

Advancing Snow Depth Monitoring with Machine Learning and L-band InSAR Data: A Case Study Using SnowEx 2017 Data

1 Ibrahim Olalekan Alabi^{1*}, Hans-Peter Marshall², Jodi Mead³, Ernesto Trujilio²

2 ¹Department of Computer Science, Boise State University, Boise, ID, USA

3 ²Department of Geoscience, Boise State University, Boise, ID, USA

4 ³Department of Mathematics, Boise State University, Boise, ID, USA

5 * **Correspondence:**

6 Ibrahim Olalekan Alabi

7 ibrahimolalekana@u.boisestate.edu

8 Abstract

9 Current snow depth mapping from space faces challenges in spatial coverage, revisit frequency, and
10 cost. Airborne Lidar, although precise, incurs high costs and has limited geographical coverage,
11 thereby necessitating the exploration of alternative, cost-effective methodologies for snow depth
12 estimation. The forthcoming NASA-ISRO Synthetic Aperture Radar (NISAR) mission, with its 12-
13 day global revisit cycle and 1.25 GHz L-band frequency, introduces a promising avenue for cost-
14 effective, large-scale snow depth estimation using L-band Interferometric SAR (InSAR) capabilities.
15 This study demonstrates InSAR's potential for snow depth estimation via machine learning. Using 3
16 m resolution L-band InSAR products over Grand Mesa, Colorado, we compared the performance of
17 XGBoost, ExtraTrees, and Neural Networks across open, forested, and mixed terrains using Root Mean
18 Square Error (RMSE), Mean Bias Error (MBE), and R^2 metrics. XGBoost emerged as the superior
19 model, registering RMSE values of 9.85 cm, 9.88 cm, and 10.46 cm for open, mixed, and forested
20 regions, respectively. Validation against in-situ snow depth measurements resulted in an RMSE of
21 approximately 16 cm. Our findings indicate that L-band InSAR, with its ability to penetrate clouds and
22 cover extensive areas, coupled with machine learning, holds promise for enhancing snow depth
23 estimation. This approach, especially with the NISAR mission, may enable high resolution (~10 m)
24 snow depth mapping over extensive areas, provided suitable training data are available, offering a cost-
25 effective complement to existing snow monitoring practices.

26 Keywords: Snow Depth, InSAR, Machine Learning, NISAR, Remote Sensing

27 1 Introduction

28 Accurately measuring snow depth is critical for applications in hydrologic science, water resource
29 management, and climate modeling (1). Seasonal snowpacks act as natural reservoirs, storing winter
30 precipitation and gradually releasing meltwater in spring and summer as the temperature rises (1).
31 Tracking snow accumulation and melt is instrumental in understanding how snowpack storage affects
32 water resources. For instance, in California, the Sierra Nevada seasonal snowpack contributes
33 approximately 70% of additional water storage to supplement the existing artificial reservoir system
34 (2–4). However, monitoring snow depth, especially on a large scale, is fraught with challenges. The
35 diverse and rugged terrain of mountainous regions makes it particularly difficult (5).

Traditionally, snow depth monitoring has relied on in-situ measurements and airborne observations (6–9). However, these methods are often limited by geographical reach, temporal frequency, and financial resources. In-situ measurements are known for their accuracy and reliability, as they are often conducted manually using instruments such as snow probes or through automated snow telemetry stations (SNOTEL) (10). However, access to mountainous regions can be difficult during winter when snow monitoring is essential. Additionally, the resources required for frequent and widespread in-situ measurements, including financial and human capital, can be prohibitively high (9). Moreover, in-situ measurements have limited support (i.e.; representativeness) and cannot fully resolve the spatial heterogeneity of snowpack properties across a landscape (e.g.; snow depth, snow water equivalent (SWE), and snow density) (11,12).

Spaceborne remote sensing technologies such as Light Detection and Ranging (lidar) provide high-precision snow depth estimates, but cost limit comprehensive monitoring (9). Airborne lidar and structure-from-motion (i.e.; photogrammetry using aerial photography) deliver detailed snow depth maps but are geographically limited (13,14). Satellite passive microwave sensors estimate basin average SWE yet suffer from coarse resolution (~25 km) (15,16). Optical remote sensing is weather-dependent and unable to penetrate dense forest cover (17,18). Additionally, while initiatives like the Airborne Snow Observatory (ASO) are making strides towards the broader use of lidar for snow mapping, snow depth data is primarily collected in the western United States (US) (19), leaving a significant geographical gap in snow monitoring. Consequently, the development of new spaceborne capabilities to map snow depth continues to be an active area of research.

Amid the limitations of traditional snow monitoring methods and the challenges of lidar acquisition on a global scale, the pursuit of alternative remote sensing technologies is imperative. The forthcoming NASA-ISRO (NASA-Indian Space Research Organization) Synthetic Aperture Radar (NISAR) mission, scheduled for launch in October 2024, appears promising in this regard. NISAR, equipped with an L-band radar, is set to observe nearly all of Earth's terrestrial and ice surfaces at an approximate resolution of 10 m, with a revisit frequency of twice every 12 days (20). Operating within the 1-2 GHz frequency range, the L-band, with wavelengths between 15 to 30 cm, is conducive to the penetration of cloud cover and several meters of snowpack, facilitating all-weather, day-night snow monitoring. Repeat pass interferometric SAR (InSAR) acquisitions have demonstrated potential for quantifying snow properties, given SAR's all-weather imaging and vegetation penetration capabilities (21,22). A recent study by Hoppinen et al.(22) employing repeat pass InSAR for monitoring SWE over Idaho found strong correlations between retrieved SWE changes from SAR images and both in-situ and modeled results. Tarricone et al. (21) used repeat-pass L-band InSAR to estimate SWE changes in an environment with both snow accumulation and ablation, showcasing the capability of L-band InSAR for tracking SWE changes. Studies have also shown promise in retrieving snow depth information from InSAR coherence, phase, and incidence angle data (23,24). Li et al. (24) used repeat-pass InSAR measurements for estimating snow depth and SWE in the Northern Piedmont Region of the Tianshan Mountains and found the snow depth estimation to be consistent with field measurements. With the promising capabilities of InSAR, the NISAR mission may be able to provide high-resolution (~10 m) snow depth estimates over large areas if appropriate snow depth observations are available for training machine learning models.

Our goal in this work is to use repeat-pass L-band InSAR products to estimate total snow depth (e.g.; bulk snowpack depth to the ground) using Machine Learning algorithms. While InSAR is inherently more related to changes in snow depth rather than total snow depth, snow accumulation patterns tend to exhibit consistency. As such, snow depth before the onset of melt displays similar patterns to snow depth changes (25). Snow distribution patterns have been shown to exhibit intra- and inter-seasonal

consistency despite differences in weather patterns and seasonal snowfall amounts (26–29). While the actual snow depths may change, the locations of deeper and shallower snow areas generally tend to be consistent.

Since the 1990s, machine learning algorithms have gained prominence in environmental remote sensing (30–32) and, over time, have proliferated across various application areas such as snow depth retrieval (33–36), snow density (37,38), and SWE (39,40) predictions. Hu et al. (41) conducted a study using machine learning algorithms to fuse gridded snow depth datasets with inputs including geolocation, topography, and in-situ observations. The Random Forest algorithm proved to be the most proficient of the three learning machines tested. Liang et al. (42) applied the Support Vector Machine (SVM) to estimate snow depth in northern Xinjiang using data from visible and infrared surface reflectance, brightness temperature, and auxiliary information. The SVM method outperformed the Che algorithm in China (43) and the Artificial Neural Networks (ANN) utilized in Finland (34).

These works and others found in the literature have highlighted the potential of machine learning in producing improved snow depth predictions. However, a common limitation across many of these studies is the constrained spatial coverage, which can be attributed to the scarcity of global snow depth data and the prohibitive costs associated with acquiring lidar data globally. This study aims to develop a snow depth prediction system using L-band InSAR products and ML algorithms. Specifically, we will compare the performance of three machine learning algorithms: eXtreme Gradient Boosting (XGBoost) (44), Extremely Randomized Tress (ExtraTrees) (45), and Artificial Neural Networks (ANN) (46). We will also investigate the impact of vegetation on snow depth prediction accuracy.

The objective of this work is divided into three broad aspects:

1. to test the effectiveness of L-band InSAR products in estimating total snow depth using ML algorithms,
2. to analyze the effect of vegetation type on the performance of the ML models and
3. to understand the relative importance of each input feature in snow depth estimation.

The study area is stratified into open areas, vegetated areas, and a combination of both to evaluate the impact of vegetation on model performance. Our research is motivated by the need for a global snow depth prediction system that is accurate, cost-effective, and scalable. L-band InSAR is a promising technology for snow depth mapping because it can penetrate through clouds and forest cover, and it can be deployed over large areas. Machine learning algorithms can be used to extract snow depth information from L-band InSAR data in a robust and efficient manner.

We believe that our research has the potential to complement existing snow monitoring practices by facilitating cost-effective, high-resolution, and extensive snow depth estimation. Our findings could lead to the development of a global snow depth prediction system that provides valuable information for water resource management, flood forecasting, and avalanche hazard assessment, provided that accurate and representative training data is available.

2 Data and Methods

2.1 Study Area

This study was conducted in Grand Mesa, a flat-topped mesa in western Colorado, USA (39.1°N, 107.9°W). Grand Mesa is one of the largest flat-topped mesas in the world (47,48) with elevations from 2,000 to 3,400 m above sea level (49). The climate is continental (49), with the coldest, warmest,

wettest, and windiest months being January, July, September, and June, respectively (Time and Date website, 2024: <https://www.timeanddate.com/weather/@5423575/climate>).

2.2 Data

The data for our study was sourced from NASA's 2017 SnowEx campaign (50)(50). The 2017 campaign aims to evaluate the sensitivity of snow remote sensing techniques through a gradient of forest densities (NSIDC: <https://nsidc.org/data/snowex>).(50) This campaign was primarily conducted in two locations in Colorado, USA: the Grand Mesa and the Senator Beck Basins. Our focus is on Grand Mesa (Figure 1). SnowEx datasets are publicly available through the National Snow and Ice Data Center (NSIDC: <https://nsidc.org/data/snowex>).(51)

The 2017 SnowEx campaign spanned from September 2016 to July 2017, with the Intense Observation Period (IOP) taking place between February 6 and February 25, 2017 (52). During this period, a multitude of measurements were collected, including data from cloud-absorption radar, ground-penetrating radar, synthetic aperture radar, lidar, airborne video, Global Navigation Satellite Systems (GNSS) measurements, and snow pit measurements (50,52). In this study, we focus on the L-band InSAR products and high-resolution airborne lidar data collected during the campaign (snow depths and bare ground elevations). Our study site within Grand Mesa spans approximately 70 km², and all data used in this study has a resolution of 3 m. The details of each data type's collection and use are outlined in the following sections.

