Author responses to comments of reviewer 2

Quantifying the volume of frozen water stored as seasonal mountain snow is a fundamental topic in hydrology and water resources management. The authors address the research need by modeling snow depth (SD) at in situ locations and for spatially distributed applications, using an XGBoost machine learning algorithms and forcing it with Sentinel-1 dual polarized synthetic aperture radar (PolSAR) data products and traditional inputs (e.g., precipitation , elevation, slope, aspect). The novelty of the work is using the PolSAR data in the ML workflow, and rigorously assessing the information gained by comparing to existing ML SD modeling workflows through model performance metrics, feature importance, and SHAP analysis-for a total of 5 ML SD models. The results indicate that the PolSAR features improved model performance at in situ and spatially distributed SD modeling locations, often a key model input in the feature importance assessments but behind meteorological and fraction snow cover area products. A key element of the manuscript is the three-fold nested cross-validation technique used for model evaluation. It is rigorous and establishes a standard for others to follow as the method supports an independent assessment of the model's spatial, temporal, and spatiotemporal prediction skill.

While the manuscript provides a publishable contribution to the hydrologic and cryosphere modelling community, the manuscript is quite long and could be much more focused. Primarily, this is the results and discussion section that is 9 full pages. While the information is useful, many of the case-by-case examples could be placed in a supplementary information section and the manuscript could communicate the key findings of the study and discuss. I suggest a minor revision, emphasizing a restructuring of the Results/Discussion section to have it more effectively communicate the key results and discuss their relevance to the snow modeling community.

First, we would like to thank the reviewer for the useful comments, which helped us improve the manuscript. Below, we listed our replies to the comments in red. The original reviewer comments are included in black, new text in the manuscript is shown in blue, with the indicated line numbers referring to the revised manuscript. As response to the main comment on the manuscript length, we have shifted some of the case-by-case examples to the Appendix. As a result, this section now comprises around 2 pages less text and is more focused towards the key findings.

Main Comments

Comment 1
The introduction has great information but could improve in flow. Key research gaps are shared throughout the section, rather than the gaps being shared at the conclusion. Revising the introduction to have each paragraph build on the prior and clearly state the research gap(s) identified in the literature review would improve the section's flow and support reader comprehension.

The introduction has been revised. Necessary missing references have been added, and the paragraphs have been restructured to improve flow and readability.

Comment 2
Additionally, I suggest additional commentary in the figure captions, highlighting the key takeaways in addition to the description of the labels. Having the authors highlight the key findings in the figure captions will support a more detailed understanding of the work for a broader audience. Lastly, I request that figure 8 includes the No Survey mapped product. Adding this plot to the figure would improve the understanding of the spatial performance of the in situ trained model for spatial prediction.

We modified the captions in the revised manuscript to highlight key findings. Next, Figure 8 of the revised manuscript includes the "No Survey" mapped product. The revised manuscript also includes a map of SD estimates across the whole study area, generated with the models obtained from the spatial nested cross-validation approach.

Comment 3
Much of the manuscript results/discussion section would be appropriate for a supplementary material section so that the main document could be clearer and concise, aiding in interpretation to a broader audience.

Many of the case-by-case examples have been moved to the Appendix. For example, we moved Section "4.2.2 Time series of SD and SHAP values" entirely to the appendix. Next, we moved the more detailed discussion of the 16 March 2017 Dischma valley snow survey to a new Appendix "Appendix D: 16 March 2017 Dischma valley snow survey SD estimates", where the results are discussed based on two new figures to resolve many of the "not shown" instances (in response to concerns from Reviewer 1).

Comment 4
The abstract could highlight the impacts more, rather than saying what is done.

