated LiDAR and RF-based approaches. Our results show that these models perform well compared to independent validation datasets, confirming the reliability of the derived snow depth distributions. Given the crucial role of snow in regulating permafrost stability, we argue that these spatial datasets provide valuable insights into the potential snow-induced thermal dynamics of palsas. While additional ground-based validation of permafrost responses would strengthen this link, our study provides an important foundation for future research in this area. We have clarified these points in the discussion to better emphasize the unique contribution of our study.
66. Ln 412-414: A fewer number of folds should be used in the model training/validation	Changed/ Answered	See Appendix A.
67. Ln 421, 425-426: A large number of input features are used in this model and the results as presented show nearly perfect model performance – are you suggesting others should be included? If others could make the model better, why were they not included?	Answered	Thank you for your comment. With our revised model approach, we now use only 12 input parameters, ensuring a more streamlined and interpretable model. Our intention was not to suggest that additional parameters should necessarily be included in this study, but rather to acknowledge that future research could explore further potentially relevant predictors. For instance, more detailed vegetation classifications - such as specific vegetation types or density indices - could enhance snow depth modelling. Additionally, there may be other influential parameters that are not directly linked to snow depth but still play a role in snow distribution patterns.

		Identifying such factors would require a dedicated study focused on assessing and selecting the most critical parameters for snow depth modelling. We have clarified this point in the discussion to ensure that our statement is not misinterpreted as a recommendation for additional parameters in the current model.
68. Ln 423-424: Good point		Thank you!
69. Once noted challenges throughout are addressed – the discussion should be re-written to align with the updated manuscript	Answered	After all changes have been made, we have adapted the discussion based on the updated manuscript.
Conclusions		
70. As presented, the paper is focused on the evaluation of the methods for snow depth mapping and on the predictors that control the depth distribution -- discussion into the influence of these characteristics on the thermal profiles is purely assumption based -- thus reframing the conclusion in line with the revised paper and the actual results/data presented will be critical in the revised version.	Answered	Based on all the revisions and refinements made throughout the manuscript, we have rewritten the conclusion to align more clearly with the revised focus of the paper and the actual results presented. This ensures that our conclusions remain grounded in the data and analyses conducted.

Appendix

In this section, we provide additional information addressing comments #4, 5, 10, 16, 22, 23, 24, 44, 49, 50, 54, 56 from Reviewer 1 and #5, 6, 8, 10, 38, 40, 41, 52, 53, 63, 66 from Reviewer 2.

We sincerely appreciate your insightful comments and suggestions, which have significantly contributed to improving both the modelling approach and the overall quality of the manuscript.

Appendix A

To ensure high-quality modelling results and accurate snow depth distribution maps derived from UAS-LiDAR, we implemented your recommendations, including the removal of vegetation from the LiDAR-derived products and a re-evaluation of the modelling approach.

Additionally, we incorporated hyperparameter tuning and cross-validation to determine the most suitable parameter settings for the Random Forest model. To further improve model robustness and prevent overfitting, we also adjusted the data splitting strategy by testing the RF model on an independent external dataset.

1. Removal of vegetation from UAS-LiDAR DSM

Our initial decision to retain vegetation in the modelling process assumed that small and dense vegetation, as present in our study sites, is difficult to remove - even from point clouds. Testing several vegetation filter algorithms, such as the *Cloth Simulation Filter* (CSF) and *Statistical Outlier Removal* (SOR) in CloudCompare, confirmed this assumption, as the vegetation was not properly removed in the resulting products.

Additionally, we considered that vegetation significantly influences snow depth distribution by enhancing snow retention capacity. Therefore, we initially decided to include vegetation in the modelling process, expecting it to be beneficial for RF modelling.

However, based on your suggestions, we tested the *Progressive Morphological Filter* (PMF) Algorithm as described by Zhang et al. (2003) and Jacobs et al. (2021) and obtained satisfactory results with an effective removal of vegetation. We applied PMF filtering using the following parameters:

- Window sizes: 0.5, 1, 2, and 3
- Thresholds: 0.05, 0.1, 0.3, and 0.5

The extracted ground and vegetation points were saved in point cloud format. Using CloudCompare, we generated a DTM for each palsa using the Rasterize function. Empty cells within the point clouds were interpolated with a triangle max edge length value of 5.0.

The newly created DTMs were then used to recalculate the snow depth distribution for all three test sites in GIS, following the methodology described in the manuscript. In our initial calculations, all negative values were set to zero. However, in this revised approach, we retained negative values to highlight areas where either the LiDAR sensor produced inaccuracies or surface degradation occurred between the summer and winter flights.

Based on these refined DTMs, we recalculated all input parameters used in the final RF model run in SAGA GIS. The following 12 parameters were included: *Aspect, Elevation, Channel Network Base Level, Channel Network Distance, Negative Openness, Positive Openness, Relative Slope Position, Slope, Topographic Position Index, Valley Depth, Wind Effect, Wind Exposition*.
A detailed description of these parameters is provided in Table 2. We have now focused on a single model run, and accordingly, we have removed descriptions of other parameters from the manuscript to ensure clarity and consistency.

2. Splitting data into training and test datasets

In the initial study design, we used the entire buffered $SD_{in\text{-}situ}$ dataset to extract the input parameters from the raster stack, resulting in a data frame with 5222 points. We then split this dataset into 70% training and 30% test data. However, this approach introduced a risk of overfitting, as each $SD_{in\text{-}situ}$ point was represented an average of 28 times in the dataset. Consequently, many points appeared in both the training and test datasets, reducing the independence of the validation process.