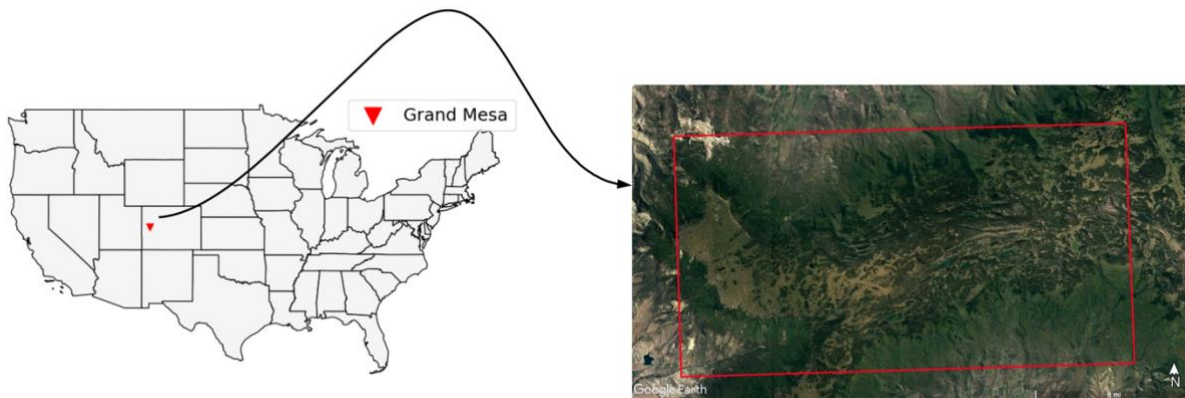


Figure 1. Study Site: Grand Mesa, Colorado. The large red rectangle outlines a 9.25 x 32 km target area that was imaged by all airborne sensors. The natural gradient of Snow Water Equivalent (SWE) increases from west to east, and the forest cover also naturally varies across the region (50).

2.2.1 Lidar Snow Depth Measurements

Lidar, an acronym for Light Detection and Ranging (53,54), is a remote sensing method that uses the travel time of a laser pulse to measure distances (55,56). The lidar data used in this study was collected by the NASA Airborne Snow Observatory (ASO) as part of the 2017 SnowEx campaign on two dates: September 20, 2016, and February 8, 2017. The September 20 flight was a "no-snow" flight, which means that the lidar data was collected before any snow had fallen on the ground. The February 8 flight was a "snow-on" flight, which means that the lidar data was collected when a snowpack had developed on the ground. The initial gridded ASO lidar product has a spatial resolution of 3 m and a vertical accuracy characterized by a root mean square difference of 8 cm when compared to near-coincident median values from 52 snow-probe transects (57). Snow depths were then calculated by differencing

the two DEMs (snow-on minus snow-off). In addition to snow depth, bare ground elevation and vegetation height to the top of the canopy at matching 3-m resolution were produced from the lidar data (Figure 2).

For this work, we divided the data into three categories based on vegetation:

- Open areas: areas with no vegetation or vegetation heights < 0.5 meters.
- Vegetated areas: areas with vegetation height ≥ 0.5 meters
- A combination of both open and forested areas.

The resulting training dataset contains 3.174×10^6 , 2.166×10^6 , and 5.341×10^6 3-m resolution snow depths for open areas, forested areas, and a combination of both, respectively. Table 1 shows the snow depth statistics for each category in the training data. The mean snow depth is approximately 1.06 m in open areas, which is slightly lower than that of the vegetated areas at 1.10 m. The range of snow depth across all areas is from 0.000168 m to a maximum of 2.99 m.

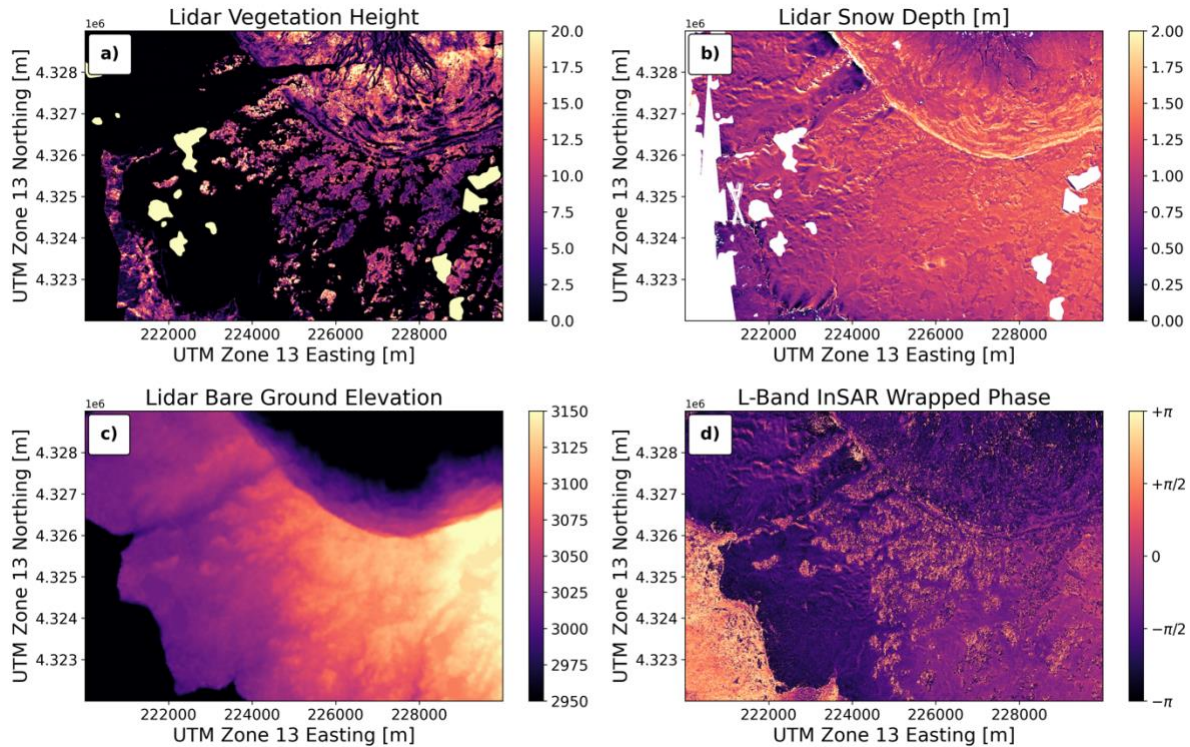


Figure 2. Lidar-derived products (a-c) from the study site. **a:** Vegetation height with a color range from 0 to 20 m, **b:** Snow depth with a color range from 0 to 2 m. **c:** Bare ground elevations with a color range spanning approximately 2950 to 3150 m. **d:** L-Band InSAR Wrapped Phase with a color scale ranging from $-\pi$ to $+\pi$, capturing ground displacements.

Table 1. Summary Statistics of L-band InSAR products, lidar snow depths, bare ground elevation, and vegetation height (all at 3 m resolution) in the training set. The table provides the count, mean, standard deviation (Std), minimum (Min), 25th percentile (25%), median (50%), 75th percentile (75%), and maximum (Max) values for each product within open, vegetated, and the full dataset subsets.

Subset	Product	Mean	Std	Min	25%	50%	75%	Max
Open Areas (n=3174384)	Amplitude	0.12	0.07	0.02	0.09	0.11	0.13	3.55
	Unwrapped Phase	-7.92	0.82	-15.22	-8.44	-8.01	-7.33	-0.06
	Coherence	0.64	0.13	0.003	0.56	0.66	0.73	0.98
	Incidence Angle	0.92	0.13	0.28	0.84	0.94	1	1.95
	Bare Earth DEM	3044.3	57.61	2492.03	3019.76	3035.7	3076.51	3157.45
	Wrapped Phase	-1.38	0.96	-3.14	-2.06	-1.54	-0.87	3.14
	Elevation (m)	3048.59	57.61	2501.13	3023.22	3039.78	3081.68	3168
	Snow Depth (m)	1.06	0.30	0.00017	0.87	1.07	1.26	3.0
Vegetated Areas (n=2166210)	Amplitude	0.39	0.18	0.01	0.28	0.35	0.46	6.11
	Unwrapped Phase	-8.73	0.81	-15.26	-9.17	-8.67	-8.23	-0.47
	Coherence	0.48	0.15	0.006	0.36	0.47	0.59	0.97
	Incidence Angle	0.79	0.22	0.14	0.65	0.82	0.94	1.95
	Bare Earth DEM	2957.81	154.61	2490.96	2869.95	3004.13	3076.61	3157.38
	Vegetation Height (m)	8.43	5.26	0.5	4.09	8.12	12.06	34.62
	Wrapped Phase	-1.03	1.67	-3.14	-2.26	-1.69	-0.09	3.14
	Elevation (m)	2969.7	153.62	2500.16	2883.21	3016.08	3088.03	3168
Mixed Dataset (n=5340595)	Snow Depth (m)	1.11	0.28	0.00041	0.95	1.11	1.26	3.0
	Amplitude	0.23	0.18	0.01	0.10	0.15	0.33	6.11
	Unwrapped Phase	-8.25	0.91	-15.26	-8.76	-8.28	-7.68	-0.06
	Coherence	0.57	0.16	0.003	0.46	0.59	0.70	0.98
	Incidence Angle	0.87	0.18	0.14	0.77	0.90	0.99	1.95
	Bare Earth DEM	3009.17	116.11	2490.91	3007.21	3031.91	3076.55	3157.44
	Vegetation Height (m)	3.47	5.30	0	0	0.18	6.35	34.62
	Wrapped Phase	-1.24	1.3	-3.14	-2.14	-1.59	-0.75	3.14
	Elevation	3016.55	114.27	2500.20	3012.51	3036.44	3084.15	3168
	Snow Depth (m)	1.08	0.29	0.00017	0.90	1.09	1.26	3.0

2.2.2 L-band InSAR Products

Interferometric Synthetic Aperture Radar (InSAR) is a radar technique that calculates the distance between the radar antenna and the ground surface by comparing the phase difference of two or more

radar images (58). Repeat-pass InSAR uses two or more SAR images of the same scene taken at different times to generate an interferogram, which measures the phase shift between the image acquisitions. InSAR can be used to measure a variety of surface changes, including deformation, subsidence, and ice sheet movement. More recently, repeat-pass InSAR has been used to measure SWE changes (21).

The L-band SAR frequency range is 1 to 2 GHz, which corresponds to a wavelength range of 30 to 15 cm (NASA: <https://www.earthdata.nasa.gov/learn/backgrounders/what-is-sar>). L-band SAR signals can penetrate clouds, snow, (21) and vegetation. As the wave moves through the snow, it slows down, causing a phase shift. This phase shift can be used to estimate changes in snow depth and SWE. L-band SAR signals are also less sensitive to atmospheric moisture than other SAR bands, which makes the InSAR data more reliable in all weather conditions. As part of the 2017 SnowEx campaign, the NASA Jet Propulsion Laboratory (JPL) flew the L-Band UAVSAR sensor over Grand Mesa in February and March. A total of five flights were conducted on February 6, 22, 25, and March 8 and 31. For this work, we used the pair of flights closest to the lidar acquisition. Since the lidar was acquired on February 8, we used the HH polarization for February 6 and 22 as the InSAR features for our ML models. UAVSAR products come at a native resolution of 2 m. However, we resampled the UAVSAR features to 3 m to match the lidar resolution.

