A Machine Learning Model for Estimating Snow Water Equivalent from Snow Depth and Seasonal Snow Climate Classes

Ibrahim Olalekan Alabi,[a] Hans-Peter Marshall,[b] Jodi Mead,[c] and Ernesto Trujillo[b]

[a] *Computing Ph.D., Boise State University, Boise, Idaho, United States*

[b] *Department of Geosciences, Boise State University, Boise, Idaho, United States*

[c] *Department of Mathematics, Boise State University, Boise, Idaho, United States*

Corresponding author: Ibrahim Olalekan Alabi, ibrahimolalekana@u.boisestate.edu

ABSTRACT: Direct snow water equivalent (SWE) measurements are time-consuming, labor-intensive, and often impractical over large areas. However, if the snowpack bulk density is known, SWE can be estimated from snow depth. Snow depth is a more accessible proxy for estimating SWE, as it can be quickly and easily measured with high precision. We propose a machine learning (ML) model that estimates snowpack bulk density from snow depth and other variables that only require the location and date of snow depth acquisitions. The estimated density can then be used to calculate SWE from depth. Our model was trained on approximately 2 million data points and tested on 544,513 testing samples from 864 SNOTEL sites in the western United States. We compared the proposed ML model to state-of-the-art statistical models, and it outperformed them all, reducing SWE estimation RMSE by up to 40% compared to the best-performing statistical model. To evaluate the ML model's transferability, we tested it using data from the Maine Snow Survey - a different snow climate in the northeastern United States. The model provided promising SWE estimates without retraining and showed improved performance after retraining with a small portion of local data. Our ML model offers a practical and broadly applicable tool for estimating SWE in regions with limited snowpack monitoring.

SIGNIFICANCE STATEMENT: This study aims to improve snow water equivalent (SWE) estimates by developing a machine learning (ML) model that uses snow depth and minimal additional data. SWE is critical for managing water resources but is challenging to measure directly, and remote sensing techniques have yet to prove globally reliable for SWE estimation. Snow depth, however, can be easily measured with high precision using lidar and in-situ probes. We propose an ML model that predicts snow density from snow depth. The estimated density can then be used to calculate SWE from depth. This method holds potential for global application and provides a convenient tool for SWE estimation in regions with limited snowpack monitoring.

1. Introduction

Recent studies have shown a decline in snow cover extent (SCE) in many regions of the world (Mote et al. 2005; Stewart 2009; Brown et al. 2010; Najafi et al. 2016; Thiebault and Young 2020; Siirila-Woodburn et al. 2021; Young 2023), which threatens the future water availability for the nearly two billion people who rely on spring snowmelt for water supply. Analyzing the MOD10C2 eight-day maximum value composite time series data, Young (2023) demonstrated that between February 2000 and March 2023, global SCE decreased at a rate two to ten times faster than it increased, depending on the season. This trend translates to a 5.12% annual reduction in global snow cover, excluding Antarctica and Greenland (Young 2023). The observed SCE reductions are primarily attributed to climate change, with warming temperatures causing winter precipitation to fall as rain rather than snow (Knowles et al. 2006; Feng and Hu 2007; Brown and Robinson 2011; Chen et al. 2016; Sun et al. 2020; Young and Young 2021). As global temperatures continue to rise, climate projections indicate further reductions in snow cover (Brown and Robinson 2011; Siirila-Woodburn et al. 2021), intensifying concerns over water scarcity in snow-dependent regions.

While SCE provides valuable insights into the presence of snow, it offers limited information about the actual volume of water stored in snowpacks (Brown 2000; Brown et al. 2003; Sturm et al. 2010). The more critical metric for understanding water availability is Snow Water Equivalent (SWE), which represents the amount of water we would get if we instantly melt an entire snowpack. Accurate SWE estimation is essential for managing water resources, predicting floods, and understanding climate dynamics in snow-dominated regions (Hedrick et al. 2020; Tsang et al. 2021). However, current SWE monitoring relies heavily on field and station measurements, such

as manual snow coring (Church 1933) and snow pillows (Beaumont 1965), both of which have limitations. Snow coring measurements are labor-intensive, time-consuming, and not feasible for continuous, around-the-clock data collection (Kinar and Pomeroy 2015). Moreover, coring is prone to human error; for instance, snow may either fall out of the coring device or become compressed during extraction, leading to measurement biases (Hill et al. 2019). Coring devices, as a general rule, are considered accurate to within ±10% of the true SWE value (Hill et al. 2019; Johnson et al. 2015). On the other hand, snow pillows, while automated, have logistic issues with respect to installation. Their large size necessitates access to suitable terrain, typically by vehicle, which limits their deployment to areas with relatively simple topography (Hill et al. 2019). Furthermore, snow pillows cannot be rapidly deployed or relocated because installation must be done before the snow season (Hill et al. 2019). Finally, despite advances in remote sensing enabling global snow cover tracking (e.g., the MODIS product), no reliable global remote sensing method for SWE estimation has yet to be established (Painter et al. 2016; Tsang et al. 2021; Meehan et al. 2024). As snow cover continues to decline, the ability to accurately monitor and predict SWE becomes increasingly critical.

Given the limitations of direct SWE measurement, researchers have increasingly turned to snow depth (h_s) as a more practical and abundant proxy for estimating snow water resources (Sturm et al. 2010; Hill et al. 2019). Manual SWE measurement is estimated to take approximately 20 times longer than manual snow depth measurement (Sturm et al. 2010). Snow depth can be measured with high precision and at high resolution over large spatial scales using remote sensing technologies such as lidar (Deems et al. 2013), or as point measurements taken manually with in-situ depth probes (Sturm and Holmgren 2018). In-situ snow depths can also be automated at a single point using low-cost ultrasonic sensors (Ryan et al. 2008), making snow depths far more accessible than direct SWE measurements. As a result, snow depth data significantly outnumber SWE data, with estimates suggesting a ratio of 30 to 1 for manual measurements (Sturm et al. 2010). This ratio is likely even higher when considering remote sensing technologies, which provide large-scale, high-resolution snow depth data.

Background

SWE is strongly correlated with snow depth (see Figure 1a), and by leveraging this relationship, we can estimate SWE from snow depth measurements. Specifically, the ratio of SWE to snow depth is referred to as the snowpack bulk density (ρ_b). Once ρ_b is known, we can easily convert h_s to SWE estimates using Equation (1):

$$SWE = h_s \times \frac{\rho_b}{\rho_w} \tag{1}$$

where ρ_w is the density of water (1 $g\ cm^{-3}$), h_s is measured in cm, ρ_b is measured in $g\ cm^{-3}$, and SWE is measured in cm of water.

This approach has important implications for regions where h_s is already measured, as SWE estimates can be derived with minimal additional labor or expense (Jonas et al. 2009; Sturm et al. 2010), simply by calculating ρ_b. Moreover, estimating ρ_b from h_s is the most practical path to potentially accurate SWE (Sturm et al. 2010). Previous studies have reported that h_s is more variable than ρ_b (Dickinson and Whiteley 1972; Pomeroy and Gray 1995; Sturm et al. 2010; McCreight and Small 2014; Bonnell et al. 2023). In our training data, h_s ranged from 5 cm to 711 cm, while ρ_b ranged from 0.05 $g\ cm^{-3}$ to 0.6 $g\ cm^{-3}$ (data not shown). When normalized by their respective means, h_s was four times more variable than ρ_b. Other studies, such as Sturm et al. (2010), have reported up to ten times the variability in h_s compared to ρ_b. Therefore, directly measuring the more variable (and easier to measure) h_s and estimating the less variable ρ_b is the most efficient, cost-effective, and practical method for achieving accurate SWE estimates. Additionally, the ability to accurately predict SWE using widely available snow depth data can lead to an expansion in global SWE datasets, providing crucial information for water resource management in regions lacking direct SWE measurements.

With the clear relationship between h_s and SWE, and the relative ease of acquiring h_s, our focus in this paper is to develop a machine learning algorithm that can predict ρ_b from h_s. Snow density varies with temperature, precipitation, elevation, and snowpack age, and typically ranges from 0.05 $g\ cm^{-3}$ for fresh, dry snow to upwards of 0.5 $g\ cm^{-3}$ for older, denser snowpacks (Paterson 1994; Golubev and Frolov 2000; Judson and Doesken 2000). Over time, snow undergoes compaction

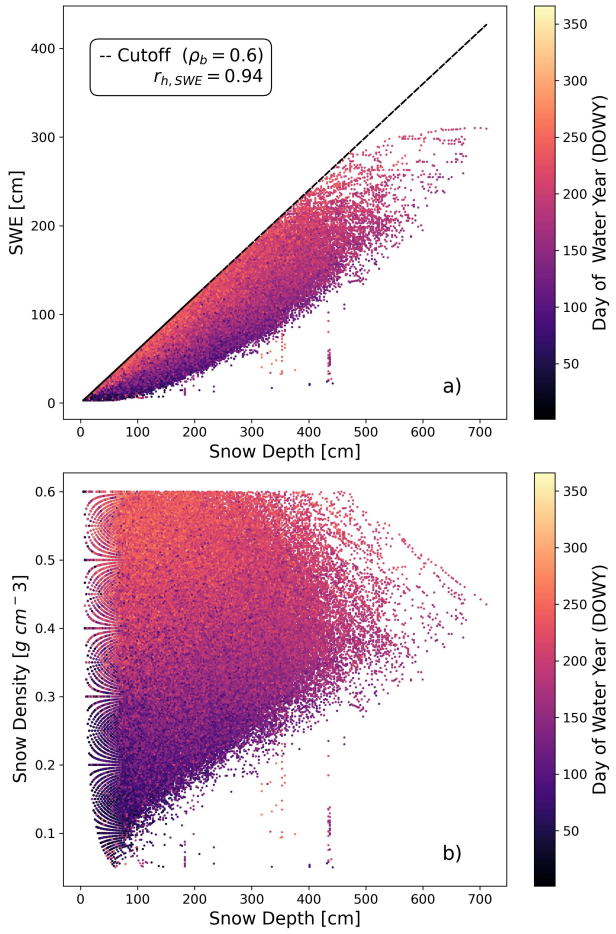


FIG. 1. **(a)** Scatter plot showing a strong positive linear correlation between h_s and SWE in the training dataset, with a dashed black line representing the cutoff at $\rho_b = 0.6\ g\ cm^{-3}$. **(b)** Scatter plot illustrating the relationship between h_s and ρ_b in the training dataset. For both plots, the data is color-coded by Day of Water Year (DOWY), where October 1st corresponds to the start of the water year. The color bar (DOWY) in both plots shows how snowpack characteristics evolve over the course of the snow season.

and metamorphism, which increases its density as the snowpack ages. Several statistical and physics-based models have been developed to estimate snow density based on these factors.

Statistical models often use empirical relationships derived from observational data. For instance, Sturm et al. (2010) proposed a model incorporating snow climate classifications (as defined by Sturm et al. (1995)), snow depth, and day of the water year to estimate snow density. The model was optimized separately for each snow climate classification, with specific parameters provided for each category. Similarly, Jonas et al. (2009) focused on estimating snow density using snow depth,

season (i.e., the month of h_s measurement), and site-specific characteristics such as altitude and region in the Swiss Alps. The authors developed a linear regression model where snow depth is the primary input, and the slope and intercept are determined by season and site altitute. Additionally, Jonas et al. (2009) provided offset values to correct for biases specific to the Swiss Alps. Pistocchi (2016) introduced a simple, time-based approach that estimates snow density as a linear function of the day of the water year. The model assumes that snowpack density starts at an initial value of 0.2 $g\ cm^{-3}$ on November 1st and increases by 0.001 $g\ cm^{-3}$ each day until the end of the snow season. For a more comprehensive review of density regression equations, readers can refer to Avanzi et al. (2015).

While statistical models have advanced our understanding of snow density estimation, they often rely on simplifying assumptions that may not fully capture the complexities of snowpack dynamics. For instance, some models assume a steady increase in snow density throughout the season (e.g., Pistocchi (2016)), while others presume h_s and ρ_b are positively correlated (e.g., Jonas et al. (2009) and Sturm et al. (2010)). The positive correlation assumption, although valid for the original application of Jonas et al. (2009) and Sturm et al. (2010) (because their training data were sampled at intervals of at least two weeks), does not always hold true. For example, ρ_b can decrease even as h_s increases due to the accumulation of fresh, low-density snow (see Figure 2; see also Figure 1 in McCreight and Small (2014)), resulting in negative correlation. Conversely, in the days following snowfall events, the freshly fallen snow undergoes compaction, leading to a decrease in h_s while ρ_b increases, resulting again in a negative correlation (see Figure 2). This observation is consistent with previous work that demonstrated that h_s and ρ_b can exhibit strong negative correlations at certain timescales (McCreight and Small 2014). Before peak SWE, h_s and ρ_b are negatively correlated over short timescales (5-10 days) and positively correlated on longer timescales (McCreight and Small 2014). After peak SWE, as the snowpack melts, ρ_b increases while h_s decreases, leading to negative correlations over both daily and longer timescales (McCreight and Small 2014). These short-term fluctuations are why the steady increase in ρ_b throughout the snow season implied by the model proposed by Pistocchi (2016) oversimplifies the dynamics of snowpack evolution. The short-term fluctuations are particularly important for high temporal resolution measurements, such as daily snow density or SWE estimates, yet they are not well captured by traditional statistical models. In fact, Jonas et al. (2009) acknowledged that their

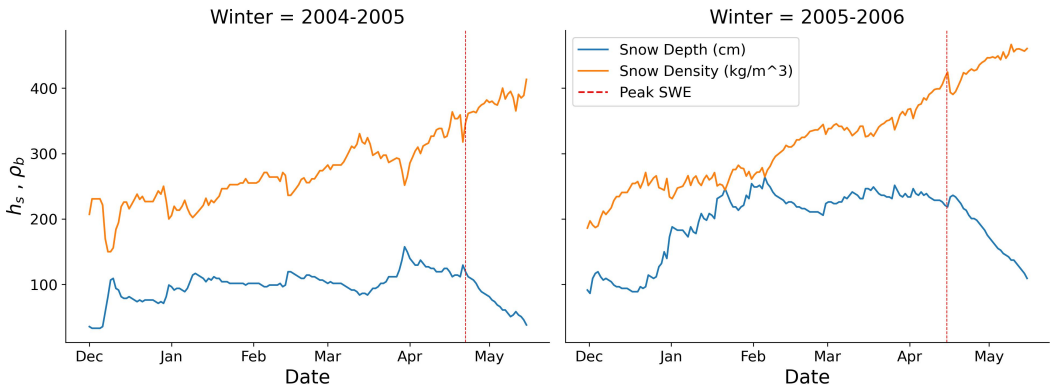


FIG. 2. Two randomly selected winter seasons (2004–2005 and 2005–2006) from the Banner Summit Snowpack Telemetry site in Boise County, Idaho. These seasons demonstrate the consistent interannual relationship between h_s and ρ_b across different years. Although ρ_b generally increases throughout the snow season due to compaction, it can temporarily decrease following snow accumulation events. After peak SWE (indicated by the red dashed line), h_s and ρ_b exhibit a negative correlation, both in the short term and long term, as snow depth decreases and snow density increases during the melt season.

model was not well-suited for daily snow density estimates, as transient phenomena like the settling of fresh snow are not adequately represented.

