

Machine learning for snow depth estimation over the European Alps, using Sentinel-1 observations, meteorological forcing data and ~~physically-based~~ process-based model simulations

Lucas Boeykens^{1,2}, Devon Dunmire², Jonas-Frederik Jans¹, Willem Waegeman³, Gabriëlle De Lannoy², Ezra Beernaert¹, Niko E.C. Verhoest¹, and Hans Lievens¹

¹Hydro-Climate Extremes Lab, Ghent University, Coupure links 653, 9000 Gent, Belgium

²Department of Earth and Environmental sciences, KU Leuven, Celestijnenlaan 200E, 3001 Leuven, Belgium

³Department of Data Analysis and Mathematical Modelling, Ghent University, Coupure links 653, 9000 Gent, Belgium

Correspondence: Lucas Boeykens (lucas.boeykens@ugent.be, lucas.boeykens@kuleuven.be)

Abstract. Seasonal mountain snow is an indispensable resource, ~~providing drinking water to more than a billion people worldwide, supporting agriculture, industry and hydropower generation, and sustaining river discharge, soil moisture and groundwater recharge. However, but~~ accurate estimates of this ~~seasonal~~ 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 ~~this issue by estimating~~ Alpine snow depth estimation at a 100 m spatial resolution and sub-weekly temporal resolution ~~with over the 2015–2024 period using multiple input configurations within~~ an extreme gradient boosting ~~model~~ (XGBoost) ~~for the time period 2015–2024~~ machine learning (ML) model. We explore the potential ~~for using of~~ Sentinel-1 C-band dual-polarized synthetic aperture radar polarimetry (PolSAR) observations ~~to improve upon backscatter-based approaches~~, and include either regionally downscaled meteorological forcing data ~~and or~~ modeled snow depth as additional inputs to further explain interannual and spatial variability. ~~To A threefold nested cross-validation scheme is used to~~ account for the spatio-temporal dependencies present in the snow depth data, ~~we conduct a threefold nested cross-validation, and incorporate spatial training data to better represent topographical patterns in snow depth variability. Finally, we utilize~~ XGBoost's internal booster and Shapley additive explanation ~~values to understand the relationship between the input features and predicted snow depths during (SHAP) values are used to relate the input features with the predictions for~~ both dry and wet snow conditions. Our results ~~demonstrate that incorporating Sentinel-1 indicate that the inclusion of~~ PolSAR observations leads to ~~more accurate snow depth retrievals compared to using backscatter alone. In addition, our analyses indicate that including 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 ~~estimates substantially improves the XGBoost snow depth estimates, both of which yield comparable accuracy. Finally, we demonstrate that the inclusion of spatial training data is~~ substantially enhances predictive performance, particularly when spatially distributed training data, proven to be essential for capturing ~~the topographic influence~~ topographic controls on snow depth ~~estimates, 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 to obtain good spatial prediction accuracy. Overall, this work
25 contributes to an improved large-scale monitoring of water stored in seasonal mountain snow 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.

1 Introduction

~~Seasonal snow plays a critical role in nature, society and climate. First, snow~~ Globally, the spatial and temporal dynamics
30 of seasonal snow have been changing over the last decades. Notarnicola (2022) and Bormann et al. (2018) report a negative
trend in global annual snow cover, while Matiu et al. (2021) observed an overall decreasing trend in monthly mean snow depth
(SD) for the months of November through May across the European Alps. These changes in seasonal snow dynamics have
implications for society, ecosystems, and the climate system, given the critical role that snow plays in each of these domains.
Snow namely serves as a natural water reservoir that contributes to river discharge, soil moisture and groundwater recharge
35 (Gascoïn et al., 2024; Dozier et al., 2016). ~~In addition, snow~~, and moreover fuels water basins used for hydropower generation
(Bormann et al., 2018). Snow also provides drinking water to over 1.2 billion people worldwide, supports agricultural irrigation
via snow melt through snowmelt-driven runoff, and meets other essential human needs (Bales et al., 2006; Barnett et al., 2005; Luce et al., 2005).
~~Snow melt runoff is also indispensable for hydropower generation, can contribute to~~ exerts substantial socioeconomic impacts,
for example through traffic delays and accidents ~~; and sustains a multi-billion dollar~~ and by sustaining a multi-billion-dollar win-
40 ter recreation industry that is the dominant revenue source for many Alpine communities (Burakowski and Magnusson, 2012; Seeherman et al., 2012).
~~(Bales et al., 2006; Barnett et al., 2005; Bormann et al., 2018; Qin et al., 2020; Sturm et al., 2017; Burakowski and Magnusson, 2012; Seeherman et al., 2012)~~

The climatic significance of snow is arguably even greater, albeit harder to monetize (Sturm et al., 2017). The combination of
the high albedo of snow, which reflects of a large portion of the incoming solar radiation, with the vast area of the world covered
45 yearly by snow, enhances global cooling, and thus influences the Earth's surface energy balance (Warren, 1982; Groisman et al.,
1994; Marks and Dozier, 1992).

Globally, the spatial and temporal dynamics of seasonal snow have been changing over the last decades. Notarnicola (2022)
and Bormann et al. (2018) report a negative trend in global annual snow cover, while Matiu et al. (2021) observed decreasing
trends in monthly mean snow depth (SD) for the months of November through May, at most of their surveyed sites across the
50 European Alps. Given the importance of seasonal snow As a result of its importance, key snowpack properties such as SD and
snow water equivalent (SWE), the latter relating to SD through snow bulk density, have been designated as essential climate
variables (Gascoïn et al., 2024; Global Climate Observing System), and various scientific institutions and international organi-
zations have prioritized their enhanced observation and monitoring (Gascoïn et al., 2024; Dozier et al., 2016; World Meteorological Organization,
2016; Dozier et al., 2016; World Meteorological Organization, 2023).

To date, mountain SD and SWE have been observed and monitored using a wide range of measurement and modeling
55 techniques, including manual and automated point measurements; airborne passive and active optical sensors; airborne and

space-borne passive microwave radiometers; ~~and~~ airborne and space-borne active microwave sensors, such as synthetic aperture radar (SAR); ~~and process-based models~~. Manual and automated point measurements offer frequent data at many locations globally ~~;~~ (e.g., [Fontrodona-Bach et al. \(2023\)](#) or [Matiu et al. \(2021\)](#)), but fall short in capturing the snowpack's spatial variability due to their relatively sparse and uneven spatial distribution within alpine regions (Miller et al., 2022; López-Moreno et al., 2015). Moreover, this type of monitoring is often conducted at relatively flat and open sites, rarely covers the highest elevations, and can disturb the snowpack (Dozier et al., 2016; Deems et al., 2013; Miller et al., 2022). Snow measurements derived from airborne passive and active optical sensors, such as photogrammetry-based SD maps (Bührlé et al., 2023; Marty et al., 2019) and lidar altimetry-derived SD and SWE products (Painter et al., 2016; Deems et al., 2013), provide high-resolution, spatially distributed measurements [within mountain ranges](#), yet their use is constrained by limited spatial coverage and sparse temporal sampling (Tsang et al., 2022; Dozier et al., 2016; Hill et al., 2019). In contrast, space-borne passive microwave radiometers provide broad spatial coverage, but their effectiveness is reduced by a coarse spatial resolution (> 25 km) and signal saturation in deep (≥ 0.8 m) snowpacks (Tedesco and Narvekar, 2010), [limiting their applicability in mountainous regions](#). Alternatively, space-borne SAR has demonstrated potential for monitoring SD or SWE at finer spatial scales, particularly with Ku- and X-band observations (Tsang et al., 2022). Unfortunately, X-band observations are not publicly available, and no Ku-band satellite mission currently exists (Tsang et al., 2022); ~~thereby constraining the development of such high-resolution SAR snow products over vast areas~~. Differently, ~~multiple process-based snow models exist to estimate both SD and SWE at different spatial scales (e.g. Snowclim (Lute et al., 2022); SNOWPACK (Bartelt and Lehning, 2002); SnowModel (Liston and Elder, 2006a); factorial snowpack model (Essery, 2015)).~~ Such models use meteorological forcings — often reanalysis data — to simulate the snowpack as one or multiple layers; solving mass and energy balance equations and mathematically representing the physical processes occurring within the snowpack. However, due to uncertainties in the forcing data and/or limitations in model representation, the modeled snow properties may differ substantially compared to in situ measurements and among each other (e.g., [Terzago et al. \(2020\)](#), [Largeron et al. \(2020\)](#), and [Ryken et al. \(2020\)](#)).

~~Recent work (Lievens et al., 2019; Brangers et al., 2024b) has demonstrated the sensitivity of~~ Recently, also SD retrieval algorithms have been developed that utilize Sentinel-1 (S1) active microwave observations ~~to snow~~ at C-band (~ 5.4 GHz). ~~This work focuses on~~ For instance, [Lievens et al. \(2022\)](#) use C-band ~~backscatter in~~ co- (VV) and cross- (VH) polarization, [polarized backscatter intensity observations in a conceptual change detection algorithm, which allows them to estimate SD across the European Alps at sub-kilometer spatial resolution. This algorithm has been developed based on the results of Lievens et al. \(2019\) and Brangers et al. \(2024b\), who demonstrated the sensitivity of active microwave observations to snow at C-band. These previous studies have focused on C-band backscatter intensity in co- and cross-polarization,](#) demonstrating that under dry snow conditions, an increasing snowpack [depth](#) leads to enhanced signal depolarization. This depolarization is thought to be related to snow volume scattering; ~~;~~ (multiple) scattering on anisotropic snow crystals or clusters of crystals; ~~;~~ and scattering at snow (or snow-ground) layer interfaces; ~~ete~~. As a result, mainly ~~cross-polarized backscatter~~ VH, or the ratio of ~~cross- to co-polarized backscatter~~ [VH to VV backscatter intensity](#) is sensitive to an increasing SD (Lievens et al., 2019; Tsang et al., 2022). Jans et al. (2025) further explored the sensitivity of ~~Sentinel-1 (S1)~~ C-band backscatter ~~satellite intensity~~ observations to seasonal patterns of SD and SWE ~~. This work emphasizes~~ [across the European Alps. In addition,](#)

Jans et al. (2025) emphasize the potential of S1 C-band dual-polarized SAR polarimetry (PolSAR) for SD retrieval, through the use of derived PolSAR variables that inform about the dominant snow scattering mechanism (i.e., a polarimetric scattering angle that increases with a growing snowpack), or about the intensity of the total received backscatter (i.e., the first Stokes parameter). Previous work has explored PolSAR for wet snow or snow cover area detection (Tsai et al., 2019; Varade et al., 2020). However, but no operational retrieval algorithms currently exist that utilize dual-polarized PolSAR observations for estimating mountain SD or SWE.

In contrast, SD retrieval algorithms that incorporate S1 C-band backscatter observations have been developed previously. For instance, Lievens et al. (2022) use C-band co- and cross-polarized backscatter observations in a conceptual change detection algorithm to estimate SD across the European Alps at sub-kilometer spatial resolution. However, Despite its ability to characterize mountain SD at high-resolution over vast areas, the algorithm's performance is limited of Lievens et al. (2022) shows limited performance during periods of wet and shallow snow, in forested areas, or after frequent freeze-thaw events (Lievens et al., 2022; Brangers et al., 2024b). To address some of these limitations, Dunmire et al. (2024) implemented machine learning (ML) to enhance S1-based SD estimates at a high spatial resolution (100 m). They utilized an extreme gradient boosting model (XGBoost; Chen and Guestrin (2016)), incorporating input features related to topography, land and snow cover, and S1 C-band backscatter intensity. For the latter, they focused on the same satellite observations used by Lievens et al. (2022), namely eo-(VV-) polarized backscatter-VV backscatter intensity, and the cross-polarization ratio, defined as the ratio of cross-(VH-) to eo-polarized backscatter-VH to VV backscatter intensity. Other studies have similarly explored the application of ML for estimating SD and SWE, both with and without integrating S1 C-band satellite observations (Xiong et al., 2022; Daudt et al., 2023; Broxton et al., 2024; Duan et al., 2024). However, no work currently exists that incorporates dual-polarized PolSAR variables for snow retrieval; not in a conceptual algorithm, nor within an ML model.

In addition to the previously mentioned conceptual, empirical, and ML approaches, physically-based models are also used to estimate snow properties, including SD and SWE (e.g., Lute et al. (2022); Bartelt and Lehning (2002); Niu et al. (2011)). Such models use meteorological forcings — often reanalysis data — to simulate the snowpack as one or multiple layers; solving mass and energy balance equations and mathematically representing the physical processes occurring within the snowpack. However, due to uncertainties in the forcing data and/or limitations in model representation, modeled SD may differ substantially compared to in situ measurements. In an attempt to further improve the estimates, multiple Similarly, multiple authors have explored the potential of ML in a hybrid physical/data-driven approach to enhance the outcomes of snow models (e.g., Wang and Tian (2024); Pomarol Moya et al. (2025)). Thereby, additional information relevant to the snowpack can be incorporated, e.g., such as S1 C-band observations. A similar approach has been used by Brangers et al. (2024a) and De Lannoy et al. (2024), De Lannoy et al. (2024) and Dunmire et al. (2025) within a data-assimilation setup, who used S1 C-band snow retrievals to enhance modeled SD. Additionally, ML models can also be used as emulators, replacing physical models by directly estimating SD from meteorological forcings (e.g., Charbonneau et al. (2025)). Nevertheless, few studies have investigated ML setups that directly integrate modeled SD with S1 C-band observations. Moreover, direct comparisons between SD ML configurations incorporating modeled SD and those using meteorological forcing data as emulator remain

Table 1. Overview of the datasets and their associated targeted resolution ~~marked by a cross(×)~~. The 100 m and 500 m resolutions refer to the spatial resolution of the WGS84 grid.

Resolution	In situ SD	S1 C-band	Snow cover	Wet snow	Meteorological	<i>Snowlim</i>	Auxilia
WGS84	photogrammetry maps (incl. photogrammetry maps)	variables	data	mask	forcing data	SD	data
100 m	×			×			×
500 m		×	×		×	×	

~~scarce, leaving unresolved whether meteorology-driven ML setups can achieve performance comparable to configurations that explicitly include modeled SD.~~

In this manuscript, we further investigate the potential of ML ~~to improve SD estimates with different input configurations to accurately estimate SD~~ across the European Alps, ~~essential for accurately quantifying the annual water stored as snow~~. To this end, we conduct various experiments comparing the performance of S1 C-band backscatter ~~intensity~~ with PolSAR observations. ~~Additionally, we use feature importance (FI) analysis to examine~~ ~~To gain insights in~~ when and where ~~which type of~~ satellite observations contribute to ~~the~~ SD predictions, ~~we use feature importance (FI) analysis~~ under both dry and wet snow conditions. We also evaluate the added value of incorporating meteorological forcing data or ~~physically-based process-based~~ snow model SD estimates as features in our ML model, to improve the representation of interannual and site-specific variability. To validate our approach, we implement a threefold nested cross-validation (nested CV) framework, which masks subsets of the data during training and predicting (testing). This framework accounts for the spatial, temporal and spatio-temporal dependencies in the data, an essential consideration when validating ML models for spatio-temporal purposes (Meyer et al., 2018). Finally, to address the limitations of relying solely on point-based training data, we compare input configurations trained with and without airborne snow survey data, and validate the predictions against ~~nine~~ airborne photogrammetry snow surveys conducted across the Dischma Valley, Switzerland.

2 Data

Various datasets were collected and reprojected (and/or resampled) to two geographic grids used in this study: the ~ 100 m ($1/1008^\circ$) resolution World Geodetic System 1984 (100 m WGS84) grid, ~~used~~ for fine-resolution datasets or to preserve more details in the target (SD) and auxiliary datasets, and the ~ 500 m ($5/1008^\circ$) WGS84 grid (500 m WGS84), used for the ~~course-scale coarse-scale~~ and satellite datasets. Table 1 provides an overview of the data products and the grid to which each was aligned to be used within this study.

2.1 In situ SD measurements

We compiled in situ SD data over the European Alps by combining stationary and point-based measurements and airborne photogrammetry snow surveys, sourced from multiple regional providers across the Alpine countries (Fig. 1). Point-based

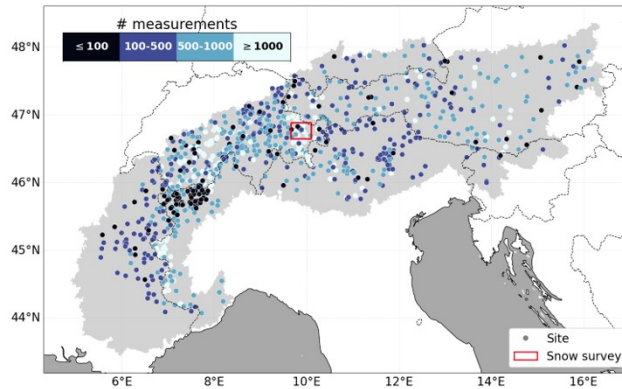


Figure 1. Spatial distribution of (dots) static [point or point-based](#) measurement sites colored by the number of available data in time, and (red box) airborne photogrammetry snow surveys conducted across the Dischma valley, used to train XGBoost and evaluate the SD predictions. The light gray [shaded](#)-area delineates the Alpine region as defined by the Alpine Convention (Aureliano, 2020).

150 measurements were collected for the time period 2015-2024, and include data from Austria (provided by Geosphere Austria, formerly ZAMG, and TIWAG), France (from measurement stations operated by partners under agreement of MétéoFrance), Italy (collected in the autonomous regions of Valle d’Aosta, Trento and Alto-Adige, Piemonte, Lombardy and data provided by the International Center for Environmental Monitoring CIMA Research Foundation), and Switzerland (manual and automated point-based SD measurements collected and managed by the Swiss Federal Institute for Forest, Snow and Landscape
 155 Research (WSL) - Institute for Snow and Avalanche Research (SLF)). In addition, we [supplemented-augmented](#) our SD dataset with measurement [sites-originating-data-sourced](#) from the Synoptic Data platform and the National Oceanic and [the](#) Atmospheric Administration’s Global Historical Climatology ([GHC](#))[network, extending our coverage to locations in](#) [Network-Daily database \(GHCNd, Menne et al. \(2012\)\)](#). [With the addition of these data, our dataset also contained sites within](#) Germany and Slovenia.

160 Next, we collected spatial SD measurements, derived from airborne photogrammetry (Marty et al., 2019; Bührlé et al., 2023) over the Dischma valley, Switzerland (Fig. 1, red box). A total of nine photogrammetry SD surveys (snow surveys) were added to our dataset, with an original resolution of 0.5 m (6 maps, measured yearly near peak SD between 2017 and 2022) or 2 m (3 maps, collected on 3 distinct days in 2016). For both resolutions, we applied a linear averaging and reprojected all surveys to the target grid of our SD estimates, the 100 m WGS84 grid.

165 2.2 S1 C-band PolSAR and backscatter [intensity](#) observations

Within our research, we collected data from the European Space Agency (ESA) and Copernicus S1A and -B satellites, which operate [in-the-at](#) C-band ([5.4 GHz](#)) and share the same orbital plane with a 180° orbital phasing difference. Both have a 12 day repeat cycle. Combined, they offer a repeat cycle of 6 days over the study area. S1B, however, has been out of operation since 24 December 2021, and thus does not cover the complete study period (2015-2024).

170 S1 dual-polarized PolSAR observations were acquired from interferometric wide swath (IW) single look complex (SLC) data
in dual-polarization (VV and VH) mode. The data were processed into two dual-polarized PolSAR variables: the polarimetric
scattering angle (α) and the first Stokes parameter (S_0), using the Sentinel application platform (SNAP) toolbox, version 9,
with processing steps described in Jans et al. (2025). These processing steps included applying precise orbit file corrections,
radiometric calibration, debursting TOPSAR bursts, and merging sub-swaths into a single image. This was followed by the
175 generation of the 2×2 covariance matrix (C2-matrix). The C2-matrix was then used as input for two distinct processing chains.
The first one involved an eigenvector/value decomposition of the C2-matrix, followed by the H/ α decomposition (Cloude and
Pottier, 1997), from which α was derived. The second processing chain utilized the compact-polarimetry Stokes parameter
generation to derive S_0 from the C2-matrix. Next, both variables were multilooked to ~ 25 m to reduce speckle and finally, a
Range Doppler terrain correction was applied to both variables, reducing (geometric) distortions due to topographical variations
180 in the scene. For the latter, the Copernicus 30 m global digital elevation model (GLO-30 DEM) was used.

Next, co- (γ_{VV}^0) and cross- (γ_{VH}^0) polarized backscatter intensity variables were retrieved from ground range detected (GRD)
amplitude data. Processing steps, described in Lievens et al. (2022) and Jans et al. (2025), were applied, including precise orbit
file application, GRD border and thermal noise removal, radiometric calibration, terrain flattening to backscatter as γ^0 (Small,
2011) and finally a Range Doppler terrain correction. As for the PolSAR variables, the Copernicus GLO-30 DEM was used
185 during the processing steps. Additionally, local incidence angle (LIA) information was preserved and the cross-polarization
ratio (γ_{CR}^0 ; calculated as the ratio between γ_{VH}^0 and γ_{VV}^0 in linear scale) was calculated.