The following L-band InSAR products, all of which are resampled to 3-m resolution, were used in this study:

- **Coherence (CO):** Coherence, also known as interferometric correlation, measures the accuracy of determining the interferometric phase. It ranges from 0 to 1, with higher coherence indicating greater confidence in the phase changes.
- **Incidence Angle (IA):** The incidence angle is the angle between the SAR beam and the snow surface. The incidence angle affects the backscatter signal from the snow surface, and it must be considered when interpreting InSAR data.
- **Amplitude (AM):** Amplitude is one of the two fundamental characteristics of radar signals. It is the magnitude/energy of the backscattered signal, and it records information on the surface roughness. For example, surfaces that are rough at the scale of the radar wavelength or have a higher dielectric constant (such as wet surfaces) will generally backscatter more of the radar signal and thus have a higher amplitude. Conversely, smoother surfaces or those with a lower dielectric constant will backscatter less of the signal, resulting in a lower amplitude.
- **Wrapped Phase (WP):** Phase is the second fundamental property of radar signals. The phase image records information on sensor-to-target distance. Each pixel in a SAR image has a phase value that corresponds to the distance between the satellite sensor and the ground target (snow surface), modulated by the wavelength of the radar signal. When two or more SAR images are acquired at different times in repeat-pass interferometry, the phase difference between corresponding pixels in the two images can be calculated. This phase difference, referred to as the interferometric phase or wrapped phase, contains valuable information about the target surface, such as elevation change or displacement. The term "wrapped" refers to the fact that the phase values are cyclic, i.e., a phase value of 2π is equivalent to a phase value of 0. Thus, the phase values "wrap" around at intervals of 2π .
- **Unwrapped Phase (UW):** The unwrapped phase is the phase difference between the two SAR images after it has been processed to remove the cycle "wrapping". It provides a continuous measure of the distance between the radar antenna and the snow surface, which is crucial for measuring changes in the snow depth.

InSAR measures changes in surface distance or surface deformation. This technology uses radar signals to precisely detect alterations in the Earth's surface, providing valuable insights into various geophysical phenomena. Therefore, to use InSAR for estimating total snow depth, we assume that the total snow depth patterns before melts starts within a season are consistent (29) with the patterns of snow depth change detected by the InSAR instrument. This assumption is generally valid, but there are instances and locations where this may not be the case depending on weather patterns, landscape and vegetation changes, and other processes affecting snow distribution. L-band InSAR data will be available from the NISAR mission at a 12-day temporal resolution, which has the potential to complement existing snow depth measurement practices. The InSAR products used in this work are shown in Figure 3, and their summary statistics are included in Table 1.

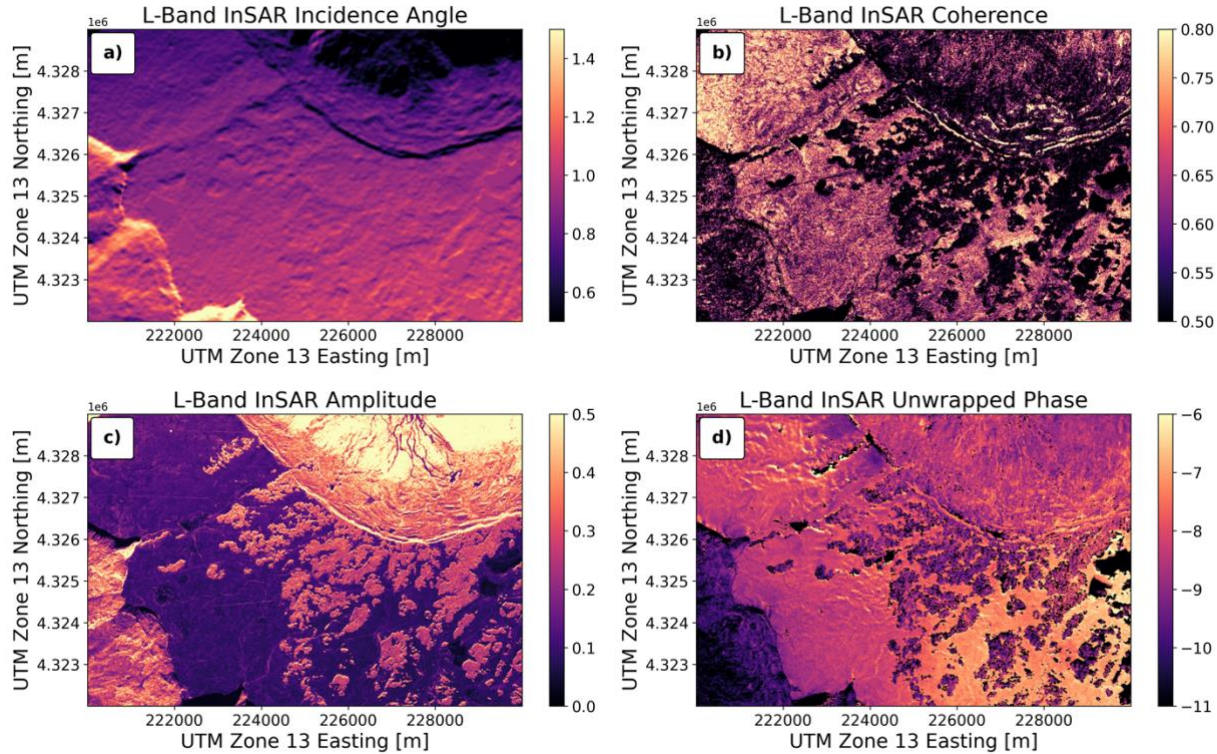


Figure 3. L-band InSAR products: (a) Incidence angles with values from 0.5 to 1.5, (b) Coherence metrics with values from 0.50 to 0.80, (c) Amplitudes with signal intensities between 0.0 and 0.5, and (d) Unwrapped Phase with phase shift values from ~ -11 to -6 .

2.3 Methods

Our methodology for estimating snow depth from InSAR products using machine learning models comprised several key steps, which are outlined below.

2.3.1 Data Preprocessing

Data preprocessing was the foundational step in our methodology, preparing the dataset for compatibility with our Machine Learning algorithms. The raw remote sensing data spanned numerous raster layers at a 3-m resolution. Each raster layer represents approximately a 7 km x 10 km area on the ground, with each pixel representing an averaged value over a 3 m x 3 m area. We first restructured the raster into pixel-wise tabular data frames to enable vector-based preprocessing. After concatenating all variables, we inspected for noise and outliers that could potentially affect the models' learning.

Noise in the data was identified through pixel values coded as -9999, which are standard no-data flags, and these values were dropped from the analyses. Outliers were defined based on the physical plausibility of measurements, with snow depth values greater than 3 m classified as outliers. This threshold was chosen after inspecting the in-situ snow depth distribution for the day (February 8), which showed a maximum snow depth of 3 m. Extreme values (e.g., snow depths of 13 m) were unrealistic and likely due to errors such as the lidar laser hitting the top of vegetation during the snow-on flight.

No missing value imputation was done as we have an abundance of data points (~ 6.7 million observations after data cleaning). Imputation is a process where missing data points are filled in using various statistical methods, which can sometimes introduce uncertainties. By avoiding imputation, we maintained dataset integrity. These data cleaning steps resulted in a dataset of 6,675,745 usable observations.

After data cleaning, we split our data into three disjoint subsets:

- Training Set: 80% of the data was used for training the models.
- Tuning Set: 10% of the data was used for hyperparameter tuning to optimize model performance and to prevent overfitting during model training. Hyperparameter tuning involves adjusting the parameters of the learning algorithm itself to find the best configuration. This process helps improve predictive accuracy and robustness. Overfitting occurs when a model learns the training data too well, capturing noise and details that do not generalize to new data. By using a tuning set, we can ensure the model generalizes well and performs effectively on unseen data. Note that using a tuning set for hyperparameter optimization is an alternative to cross-validation for large datasets.
- Testing Set: The remaining 10% of the data was used to test the models' performance.

Following the partitioning, all input features were converted into a common dimensionless scale with a mean of zero and a standard deviation of one using the z-score standardization approach following Eq. (4). This will prevent potential domineering effects from variables with larger raw value ranges during model training. We have chosen the z-score standardization approach because it ensures outliers are handled more properly (59). The normalized version of every observation $X_{i,j}$ in feature X_j is obtained by:

$$X_{i,j}^z = \frac{X_{i,j} - \mu_j}{\sigma_j} \quad (1)$$

Where $X_{i,j}^z$ is the standardized value, μ_j and σ_j are the mean and standard deviation of variable X_j from the training set.

For each data stratum (open, vegetated, and mixed), we fitted three models:

- Model 1: The first model uses only bare ground elevation as input, and this model serves as the baseline model in this work. Mathematically, we write:

$$SD = f(BGE) \quad (2)$$

- Model 2: This model uses InSAR products and bare ground elevation as input features. The model is represented mathematically as follows:

$$SD = f(CO, IA, BGE, WP, AM, UP) \quad (3)$$

- Model 3: This model combines the InSAR products, bare ground elevation, and vegetation height as features. This configuration is used to evaluate the effect of vegetation height on the model performance.

$$SD = f(CO, IA, BGE, WP, AM, UP, VH) \quad (4)$$

Where f is the various learning algorithms compared, VH is the vegetation height, BGE is the bare ground elevation, and SD is the snow depth. Note that the open areas only have models 1 and 2 because, in the open areas, vegetation height is assumed to be 0 when buried under the snow cover.

2.3.2 Model Selection and Training

Once the data were appropriately prepared, we proceeded to the next phase: model selection and training. We evaluated and compared three machine learning algorithms: Extremely Randomized Trees (Extra Trees), eXtreme Gradient Boosting (XGBoost), and artificial neural networks (ANN). These models were chosen because they have been shown to perform well on a variety of remote sensing data (60,61). For hyperparameter tuning, we used the Optuna framework (62). Optuna is an open-source hyperparameter optimization framework in Python that is designed to optimize the hyperparameters for machine learning models. Unlike traditional grid search, Optuna utilizes a Tree-structured Parzen Estimator (TPE) algorithm (63), a Bayesian optimization algorithm, which tends to find optimal hyperparameters faster and with fewer function evaluations compared to grid search. This efficient search approach proved to be especially advantageous given the large dataset involved in our study. The details of the hyperparameters tuned for each model can be found in Table A1 in the appendix.