Physics-based models offer an alternative approach by simulating the physical processes governing snowpack evolution. These models solve the surface energy balance and track the metamorphism and settlement of individual snow layers over time (Brun et al. 1989; Liston and Elder 2006; Lehning et al. 2006). For example, the Crocus model developed by Brun et al. (1989) and later enhanced by Brun et al. (1992) simulates detailed snowpack stratigraphy and properties by accounting for energy exchanges and snow microstructure changes. Similarly, the SnowModel by Liston and Elder (2006), and the SNOWPACK model by Bartelt and Lehning (2002); Lehning et al. (2002) use meteorological inputs to simulate snowpack processes at high temporal resolution. Although physics-based models can simulate detailed snowpack processes and potentially capture short-term variations, their extensive meteorological data requirements (e.g., temperature, precipitation, wind speed, humidity, radiation, et cetera) and computational complexity make them impractical for large-scale or data-sparse applications. Even simpler physics-based models, such as SNOW-17 (Anderson 1973) or the model by De Michele et al. (2013), require time series of meteorological variables like temperature and precipitation. However, many snow depth measurements, including

automated observations, do not have accompanying meteorological data, and obtaining such data can represent significant additional expense and effort (McCreight and Small 2014).

In this paper, we aim to develop a machine learning (ML) model capable of producing snow density estimates for any given day within the snow season using snow depth and other features that can be measured along with snow depth with minimal extra effort. McCreight and Small (2014) introduced a regression model for daily snow density predictions, but their approach requires a continuous time series of depth measurements, which limits its applicability to locations with constant monitoring. Given that our goal is to develop a model that is applicable everywhere snow depth is measured, relying on a time series of depth as inputs is not feasible.

Our method aims to retain the practical advantages of statistical models—such as requiring minimal input data—while overcoming some of their limitations through machine learning techniques. The motivation for using machine learning lies in its ability to model complex, non-linear relationships that might be difficult to capture with traditional statistical approaches. Given that the relationship between snow density and snow depth is inherently non-linear (see Figure 1b; see also Figure 4 in Sturm et al. (2010)), an ML approach is justified. Our approach is most similar to the work of Jonas et al. (2009) and Sturm et al. (2010), but with three key differences. First, our model is trained on daily snow depth data, allowing it to capture the full range of snowpack variability throughout the snow season. This daily granularity enables the model to account for short-term fluctuations in snowpack conditions that previous models, which used less frequent observations, might miss. Second, our model uses a machine learning framework, which allows it to learn directly from the data without relying on predefined assumptions or functional forms. This flexibility enables our model to recognize general pattern in snowpack behavior and capture the dynamic interactions between snow density and the input features over the course of the season. Third, unlike Sturm et al. (2010), who provided different coefficients for different snow climate classifications, and Jonas et al. (2009), who provided different coefficients for different months and altitudes, our ML model can generalize across these conditions without the need for separate coefficients. This makes our model more flexible and potentially easier to apply across a wide range of conditions.

2. Materials and Methods

a. Data

The goal of this study is to develop a machine learning (ML) model that can estimate snowpack bulk density (ρ_b) and, subsequently, snow water equivalent (SWE) from snow depth (h_s). To ensure the model is broadly applicable, the main model uses snow depth, snow climate class (Sturm and Liston 2021; Liston and Sturm 2021), day of water year (DOWY), and elevation as input features. The additional features only require recording the date and location of h_s acquisitions.

However, to explore the potential benefits of including additional climate information, a secondary model is developed that incorporates the two-week average of temperature and precipitation as additional features. This enables us to assess whether adding these variables improves the accuracy of ρ_b estimates in regions where they are available. Therefore, although climate variables (temperature and precipitation) are included in the dataset, the main model does not depend on them. This design ensures that the model can be applied in regions where climate data are not available.

The dataset used in this study comes from two sources: (1) in-situ measurements from the Snowpack Telemetry network (Serreze et al. 1999; Dressler et al. 2006) and (2) the Snow Climate Classifications (Sturm et al. 1995), which assigns each SNOTEL site to a specific snow climate class.

1) USA NRCS Snowpack Telemetry Network

The primary dataset for this study comes from Snow Telemetry (SNOTEL) sites. The SNOTEL network, operated by the Natural Resources Conservation Service (NRCS), is an extensive system of automated data collection sites strategically placed in remote, high-elevation mountain watersheds across the western United States. The automated SNOTEL system was created in 1977 (Helms et al. 2008) and has become vital for monitoring snowpack conditions.

Each SNOTEL site is equipped with sensors to measure various parameters critical to understanding snowpack dynamics. For this study, we selected SWE, h_s, air temperature, and precipitation as variables of interest. SNOTEL sites do not directly measure ρ_b. Instead, ρ_b was calculated as the ratio of SWE to h_s (see Equation (1)). SWE is measured using snow pillows with a precision of 0.254 cm, while h_s is recorded using ultrasonic sensors with a precision of 1.27 cm. Air

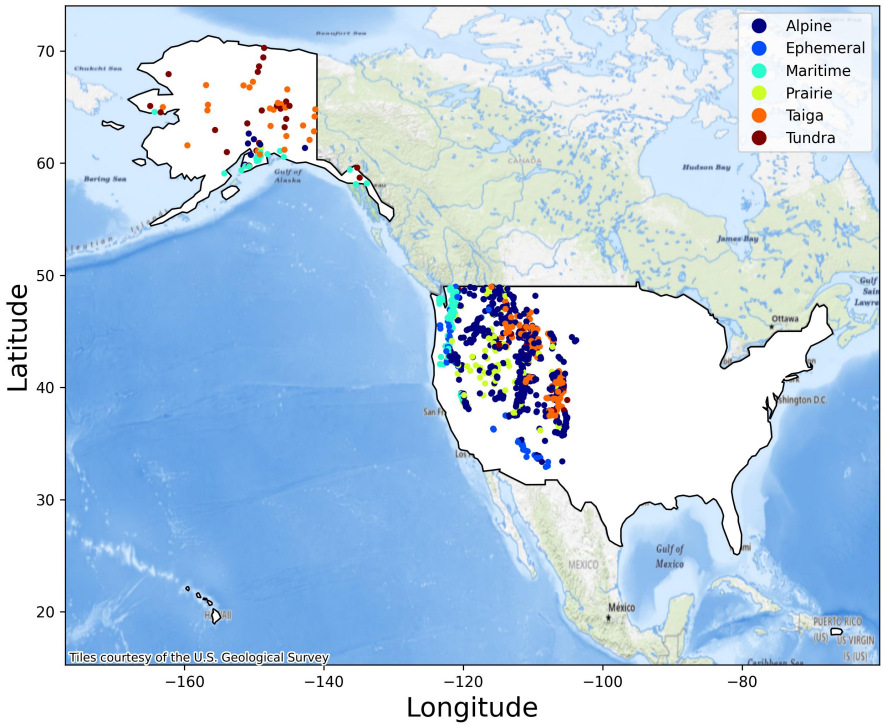
temperature is obtained via a shielded thermistor with a precision of 0.1 °C, and precipitation is measured using either tipping bucket gauges with a precision of 0.0254 cm or storage-type gauges with a precision of 0.254 cm.

We used data spanning from January 2000 to December 2023 from 864 active SNOTEL sites. The SNOTEL network currently consists of 911 active sites (NRCS: `https://wcc.sc.egov.usda.gov/nwcc/yearcount?network=sntl&counttype=yearcount&state=`). However, several stations were excluded from the analysis due to data completeness issues. Specifically, because ρ_b was derived as the ratio of SWE to h_s, data points were dropped if either SWE or h_s was missing or recorded as zero, as this results in an undefined ρ_b calculation (e.g., division by zero). Consequently, some SNOTEL sites had no valid data points remaining after this step and were excluded from the analysis. After this initial data cleaning, we were left with 864 SNOTEL sites that had at least one valid data point for calculating ρ_b. The distribution of these 864 sites across the western United States and their elevation ranges are displayed in Figures 3 and 4b, respectively.

Additional data cleaning steps were applied to ensure that retained data points were physically plausible. Snow density values below 0.05 $g\ cm^{-3}$ and above 0.6 $g\ cm^{-3}$ were dropped. These thresholds were selected based on typical snow density ranges reported in the literature (Paterson 1994; Golubev and Frolov 2000; Judson and Doesken 2000; Sturm et al. 2010). According to Paterson (1994), fresh, dry snow typically has a density between 0.05 $g\ cm^{-3}$ and 0.07 $g\ cm^{-3}$, while Sturm et al. (2010) noted that the maximum density of seasonal snow without water infiltration approaches 0.6 $g\ cm^{-3}$. Additionally, we retained only data points with $h_s > 5$ cm and SWE > 3 cm to remove extremely shallow snowpack conditions that could lead to unreliable ρ_b estimates due to measurement uncertainties and instrument limitations. The final dataset comprised 2,722,562 data points, with each data point representing a unique combination of SWE, ρ_b, h_s, air temperature, and precipitation.

2) Seasonal Snow Climate Classification

In addition to the SNOTEL data, each station was classified into a seasonal snow climate class using the revised seasonal snow climate classification system (SSCCS) published in 2021 (Sturm and Liston 2021; Liston and Sturm 2021). This updated SSCCS provides a higher resolution of 300 m (10-arc-second) grid, compared to the original 50 km grid introduced in Sturm et al. (1995).



239 FIG. 3. Geographic distribution of the SNOTEL stations used in this study across the western United States
240 and Alaska, categorized by snow climate class.

260 The updated SSCCS classifies snow into eight distinct classes: alpine, taiga, maritime, prairie,
261 ephemeral, tundra, ice, and ocean. For this study, we considered only the first six classes, as the
262 ice and ocean classes pertain to glaciers and frozen water bodies, which do not represent seasonal
263 snow.

264 Using this revised classification, each SNOTEL site in our study was assigned to one of the six
265 relevant classes based on its location (latitude and longitude). This classification serves as an input
266 feature in our machine learning model to account for variations in snowpack characteristics across
267 different regions. Figures 3 and 4a illustrate the distribution of SNOTEL sites by snow climate
268 class across the western United States. The alpine class was the most prevalent, accounting for
269 53% of the sites, followed by taiga (17%), maritime (11%), prairie (9%), ephemeral (6%), and
270 tundra (4%).

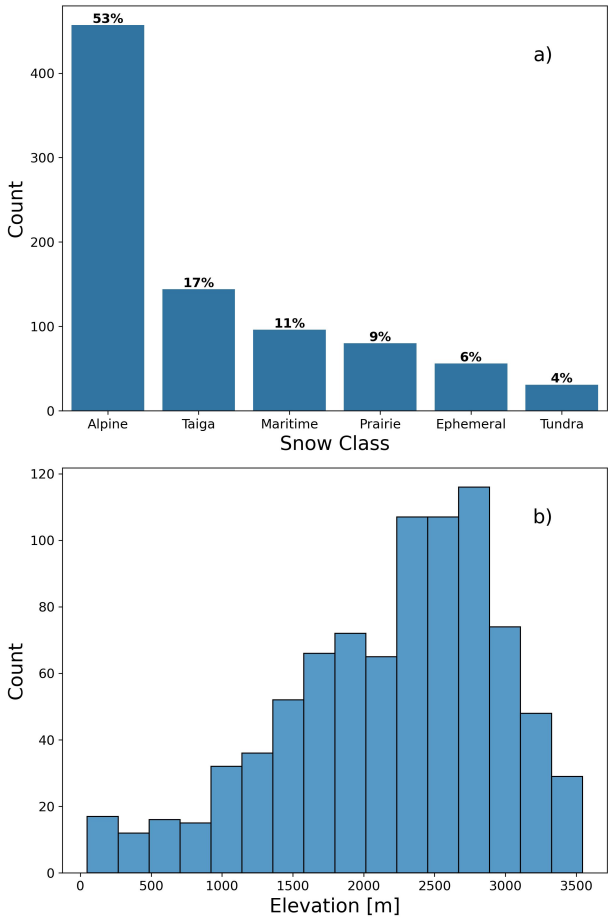


FIG. 4. Distribution of SNOTEL stations in the study by **(a)** snow climate class, and **(b)** elevation (meters). The alpine class dominates the snow class distribution, followed by taiga and maritime. The stations range from 50 m to 3500 m above sea level.

b. Model Setup

The primary objective of our study is to develop a practical and broadly applicable machine learning (ML) model to estimate ρ_b from h_s, which can then be used to compute SWE. For the model to be effective in operational settings, especially for large-scale snow surveys and remote sensing applications, it should rely on input variables that are easy to obtain with minimal additional effort during h_s measurements. Therefore, we focused on using variables that can be readily recorded or derived from the date and location of h_s observations.

Snow depth (h_s) is the primary input feature in our ML model for reasons discussed in the introduction. In addition to h_s, other factors such as snowpack aging, elevation, and climate

conditions (e.g., temperature, precipitation, and wind) also influence ρ_b. While snowpack aging and elevation can be included directly in the model, incorporating climate factors directly would limit the model's applicability to areas with meteorological stations. To maintain broad applicability, these climatic factors are captured indirectly through the snow climate class (C_s), which acts as a proxy for temperature, precipitation, and wind, based on historical climatology.

Snowpack aging is represented by the day of the water year (DOWY), which tracks the temporal progression of the snow season. DOWY is the number of days since October 1st (the start of the water year). The age of the snow cover influences ρ_b because the snow undergoes continuous metamorphism throughout the season (Venäläinen et al. 2021; Male and Gray 1981). By including DOWY as a model input, the model can adjust for temporal variations in ρ_b and capture the evolving characteristics of the snowpack over time.

Elevation (z) accounts for topographic effects on snowpack characteristics. Elevation impacts atmospheric conditions such as temperature and precipitation types, which in turn affect snow accumulation and metamorphism processes. For instance, ρ_b has been observed to increase with altitude in early winter and decrease with altitude in late winter (Jonas et al. 2009). Elevation data for each SNOTEL site were obtained from the NRCS SNOTEL metadata.

Climatic factors are included indirectly through the snow climate class (C_s). The Seasonal Snow Classification System (SSCCS) (Sturm and Liston 2021) assigns a snow climate class to each location based on climatological variables such as air temperature, precipitation, and wind speed. By using C_s as an input, we account for regional climatic influences on snowpack formation and evolution without relying on direct measurements of climatic variables.

Given that elevation (z), DOWY, and snow climate class (C_s) are easily recorded or derived from the date and location of h_s measurements, they are practical input features for the ML model. Our main model can thus be expressed as:

$$\rho_b = f_{\texttt{MAIN}}(h_s,\ \texttt{DOWY},\ z,\ C_s) \tag{2}$$

where $f_{\texttt{MAIN}}$ is the ML model that estimates ρ_b from h_s (cm), DOWY, z (m), and C_s.

To explore the potential for improved performance, a secondary model was developed that extends the main model to include two-week averages of temperature and precipitation. Including these climatic variables directly allows the model to capture recent weather conditions that are known to

influence ρ_b (Bormann et al. 2013; Zhao et al. 2023; Lee et al. 2024). However, since temperature and precipitation data are only available in locations with weather stations, the secondary model is less broadly applicable than the main model.