We have modified the abstract in the revised manuscript, highlighting more results and how these results relate to the last sentence of the original abstract. Therefore, we changed the abstract to:

L1-18 "Seasonal mountain snow is an indispensable resource, but accurate estimates of this water storage remain limited, even in the European Alps, where there is a dense network of in situ monitoring stations. In this study, we address Alpine snow depth estimation at a 100 m spatial resolution and sub-weekly temporal resolution over the 2015–2024 period using multiple input configurations within an extreme gradient boosting (XGBoost) machine learning (ML) model. We explore the potential of Sentinel-1 C-band dual-polarized synthetic aperture radar polarimetry (PolSAR) observations, and include either regionally downscaled meteorological forcing data or modeled snow depth as additional inputs to further explain interannual and spatial variability. A threefold nested cross-validation scheme is used to account for the spatio-temporal dependencies present in the snow depth data. XGBoost's internal booster and Shapley additive explanation (SHAP) values are used to relate the input features with the predictions for both dry and wet snow conditions. Our results indicate that the inclusion of PolSAR observations leads to modest improvements over a backscatter-intensity-based configuration, whereas the SHAP-based feature attribution reveals a high reliance of XGBoost on the polarimetric scattering angle and co-polarized (VV) backscatter intensity. Next, incorporating either meteorological forcing data or modeled snow depth substantially enhances predictive performance, particularly when spatially distributed training data, proven to be essential for capturing topographic controls on snow depth variability, are included. When supplemented with spatial training data and either meteorological forcing data or modeled snow depth estimates, XGBoost shows good agreement with nine snow surveys conducted in the Dischma valley (Switzerland), achieving correlation coefficients (R) of 0.76 and 0.78 and mean biases of 0.07 and 0.17 m, respectively. When applied to unseen locations across the Alps, the performance remains high, with R = 0.80 and biases of −0.04 and −0.03 m, respectively."

Comment 5
The abstract shares a complete summary, but covers too much. It could be more concise with respect to the main experiments, results, and conclusions.

Within the revised abstract, we put more focus on the key takeaways. We still opted to include the main experiments conducted, to give the reader a complete overview of what to expect when consulting our manuscript.

Comment 6
The manuscript would substantially benefit from being more focused and highlighting key conclusions. Additional information could be referenced and placed in a supplemental information section.

Addressed by moving some of the detailed discussions to the appendix.

Comment 7
I suggest in Table 2 that the authors provide a definition of the variables (part of the ML configuration). This would support a broader audience's understanding of the work and provide a clear reference for variables mentioned throughout the manuscript.

Addressed.

Comment 8
In the conclusion (line 582-583), the authors state "These results showcase the potential of using PolSAR observations within non-machine learning applications, e.g., within a conceptual model." This is new information for the reader, and I do not see where the results support this conclusion.

We agree that this suggestion is not previously discussed throughout the manuscript and not necessarily relevant for the work presented in our manuscript. We have therefore chosen to remove this sentence.

Comments by line number

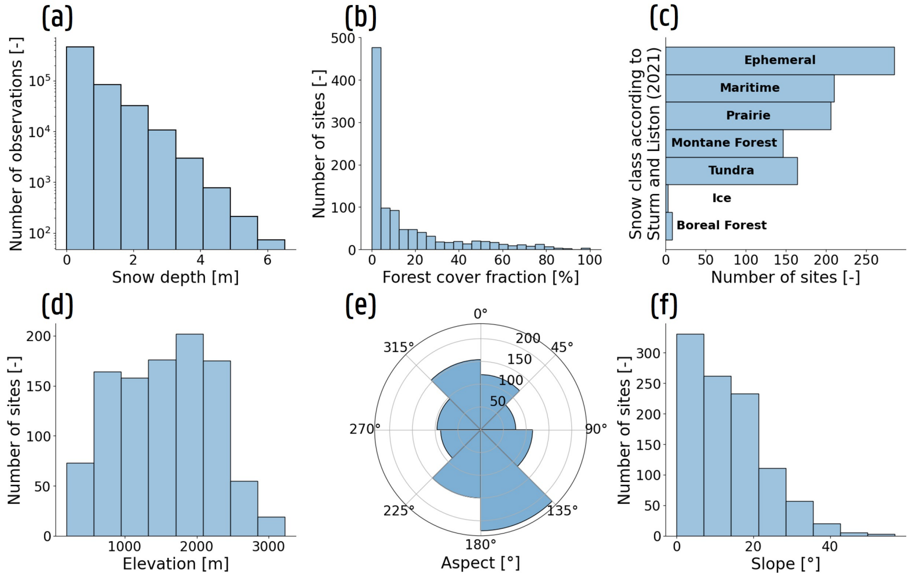
107-108, 118: How do the in situ and point-based measurements represent the variability of SD throughout the study domain (European Alps). What is the range in snow environments (e.g., Sturm snow classification), elevation, aspect, slope, etc observed in the data. Sharing the distribution of training data would support more nuanced conclusions related to the model output, such as providing more context to model skill at locations other than in situ locations.