2.3.2.1 Extremely Randomized Trees

Extra Trees is an ensemble algorithm that averages predictions across a pre-defined number of randomized decision trees to improve accuracy and control overfitting. The "extra" in Extra Trees stands for extremely randomized, indicating that at each split in the learning process, the features and cut points are chosen in a random manner, hence reducing the variance of the model. Due to this extreme randomization, Extra Trees are faster than Random Forest and hence suitable for large datasets. Mathematically, the prediction $\hat{f}_{EXT}(\mathbf{x})$ from an extra tree regressor can be expressed as:

$$\hat{f}_{EXT}(\mathbf{x}) = \frac{1}{T} \sum_{t=1}^T \hat{f}^t(\mathbf{x}) \quad (1)$$

Where T is the total number of trees in the ensemble and $\hat{f}^t(\mathbf{x})$ is the prediction of the t -th tree for the input vector \mathbf{x} .

In our analysis, the optimal hyperparameters for the Extra Trees model were identified using Optuna. The optimal number of trees in the forest was found to be 150. The maximum depth of the trees was set to None, indicating that the nodes are expanded until they contain fewer than the minimum samples required to split, allowing the trees to grow to their full depth. Finally, the mean squared error (MSE) was used as the measure of split quality at each node. These hyperparameter settings were found to

326 provide the best performance on the tuning set. The details of these hyperparameters can be found in
327 Table A1 in the appendix.

328 **2.3.2.2 eXtreme Gradient Boosting**

329 XGBoost is another ensemble algorithm, but unlike Extra Trees, it builds a sequence of decision trees
330 instead of a forest of decision trees. XGBoost operates by sequentially constructing weak learners
331 (decision trees), with each tree aiming to correct the errors made by the previous one. This process of
332 sequential error correction is known as Boosting. At each iteration, a weak learner is trained to
333 approximate the gradient of the loss function (the residual errors). Boosting is then achieved by
334 iteratively updating the residual errors when a new learner is added to the ensemble. This methodology
335 of leveraging the gradient of the loss function to guide the boosting process is known as Gradient
336 Boosting. XGBoost uses a variant of the Gradient Boosting algorithm called Newton boosting, which
337 attaches weights to the residuals through the Hessian (second-order derivative of the loss function).
338 With this, observations with larger errors have more weight. XGBoost takes Newton boosting to the
339 extreme by regularizing the loss functions and introducing efficient tree learning with parallelizable
340 implementation. This “extreme” attribute of XGBoost makes it suitable for large datasets. Moreover,
341 XGBoost can also benefit from GPU acceleration, making it suitable for large datasets.

342 In our analysis, the optimal hyperparameters for the XGBoost model were identified using Optuna, as
343 detailed in Table A1 in the appendix. The objective was set to minimize the MSE between the predicted
344 and true snow depth values. We used a learning rate of 0.05 to control the step size at each iteration
345 while moving toward a minimum of the loss function. The depth of the trees was set to grow
346 unrestricted to allow the model to learn complex relationships in the data. A total of 1000 trees were
347 grown using a histogram-based training method to accelerate the training process.

348 **2.3.2.3 Artificial Neural Network**

349 Artificial Neural Network (ANN) is an ensemble of linked artificial neurons organized into input,
350 hidden, and output layers. Each neuron receives inputs, performs mathematical operations on these
351 inputs, and passes the output to the next layer. Every connection between nodes has an associated
352 weight; the optimal weights are learned during training. The input layer holds the features, with one
353 node per input feature, and the output layer holds the network’s prediction. The number of neurons in
354 the hidden layers is determined through hyperparameter tuning. In this study, we employed a Feed-
355 forward Neural Network (FNN), a type of ANN where connections between nodes do not form a cycle.
356 Specifically, we used the FNN because it is the gold standard for tabular datasets.

357 We designed a five-layer FNN using the PyTorch framework for our analysis. The architecture
358 comprises one input layer, three densely connected hidden layers with 2048, 1500, and 1000 nodes,
359 respectively, and one output layer. Rectified Linear Unit (ReLU) activation functions were used in the
360 hidden layers to introduce non-linearity, while a linear activation function was employed in the output
361 layer for snow depth estimation. The model was trained using the Adam (64) optimization algorithm
362 with a learning rate of 0.0001 to minimize the Mean Squared Error Loss (MSELoss) between the
363 predicted and actual snow depths. The training was conducted over 15 epochs with a batch size of 128.
364 The hyperparameters for the FNN model were optimized using Optuna, with the details of the tuned
365 hyperparameters presented in Table A1 in the appendix.

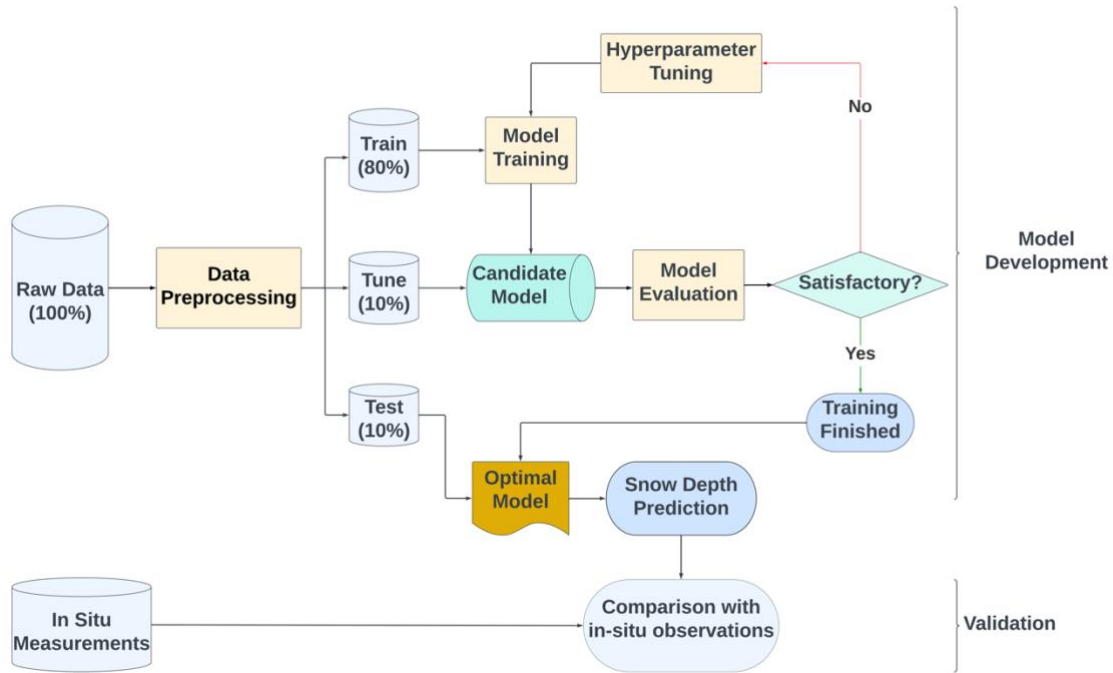


Figure 4. A flowchart illustrating the step-by-step methodology of our model training.

Model performance was evaluated using root mean squared error (RMSE), mean bias error (MBE), and coefficient of determination (R^2). RMSE measures the average magnitude of the errors between the predicted and observed values, providing insight into the model's overall accuracy. MBE assesses the average bias in the predictions, indicating whether the model systematically overestimates or underestimates the true values. R^2 measures how well the predicted values match the true values using a 1:1 line, with higher values indicating a better fit. RMSE has a lower bound of zero, with smaller values indicating better performance. MBE is unbounded and can be positive or negative, with values closer to zero in absolute value being preferable. Note that R^2 , in this case, is not the square of the correlation coefficient; hence, it can be negative for very bad models.

2.4 Feature Importance

To quantify the contribution of individual features to snow depth prediction, we employed two complementary methods: the gain metric from the XGBoost model and SHapley Additive exPlanations (SHAP) (65). The *gain*, derived from the XGBoost framework, measures the average contribution of each feature to reducing the mean squared error loss across all trees within the model. This method offers an initial insight into the relative importance of features based on their utility in constructing the predictive model. However, gain-based importance can be misleading for high cardinality (many unique values) features and may not fully capture the nuanced interactions between features. To address this, we also utilized SHAP values, which provide a more comprehensive and stable measure of feature importance by considering the contribution of each feature to every possible combination of features in the dataset (66). SHAP originates from concepts in cooperative game theory, and it computes the importance of a feature by distributing the predictive contribution among features in a model, akin to dividing payoffs among collaborating players. Although SHAP analysis is computationally expensive, particularly for large datasets, its ability to offer clear and consistent interpretations of the features' impact on the model's output makes it a valuable tool for in-depth feature importance analysis.

394

395 **3 Results**

396 **3.1 Snow Depth Estimation**

397 Comparing the performance of Models 1, 2, and 3 allowed us to evaluate the incremental value of
398 remote sensing predictors for improving snow depth estimation beyond using bare ground elevation
399 alone. Across most configurations and areas, XGBoost consistently outperformed other models.
400 Therefore, our discussion will focus on the XGBoost results.

401 The model comparison revealed distinct performance across different model configurations. The
402 baseline model (Model 1) provided a foundational understanding but demonstrated limited predictive
403 power with R^2 values ranging from 0.21 to 0.33 across the open, vegetated, and mixed areas. However,
404 incorporating InSAR-derived features (Model 2) markedly increased the precision of snow depth
405 estimates.

406 Table 2: Comparative Performance of FNN, Extra Trees, and XGBoost across Vegetated, Open, and
407 Mixed Areas on the Held-out Test Set. Values in bold fonts are the best for each metric.

Model	Data	Configuration	RMSE (m)	MBE (m)	R^2
FNN	Open Areas	Model 1	0.2493	0.0003	0.321
		Model 2	0.1646	-0.0035	0.704
	Mixed Areas	Model 1	0.2613	0.001	0.2047
		Model 2	0.1738	0.0095	0.6481
		Model 3	0.1617	-0.004	0.6952
	Vegetated Areas	Model 1	0.2425	-0.0025	0.2368
		Model 2	0.1802	-0.0094	0.5787
		Model 3	0.1611	0.0071	0.6631
Extra Trees	Open Areas	Model 1	0.2365	0.0001	0.3891
		Model 2	0.104	0.0001	0.8818
	Mixed Areas	Model 1	0.2483	0.0003	0.2815
		Model 2	0.1085	0.0002	0.8629
		Model 3	0.1052	0.0003	0.871
	Vegetated Areas	Model 1	0.2201	-0.0004	0.3712
		Model 2	0.1039	0.00004	0.8441
		Model 3	0.1118	0.0008	0.8379
XGBoost	Open Areas	Model 1	0.2483	0.0007	0.3265
		Model 2	0.0985	-0.0001	0.894
	Mixed Areas	Model 1	0.2606	0.0004	0.2087
		Model 2	0.104	-0.000009	0.8739
		Model 3	0.0988	0.0001	0.8863
	Vegetated Areas	Model 1	0.2421	0.0001	0.2393
		Model 2	0.1098	-0.0002	0.8436

408

409

410 Model 2 increased the R^2 value from 0.33 to 0.89 in the open areas, from 0.24 to 0.84 in the vegetated
 411 areas, and from 0.21 to 0.87 in the mixed areas. This improvement underscores the potential of InSAR
 412 data in capturing key variables influencing snow depth. The highest R^2 was observed in open areas.
 413 The performance in vegetated areas also improved, which indicates that InSAR products contribute
 414 valuable information even in vegetated landscapes. For the mixed (open + forested) areas, Model 2
 415 attained an R^2 of 0.87, which is also a substantial increase from the baseline performance.