To address this issue, we revised our study design by first separating 70% of the point features from each $SD_{in\text{-}situ}$ dataset for training and 30% for testing. Only after this separation did we extract the input parameter values for the training dataset, ensuring a clear distinction between training and validation data. The test dataset was reserved exclusively for model validation. The following extract from the R script illustrates these steps:

```r
#===== Function to split training and test dataset =====
split_shapefile <- function(shp) {
  set.seed(42)  # Ensure reproducibility
  num_samples <- nrow(shp)  # Get the number of samples
  train_indices <- sample(num_samples, size = round(0.7 * num_samples))  # Select 70% of the samples for training
  test_indices <- setdiff(1:num_samples, train_indices)  # The remaining 30% for testing

  shp_train <- shp[train_indices, ]  # Create training dataset
  shp_test <- shp[test_indices, ]  # Create test dataset

  return(list(train = shp_train, test = shp_test))  # Return the split datasets as a list
}

# Splitting the dataset for all three locations
split_pousu <- split_shapefile(shp_pousu)
split_peera <- split_shapefile(shp_peera)
split_puolikkoniva <- split_shapefile(shp_puolikkoniva)

# Combine training and test datasets for all Palsas
shp_train_all <- rbind(split_pousu$train, split_peera$train, split_puolikkoniva$train)  # Merge training datasets
shp_test_all <- rbind(split_pousu$test, split_peera$test, split_puolikkoniva$test)  # Merge test datasets
```

After extracting the input parameters from the raster stack, the final dataset consisted of:

- Training dataset: 3,645 points (Puolikkoniva: 1,983; Pousu: 905; Peera: 757)
- Test dataset: 1,577 points (Puolikkoniva: 836; Pousu: 401; Peera: 340)

To prevent errors and miscalculations, all NoData values were removed from the datasets, resulting in a final training dataset of 3,504 points and a final test dataset of 1,548 points for further modelling and validation.

3. Hyperparameter tuning and cross validation

To determine the optimal values for *mtry*, *min.node.size*, and *sample fraction*, we performed hyperparameter tuning using the *mlr* package in R (Bischl et al., 2016).

To prevent overfitting, we restricted the search range for *min.node.size* to 10–15 and for *sample fraction* to 0.7–0.85, following the recommendations of Probst et al. (2019) and Breiman (2001). Allowing an unlimited search range initially resulted in better model performance, but at the cost of reduced generalization, indicating signs of overfitting. We selected the final search range based on multiple test runs with different settings.

For cross-validation, we tested different fold sizes to identify the most effective configuration. The best results were achieved using a 4-fold cross-validation. The following R script extract provides details on the tuning process:

```r
#===== Hyperparameter Tuning with tuneRanger (Regression) =====

# Define the regression task
task <- makeRegrTask(data = all_train, target = "Class")

# Define the cross-validation strategy
cv_desc <- makeResampleDesc("CV", iters = 4)  # 4-fold cross-validation

# Define the Random Forest learner with hyperparameters as tuning options
learner <- makeLearner("regr.ranger", num.trees = 1000)

# Define the hyperparameter search space
param_set <- makeParamSet(
  makeIntegerParam("mtry", lower = 2, upper = ncol(all_train) - 1),  # Number of variables to consider at each split
  makeIntegerParam("min.node.size", lower = 10, upper = 15),  # Minimum number of observations per node
  makeNumericParam("sample.fraction", lower = 0.7, upper = 0.85)  # Proportion of samples used in each tree
)

# Define the tuning control (e.g., Bayesian optimization or random search)
control <- makeTuneControlRandom(maxit = 70)  # 70 iterations for tuning

# Hyperparameter tuning with cross-validation
tuned_params <- tuneParams(
  learner = learner,
  task = task,
  resampling = cv_desc,  # 4-fold CV
  par.set = param_set,
  control = control,
  measures = rmse  # Root Mean Squared Error as the performance metric
)

# Display results
print(tuned_params)

# Best Random Forest model with tuned parameters
best_learner <- setHyperPars(learner, par.vals = tuned_params$x)
```

The final tuned hyperparameters were as follows:

- mtry: 9
- min.node.size: 10
- sample fraction: 0.79

4. Permutation Importance (PI)

In our initial study design, we conducted the RF modelling once and directly used the permutation importance (PI) values provided by the model.

In our revised approach, we refined this process by repeating the calculation 100 times to obtain a mean PI value for each input parameter, ensuring more robust and reliable importance rankings.

The following R script extract details the implementation of this approach:

```r
#========== Permutation Importance
num_repeats <- 100
importance_values <- matrix(NA, nrow = num_repeats, ncol = ncol(all_train) - 1)

for (i in 1:num_repeats) {
  cat("Iteration:", i, "\n")

  # Train the model using the identical hyperparameters from tuning
  temp_model <- ranger(
    x = all_train[, -ncol(all_train)],
    y = all_train$Class,
    mtry = tuned_params$x$mtry,  # Optimized mtry value
    min.node.size = tuned_params$x$min.node.size,  # Optimized min.node.size
    sample.fraction = tuned_params$x$sample.fraction,  # Optimized sample.fraction
    num.trees = 1000,
    importance = "permutation",
    seed = i  # Different seed per run for robustness
  )

  # Store the feature importances in the matrix
  importance_values[i, ] <- importance(temp_model)
}

# Compute the mean Permutation Importance over the 100 runs
mean_importance <- colMeans(importance_values)
```

We modified Figure 7 to display only the 12 selected parameters along with their respective mean PI values over 100 iterations. Additionally, we normalized the values, setting the most important parameter (Topographic Position Index) to 1.