Finally, both the PolSAR and backscatter intensity observations were rescaled and summer-corrected, to reduce inter-orbital
and interannual start-of-season differences, respectively. To this end, we first reprojected the satellite observations onto the
100 m WGS84 grid using linear averaging, and resampled to the 500 m WGS84 grid by taking the mean. Then, we performed
190 a first and second-order moment scaling (to correct for differences in the mean and variance) between orbits, ~~followed by~~
~~subtracting~~. Finally, from each scaled time series for each snow season (September-June), we subtracted the mean summer
value (July-September) ~~from the satellite observations for each snow season (September-June)~~ of the summer preceding that
snow season, thereby reducing interannual start-of-season differences that are not related to snow conditions (e.g., changing
vegetation). To indicate that these variables have been rescaled, we further append a superscript "s" (^s) to the satellite variable
195 notation.

2.3 Snow cover data and wet snow mask

Snow cover data were obtained from two sources: binary snow cover information from the interactive multisensor snow and
ice mapping system (IMS; U.S. National Ice Center (2008)) at 1 km resolution, and fractional snow cover (FSC) information
from the moderate resolution imaging spectroradiometer (MODIS; Hall and Riggs (2020a, b)) at 500 m resolution. For the
200 latter, we collected cloud gap-filled FSC data separately from the Terra and Aqua satellites, which we both reprojected onto
the 500 m WGS84 grid using linear averaging. Subsequently, we combined the individual products using a weighted average,
which accounted for the quality flags of each product, the time lag between the last cloud-free FSC observation and the current
date, and the time lag of the individual satellite observations relative to each other (Appendix A). Next, a cutoff was set on the

combined product, only allowing for FSC information less than 7 days before the current date. Finally, remaining gaps (3.95%
 205 of the data points across the whole Alps and study period) were filled using IMS' binary snow cover information, reprojected
 to the same 500 m WGS84 grid using majority resampling. Hereby, a no-snow indication of IMS was set to an FSC value of
 0%, while an IMS snow covered pixel received a value of 100%.

Additionally, a wet snow mask was acquired to differentiate between dry and wet snow periods. This dataset, based on S1
 C-band satellite observations and obtained from the Copernicus Land Monitoring Service (CLMS) High Resolution Snow and
 210 Ice Monitoring (HRSIM) SAR Wet Snow in high mountains (SWS) product (Copernicus Land Monitoring Service, 2024),
 provides a binary classification between wet snow conditions, and dry, patchy or snow-free conditions. The data, available at
 60 m resolution from September 2016 onward, were first reprojected onto the 100 m WGS84 grid using majority resampling,
~~meanwhile excluding~~ whereby the most frequently occurring SWS value within each 100 m target pixel was selected while
excluding any no-data values. Subsequently, all non-wet pixels were classified as dry, while quality flags were retained to
 215 identify locations and periods with no (reliable) classification.

2.4 Meteorological forcing data

We downscaled ~~three-hourly~~ 3-hourly meteorological forcing data over the European Alps to a 500 m spatial resolution, to
 serve as additional input to the ML model. To this end, we first collected coarse three-hourly meteorological forcing data with
 a 0.1° spatial resolution. Precipitation data were sourced from the multi-source weighted-ensemble precipitation (MSWEP,
 220 Beck et al. (2019)) dataset, which combines various state-of-the-art reanalysis products with gauge and satellite observations
 (Beck et al., 2019). Other variables, including downward shortwave and longwave radiation, 2-m air temperature, wind speed,
 and 2-m relative humidity, were obtained from the multi-source weather (MSWX, Beck et al. (2022)) product. This dataset is
 derived from coarse-scale European centre for medium-range weather forecasts reanalysis ~~v5~~ version 5 (ERA5) data refined
 through bias correction and downscaling using monthly 0.1° reference climatologies (Beck et al., 2022).

225 Subsequently, we applied bilinear interpolation in combination with ~~other downscaling techniques~~ different downscaling
techniques, explained below, to account for local terrain features. To this end, the Copernicus GLO-30 DEM was employed,
 linearly averaged to the 500 m WGS84 grid to match the target resolution for the meteorological forcings. ~~For First, coarse-scale~~
precipitation (P_{coarse} ; $[\text{mm } 3\text{h}^{-1}]$) 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
 230 study. Thus, for each location x within the 500 m grid and at time step t , ~~the coarse-scale precipitation ($P_{\text{coarse},t}$; $[\text{mm } 3\text{h}^{-1}]$)~~
~~data were was~~ downscaled using the following ~~method:~~ equation:

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

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°
 235 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,t}}\$ values.](#)

The other forcings were adjusted as follows. The temperature was corrected using a seasonally varying lapse rate, following
240 Liston and Elder (2006b), to account for elevation-dependent temperature variations. For relative humidity, we first converted
to dewpoint temperature, then applied a lapse rate correction —similar as for the temperature—, and finally converted the
corrected dewpoint temperature back to relative humidity (Liston and Elder, 2006b). Downward shortwave radiation was
adjusted following the method of Fiddes and Gruber (2014), which involved partitioning the radiation data into direct and
diffuse components, accounting for the elevation effect on the direct component, and applying a topographic correction to both
245 components. In contrast, downward longwave radiation and wind speed were not adjusted after bilinear interpolation.

2.5 Physically-based snow model *Snowclim*

Modeled SD estimates were obtained using *Snowclim* (Lute et al., 2022), a ~~physically-based~~ [process-based](#) enhanced single
layer snow model utilizing a fully distributed energy and mass balance. We ran the model over the entire European Alps with
the downscaled meteorological forcing data at a 500 m spatial and 3-hourly temporal resolution. To obtain daily values, the
250 3-hourly SD estimates were averaged to a daily time step. A description of the model’s parameter settings, how they have been
calibrated, and the model’s performance when validated against in situ measurements, is provided in Appendix B.

2.6 Time-independent auxiliary data

In addition to time-dependent data, we employed time-independent auxiliary data, sourced from various providers. As with the
other datasets (except the meteorological forcing and *Snowclim* data), these data were first projected onto the 100 m WGS84
255 grid, using linear averaging for continuous variables and majority resampling for categorical data. Forest cover fraction (FCF)
data and land cover information, including a water mask, were acquired from the 2018 epoch of the 100 m Copernicus PROBA-
V global land cover dataset (Buchhorn et al., 2020). Next, we used the Copernicus GLO-30 DEM to extract elevation and three
topographic features: slope, aspect, and topographic position index (TPI), the latter quantifying the relative elevation of a
grid cell concerning its surroundings within a predefined diameter. Slope and aspect were computed in MATLAB using the
260 Geodetic (Craymer, 2022) and TopoToolbox (Schwanghart and Scherler, 2014) toolboxes, while TPI was calculated following
the methodologies of Weiss (2001) and Lundblad et al. (2006). For this study, we computed the TPI using only the neighboring
pixels of a specific grid cell, corresponding to a 3×3 grid window. Finally, we included a glacier mask, retrieved from the
Randolph Glacier Inventory version 7.0 (RGI 7.0 Consortium, 2023), [and collected the Sturm and Liston \(2021\) snow classes.](#)

3 Methods

265 3.1 [XGBoost model selection](#)

Within this study, we deployed XGBoost to estimate SD across the European Alps. XGBoost is a tree-based traditional ML algorithm that constructs multiple decision trees in a sequential order, to minimize a differentiable loss function (Chen and Guestrin, 2016). Thereby, new trees are trained to fit the residual errors of the previously fitted trees, while incorporating regularization terms to reduce model complexity, and including additional mechanisms that contribute to its computational efficiency (Chen and Guestrin, 2016). We chose this model after comparative experiments with other tree-based models, including Random Forest, LightGBM and CatBoost; and because of its use in recent snow studies (e.g., Dunmire et al. (2024) or Goodarzi et al. (2025)).

3.2 Dataset preparation

We trained, validated, and tested XGBoost with various configurations on a dataset containing input features associated with each SD measurement (predictor variable) from the in situ measurement sites and snow surveys. To construct this dataset, we first excluded those stationary measurement sites located within glaciated areas, as well as those with less than 5% non-zero SD observations. Subsequently, we identified the unique locations of the stationary sites within the 100 m WGS84 grid, and averaged time series from sites within the same grid cell. Furthermore, we excluded SD measurements during the summer months July and August, and resampled the data with a 500 m spatial resolution to the 100 m WGS84 grid using value replication.

The S1 data spatially and temporally coinciding with available SD measurements on the 100 m WGS84 grid include both ascending and descending orbits. This allows a single location to have two S1 observation values for the same SD measurement on a given date. Importantly, we chose not to exclude satellite observations with high or low LIA, as we assumed that XGBoost could effectively learn the relationships between these satellite observations and the corresponding SD. For each location and date with an available satellite observation, we also selected the corresponding (nearest) time-dependent non-satellite data and time-independent auxiliary data. This resulted in a dataset comprising 1022 measurement sites, with the distributions of observed SD and selected auxiliary variables shown in Fig. E1.

A similar procedure was applied for the photogrammetry snow survey data: for each SD measurement, the nearest input feature values within the 100 m WGS84 grid were selected and compiled into a dataset, linking each SD observation with its corresponding input features. Since satellite acquisitions did not always coincide with the exact date of the snow surveys, we used the most recent S1 observation prior to the survey, with a maximum offset of two days. Lastly, to represent the temporal component of our datasets, we computed the day of the year (DOY) for each instance. To account for the cyclic nature of the calendar year and prevent discontinuity between 31 December and 1 January, we transformed the DOY into two separate features using sine (DOY_{\sin}) and cosine (DOY_{\cos}) functions:

$$DOY_{\sin} = \sin\left(\frac{2\pi DOY}{n} \frac{2 \cdot \pi \cdot DOY}{n}\right) \quad (2)$$

$$DOY_{\cos} = \cos\left(\frac{2\pi DOY}{n} \frac{2 \cdot \pi \cdot DOY}{n}\right) \quad (3)$$

with n the number of days in the corresponding year.

3.3 Feature configurations and SD prediction

3.3.1 S1 PolSAR vs. backscatter intensity observations

~~XGBoost was trained and validated with three different configurations to compare~~ To assess the performance of ML setups ~~with PolSAR and backscatter observations using S1 PolSAR versus backscatter intensity observations, XGBoost was trained and validated under three distinct configurations~~ (Table 2), without including meteorological forcing data or *Snowclim* SD estimates in these experiments. For each configuration, XGBoost incorporated the same time-independent auxiliary features: elevation, slope, aspect, TPI, and FCF. Next, DOY features, FSC data and LIA information were included, the latter to indirectly account for orbital differences. The first configuration, referred to as PolSAR_{ML}, incorporated α^s and S_0^s next to the shared baseline input features. ~~Conversely, the configuration~~ The second configuration, focusing on backscatter intensity observations (Backscatter_{ML} configuration), included $\gamma_{VV}^{0,s}$ and $\gamma_{CR}^{0,s}$ as satellite input features. In addition, we combined the PolSAR and backscatter intensity satellite input features (Combination_{ML} configuration), to identify the satellite variables most effectively used within XGBoost.

3.3.2 S1 observations plus meteorological or *Snowclim* SD estimates

The satellite observations were further complemented with either meteorological forcing data (Weather_{ML} configuration) or *Snowclim* SD estimates (*Snowclim*_{ML} configuration) to better resolve interannual and in-between-site variability (Table 2). The Weather_{ML} configuration was an extension of the Combination_{ML} setup, now including cumulative snowfall, cumulative wind speed, cumulative shortwave radiation, a cumulative number of melt days, and the mean daily temperature (T_a). Cumulative snowfall ($P_{s,c}$) was estimated by first calculating the daily fraction of precipitation falling as snow, based on temperature, precipitation and relative humidity using the bivariate logistic regression model described by Jennings et al. (2018), and subsequently summing these fractions from the 1st of September of the corresponding snow year up to the prediction date. Similarly, the cumulative wind speed ($U_{a,c}$) was derived as the sum of the daily wind speed from the 1st of September of the corresponding snow year up to the prediction date. Cumulative shortwave radiation (SWd₇), on the other hand, was summed over the seven days preceding the prediction date. Similarly, the cumulative number of melt days (MD₇) was defined as the number of days with a maximum temperature above 0°C within the 7 days preceding the prediction date. The *Snowclim*_{ML} configuration did not include meteorological forcing data as direct inputs, but instead makes use of modeled SD estimates from *Snowclim* (SD_{SC}), which are driven by meteorological conditions.

3.3.3 ~~Snow~~ XGBoost snow depth prediction procedure

Prior to being used in XGBoost for SD prediction, each feature was standardized individually, using the mean and standard deviation calculated from the training set values. This standardization was consistently applied to the validation and predicting (test) datasets as well, ensuring compatibility across all data splits. Input features from ascending and descending orbits were

Table 2. Features included in the ~~different~~ ML configurations tested ~~within in~~ this ~~manuscript~~ study: polarimetric scattering angle (α^s); first Stokes parameter (S_0^s); cross-polarization ratio ($\gamma_{CR}^{0,s}$); co-polarized backscatter intensity ($\gamma_{VV}^{0,s}$); cumulative snow precipitation and wind speed ($P_{s,c}$ and $U_{a,c}$, summed between September 1st of the corresponding snow year and the prediction date); cumulative shortwave radiation and melt days over the preceding 7 days (SWd_7 and MD_7); mean daily temperature on the prediction day (T_a); and Snowclim SD estimates (SD_{SC}). A cross (\times) denotes the presence of a feature in a given configuration. Additionally, all configurations include the following ~~baseline~~ input features, ~~which are not listed~~ indicated as *baseline* in the table: elevation, slope, aspect, topographic position index (TPI), forest cover fraction (FCF), fractional snow cover (FSC), sine and cosine values of the day of year (DOY_{sin} and DOY_{cos}), and the local incidence angle (LIA).

ML configuration	<i>baseline</i>	α^s	S_0^s	$\gamma_{CR}^{0,s}$	$\gamma_{VV}^{0,s}$	$P_{s,c}$	SWd_7 $U_{a,c}$	T_a SWd_7	MD_7	T_a	SD_{SC}
PolSAR _{ML}	\times	\times	\times								
Backscatter _{ML}	\times			\times	\times						
Combination _{ML}	\times	\times	\times	\times	\times						
Weather _{ML}	\times	\times	\times	\times	\times	\times	\times	\times	\times	\times	
Snowclim _{ML}	\times	\times	\times	\times	\times						\times

fed to XGBoost separately. In case a location had both an ascending and descending satellite observation on a given date, the mean of both XGBoost SD outputs was taken.

3.4 Spatio-temporal nested cross-validation and hyperparameter tuning

330 The concept of nested cross-validation (nested CV) has been extensively utilized in other studies (Dora et al., 2018; Abdulaal et al., 2018; Parvande et al., 2020), since nested CV allows to characterize the generalization error of an ML model correctly (Blancas, 2021). Figure 2 illustrates that nested CV involves both inner and outer resampling to, respectively, train and tune the ML model (and its hyperparameters), and evaluate predictions on unseen and independent test data. Considering the susceptibility of SD measurements to spatial and temporal autocorrelation, we implemented the ~~cross-validation~~ CV strategy described
335 in Landschoot et al. (2012) within a threefold nested CV framework for the site data (excluding the snow surveys), in which subsets of the data are masked during training, validation and predicting (testing). To this end, the data of the site measurements were partitioned into spatial, temporal and spatio-temporal folds.

The spatial folds were constructed using a two-step approach: first, ~~the~~ sites located within five km of one another were grouped into clusters, ~~which were then thereby preventing nearby sites with similar (climatic) characteristics and SD patterns~~
340 ~~from being split across training and test sets. Subsequently, these clusters were~~ randomly assigned to five unique folds, ensuring that each fold contained a comparable amount of data and ~~maintained~~ ~~preserved~~ a similar SD distribution. For the temporal folds, ~~measurement sites were kept separate, but data from all sites within a single sites were not clustered; instead,~~ ~~all observations from a given~~ snow season (September–June) ~~were grouped and then divided and from across the study area~~ ~~were grouped into nine blocks (corresponding with the number of snow seasons for which SD data is available), which were~~

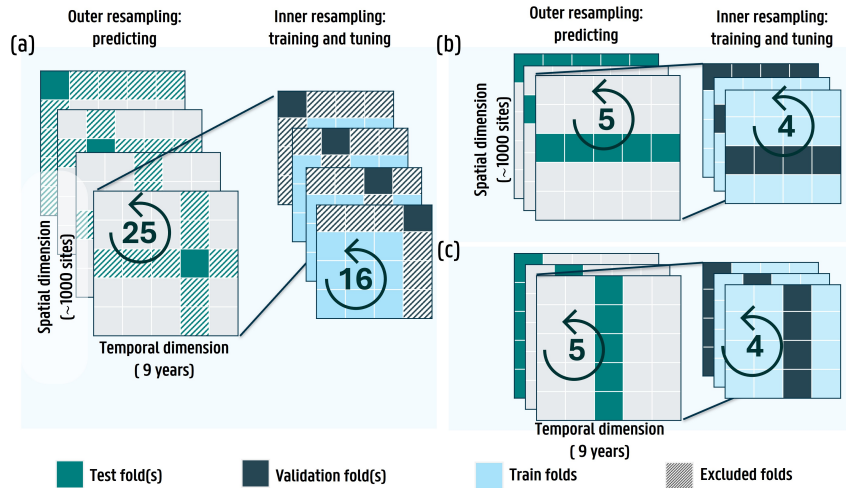


Figure 2. Principle of nested CV for (a) the spatio-temporal setup, (b) the spatial setup and (c) the temporal setup. The predicting (test) and training folds during outer resampling are colored in turquoise and light gray respectively. The validation and training folds during inner resampling are colored in dark and light blue. The striped boxes in (a) represent the excluded folds during training and testing within the spatio-temporal setup. During outer resampling, predictions are made with the trained ML model, with hyperparameters that are tuned during inner resampling. The numbers indicated in the arrows represent the amount of times outer and inner resampling are performed, respectively. For (a), four iterations are shown, while for (b) and (c), three iterations are displayed.

345 [then partitioned](#) into five folds, again ensuring comparable fold sizes and [a consistent SD](#) distribution. Finally, by combining [spatial and temporal folds](#) both [fold-creation techniques](#), we constructed 25 unique spatio-temporal folds.

For each unique fold (five for spatial and temporal nested CV; 25 for spatio-temporal nested CV), we iteratively designated one fold as the test set during outer resampling, using the remaining folds for training and validation. To ensure independence of the test data, all data from the sites and/or years present in the test fold were excluded from the training and validation sets
350 (Fig. 2).

The same approach was applied during inner resampling, to split the training and validation data used for tuning XGBoost’s hyperparameters. To tune the hyperparameters, we utilized Scikit-learn’s (Pedregosa et al., 2011) *RandomizedSearchCV* algorithm, employing the mean squared error (MSE) as loss function. A different random seed was set for each outer resampling loop, introducing variability in the selection of optimal hyperparameter combinations. During each loop, 150 hyperparameter
355 combinations were chosen from a predefined tuning grid, based on the one provided in PyCaret (PyCaret, 2020). The hyperparameters yielding the best mean MSE score across the validation folds were selected. This approach produced independent predictions for every instance in the dataset, enabling the calculation of performance metrics and the construction of FI scores.

3.5 Feature importance scores

We used both XGBoost's booster and Shapley additive explanations (SHAP; Lundberg and Lee (2017)) values to assess the feature's impact on the SD estimates. For the booster, we selected the *gain* FI, which informs about the contribution of each input feature in minimizing the loss function during model training. As such, for every outer resampling loop within the nested CV framework, we computed the ~~gain~~gain FI of the individual features during model training.

SHAP values, on the other hand, quantify the contribution of each input feature to an individual (SD) prediction, measured relative to an expected prediction (SD) value. Within the nested CV framework, SHAP values thus indicate how each feature influences the SD predictions made for the outer resampling predicting sets, relative to the average (SD) prediction computed from the training set. Consequently, SHAP values can be both positive and negative. To assess the overall global SHAP value FI during SD prediction, mean absolute SHAP values were computed for each input feature during every outer resampling loop. The latter were converted into relative contributions by normalizing them with the sum across all features.

We further used SHAP values to assess the contribution of the S1 PolSAR and backscatter ~~satellite~~intensity observations to the SD predictions throughout the seasonal evolution of the snowpack. To this end, for each site and snow season in the test sets, we first identified the periods during which a snowpack was present, based on the measured SD. The season start date was identified as the first date with $SD \geq 10$ cm, that was followed by nine consecutive days with $SD \geq 1$ cm. For low elevation sites with (very) shallow snowpacks, where this condition was not met for an entire snow season, the start of the snow season was marked by the first date with $SD \geq 10$ cm. Conversely, we defined the end of a site's snow season as the first day of a 10-day period with 0 cm SD, near the end of snowpack presence. In addition, we used the CLMS SWS product to distinguish, where possible, between dry and wet snow conditions. As this product is only available from September 2016 onward, the 2015-2016 snow season was omitted from the corresponding analyses.