416 Introducing vegetation height as an additional predictor alongside InSAR products in Model 3 yielded
 417 minimal performance gains in model accuracy across vegetated and mixed areas (assuming open areas
 418 have no vegetation). In vegetated areas, Model 3 increased the R^2 from 0.84 to 0.86, while in mixed
 419 areas, the R^2 improved from 0.87 to 0.89. Despite these incremental improvements, the overall best
 420 performance was observed in open areas (Model 2: RMSE=9.85 cm, R^2 =0.894), followed closely by
 421 mixed areas (Model 3: RMSE=9.88 cm, R^2 =0.886). Vegetated areas exhibited the least performance
 422 (Model 3: RMSE=10.46 cm, R^2 =0.858), although still maintaining a high level of accuracy.

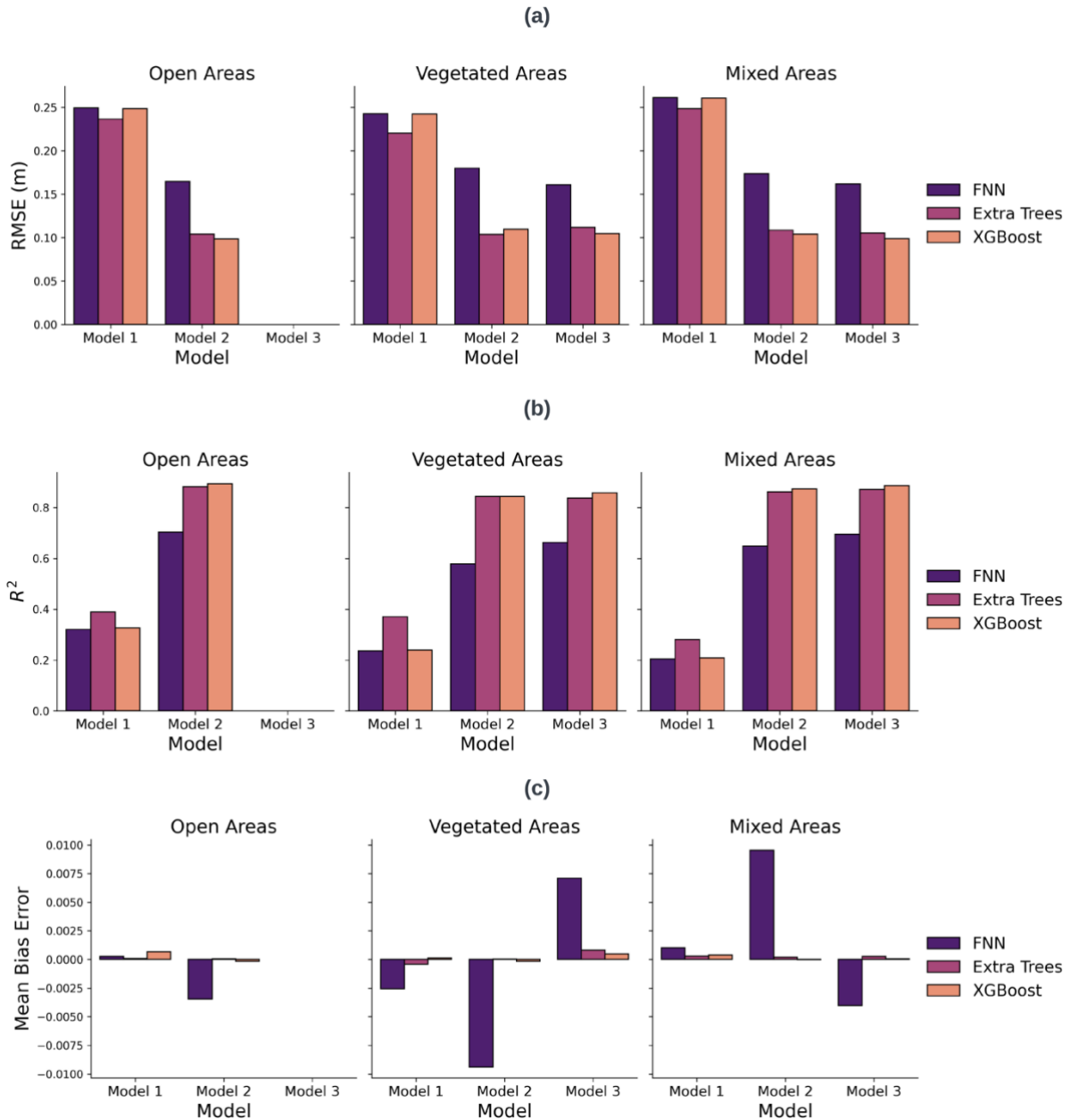


Figure 5: Comparative performance metrics of snow depth estimation Models on the testing set. Panel (a) displays the RMSE across all models and data strata, illustrating the accuracy of snow depth predictions. Panel (b) presents the R^2 values, and Panel (c) presents the MBE.

Based on the performance comparison, Model 3 for the mixed areas appears to be the optimal choice. This model demonstrated comparable results (Table 2) to those obtained in open areas (Model 2), effectively eliminating the need for separate models for open and vegetated regions. Adopting a single model approach offers several advantages, including saving time and resources while allowing more focus on refining and optimizing this unified model. Moving forward, the residual analysis will be reported based on Model 3 for the mixed areas, as it provides a comprehensive and efficient solution for our application.

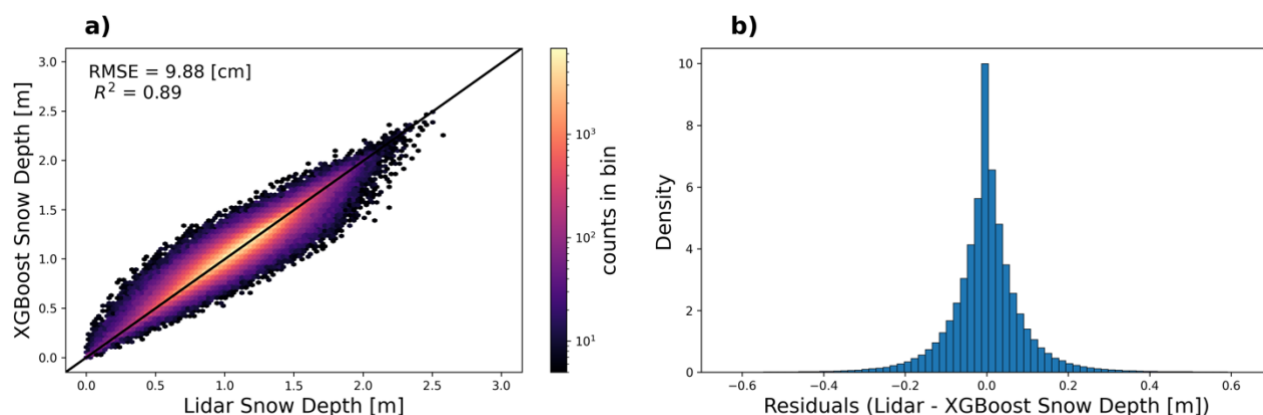


Figure 6. XGBoost Residual Analysis (Model 3 on mixed areas). (a) Hexbin density plot comparing lidar measured snow depths with predictions from the XGBoost model. (b) Histogram of the residuals (lidar minus XGBoost predicted snow depths)

The XGBoost modeled snow depths showed strong correspondence with lidar snow depths at the 3-m scale ($RMSE=9.88$ cm, $R^2=0.89$; Figure 6a), illustrating the model's ability to reproduce the observed snow depth variability. Similarly, the residuals (i.e., lidar minus XGBoost predicted snow depth; Figure 6b) are concentrated around zero with a mean of -7.9×10^{-3} cm and a standard deviation of 9.9 cm, indicating minimal errors and biases.

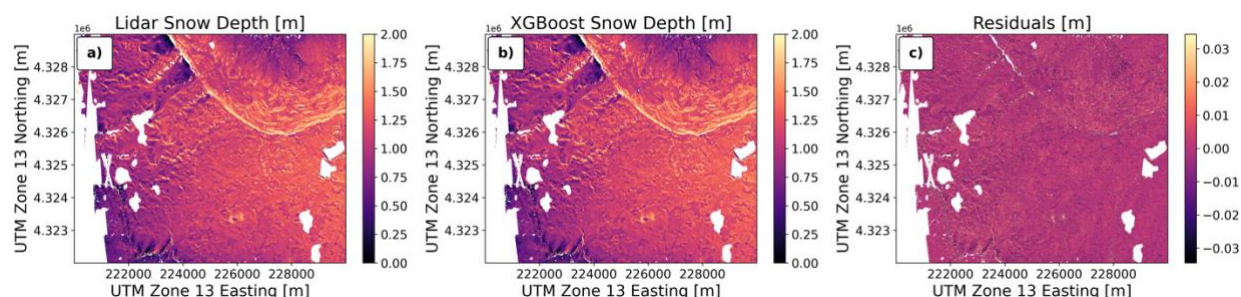


Figure 7. Spatial Residual Analysis. (a) Lidar-derived snow depth. (b) The XGBoost predicted snow depth. (c) The prediction error (lidar minus predicted snow depths).

Figure 7b (predicted depths) visually appears to capture the spatial variability observed in the lidar data (Figure 7a), suggesting good model performance. Figure 7c quantifies the prediction error, visually encoded to emphasize areas of underestimation or overestimation by the model. Figure 7c indicates minimal bias based on the color map.

3.2 Feature Importance

Using SHAP values (Figure 8a), the most influential variables were unwrapped phase, bare ground elevation, and amplitude. Gain-based importance (Figure 8b) also showed unwrapped phase, bare ground elevation, and amplitude as the top 3 predictors, aligning with SHAP. This indicates their high explanatory power for estimating snow depth from InSAR.