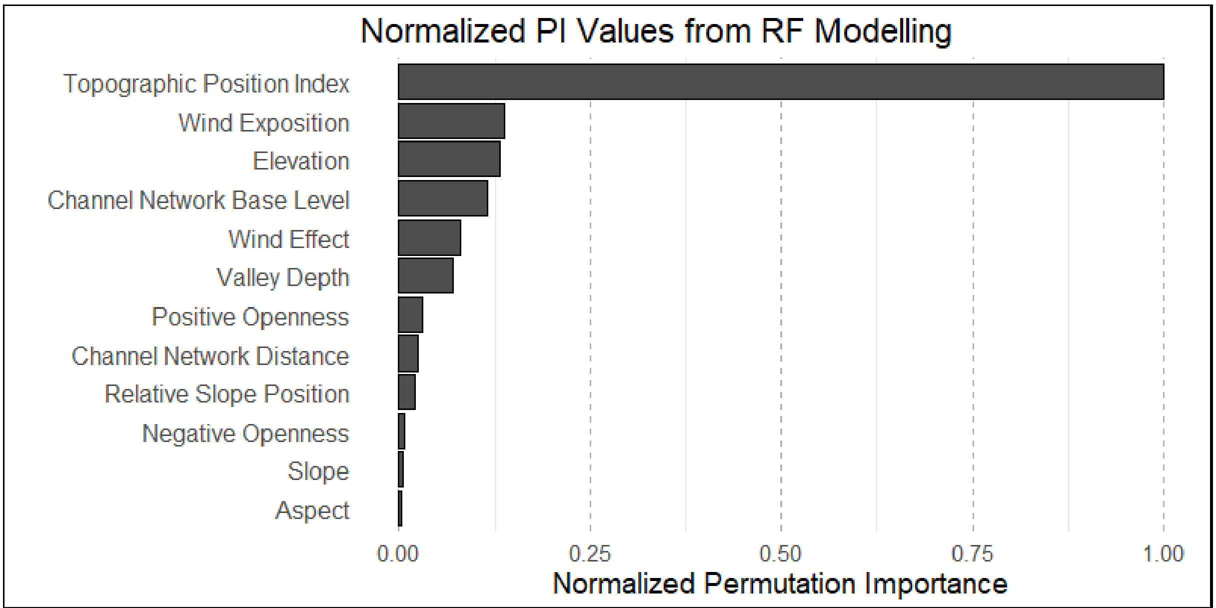

Figure 7. Overview of normalized mean Permutation Importance values from RF modelling over 100 iterations.

5. Final results and validation

Both the RF-based and UAS-LiDAR-based results were validated using the initially separated test dataset. Additionally, we conducted three further RF model runs, where in each iteration, two palsa sites were used as the training dataset, and one was used as the test dataset. This approach further validated the generalization capability of the model.

The validation results indicate that the RF-based approach now exhibits lower peak accuracies compared to the initial study design. However, by reducing overfitting, the results are more plausible and robust, while still achieving high accuracy and outperforming the UAS-LiDAR-based approach:

Table 3. Overview of the calculated Root Mean Square Error (RMSE), Coefficient of Determination (R2), Mean Absolute Error (MAE) and Standard Deviation (SD) for RF- and UAS-LiDAR-based snow depth estimations. Additionally, external validation results (RMSE and R^2) for RF-based snow depth at each palsa site (Peera RF, Pousu RF, Puolikkoniva RF) are provided.

Parameter	***RF***	***LiDAR UAS***	***Peera RF***	***Pousu RF***	***Puolikkoniva RF***
RMSE	18.33	23.49	16.67	21.31	27.13
R^2	0.77	0.691	0.628	0.767	0.578
MAE	13.26	17.49	-	-	-
SD	18.11	20.84	-	-	-

We recalculated all metrics for different point groups and included the number of points per group. These groups were classified visually, based on orthophotos, slope data, and elevation characteristics of the respective locations.

The results show that the accuracy differences between RF and UAS-LiDAR-based approaches are now less pronounced. However, in certain categories, such as *Thermokarst* and *Open Area*, the UAS-LiDAR-based results show lower accuracy, likely due to measurement inaccuracies caused by water surfaces and irregularities in areas with higher vegetation.

Table 4. Overview of RMSE, R^2, MAE and SD divided by validation point locations within the investigation areas.

	RMSE		**R^2**		**MAE**		**SD**	
	RF	*LiDAR*	*RF*	*LiDAR*	*RF*	*LiDAR*	*RF*	*LiDAR*
On Top (n = 69)	8.33	8.33	0.841	0.730	3.84	3.84	8.32	10.83
Edge (n = 66)	13.12	13.12	0.894	0.768	5.85	5.85	12.82	19.09
Thermokarst (n = 16)	10.99	33.73	0.893	0.592	5.42	30.35	10.69	25.08
Open Area (n = 26)	4.54	14.23	0.926	0.519	1.56	9.84	4.40	12.59

Figures 5, 6, 8, and 9 have been updated based on the new results.

Figure 5 now includes the recalculated snow depth maps. We have incorporated all areas where SD_{LiDAR} values are below 0, visualizing these parts in red to highlight regions where the LiDAR sensor may have measured incorrectly or where degradation has occurred between flights.