3.6 Performance metrics

In addition to the FI and SHAP value analysis, several performance metrics were used to validate XGBoost SD predictions with measured SD. For each configuration and across all outer resampling predicting sets (thus the total dataset), we computed the ~~spatiotemporal~~spatio-temporal Pearson correlation coefficient (R), mean absolute error (MAE), root mean squared error (RMSE) and bias between predicted and measured SD (all averaged over time and space). Since cross-validation is conducted in both space and time, the performance metrics reflect the ML model's ability to generalize across purely spatial, purely temporal, and combined spatio-temporal domains. Additionally, certain metrics were calculated on a per-site basis, such as the MAE.

Next, we evaluated the XGBoost's capability to generate spatial SD predictions by comparing them with the photogrammetry snow surveys conducted in the Dischma Valley, Switzerland. For this case, we trained XGBoost exclusively on ~~measurement site data~~stationary and point-based measurements, excluding spatial data from the snow surveys. Model training and hyperparameter tuning were performed using fivefold spatial ~~cross-validation~~CV, as this approach best reflects the intended application of the ML model: predicting SD on unseen locations across the European Alps within the time period of our S1 data collection.

After training, XGBoost was applied to predict SD for each of the nine snow surveys, and performance metrics of ~~of predicted vs. observed SD~~ were computed.

We then repeated this process, but now incorporating all snow survey data except those of the targeted date into the training dataset. This resulted in nine independently trained and tuned XGBoost models, each applied to predict SD for the correspond-
395 ing unseen snow survey. In this approach, we also employed spatial ~~cross-validation CV~~ for training and tuning, where every training snow survey was randomly assigned to one of the five folds. This ensured that the distribution of training and validation data remained consistent across folds.

4 Results and discussion

4.1 Performance of ML configurations with S1 PolSAR and backscatter ~~intensity~~ variables

400 ~~Across the three nested CV frameworks and~~ In general, all configurations (i.e., PolSAR_{ML}, Backscatter_{ML} and Combination_{ML}) ~~, performance is highest for the temporal~~ achieve best results within the temporal nested CV framework (Table E1), and performance progressively deteriorates in the spatial and spatio-temporal frameworks. ~~The differences between the configurations are moreover most pronounced for the temporal nested CV framework, with significant improvements in site MAE when comparing the PolSAR_{ML} or Combination_{ML} configurations with the Backscatter_{ML} configuration ($p \ll 0.05$, 95% confidence level). While this indicates that at the training sites, PolSAR variables, or a combination of PolSAR and backscatter variables, seem to better capture the seasonal evolution of SD, the temporal framework overestimates model performance relative to the overall objective of this study, namely predicting SD at~~ (Table 3). This behavior is expected, as SD patterns within the same area (i.e., at the same sites) tend to recur across different years, such that SD prediction for an unseen snow season (i.e., the temporal framework) is inherently more favorable for XGBoost than prediction at unobserved locations (i.e., the
410 spatial framework). Predicting at unobserved locations and during unseen snow seasons (i.e., the spatio-temporal framework), including times outside the study period is even more challenging, as XGBoost cannot exploit season-specific information from the training data nor rely on recurrent site-specific SD patterns. Consequently, because the intended application of XGBoost is to predict SD at previously unobserved locations across the European Alps ~~–~~

~~The spatial and spatio-temporal frameworks provide a more realistic~~ within the training dataset time period (e.g., Fig. 3a),
415 reliance on the temporal framework would lead to an overly optimistic evaluation of model performance ~~for this study. For,~~ whereas the spatio-temporal framework would provide a more conservative assessment.

The absolute improvements of using PolSAR observations, as a replacement for (or in combination with) backscatter intensities, are small within all nested CV frameworks, or even minimal within some CV frameworks. Within the spatial framework, the PolSAR_{ML} and Backscatter_{ML} configurations display marginal differences in performance metrics (Table 3):
420 ~~This is likely the result of the,~~ which likely stems from the similar relationship of the variables with SD, and the spatial noise present in both types of S1 variables, ~~and the similar relationship of the variables with SD. However, the.~~ The Combination_{ML} configuration exhibits also shows similar overall performance compared to the Backscatter_{ML} configuration, yet the improvements in MAE, computed per site, are significant ($p \ll 0.05$, 95% confidence level) ~~improvements in site~~

MAE, resulting in an overall MAE of 35 cm, a mean site MAE of 29 cm, and a mean site bias of 11 mm (excluding zero-measured SD). These values are slightly higher than those reported by Dunmire et al. (2024), which may stem from differences in satellite observation scaling, input feature configurations (e.g., a cumulative FSC). PolSAR observations moreover improve model performance for the spatio-temporal framework, with significant improvements in site MAE for both the PolSAR_{ML} and Combination_{ML} configurations ($p \ll 0.05$, 95% confidence interval). This improvement suggests that PolSAR observations, alone or in combination with backscatter intensities, more effectively capture the seasonal evolution of SD, a conclusion that is also supported by statistically significant gains ($p \ll 0.05$, and the larger number of sites and years included in the present evaluation; 95% confidence level) observed in the temporal nested CV framework. Nonetheless, also here the absolute improvements remain limited, and the increased computational cost of processing S1 SLC data into PolSAR variables relative to deriving backscatter intensity from GRD data must be considered.

Additionally, all three configurations exhibit difficulties in predicting high SD values at unobserved sites (Fig. ??a3), which we attribute to the relatively small number of observation sites with a large record of observed SD – likely arises from the scarcity of high SD observations in the training dataset (Fig. E1a). For the spatial nested CV framework, the PolSAR_{ML}, Backscatter_{ML} and Combination_{ML} configurations exhibit a bias of -1.46, -1.51 and -1.48 m for observed $SD \geq 2.5$ m. Specifically, at those, which further deteriorates in the spatio-temporal framework. This is also visible in Fig. 3a, when comparing the observations (squares) with the SD prediction, and in Fig. 3b, that displays sites with a site-bias ≤ -0.5 m (mostly located in Switzerland (-0.5 m for the different configurations within the spatial nested CV framework (zero-measured SD excluded). Interestingly, many of the sites displayed in Fig. ??b)) 14.2% of the observed SD measurements exceed 3 m, compared to only 1% across the entire training dataset. The limited 3b also appear to have strong negative biases in Fig. 2a of Dunmire et al. (2024). Consequently, the reduced ability of XGBoost models, trained with S1 variables to explain interannual and site-specific variability, to predict high SD values must be considered when applying taken into account when deploying the model across the European Alps, especially when one wants to apply the model to predict outside the training dataset time period. For the latter, Table 3 indicates a further deterioration in model performance for the spatio-temporal nested CV framework, which must be seen as an accurate representation of the model performance when applying XGBoost to predict across the Alps outside the training period. Nonetheless, the PolSAR variables also improved model outcomes for this scenario, with significant ($p \ll 0.05$, 95% confidence interval) improvements in site MAE for both the PolSAR_{ML} and Combination_{ML} configurations, particularly when applying it beyond the temporal range covered by the training data.

4.2 Feature importance for ML configurations with S1 PolSAR and backscatter intensity variables

Figure 4 displays the importance of the various input features across the PolSAR_{ML}, Backscatter_{ML} and Combination_{ML} configurations, as determined by (a) XGBoost's internal gain metric and (b) SHAP values computed from the outer resampling test sets. Across the configurations and nested CV frameworks, FSC and elevation consistently emerge as the most important features, similar to what has been reported by Dunmire et al. (2024). In addition, the, indicated by both XGBoost's internal gain -based FI indicates an increased importance of FSC metric and SHAP values (Fig. 4a and b). FSC is particularly important, which makes the predictions prone to potential errors in this input feature (e.g., the yellow box in Fig. 3a). The gain

Table 3. Overall performance metrics for the PolSAR_{ML}, Backscatter_{ML} and Combination_{ML} configurations for the spatial and spatio-temporal nested CV frameworks, when evaluated over all sites displayed in Fig. 1. The values inside the brackets denote the metric values when zero-measured SD are included in the evaluation.

Metric	Spatial			Spatio-temporal		
	PolSAR _{ML}	Backscatter _{ML}	Combination _{ML}	PolSAR _{ML}	Backscatter _{ML}	Combination _{ML}
R [-]	0.73 (0.81)	0.73 (0.81)	0.74 (0.82)	0.70 (0.79)	0.69 (0.79)	0.70 (0.79)
RMSE [m]	<u>0.53 (0.41)</u>	<u>0.53 (0.42)</u>	<u>0.52 (0.41)</u>	<u>0.55 (0.43)</u>	<u>0.56 (0.44)</u>	<u>0.55 (0.43)</u>
MAE [m]	0.35 (0.23)	0.35 (0.23)	0.35 (0.22)	0.37 (0.24)	0.37 (0.24)	0.37 (0.24)
bias [m]	<u>0.53 (0.41) 0.04 (0.00)</u>	<u>0.53 (0.42) 0.04 (-0.00)</u>	<u>0.52 (0.41) 0.04 (0.00)</u>	<u>0.55 (0.43) 0.04 (0.00)</u>	<u>0.56 (0.44) 0.04 (0.00)</u>	<u>0.55 (0.43) 0.04 (0.00)</u>

FI importance (Fig. 4a) moreover indicates an increasing FSC importance when transitioning from temporal to spatio-temporal nested CV, ~~whereas elevation follows frameworks (~25% to > 40%),~~ with an opposite trend. ~~In temporal nested CV, the ML model can leverage recurring SD patterns at sites within the training data, where — mainly at high-elevation sites — interannual variability is relatively limited and SD tends to follow similar trajectories across multiple years. As a result, the ML model may place greater emphasis on elevation for elevation (~20 to ~10%).~~ When predicting at unseen locations, XGBoost seems to adjust the predictions. Conversely, in spatio-temporal nested CV, XGBoost may rely less on recurring SD patterns and therefore, seems to place greater emphasis on FSC and less on elevation rely more heavily on FSC to distinguish snow-covered from snow-free periods. Figure 4b, which displays the FI based on the training data, on the other hand, indicates a stronger importance for FSC, but shows a similar increase in fSC-FI for spatio-temporal nested CV. Differently, Fig 4b displays a less clear pattern for the elevation feature. Nevertheless, non-relative mean absolute SHAP values display similar patterns as periods, and cannot rely as much on recurring SD patterns that display a strong relationship with elevation. Finally, the *gain* FI.

Figure DOY features, which help capturing the seasonal variations in the estimates, also emerge as important inputs, as reflected across the configurations in both Fig. 4a ~~also reveals that within PolSAR_{ML} and Backscatter_{ML} configurations, the~~ and b.

When used in the separate configurations, The S1 PolSAR and backscatter intensity features are similarly used during model training. Specifically, S_0^s and $\gamma_{VV}^{0,s}$ appear to be less important than α^s and $\gamma_{CR}^{0,s}$ in the PolSAR_{ML} and Backscatter_{ML} configurations, respectively. In addition, (Fig. 4b shows that a). However, the *gain* FI indicates a stronger importance of α^s is more important in within the PolSAR_{ML} configuration than over $\gamma_{CR}^{0,s}$ is in the Backscatter_{ML} one, which indicates. This observation is further confirmed in Fig. 4b across all nested CV frameworks, indicating that XGBoost seems to effectively extract more information from α^s during model training and SD prediction. Moreover, Consistent with this finding, the Combination_{ML} configuration also exhibits a clear preference of α^s over $\gamma_{CR}^{0,s}$. In addition, both the *gain*-based and SHAP value FI analyses reveal a less distinct separation between the S1 input features in the Backscatter_{ML} configuration, with the SHAP values even suggesting that $\gamma_{VV}^{0,s}$ contributes as much as, or slightly more than, $\gamma_{CR}^{0,s}$. This implies that despite the lower

importance during model training (i.e., the *gain*-based FI), XGBoost seems to rely more on $\gamma_{VV}^{0,s}$ when predicting during SD prediction at unseen locations and/or time periods.

485 Finally, both Fig. 4a and b showcase a consistent preference of the Combination_{ML} configuration for α^s over $\gamma_{CR}^{0,s}$. Although Fig. 4 provides an aggregated view across all sites and the full snow season, including dry and wet snow conditions as well as zero-measured SD, the result aligns with findings from Jans et al. (2025), who reported stronger correlations between α and modeled SD during the accumulation (dry snow) period at most locations in their study area, compared to γ_{CR}^0 . In contrast, our results indicate that S_0 contributes less than γ_{VV}^0 during both training and prediction, despite its stronger correlation with dry snow reported by Jans et al. (2025). This difference may stem from several factors, including a different satellite resolution
490 (1 km in the work of Jans et al. (2025)), the input scaling applied before model training, variations in land cover across sites, and especially the inclusion of both dry and wet snow periods.

4.2.1 SHAP value analysis during snowpack presence

~~We further assessed the impact~~ The SHAP value FI of the S1 variables (Fig. 4b) were further evaluated during snow-covered periods, during both dry and wet snow conditions. To this end, we focused on — encompassing both wet and dry snow periods
495 — within the spatial nested CV framework, as it best aligns this framework matches best with the objective of this research; i.e., predicting SD at unseen locations within the the study: SD prediction across the Alps using the available S1 archive time period. Since the wet snow mask is only available from September 2016 onward, the 2015–2016 snow year was excluded from this analysis.

Fig. 5a showcases that in the presents the FI for both the full set of predictions (fully colored bars), and the subset
500 remaining after excluding the non-valid areas masked within the wet snow product (dark gray lines). Excluding the non-valid areas increases the overall and dry-period FI for both α^s and $\gamma_{CR}^{0,s}$ within the PolSAR_{ML} configuration, the contribution of α^s is similar under both dry and wet snow conditions ($\sim 9\%$), with a slightly lower contribution observed during dry conditions. However, this result is based on a binary classification between the snow conditions that does not account for masked areas, and therefore still includes lower-elevation locations and/or densely forested locations. Without the masked
505 data, the dry-condition relative importance of α^s increases to approximately 13% (Fig. 5a), while the relative importance of the elevation feature decreases (not shown). Since the masked data points predominantly correspond to and Backscatter_{ML} configuration, respectively. As these non-valid areas include low-elevation areas (i.e., valleys) and densely-forested regions, locations where typically low SD values (≤ 1 m), the SHAP value FI suggests that are observed, these results suggests that XGBoost places relatively more importance on these S1 variables at bare, higher-elevation locations with deeper snowpacks;
510 XGBoost places relatively more importance on α^s , a pattern also observed for $\gamma_{CR}^{0,s}$ in the Backscatter_{ML} configuration. that are often characterized by deeper snowpacks. Conversely, the decrease in SHAP value FI when including these non-valid areas indicates the limited added value of the S1 variables to predict SD at low-elevation and/or densely-forested sites. Although no valid wet-snow classification is available for such areas (e.g., Fig. 5d where only low FCF values appear for the wet-snow subplot) and more advanced classification approaches incorporating meteorological data could enable a more
515 accurate separation of dry and wet snow periods, we argue that the current analysis is sufficient given the limited contribution

of S1 variables in these environments. This marginal contribution was also reported by Dunmire et al. (2024) and is further evident in Fig. C1d. As this limitation arises from the raw S1 observations rather than the type of processed S1 observations (i.e., PolSAR vs. backscatter intensities), this limitation should be addressed by incorporating additional input features (e.g., Fig. C2).

520 Figure 5a also explains the more pronounced difference in SHAP value FI between the PolSAR variables, as XGBoost relies more heavily on α^s than on S_0^s across all snow conditions. In contrast, ~~Fig. 5a displays that for~~ the Backscatter_{ML} configuration, ~~XGBoost~~ places greater emphasis on $\gamma_{VV}^{0,s}$ than on $\gamma_{CR}^{0,s}$ during wet snow conditions, which accounts for their nearly equivalent overall SHAP value FI (Fig. 4). In addition, the FI patterns observed in the Combination_{ML} configuration closely resemble those of the PolSAR_{ML} and Backscatter_{ML} configurations, with the notable exception of $\gamma_{CR}^{0,s}$. ~~We attribute this~~
525 ~~to overlapping information between $\gamma_{CR}^{0,s}$ and α^s , with XGBoost now relying~~ Here, XGBoost seems to rely more heavily on α^s at the expense of $\gamma_{CR}^{0,s}$. ~~Additionally, internal compensations arising from the joint use of both PolSAR and backscatter inputs may further reduce the relative importance of $\gamma_{CR}^{0,s}$,~~ which was also evident from Fig. 4, and is also observed at a measurement site located near Prato, Switzerland (46.47°N, 8.72°E; Fig C1c). Similar observations were made by Jans et al. (2025), who reported stronger correlations between α and modeled SD during the accumulation (dry snow) period at most locations in their
530 study area than between modeled SD and γ_{CR}^0 . In addition, S_0 shows a similar SHAP value FI importance as γ_{VV} during dry snow conditions (Fig. 5a, Combination_{ML} configuration), and is even more important at bare high-elevation sites (dark gray lines), further confirming the findings of Jans et al. (2025). In contrast, γ_{VV} is more extensively used during wet snow conditions, resulting in the slightly higher observed overall FI for this S1 variable.

Figure 5b further illustrates the SHAP value contribution of α^s within the ~~Within the~~ PolSAR_{ML} configuration. ~~During dry~~
535 ~~conditions~~(Fig. 5b), XGBoost has learned a positive relationship between α^s and its contribution to the SD estimates. ~~This can be related to,~~ consistent with the findings of Jans et al. (2025), who ~~found that reported increasing α values with dry snowpack accumulation~~ at sites with medium to high LIA, ~~α progressively augments with an increasing snowpack during the accumulation period.~~ However, at certain medium-high elevation sites (1500 - 2500 m; ~~not shown~~), ~~mean e.g., Fig. C1b),~~ both α^s and ~~corresponding its~~ SHAP value contributions remain low, ~~even though predicted (and observed) SD are still~~
540 ~~relatively high.~~ At these locations, there is often no clear increase in α^s with increasing SD; instead, a rise in or even negative (Fig. C1b) despite relatively high predicted and observed SD. Instead, increases in $\gamma_{VV}^{0,s}$ or S_0^s may be observed. ~~We hypothesize that this pattern is more common~~ are observed, a pattern we hypothesize to be more prevalent at sites with low LIA, where Jans et al. (2025) similarly observed decreasing values of α . Furthermore, Fig. 5b and d illustrate that, while for α^s values the associated SHAP values and SHAP value FI remain elevated under wet snow conditions, this is less the case for $\gamma_{CR}^{0,s}$ within
545 the Backscatter_{ML} configuration, suggesting that SD estimation based on α^s might be less sensitive to the presence of liquid water than when based on $\gamma_{CR}^{0,s}$.

Conversely, Fig. 5c displays that XGBoost has learned a dual relationship between the observed values for $\gamma_{VV}^{0,s}$ and the SHAP value contribution. The lower importance of $\gamma_{VV}^{0,s}$ during dry snow conditions can be explained by the findings of Lievens et al. (2019), who noted that volume scattering in VV-polarization within a dry snowpack at C-band is generally too
550 weak to produce a significant impact. However, at certain locations — mostly with low-LIA satellite overpasses — increasing

$\gamma_{VV}^{0,s}$ and associated positive SHAP values may be observed. We hypothesize that in certain similar locations, the where S1 observations are hypothesized to be more sensitive to surface scattering from snow layer interfaces, resulting in rising values for snow layer interfaces, and where Jans et al. (2025) similarly observed decreasing α -values. This also explains the dual relationship revealed for $\gamma_{VV}^{0,s}$; while both α^s and associated SHAP values tend to decrease. During wet snow conditions, however, the $\gamma_{VV}^{0,s}$ SHAP value relationship is the opposite. During these conditions, XGBoost seems to leverage the information contained within the low observed $\gamma_{VV}^{0,s}$ values due to liquid water presence, to positively contribute to the predicted SD.

One of the main limitations of our current approach is that only valid wet snow classifications are available for higher-elevation and sparsely forested sites (Fig 5d), while some pixels labeled as dry may in fact correspond to wet snow conditions, or to patchy snow conditions. Although a more sophisticated classification method that incorporates meteorological data could provide a more accurate separation of dry and wet snow periods, we argue that the current approach remains suitable for identifying the contribution of the satellite variables within XGBoost under varying snow conditions. The wet snow mask provides a reliable classification primarily at higher-elevation sites — locations that typically exhibit deeper snowpacks — where satellite variables contribute more substantially to model performance. At masked locations, by contrast, the contribution of satellite variables is generally limited. This is consistent with findings by Dunmire et al. (2024), who showed that S1 C-band backscatter is most informative in non-forested terrain with $SD > 1$ m, and has limited influence in areas where the wet snow mask does not provide a valid classification.