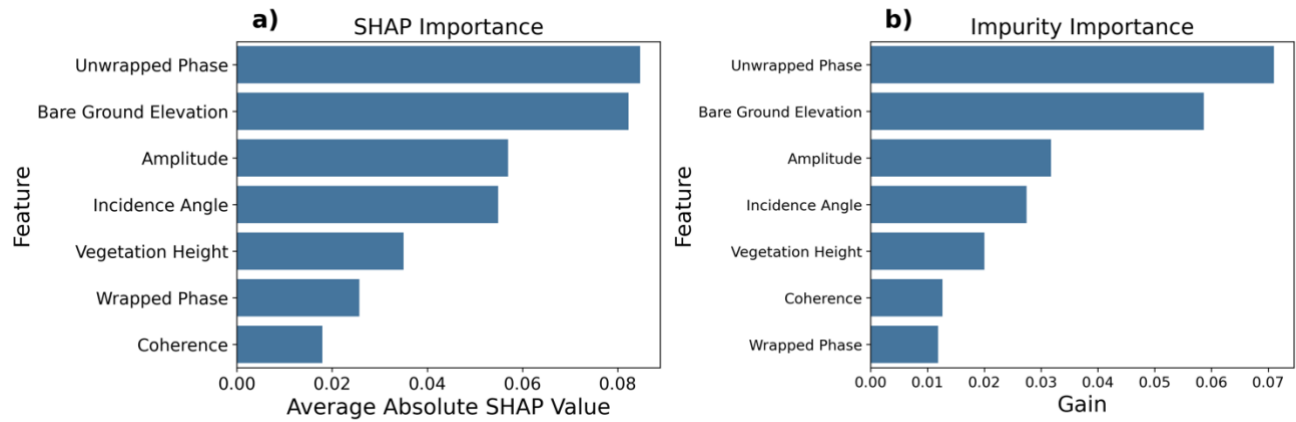


Figure 8. Feature Importance for Model 3 (mixed areas): (a) SHAP importance (b) XGBoost impurity-based importance.

Amplitude backscatter ranked 3rd, providing information on surface roughness and structure. When stratified into open and vegetated areas, the feature importance ranks were the same as Figure 8 in open areas. However, for vegetated areas, bare ground elevation was the leading feature, highlighting the complex relationship between the topography and snow accumulation beneath the canopy. Vegetation height also became more important in forests, ranking 3rd compared to 5th overall. This confirms vegetation's role in attenuating radar signals in dense cover. Feature importance plots for forested and open areas can be found in Figure A3 in the appendix.

3.3 In-Situ Validation

To provide an independent assessment of model accuracy, predicted snow depths were validated against in-situ observations collected as part of the 2017 NASA SnowEx campaign (67). Manual measurements were taken using either a standard, handheld 3-m snow probes or a shorter GPS-equipped 1.2-m MagnaProbe. The GPS technology in the MagnaProbe provides a position accuracy of ± 2.5 m (68). During the intense observation period (February 6-25), a total of 27,081 snow depth measurements were taken at intervals of approximately 3 m.

Validation was done using a 3-m buffer approach, where the average snow depth (lidar and predicted) within a 3-m radius of each in-situ observation was calculated. The lidar snow depths used for developing our models were from February 8. Hence, the data used for validation were measurements taken on February 8 alone (1773 snow depth measurements).

The buffer analysis yielded 234 observations for comparison against the lidar dataset and the complete model predictions (training and testing predictions). However, when the buffer analysis was limited to the testing set alone, the number of observations reduced to 61.

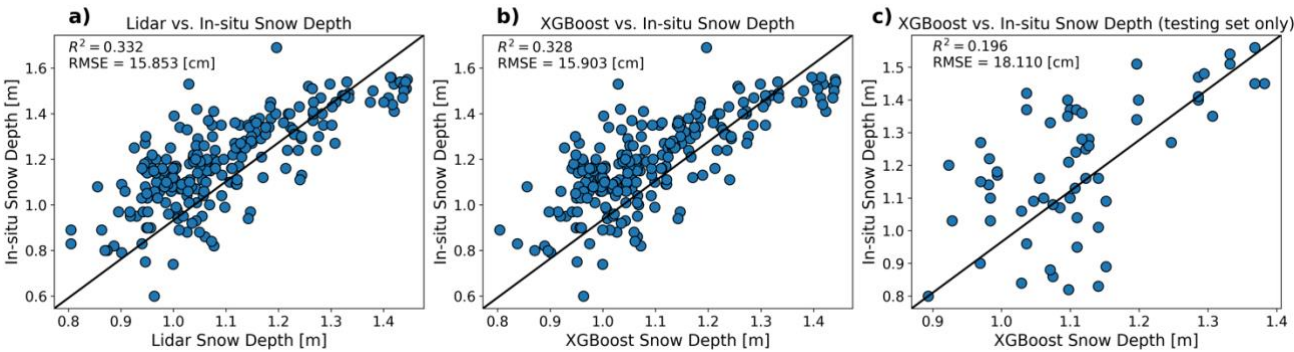
When compared to in-situ measurements (Table 3), the lidar-derived snow depths achieved an RMSE of 15.85 cm and R^2 of 0.332, while snow depths predicted by XGBoost across the entire study area (including both training and testing predictions) achieved a slightly worse performance (RMSE = 15.95 cm and $R^2 = 0.328$). When considering the testing set alone, XGBoost achieved an RMSE of 19.705 cm and an R^2 of 0.095, indicating decreased predictive accuracy.

486

Table 3: In-situ validation results

Model	RMSE (m)	R^2	Sample Size
Lidar	0.1585	0.332	234
XGBoost (whole)	0.1595	0.328	234
XGBoost (test)	0.1811	0.196	61

487 Overall, the in-situ validation provides confidence in the modeling framework, with XGBoost
488 predictions showing accuracy approaching the lidar training data.



489

490 Figure 9. Snow depth validation using in-situ measurements: The closer the data aligns with the
491 diagonal line, the more accurate the snow depth estimates are in comparison to in-situ measurements.
492 (a): Lidar-derived depths versus in-situ depths. (b) XGBoost predicted (both training and testing)
493 depths versus in-situ depths. (c): XGBoost depths (testing set only) versus in-situ depths.

494 **4. Discussion**

495 **4.1 Model Performance**

496 This study explored the potential of machine learning for snow depth prediction using data from NASA
497 JPL's UAVSAR sensor, which employs L-band radar for InSAR measurements. We developed and
498 compared three machine learning algorithms – Extra Trees, XGBoost, and FNN, each selected for their
499 robustness in handling complex datasets and their fast training and prediction time, with XGBoost and
500 FNN capable of benefiting from GPU acceleration. XGBoost consistently outperforms other
501 algorithms. Hence, the results are based on the XGBoost model. To investigate the potential differences
502 in model performance across open and vegetated areas and assess the feasibility of using a single model
503 for snow depth prediction across the entire study area, we partitioned the study area into three distinct
504 categories: vegetated regions, open regions, and a combination of both (mixed).

505 Within each category, we developed three distinct models: a baseline model using only elevation data
506 (model 1), a second model incorporating InSAR parameters and elevation (model 2), and a third model
507 combining InSAR parameters, elevation, and vegetation height (model 3). The addition of vegetation
508 height improved model performance for both vegetated and mixed areas (assuming open areas have no

vegetation). In vegetated areas, including vegetation height as a feature reduced the RMSE from 10.98 cm to 10.46 cm, while in mixed areas, the RMSE reduced from 10.40 cm to 9.88 cm.

These findings suggest that while vegetation height contributes to the model's predictive power, its impact is less pronounced compared to the inclusion of InSAR-derived features. In open areas, adding InSAR parameters reduced the RMSE from 24.83 cm to 9.85 cm compared to the baseline; in vegetated areas, the RMSE reduced from 24.21 cm to 10.98 cm; and in the mixed areas, the RMSE reduced from 26.06 cm to 10.40 cm. The marginal improvements in model accuracy from model 2 to model 3 indicate that the InSAR parameters capture a larger portion of the variability in snow depth, even in the presence of vegetation.

The(69,70) model performed better in open areas than in vegetated areas. This is likely because forests have a more complex structure than open areas, which can scatter and attenuate the InSAR signal. Additionally, trees can intercept snowfall, making it more difficult to measure snow depth beneath the canopy accurately. Although our models performed better in open areas, the results from the vegetated areas also maintained high accuracy. For model 2, the difference in RMSE between open and vegetated areas is 1.13 cm. The high accuracy in vegetated areas can be attributed to the fact that L-band data is less sensitive to the forest structure. This finding is consistent with the work of Hosseini (71), where he found the L-band to be less sensitive to forest structure than the P-band.

Considering the performance metrics across different areas, Model 3 for the mixed areas emerges as the best choice, offering a balance between accuracy and practicality. By incorporating both InSAR-derived attributes and vegetation height, Model 3 effectively captures the variability in snow depth across open and vegetated areas, achieving an impressive R^2 of 0.89 and an RMSE of 9.88 cm. This single model approach for mixed areas eliminates the need to develop separate models for open and vegetated regions, saving time and resources while maintaining high predictive power. The inclusion of vegetation height in Model 3 successfully differentiates between open and vegetated areas, as evidenced by the comparable performance metrics obtained when the mixed area predictions are divided into their respective subsets (vegetated areas: RMSE=10.46 m, R^2 =0.86; open areas: RMSE=9.5 cm, R^2 =0.90). These results further validate the robustness and versatility of Model 3 in accurately estimating snow depth across diverse landscapes. This enhanced performance implies the potential of a single and robust model to reliably estimate snow depth in both vegetated and open areas, which can enhance the efficiency of large-scale snow monitoring efforts.

4.2 Feature Importance

We conducted feature importance using SHAP and gain metrics from XGBoost. This is because the SHAP feature importance provides a more reliable and stable ranking, and the gain metrics can be misleading for features with high cardinality. We found an agreement between the importance of the feature from both SHAP's and XGBoost's gain (Figure 8). Unwrapped phase, bare earth elevation, and amplitude are the top three influential features. Unwrapped phase emerged as the most influential predictor, as it is a direct indicator of snow depth changes, hence its strong influence on the model's predictive power. The bare ground elevation's importance is similarly intuitive; it represents the underlying topography, which is fundamental in understanding snow accumulation patterns. Topography relates to snow depth through factors like elevation, slope, and aspect (69,72). Higher elevations experience more snowfall, north-facing slopes retain snowpack longer in the northern hemisphere, and leeward areas develop drifts. The bare ground elevation provides the terrain context to model these topological influences on snow accumulation. Amplitude, ranking third in importance, suggests that the backscatter intensity, which is affected by surface characteristics, including roughness and snow density, is also an important predictor. This aligns with findings in literature where

backscatter properties have been directly correlated with snow depth. Kings et al. (73) found a strong correlation between Ku-band backscatter and snow depth in the tundra. The remaining features, like incidence angle and coherence, showed lower importance. Further analysis is warranted to fully explain the relative feature contributions. Additionally, from a remote sensing perspective, SAR images contain information on the Earth's surface in the form of the amplitude and phase components of the backscattered radar signal. Given the rich information contained in the amplitude and phase components of SAR images, it is reasonable to expect that these features would rank high on the feature importance scale when used in various remote sensing applications, such as snow depth estimation.