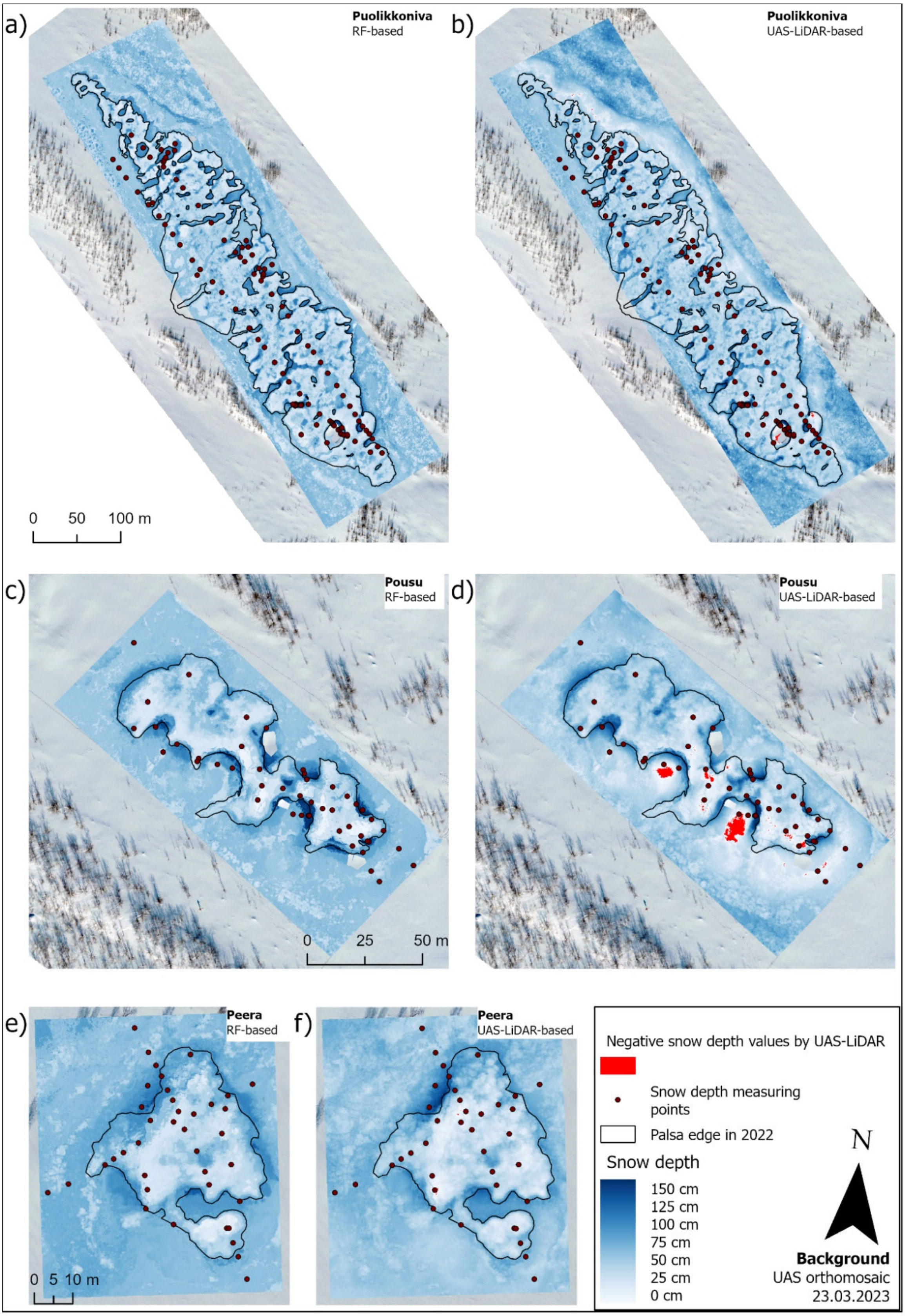

Figure 5. Snow depth predictions based on the RF model (left) and the UAS-LiDAR (right) at site Puolikkoniva (a, b), Pousu (c, d) and Peera (e, f) palsas. Red points are showing the in-situ snow depth measurement locations.

In Figure 6 we inserted the new calculated difference maps and we also included the parts with negative values in red:

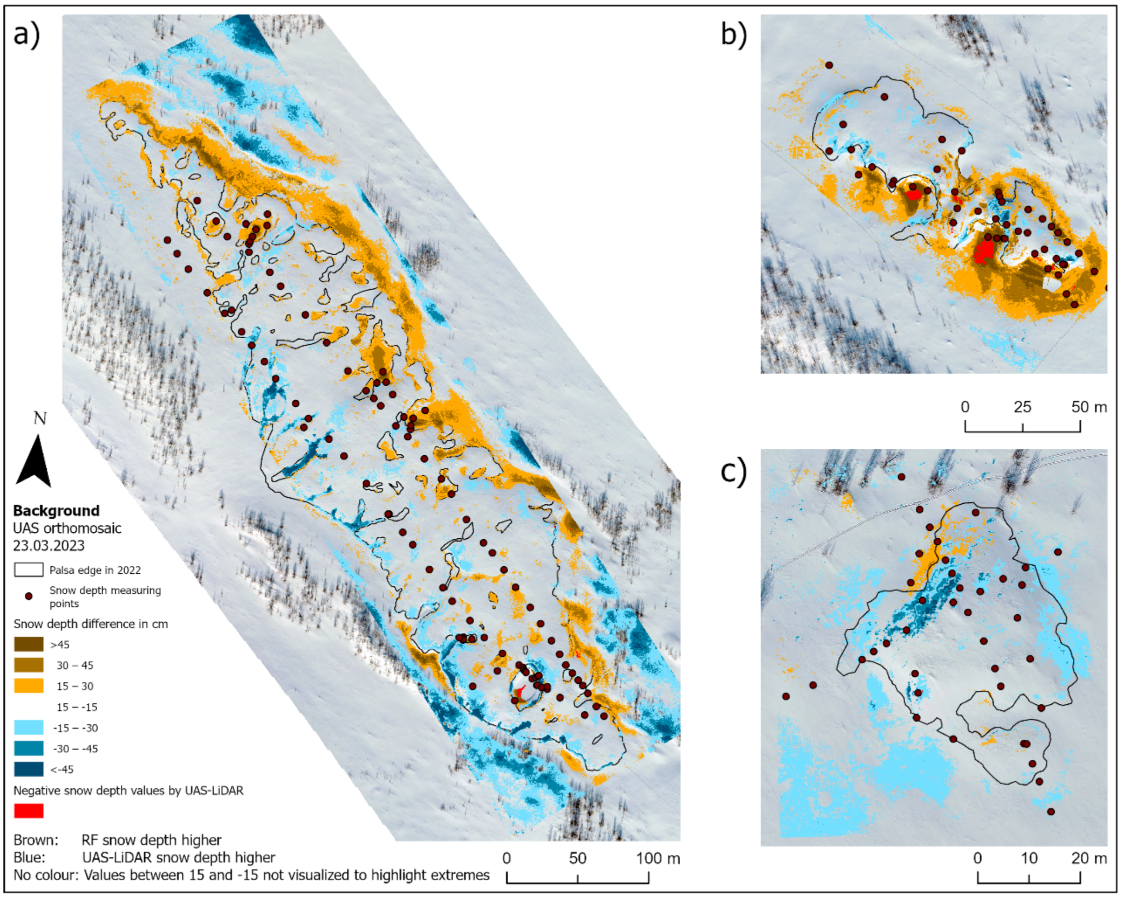

Figure 6. Snow depth differences between modelled and UAS LiDAR results at a) Puolikkoniva, b) Pousu and c) Peera palsas.