SHAP value analysis for the spatial nested CV framework during snow presence. (a) Relative SHAP value FI of the satellite input features for the PolSAR_{ML}, Backscatter_{ML} and Combination_{ML} configurations. The dark gray lines represent the FI after masking out non-valid dry snow classifications. (b) Scatter plot of mean α^s values and corresponding SHAP values for the PolSAR_{ML} configuration, taken as the mean per site and snow season during dry and wet snow conditions (only valid classifications). The dots are colored according to the mean predicted SD (SD_{pred}). (c) Same as (b), but for $\gamma_{VV}^{0,s}$ within the Backscatter_{ML} configuration. (d) Same as (c) for $\gamma_{CR}^{0,s}$, but colorized to the site FCF, indicating that the classification primarily targets higher-elevation non-forested areas.

4.2.2 Time series of SD and SHAP values

Finally, we assessed the SHAP value contribution of the S1 variables within the Combination_{ML} configuration during the temporal evolution of the snowpack. To this end, Fig. C1 displays the results for two snow seasons at three distinct sites in Switzerland, all operated by SLF, whereas Fig. ?? represents an additional site near Grindelwald, Switzerland. To provide a realistic overview of the model performance within this study, the time series show the results of the spatial nested CV framework.

Figure C1a shows the time series for a measurement site (46.41N, 8.92E) near Arolla, Valais, located in a bare rock area with a 0% forest cover fraction and surrounded by mountains to the east, north, and west. At this high-elevation and bare site, the added value of α^s under dry snow conditions and $\gamma_{VV}^{0,s}$ under in Fig. 5c: at certain low-LIA locations, there is a limited positive contribution, whereas at the majority of the sites, a strong inverse relationship emerges during wet snow conditions becomes

585 clear. During the 2017–2018 snow season, namely, a clear increase in α^s is observed during the accumulation phase. After reaching peak values around late January, associated with a large snowfall event, α^s and its associated SHAP values remain elevated until the snowpack starts to wet near the end of March. However, the peak occurs before the maximum measured SD in early April, a phenomenon not unique to this site and also reported by Jans et al. (2025), who found an average lag of around 60 days between peak α and peak SD in their study area. In contrast, the observed $\gamma_{VV}^{0,s}$ values do not increase during
590 the accumulation phase, consistent with the findings of Lievens et al. (2019), and contribute minimally to the SD predictions. At certain sites — usually with low-LIA satellite overpasses — however, $\gamma_{VV}^{0,s}$ and its SHAP values do increase during the accumulation period (e.g., Fig. ??). As such, depending on the LIA, α^s and $\gamma_{VV}^{0,s}$ might contribute more or less to the predictions.

During the ablation phase, marked by a decline in $\gamma_{VV}^{0,s}$ in early April, the contribution of α^s drops sharply, in line with
595 decreasing observed values. In contrast, $\gamma_{VV}^{0,s}$ emerges as the most informative, when typically a drop in this S1 feature under these wet snow conditions, with positive SHAP values peaking near its lowest observed backscatter values. Beyond this point, as $\gamma_{VV}^{0,s}$ increases due to the growing influence of superficial scattering processes (Marin et al., 2020), the contribution gradually declines. Similar patterns have been observed at other sites variable is observed (e.g., 46.42N, 8.23E and 46.39N, 7.97E) and is also evident from Fig. 5c. In contrast, the behavior of α^s and its contribution to the SD estimates during the ablation period
600 is less consistent. At some sites (e.g., 46.08N, 7.92E), the drop in $\gamma_{VV}^{0,s}$ coincides with a sudden increase in α^s , whereas at other sites, no such increase is observed or an increase is linked to late-season snowfall events. Further research is required to better understand the interaction between wet snow and α , though this is beyond the scope of this study.

The Arolla site (Fig. C1a) shows a similar seasonal evolution of $\gamma_{VV}^{0,s}$ in 2018–2019 as in 2017–2018, further supporting the findings of . Finally, Fig. 5 on the role of this S1 feature under different snow conditions. In contrast, although b and d
605 illustrate that under wet snow conditions, α^s increases slightly in mid-March, it begins to decline as early as late February, whereas no corresponding sharp decrease is observed in $\gamma_{VV}^{0,s}$ (not shown). While difficult to confirm, this may be related to melt-freeze events and/or metamorphism processes happening in the snowpack starting in early March. Nevertheless, α is also here strongly contribution during snow accumulation, when it is most informative.

The potentially added value of α^s to the SD predictions at higher-elevated, bare locations is further supported by Fig. C1b,
610 that displays the time series at a measurement site located near Prato, Ticino (46.47N, 8.72E), with a similar 0% forest cover fraction but lower elevation (2222 m above sea level). Unlike the Arolla site, however, α^s is not only mainly informative during snow accumulation, but remains relatively elevated throughout February and March for both snow seasons, when the snowpack exceeds 2 m of snow. In addition, following the drop in $\gamma_{VV}^{0,s}$ near the beginning of the accumulation period, a (very) slight increase of $\gamma_{VV}^{0,s}$ is observed toward the ablation phase at this site (not shown), a pattern also observed by Jans et al. (2025) at
615 some locations with very deep (and dry) snowpacks. In addition, Fig. C1b helps to explain the relative SHAP value FI of $\gamma_{CR}^{0,s}$ in Fig. 5a. Although XGBoost captures the increase in remains more elevated compared to $\gamma_{CR}^{0,s}$, leading to higher associated α^s SHAP values in the PolSAR_M configuration than the corresponding $\gamma_{CR}^{0,s}$ during the accumulation period to contribute positively to the predicted SD, the SHAP value contribution is way smaller compared to α^s , indicating a preference for α^s over $\gamma_{CR}^{0,s}$ by XGBoost.

620 Finally, the bottom time series (Fig. C1c) corresponds to a forested (63% FCF, 1868 m above sea level) site in Sobrio, Ticino (46.41N, 8.92E). Despite being consistently classified as dry snow in the time series, this area is actually masked as forest in the CLMS-SWS wet snow product. This site was also analyzed by Dunmire et al. (2024), who concluded that, given the low FI of $\gamma_{VV}^{0,s}$ and $\gamma_{CR}^{0,s}$ in their study, S1 satellite observations should not be used for this location. Our findings support the limited contribution of SHAP values in the S1 features at similar lower and/or forested areas, as we observe minimal added value from both Backscatter_{ML} configuration. This suggests that SD estimation based on α^s and $\gamma_{VV}^{0,s}$ at this site, and the predictions from XGBoost trained without satellite observations are similar to, or even outperform, those based on satellite inputs for the time series at this site (not shown) might be less sensitive to the presence of liquid water than SD estimation based on $\gamma_{CR}^{0,s}$.

4.3 Added value of meteorological forcings and Snowclim SD estimates

Figure 6 illustrates the improvements when incorporating either regionally downscaled meteorological forcing data (Fig. 6b) or Snowclim SD estimates (when predicting at unseen locations and time periods; Fig. 6ea and b), with the most important improvements for high observed SD values SD observations (≥ 2.5 m) when predicting at unseen locations (and time periods). As an example, the bias for SD. The bias for observations ≥ 2.5 m improves by 27 cm when comparing the Weather_{ML} to the Combination_{ML} configuration within the spatial nested CV framework (Fig. 6a), decreasing from -1.47 m to -1.20 m. Incorporating Snowclim SD estimates further reduces this bias to -1.12 m, likely reflecting a stronger correspondence between the modeled and observed high SD values Snowclim estimates and high SD observations than what is captured through derived meteorological variables. Despite these gains, the number of sites with a mean bias ≤ -0.5 m remains similar in both the Weather_{ML} and Snowclim_{ML} configurations, with most of these sites again located in Switzerland (not shown). Although a smaller clustering radius during fold creation could mitigate this underestimation through more balanced fold SD distributions, the current clustering offers a fair evaluation of model performance, indicating a Fig. E2 — reflecting the need for expanded data collection of high observed SD values.

The high SD data collection. Furthermore, while the inclusion of meteorological forcing data and Snowclim SD estimates also help XGBoost better estimate low snow and snow-free conditions at unseen locations, particularly at low-elevation sites (< 1000 m) near the end of the snow season. While the Snowclim_{ML} configuration performs best at high-elevation sites (> 2500 m) during peak snow depth and the ablation period, it tends to overestimate SD at mid-elevation sites (1000–2500 m) near the end of the season, and at high-elevation sites at the beginning of the season, compared to the Combination_{ML} and Weather_{ML} configurations (not shown). Overall or modeled SD estimates reduces the relative SHAP value FI of the S1 input features, these features continue to influence SD predictions (Fig. 6d), however, accounting in total for approximately 9% and 7% of the SHAP value FI in the Weather_{ML} and Snowclim_{ML} configurations, respectively, when predicting at unseen locations.

Overall, the Snowclim_{ML} configuration yields the best performance (Fig. 6ea and b). Under the spatio-temporal nested CV framework, site MAE during snow-covered periods decreases by at least 2.5 cm at 55% of sites, while only 16% exhibit an increase of 2.5 cm or more, relative to the Combination_{ML} configuration. Compared to the Weather_{ML} configuration, the difference in site MAE is minimal, and only significant for the spatio-temporal framework. These results thus indicate that

directly using meteorological forcings can achieve comparable predictive performance, underscoring the potential of machine learning as a valuable complement to physically-based snow models, provided sufficient training data are available.

655 ~~Although the inclusion of~~ However, the *Snowclim* SD estimates appear to more accurately represent conditions at low-elevation sites (< 1000 m) during the early snow season (Fig. 6c), highlighting the importance of reliable FSC information to correctly indicate snow-free conditions when relying solely on meteorological forcing data ~~or modeled SD reduces the relative SHAP value FI of the S1 input features, these features still influence the SD predictions. In the spatial and spatio-temporal nested CV frameworks, S1 features account for approximately 9% and $\sim 7\%$ of the total SHAP value FI for the .~~ In addition, while the *Weather_{ML}* configuration displays the lowest bias of the three configurations at medium-elevation (1000-2500 m) sites (Fig. 6c), the *Snowclim_{ML}* configuration still performs best at high-elevation sites (> 2500 m) during peak SD and the ablation period. Nonetheless, both the *Weather_{ML}* and *Snowclim_{ML}* configurations ~~, respectively, with α^s and $\gamma_{VV}^{0,s}$ contributing the most. Similarly, the relative gain-based FI reveals a non-negligible role for the S1 features, with contributions of around 12% for both configurations and nested CV frameworks~~ seem to underestimate SD near March-May, with the differences being less pronounced for the 1500-2500 m elevation class. As such, the results indicate that directly using meteorological forcings can achieve comparable predictive performance within the applied XGBoost setup. This would eliminate the need to run a snow model, which reduces both computational cost and model complexity.

4.4 Spatial snow depth prediction

To address the findings of López-Moreno et al. (2015) and Miller et al. (2022), who emphasized that stationary site SD measurements are insufficient to represent the spatial variability within the snowpack. To address this limitation, Fig. 7 displays the SD predictions for the nine Dischma valley snow surveys, made with (a and e) the *Combination_{ML}* (b and f) the and point-based SD measurements inadequately capture spatial snowpack variability, we trained the *Combination_{ML}*, *Weather_{ML}* and (e and g) the *Snowclim_{ML}* configurations , all trained with spatial SD data included. Among these, with additional spatially distributed SD training (survey) data, and compared its impact over the Dischma valley. Fig. 7a-c display the results of the *Combination_{ML}* configuration exhibits the highest MAE for the 16 March 2017 snow survey (Fig. 7a), also when trained solely with stationary site data (Fig. E3). However, by including spatial training data, d), in which mountain ridges appear more distinctly in the predictions, and high SD predictions estimates at steep locations and near ridges are more accurately corrected. This effect is also reflected in compared to Fig. 7h, which shows that including E3a-e, which depict results obtained without spatial training data leads to a generally lower SHAP value FI for the topographic input features, with a pronounced drop for the slope. When trained without spatial data, XGBoost namely had incorrectly established a relationship between increasing slopes and predicted SD — explaining the previously high slope FI — leading to high SD. Negative adjustments can be observed at high-elevation areas with (relatively) steep slopes (Fig. E3a). This misrepresentation likely stemmed from the training dataset, that contains few high-elevation stationary measurement sites in steep-slope terrain — sites that moreover display deep snow accumulation during March (data not shown).

685 SD predictions further improve for both the for all configurations, but is most pronounced for the *Combination_{ML}* configuration (Appendix D). For the latter, the negative corrections are reflected in a reduced SHAP value FI of the topographic inputs

(Fig. 7b, survey data; Appendix D). In contrast, the SHAP value FI of the topographic inputs increases in the $Weather_{ML}$ and $Snowclim_{ML}$ configurations, as shown not only in the 16 March 2017 snow survey, but also in the 2D histograms (Fig. 7b, e, f and g), and what was already clear from Fig. 6. Including spatial SD training data mainly leads to better estimates for lower observed SD, whereas the differences with the same configurations, trained without spatial data, are less clear for high (≥ 2.5 m) SD (Fig. E3f and g). Although high SD values are generally better captured by including i and j). Here, the topographic features contributed less when the models were trained solely on point-based measurements but become increasingly influential once spatial training data are included, enabling improved representation of topographic controls on the SD estimates. Notably, α^s remains among the more influential features for these configurations, indicating that despite its reduced relative importance, the inclusion of this S1 variable (and potentially other S1 variables) continues to affect the final SD predictions.

Despite the improvements, SD remains overestimated for the 16 March 2017 snow survey across all configurations. This overestimation is most pronounced for the $Combination_{ML}$ configuration and exceeds the bias reported by Dunmire et al. (2024). Although it is not the objective of this study to perform a direct comparison, the persistence of the $Combination_{ML}$ bias indicates potential limitations in the current feature selection or training dataset. Nonetheless, both the $Weather_{ML}$ and $Snowclim_{ML}$ configurations outperform the results reported by Dunmire et al. (2024). For the 9 March 2016 and 16 March 2017 snow surveys used by Dunmire et al. (2024) for model validation, the $Weather_{ML}$ and $Snowclim_{ML}$ configurations achieve higher correlation coefficients (0.68 and 0.64) and lower mean absolute errors (31 and 36 cm, respectively) than those reported by Dunmire et al. (2024) (0.56 and 41 cm). For the 16 March 2017 snow survey, particularly at locations ≥ 1500 m. The overestimation introduced by survey specifically, Dunmire et al. (2024) reported a mean SD difference (predicted minus observed) of 16 cm, whereas the $Weather_{ML}$ and $Snowclim_{ML}$ configurations exhibit differences of -5 and 14 cm, respectively. Thereby, the more pronounced overestimation of the $Snowclim_{ML}$ SD estimates — on average 53 cm higher than observed values within the $1-2$ m range — configuration likely results from the systematic overestimation in the $Snowclim$ SD estimates for this survey (Fig. D2d), which may also explain its 6 cm higher overall MAE compared to the $Weather_{ML}$ configuration (Fig. 7b and c). Between approximately 1500 and 2500 m, the $Weather_{ML}$ configuration shows a slightly negative bias of 4 cm, whereas $Snowclim_{ML}$ maintains a consistent positive bias of around 17 cm. In the 2500 to 3200 m elevation range a similar pattern is noted, with the $Weather_{ML}$ configuration underestimating the observed values by approximately 7 cm, while $Snowclim_{ML}$ overestimates them by around 13 cm. At the highest elevations, both configurations display less consistent patterns, with instances of both under- and overestimation.

Nevertheless, even with spatial training data included, Across the nine snow surveys, the inclusion of spatially distributed SD training data primarily improves predictions for lower observed SD values (Fig. 7 and E3e–g), whereas differences for higher SD observations (≥ 2.5 m) remain less pronounced, albeit best captured by the $Snowclim_{ML}$ configuration. Despite a persistent positive bias across all configurations, the $Weather_{ML}$ configuration exhibits a bias reduction of approximately 10 cm relative to the $Weather_{ML}$ and $Snowclim_{ML}$ configuration, suggesting that explicitly running $Snowclim$ may not be required when the primary objective is to obtain accurate SD estimates across the European Alps. Indeed, XGBoost effectively learns the relationship between the meteorological forcing data and observed SD, without explicitly representing the physical

processes governing snowpack evolution. For the *Snowlim*_{ML} configurations still overestimate observed values, leading to overall biases of 7 cm and 17 cm across the nine snow surveys, respectively. Albeit it is harder to pinpoint the underlying cause for the configuration, the remaining overestimation largely originates from biases in the *Snowlim* SD estimates themselves (Fig. B1b and Fig. D2d). In contrast, the origin of the positive bias in the *Weather*_{ML} configuration, the (strong) positive discrepancies may be driven by is less straightforward and may reflect either an overestimation of cumulative snowfall ($P_{s,c}$). Alternatively, it is also possible that XGBoost is unable to (indirectly) account for increases or limitations in XGBoost's ability to implicitly represent variations in snow bulk density, such that $P_{s,c}$ may be accurate, but SD is nonetheless overestimated. In the latter case, even accurate snowfall forcing may still result in SD overestimation due to an implicit underestimation of snow density underrepresentation of snow densification processes. Further research should attempt to include factors or use ML models that govern variations in snow density, especially to enhance the potential of predicting short-term variations in SD. Finally, large elevation differences within the relative elevation mismatches within the relatively coarse 500 m forcing data might also lead to inaccurate temperature representations, resulting in incorrect snowfall estimates or an under-/overestimation of the meteorological forcing data may introduce temperature biases, thereby affecting snowfall estimates and the inferred number of melt days.

Similar as for the *Combination*_{ML} configuration, model training with spatial SD data decreases the elevation SHAP value FI for both the *Weather*_{ML} and *Snowlim*_{ML} configurations (Fig. 7i and j). However, here, this reduction appears to be driven by the increased topographic complexity in the training dataset, which weakens the relationship between elevation and SD. Indeed, for both configurations, elevation contributes less consistently across space, while other topographic features play a greater role in adjusting predictions for slopes, mountain ridges and/or local depressions (not shown). In addition, α^s still emerges among the more important features, indicating that despite the decreased importance of this S1 variable, including it (and/or other S1 variables) can still influence the final predictions.

Finally, Fig. 8 illustrates how the estimates are adjusted outside the Dischma valley, following the inclusion of Figure 8 further illustrates the impact of the spatial training data outside the Dischma valley. The top row displays the SD estimates across the Ötztal region (Austria) on 29 January 2018, while the bottom row showcases the results for the Lötschental region (Switzerland), on the same date. For both valleys, training XGBoost with spatial SD data leads to improved the inclusion of spatial training data improved the spatial patterns, manifested by the appearance of mountain ridges and local depressions (Fig. 8a and db and f). These patterns also appear in the difference maps (Fig. 8b and ec and g), and seem to be related to the TPI (Fig. 8e and fd and h). Specifically, it appears that there is negative relationship between increasing TPI, a proxy for the microtopography present in the area, and the adjusted SD estimates. This indicates that after incorporating spatial training data, XGBoost assigns higher SD values to areas with negative TPI (i.e., locations Fig. 8d and h indicate that SD estimates at areas situated below their surroundings) than it would when trained solely on stationary site data. On the contrary, places with high TPI values (i.e., mountain ridges) have been negatively adjusted now. While the relationship with the other topographic features (e.g., aspect) is less clear, Fig. 8e and f also (negative TPI) are positively corrected, whereas the opposite is true near mountain ridges. Figure 8d and h moreover indicate a dependency with the slope of the area. Steeper slopes appear to be corrected more negatively, while the opposite is true for more flat locations, with steeper areas exhibiting more pronounced

negative corrections, while flatter regions show the opposite tendency. The relationship with the other topographic features (e.g., aspect), however, is less clear.

5 Conclusions

760 ~~In this study, we applied an~~ An extreme gradient boosting model (XGBoost) ~~that integrates~~ was applied to assess whether
the integration of Sentinel-1 (S1) C-band synthetic aperture radar (SAR) ~~backscatter or~~ dual-polarized SAR polarimetry
polarimetric (PolSAR) observations, ~~or both, along with regionally downscaled~~ — either as a replacement for, or in combination
with S1 backscatter intensity observations — improves SD predictions across the European Alps relative to a backscatter-intensity-based
configuration. In addition, two extended configurations were evaluated in which the S1 observations were supplemented with
765 either meteorological forcing data ~~, and snow depth (SD) or~~ SD estimates from the ~~physically-based model~~ process-based
Snowclim, ~~to further advance SD predictions across the European Alps. We implemented a threefold nested cross-validation~~
~~(nested CV) framework to analyse model performance when predicting at either unseen time periods, unseen locations, or~~
~~at unseen locations~~ model, to assess both the associated performance gains and the continued contribution of S1 observations
to SD prediction. Results were evaluated at locations and/or during time periods not covered by the training data. We also
770 ~~evaluated XGBoost's ability to generate spatial predictions, assessed the impact of spatial training data on SD estimates, and~~
~~used~~ using a threefold nested cross-validation (nested CV) scheme, while feature importance analysis ~~to better understand~~
~~the relationship between the input features and model output. was employed to assess the contribution of the different input~~
variables during both dry and wet snow periods. In addition, nine photogrammetry snow surveys were used to validate the
spatial prediction capacity of the configurations.