We tested different averaging windows to determine the optimal period for capturing the influence of temperature and precipitation. Using correlation analysis and model performance on the tuning dataset, we found that a two-week (14-day) window provided the best results. The secondary model can be expressed as:

$$\rho_b = f_{\texttt{CLIMATE}}(h_s,\ \texttt{DOWY},\ z,\ C_s,\ t_{14},\ p_{14}) \tag{3}$$

where $f_{\texttt{CLIMATE}}$ is the ML model that estimates ρ_b from h_s (cm), DOWY, z (m), C_s, and the two-week averages of temperature (t_{14} (°C)) and precipitation (p_{14} (cm)).

c. Experimental Design

Our experimental design comprised five stages: model selection, data splitting, data preprocessing, model training and optimization, and feature importance analysis.

1) Model Selection

To model the snowpack bulk density, we explored four tree-based ensemble machine learning models: Random Forest (Breiman 2001), Extremely Randomized Trees (Geurts et al. 2006), Light Gradient Boosting Machine (Ke et al. 2017), and Extreme Gradient Boosting (Chen and Guestrin 2016). We focused on these models due to their demonstrated strong performance on tabular datasets. Studies have shown that tree-based ensemble models often outperform deep learning models on tabular data and require fewer hyperparameters tuning iterations to achieve good results (Lundberg et al. 2019; Shwartz-Ziv and Armon 2022; Grinsztajn et al. 2022; Zoppi et al. 2024). Additionally, their lower computational requirements make them suitable for operational deployment, including in remote sensing applications where hardware resources may be limited.

(i) Random Forest (RF) Random Forest constructs a "forest" of regression trees using bootstrapped samples of the training dataset and random subsets of the input features (Breiman 2001). The final output of RF is the average of the predictions from all the trees, which helps reduce overfitting and improves performance on new (e.g., test) datasets. Overfitting occurs when a model

performs well on the training data but poorly on unseen data. RF is well-established in the geoscience literature and has been extensively applied in snow science studies (Havens et al. 2012; Tinkham et al. 2014; Gislason et al. 2006, 2004; Lary et al. 2016).

(ii) Extremely Randomized Trees (Extra Trees) Extra Trees is an extension of RF that introduces additional randomness into the tree-building process (Geurts et al. 2006). Unlike RF, which selects the best split among a random subset of features, Extra Trees selects random thresholds for each feature at each node of the regression trees. This means that the features and the split thresholds are chosen randomly. This additional randomness reduces variance and makes Extra Trees computationally efficient and faster to train than RF. Extra Trees have shown promising results in snow science and hydrologic applications (Yang et al. 2024; Galelli and Castelletti 2013).

(iii) Light Gradient Boosting Machine (LightGBM) LightGBM is a gradient boosting framework that builds an ensemble of regression trees sequentially, with each new tree correcting the errors of the previous ones—a process known as boosting (Ke et al. 2017; James et al. 2013). In gradient boosting, the model minimizes the loss function by adding weak learners (regression trees) trained on the residuals of the preceding learners. LightGBM is designed to be efficient and scalable. It uses histogram-based algorithms to bin continuous features into discrete bins, which speeds up computation and reduces memory usage. Finally, LightGBM is "light" because it employs the Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB) techniques to reduce the number of training samples and features, respectively (Kunapuli 2023). These optimizations make LightGBM faster and more memory-efficient, making it ideal for large-scale datasets.

(iv) Extreme Gradient Boosting (XGBoost) XGBoost is another gradient boosting framework that builds an ensemble of regression trees sequentially (Chen and Guestrin 2016). XGBoost enhances the standard gradient boosting algorithm by incorporating both first and second-order derivatives (gradients and Hessians) of the loss function to improve model performance. XGBoost is considered "extreme" because it incorporates advanced regularization techniques ($\mathbb{L}_1$ and $\mathbb{L}_2$) to prevent overfitting, feature efficient tree learning, and support parallelizable implementation (Kunapuli 2023). XGBoost's extreme attributes make it scalable, fast to train, and suitable for large-scale tabular datasets. XGBoost and LightGBM are state-of-the-art models that have shown

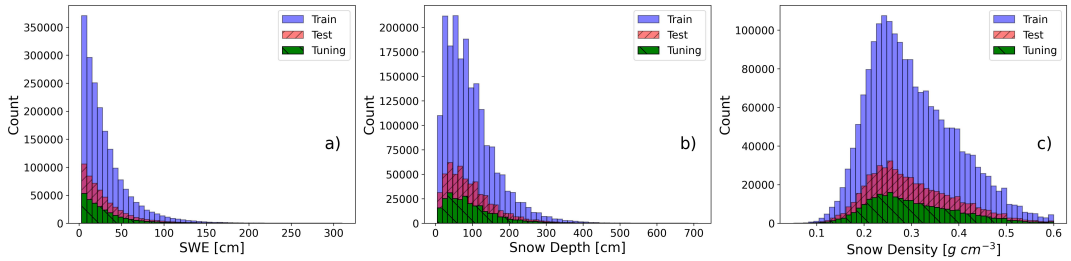


FIG. 5. Distribution of **(a)** SWE, **(b)** Snow Depth, and **(c)** Snow Density in the training, tuning, and testing datasets. The histograms show similar distributions across all three sets, indicating successful stratifcation by snow climate class, which can potentially improve model generalization.

promising results in snow science applications (Iban and Bilgilioglu 2023; Zhao et al. 2024; Alabi et al. 2023; Ofekeze et al. 2023).

For further details about these models, readers may refer to (Clarke et al. 2009) and (Kunapuli 2023).

2) Data Splitting

Following the data cleaning steps described in the *Data* section, the dataset was divided into training, tuning, and testing sets. We used a 70/10/20 split, with 70% (1,905,792 samples) of the data reserved for training, 10% (272,257 samples) for tuning, and 20% (544,513 samples) for testing. The training set was used to train the ML models, the tuning set was used to optimize hyperparameters, and the testing set was used to evaluate the model's performance on unseen data. Hyperparameters are parameters that control the learning process of an ML model; they are not learned from the data and thus must be specified prior to model training.

To maintain the proportional representation of each snow class across all data subsets, we used stratified sampling based on C_s. This stratification is crucial to prevent class imbalance, which could bias the model and adversely affect its generalization to underrepresented classes. The distribution of the training, tuning, and test sets across SWE, h_s, and ρ_b is shown in Figure 5.

3) Data Preprocessing

After data splitting, the snow class (C_s) was encoded to convert it into a numerical format that the ML models could process. We compared three encoding methods: one-hot encoding, target encoding (Micci-Barreca 2001), and CatBoost encoding (Dorogush et al. 2018; Prokhorenkova

et al. 2018). One-hot encoding creates a binary column for each category, while target encoding replaces each category with the smoothed average of the target variable (ρ_b) for that category. CatBoost encoding is a variant of target encoding that uses ordered target statistics to mitigate prediction shift casued by target leakage.

For each encoding method, we fitted the encoder on the training data and trained the models using the encoded training set. The tuning set was then transformed using the encoding learned from the training data, and model performance was evaluated on the transformed tuning set. Target encoding provided the best performance on the tuning set and was thus chosen to encode C_s in our analyses. Since tree-based ensemble models are generally insensitive to feature scaling, we did not apply scaling or standardization to the numerical features.

4) Model Training and Optimization

Each ML model has hyperparameters that control its learning process and complexity. Hyperparameter optimization involves selecting the best combination of these hyperparameters to improve model performance. In our study, hyperparameter tuning was conducted using Optuna (Akiba et al. 2019). Optuna is an automatic optimization framework that leverages the Tree-structured Parzen Estimator (TPE) (Bergstra et al. 2011; Watanabe 2023), a Bayesian optimization algorithm, to efficiently search the hyperparameter space. Compared to traditional grid search (e.g., Larochelle et al. (2007)) and random search (Bergstra and Bengio 2012), TPE is more efficient, as it adaptively selects hyperparameters based on performance from previously sampled configurations. This adaptability allows the search to focus on promising regions of the hyperparameter space.

The hyperparameters tuned for each model and their optimal values are listed in Table A1 in Appendix A. For all models, we optimized the squared error loss function. In the case of the gradient boosting models (LightGBM and XGBoost), we used early stopping to prevent overfitting. This is possible because gradient boosting builds ensembles sequentially, and we can halt the addition of trees once the model stops improving. Early stopping was configured to terminate training if performance on the tuning set did not improve for 50 consecutive iterations. We performed 100 trials for each model and selected the Hyperparameter combination that yielded the lowest root mean squared error (RMSE) on the tuning set. The final models were trained using the best hyperparameters on the combined training and tuning sets.

The overall best-performing model was selected based on its performance on the testing set, using three evaluation metrics:

- **Root Mean Square Error (RMSE)**: Measures the average magnitude of error (best value = 0; worst value = $+\infty$).
- **Mean Bias Error (MBE)**: Quantifies the average bias in the model predictions. It indicates whether the model tends to overpredict (positive MBE) or underpredict (negative MBE) the target variable (best value = 0; worst value = $\texttt{abs}(\pm\infty)$).
- **Coefficient of Determination (R^2)**: Measures the proportion of variance in the target variable that is predictable from the input features (worst value = $-\infty$; best value = +1). Note that R^2 can be negative when the model fits the data poorly. For further details on negative R^2 values, refer to Chicco et al. (2021).

The overall best model was subsequently used for feature importance analysis.

5) Feature Importance Analysis

To quantify the influence of each input variable on the snowpack bulk density (ρ_b) estimates, we conducted a feature importance analysis using the overall best-performing model. Two complementary methods were employed: XGBoost's gain-based importance and SHapley Additive exPlanations (SHAP) (Lundberg 2017; Lundberg et al. 2019).

XGBoost's gain importance quantifies the relative importance of each input feature by measuring the improvement in the loss function (i.e., reduction in squared error) when a feature is used in a splitting nodes during tree construction. The gain-based importance for each feature is calculated as the average gain across all splits where the feature is used. The gain importance $\texttt{I}_j$ for feature j is given by:

$$\texttt{I}_j = \frac{\sum_{t=1}^{T} \sum_{n \in N_j^{(t)}} \Delta L_n^{(t)}}{\sum_{t=1}^{T} |N_j^{(t)}|} \tag{4}$$

where:

- T is the total number of trees in the ensemble,

- $N_j^{(t)}$ is the set of nodes where feature j is used for splitting in tree t,
- $|N_j^{(t)}|$ is the number of nodes where feature j is used for splitting in tree t,
- $\Delta L_n^{(t)}$ is the gain (reduction in loss) achieved by splitting node n in tree t.

Features with higher gain importance values are considered more influential. However, gain-based importance can be biased toward features with higher cardinality (Strobl et al. 2007). To address this limitation, we complemented the analysis with SHAP values.

SHAP values are based on Shapley values from cooperative game theory and provide a unified measure of feature importance that is consistent across different models (Lundberg 2017). SHAP values estimates the marginal contribution of each feature to model predictions by considering all possible combinations of input features. The SHAP value for feature j and instance n is given by:

$$\texttt{SHAP}_{j,n} = \sum_{S \subseteq J \setminus \{j\}} \frac{|S|!(|J| - |S| - 1)!}{|J|!} [f(S \cup \{j\}) - f(S)] \tag{5}$$

where:

- J: the set of all features,
- $|J|$: the total number of features,
- S: any subset of features not containing feature j (i.e., $S \subseteq J \setminus \{j\}$),
- $|S|$: the number of features in S,
- $f(S)$: the model's prediction when only features in S are used,
- $f(S \cup \{j\})$: the model's prediction when features in S and j are used.

The global importance is computed using the average absolute SHAP values across all instances in the dataset:

$$\texttt{SHAP}_j = \frac{1}{N} \sum_{n=1}^{N} |\texttt{SHAP}_{j,n}| \tag{6}$$

Where N is the total number of instances in the dataset. Features with higher average absolute SHAP value are considered more influential in the model's predictions. Our modeling workflow is summarized in Figure 6.

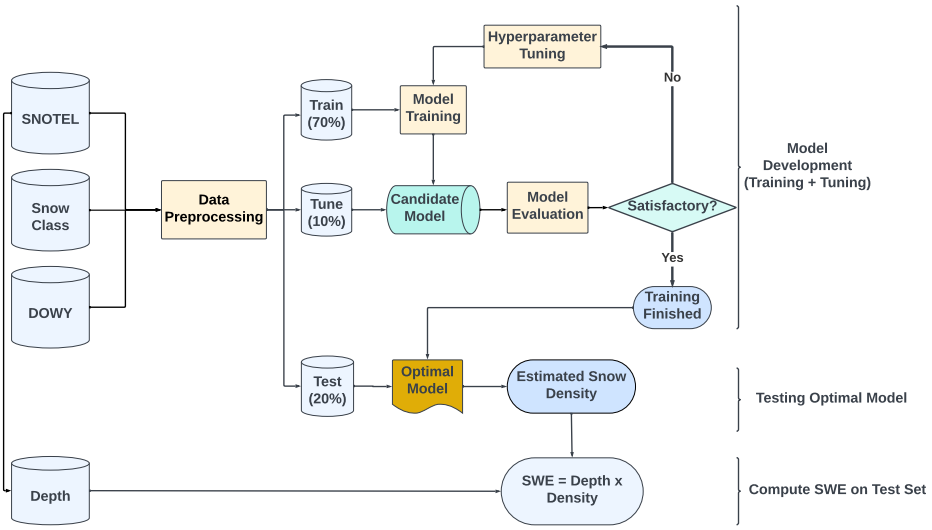


FIG. 6. A flowchart representing the modeling framework for estimating ρ_b and SWE from h_s. The process begins with data preprocessing, followed by model training, hyperparameter tuning, and evaluation. The optimal model is then tested and used to compute SWE on the test set.

3. Results

a. Bulk Density and SWE Estimation

The performance of the four machine learning models—RF, Extra Trees, LightGBM, and XGBoost—was evaluated for predicting snow density and SWE under two model configurations: $f_{\texttt{MAIN}}$ (Equation (2)) and $f_{\texttt{CLIMATE}}$ (Equation (3)). The results are summarized in Table 1.

Under the $f_{\texttt{MAIN}}$ configuration, XGBoost and LightGBM achieved the lowest RMSE of 0.043 g/cm^3 for snow density prediction, indicating superior performance compared to RF and Extra Trees. XGBoost slightly outperformed LightGBM in terms of R^2, explaining 78.3% of the variance in ρ_b compared to 78% for LightGBM. The mean bias error (MBE) was effectively zero across all models, suggesting no systematic overprediction or underprediction of ρ_b.

For SWE estimation using $f_{\texttt{MAIN}}$, XGBoost attained the lowest RMSE of 4.06 cm, closely followed by LightGBM with an RMSE of 4.1 cm. Both models achieved the highest R^2 value of 0.98, indicating excellent predictive capability and explaining approximately 98% of the variance

TABLE 1. Performance metrics for RF, Extra Trees, LightGBM, and XGBoost under the main ($f_{\texttt{MAIN}}$) and climate-enhanced configurations ($f_{\texttt{CLIMATE}}$) for predicting snow density (ρ_b) and SWE on the held-out testing set. Bold values indicate the best performance for each metric.