Figure 8 shows the scatter plots based on the 30% test dataset. Here we used only the single values of the $SD_{in-situ}$, not considering the values within the buffer areas of the test data. We decided to do it like that, to obtain a very fine validation of both methods:

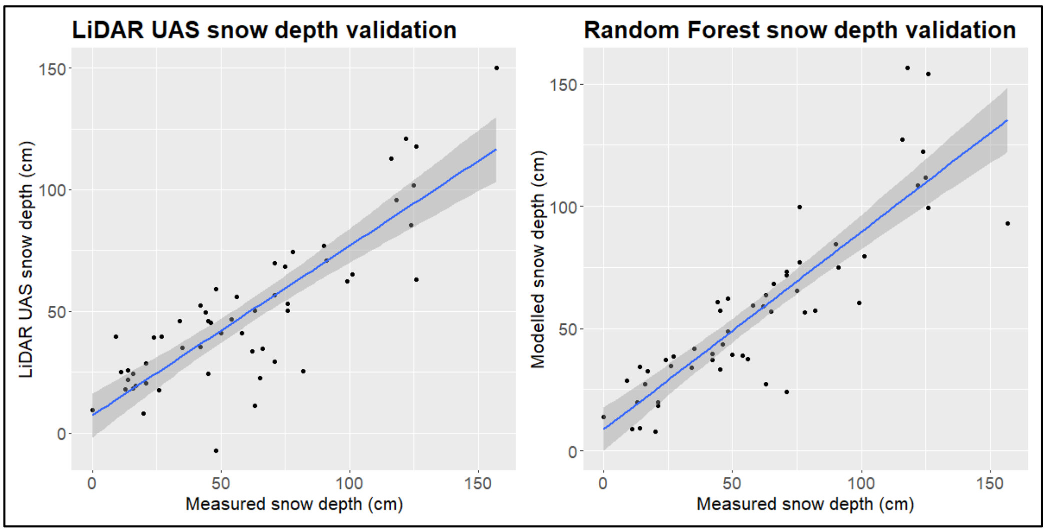

Figure 8. Scatter plots with regression lines for UAS-LiDAR-derived and RF-modelled snow depths, based on the external test dataset.

Figure 9 has been updated to reflect the new results. Additionally, we have incorporated the calculated slope derived from the DTM of Pousu palsa.

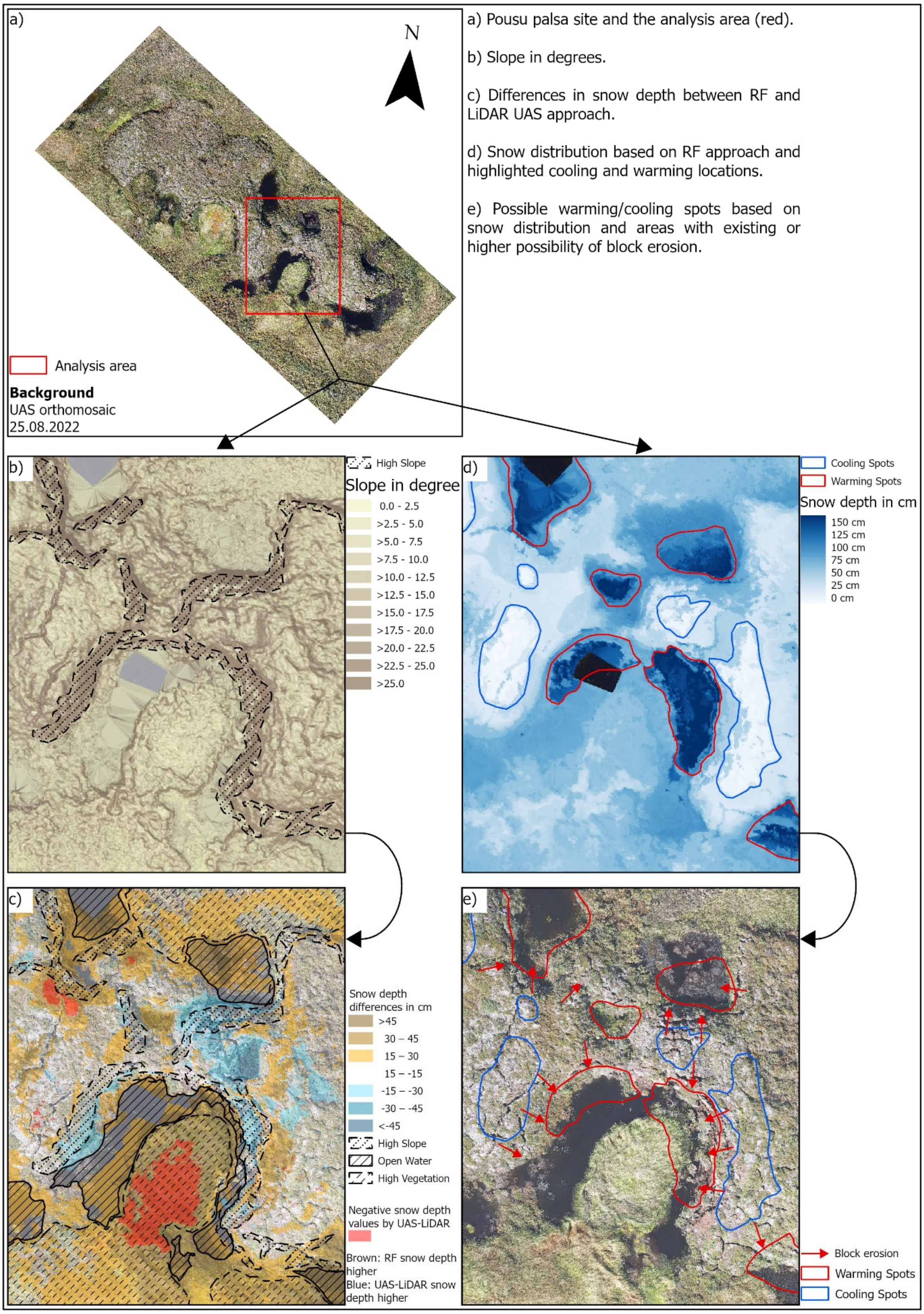

Figure 9. Explanation of differences between UAS LiDAR-derived and RF-modelled snow depths.