775 Our results ~~show that incorporating S1 C-band PolSAR observations, either in addition to or in place of backscatter, can~~
~~slightly improve SD predictions compared to utilizing an ML model trained with only backscatter as S1 satellite variables.~~
~~Although modest — likely due to the similar informational content and noise in both types~~ showcase modest improvements in
estimated SD with the inclusion of S1 variables — ~~the improvements are most evident when predicting at known locations for~~
~~unseen C-band PolSAR observations, with evident improvements at unseen locations during uncovered~~ time periods, and less
780 ~~pronounced when generalizing to~~ whereas no significant improvements are observed at ~~unseen locations. Thereby~~ However,
feature importance analysis indicates ~~that among the S1 variables, the polarimetric scattering angle (α) contributes the most.~~
~~In addition, FI reveals that XGBoost consistently relies more~~ a stronger reliance on α than on the cross-polarization ra-
tio (γ_{CR}^0) ~~under both wet and dry snow conditions when both PolSAR and backscatter observations are included. XGBoost~~
~~moreover actively exploits~~, during both dry and wet snow periods, while the drop in co-polarized (VV) backscatter intensity
785 (γ_{VV}^0) seems to be actively used during the ablation period ~~to refine its SD estimates. These results showcase the potential of~~
~~using PolSAR observations within non-machine learning applications, e.g., within a conceptual model.~~

~~Next, we revealed that incorporating~~. Incorporating meteorological forcing data or Snowclim SD estimates substantially
improve the predictions, because both help to capture interannual and site-specific variability. Besides improving the predictions
for low-snow and snow-free conditions at low-elevation sites, particularly near the beginning and end of the snow season,

790 ~~these additional input features can help mitigate the underestimation of high observed (≥ 2.5 m) SD values~~ further improved XGBoost's performance substantially, while the model continued to make use of information from the S1 observations, particularly α . Nonetheless, ~~our results reveal that high SD values~~ high SD observations remain systematically underestimated, primarily due to their limited representation in the training dataset, ~~and that the S1 C-band observations may still influence the predictions, reflected through their FI.~~

795 . Finally, our results ~~underscore~~ depict the importance of ~~incorporating spatial~~ spatially distributed SD training data to ~~more accurately~~ capture topographic variability ~~in SD estimates. This is primarily reflected in~~, reflected through a negative relationship between topographic position index (TPI) and adjustments in SD estimates, ~~relative to predictions from the same model~~ XGBoost trained without spatial data. ~~With the inclusion of spatial training data, and meteorological forcings or modeled SD as additional input features, XGBoost predictions achieve a spatiotemporal Pearson correlation coefficient of 0.77 and a~~
800 ~~mean absolute error of approximately 40 cm when compared to airborne photogrammetry snow surveys conducted in the Dischma Valley, Switzerland~~ training data.

Despite these advances, SD values above ~~3~~ approximately 2.5 m remain underestimated, emphasizing the need for more extensive snow monitoring and additional spatial snow surveys within the European Alps. Furthermore, XGBoost operates at the pixel level, limiting its ability to incorporate information from neighboring pixels. Finally, while improving SD estimates is
805 a step toward addressing snow mass knowledge gaps, future research should focus ~~directly on on~~ direct snow mass estimation and integrating spatial dependencies through methods that incorporate information from adjacent pixels.

Code availability. The datasets with the ML experiments, used for visualization and analyses, will be made freely available after review.

Data availability. The Sentinel-1 data is freely available on the Alaska Satellite Facility's Vertex data portal (<https://search.asf.alaska.edu>). IMS and MODIS snow cover data can be downloaded at <https://cmr.earthdata.nasa.gov/search/concepts/C1386246258-NSIDCV0.html> and
810 by using the earthaccess Python package ("MOD10A1F" and "MYD10A1F" collections), respectively. The SAR Wet Snow product can be sourced from <https://land.copernicus.eu/en/products/snow/high-resolution-sar-wet-snow>. Instructions to download the MSWX and MSWEP meteorological data can be found at <https://www.gloh2o.org/>. The Copernicus 30 m digital elevation model can be retrieved from the STAC collection described at https://hda.data.destination-earth.eu/ui/dataset/EO.DEM.DAT.COP-DEM_GLO-30-DTED. Finally, land cover data can be sourced from the Copernicus Land Monitoring Service via: <https://land.copernicus.eu/en/products/global-dynamic-land-cover/copernicus-global-land-service-land-cover-100m-collection-3-epoch-2018-globe>.
815

The majority of the in situ snow depth measurements can be freely downloaded from the various providers across the European Alps:

- IMIS measuring network (Switzerland)
- Manual measuring network (Switzerland)
- METEO FRANCE (France)
- 820 – MeteoTrentino (Italy)

- Open Meteo Data V1 (Italy)
- GeoSphere Austria (Austria)

Additional snow depth measurements were collected through personal communication.

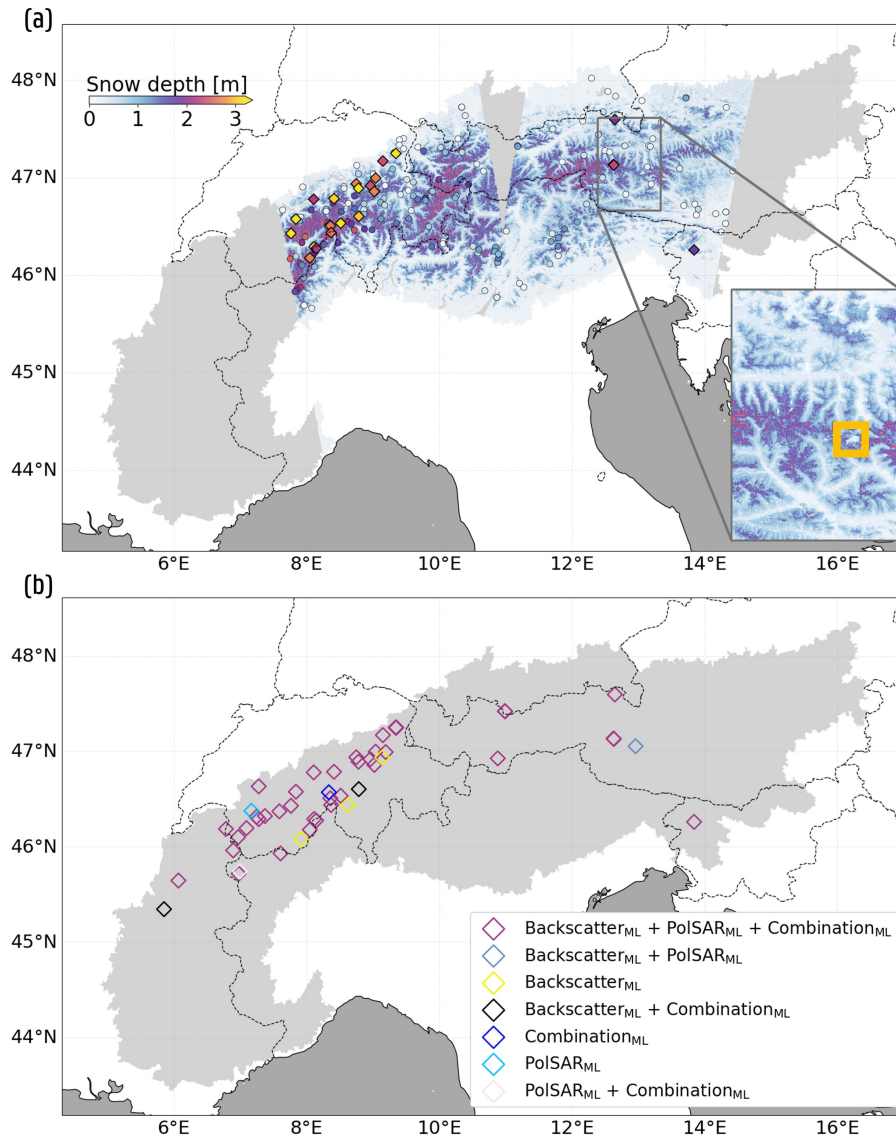


Figure 3. Spatial SD estimate and underestimation of the Combination_{ML} configuration. (a) Prediction on 18 January 2018 encompassing both an ascending and descending S1 observation, derived as the mean SD from the five XGBoost models trained within the spatial nested CV framework. Observations are indicated by dots and squares, with squares representing sites exhibiting a site bias ≤ -0.5 m for this framework (displayed in (b)). Estimates over Austria are highlighted, with the yellow square marking an area of underestimation attributable to errors in the SCF product used. (b) Measurement stations exhibiting a mean site bias ≤ -0.5 m within the spatial nested CV framework (zero-measured SD excluded). Sites with fewer than 10 observations are excluded. Colors indicate the configurations for which this underestimation occurs. Across the displayed sites, 23% of the observed SD measurements ≥ 2.5 m, compared to only 2% across the full training dataset.

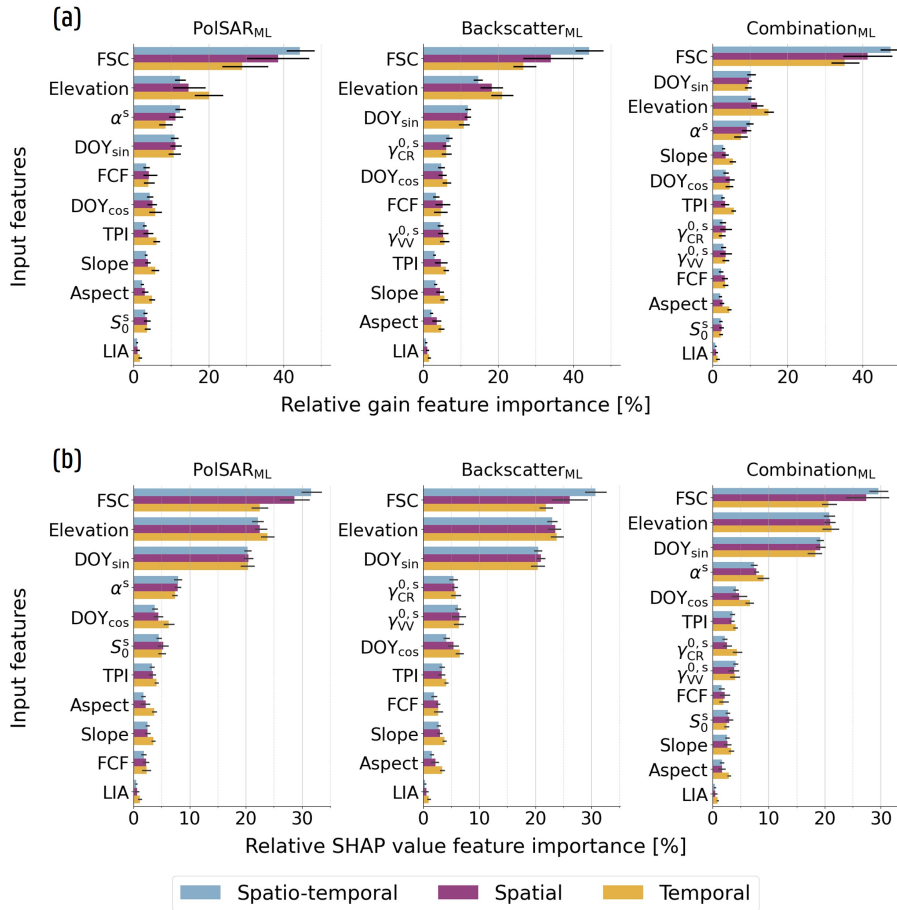


Figure 4. Upper row (a): Relative gain-based FI Feature importance for the PoISAR_{ML}, Backscatter_{ML} and Combination_{ML} configurations, indicating the strong influence of FSC, elevation and DOY on the model outcome (a and b), the elevated usage of α^s and $\gamma_{VV}^{0,s}$ during SD prediction (b), and the preference of the Combination_{ML} configuration for α^s over $\gamma_{CR}^{0,s}$ (a and b). Upper row (a): Relative gain-based FI. Bars represent the mean FI across the XGBoost models trained during inner resampling, with colors indicating the type of nested CV framework. Lower row (b): Similar to (a), but here FI is quantified as the mean absolute SHAP value per feature (relative values), computed separately for each test fold during outer resampling. The black lines in the bars denote the 95% confidence interval.

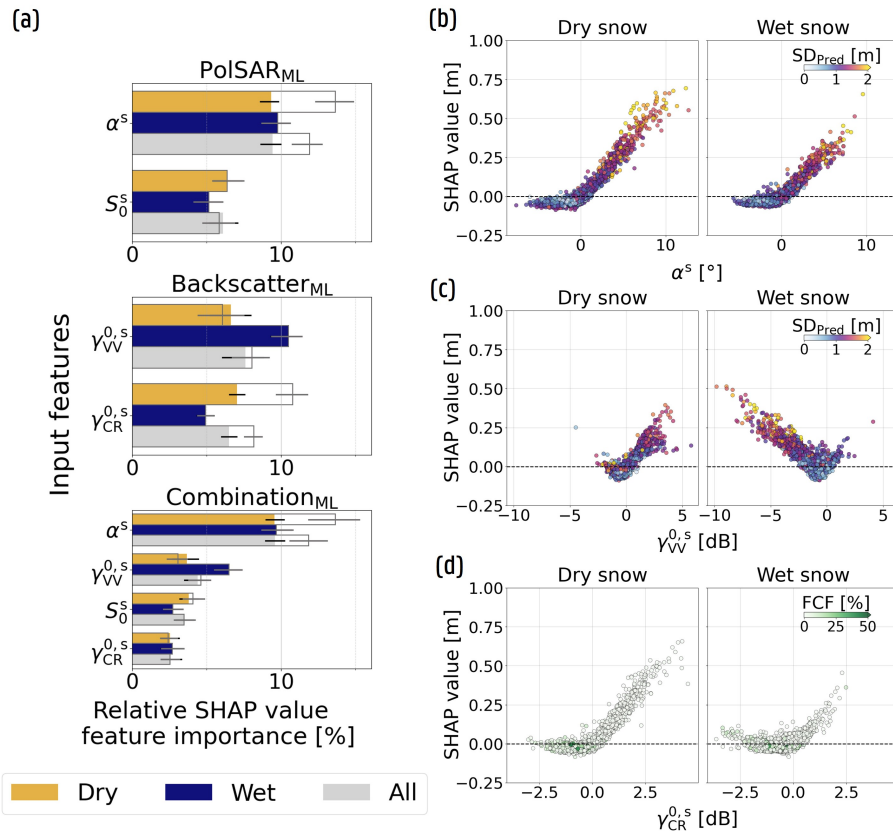


Figure 5. Time-series SHAP value analysis for the $\text{Combination}_{\text{ML}}$ configuration within the spatial nested CV framework during snow presence, indicating XGBoost's stronger reliance of, and preference for three sites within Switzerland, α^s compared to $\gamma_{\text{CR}}^{0,s}$, and the notable use of $\gamma_{\text{VV}}^{0,s}$ during wet snow periods. The time-series display measured (gray) and predicted (light blue) SD. Relative SHAP value FI of the satellite input features for the 2017-2018 $\text{PolSAR}_{\text{ML}}$, $\text{Backscatter}_{\text{ML}}$ and 2018-2019 $\text{Combination}_{\text{ML}}$ configurations. The dark gray lines represent the FI after masking out non-valid dry snow seasons classifications. (b) Scatter plot of mean α^s values and corresponding SHAP values for two-satellite-variable the $\text{PolSAR}_{\text{ML}}$ configuration, representing their contribution in time, are plotted taken as well, the mean per site and snow season during dry and wet snow conditions (only valid classifications). The dots are colored according to their input value in XGBoost the mean predicted SD (SD_{Pred}). The stripes below (c) Same as (b), but for $\gamma_{\text{VV}}^{0,s}$ within the time-series indicate dry $\text{Backscatter}_{\text{ML}}$ configuration. (light gray) or wet (dark blue) snow conditions according for $\gamma_{\text{CR}}^{0,s}$, but colored to CLMS-SWS wet snow product the site FCF, indicating that the classification primarily targets higher-elevation non-forested areas.

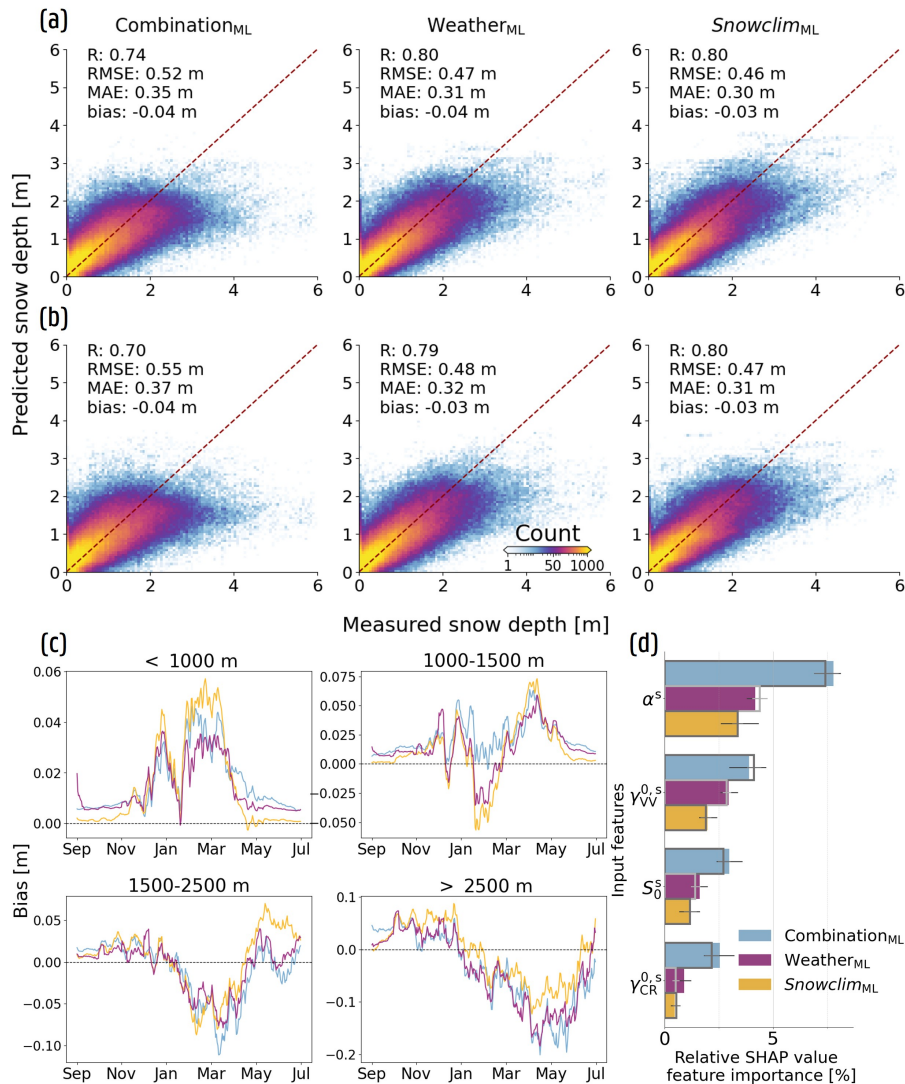


Figure 6. Improvements brought by additional input of meteorological forcing data or *Snowclim* SD estimates when predicting at unseen locations and time periods. (a) 2D histograms of measured vs. predicted SD for different ML configurations and nested CV frameworks: (a) the Combination_{ML}, (b) Weather_{ML}, and (c) Snowclim_{ML} configurations for the spatial nested CV framework. Performance metrics, excluding are computed with exclusion of zero-measured SD. (b) same as (a), are shown in but for the left-upper corner spatio-temporal framework. (c) 7-day mean bias (including zero-measured SD) over the course of a general snow season, for different elevation classes. (d) relative SHAP value FI for the subplots different configurations for the spatial (full colored bars) and the spatio-temporal nested CV framework (gray line).