Model Configuration	Variable	Metric	RF	Extra Trees	LightGBM	XGBoost
$f_{\texttt{MAIN}}$	ρ_b	RMSE (g/cm^3)	0.044	0.048	**0.043**	**0.043**
		MBE (g/cm^3)	**0**	**0**	**0**	**0**
		R^2	0.771	0.734	0.780	**0.783**
	SWE	RMSE (cm)	4.2	4.611	4.096	**4.057**
		MBE (cm)	0.016	0.014	**0.009**	0.011
		R^2	0.978	0.973	**0.979**	**0.979**
$f_{\texttt{CLIMATE}}$	ρ_b	RMSE (g/cm^3)	0.036	0.038	0.038	**0.033**
		MBE (g/cm^3)	**0**	**0**	**0**	**0**
		R^2	0.845	0.831	0.827	**0.868**
	SWE	RMSE (cm)	3.258	3.417	3.488	**2.932**
		MBE (cm)	0.022	0.022	0.009	**0.005**
		R^2	0.987	0.985	0.985	**0.989**

in SWE. The MBE values were minimal for all models, with LightGBM showing the lowest MBE of 9×10^{-3} cm.

All models showed improved performance under the $f_{\texttt{CLIMATE}}$ configuration, which included two-week averages of temperature and precipitation. XGBoost outperformed the other models, achieving the lowest RMSE of 0.033 g/cm^3 for ρ_b and the highest R^2 of 0.87. The MBE remained effectively zero across all models. For SWE estimation, XGBoost again achieved the best result, with an RMSE of 2.93 cm and an R^2 of 0.99. It also had the lowest MBE of 0.005 cm, indicating minimal bias in the predictions.

Overall, XGBoost consistently demonstrated the best performance among the models across both configurations and target variables, making it the top-performer in this study. Consequently, XGBoost was selected for further analysis, including feature importance evaluation and residual analysis by snow class.

b. Comparison with Statistical Models

To assess the performance of the machine learning approach against established statistical models in snow science literature, we compared our best-performing model, XGBoost, with several state-of-the-art statistical models. Specifically, we included the models developed by Sturm et al. (2010)

TABLE 2. Performance comparison of the XGBoost (top-performing ML) models and state-of-the-art statistical models in predicting ρ_b and SWE on the testing set. Only instances where all models provided valid predictions were considered. Note that $f_{\texttt{CLIMATE}}$ results are provided for reference, as it includes additional climate variables not used by the statistical models.

Model	ρ_b			SWE		
	RMSE (g/cm^3)	MBE (g/cm^3)	R^2	RMSE (cm)	MBE (cm)	R^2
XGBoost ($f_{\texttt{CLIMATE}}$)	0.032	0	0.87	2.812	-0.007	0.989
XGBoost ($f_{\texttt{MAIN}}$)	0.042	0	0.786	3.87	-0.027	0.978
Jonas et al. (2009)	0.068	-0.011	0.432	6.496	-0.193	0.939
Sturm et al. (2010)	0.081	-0.01	0.2	9.41	-0.628	0.872
Pistocchi (2016)	0.095	0.067	-0.121	9.593	6.463	0.867
Hill et al. (2019)	N/A	N/A	N/A	11.157	6.527	0.821

and Jonas et al. (2009), which are widely cited in snow science literature. Additionally, we considered the simpler model proposed by Pistocchi (2016), which assumes an initial snow density that increases linearly over time, and the model by Hill et al. (2019), which predicts SWE directly without estimating snow density. All statistical models are described in Appendix B.

Due to the validity domains of the statistical models, direct comparison required careful selection of test instances. For example, the model by Sturm et al. (2010) is valid only from October to June and does not provide predictions for the Ephemeral class, while the model by Jonas et al. (2009) excludes October and is only valid for altitudes above 2000 m during June and July (see Table B1 in Appendix B). To ensure a fair comparison, we restricted the comparison to the subset of the test data where all models could generate predictions, which constituted approximately 90% (491,106 test instances) of the original test set.

It is important to note that our climate-enhanced model configuration ($f_{\texttt{CLIMATE}}$) incorporates temperature and precipitation variables, which provides it with additional information not available to the statistical models. Consequently, while we report the performance of $f_{\texttt{CLIMATE}}$ for completeness, the primary comparison focuses on the $f_{\texttt{MAIN}}$ configuration, which utilizes a similar set of input variables as the statistical models.

The results in Table 2 show that XGBoost ($f_{\texttt{MAIN}}$) outperformed all statistical models for ρ_b and SWE estimation. The $f_{\texttt{MAIN}}$ model achieved an RMSE of 0.042 g/cm^3 for predicting ρ_b,

outperforming the statistical models by Sturm et al. (2010) and Jonas et al. (2009), which had RMSE values of 0.081 g/cm^3 and 0.068 g/cm^3, respectively.

$f_{\texttt{MAIN}}$ also achieved the highest R^2 value of 0.79. In contrast, the statistical models by Sturm et al. (2010) and Jonas et al. (2009) had lower R^2 values of 0.2 and 0.43, respectively. The MBE for $f_{\texttt{MAIN}}$ was effectively zero, suggesting minimal bias in the predictions. For SWE estimation, $f_{\texttt{MAIN}}$ achieved an RMSE of 3.87 cm and an R^2 of 0.98, outperforming the Sturm model (RMSE=9.41 cm, R^2=0.87) and the Jonas model (RMSE=6.50 cm, R^2=0.94). The models by Pistocchi (2016) and Hill et al. (2019) showed higher RMSE values and lower R^2 values, indicating less accurate predictions.

While the XGBoost model configured with climate variables ($f_{\texttt{CLIMATE}}$) demonstrated even better performance (RMSE of 0.032 g/cm^3 for ρ_b and 2.8 cm for SWE), its use of climatology renders it not directly comparable to the statistical models. Nonetheless, these results underscore the potential benefits of integrating climatological variables into predictive models for ρ_b. The model predictions versus the observed values and their corresponding residual distribution for ρ_b and SWE are visualized in Figures 7 and 8, respectively.

c. Model Evaluation by Snow Climate Class

To evaluate the performance of the XGBoost and statistical models across different snow climate classes, we stratified the test data by snow class and computed the RMSE, MBE, and R^2 for ρ_b and SWE without retraining the model. The results are summarized in Table 3 and Figure 9. All results are reported on the testing set, considering only instances where all models provided valid predictions. While the $f_{\texttt{CLIMATE}}$ configuration was evaluated and generally yielded superior performance, our focus here remains on the $f_{\texttt{MAIN}}$ configuration to provide a fair comparison with the statistical models, which also lack climate-specific input features.

Across all snow classes (Table 3), XGBoost ($f_{\texttt{MAIN}}$) achieved lower RMSE values and higher R^2 values compared to the statistical models for both ρ_b and SWE estimation. For instance, in the Alpine class, $f_{\texttt{MAIN}}$ obtained an RMSE of 0.042 g/cm^3 for ρ_b and 3.87 cm for SWE, outperforming the Sturm et al. (2010) and Jonas et al. (2009) models, which had RMSE values of 0.065 g/cm^3 and 0.071 g/cm^3, respectively. The climate-enhanced XGBoost model ($f_{\texttt{CLIMATE}}$) further reduced the RMSE to 0.033 g/cm^3 for ρ_b and 2.617 cm for SWE. In terms of bias, the MBE for XGBoost

models were effectively zero across all snow classes for ρ_b and close to zero for SWE, indicating minimal systematic errors in the predictions. The statistical models, on the other hand, exhibited varying levels of bias, with the Pistocchi (2016) model showing the highest bias in most cases (see Table 3). Similar trends were observed in the other snow classes, with the machine learning models consistently outperforming the statistical models across all metrics. Across all snow climate classes, the inclusion of climatological variables in the $f_{\texttt{CLIMATE}}$ configuration led to improved performance for both ρ_b and SWE estimation.

d. Feature Importance Analysis

Feature importance analysis reveals the relative influence of each input feature on the snow density (ρ_b) predictions for both the $f_{\texttt{MAIN}}$ and $f_{\texttt{CLIMATE}}$ models. Figure 10 displays results based on gain-based importance and SHAP values for each model configuration.

Using both SHAP and gain-based importance, the $f_{\texttt{MAIN}}$ model (Figures 10c and 10d) identifies DOWY, snow climate class, and elevation as the top three most influential features for predicting ρ_b. When climatological variables are incorporated in the $f_{\texttt{CLIMATE}}$ model (Figures 10a and 10b), a shift in feature importance is observed. DOWY remains a key predictor, ranking highest in importance according to SHAP values (Figure 10b) and second according to gain-based importance (Figure 10a). However, the inclusion of climatological variables brought the 14-day lagged average temperature to prominence. This variable emerged as a key predictor, ranking second in importance according to SHAP values (Figure 10b) and fourth according to gain-based importance (Figure 10a). The snow climate class remained an influential feature in the $f_{\texttt{CLIMATE}}$ model, ranking first according to the gain-based metric and fourth according to SHAP. Other variables, such as elevation and 14-day lagged precipitation, also contributed to the model's predictions in the $f_{\texttt{CLIMATE}}$ configuration, albeit to a lesser extent.

4. Discussion

a. Bulk Density and SWE Estimation

In this study, we proposed a machine learning (ML) model that predicts snowpack bulk density (ρ_b) from snow depth (h_s) and other variables that are easily measured alongside h_s. The core strength of this approach is its simplicity. This simplistic approach offers several advantages: it is

easy to deploy operationally, suitable for regions where access is difficult during the snow season, applicable in areas lacking weather stations, and computationally efficient compared to physics-based models that may be too resource-intensive for large basins. Even without climate variables, the ML model ($f_{\texttt{MAIN}}$) achieved an impressive accuracy, estimating SWE with an RMSE value of 3.87 cm (see Table 2). This high level of accuracy is noteworthy given the simplicity of the model and the limited input data required.

The success of this approach can be attributed to the conservative nature of ρ_b as a physical property. The density of snow without water infiltration tends to fall within a predictable range - between 0.05 g/cm^3 for fresh snow and 0.60 g/cm^3 for older snow (Paterson 1994; Sturm et al. 2010). In our raw dataset, 90% of the observed ρ_b values fell between 0.15 g/cm^3 and 0.49 g/cm^3 (data not shown). This relatively narrow range suggests that even using the average ρ_b from the traning set can yield a reasonable SWE estimate.

Evidence of this is seen in the performance of the Pistocchi model (Pistocchi 2016), which, despite achieving a negative R^2 value of -0.12 for ρ_b prediction (see Table 2) on the testing set, still produced a reasonable SWE estimate. A negative R^2 value indicates that the model performs worse than a simple average of the observed ρ_b values in the training set. Yet, it still achieved a SWE RMSE of 9.59 cm (see Table 2), showing that even poor density estimates can lead to reasonable SWE predictions. This result reinforces the idea that ρ_b is a "conservative" variable.

Modeling ρ_b from h_s and then converting it to SWE using the relationship in Equation (1) proved more effective than directly modeling SWE from h_s. The poor performance of the Hill model (Hill et al. 2019), which directly predicts SWE, compared to the Pistocchi model supports this conclusion. The Hill model achieved an RMSE of 11.16 cm for SWE, which is higher than the Pistocchi model's RMSE of 9.59 cm (see Table 2). Despite Pistocchi's poor ρ_b estimates, its density-based approach outperformed Hill's direct SWE estimation. This findings align with the work of Sturm et al. (2010), who arugued that measuring h_s and inferring SWE through density is more appropriate than modeling SWE directly from h_s. The Hill model's poor performance is likely due to the inherent variability in SWE, which is influenced by factors such as snow density, which it doesn't account for.

While the approach of deriving ρ_b from h_s and then estimating SWE is not new—statistical models like those of Sturm et al. (2010) and Jonas et al. (2009) are based on this principle—our

contribution lies in enhancing model performance using machine learning (ML) while retaining simplicity. Our ML model outperformed the statistical models by handling environmental variability (e.g., snow class and elevation) in a continuous rather than discrete manner and by leveraging XGBoost, which is scalable and relatively easy to deploy in practice.

In our experiments, the Jonas model (Jonas et al. 2009) emerged as the best-performing statistical model (see Table 2). These findings align with previous studies (Bonnell et al. 2023; Hill et al. 2019), which similarly identified the Jonas model as outperforming the Sturm model for SWE estimation. However, our ML model without climatological variables reduced the Jonas model's RMSE by approximately 38% and 40% for ρ_b and SWE, respectively (see Table 2). To contextualize this improvement, consider a basin reliant on mountainous snowmelt (e.g., the South Platte River Basin in Colorado), with an average annual flow of 1 million acre-feet. If the Jonas model's RMSE were 5% of this flow (50,000 acre-feet), reducing the RMSE by 40% would translate to an error reduction of 20,000 acre-feet. This volume is sufficient to supply approximately 217,000 people in the United States for a year, based on an average household water usage of 82 gallons per person per day. This level of improvement demonstrates how ML can enhance the density-based SWE modeling approach and improve water management in snow-dependent regions.

The Pistocchi model exhibited the least performance in our evaluation, which is expected due to its inherent limitation. By design, the Pistocchi model cannot estimate ρ_b values less than 0.26 g/cm^3 because of its functional form (see Equation (B3) in appendix B). Consequently, densities below 0.26 g/cm^3 ($\sim$ 39% of the test instances) are overestimated, leading to the positive bias observed (see Table 2). This positive bias can be seen in the scatter plot in Figure 7i and the residual histogram in Figure 7j. Despite these challenges, Pistocchi (2016) described the model as a "first guess" tool, and our findings agree with this characterization.

The improved performance of our ML model has practical implications for remote sensing applications. Radar-based remote sensing technique, such as the interferometric synthetic aperture radar (InSAR) have demonstrated the potential to estimate snow depth (e.g., Marshall et al. (2021)). However, converting h_s to SWE requires reliable ρ_b predictions. With missions like the NASA-ISRO Synthetic Aperture Radar (NISAR) on the horizon, combining modeled ρ_b with remotely-sensed h_s could enable high-resolution SWE estimation across large areas. Furthermore, the NASA Decadal Survey Water Panel targets a SWE error of 10% or less (National Academies of

Sciences, Engineering, and Medicine 2019). Our main ML model met this benchmark, with 61% of our predicted SWE values falling within the ±10% of the observed SWE (data not shown). In comparison, the Jonas model achieved 41%, the Sturm model 31%, the Pistocchi model 19%, and the Hill model 22%. This highlights the promise of ML approaches in delivering accurate SWE measurements to meet water resource management needs.

Finally, the inclusion of climatological variables in our climate-enhanced model ($f_{\texttt{CLIMATE}}$) further improved performance across all metrics (see Table 2), with 74% of predicted SWE values within the 10% threshold. This suggests that incorporating temperature and precipitation data where available can improve model accuracy. Previous studies, such as Bormann et al. (2013), have also identified winter precipitation and air temperature as influential factors in estimating ρ_b. Our findings corroborate these observations.

b. Model Evaluation by Snow Climate Class

The XGBoost models consistently outperformed the statistical models across all metrics and snow climate classes (Table 3, Figure 9). Among the statistical models, the Jonas model generally performed best, except in the Maritime and Ephemeral snow classes, where the Pistocchi model showed comparatively stronger results.

In maritime and ephemeral snow classes, the Pistocchi model was the best among the statistical models for both ρ_b and SWE predictions. Pistocchi (2016) noted that the model "appears robust for regions with snow of Alpine and Maritime characteristics." Our results align with its robustness in Maritime snow but not in Alpine snow, suggesting that the Pistocchi model should be treated as a first-guess tool in specific climates, as recommended by the author.