Appendix B

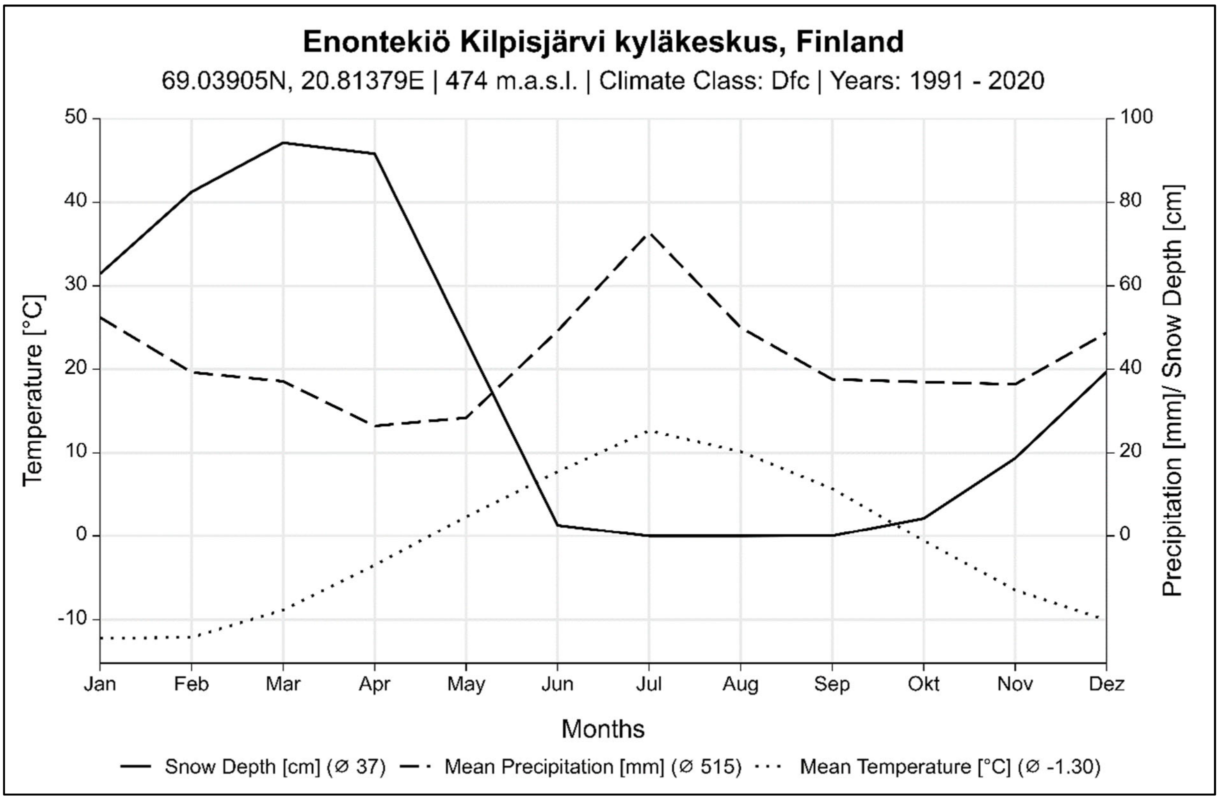

Figure 2. Climate chart of Kilpisjärvi (FMI, 2022). Dotted line shows 2 m above ground temperature in °C, dashed line shows precipitation in mm and solid line shows snow depth in cm.