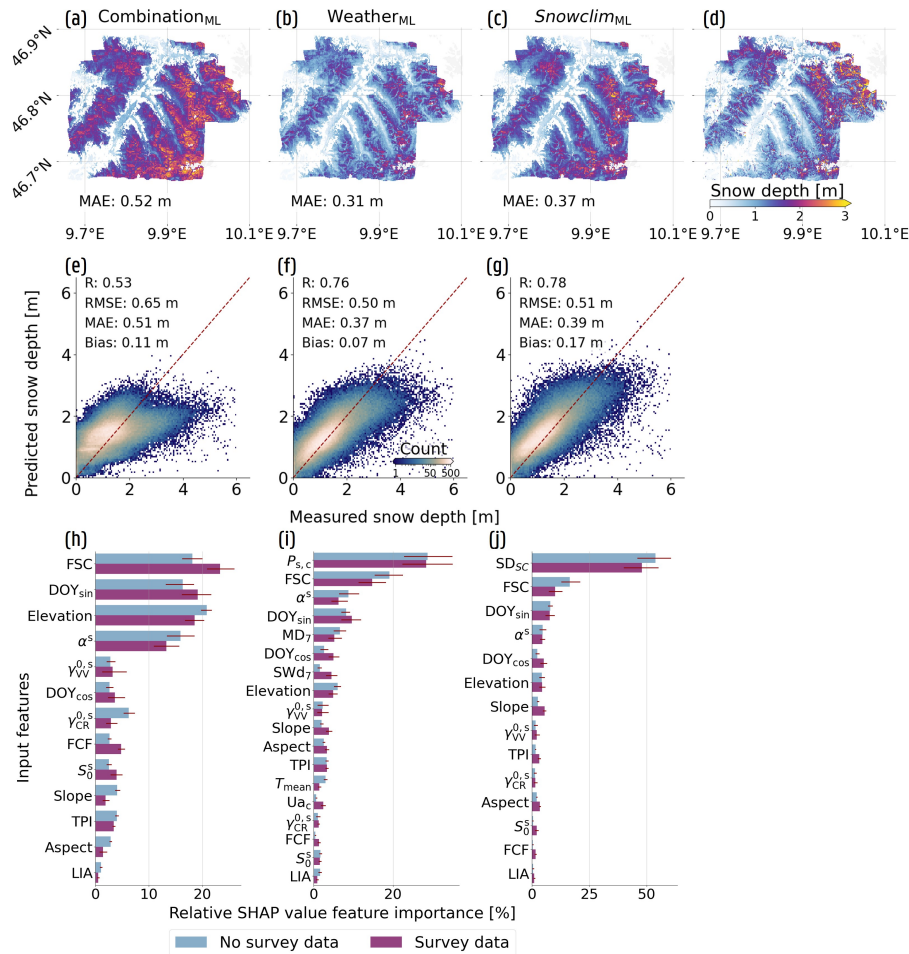


Figure 7. XGBoost spatial prediction performance for Improvements obtained by incorporating spatially distributed SD training data when predicting across the Dischma valley snow surveys Valley (Switzerland), including their impact on SHAP-based FI. The first column (a, e and h) represents the results for the Combination_{ML} configuration, while the second (b, f and i) and third column (c, g and j) show the results for the Weather_{ML} and Snowclim_{ML} configurations, respectively. (a), (b) and (c) Display display the spatial predictions for the 2017-03-16 snow survey, with (d) representing the measured SD. (e), (f) and (g) Show 2D histograms and performance metrics across all nine conducted snow surveys. Finally, (h), (i) and (j) display relative SHAP value FI, both when trained with or without spatial SD (survey) data. The 95% confidence interval is indicated with a red bar.

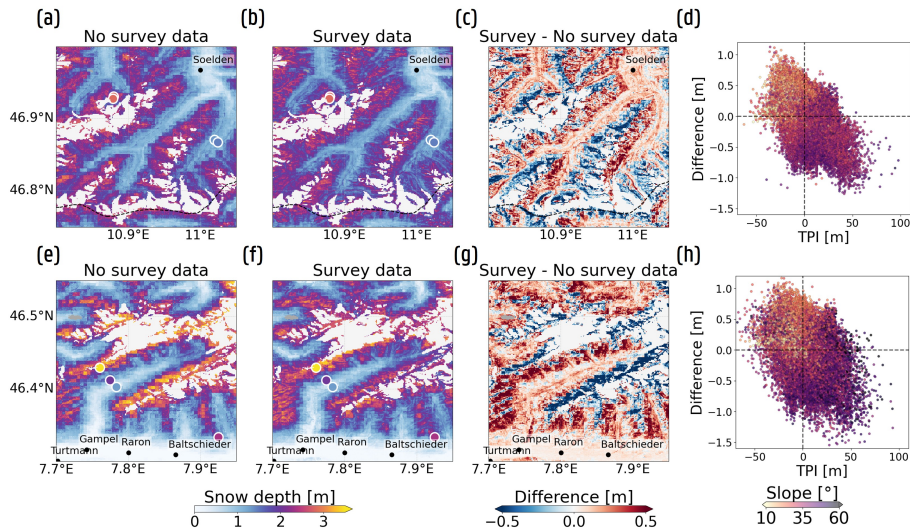


Figure 8. XGBoost-Improved representation of topographical controls on SD predictions-estimates for the Weather_{ML} configuration on 29 January 2018 over the Ötztal, Austria (top row) and the Lötschental, Switzerland (bottom row),-using the Weather_{ML} configuration. (a) and (e) mean-Mean SD estimates predicted with the nine separately trained XGBoost models utilizing no spatial (No survey) training data. (b) and (f) Same as (a) and (e), but with inclusion of spatial SD training data (Survey). (c) and (g) Difference maps between the mean SD predictionpredictions, trained using survey data and SD estimates predicted with XGBoost-trained and without spatial (survey) SD training data. (d) and (h) Scatter plots of TPI and SD differences, colored according to aspectslope. Observed SD at the stationary sites are displayed in (a), (b), (e) and (f) as dots. Glaciers and lakes located within the valleys are colored in white and dark gray, respectively, and are excluded from the analyses.

Appendix A: Combining Terra and Aqua fractional snow cover data

825 To generate a combined daily fractional snow cover (FSC) product from the MODIS Terra and Aqua satellites, we applied a weighted averaging approach to merge the individual satellite observations. As such, we calculated for each location i and day t the weights of the individual FSC data ($W_{\text{sat},i,t}$) using the following formula:

$$W_{\text{sat},i,t} = \frac{1}{W_{\text{QA}} W_{\text{CP}} W_{\text{DD}}} \frac{1}{W_{\text{QA}} \cdot W_{\text{CP}} \cdot W_{\text{DD}}} \quad (\text{A1})$$

830 in which W_{QA} is the weight associated with the quality flag (QA) ~~of the satellite data, the latter ranging from~~, that ranges between 0 to and 2 (with 0 indicating the highest quality) of the satellite data; W_{CP} is the weight based on the number of days since the last cloud-free observation (CP; values between 0 and ∞); and W_{DD} is the weight accounting for the difference in days between the most recent cloud-free Terra and Aqua observations (DD; values between 0 and ∞), where 0 corresponds to the satellite with the most recent cloud-free data). The individual weights were ~~calculated~~ computed using the following ~~formulas~~ formulations, which were designed to ensure that none of the weights attain zero:

$$835 \quad W_{\text{QA}} = \frac{1}{\text{QA} + \frac{(2-\text{QA})}{2}} \quad (\text{A2})$$

$$W_{\text{CP}} = \frac{1}{(\text{CP} + 1)^2} \quad (\text{A3})$$

$$W_{\text{DD}} = \frac{1}{(\text{DD} + 1)} \quad (\text{A4})$$

Appendix B: *Snowlim* parameter ~~setting~~ settings and performance

Snowlim model parameters were calibrated with a two-step approach, involving a subset of the snow dataset. First, we identified the best working downscaling techniques ~~by assessing model performance using standard parameters~~ multiple options were assessed to downscale temperature, precipitation, relative humidity, and downward shortwave radiation — by comparing model performance with measured SD using the standard parameter set used by Lute et al. (2022) (Table 2, superscript ^c) in their full model run. Next, we used the downscaled forcing data to select the optimal parameters out of 1276 tested combinations, by applying a similar approach as described in Lute et al. (2022). A list of the final calibrated parameters is provided in 845 Table B1.

Figure B1 shows the performance across the ~~snow depth dataset, incorporating both SD dataset, for (a)~~ data from the stationary ~~sites and the~~ and point-based measurements, and (b) the Dischma valley photogrammetry snow surveys. Performance metrics are computed excluding zero-measured SD. When ~~including zero values~~ zero values and both stationary site and survey data are included, the Pearson correlation coefficient ~~increases to 0.85~~ is 0.84, while the RMSE and MAE ~~decrease to 0.60~~ display values of 0.64 m and 0.360.40 m, respectively. Compared to the stationary sites alone, the *Snowlim* SD estimates show higher accuracy, with ~~a MAE of 0.22~~ an MAE of 0.40 m (excluding zero-measured SD). In contrast, when only the nine photogrammetry snow surveys are compared with the corresponding *Snowlim* estimates, the MAE deteriorates to ~~0.81~~ 0.80 m.

Table B1. Parameter settings used to generate [snow depth SD](#) estimates across the European Alps. For the detailed description of the separate parameters, we refer to Lute et al. (2022).

Parameter	Abbreviated name	Calibrated values/methods	Units
Albedo algorithm	<i>albedo_opt</i>	Essery ^a	-
Momentum roughness length	z_0	10^{-5}	m
Heat and vapor roughness length	z_h	$z_0/10$	m
Maximum albedo	<i>albedo_max</i>	0.90	-
Maximum liquid water fraction	<i>lw_max</i>	0.1	-
Windless heat exchange coefficient	<i>EO</i>	0.1	$\text{Wm}^{-2}\text{K}^{-1}$
Windless heat exchange coefficient flux application	<i>EO_app</i>	1	-
Windless heat exchange coefficient stability condition	<i>EO_stability</i>	2	-
Cold content threshold at which to start energy tax	cc_0	-5000	kJm^{-2}
Cold content range to tax	cc_1	-5000	kJm^{-2}
Maximum tax to apply to snow cover energy	maxtax	0.9	-
Snow cover energy flux smoothing window	<i>smooth_hrs</i>	24	h
Snow surface temperature augmentation	T_{add}	2	$^{\circ}\text{C}$

^a Essery et al. (2013).

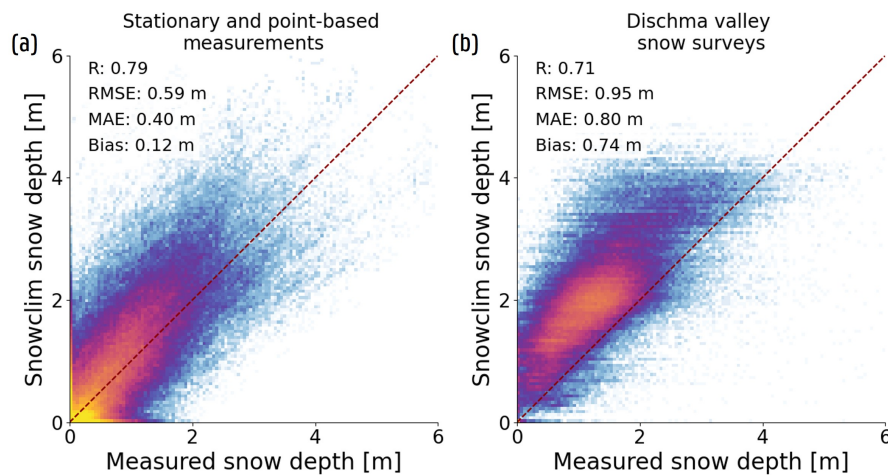


Figure B1. Performance of the process-based model *Snowlim* for the snow measurements dataset, indicating a general overestimation, particularly for the Dischma valley snow surveys. (a) 2D histogram comparing measured SD (at the stationary sites and photogrammetry snow surveys) with the *Snowlim* SD estimates of. (b) Same as (a), but for the physically-based model *Snowlim* in the Dischma valley snow surveys. Performance-Displayed performance metrics do not incorporate zero-measured SD.

Appendix C: Time series of SD and SHAP values

855 The SHAP value contributions of the S1 variables within the Combination_{ML} configuration were assessed during the 2017–2018 and 2018–2019 snow seasons at four distinct sites in Switzerland (Fig. C1). All sites are operated by the Swiss Institute for Snow and Avalanche Research (SLF; Section 2.1), and the results shown correspond to the spatial nested CV framework. Figure C1a shows the time series for a measurement site near Arolla, Valais (46.41°N, 8.92°E), located in a bare rock area with a 0% FCF and surrounded by mountains to the east, north, and west. At this high-elevation and bare site, the added value of α^s under dry snow conditions and $\gamma_{VV}^{0,s}$ under wet snow conditions becomes clear. During the 2017–2018 snow season, 860 namely, a clear increase in α^s is observed during the accumulation phase. After reaching peak values around late January, associated with a large snowfall event, α^s and its associated SHAP values remain elevated until the snowpack starts to wet near the end of March. However, the peak occurs before the maximum measured SD in early April, a phenomenon not unique to this site and also reported by Jans et al. (2025), who found an average lag of around 60 days between peak α and peak SD in their study area. In contrast, the observed $\gamma_{VV}^{0,s}$ values do not increase during the accumulation phase, consistent with the findings 865 of Lievens et al. (2019), and contribute minimally to the SD predictions. At certain sites — usually with low-LIA satellite overpasses — however, $\gamma_{VV}^{0,s}$ and its SHAP values do increase during the accumulation period (e.g., a site near Grindelwand shown in Fig. C1b). As such, depending on the LIA, α^s and $\gamma_{VV}^{0,s}$ might contribute more or less to the predictions during dry snow conditions.

During the ablation phase, marked by a decline in $\gamma_{VV}^{0,s}$ in early April, the contribution of α^s drops sharply, in line with 870 decreasing observed values. In contrast, $\gamma_{VV}^{0,s}$ emerges as the most informative S1 feature under these wet snow conditions, with positive SHAP values peaking near its lowest observed backscatter intensities. Beyond this point, as $\gamma_{VV}^{0,s}$ increases due to the growing influence of superficial scattering processes (Marin et al., 2020), the contribution gradually declines. Similar patterns have been observed at other sites (e.g., 46.42°N, 8.23°E and 46.39°N, 7.97°E) and is also evident from Fig. 5c. In contrast, the behavior of α^s and its contribution to the SD estimates during the ablation period is less consistent. At some sites 875 (e.g., 46.08°N, 7.92°E), the drop in $\gamma_{VV}^{0,s}$ coincides with a sudden increase in α^s , whereas at other sites, no such increase is observed or an increase is linked to late-season snowfall events. Further research is required to better understand the interaction between wet snow and α , though this is beyond the scope of this study.

The Arolla site (Fig. C1a) shows a similar seasonal evolution of $\gamma_{VV}^{0,s}$ in 2018–2019 as in 2017–2018, further supporting the findings of Fig. 5 on the role of this S1 feature under different snow conditions. In contrast, although α^s increases slightly in 880 mid-March, it begins to decline as early as late February, whereas no corresponding sharp decrease is observed in $\gamma_{VV}^{0,s}$. While difficult to confirm, this may be related to melt-freeze events and/or metamorphism processes happening in the snowpack starting in early March. Nevertheless, α^s also here strongly contributes during snow accumulation, when it is most informative.

The potentially added value of α^s to the SD predictions at higher-elevated, bare locations is further supported by Fig. C1c, 885 that displays the time series at a measurement site located near Prato, Ticino (46.47°N, 8.72°E), with a similar 0% FCF but lower elevation (2222 m above sea level). Unlike the Arolla site, however, α^s is not only mainly informative during snow

890 accumulation, but remains relatively elevated throughout February and March for both snow seasons, when the snowpack exceeds 2 m of snow. Nonetheless, the SD predictions appear to level off with only a limited response to new snowfall events, which is observed at the Arolla site during the 2017-2018 snow season as well. One explanation lies in the limited number of high SD observations used to train XGBoost (Fig. E1a), which constrains the model's ability to represent higher SD values and results in a flattening of the predicted SD range (Fig. 6a). Differently, there might be a saturation, or even already a decrease, in the S1 variables that respond to increasing SD at the beginning of the snow season, so that XGBoost cannot rely on these features to account for new snow events. This leveling-off should be taken into account when deploying XGBoost to predict SD across the Alps, especially when relying solely on S1 variables and FSC to explain interannual variability. Additionally, 895 Fig. C1c helps to explain the relative SHAP value FI of $\gamma_{CR}^{0,s}$ in the bottom panel of Fig. 5a. Although XGBoost captures the increase in $\gamma_{CR}^{0,s}$ during the accumulation period to contribute positively to the predicted SD, the SHAP value contribution is way smaller compared to α^s , indicating a preference for α^s over $\gamma_{CR}^{0,s}$ by XGBoost.

Finally, the bottom time series (Fig. C1d) corresponds to a forested (63% FCF, 1868 m above sea level) site in Sobrio, Ticino (46.41°N, 8.92°E). Despite being consistently classified as dry snow in the time series, this area is actually masked as forest in 900 the CLMS SWS wet snow product. This site was also analyzed by Dunmire et al. (2024), who concluded that, given the low FI of $\gamma_{VV}^{0,s}$ and $\gamma_{CR}^{0,s}$ in their study, S1 satellite observations should not be used for this location. Our findings support the limited contribution of the S1 features at similar lower and/or forested areas, as we observe minimal added value from both α^s and $\gamma_{VV}^{0,s}$ at this site. Within such sites, the predictions from XGBoost trained without satellite observations (No S1) are similar to, or even outperform, those based on satellite inputs (Fig. C2). Differently, additional inputs such as meteorological forcings can 905 also improve the results (Fig. C2).

Appendix D: 16 March 2017 Dischma valley snow survey SD estimates

Figure D1 illustrates differences in predicted SD values for the Combination_{ML} configuration trained without (No survey; b) and with (Survey; c) spatially distributed SD training data. Although substantial overestimation persists after incorporating spatial SD training data, excessively high SD predictions at steep slopes and near mountain ridges are more effectively corrected. Two 910 regions exhibiting pronounced corrections are highlighted in Fig. D1a and e. Figure D1a shows the difference between panels (b) and (c), together with the associated slope values, and the corresponding slope SHAP values for the configuration trained without spatially distributed SD data. Areas characterized by relatively steep slopes exhibit strong positive SHAP values, indicating that XGBoost previously learned an erroneous relationship between increasing slope and positive contributions to SD estimates, leading to inflated SD predictions at this high-elevation area with steep terrain. Such regions are negatively 915 adjusted in Fig. D1c, as evident in both zoomed-in areas. Figure D1e, however, further highlights the influence of the TPI, in addition to the slope. Comparison of the central and right panels of Fig. D1e shows that areas characterized by strongly positive TPI values are also substantially negatively adjusted. Conversely, SD estimates are substantially positively adjusted in regions characterized by lower slope values following the inclusion of spatially distributed SD training data.

Table E1. Overall performance metrics for the PolSAR_{ML}, Backscatter_{ML} and Combination_{ML} configurations for the temporal nested CV framework. The values inside the brackets denote the metric values when zero-measured SD are included in the evaluation.

Metric	PolSAR _{ML}	Backscatter _{ML}	Combination _{ML}
R [-]	0.83 (0.88)	0.83 (0.88)	0.84 (0.89)
MAE [m]	0.30 (0.19)	0.30 (0.19)	0.29 (0.18)
RMSE [m]	0.43 (0.33)	0.44 (0.34)	0.42 (0.33)
Bias [m]	<u>-0.02 (0.01)</u>	<u>-0.01 (0.01)</u>	<u>0.02 (0.01)</u>

920 Including meteorological forcing data or *Snowclim* SD estimates further improves SD predictions for the 16 March 2017 snow survey, as shown in Fig. 7b and c, and is reflected in Fig. D2a. Without spatially distributed SD training data (No survey), both the Weather_{ML} and *Snowclim*_{ML} configurations generally overestimate SD values, except at the lowest elevations (<1500 m). Here, an underestimation is observed for both configurations (Fig. D2a), particularly for the Ephemeral snow class (Fig. D2b). Also the *Snowclim* estimates reveal an underestimation in these areas (Fig. D2d), whereas the meteorological forcings may display overly warm temperature estimates or an undercatch of preceding snowfall events.

925 For the mid-elevation regions between 1500 and 2500 m the Weather_{ML} configuration shows an approximately zero bias, whereas the *Snowclim*_{ML} configuration retains a persistent overestimation of about 17 cm. These areas are dominated by the Prairie, Montane Forest, and Maritime snow classes (Fig. D2c), which explains the biases observed for the Montane Forest and Prairie snow classes in Fig. D2b. At the highest elevations, dominated by the Tundra snow class with only a few Ice-class locations, the *Snowclim*_{ML} configuration continues to overestimate SD, while the Weather_{ML} configuration again
 930 exhibits a slight underestimation, as reflected in Fig. D2b. Nonetheless, both configurations underestimate high (> 2.5 m) SD observations, resulting in an underestimation of -1.34 and -1.16 m for the Weather_{ML} and *Snowclim*_{ML} configurations.

Appendix E: Additional tables and figures

XGBoost prediction performance at unseen locations (spatial nested CV framework) across the Backscatter_{ML}, PolSAR_{ML}, and Combination_{ML} configurations. (a) 2D histograms of observed versus predicted SD. (b) Sites with a mean site bias (observed minus predicted SD) > 0.5 m. Measurement stations with fewer than 10 observations are excluded. Colors indicate the configurations for which the site bias exceeds 0.5 m.