The Sturm model (Sturm et al. 2010) showed its weakest performance in the Taiga class, with an RMSE of 0.11 g/cm^3 for ρ_b and 14.62 cm for SWE (see Table 3). This poor performance stems from the model's structure in the Taiga class, where it predicts a constant ρ_b value of 0.217 g/cm^3 (see Table B2 and Equation (B2) in Appendix B) regardless of snow depth or DOWY. This lack of variability leads to underperformance, as evidenced by the flat distribution of the Sturm model's predictions for the Taiga class in Figure 9a. However, due to the conservative nature of snow density, the model still achieved a reasonable SWE R^2 of 0.60 (see Table 3).

Notably, all models faced challenges predicting SWE in the Maritime snow class (see Table 3). The Hill model (Hill et al. 2019) exhibited the highest RMSE of 26.16 cm for SWE, while our climate-enhanced ML model ($f_{\texttt{MAIN}}$) achieved the lowest RMSE of 4.45 cm. The XGBoost model with the $f_{\texttt{MAIN}}$ configuration had an RMSE of 6.65 cm. Among the density-based statistical models, RMSE values ranged from 10.55 cm (Pistocchi) to 12.47 cm (Jonas). The reduced performance in the Maritime class may be attributed to the characteristics of maritime snowpacks, which are typically deep and wet (see Table 4 in Sturm and Liston (2021)). Additionally, the warmer temperatures and high precipitation levels in maritime environments introduce complexities that are difficult to capture without incorporating detailed climatological data.

Furthermore, the ML models' predictions exhibited a spread similar to the observed data across all snow classes (see Figure 9). This suggests that the models captured the variability in the observed data more closely than the statistical models, which often showed narrower distributions (e.g., the Sturm model in the Taiga class). The ability to replicate not just the median but also the distribution of observed data is a strong indicator of model reliability.

Finally, stratifying the results by snow class revealed performance differences that may be smoothed out when considering the aggregated results for the entire test set. This class-specific analysis allowed us to identify particular strengths and weaknesses of each model in different snow enivironment. For instance, it highlighted the improved performance of our ML models in challenging classes like Maritime snow, where statistical models struggled. Overall, the ML models show consistent improvement over the statistical models across all snow classes.

c. Feature Importance Analysis

The feature importance results provide valuable insights into the factors influencing ρ_b predictions and how they shift with the inclusion of climatological variables. For the $f_{\texttt{MAIN}}$ model, both the SHAP values and gain-based importance metrics identified DOWY, snow climate class, and elevation as the top three most influential features (see Figures 10c and 10d). This indicates that snowpack age, regional climatic characteristics, and altitude are key determinants of ρ_b.

The prominence of DOWY underscores the substantial impact of age (i.e., seasonal changes) on ρ_b. As the snow season advances, snow undergoes metamorphism processes like compaction, recrystallization, and melt-freeze cycles, that leads to densification. DOWY is a proxy for the tem-

poral progression of these processes, making it a critical predictor. The importance of snow climate class highlights the role of regional climatic conditions in influencing snowpack characteristics. Different snow climates exhibit distinct properties due to variations in atmospheric conditions (e.g.,temperature regimes, precipitation patterns, and wind). By accounting for these differences through the snow climate classification, the model indirectly incorporates regional variability. The influence of elevation emphasizes how altitude affects snow density. Elevation influences temperature and precipitation patterns, which in turn affect snow accumulation and densification rates. Higher elevations typically experience colder temperatures and different precipitation types (e.g., snowfall vs. rain), leading to variations in snowpack structure and density.

The inclusion of climate variables in the $f_{\texttt{CLIMATE}}$ model introduced a shift in the relative influence of predictors (see Figure 10). While DOWY remained a key feature, the 14-day lagged average temperature rose to prominence, ranking second and fourth according to SHAP and gain-based metrics, respectivey (see Figures 10a and 10b). This shift underscores the influence of temperature on snow density, with warmer conditions generally leading to denser snow. The lagged temperature variable captures the cumulative effect of temperature on snowpack properties, reflecting the impact of sustained warm or cold periods on snow density. The snow climate class retained its importance in the $f_{\texttt{CLIMATE}}$ model, indicating the continued importance of regional climatic characteristics. Elevation and 14-day lagged precipitation also contributed to the model's predictions, albeit to a lesser extent. The reduced prominence of elevation in the $f_{\texttt{CLIMATE}}$ model may be due to the temperature variable capturing some of the influence that elevation has on snow density, given that temperature often decreases with increasing altitude. The 14-day lagged precipitation variable, while the least influential (see Figures 10a and 10b), still plays a role in shaping snow density, reflecting the impact of recent precipitation events on snowpack properties.

d. Model Transferability

To evaluate the transferability of our machine learning model to a different snow climate, we tested its performance using data from the Maine Snow Survey. This dataset is from the northeastern United States, a different climatic region than the western United States, where the model was initially trained. The Maine Snow Survey collects snow depth and SWE measurements from December through May. The dataset does not include climatological variables (i.e., temperature

and precipitation), so we focused on models that are universally applicable (i.e., $f_{\texttt{MAIN}}$, Jonas, Sturm, and Pistocchi). Additionally, elevation data, which is a feature in our model and Jonas', was not provided in the survey. We extracted elevation values from the Shuttle Radar Topography Mission (SRTM) dataset using the latitude and longitude of each measurement location. After filtering the dataset to exclude ocean snow classes and erroneous observations (e.g., $\rho_b > 1$ g/cm^3), we obtained a final dataset of 24,235 observations.

Without retraining, the $f_{\texttt{MAIN}}$ model achieved an RMSE of 3.40 cm, an MBE of 0.18 cm, and an R^2 of 0.79 for SWE predictions on the Maine dataset (see Table 4). These results indicate that the model provides promising SWE estimates even when applied to a different snow climate. However, retraining the $f_{\texttt{MAIN}}$ model using 10% of the Maine data (2,423 observations) improved its performance, reducing the RMSE to 2.824 cm and increasing R^2 to 0.852 when tested on the remaining 90% of the dataset (21,812 samples). This finding highlights the importance of retraining to adapt the model to local conditions for improved accuracy. The $f_{\texttt{MAIN}}$ model outperformed the statistical models across all metrics both before and after retraining (see Table 4).

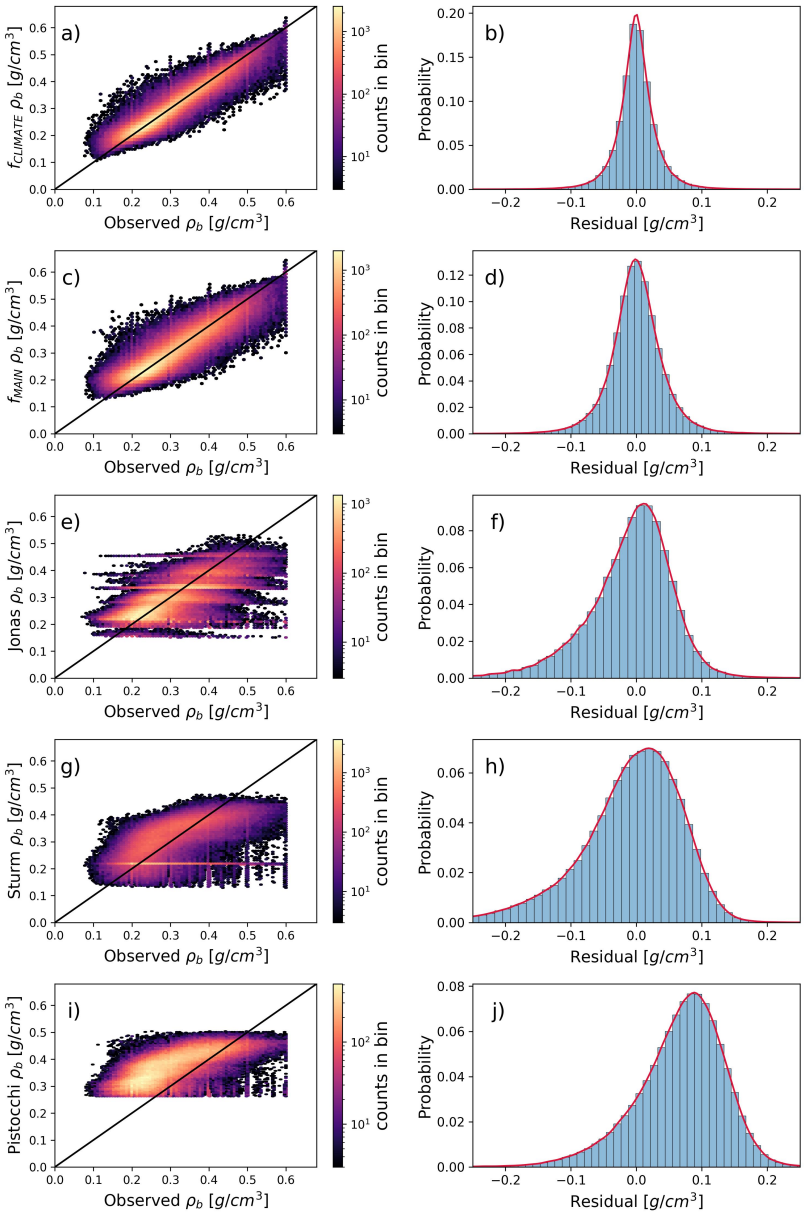
To assess how much data are needed for effective retraining, we conducted a learning curve analysis. A learning curve shows the relationship between training set size and model performance. By holding out 20% of the Maine data as a test set and incrementally training the model on 10%, 20%, up to 80% of the remaining data, we observed that the RMSE marginally decreased from 0.0676 g/cm^3 to 0.0620 g/cm^3 as the training set increased (Figure 11). This suggests that even a small fraction of local data may be sufficient to retrain the model for improved performance. Overall, the results show that while the ML model structure transfers to other snow climate, retraining with local data is recommended for optimal performance.

5. Conclusions

In this study, we developed a machine learning model to estimate snow water equivalent from snow depth observations, requiring no additional measurements beyond variables that can be easily collected alongside snow depth. By focusing on snow depth, day of water year, snow climate class, and elevation, we created a model that predicts snowpack bulk density and subsequently estimates SWE using the relationship in Equation (1). Our model was developed using records from 864 SNOTEL sites across the western United States, and we validated the model's transferability using

data from the Maine Snow Survey. Our model outperformed well-cited existing statistical models based on RMSE, MBE, and R^2 metrics.

The model works well because snow density is a conservative variable constrained within relatively narrow limits, while snow depth varies over a much greater range. As such, much of the variability in SWE can be attributed to snow depth. Including climatological variables further enhanced performance, suggesting that, when available, climate data can improve snow density predictions. Our method is easy to apply and requires only simple input data; however, we recommend retraining the model with local data for optimal performance. The simplicity and scalability of this approach make it suitable for various applications, from localized studies in individual basins to potentially large-scale remote sensing integrations.



527 FIG. 7. Comparison of observed vs. predicted snow density (ρ_b) and residual distributions for the XGBoost
528 models ($f_{\texttt{MAIN}}$ and $f_{\texttt{CLIMATE}}$) and the statistical models (Jonas et al. (2009), Sturm et al. (2010), and Pistocchi
529 (2016)). Panels **(a)**, **(c)**, **(e)**, **(g)**, and **(i)** display hexbin plots of predicted vs. observed snow density. The color
530 scale indicates the count of data points within each bin (minimum count = 3), and the 1:1 line (black diagonal)
531 represents perfect predictions. Panels **(b)**, **(d)**, **(f)**, **(h)**, and **(j)** show the corresponding residual (predicted ρ_b
532 - observed ρ_b) distributions, with histograms overlaid by a kernel density estimate (red line). The ML models
533 generally exhibit higher counts along the 1:1 line and more centralized residuals, indicating better accuracy
534 compared to the statistical models. The statistical models show a negative skewness in their residuals, suggesting
535 systematic bias in the predictions.

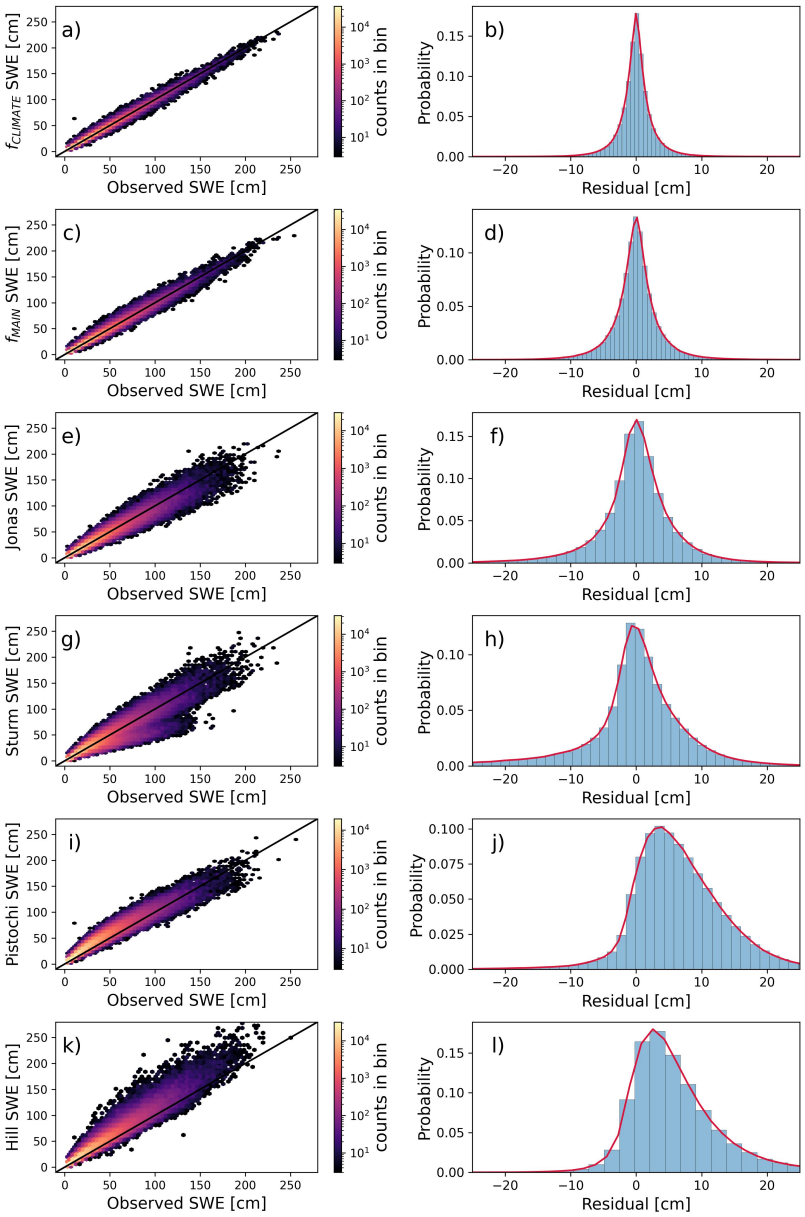
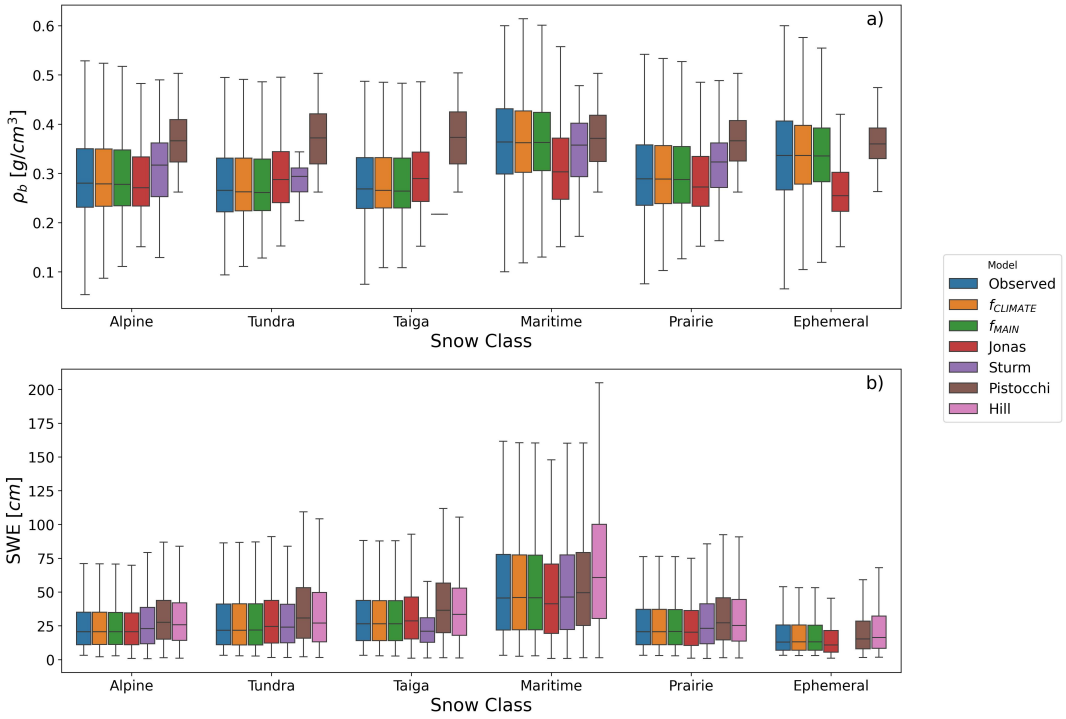