Appendix C

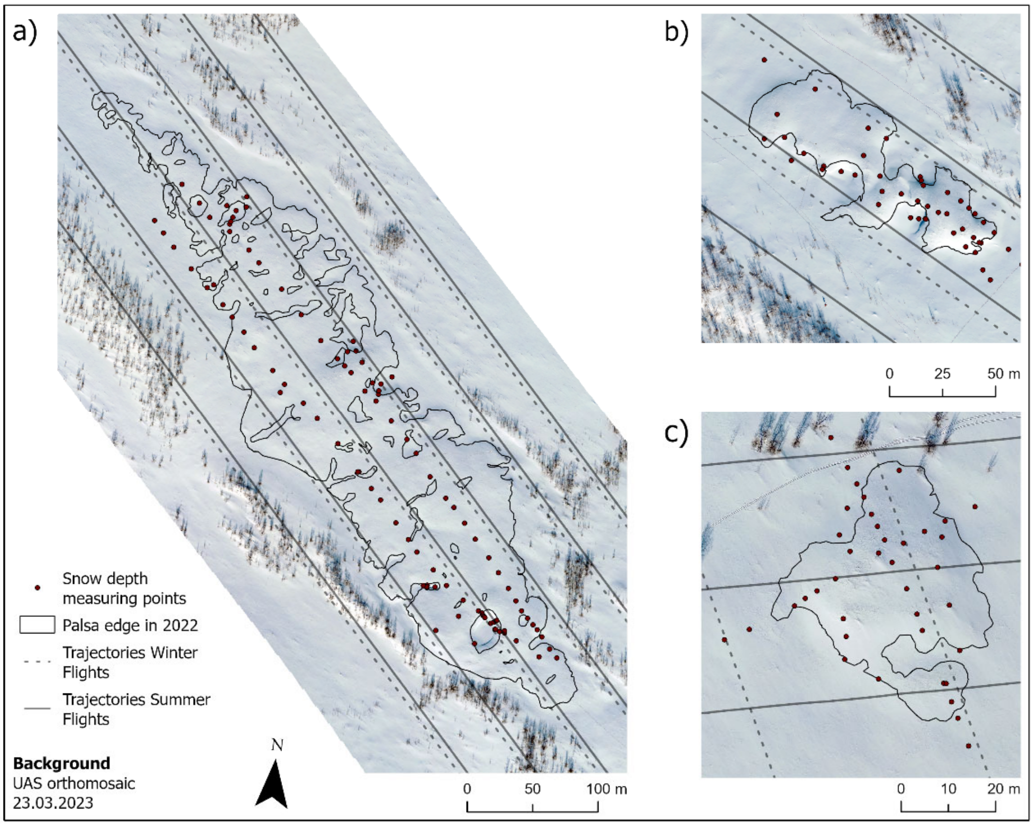

Figure 4. Snow depth measuring points within the investigation sites at Puolikkoniva (a), Pousu (b) and Peera (c) palsa illustrating different methods for recording snow depth (transects, randomized, crossed).

Appendix D

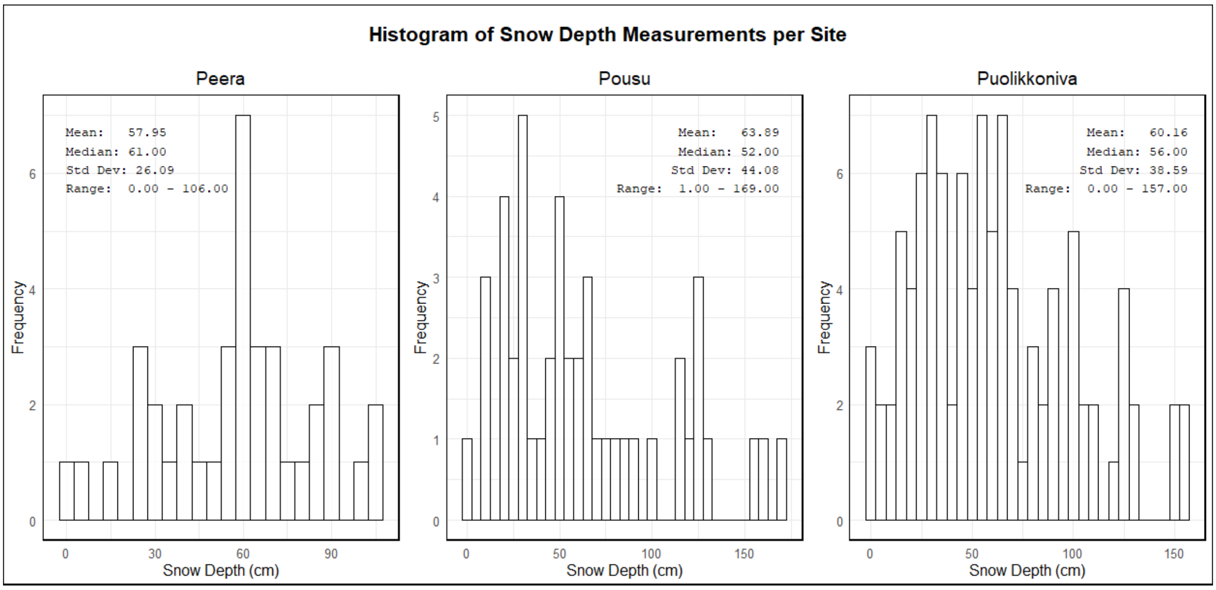

Figure A1. Histogram of $SD_{in\text{-}situ}$ points and respective statistics per palsa site.

Appendix E

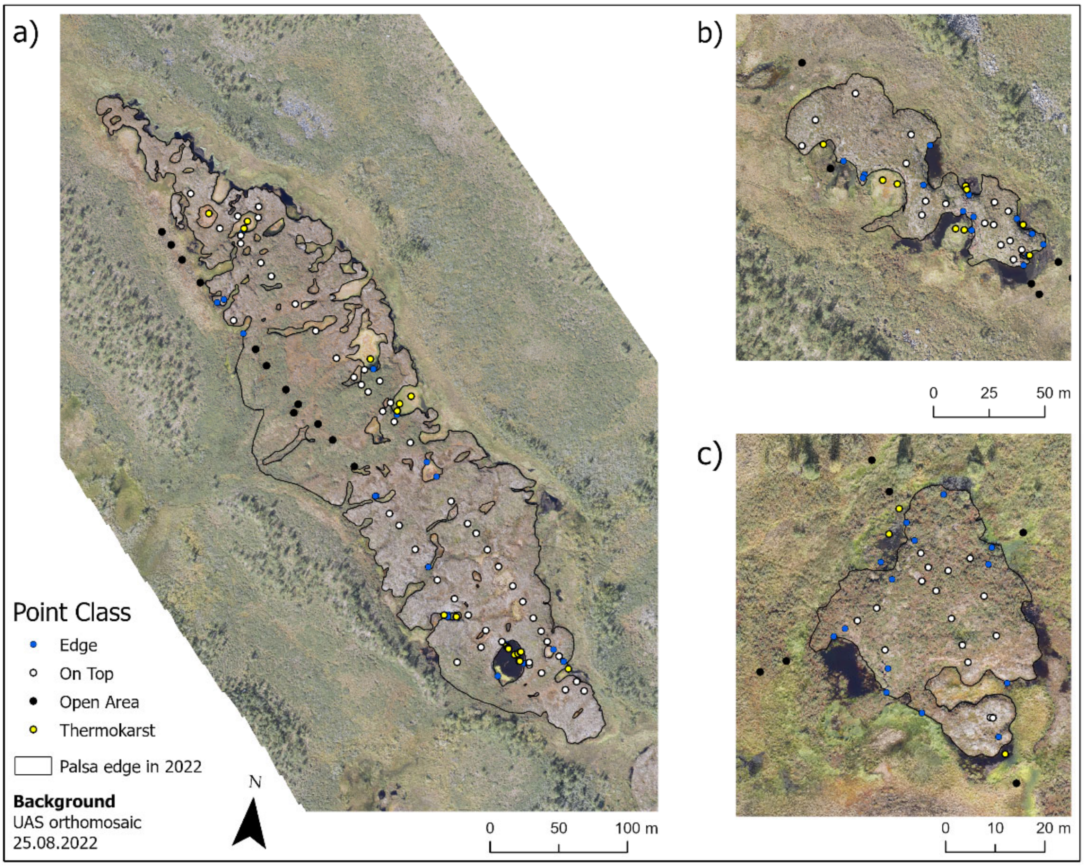

Figure A2. Overview of classification of all $SD_{in\text{-}situ}$ points into classes Edge, On Top, Open Area and Thermokarst.