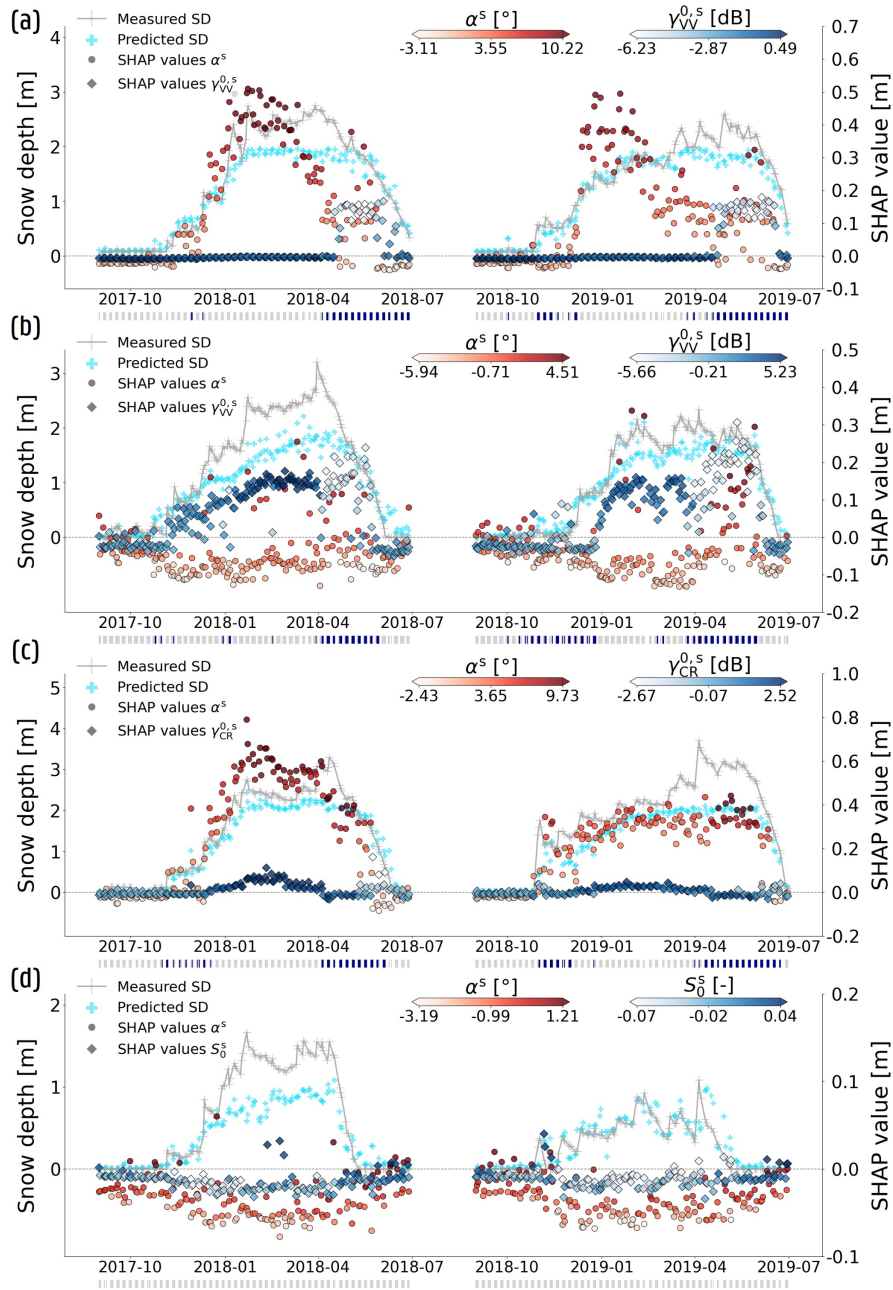


Figure C1. SD predictions (spatial nested CV framework) and associated contribution of various S1 input features for the 2017-2018 and 2018-2019 snow seasons, at four SLF sites within Switzerland: (a) Arolla: 46.41°N, 8.92°E, (b) Grindelwand: 46.67°N, 8.06°E, (c) Prato: 46.47°N, 8.72°E, and (d) Sobrio: 46.41°N, 8.92°. Measured SD is shown in gray, while predictions are denoted by light blue crosses. SHAP values for two S1 variables, representing their contribution in time, are plotted as well, and colored according to their input value in XGBoost. The stripes below the time series indicate dry (light gray) or wet (dark blue) snow conditions according to CLMS SWS wet snow product.

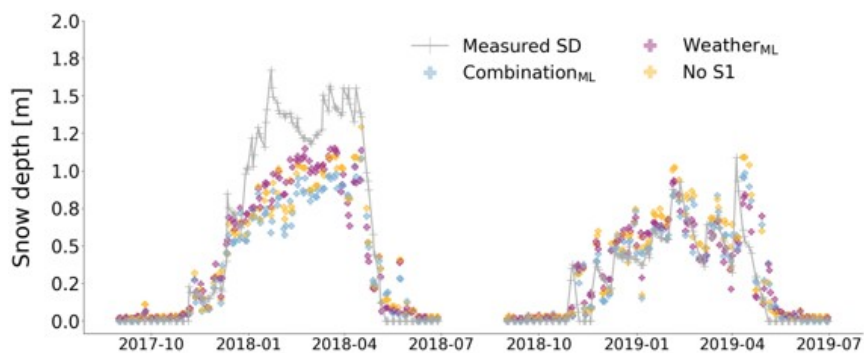


Figure C2. Time series for the spatial nested CV framework at an SLF site near Prato, Ticino in Switzerland (46.41°N, 8.92°), displaying the added value of meteorological forcings, and the performance without inclusion of S1 variables. The time series display measured (gray) and predicted (crosses) SD for the 2017-2018 and 2018-2019 snow seasons. The colors indicate the configuration used: Combination_{ML} (blue), Weather_{ML} (purple) and a configuration using no S1 variables, nor meteorological forcings or *Snowclim* SD estimates (No S1; yellow).

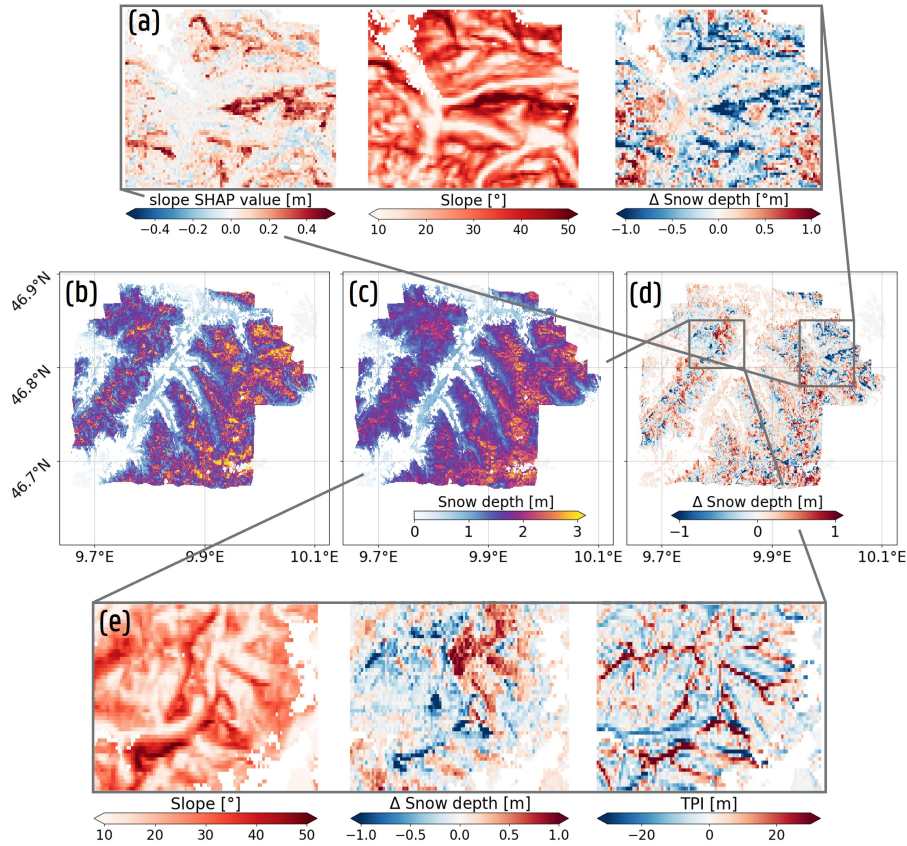


Figure D1. Improvements obtained by adding spatially distributed SD training data (Survey) within the $\text{Combination}_{\text{ML}}$ configuration for the 16 March 2017 Dischma valley snow survey. (a) Zoom-in of the difference map (Survey (b) - No survey (c) SD estimates), with corresponding slope and slope SHAP values. The slope SHAP values correspond with the No survey SD estimates. (b) SD estimates from the $\text{Combination}_{\text{ML}}$ configuration when no spatially distributed SD data is used during training. (c) Same as (b), but with inclusion of spatial SD training data. (d) Difference map of predicted SD between (b) and (c). (e) Zoom-in of (d), with corresponding slope and TPI values for the area.

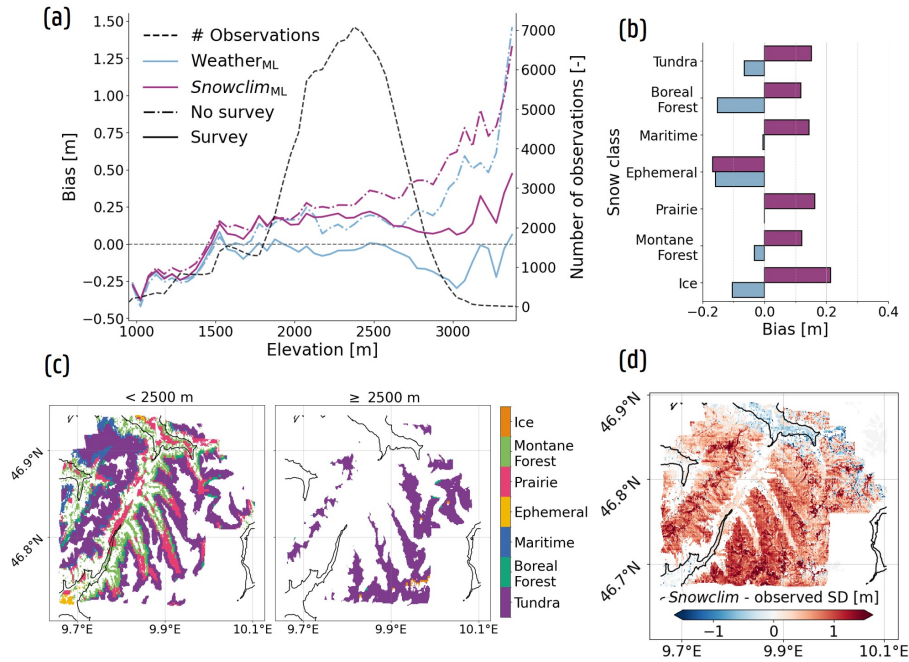


Figure D2. Improvement brought by adding spatially distributed SD training data (Survey) within the Weather_{ML} and Snowclim_{ML} configurations for the 16 March 2017 Dischma valley snow survey. (a) Bias for the two configurations within elevation bins of 50 m, trained without (No survey) and with (Survey) spatial SD training data. The black curve displays the amount of observations per elevation bin. (b) Bias displayed per Sturm and Liston (2021) snow class. (c) Spatial distribution of the snow classes within different elevation areas. (d) Difference map between Snowclim SD estimates and measured SD. The black lines in (c) and (d) denote the contour line of 1500 m elevation.

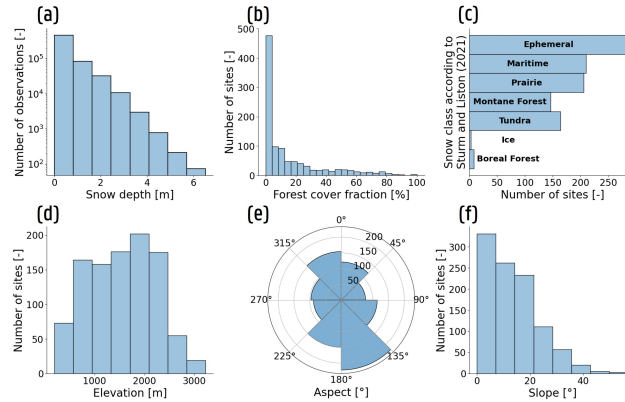


Figure E1. Distributions of measured SD and static features of the in situ measurement sites used to train, validate and test XGBoost. (a) Distribution of measured SD. (b) Distribution of forest cover fraction of the unique measurement sites. (c), (d) (e) and (f) Same as (b), but for the snow classes described in Sturm and Liston (2021), elevation, aspect and slope, respectively.

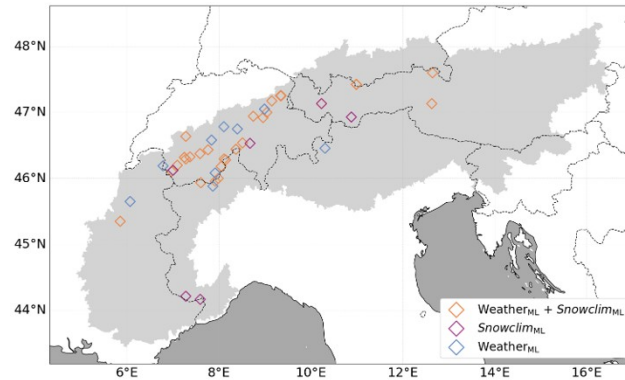


Figure E2. Time-series for Measurement stations exhibiting a measurement mean site located in Switzerland (46.67N, 8.06E). Predictions are made with the Combination_{ML} configuration bias < -0.5 m within the spatial nested CV framework. Measured (gray) and predicted (light blue) SD are displayed, along with SHAP values for two satellite variables, representing their contribution in time the Weather_{ML} and Snowclim_{ML} configurations. The SHAP values Sites with fewer than 10 observations are colored according to their input value in the model excluded. The stripes below the time series Colors indicate dry (light gray) or wet (dark blue) snow conditions according to the CLMS SWS wet snow product for which configurations a mean site bias < -0.5 m is observed.

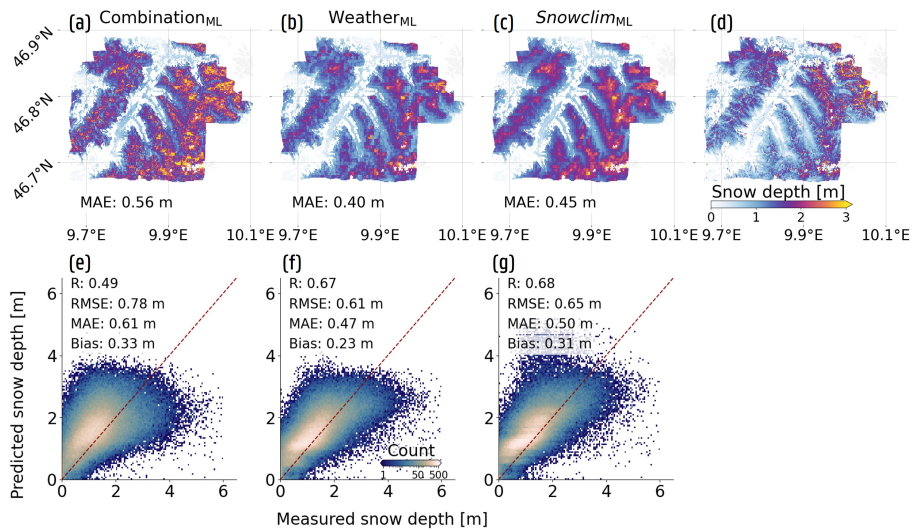


Figure E3. XGBoost spatial prediction performance for the Dischma valley snow surveys with no spatial data included during model training and tuning. The first column (a and e) represents the $Combination_{ML}$ configuration, while the second and third column show the results for the $Weather_{ML}$ (b and f) and $Snowclim_{ML}$ (c and g) configurations, respectively. (a), (b) and (c) Display the spatial predictions for the 2017-03-16 snow survey, with (d) representing the measured SD. (e), (f) and (g) Show 2D histograms and performance metrics across all nine conducted snow surveys.

Author contributions. LB, HL, DD, WW and GDL designed the study. LB and EB worked on the data preparation, and LB conducted the experiments discussed in the manuscript. The analyses, original draft preparation and visualization was performed by LB. All other authors reviewed and edited the manuscript before first submission.

940 *Competing interests.* The authors declare that there are no competing interests.

Acknowledgements. This work was funded by the SNOWTRANE project (SR/00/407) of the Belgian Science Policy (~~Belspe~~[Office \(BELSPO\)](#)). The Flemish ~~Supereom-410-puter~~[Supercomputer](#) Center, funded by ~~FWO~~[Research Fund – Flanders \(FWO\)](#) and the Flemish Government, was used for computational resources and services. Next, the authors employed chatGPT to further improve the readability and language of the manuscript. Before submission, the authors edited and reviewed the manuscript, and take full responsibility for its content.

945 References

- Abdulaal, M. J., Casson, A. J., and Gaydecki, P.: Performance of nested vs. non-nested SVM cross-validation methods in visual BCI: Validation Study, in: 2018 26th European Signal Processing Conference (EUSIPCO), pp. 1680–1684, <https://doi.org/10.23919/EUSIPCO.2018.8553102>, 2018.
- Aureliano, P.: Perimeter of the Alpine Convention, [https://www.atlas.alpconv.org/layers/geonode:Alpine_Convention_Perimeter_2018_v2#](https://www.atlas.alpconv.org/layers/geonode:Alpine_Convention_Perimeter_2018_v2#more)
950 more, 2020.
- Bales, R. C., Molotch, N. P., Painter, T. H., Dettinger, M. D., Rice, R., and Dozier, J.: Mountain hydrology of the western United States, *Water Resources Research*, 42, <https://doi.org/10.1029/2005WR004387>, 2006.
- Barnett, T. P., Adam, J. C., and Lettenmaier, D. P.: Potential impacts of a warming climate on water availability in snow-dominated regions, *Nature*, 438, 303–309, <https://doi.org/10.1038/nature04141>, 2005.
- 955 Bartelt, P. and Lehning, M.: A physical SNOWPACK model for the Swiss avalanche warning: Part I: numerical model, *Cold Regions Science and Technology*, 35, 123–145, [https://doi.org/https://doi.org/10.1016/S0165-232X\(02\)00074-5](https://doi.org/https://doi.org/10.1016/S0165-232X(02)00074-5), 2002.
- Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., van Dijk, A. I. J. M., McVicar, T. R., and Adler, R. F.: MSWEP V2 global 3-Hourly 0.1° precipitation: methodology and quantitative assessment, *Bulletin of the American Meteorological Society*, 100, 473 – 500, <https://doi.org/10.1175/BAMS-D-17-0138.1>, 2019.
- 960 Beck, H. E., van Dijk, A. I. J. M., Larraondo, P. R., McVicar, T. R., Pan, M., Dutra, E., and Miralles, D. G.: MSWX: global 3-Hourly 0.1° bias-corrected meteorological data including near-real-time updates and forecast ensembles, *Bulletin of the American Meteorological Society*, 103, E710 – E732, <https://doi.org/10.1175/BAMS-D-21-0145.1>, 2022.
- Blancas, E.: Model selection done right: A gentle introduction to nested cross-validation, <https://ploomber.io/blog/nested-cv/>, accessed: 2025-06-02, 2021.
- 965 Bormann, K. J., Brown, R. D., Derksen, C., and Painter, T. H.: Estimating snow-cover trends from space, *Nature Climate Change*, 8, 924–928, <https://doi.org/10.1038/s41558-018-0318-3>, 2018.
- Brangers, I., Lievens, H., Getirana, A., and De Lannoy, G. J. M.: Sentinel-1 snow depth assimilation to improve river discharge estimates in the western European Alps, *Water Resources Research*, 60, e2023WR035 019, <https://doi.org/https://doi.org/10.1029/2023WR035019>, 2024a.
- 970 Brangers, I., Marshall, H.-P., De Lannoy, G., Dunmire, D., Mätzler, C., and Lievens, H.: Tower-based C-band radar measurements of an alpine snowpack, *The Cryosphere*, 18, 3177–3193, <https://doi.org/10.5194/tc-18-3177-2024>, 2024b.
- Broxton, P., Ehsani, M. R., and Behrangi, A.: Improving mountain snowpack estimation using machine learning with Sentinel-1, the Airborne Snow Observatory, and University of Arizona snowpack data, *Earth and Space Science*, 11, e2023EA002 964, <https://doi.org/10.1029/2023EA002964>, 2024.
- 975 Buchhorn, M., Smets, B., Bertels, L., De Roo, B., Lesiv, M., Tsendbazar, N.-E., Herold, M., and Fritz, S.: Copernicus global land service: land cover 100 m: collection 3: epoch 2018: Globe, <https://doi.org/10.5281/zenodo.3518038>, 2020.
- Bührlé, L. J., Marty, M., Eberhard, L. A., Stoffel, A., Hafner, E. D., and Bühler, Y.: Spatially continuous snow depth mapping by aeroplane photogrammetry for annual peak of winter from 2017 to 2021 in open areas, *The Cryosphere*, 17, 3383–3408, <https://doi.org/10.5194/tc-17-3383-2023>, 2023.
- 980 Burakowski, E. and Magnusson, M.: Climate impacts on the winter tourism economy in the United States, 2012.