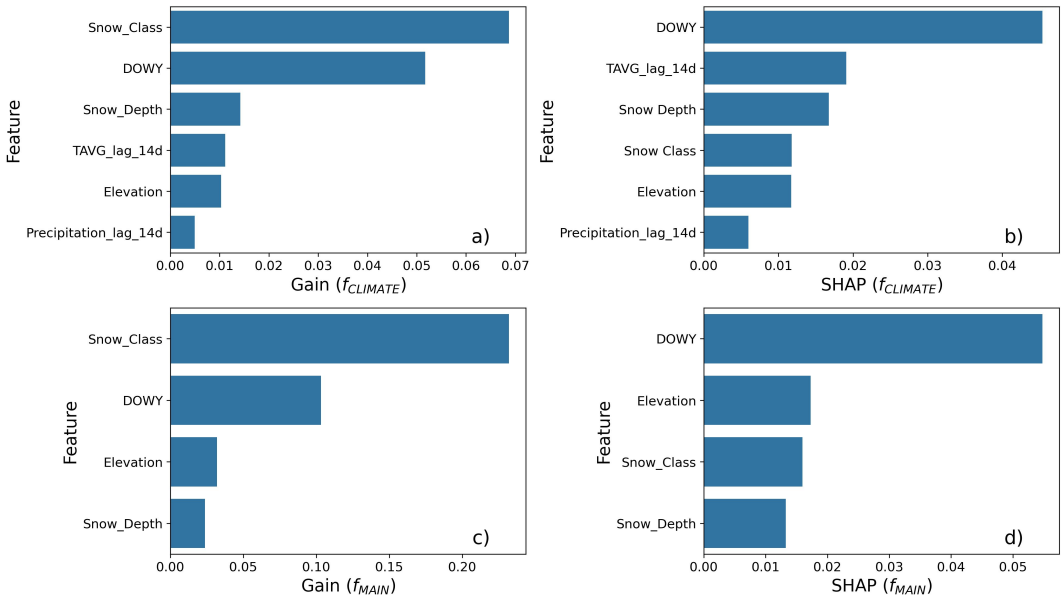
FIG. 8. Hexbin plots and residual histograms showing the performance of the XGBoost and statistical models for SWE estimation. Panels **(a, c, e, g, i, k)** show the predicted vs. observed SWE (in cm), with hexagonal binning representing the density of data points (minimum count = 3 per bin). Panels **(b, d, f, h, j, l)** show the corresponding residual histograms. The XGBoost models demonstrate tighter clustering around the 1:1 line and a symmetric residual distribution, which indicates minimal systematic bias. The Jonas and Sturm models exhibit slight negative skewness in their residuals, while the Pistocchi and Hill models exhibit heavy negative skewness. Overall, the XGBoost models demonstrate narrower residual distributions, highlighting their superior predictive accuracy compared to the statistical models.

TABLE 3. Performance metrics (RMSE, MBE, and R^2) of XGBoost models and statistical models for predicting ρ_b and SWE across different snow climate classes on the testing set. N/A indicates cases where the statistical models were not designed to provide ρ_b or SWE estimates. The number of observations (n) for each class is indicated. Bold values represent the best performance for each metric.

Snow Class	Variable	Metric	XGBoost ($f_{\texttt{CLIMATE}}$)	XGBoost ($f_{\texttt{MAIN}}$)	Jonas	Strum	Pistocchi	Hill
		RMSE (g/cm^3)	0.033	0.042	0.065	0.071	0.096	N/A
	ρ_b	MBE (g/cm^3)	0	0	-0.012	0.010	0.071	N/A
Alpine		R^2	0.855	0.766	0.437	0.341	-0.220	N/A
(n=268,667)		RMSE (cm)	2.617	3.501	5.675	6.347	8.628	8.787
	SWE	MBE (cm)	-0.001	-0.029	-0.455	1.987	6.036	5.378
		R^2	0.987	0.977	0.940	0.925	0.861	0.856
		RMSE (g/cm^3)	0.028	0.036	0.056	0.106	0.104	N/A
	ρ_b	MBE (g/cm^3)	0	0	0.014	-0.069	0.087	N/A
Taiga		R^2	0.879	0.805	0.511	-0.727	-0.65	N/A
(n=103,726)		RMSE (cm)	2.715	3.657	5.885	14.623	11.619	9.743
	SWE	MBE (cm)	-0.015	-0.028	1.774	-8.769	9.139	6.791
		R^2	0.986	0.974	0.934	0.596	0.745	0.82
		RMSE (g/cm^3)	0.037	0.050	0.094	0.079	0.072	N/A
	ρ_b	MBE (g/cm^3)	0	0	-0.053	-0.021	0.006	N/A
Maritime		R^2	0.847	0.724	0.017	0.313	0.429	N/A
(n=59,363)		RMSE (cm)	4.446	6.648	12.466	11.0	10.553	26.160
	SWE	MBE (cm)	-0.066	-0.068	-4.927	-0.212	1.495	17.146
		R^2	0.989	0.974	0.910	0.930	0.935	0.602
		RMSE (g/cm^3)	0.033	0.042	0.069	0.073	0.092	N/A
	ρ_b	MBE (g/cm^3)	0	0	-0.016	0.015	0.065	N/A
Prairie		R^2	0.863	0.771	0.396	0.317	-0.089	N/A
(n=42,303)		RMSE (cm)	2.740	3.830	6.377	8.027	8.547	9.753
	SWE	MBE (cm)	0.016	-0.018	-0.481	2.834	5.672	5.589
		R^2	0.991	0.982	0.950	0.922	0.911	0.884
		RMSE (g/cm^3)	0.026	0.035	0.060	0.070	0.107	N/A
	ρ_b	MBE (g/cm^3)	0	0	0.015	0.002	0.087	N/A
Tundra		R^2	0.903	0.830	0.500	0.323	-0.593	N/A
(n=17,047)		RMSE (cm)	2.578	3.642	6.162	8.350	10.613	8.692
	SWE	MBE (cm)	0.017	0.033	1.316	-0.615	7.830	5.409
		R^2	0.99	0.981	0.944	0.898	0.835	0.889
		RMSE (g/cm^3)	0.047	0.063	0.113	N/A	0.087	N/A
	ρ_b	MBE (g/cm^3)	0.001	0	-0.069	N/A	0.023	N/A
Ephemeral		R^2	0.767	0.583	-0.355	N/A	0.194	N/A
(n=16,662)		RMSE (cm)	2.251	3.105	5.671	N/A	5.004	8.671
	SWE	MBE (cm)	-0.022	-0.049	-3.044	N/A	1.811	4.584
		R^2	0.983	0.968	0.892	N/A	0.916	0.748



575 FIG. 9. Boxplots showing the distribution of observed and predicted **(a)** snow density (ρ_b) and **(b)** SWE across
576 different snow climate classes. Predictions from the main ML model ($f_{\texttt{MAIN}}$), the climate-enhanced ML model
577 ($f_{\texttt{CLIMATE}}$), and the statistical models (Jonas et al. (2009), Sturm et al. (2010), Pistocchi (2016), and Hill et al.
578 (2019)) are compared with observed values. The Sturm model does not provide ρ_b estimates for the Ephemeral
579 class, resulting in the absence of boxplots for this class in both **(a)** and **(b)**. Additionally, the Sturm model predicts
580 a constant ρ_b value of 0.217 for the Taiga class (see Table B2 in Appendix B), which leads to a flat boxplot for
581 the Taiga class in **(a)**. Note that all outliers have been removed for clarity.



586 FIG. 10. Feature importance for the XGBoost models in predicting snow density (ρ_b) based on two methods:
587 gain-based importance (left column) and SHAP values (right column). Panels **(a)** and **(b)** represent the climate-
588 enhanced model ($f_{\texttt{CLIMATE}}$), while panels **(c)** and **(d)** show results for the main model ($f_{\texttt{MAIN}}$).

774 TABLE 4. Performance metrics of the $f_{\texttt{MAIN}}$ model and statistical models on the Maine Snow Survey dataset
775 before and after retraining. The metrics reported before retraining are based on applying the $f_{\texttt{MAIN}}$ model, trained
776 on the western United States SNOTEL data, to 100% of the Maine dataset without any retraining. The metrics
777 reported after retraining are based on retraining the $f_{\texttt{MAIN}}$ model using 10% of the Maine data (2,423 observations)
778 and testing on the remaining 90% (21,812 observations). Values in bold fonts are the best for each metric.

Model	SWE (Before Retraining)			SWE (After Retraining)		
	RMSE (cm)	MBE (cm)	R^2	RMSE (cm)	MBE (cm)	R^2
XGBoost ($f_{\texttt{MAIN}}$)	**3.396**	**0.180**	**0.786**	**2.824**	**-0.097**	**0.852**
Jonas et al. (2009)	4.498	2.492	0.625	4.487	2.478	0.626
Sturm et al. (2010)	4.790	2.751	0.545	4.776	2.735	0.577
Pistocchi (2016)	7.433	5.617	-0.024	7.414	5.599	-0.020

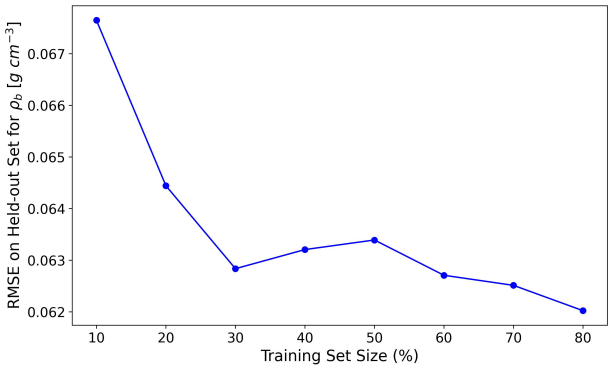


FIG. 11. Learning curve showing the relationship between training set size and model performance (RMSE) for snow density (ρ_b) predictions on the held-out test set using the Maine Snow Survey dataset. The RMSE decreases as the training set size increases, demonstrating the potential for improved model performance with additional training data.

Acknowledgments. We would like to thank the USDA NRCS for providing the SNOTEL data. We would also like to thank M3Works for their `metloom` package, which we used to download the SNOTEL data. We appreciate the Maine Geological Survey and the United States Geological Survey for providing the Maine Snow Survey data. Additionally, we acknowledge the creators of the `srtm.py` package for their open-source tool, which we used to obtain the SRTM elevation data.

This work was funded by the U.S. Army CRREL under the project titled "Advancement of Snow Monitoring for Water Resources, Vehicle Mobility, and Hazard Mitigation: Using Optical, Microwave, Acoustic, and Seismic Techniques" (BAA W913E520C0017).

Data availability statement. All data and code used in this study are available in the publicly accessible GitHub repository: `https://github.com/Ibrahim-Ola/snotel-density-project`. This repository includes the full codebase, datasets, and links to any external data sources used in the analysis. Users can access, download, and use the materials provided under the repository's license.

APPENDIX A

Hyperparameter Tuning Details

The description of the tuned hyperparameters (see Table A1) are as follows:

- **Number of Estimators**: The number of trees in the ensemble.
- **Maximum Depth**: The maximum depth of each tree in the ensemble.
- **Learning Rate**: The step size shrinkage at each iteration of the gradient boosting algorithm to prevent overfitting.
- **Minimum Samples Split**: The minimum number of samples required to split an internal node.
- **Minimum Child Weight**: The minimum sum of instance weight needed in a child.
- **Minimum Samples Leaf**: The minimum number of samples required to be at a leaf node.
- **Maximum Bin**: Maximum number of bins for discretizing continuous features.