The point-based in situ measurement sites shown in Fig. 1 of the original manuscript represent the snow classes defined in Sturm and Liston (2021) well, with the exception of the Ice and Boreal Forest classes (Review Fig. 1). However those two classes are not often observed within the European Alps and, as outlined on line 216 of the original manuscript, we excluded sites located in glacial areas following the Randolph Glacier Inventory version 7.0. Nonetheless, some sites that did not appear in the Randolph Glacier Inventory are classified as Ice by Sturm and Liston (2021). Next, the measurement sites represent most elevations found within the European Alps, with the exception of elevations above 3000 m, and the lowest elevations. The training dataset contains fewer east and west facing slopes, and has roughly the same amount of north and south facing slopes. Finally, the amount of sites with steep slopes is limited, as was described within the original manuscript (lines 524-526), and only 2% of the measured snow depths is $\geq$ 2.5 m.

In order to give a better overview of the distribution of the training data, we added Review Fig. 1 of this response file to the Appendix of the revised manuscript.

180-181: Can the authors provide a citation for their precipitation downscaling method?

Equation 1 of the original manuscript is an adaptation from Equation 2 in the study of Huss et al. (2013), who used meteorological forcing data to model accumulation and melt for two glaciers in Switzerland. Within the study, the authors corrected observed (in situ) precipitation data for gauge



Review Figure 1: Distributions of measured SD and static features of the in situ measurement sites used to train, validate and test XGBoost. (a) Distribution of measured SD. (b) Distribution of forest cover fraction of the unique measurement sites. (c), (d) (e) and (f), same as (b), but for the snow classes described in Sturm and Liston (2021), elevation, aspect and slope, respectively.

undercatch errors using a scaling factor c_{prep} and an altitudinal gradient. Within our study, however, we did not address gauge undercatch. Instead, we adapted equation 2 of Huss et al. (2013) to downscale coarse-scale precipitation data to account for orographic precipitation. To this end, we modified this equation to 1) focus on areas with major elevation differences (by introducing D_{dif} and limiting D to 1, so that Equation 1 within mountainous areas becomes $P_{\text{coarse}_{x,t}} \cdot \left[0.75 + 0.5\frac{z_x - z_{\text{min}}}{z_{\text{max}} - z_{\text{min}}}\right]$) and 2) by limiting the precipitation adjustments to 75-125% of the non-downscaled values. The original manuscript has been adapted as follows:

L192-201 "First, coarse-scale precipitation (P_{coarse}; [mm 3h^{-1}]) was corrected as a function of elevation to account for orographic effects, using a rescaling function adapted from Huss et al. (2013), who corrected in situ precipitation data for gauge undercatch in a glacier mass-balance study. Thus, for each location x within the 500 m grid at time step t, $P_{\text{coarse}_{x,t}}$ was downscaled using the following equation:

$$P_{x,t} = (1 - D) \cdot P_{\text{coarse}_{x,t}} + D \cdot P_{\text{coarse}_{x,t}} \cdot \left[0.75 + 0.5\frac{z_x - z_{\text{min}}}{z_{\text{max}} - z_{\text{min}}}\right] \text{ with } D = \frac{z_{\text{max}} - z_{\text{min}}}{D_{\text{dif}}} \quad 0 \leq D \leq 1 \tag{1}$$

with z_x the elevation [m] of the 500 m Copernicus GLO-30 DEM, z_{min} and z_{max} the minimum and maximum elevation [m] within an interpolation window — centered on the location x and spanning an area roughly matching the original 0.1° grid size — and D_{dif} a user defined difference in elevation, set to 250 m. The user defined difference was introduced to focus the corrections on the study area, with minor adjustments for areas with small elevation differences. Different from Huss et al. (2013), Equation 1 limits the downscaled precipitation values between 75 and 125% of the original $P_{\text{coarse}_{x,t}}$ values."