When the data was stratified by open versus vegetated land cover, some notable differences emerged in the feature importance (Figure A3). In open areas, unwrapped phase and bare earth elevation remained the top two predictors, followed by incidence angle and then amplitude. This aligns with the overall importance rankings and reinforces the primacy of phase and topography for snow depth estimation in open regions. However, in vegetated regions, bare earth elevation was the top performer, followed by incidence angle and vegetation height. Unwrapped phase dropped to fourth. The decreased ranking of the unwrapped phase in forests is likely because dense vegetation attenuates and scatters the radar signal, reducing the phase's sensitivity to snow depth variations below the canopy. Meanwhile, vegetation height became the third most important variable, confirming its role in correcting for signal attenuation effects in forested areas. These land cover-specific feature importance findings provide useful insights. The unwrapped phase appears most valuable in open areas where radar penetration is uncompromised, while bare earth elevation and vegetation structure take precedence in forests. These results point to potential pathways for improving InSAR snow depth retrieval through optimal parameterization tailored to different land cover regimes. Additionally, these results can guide future research and data collection efforts, making them more focused, and may lead to more sophisticated physics-based models and physics-informed machine learning.

4.3 In-situ Validation

Validation against in-situ snow depth measurements provides an independent assessment of model accuracy. The lidar training data achieved 15.9 cm RMSE versus in-situ points, aligned with its expected accuracy. For the optimized XGBoost model, predictions across the entire study area (training + testing) achieved 16.0 cm RMSE compared to in-situ data, indicating that the model effectively learned real relationships between InSAR signals and physical snow depth. When evaluating only the reserved 10% testing subset, XGBoost RMSE increased to 18.1 cm. This highlights some generalization errors, likely because the testing data within the 3 m buffer of in-situ depths did not fully represent the diversity of land cover (Figure 7) and terrain conditions across the study area. However, further investigation is warranted to ascertain this claim. In summary, the close performance of the XGBoost model to the lidar snow depth data in terms of RMSE is encouraging, particularly considering the complexities associated with modeling snow depth. Future work could explore the integration of additional features, model refinement, and an expanded in-situ measurement dataset to further enhance model accuracy and reliability.

5. Limitation of the Study

While this study provides valuable insights into the use of InSAR products for estimating total snow depth, there is a limitation that must be acknowledged. A key assumption of this approach is that the spatial patterns of total snow depth remain consistent between the time of snow accumulation and the subsequent InSAR acquisition. In other words, the snow depth change detected by the InSAR instrument is assumed to be representative of the total snow depth distribution. However, if external factors, such as strong winds, alter the snow accumulation pattern during this time interval, the

599 accuracy of the total snow depth estimates may be compromised. Consequently, the effectiveness of
600 this approach relies heavily on the stability of snow depth patterns between InSAR acquisitions.

601 **6. Conclusion and Future Work**

602 This study serves as a proof of concept for the potential of machine learning to estimate snow depth
603 using L-band InSAR data. The XGBoost model demonstrated promising performance, and the feature
604 importance analysis provided valuable insights into the relationships between L-band InSAR features
605 and snow depth. The upcoming NISAR mission, with its global L-band InSAR coverage, presents a
606 unique opportunity to further advance this approach. With the availability of NISAR data, we can
607 expand the training dataset, incorporate additional polarizations, and explore alternative machine
608 learning approaches, potentially leading to even more accurate snow depth estimation. The success of
609 this approach hinges on the availability of representative training data. With sufficient and diverse
610 training data, machine learning models can effectively capture the complex relationships between snow
611 depth and its influencing factors, enabling accurate snow depth prediction. With the advent of NISAR
612 and continued research efforts, we can harness the power of machine learning to potentially improve
613 water resource management.

614 Moving forward, we recommend that future work focus on:

- 615 • Exploring model transferability: Can we develop models that perform well on new datasets
616 without the need for retraining?
- 617 • Analyzing model performance at 10-50 meter resolution: Can we achieve similar performance
618 using lower-resolution data, which is more widely available?
- 619 • Incorporating advanced computer vision approaches: Can we improve performance using 2D
620 estimation methods such as Convolutional Neural Networks (74) and Vision Transformers (75)
621 instead of pixel-wise approaches?

622 Addressing these questions will allow us to further advance the field of machine learning-based snow
623 depth prediction and contribute to more accurate and reliable snow monitoring and forecasting systems.

624

625 7. References

- 626 1. Lievens H, Brangers I, Marshall HP, Jonas T, Olefs M, De Lannoy G. Sentinel-1 snow depth
627 retrieval at sub-kilometer resolution over the European Alps. *Cryosphere*. 2022;16(1):159–77.
- 628 2. Dettinger MD, Anderson ML. Storage in California's reservoirs and snowpack in this time of
629 drought. *San Francisco Estuary and Watershed Science*. 2015;13(2).
- 630 3. Hedrick AR, Marks D, Havens S, Robertson M, Johnson M, Sandusky M, et al. Direct
631 insertion of NASA Airborne Snow Observatory-derived snow depth time series into the
632 iSnobal energy balance snow model. *Water Resour Res*. 2018;54(10):8045–63.
- 633 4. Henn B, Musselman KN, Lestak L, Ralph FM, Molotch NP. Extreme runoff generation from
634 atmospheric river driven snowmelt during the 2017 Oroville Dam spillways incident. *Geophys*
635 *Res Lett*. 2020;47(14):e2020GL088189.
- 636 5. Deems JS, Fassnacht SR, Elder KJ. Fractal distribution of snow depth from LiDAR data. *J*
637 *Hydrometeorol*. 2006;7(2):285–97.
- 638 6. Webb RW, Marziliano A, McGrath D, Bonnell R, Meehan TG, Vuyovich C, et al. In situ
639 determination of dry and wet snow permittivity: improving equations for low frequency radar
640 applications. *Remote Sens (Basel)*. 2021;13(22):4617.
- 641 7. Marshall H, Hawley RL, Tedesco M. Field measurements for remote sensing of the
642 cryosphere. *Remote Sensing of the Cryosphere*. 2015;345–81.
- 643 8. Dong C. Remote sensing, hydrological modeling and in situ observations in snow cover
644 research: A review. *J Hydrol (Amst)*. 2018;561:573–83.
- 645 9. Deems JS, Painter TH, Finnegan DC. Lidar measurement of snow depth: a review. *Journal of*
646 *Glaciology*. 2013;59(215):467–79.
- 647 10. Serreze MC, Clark MP, Armstrong RL, McGinnis DA, Pulwarty RS. Characteristics of the
648 western United States snowpack from snowpack telemetry (SNOTEL) data. *Water Resour*
649 *Res*. 1999;35(7):2145–60.
- 650 11. Bühler Y, Adams MS, Bösch R, Stoffel A. Mapping snow depth in alpine terrain with
651 unmanned aerial systems (UASs): potential and limitations. *Cryosphere*. 2016;10(3):1075–88.
- 652 12. Trujillo E, Lehning M. Theoretical analysis of errors when estimating snow distribution
653 through point measurements. *Cryosphere*. 2015;9(3):1249–64.
- 654 13. Harder P, Pomeroy JW, Helgason WD. Improving sub-canopy snow depth mapping with
655 unmanned aerial vehicles: lidar versus structure-from-motion techniques. *Cryosphere*.
656 2020;14(6):1919–35.

- 657 14. Meyer J, Deems JS, Bormann KJ, Shean DE, Skiles SM. Mapping snow depth and volume at
658 the alpine watershed scale from aerial imagery using Structure from Motion. *Front Earth Sci*
659 (Lausanne). 2022;10:989792.
- 660 15. Chen L, Wang L. Recent advance in earth observation big data for hydrology. *Big Earth Data*.
661 2018;2(1):86–107.
- 662 16. Taheri M, Mohammadian A. An overview of snow water equivalent: Methods, challenges, and
663 future outlook. *Sustainability*. 2022;14(18):11395.
- 664 17. Sinha S, Jeganathan C, Sharma LK, Nathawat MS. A review of radar remote sensing for
665 biomass estimation. *International Journal of Environmental Science and Technology*.
666 2015;12:1779–92.
- 667 18. Aquino C, Mitchard E, McNicol I, Carstairs H, Burt A, Vilca BLP, et al. Using Experimental
668 Sites in Tropical Forests to Test the Ability of Optical Remote Sensing to Detect Forest
669 Degradation at 0.3–30 M Resolutions. In: 2021 IEEE International Geoscience and Remote
670 Sensing Symposium IGARSS. IEEE; 2021. p. 677–80.
- 671 19. Ferraz A, Saatchi S, Bormann KJ, Painter TH. Fusion of NASA Airborne Snow Observatory
672 (ASO) lidar time series over mountain forest landscapes. *Remote Sens (Basel)*.
673 2018;10(2):164.
- 674 20. LAL P, Singh G, Das NN, Entekhabi D. A data-driven snapshot algorithm for high-resolution
675 soil moisture retrievals for the upcoming NISAR mission. In: AGU Fall Meeting Abstracts.
676 2022. p. H42G--1379.
- 677 21. Tarricone J, Webb RW, Marshall HP, Nolin AW, Meyer FJ. Estimating snow accumulation
678 and ablation with L-band interferometric synthetic aperture radar (InSAR). *Cryosphere*.
679 2023;17(5):1997–2019.
- 680 22. Hoppinen Z, Oveisgharan S, Marshall H peter, Mower R, Elder K, Vuyovich C. Snow Water
681 Equivalent Retrieval Over Idaho , Part B : Using L-band UAVSAR Repeat-Pass
682 Interferometry. 2023;(August):1–24.
- 683 23. Ohfwurqlf RI, Ri DQG, Ri K, Qjlqhulqj X, Vkrzv Z, Frkhuhqfh ORZ, et al. Snow Extraction
684 Using X Band Multi Temporal Coherence Based on Insar Technology & Klqd (Pdlo
685 Xhvwfjxrfdl] Khqj # Frp. :18–21.
- 686 24. Li H, Wang Z, He G, Man W. Estimating snow depth and snow water equivalence using
687 repeat-pass interferometric SAR in the northern piedmont region of the Tianshan Mountains. *J*
688 *Sens*. 2017;2017.
- 689 25. Mason MA. Snow Depth Distribution Patterns and Consistency from Airborne Lidar Time
690 Series. 2020;
- 691 26. Deems JS, Fassnacht SR, Elder KJ. Interannual consistency in fractal snow depth patterns at
692 two Colorado mountain sites. *J Hydrometeorol*. 2008;9(5):977–88.