- Charbonneau, A., Deck, K., and Schneider, T.: A Physics-Constrained Neural Differential Equation Framework for Data-Driven Snowpack Simulation, *Artificial Intelligence for the Earth Systems*, 4, <https://doi.org/10.1175/aies-d-24-0040.1>, 2025.
- Chen, T. and Guestrin, C.: XGBoost: A scalable tree boosting system, in: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 785–794, Association for Computing Machinery, New York, NY, USA, 985 <https://doi.org/10.1145/2939672.2939785>, 2016.
- Cloude, S. and Pottier, E.: An entropy based classification scheme for land applications of polarimetric SAR, *IEEE Transactions on Geoscience and Remote Sensing*, 35, 68–78, <https://doi.org/10.1109/36.551935>, 1997.
- Copernicus Land Monitoring Service: SAR wet snow 2016-present (raster 60 m), Europe, daily, <https://land.copernicus.eu/en/products/snow/high-resolution-sar-wet-snow>, 2024.
- 990 Craymer, M.: Geodetic toolbox, MATLAB Central File Exchange, 2022.
- Daudt, R. C., Wulf, H., Hafner, E. D., Bühler, Y., Schindler, K., and Wegner, J. D.: Snow depth estimation at country-scale with high spatial and temporal resolution, *ISPRS Journal of Photogrammetry and Remote Sensing*, 197, 105–121, <https://doi.org/https://doi.org/10.1016/j.isprsjprs.2023.01.017>, 2023.
- De Lannoy, G. J. M., Bechtold, M., Busschaert, L., Heyvaert, Z., Modanesi, S., Dunmire, D., Lievens, H., Getirana, A., and Mas- 995 sari, C.: Contributions of irrigation modeling, soil moisture and snow data assimilation to high-resolution water budget estimates over the Po basin: Progress towards digital replicas, *Journal of Advances in Modeling Earth Systems*, 16, e2024MS004433, <https://doi.org/https://doi.org/10.1029/2024MS004433>, e2024MS004433 2024MS004433, 2024.
- Deems, J. S., Painter, T. H., and Finnegan, D. C.: Lidar measurement of snow depth: A review, *Journal of Glaciology*, 59, 467–479, <https://doi.org/10.3189/2013JoG12J154>, 2013.
- 1000 Dora, L., Agrawal, S., Panda, R., and Abraham, A.: Nested cross-validation based adaptive sparse representation algorithm and its application to pathological brain classification, *Expert Systems with Applications*, 114, 313–321, <https://doi.org/https://doi.org/10.1016/j.eswa.2018.07.039>, 2018.
- Dozier, J., Bair, E. H., and Davis, R. E.: Estimating the spatial distribution of snow water equivalent in the world’s mountains, *Wiley Interdisciplinary Reviews: Water*, 3, 461–474, <https://doi.org/10.1002/wat2.1140>, 2016.
- 1005 Duan, S., Ullrich, P., Risser, M., and Rhoades, A.: Using temporal deep learning models to estimate daily snow water equivalent over the Rocky Mountains, *Water Resources Research*, 60, e2023WR035009, <https://doi.org/10.1029/2023WR035009>, 2024.
- Dunmire, D., Lievens, H., Boeykens, L., and De Lannoy, G. J.: A machine learning approach for estimating snow depth across the European Alps from Sentinel-1 imagery, *Remote Sensing of Environment*, 314, 114369, <https://doi.org/10.1016/j.rse.2024.114369>, 2024.
- Dunmire, D., Bechtold, M., Boeykens, L., and De Lannoy, G. J. M.: Advancing snow data assimilation with a dynamic observation uncer- 1010 tainty, *EGUsphere*, 2025, 1–26, <https://doi.org/10.5194/egusphere-2025-2306>, 2025.
- Essery, R.: A factorial snowpack model (FSM 1.0), *Geoscientific Model Development*, 8, 3867–3876, <https://doi.org/10.5194/gmd-8-3867-2015>, 2015.
- Essery, R., Morin, S., Lejeune, Y., and B Ménard, C.: A comparison of 1701 snow models using observations from an alpine site, *Advances in Water Resources*, 55, 131–148, <https://doi.org/https://doi.org/10.1016/j.advwatres.2012.07.013>, 2013.
- 1015 Fiddes, J. and Gruber, S.: TopoSCALE v. 1.0: Downscaling gridded climate data in complex terrain, *Geoscientific Model Development*, 7, 387–405, <https://doi.org/10.5194/gmd-7-387-2014>, 2014.

- Fontrodona-Bach, A., Schaeffli, B., Woods, R., Teuling, A. J., and Larsen, J. R.: NH-SWE: Northern Hemisphere Snow Water Equivalent dataset based on in situ snow depth time series, *Earth System Science Data*, 15, 2577–2599, <https://doi.org/10.5194/essd-15-2577-2023>, 2023.
- 1020 Gascoin, S., Luojus, K., Nagler, T., Lievens, H., Masiokas, M., Jonas, T., Zheng, Z., and De Rosnay, P.: Remote sensing of mountain snow from space: Status and recommendations, *Frontiers in Earth Science*, 12, 1381–1383, <https://doi.org/10.3389/feart.2024.1381323>, 2024.
- Global Climate Observing System: Snow — Essential Climate Variables, <https://gcos.wmo.int/site/global-climate-observing-system-gcos/essential-climate-variables/snow>, accessed: 2026-01-21.
- Goodarzi, M. R., Barzkar, A., Sabaghzadeh, M., Ghanbari, M., and Fathollahzadeh Attar, N.: Uncertainty Analysis of Machine Learning Methods To Estimate Snow Water Equivalent Using Meteorological and Remote Sensing Data, *Water Resources Management*, 39, 4471–4491, <https://doi.org/10.1007/s11269-025-04164-z>, 2025.
- 1025 Groisman, P. Y., Karl, T. R., Knight, R. W., and Stenchikov, G. L.: Changes of snow cover, temperature, and radiative heat balance over the Northern Hemisphere, *Journal of Climate*, 7, 1633–1656, [https://doi.org/10.1175/1520-0442\(1994\)007<1633:COCTA>2.0.CO;2](https://doi.org/10.1175/1520-0442(1994)007<1633:COCTA>2.0.CO;2), 1994.
- Hall, D. and Riggs, G.: MODIS/Aqua CGF Snow Cover Daily L3 Global 500m SIN Grid, Version 61, <https://doi.org/10.5067/MODIS/MYD10A1F.061>, 2020a.
- 1030 Hall, D. and Riggs, G.: MODIS/Terra CGF Snow Cover Daily L3 Global 500m SIN Grid, Version 61, <https://doi.org/10.5067/MODIS/MOD10A1F.061>, 2020b.
- Hill, D. F., Burakowski, E. A., Crumley, R. L., Keon, J., Hu, J. M., Arendt, A. A., Wikstrom Jones, K., and Wolken, G. J.: Converting snow depth to snow water equivalent using climatological variables, *The Cryosphere*, 13, 1767–1784, <https://doi.org/10.5194/tc-13-1767-2019>, 2019.
- 1035 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.
- Jans, J.-F., Beernaert, E., De Breuck, M., Brangers, I., Dunmire, D., De Lannoy, G., and Lievens, H.: Sensitivity of Sentinel-1 C-band SAR backscatter, polarimetry and interferometry to snow accumulation in the Alps, *Remote Sensing of Environment*, 316, 114477, <https://doi.org/10.1016/j.rse.2024.114477>, 2025.
- 1040 Jennings, K. S., Winchell, T. S., Livneh, B., and Molotch, N. P.: Spatial variation of the rain–snow temperature threshold across the Northern Hemisphere, *Nature Communications*, 9, 1148, 2018.
- Landschoot, S., Waegeman, W., Audenaert, K., Vandepitte, J., Haesaert, G., and De Baets, B.: Toward a reliable evaluation of forecasting systems for plant diseases: A case study using Fusarium head blight of wheat, *Plant Disease*, 96, 889–896, <https://doi.org/10.1094/PDIS-08-11-0665>, 2012.
- 1045 LARGERON, C., DUMONT, M., MORIN, S., BOONE, A., LAFAYSSÉ, M., METREF, S., COSME, E., JONAS, T., WINSTRAL, A., and MARGULIS, S. A.: Toward Snow Cover Estimation in Mountainous Areas Using Modern Data Assimilation Methods: A Review, *Frontiers in Earth Science*, Volume 8 - 2020, <https://doi.org/10.3389/feart.2020.00325>, 2020.
- Lievens, H., Demuzere, M., Marshall, H.-P., Reichle, R. H., Brucker, L., Brangers, I., de Rosnay, P., Dumont, M., Giroto, M., Immerzeel, W. W., Jonas, T., Kim, E. J., Koch, I., Marty, C., Saloranta, T., Schöber, J., and De Lannoy, G. J. M.: Snow depth variability in the Northern Hemisphere mountains observed from space, *Nature Communications*, 10, 1–12, <https://doi.org/10.1038/s41467-019-12566-y>, 2019.
- 1050 Lievens, H., Brangers, I., Marshall, H.-P., Jonas, T., Olefs, M., and De Lannoy, G.: Sentinel-1 snow depth retrieval at sub-kilometer resolution over the European Alps, *The Cryosphere*, 16, 159–177, <https://doi.org/10.5194/tc-16-159-2022>, 2022.

- Liston, G. E. and Elder, K.: A Distributed Snow-Evolution Modeling System (SnowModel), *Journal of Hydrometeorology*, 7, 1259 – 1276, 1055 <https://doi.org/10.1175/JHM548.1>, 2006a.
- Liston, G. E. and Elder, K.: A meteorological distribution system for high-resolution terrestrial modeling (MicroMet), *Journal of Hydrometeorology*, 7, 217–234, <https://doi.org/10.1175/JHM486.1>, 2006b.
- Luce, C., Staab, B., Kramer, M., Wenger, S., Isaak, D., and McConnell, C.: Sensitivity of summer stream temperatures to climate variability in the Pacific Northwest, *Water Resources Research*, 50, 3428–3443, <https://doi.org/10.1002/2013WR014329>, 2014.
- 1060 Lundberg, S. M. and Lee, S.-I.: A unified approach to interpreting model predictions, in: *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17*, p. 4768–4777, Curran Associates Inc., Red Hook, NY, USA, ISBN 9781510860964, 2017.
- Lundblad, E. R., Wright, D. J., Miller, J., Larkin, E. M., Rinehart, R., Naar, D. F., Donahue, B. T., Anderson, S. M., and Battista, T.: A benthic terrain classification scheme for American Samoa, *Marine Geodesy*, 29, 89–111, <https://doi.org/10.1080/01490410600738021>, 2006.
- 1065 Lute, A. C., Abatzoglou, J., and Link, T.: SnowClim v1.0: High-resolution snow model and data for the western United States, *Geoscientific Model Development*, 15, 5045–5071, <https://doi.org/10.5194/gmd-15-5045-2022>, 2022.
- López-Moreno, J. I., Revuelto, J., Fassnacht, S. R., Azorín-Molina, C., Vicente-Serrano, S. M., Morán-Tejeda, E., and Sexstone, G. A.: Snowpack variability across various spatio-temporal resolutions, *Hydrological Processes*, 29, 1213–1224, <https://doi.org/https://doi.org/10.1002/hyp.10245>, 2015.
- 1070 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.
- Marks, D. and Dozier, J.: Climate and energy exchange at the snow surface in the alpine region of the Sierra Nevada: 2. Snow cover energy balance, *Water Resources Research*, 28, 3043–3054, <https://doi.org/10.1029/92WR01483>, 1992.
- 1075 Marty, M., Bühler, Y., and Ginzler, C.: Snow depth mapping, <https://doi.org/http://dx.doi.org/10.16904/envidat.62>, 2019.
- Matiu, M., Crespi, A., Bertoldi, G., Carmagnola, C. M., Marty, C., Morin, S., Schöner, W., Cat Berro, D., Chiogna, G., De Gregorio, L., Kotlarski, S., Majone, B., Resch, G., Terzago, S., Valt, M., Beozzo, W., Cianfarra, P., Gouttevin, I., Marcolini, G., Notarnicola, C., Petitta, M., Scherrer, S. C., Strasser, U., Winkler, M., Zebisch, M., Cicogna, A., Cremonini, R., Debernardi, A., Faletto, M., Gaddo, M., Giovannini, L., Mercalli, L., Soubeyroux, J., Sušnik, A., Trenti, A., Urbani, S., and Weilguni, V.: Observed snow depth trends in the 1080 European Alps: 1971 to 2019, *The Cryosphere*, 15, 1343–1382, <https://doi.org/10.5194/tc-15-1343-2021>, 2021.
- Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E., and Houston, T. G.: An Overview of the Global Historical Climatology Network-Daily Database, *Journal of Atmospheric and Oceanic Technology*, 29, 897 – 910, <https://doi.org/10.1175/JTECH-D-11-00103.1>, 2012.
- Meyer, H., Reudenbach, C., Hengl, T., Katurji, M., and Nauss, T.: Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation, *Environmental Modelling & Software*, 101, 1–9, <https://doi.org/10.1016/j.envsoft.2017.12.001>, 2018.
- 1085 Miller, Z. S., Peitzsch, E. H., Sproles, E. A., Birkeland, K. W., and Palomaki, R. T.: Assessing the seasonal evolution of snow depth spatial variability and scaling in complex mountain terrain, *The Cryosphere*, 16, 4907–4930, <https://doi.org/10.5194/tc-16-4907-2022>, 2022.
- Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., and Xia, Y.: The community Noah land surface model with multiparameterization options (Noah-MP): 1090 1. Model description and evaluation with local-scale measurements, *Journal of Geophysical Research: Atmospheres*, 116, <https://doi.org/https://doi.org/10.1029/2010JD015139>, 2011.

- Notarnicola, C.: Overall negative trends for snow cover extent and duration in global mountain regions over 1982–2020, *Scientific Reports*, 12, 13 731, <https://doi.org/10.1038/s41598-022-16743-w>, 2022.
- 1095 Painter, T. H., Berisford, D. F., Boardman, J. W., Bormann, K. J., Deems, J. S., Gehrke, F., Hedrick, A., Joyce, M., Laidlaw, R., Marks, D., et al.: The Airborne Snow Observatory: Fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo, *Remote Sensing of Environment*, 184, 139–152, <https://doi.org/10.1016/j.rse.2016.06.018>, 2016.
- Parthum, B. and Christensen, P.: A market for snow: Modeling winter recreation patterns under current and future climate, *Journal of Environmental Economics and Management*, 113, 102 637, <https://doi.org/10.1016/j.jeem.2022.102637>, 2022.
- 1100 Parvande, S., Yeh, H.-W., Paulus, M. P., and McKinney, B. A.: Consensus features nested cross-validation, *Bioinformatics*, 36, 3093–3098, <https://doi.org/10.1093/bioinformatics/btaa046>, 2020.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V. and Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E.: Scikit-learn: Machine learning in Python, *Journal of Machine Learning Research*, 12, 2825–2830, 2011.
- 1105 Pomarol Moya, O., Nussbaum, M., Mehrkanoon, S., Kraaijenbrink, P. D. A., Gouttevin, I., Karssenber, D., and Immerzeel, W. W.: Improving forecasts of snow water equivalent with hybrid machine learning, *EGU sphere*, 2025, 1–23, <https://doi.org/10.5194/egusphere-2025-1845>, 2025.
- PyCaret: PyCaret: An open-source, low-code machine learning library, Version 1.0.0, Available at <https://www.pycaret.org/>, 2020.
- Qin, Y., Abatzoglou, J. T., Siebert, S., Huning, L. S., AghaKouchak, A., Mankin, J. S., Hong, C., Tong, D., Davis, S. J., and Mueller, N. D.: Agricultural risks from changing snowmelt, *Nature Climate Change*, 10, 459–465, <https://doi.org/10.1038/s41558-020-0746-8>, 2020.
- 1110 RGI 7.0 Consortium: Randolph glacier inventory - A dataset of global glacier outlines, Version 7.0, <https://doi.org/10.5067/f6jmovy5navz>, online access: <https://doi.org/10.5067/f6jmovy5navz>, 2023.
- Ryken, A., Bearup, L. A., Jefferson, J. L., Constantine, P., and Maxwell, R. M.: Sensitivity and model reduction of simulated snow processes: Contrasting observational and parameter uncertainty to improve prediction, *Advances in Water Resources*, 135, 103 473, <https://doi.org/10.1016/j.advwatres.2019.103473>, 2020.
- 1115 Schwanghart, W. and Scherler, D.: TopoToolbox 2–MATLAB-based software for topographic analysis and modeling in Earth surface sciences, *Earth Surface Dynamics*, 2, 1–7, <https://doi.org/10.5194/esurf-2-1-2014>, 2014.
- Seeherman, J. and Liu, Y.: Effects of extraordinary snowfall on traffic safety, *Accident Analysis & Prevention*, 81, 194–203, <https://doi.org/10.1016/j.aap.2015.04.029>, 2015.
- 1120 Small, D.: Flattening gamma: Radiometric terrain correction for SAR imagery, *IEEE Transactions on Geoscience and Remote Sensing*, 49, 3081–3093, <https://doi.org/10.1109/TGRS.2011.2120616>, 2011.
- 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.
- Sturm, M., Goldstein, M. A., and Parr, C.: Water and life from snow: A trillion dollar science question, *Water Resources Research*, 53, 3534–3544, <https://doi.org/10.1002/2017WR020840>, 2017.
- 1125 Tedesco, M. and Narvekar, P. S.: Assessment of the NASA AMSR-E SWE product, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 3, 141–159, <https://doi.org/10.1109/JSTARS.2010.2040462>, 2010.
- Terzago, S., Andreoli, V., Arduini, G., Balsamo, G., Campo, L., Cassardo, C., Cremonese, E., Dolia, D., Gabellani, S., von Hardenberg, J., Morra di Cella, U., Palazzi, E., Piazzi, G., Pogliotti, P., and Provenzale, A.: Sensitivity of snow models to the accuracy of meteorological

- forcings in mountain environments, *Hydrology and Earth System Sciences*, 24, 4061–4090, <https://doi.org/10.5194/hess-24-4061-2020>, 2020.
- 1130 Tsai, Y.-L. S., Dietz, A., Oppelt, N., and Kuenzer, C.: Wet and dry snow detection using Sentinel-1 SAR data for mountainous areas with a machine learning technique, *Remote Sensing*, 11, 895, <https://doi.org/10.3390/rs11080895>, 2019.
- Tsang, L., Durand, M., Derksen, C., Barros, A. P., Kang, D.-H., Lievens, H., Marshall, H.-P., Zhu, J., Johnson, J., King, J., Lemmetyinen, J., Sandells, M., Rutter, N., Siqueira, P., Nolin, A., Osmanoglu, B., Vuyovich, C., Kim, E., Taylor, D., Merkouriadi, I., Brucker, L., Navari, M., Dumont, M., Kelly, R., Kim, R. S., Liao, T.-H., Borah, F., and Xu, X.: Review article: Global monitoring of snow water equivalent using high-frequency radar remote sensing, *The Cryosphere*, 16, 3531–3573, <https://doi.org/10.5194/tc-16-3531-2022>, 2022.
- 1135 U.S. National Ice Center: IMS daily Northern Hemisphere snow and ice analysis at 4 km and 24 km resolution, National Snow and Ice Data Center, 2008.
- Varade, D., Singh, G., Dikshit, O., and Manickam, S.: Identification of snow using fully polarimetric SAR data based on entropy and anisotropy, *Water Resources Research*, 56, e2019WR025 449, <https://doi.org/10.1029/2019WR025449>, 2020.
- 1140 Wang, F. and Tian, D.: Multivariate bias correction and downscaling of climate models with trend-preserving deep learning, *Climate Dynamics*, 62, 9651–9672, <https://doi.org/10.1007/s00382-024-07406-9>, 2024.
- Warren, S. G.: Optical properties of snow, *Reviews of Geophysics*, 20, 67–89, <https://doi.org/10.1029/RG020i001p00067>, 1982.
- Weiss, A.: Topographic position and landforms analysis, in: Poster presentation, ESRI user conference, San Diego, CA, vol. 200, 2001.
- 1145 World Meteorological Organization: Rapid changes in cryosphere demand urgent, coordinated action, <https://wmo.int/news/media-centre/rapid-changes-cryosphere-demand-urgent-coordinated-action>, 2023.
- Xiong, C., Yang, J., Pan, J., Lei, Y., and Shi, J.: Mountain snow depth retrieval from optical and passive microwave remote sensing using machine learning, *IEEE Geoscience and Remote Sensing Letters*, 19, 1–5, <https://doi.org/10.1109/LGRS.2022.3226204>, 2022.