APPENDIX B

Details of the Statistical Models

a. Jonas Model

The Jonas model (Jonas et al. 2009) estimates snowpack bulk density (ρ_b) as a function of snow depth (h_s) using Equation (B1). Jonas et al. (2009) provided regression coefficients (b, a) for different months of the year and elevation ranges, as shown in Table B1. The authors also provided offset values specific to regions in the Swiss Alps, which are not included in Table B1. Our analysis did not use these offset values because the model was applied to a different region (western and northeastern United States).

$$\rho_b = a \times h_s + b \tag{B1}$$

where a and b are in kg/m^3 and h_s is in meters.

b. Sturm Model

The Sturm model (Sturm et al. 2010) estimates ρ_b as a function of h_s and day of water year (DOWY) using Equation (B2). The authors provided coefficients (ρ_0, $\rho_{\max}$, k_1, k_2) specific to different snow classes, as shown in Table B2. For the Sturm's model, `DOWY` runs from -92 (October 1) to +181 (June 30).

$$\rho_b = (\rho_{\max} - \rho_0) \times \left[1 - e^{(-k_1 \times h_s - k_2 \times \texttt{DOWY})}\right] + \rho_0 \tag{B2}$$

c. Pistocchi Model

The Pistocchi model (Pistocchi 2016) is a simple, time-based regression model that estimates ρ_b as a linear function of DOWY using Equation (B3). The model assumes that ρ_b starts at a base value ($\rho_0 = 200$ kg/m^3) on November 1 (DOWY = 1), and increases linearly by $K = 1$ kg/m^3 over the water year.

$$\rho_b = \rho_0 + K \times (\texttt{DOWY} + 61) \tag{B3}$$

d. Hill Model

Hill et al. (2019) proposed three models - one-equation, two-equation, and multivariable two-equation models - that estimate SWE from h_s and climatological variables. In this work, we used the best-performing (multivariable two-equation) model as seen in Table 2 of Hill et al. (2019). The multivariable two-equation model estimates SWE as a function of accumulated and ablation SWE using Equation (B4). The accumulated and ablation SWE are calculated using Equations (B5) and (B6). The model assumes that the snowpack reaches its maximum SWE on $\texttt{DOWY}^\star = 180$.

$$\begin{aligned} \text{SWE} = \text{SWE}_{\text{acc}} \times \frac{1}{2}\left(1 - \tanh\left[0.01 \times \left\{\texttt{DOWY} - \texttt{DOWY}^\star\right\}\right]\right) \\ + \ \text{SWE}_{\text{abl}} \times \frac{1}{2}\left(1 + \tanh\left[0.01 \times \left\{\texttt{DOWY} - \texttt{DOWY}^\star\right\}\right]\right) \end{aligned} \tag{B4}$$

where SWE_{acc} and SWE_{abl} are the accumulated and ablation SWE, respectively, and $\texttt{DOWY}^\star$ is the day of water year when the snowpack reaches its maximum SWE. SWE_{acc} and SWE_{abl} are calculated as follows:

$$\text{SWE}_{\text{acc}} = A h_s^{a_1} \times \text{PPTWT}^{a_2} \times \text{TD}^{a_3} \times \texttt{DOWY}^{a_4}; \quad \texttt{DOWY} < \texttt{DOWY}^\star \tag{B5}$$

$$\text{SWE}_{\text{abl}} = B h_s^{b_1} \times \text{PPTWT}^{b_2} \times \text{TD}^{b_3} \times \texttt{DOWY}^{b_4}; \quad \texttt{DOWY} \geq \texttt{DOWY}^\star \tag{B6}$$

Where A, a_1, a_2, a_3, a_4, B, b_1, b_2, b_3, and b_4 are the regression coefficients, PPTWT is is the winter (sum of December, January, February) precipitation (mm), and TD is the difference between the mean temperature of the warmest month and the mean temperature of the coldest month (°C). The regression coefficients were given as (A, a_1, a_2, a_3, a_4) = (0.0533, 0.9480, 0.1701, -0.1314, 0.2922) and (B, b_1, b_2, b_3, b_4) = (0.0481, 1.0395, 0.1699, -0.0461, 0.1804).

References

Akiba, T., S. Sano, T. Yanase, T. Ohta, and M. Koyama, 2019: Optuna: A next-generation hyperparameter optimization framework. *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2623–2631, https://doi.org/0.1145/3292500.3330701.

Alabi, I. O., H.-P. Marshall, J. Mead, E. Trujillo, and E. Ofekeze, 2023: Harnessing l-band insar and lidar data through machine learning for accurate snow depth estimation in grand mesa, colorado. *AGU Fall Meeting Abstracts*, Vol. 2023, C51G–102.

Anderson, E. A., 1973: *National Weather Service river forecast system: Snow accumulation and ablation model*, Vol. 17. US Department of Commerce, National Oceanic and Atmospheric Administration, National Weather Service.

Avanzi, F., C. De Michele, and A. Ghezzi, 2015: On the performances of empirical regressions for the estimation of bulk snow density. *Geografia Fisica e Dinamica Quaternaria*, **38 (2)**, 105–112, https://doi.org/10.4461/GFDQ.2015.38.10.

Bartelt, P., and M. Lehning, 2002: A physical snowpack model for the swiss avalanche warning: Part i: numerical model. *Cold Regions Science and Technology*, **35 (3)**, 123–145, https://doi.org/10.1016/S0165-232X(02)00074-5.

Beaumont, R. T., 1965: Mt. Hood pressure pillow snow gage. *Journal of Applied Meteorology and Climatology*, **4 (5)**, 626–631, https://doi.org/10.1175/1520-0450(1965)004⟨0626:MHPPSG⟩2.0.CO;2.

Bergstra, J., R. Bardenet, Y. Bengio, and B. Kégl, 2011: Algorithms for hyper-parameter optimization. *Advances in Neural Information Processing Systems*, J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Weinberger, Eds., Curran Associates, Inc., Vol. 24, URL https://proceedings.neurips.cc/paper_files/paper/2011/file/86e8f7ab32cfd12577bc2619bc635690-Paper.pdf.

Bergstra, J., and Y. Bengio, 2012: Random search for hyper-parameter optimization. *Journal of machine learning research*, **13 (2)**, https://doi.org/10.1145/2188385.2188395.

Bonnell, R., and Coauthors, 2023: Snowpack relative permittivity and density derived from near-coincident lidar and ground-penetrating radar. *Hydrological Processes*, **37 (10)**, e14 996, https://doi.org/10.1002/hyp.14996.

Bormann, K. J., S. Westra, J. P. Evans, and M. F. McCabe, 2013: Spatial and temporal variability in seasonal snow density. *Journal of Hydrology*, **484**, 63–73, https://doi.org/10.1016/j.jhydrol.2013.01.032.

Breiman, L., 2001: Random forests. *Machine learning*, **45**, 5–32, https://doi.org/10.1023/A:1010933404324.

Brown, R., C. Derksen, and L. Wang, 2010: A multi-data set analysis of variability and change in Arctic spring snow cover extent, 1967–2008. *Journal of Geophysical Research: Atmospheres*, **115 (D16)**, https://doi.org/10.1029/2010JD013975.

Brown, R. D., 2000: Northern Hemisphere snow cover variability and change, 1915–97. *Journal of climate*, **13 (13)**, 2339–2355, https://doi.org/10.1175/1520-0442(2000)013⟨2339:NHSCVA⟩2.0.CO;2.

Brown, R. D., B. Brasnett, and D. Robinson, 2003: Gridded North American monthly snow depth and snow water equivalent for GCM evaluation. *Atmosphere-Ocean*, **41 (1)**, 1–14, https://doi.org/10.3137/ao.410101.

Brown, R. D., and D. A. Robinson, 2011: Northern Hemisphere spring snow cover variability and change over 1922–2010 including an assessment of uncertainty. *The Cryosphere*, **5 (1)**, 219–229, https://doi.org/10.5194/tc-5-219-2011.

Brun, E., P. David, M. Sudul, and G. Brunot, 1992: A numerical model to simulate snow-cover stratigraphy for operational avalanche forecasting. *Journal of Glaciology*, **38 (128)**, 13–22, https://doi.org/10.3189/S0022143000009552.

Brun, E., E. Martin, V. Simon, C. Gendre, and C. Coleou, 1989: An energy and mass model of snow cover suitable for operational avalanche forecasting. *Journal of glaciology*, **35 (121)**, 333–342, https://doi.org/10.3189/S0022143000009254.

Chen, T., and C. Guestrin, 2016: Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785–794, https://doi.org/10.1145/2939672.2939785.

Chen, X., S. Liang, and Y. Cao, 2016: Satellite observed changes in the Northern Hemisphere snow cover phenology and the associated radiative forcing and feedback between 1982 and 2013. *Environmental Research Letters*, **11 (8)**, 84 002, https://doi.org/10.1088/1748-9326/11/8/084002.

Chicco, D., M. J. Warrens, and G. Jurman, 2021: The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. *Peerj computer science*, **7**, e623, https://doi.org/10.7717/peerj-cs.623.

Church, J. E., 1933: Snow surveying: its principles and possibilities. *Geographical Review*, **23 (4)**, 529–563, https://doi.org/10.2307/209242.

Clarke, B., E. Fokoue, and H. H. Zhang, 2009: *Principles and theory for data mining and machine learning*. Springer Science & Business Media, 792 pp., https://doi.org/10.1007/978-0-387-98135-2.

De Michele, C., F. Avanzi, A. Ghezzi, and C. Jommi, 2013: Investigating the dynamics of bulk snow density in dry and wet conditions using a one-dimensional model. *The Cryosphere*, **7 (2)**, 433–444, https://doi.org/10.5194/tc-7-433-2013.

Deems, J. S., T. H. Painter, and D. C. Finnegan, 2013: Lidar measurement of snow depth: a review. *Journal of Glaciology*, **59 (215)**, 467–479, https://doi.org/10.3189/2013JoG12J154.

Dickinson, W. T., and H. R. Whiteley, 1972: A sampling scheme for shallow snowpacks. *Hydrological Sciences Journal*, **17 (3)**, 247–258, https://doi.org/10.1080/02626667209493832.

Dorogush, A. V., V. Ershov, and A. Gulin, 2018: Catboost: gradient boosting with categorical features support. *arXiv preprint arXiv:1810.11363*, https://doi.org/10.48550/arXiv.1810.11363.

Dressler, K., S. Fassnacht, and R. Bales, 2006: A comparison of snow telemetry and snow course measurements in the colorado river basin. *Journal of hydrometeorology*, **7 (4)**, 705–712, https://doi.org/10.1175/JHM506.1.

Feng, S., and Q. Hu, 2007: Changes in winter snowfall/precipitation ratio in the contiguous united states. *Journal of Geophysical Research: Atmospheres*, **112 (D15)**, https://doi.org/10.1029/2007JD008397.

Galelli, S., and A. Castelletti, 2013: Assessing the predictive capability of randomized tree-based ensembles in streamflow modelling. *Hydrology and Earth System Sciences*, **17 (7)**, 2669–2684, https://doi.org/10.5194/hess-17-2669-2013.

Geurts, P., D. Ernst, and L. Wehenkel, 2006: Extremely randomized trees. *Machine learning*, **63**, 3–42, https://doi.org/10.1007/s10994-006-6226-1.

Gislason, P. O., J. A. Benediktsson, and J. R. Sveinsson, 2004: Random forest classification of multisource remote sensing and geographic data. *IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, Vol. 2, 1049–1052, https://doi.org/10.1109/IGARSS.2004.1368591.

Gislason, P. O., J. A. Benediktsson, and J. R. Sveinsson, 2006: Random forests for land cover classification. *Pattern recognition letters*, **27 (4)**, 294–300, https://doi.org/10.1016/j.patrec.2005.08.011.

Golubev, V. N., and A. D. Frolov, 2000: Model of structure and mechanical properties of dry granular snow. *Annals of Glaciology*, **31**, 434–438, https://doi.org/10.3189/172756400781819815.

Grinsztajn, L., E. Oyallon, and G. Varoquaux, 2022: Why do tree-based models still outperform deep learning on typical tabular data? *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, Eds., Curran Associates, Inc., Vol. 35, 507–520, URL https://proceedings.neurips.cc/paper_files/paper/2022/file/0378c7692da36807bdec87ab043cdadc-Paper-Datasets_and_Benchmarks.pdf.

Havens, S., H.-P. Marshall, C. Pielmeier, and K. Elder, 2012: Automatic grain type classification of snow micro penetrometer signals with random forests. *IEEE transactions on geoscience and remote sensing*, **51 (6)**, 3328–3335, https://doi.org/10.1109/TGRS.2012.2220549.

Hedrick, A. R., and Coauthors, 2020: From drought to flood: A water balance analysis of the tuolumne river basin during extreme conditions (2015–2017). *Hydrological Processes*, **34 (11)**, 2560–2574, https://doi.org/10.1002/hyp.13749.

Helms, D., S. E. Phillips, and P. F. Reich, 2008: *The history of snow survey and water supply forecasting: Interviews with US Department of Agriculture pioneers*. 8, US Department of Agriculture, Natural Resources Conservation Service . . . , https://doi.org/10.2737/RMRS-GTR-202.

Hill, D. F., E. A. Burakowski, R. L. Crumley, J. Keon, J. M. Hu, A. A. Arendt, K. Wikstrom Jones, and G. J. Wolken, 2019: Converting snow depth to snow water equivalent using climatological variables. *The Cryosphere*, **13 (7)**, 1767–1784, https://doi.org/10.5194/tc-13-1767-2019.

Iban, M. C., and S. S. Bilgilioglu, 2023: Snow avalanche susceptibility mapping using novel tree-based machine learning algorithms (xgboost, ngboost, and lightgbm) with explainable artificial intelligence (xai) approach. *Stochastic Environmental Research and Risk Assessment*, **37 (6)**, 2243–2270, https://doi.org/10.1007/s00477-023-02392-6.

James, G., D. Witten, T. Hastie, R. Tibshirani, and Coauthors, 2013: *An introduction to statistical learning*, Vol. 112. Springer.

Johnson, J. B., A. B. Gelvin, P. Duvoy, G. L. Schaefer, G. Poole, and G. D. Horton, 2015: Performance characteristics of a new electronic snow water equivalent sensor in different climates. *Hydrological Processes*, **29 (6)**, 1418–1433, https://doi.org/10.1002/hyp.10211.

Jonas, T., C. Marty, and J. Magnusson, 2009: Estimating the snow water equivalent from snow depth measurements in the Swiss Alps. *Journal of Hydrology*, **378 (1-2)**, 161–167, https://doi.org/10.1016/j.jhydrol.2009.09.021.

Judson, A., and N. Doesken, 2000: Density of freshly fallen snow in the central rocky mountains. *Bulletin of the American Meteorological Society*, **81 (7)**, 1577–1588, https://doi.org/10.1175/1520-0477(2000)081⟨1577:DOFFSI⟩2.3.CO;2.

Ke, G., Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, 2017: Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., Curran Associates, Inc., Vol. 30, URL https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf.

Kinar, N., and J. Pomeroy, 2015: Measurement of the physical properties of the snowpack. *Reviews of Geophysics*, **53 (2)**, 481–544, https://doi.org/10.1002/2015RG000481.

Knowles, N., M. D. Dettinger, and D. R. Cayan, 2006: Trends in snowfall versus rainfall in the western united states. *Journal of Climate*, **19 (18)**, 4545–4559, https://doi.org/10.1175/JCLI3850.1.

Kunapuli, G., 2023: *Ensemble methods for machine learning*. Simon and Schuster, 354 pp.

Larochelle, H., D. Erhan, A. Courville, J. Bergstra, and Y. Bengio, 2007: An empirical evaluation of deep architectures on problems with many factors of variation. *Proceedings of the 24th international conference on Machine learning*, 473–480.

Lary, D. J., A. H. Alavi, A. H. Gandomi, and A. L. Walker, 2016: Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, **7 (1)**, 3–10, https://doi.org/10.1016/j.gsf.2015.07.003.

Lee, W. Y., H.-J. Gim, and S. K. Park, 2024: Parameterizations of snow cover, snow albedo and snow density in land surface models: A comparative review. *Asia-Pacific Journal of Atmospheric Sciences*, **60 (2)**, 185–210, https://doi.org/10.1007/s13143-023-00344-2.

Lehning, M., P. Bartelt, B. Brown, C. Fierz, and P. Satyawali, 2002: A physical snowpack model for the swiss avalanche warning: Part ii. snow microstructure. *Cold regions science and technology*, **35 (3)**, 147–167, https://doi.org/10.1016/S0165-232X(02)00073-3.

Lehning, M., I. Völksch, D. Gustafsson, T. A. Nguyen, M. Stähli, and M. Zappa, 2006: Alpine3d: a detailed model of mountain surface processes and its application to snow hydrology. *Hydrological Processes: An International Journal*, **20 (10)**, 2111–2128, https://doi.org/10.1002/hyp.6204.

Liston, G. E., and K. Elder, 2006: A distributed snow-evolution modeling system (snowmodel). *Journal of Hydrometeorology*, **7 (6)**, 1259–1276, https://doi.org/10.1175/JHM548.1.