- 693 27. Pflug JM, Lundquist JD. Inferring distributed snow depth by leveraging snow pattern
694 repeatability: Investigation using 47 lidar observations in the Tuolumne watershed, Sierra
695 Nevada, California. *Water Resour Res.* 2020;56(9):e2020WR027243.
- 696 28. Schirmer M, Lehning M. Persistence in intra-annual snow depth distribution: 2. Fractal
697 analysis of snow depth development. *Water Resour Res.* 2011;47(9).
- 698 29. Sturm M, Wagner AM. Using repeated patterns in snow distribution modeling: An Arctic
699 example. *Water Resour Res.* 2010;46(12).
- 700 30. Reichstein M, Camps-Valls G, Stevens B, Jung M, Denzler J, Carvalhais N. Deep learning and
701 process understanding for data-driven Earth system science. *Nature.* 2019;566(7743):195–204.
- 702 31. Yuan Q, Shen H, Li T, Li Z, Li S, Jiang Y, et al. Deep learning in environmental remote
703 sensing: Achievements and challenges. *Remote Sens Environ.* 2020;241:111716.
- 704 32. Zhang B, Chen Z, Peng D, Benediktsson JA, Liu B, Zou L, et al. Remotely sensed big data:
705 Evolution in model development for information extraction [point of view]. *Proceedings of the*
706 *IEEE.* 2019;107(12):2294–301.
- 707 33. Ofekeze E, Marshall HP, Mead J, Alabi IO. A Machine Learning Framework for Active and
708 Passive Microwave Observation from Snow Water Equivalent Synthetic Aperture Radar And
709 Radiometer (SWESARR) for Snow Depth Estimation. In: *AGU Fall Meeting Abstracts.* 2022.
710 p. C52C--0364.
- 711 34. Tedesco M, Pulliainen J, Takala M, Hallikainen M, Pampaloni P. Artificial neural network-
712 based techniques for the retrieval of SWE and snow depth from SSM/I data. *Remote Sens*
713 *Environ.* 2004;90(1):76–85.
- 714 35. Ofekeze E, Marshall HP, Mead J, Trujillo E, Alabi IO. Advancing Snow Water Equivalent
715 Estimation with SWESARR. *AGU23.* 2023;
- 716 36. Alabi IO, Marshall HP, Mead J, Trujillo E, Ofekeze E. Harnessing L-Band InSAR and Lidar
717 Data Through Machine Learning for Accurate Snow Depth Estimation in Grand Mesa,
718 Colorado. *AGU23.* 2023;
- 719 37. Alabi IO, Marshall HP, Mead J, Ofekeze E. How Transferable are Our Models? A Case Study
720 of Idaho SNOTEL Sites. In: *AGU Fall Meeting Abstracts.* 2022. p. C52C--0365.
- 721 38. Feng T, Zhu S, Huang F, Hao J, Mind'je R, Zhang J, et al. Spatial variability of snow density
722 and its estimation in different periods of snow season in the middle Tianshan Mountains,
723 China. *Hydrol Process.* 2022;36(8):e14644.
- 724 39. Bair EH, Abreu Calfa A, Rittger K, Dozier J. Using machine learning for real-time estimates
725 of snow water equivalent in the watersheds of Afghanistan. *Cryosphere.* 2018;12(5):1579–94.
- 726 40. Broxton PD, Van Leeuwen WJD, Biederman JA. Improving snow water equivalent maps with
727 machine learning of snow survey and lidar measurements. *Water Resour Res.*
728 2019;55(5):3739–57.

- 729 41. Hu Y, Che T, Dai L, Xiao L. Snow depth fusion based on machine learning methods for the
730 northern hemisphere. *Remote Sens (Basel)*. 2021;13(7):1–24.
- 731 42. Liang J, Liu X, Huang K, Li X, Shi X, Chen Y, et al. Improved snow depth retrieval by
732 integrating microwave brightness temperature and visible/infrared reflectance. *Remote Sens*
733 *Environ [Internet]*. 2015;156:500–9. Available from:
734 <http://dx.doi.org/10.1016/j.rse.2014.10.016>
- 735 43. Che T, Li X, Jin R, Armstrong R, Zhang T. Snow depth derived from passive microwave
736 remote-sensing data in China. *Ann Glaciol*. 2008;49:145–54.
- 737 44. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd*
738 *acm sigkdd international conference on knowledge discovery and data mining*. 2016. p. 785–
739 94.
- 740 45. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Mach Learn*. 2006;63:3–42.
- 741 46. Yegnanarayana B. *Artificial neural networks*. PHI Learning Pvt. Ltd.; 2009.
- 742 47. Chesnutt JM, Wegmann KW, Cole RD, Byrne PK. Landscape Evolution Comparison between
743 Sacra Mensa, Mars and the Grand Mesa, Colorado, USA. In: *AGU Fall Meeting Abstracts*.
744 2017. p. EP53B-1690.
- 745 48. Gatebe C, Li W, Chen N, Fan Y, Poudyal R, Brucker L, et al. Snow-covered area using
746 machine learning techniques. In: *IGARSS 2018-2018 IEEE International Geoscience and*
747 *Remote Sensing Symposium*. IEEE; 2018. p. 6291–3.
- 748 49. Kulakowski D, Veblen TT, Drinkwater S. The persistence of quaking aspen (*Populus*
749 *tremuloides*) in the Grand Mesa area, Colorado. *Ecological Applications*. 2004;14(5):1603–14.
- 750 50. Kim E, Gatebe C, Hall D, Newlin J, Misakonis A, Elder K, et al. OVERVIEW OF SNOWEX
751 YEAR 1 ACTIVITIES NASA Goddard Space Flight Center , 2 USRA , 3 Aerospace Corp ., 4
752 ATA Aerospace , 5 US Forest Service ., *Igarss*. 2017;1388–90.
- 753 51. Arendt AA, Marshall HP, Henderson ST, Vuyovich C, Haley C, Reckase G, et al. Lessons
754 Learned from SnowEx 2022 Hackweek for Fostering Open Science Communities. In: *AGU*
755 *Fall Meeting Abstracts*. 2022. p. ED16B-04.
- 756 52. Brucker L, Gsfc N, Gestar U, Hiemstra C, Marshall H peter, Uni BS, et al. A FIRST
757 OVERVIEW OF SNOWEX GROUND-BASED REMOTE SENSING ACTIVITIES
758 DURING THE WINTER 2016-2017 Colorado ; Richard Kelly , Uni . Waterloo ; Jason Kraft ,
759 NASA GSFC ; Alexandre Langlois , Uni . Sherbrooke ; Daniel McGrath , Colorado State Uni
760 .; Chelsea Merr. 2017;(Figure 3):1391–4.
- 761 53. Brucker L, Hiemstra C, Marshall HP, Elder K, De Roo R, Mousavi M, et al. A first overview
762 of SnowEx ground-based remote sensing activities during the winter 2016--2017. In: *2017*
763 *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. 2017. p. 1391–4.
- 764 54. King F, Kelly R, Fletcher CG. Evaluation of lidar-derived snow depth estimates from the
765 iPhone 12 pro. *IEEE Geoscience and Remote Sensing Letters*. 2022;19:1–5.

- 766 55. Killinger DK. Lidar (light detection and ranging). In: Laser spectroscopy for sensing. Elsevier;
767 2014. p. 292–312.
- 768 56. Revuelto J, López-Moreno JI, Azor\`in-Molina C, Zabalza J, Arguedas G, Vicente-Serrano
769 SM. Mapping the annual evolution of snow depth in a small catchment in the Pyrenees using
770 the long-range terrestrial laser scanning. *J Maps*. 2014;10(3):379–93.
- 771 57. Currier WR, Pflug J, Mazzotti G, Jonas T, Deems JS, Bormann KJ, et al. Comparing aerial
772 lidar observations with terrestrial lidar and snow-probe transects from NASA’s 2017 SnowEx
773 campaign. *Water Resour Res*. 2019;55(7):6285–94.
- 774 58. Lu Z, Kwoun O, Rykhus R. Interferometric synthetic aperture radar (InSAR): its past, present
775 and future. *Photogramm Eng Remote Sensing*. 2007;73(3):217.
- 776 59. Ozdemir S. Feature Engineering Bookcamp. Simon and Schuster; 2022.
- 777 60. Meloche J, Langlois A, Rutter N, McLennan D, Royer A, Billecocq P, et al. High-resolution
778 snow depth prediction using Random Forest algorithm with topographic parameters: A case
779 study in the Greiner watershed, Nunavut. *Hydrol Process*. 2022;36(3).
- 780 61. Maxwell AE, Warner TA, Fang F. Implementation of machine-learning classification in
781 remote sensing: An applied review. *Int J Remote Sens* [Internet]. 2018;39(9):2784–817.
782 Available from: <https://doi.org/10.1080/01431161.2018.1433343>
- 783 62. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: A Next-generation Hyperparameter
784 Optimization Framework. *Proceedings of the ACM SIGKDD International Conference on*
785 *Knowledge Discovery and Data Mining*. 2019;2623–31.
- 786 63. Zhao M, Li J. Tuning the hyper-parameters of CMA-ES with tree-structured Parzen
787 estimators. *Proceedings - 2018 10th International Conference on Advanced Computational*
788 *Intelligence, ICACI 2018*. 2018;613–8.
- 789 64. Kingma DP, Ba JL. Adam: A method for stochastic optimization. *3rd International Conference*
790 *on Learning Representations, ICLR 2015 - Conference Track Proceedings*. 2015;1–15.
- 791 65. Zhang K gang, Zhang Y dong, Wang M. A Unified Approach to Interpreting Model
792 Predictions Scott. *Nips*. 2012;16(3):426–30.
- 793 66. Man X, Chan E. The best way to select features? comparing mda, lime, and shap. *The Journal*
794 *of Financial Data Science Winter*. 2021;3(1):127–39.
- 795 67. Brucker L, Hiemstra C, Marshall H, Elder K. SnowEx17 Community Snow Depth Probe
796 Measurements, Version 1 [Internet]. NASA National Snow and Ice Data Center Distributed
797 Active Archive Center; 2018. Available from: https://nsidc.org/data/SNEX17_SD/versions/1
- 798 68. Sturm M, Holmgren J. An Automatic Snow Depth Probe for Field Validation Campaigns.
799 *Water Resour Res*. 2018;54(11):9695–701.

800 69. Trujillo E, Ramirez JA, Elder KJ. Topographic, meteorologic, and canopy controls on the
801 scaling characteristics of the spatial distribution of snow depth fields. *Water Resour Res.*
802 2007;43(7).

803 70. Musselman KN, Molotch NP, Brooks PD. Effects of vegetation on snow accumulation and
804 ablation in a mid-latitude sub-alpine forest. *Hydrological Processes: An International Journal.*
805 2008;22(15):2767–76.

806 71. Hosseini S, Garestier F. Pol-InSAR sensitivity to hemi-boreal forest structure at L-and P-
807 bands. *International Journal of Applied Earth Observation and Geoinformation.*
808 2021;94:102213.

809 72. Hojatimalekshah A, Uhlmann Z, Glenn NF, Hiemstra CA, Tennant CJ, Graham JD, et al. Tree
810 canopy and snow depth relationships at fine scales with terrestrial laser scanning. *Cryosphere.*
811 2021;15(5):2187–209.

812 73. King J, Kelly R, Kasurak A, Duguay C, Gunn G, Rutter N, et al. Spatio-temporal influence of
813 tundra snow properties on Ku-band (17.2 GHz) backscatter. *Journal of Glaciology.*
814 2015;61(226):267–79.

815 74. O’Shea K, Nash R. An introduction to convolutional neural networks. *arXiv preprint*
816 *arXiv:151108458.* 2015;

817 75. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An
818 image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint*
819 *arXiv:201011929.* 2020;

820