Comment on “Machine Learning and Explainable AI for Sentinel-1–Based Snow Depth Estimation: Insights from Three Different Mountainous Regions” by R. Chandra Prabha and RAAJ Ramsankaran

General comments

The manuscript presents the development and validation of a Machine Learning (ML) model used to estimate snow depth (SD) from Sentinel-1 backscatter data and auxiliary variables. The modeling framework is applied over three different mountain ranges: The European Alps, Colorado Rocky Mountains and Indian West Himalayas. Overall, the reported results are comparable and even appear as improvement over previous Sentinel-1 based SD estimates. The ML model conception, training and evaluation strategies seem standard and quite robust.

My main comments are related to the explainable AI (SHAP values) analysis, and the conclusions that can stem from it. I think this analysis is useful in determining where and when can backscatter related variables be more informative about snow depth. However, these insights appear as more statistical rather than physical as stated in the objectives, because the latter cannot be assessed without observed or at least estimated snowpack conditions (e.g. Marin et al., 2020). I suggest rewording across the whole manuscript to reflect this.

I think that the relatively low SHAP values of the backscatter variables compared to auxiliary snow cover duration (SCD), elevation (DEM) and day of water year (DOW) are concerning. I suggest directly addressing this mismatch, because it works against the manuscript’s goal. It could be interpreted as a potential “conflict” between maximizing model performance and gaining insights from backscatter variables. I’d recommend either limiting the relevance of these variables in the model output, or reporting the performance when excluding each of these auxiliary datasets.

Given the hydrological relevance, I found that metrics about annual peak SD and peak SD timing were missing. It represents an important benchmark for the model and the information it can deliver, although the analysis may be limited to certain sites/regions due to data scarcity. Additionally, model evaluation in distinct hydrological conditions (e.g. drought and wet years) can give further insight into model applicability and usefulness.

In general, the results appear to confirm the prior knowledge about the relationship between backscatter and canopy, and shallow, deep or wet snow. Regarding the latter, I think the manuscript could be enriched by including either wet snow or runoff onset information (Gagliano et al., 2026) to quantify what kind of performance the ML model can deliver under these conditions (e.g. is it able to compensate through other variables like DOW?). A drop-off in the SHAP value for the backscatter variables would be a nice verification.

Overall, I think the manuscript needs to address the points raised above in order to reach an originality level fit for publication in The Cryosphere. In its current form, too much emphasis is given to the ML model and its performance. The actual knowledge gap that this study can explore is the information content and predictability of SD within the backscatter variables, rather than developing a perfect model which relies mostly on auxiliary data.

Specific comments

The final sentence (Line 22) on the abstract reads as if an interannual SHAP analysis was conducted ("...across seasons,"). I'd recommend changing it to "within seasons" or something similar to represent the 15 day-of year window SHAP analysis.

On figure 1 (a), I'd suggest either increasing the resolution of the global map or use a simpler map. The satellite image with national and state/provinces borders and names looks too busy and blurry.

In figure 5, the correlation colorbar could be unique for all the plots related to that metric, given that all 3 cover the same -1 to 1 range of values. Similarly for the bias metric, the colormaps are graded to show intra-region bias variability. Maybe showing a single color grade for all stations across the three ranges could be more effective at communicating performance disparities across these mountain ranges. This could also free up some room, allowing for larger font sizes on the colorbars and legend.

References

Gagliano, E., Shean, D., and Henderson, S.: A global high-resolution dataset of snowmelt runoff onset timing from Sentinel-1 SAR, 2015–2024, Earth Syst. Sci. Data, 18, 5871–5894, https://doi.org/10.5194/essd-18-5871-2026, 2026.

Marin, C., Bertoldi, G., Premier, V., Callegari, M., Brida, C., Hürkamp, K., Tschiersch, J., Zebisch, M., and Notarnicola, C.: Use of Sentinel-1 radar observations to evaluate snowmelt dynamics in alpine regions, The Cryosphere, 14, 935–956, https://doi.org/10.5194/tc-14-935-2020, 2020.