Appendix F

Table A3. Correlation between each input parameter and RF-modelled snow depth.

Parameter	Correlation to SD_{RF}	Parameter	Correlation to SD_{RF}
Aspect	0.09	Relative Slope Position	-0.49
Elevation	-0.12	Slope	0.08
Channel Network Base Level	-0.09	Topographic Position Index	-0.87
Channel Network Distance	-0.45	Valley Depth	0.50
Negative Openness	0.22	Wind Effect	-0.55
Positive Openness	-0.50	Wind Exposition	-0.80

References

Avanzi, F., Zheng, Z., Coogan, A., Rice, R., Akella, R., & Conklin, M. H. (2020). Gap-filling snow-depth time-series with Kalman Filtering-Smoothing and Expectation Maximization: Proof of concept using spatially dense wireless-sensor-network data. *Cold Regions Science and Technology*, *175*. https://doi.org/10.1016/j.coldregions.2020.103066

Bergamo, T. F., Sampaio de Lima, R., Kull, T., Ward, R. D., Sepp, K., & Villoslada, M. (2023). From UAV to PlanetScope: Upscaling fractional cover of an invasive species Rosa rugosa. *Journal of Environmental Management*, *336*. https://doi.org/10.1016/j.jenvman.2023.117693

Bischl, B., Lang, M., Kotthoff, L., Schiffner, J., Richter, J., Studerus, E., Casalicchio, G., & Jones, Z. M. (2016). mlr: Machine Learning in R. In *Journal of Machine Learning Research* (Vol. 17). https://github.com/mlr-org/mlr

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*, 5–32.

Harder, P., Pomeroy, J. W., Helgason, W. D., & Helgason, W. D. (2020). Improving sub-canopy snow depth mapping with unmanned aerial vehicles: Lidar versus structure-from-motion techniques. *Cryosphere*, *14*(6), 1919–1935. https://doi.org/10.5194/tc-14-1919-2020

Jacobs, J. M., Hunsaker, A. G., Sullivan, F. B., Palace, M., Burakowski, E. A., Herrick, C., & Cho, E. (2021). Snow depth mapping with unpiloted aerial system lidar observations: A case study in Durham, New Hampshire, United States. *Cryosphere*, *15*(3), 1485–1500. https://doi.org/10.5194/tc-15-1485-2021

Leppiniemi, O., Karjalainen, O., Aalto, J., Luoto, M., & Hjort, J. (2023). Environmental spaces for palsas and peat plateaus are disappearing at a circumpolar scale. *The Cryosphere*, *17*(8), 3157–3176. https://doi.org/10.5194/tc-17-3157-2023

Luo, J., Dong, C., Lin, K., Chen, X., Zhao, L., & Menzel, L. (2022). Mapping snow cover in forests using optical remote sensing, machine learning and time-lapse photography. *Remote Sensing of Environment*, *275*, 113017. https://doi.org/10.1016/J.RSE.2022.113017

Luoto, M., Heikkinen, R. K., & Carter, T. R. (2004). Loss of palsa mires in Europe and biological consequences. In *Environmental Conservation* (Vol. 31, Issue 1, pp. 30–37). https://doi.org/10.1017/S0376892904001018

Markkula, I., Turunen, M., & Rasmus, S. (2019). A review of climate change impacts on the ecosystem services in the Saami Homeland in Finland. In *Science of the Total Environment* (Vol. 692, pp. 1070–1085). Elsevier B.V. https://doi.org/10.1016/j.scitotenv.2019.07.272

Meriö, L. J., Rauhala, A., Ala-Aho, P., Kuzmin, A., Korpelainen, P., Kumpula, T., Kløve, B., & Marttila, H. (2023). Measuring the spatiotemporal variability in snow depth in subarctic environments using UASs - Part 2: Snow processes and snow-canopy interactions. *The Cryosphere*, *17*(10), 4363–4380. https://doi.org/10.5194/tc-17-4363-2023

Panda, S., Anilkumar, R., Balabantaray, B. K., Chutia, D., & Bharti, R. (2022). Machine Learning-Driven Snow Cover Mapping Techniques using Google Earth Engine. *INDICON 2022 - 2022 IEEE 19th India Council International Conference*. https://doi.org/10.1109/INDICON56171.2022.10040153

Probst, P., Wright, M. N., & Boulesteix, A. L. (2019). Hyperparameters and tuning strategies for random forest. In *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* (Vol. 9, Issue 3). Wiley-Blackwell. https://doi.org/10.1002/widm.1301

Quante, L., Willner, S. N., Middelanis, R., & Levermann, A. (2021). Regions of intensification of extreme snowfall under future warming. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-95979-4

Rauhala, A., Meriö, L. J., Kuzmin, A., Korpelainen, P., Ala-Aho, P., Kumpula, T., Kløve, B., & Marttila, H. (2023). Measuring the spatiotemporal variability in snow depth in subarctic environments using UASs - Part 1: Measurements, processing, and accuracy assessment. *The Cryosphere*, *17*(10), 4343–4362. https://doi.org/10.5194/tc-17-4343-2023

Zhang, K., Chen, S. C., Whitman, D., Shyu, M. L., Yan, J., & Zhang, C. (2003). A progressive morphological filter for removing nonground measurements from airborne LIDAR data. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(4 PART I), 872–882. https://doi.org/10.1109/TGRS.2003.810682