259: Table 2. Can the authors include a description of the model variable here? The would provide a reference for the reader as they move through the manuscript and want to know what variables the authors are related to model skill and feature importance.

We adjusted the caption of Table 2 in the revised manuscript, including a description of the variables.

362/493: Table 3. The authors mention Bias as a model evaluation metric, but it is not in any table or figure. Can the authors add Bias throughout?

We have included bias in the revised Figures and Tables.

508: “SHAP value FI”, I think this is referring to Figure 7. Can the authors clarify what FI figure to look at, or include it if it is not present?

The SHAP value FI referred to in line 508 does not refer to Fig. 7 within the original manuscript. We admit that this was not clear in the original manuscript. As part of other comments, we revised Fig. 6 of the original manuscript. The Figure now displays SHAP values for the three configurations for both the spatial and spatio-temporal nested CV schemes. From this revised Figure, our description in lines 508-511 of the original manuscript should be clear. Thereby, we omitted line 512 to focus the text more to how the S1 features are used during SD prediction.

534-536: The elevation bands do not provide meaningful information without connecting to a snow environment (e.g., alpine, tundra, sub-alpine…) as, depending on the location, this could be alpine or forest. I suggest coupling the elevation band ranges with a snow environment type. Lastly, 1000m is substantial and could cover several snow environments. A more representative snow environment communication other than elevation will be more impactful to the readers.

Thank you for this comment. As we have the snow class information from Sturm and Liston (2021) available, we have adapted this part of the discussion in the revised text. This discussion has been moved to the new Appendix ”Appendix D: 16 March 2017 Dischma valley snow survey SD estimates” (See main comment 3).

542: Figure 8: I suggest adding the No Survey Data spatial map so the readers can visualize the snow depth prediction across the landscape. Based on the difference map (b,e) it appears that the No survey data map product may estimate the same snow depth for all pixels, not acknowledging terrain impacts on mountain snow depth distribution. Adding two more maps would show that this is not true, or if it is, highlight the need for spatial data to be included in models. This would also help support the claim of spatial SD leading to improved spatial patterns (line 556)

The model output from the ”No survey Data” does not produce the same snow depth for all pixels, as it still includes topographic features during model training. The main difference with the ”Survey data” models (nine in total), is that the latter are trained on a more versatile dataset with respect to topographical features. The ’original’ training dataset, not including the snow-survey SD maps, contains more low-slope values, and more locations with low TPI-values. As such, an ML model trained solely on the data of the stationary sites may already perform well. Nonetheless, including spatially distributed data within the training procedure improves the representation of topographical impacts on the predictions, which is evident from Fig. 8 in the revised manuscript.

582-583: I do not know how the authors came to this conclusion and do not recall (nor can I find) supporting information within the manuscript to come to this conclusion. I suggest either removing or ensure supporting findings are mentions earlier in the document.

Thank you for pointing this out. You are right and this has now been removed in the revised manuscript.

References

Huss, M., Sold, L., Hoelzle, M., Stokvis, M., Salzmann, N., Farinotti, D., and Zemp, M.: Towards remote monitoring of sub-seasonal glacier mass balance, Annals of Glaciology, 54, 75–83, https://doi.org/10.3189/2013AoG63A427, 2013.

Sturm, M. and Liston, G. E.: Revisiting the Global Seasonal Snow Classification: An Updated Dataset for Earth System Applications, Journal of Hydrometeorology, 22, 2917 – 2938, https://doi.org/10.1175/JHM-D-21-0070.1, 2021.