Liston, G. E., and M. Sturm, 2021: Global seasonal-snow classification, version 1 [data set]. NASA National Snow and Ice Data Center Distributed Active Archive Center., date Accessed: 2024-10-19, https://doi.org/10.5067/99FTCYYYLAQ0.

Lundberg, S., 2017: A unified approach to interpreting model predictions. *arXiv preprint arXiv:1705.07874*, https://doi.org/10.48550/arXiv.1705.07874.

Lundberg, S. M., and Coauthors, 2019: Explainable ai for trees: From local explanations to global understanding. *arXiv preprint arXiv:1905.04610*, https://doi.org/10.48550/arXiv.1905.04610.

Male, D., and D. Gray, 1981: *Handbook of snow: Principles, processes, management & use*. Pergamon Press.

Marshall, H.-P., E. Deeb, R. Forster, C. Vuyovich, K. Elder, C. Hiemstra, and J. Lund, 2021: L-band insar depth retrieval during the nasa snowex 2020 campaign: Grand mesa, colorado. *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, IEEE, 625–627, https://doi.org/10.1109/IGARSS47720.2021.9553852.

McCreight, J. L., and E. E. Small, 2014: Modeling bulk density and snow water equivalent using daily snow depth observations. *The Cryosphere*, **8 (2)**, 521–536, https://doi.org/10.5194/tc-8-521-2014.

Meehan, T. G., and Coauthors, 2024: Spatially distributed snow depth, bulk density, and snow water equivalent from ground-based and airborne sensor integration at grand mesa, colorado, usa. *The Cryosphere*, **18 (7)**, 3253–3276, https://doi.org/10.5194/tc-18-3253-2024.

Micci-Barreca, D., 2001: A preprocessing scheme for high-cardinality categorical attributes in classification and prediction problems. *ACM SIGKDD explorations newsletter*, **3 (1)**, 27–32, https://doi.org/10.1145/507533.507538.

Mote, P. W., A. F. Hamlet, M. P. Clark, and D. P. Lettenmaier, 2005: Declining mountain snowpack in western north america. *Bulletin of the American meteorological Society*, **86 (1)**, 39–50, https://doi.org/10.1175/BAMS-86-1-39.

Najafi, M. R., F. W. Zwiers, and N. P. Gillett, 2016: Attribution of the spring snow cover extent decline in the Northern Hemisphere, Eurasia and North America to anthropogenic influence. *Climatic Change*, **136**, 571–586, https://doi.org/10.1007/s10584-016-1632-2.

National Academies of Sciences, Engineering, and Medicine, 2019: *Thriving on Our Changing Planet: A Decadal Strategy for Earth Observation from Space: An Overview for Decision Makers and the Public*. National Academies Press, Washington, DC, https://doi.org/10.17226/25437.

Ofekeze, E., H.-P. Marshall, J. Mead, E. Trujillo, and I. O. Alabi, 2023: Advancing snow water equivalent estimation with swesarr. *AGU Fall Meeting Abstracts*, Vol. 2023, C51G–1017.

Painter, T. H., and Coauthors, 2016: 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.

Paterson, W. S. B., 1994: *Physics of glaciers*. Butterworth-Heinemann.

Pistocchi, A., 2016: Simple estimation of snow density in an Alpine region. *Journal of Hydrology: Regional Studies*, **6**, 82–89, https://doi.org/10.1016/j.ejrh.2016.03.004.

Pomeroy, J., and D. M. Gray, 1995: Snowcover: Accumulation, relocation, and management. Tech. rep., National Hydrology Research Institute.

Prokhorenkova, L., G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, 2018: Catboost: unbiased boosting with categorical features. *Advances in neural information processing systems*, **31**, https://doi.org/10.48550/arXiv.1706.09516.

Ryan, W. A., N. J. Doesken, and S. R. Fassnacht, 2008: Evaluation of ultrasonic snow depth sensors for us snow measurements. *Journal of Atmospheric and Oceanic Technology*, **25 (5)**, 667–684, https://doi.org/10.1175/2007JTECHA947.1.

Serreze, M. C., M. P. Clark, R. L. Armstrong, D. A. McGinnis, and R. S. Pulwarty, 1999: Characteristics of the western united states snowpack from snowpack telemetry (snotel) data. *Water Resources Research*, **35 (7)**, 2145–2160, https://doi.org/10.1029/1999WR900090.

Shwartz-Ziv, R., and A. Armon, 2022: Tabular data: Deep learning is not all you need. *Information Fusion*, **81**, 84–90, https://doi.org/10.1016/j.inffus.2021.11.011.

Siirila-Woodburn, E. R., and Coauthors, 2021: A low-to-no snow future and its impacts on water resources in the western united states. *Nature Reviews Earth & Environment*, **2 (11)**, 800–819, https://doi.org/10.1038/s43017-021-00219-y.

Stewart, I. T., 2009: Changes in snowpack and snowmelt runoff for key mountain regions. *Hydrological Processes: An International Journal*, **23 (1)**, 78–94, https://doi.org/10.1002/hyp.7128.

Strobl, C., A.-L. Boulesteix, A. Zeileis, and T. Hothorn, 2007: Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC bioinformatics*, **8**, 1–21, https://doi.org/10.1186/1471-2105-8-25.

Sturm, M., and J. Holmgren, 2018: An automatic snow depth probe for field validation campaigns. *Water Resources Research*, **54 (11)**, 9695–9701, https://doi.org/10.1029/2018WR023559.

Sturm, M., J. Holmgren, and G. E. Liston, 1995: A seasonal snow cover classification system for local to global applications. *Journal of climate*, **8 (5)**, 1261–1283, https://doi.org/https://doi.org/10.1175/1520-0442(1995)008⟨1261:ASSCCS⟩2.0.CO;2.

Sturm, M., and G. E. Liston, 2021: Revisiting the global seasonal snow classification: An updated dataset for earth system applications. *Journal of Hydrometeorology*, **22 (11)**, 2917–2938, https://doi.org/10.1175/JHM-D-21-0070.1.

Sturm, M., B. Taras, G. E. Liston, C. Derksen, T. Jonas, and J. Lea, 2010: Estimating snow water equivalent using snow depth data and climate classes. *Journal of Hydrometeorology*, **11 (6)**, 1380–1394, https://doi.org/10.1175/2010JHM1202.1.

Sun, Y., T. Zhang, Y. Liu, W. Zhao, and X. Huang, 2020: Assessing snow phenology over the large part of eurasia using satellite observations from 2000 to 2016. *Remote Sensing*, **12 (12)**, 2060, https://doi.org/10.3390/rs12122060.

Thiebault, K., and S. Young, 2020: Snow cover change and its relationship with land surface temperature and vegetation in northeastern North America from 2000 to 2017. *International Journal of Remote Sensing*, **41 (21)**, 8453–8474, https://doi.org/10.1080/01431161.2020.1779379.

Tinkham, W. T., A. M. Smith, H.-P. Marshall, T. E. Link, M. J. Falkowski, and A. H. Winstral, 2014: Quantifying spatial distribution of snow depth errors from lidar using random forest. *Remote Sensing of Environment*, **141**, 105–115, https://doi.org/10.1016/j.rse.2013.10.021.

Tsang, L., and Coauthors, 2021: Global monitoring of snow water equivalent using high frequency radar remote sensing. *The Cryosphere Discussions*, **2021**, 1–57, https://doi.org/10.5194/tc-16-3531-2022.

Venäläinen, P., K. Luojus, J. Lemmetyinen, J. Pulliainen, M. Moisander, and M. Takala, 2021: Impact of dynamic snow density on globsnow snow water equivalent retrieval accuracy. *The Cryosphere*, **15 (6)**, 2969–2981, https://doi.org/10.5194/tc-15-2969-2021.

Watanabe, S., 2023: Tree-structured parzen estimator: Understanding its algorithm components and their roles for better empirical performance. *arXiv preprint arXiv:2304.11127*, https://doi.org/10.48550/arXiv.2304.11127.

Yang, Q., L. Yue, Y. Peng, Q. Yuan, and Coauthors, 2024: Estimating high-resolution snow depth over the north hemisphere mountains utilizing active microwave backscatter and machine learning. *Journal of Hydrology*, 132203, https://doi.org/10.1016/j.jhydrol.2024.132203.

Young, S. S., 2023: Global and regional snow cover decline: 2000–2022. *Climate*, **11 (8)**, 162, https://doi.org/10.3390/cli11080162.

Young, S. S., and J. S. Young, 2021: Overall warming with reduced seasonality: Temperature change in New England, USA, 1900–2020. *Climate*, **9 (12)**, 176, https://doi.org/10.3390/cli9120176.

Zhao, G.-y., K. Ohsu, H. K. Saputra, T. Okada, J. Suzuki, Y. Kuwahara, and M. Fujita, 2024: Enhancing interpretability of tree-based models for downstream salinity prediction: decomposing feature importance using the shapley additive explanation approach. *Results in Engineering*, 102373, https://doi.org/10.1016/j.rineng.2024.102373.

Zhao, W., C. Mu, L. Han, W. Sun, Y. Sun, and T. Zhang, 2023: Spatial and temporal variability in snow density across the northern hemisphere. *Catena*, **232**, 107 445, https://doi.org/10.1016/j.catena.2023.107445.

Zoppi, T., S. Gazzini, and A. Ceccarelli, 2024: Anomaly-based error and intrusion detection in tabular data: No dnn outperforms tree-based classifiers. *Future Generation Computer Systems*, **160**, 951–965, https://doi.org/10.1016/j.future.2024.06.051.

TABLE A1. Hyperparameter distributions used for tuning each machine learning (ML) model for the two model configurations ($f_{\texttt{MAIN}}$ and $f_{\texttt{CLIMATE}}$), along with the optimal values obtained. The best hyperparameter combinations were selected based on minimizing the RMSE on the tuning set. Note: 'N/A' indicates that the hyperparameter does not apply to that model.

Hyperparameter	Range Searched	Optimal Value
Number of Estimators	**RF**: `Integer Uniform` [50, 400]	**RF**: ($f_{\texttt{MAIN}}$, $f_{\texttt{CLIMATE}}$) = 58
	Extra Trees: `Integer Uniform` [50, 400]	**Extra Trees**: ($f_{\texttt{MAIN}}$, $f_{\texttt{CLIMATE}}$) = 330
	LightGBM: `Integer Uniform` [50, 3000]	**LightGBM**: ($f_{\texttt{MAIN}}$, $f_{\texttt{CLIMATE}}$) = 1500
	XGBoost: `Integer Uniform` [50, 500]	**XGBoost**: ($f_{\texttt{MAIN}}$, $f_{\texttt{CLIMATE}}$) = 300, 250
Maximum Depth	**RF**: `Integer Uniform` [2, 25]	**RF**: ($f_{\texttt{MAIN}}$, $f_{\texttt{CLIMATE}}$) = 24
	Extra Trees: `Integer Uniform` [2, 25]	**Extra Trees**: ($f_{\texttt{MAIN}}$, $f_{\texttt{CLIMATE}}$) = 24
	LightGBM: `Integer Uniform` [2, 25]	**LightGBM**: ($f_{\texttt{MAIN}}$, $f_{\texttt{CLIMATE}}$) = 15, 20
	XGBoost: `Integer Uniform` [2, 25]	**XGBoost**: ($f_{\texttt{MAIN}}$, $f_{\texttt{CLIMATE}}$) = 16, 20
Learning Rate	**RF**: N/A	**RF**: N/A
	Extra Trees: N/A	**Extra Trees**: N/A
	LightGBM: `Uniform` [10^{-4}, 5×10^{-1}]	**LightGBM**: ($f_{\texttt{MAIN}}$, $f_{\texttt{CLIMATE}}$) = 0.15, 0.1
	XGBoost: `Uniform` [10^{-4}, 5×10^{-1}]	**XGBoost**: ($f_{\texttt{MAIN}}$, $f_{\texttt{CLIMATE}}$) = 0.08, 0.1
Minimum Samples Split	**RF**: `Integer Uniform` [2, 20]	**RF**: ($f_{\texttt{MAIN}}$, $f_{\texttt{CLIMATE}}$) = 19
	Extra Trees: `Integer Uniform` [2, 20]	**Extra Trees**: ($f_{\texttt{MAIN}}$, $f_{\texttt{CLIMATE}}$) = 6
	LightGBM: N/A	**LightGBM**: N/A
	XGBoost: N/A	**XGBoost**: N/A
Minimum Child Weight	**RF**: N/A	**RF**: N/A
	Extra Trees: N/A	**Extra Trees**: N/A
	LightGBM: `Uniform` [10^{-3}, 10^{2}]	**LightGBM**: ($f_{\texttt{MAIN}}$, $f_{\texttt{CLIMATE}}$) = 0.1
	XGBoost: `Uniform` [10^{-3}, 10^{2}]	**XGBoost**: ($f_{\texttt{MAIN}}$, $f_{\texttt{CLIMATE}}$) = 9, 7
Minimum Samples Leaf	**RF**: `Integer Uniform` [1, 20]	**RF**: ($f_{\texttt{MAIN}}$, $f_{\texttt{CLIMATE}}$) = 2
	Extra Trees: `Integer Uniform` [1, 20]	**Extra Trees**: ($f_{\texttt{MAIN}}$, $f_{\texttt{CLIMATE}}$) = 3
	LightGBM: N/A	**LightGBM**: N/A
	XGBoost: N/A	**XGBoost**: N/A
Maximum Bin	**RF**: N/A	**RF**: N/A
	Extra Trees: N/A	**Extra Trees**: N/A
	LightGBM: `Uniform` [255, 6000]	**LightGBM**: ($f_{\texttt{MAIN}}$, $f_{\texttt{CLIMATE}}$) = 3500
	XGBoost: `Uniform` [255, 6000]	**XGBoost**: ($f_{\texttt{MAIN}}$, $f_{\texttt{CLIMATE}}$) = 5096, 1200

TABLE B1. Regression coefficients for the Jonas model (Equation B1) as provided in Table 1 of Jonas et al. (2009). N/A indicates that no co-efficient was provided for the month-elevation combination. All coefficients are in kg/m^3.

Month	$z <$ 1400 m asl. $[b,\ a]$	1400 m asl. $\leq z <$ 2000 m asl. $[b,\ a]$	$z \geq$ 2000 m $[b,\ a]$
January	[235, 31]	[208, 47]	[206, 52]
February	[279, 9]	[218, 52]	[217, 46]
March	[333, 3]	[218, 31]	[272, 26]
April	[347, 25]	[354, 15]	[331, 9]
May	[413, 19]	[409, 29]	[378, 21]
June	N/A	N/A	[452, 8]
July	N/A	N/A	[470, 15]
August	N/A	N/A	N/A
September	N/A	N/A	N/A
October	N/A	N/A	N/A
November	[149, 37]	[183, 35]	[206, 47]
December	[201, 26]	[190, 47]	[203, 52]

asl. = above sea level.

TABLE B2. Model coefficients for the Sturm model (Equation B2) as provided in Table 4 of Sturm et al. (2010). All coefficients are in g/cm ^{3}

Snow Class	ρ_{max}	ρ_0	k_1	k_2
Alpine	0.5975	0.2237	0.0012	0.0038
Maritime	0.5979	0.2578	0.0010	0.0038
Prairie	0.5940	0.2332	0.0016	0.0031
Tundra	0.3630	0.2425	0.0029	0.0049
Taiga	0.2170	0.2170	0.0000	0.0000