ADVANCED MACHINE LEARNING APPLICATIONS FOR SNOWPACK AND SWE PREDICTION IN THE ABSAROKA–BEARTOOTH WILDERNESS, MONTANA

by

Mary Kathleen O'Flaherty

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Earth Sciences

MONTANA STATE UNIVERSITY
Bozeman, Montana

December 2025

DEDICATION

This thesis is dedicated to the Montana State University Earth Sciences graduate students who have walked beside me these past few years. Your camaraderie, support, and resilience made every challenge surmountable. This work is in honor of Edward and Marlene Turner, whose constant love and unwavering presence have been with me through every step. I would also like to honor Claudia Albrecht, whose kindness and boundless enthusiasm lit up our department. Her guidance and generosity made a profound impact, and I am deeply thankful to have known her. To my family - my mom, dad, Thomas, Ryan, Sissy and Stanley - thank you for your endless encouragement, love, and belief in me. .

ACKNOWLEDGEMENTS

Andrew K. Laskowski, Anna Schweiger, and Robyn Gotz provided initial edits to this thesis. Funding was provided by the Montana Association of GIS Professionals (MAGIP), the MontanaView Committee, and the Department of Earth Sciences in the College of Letters and Science. I am thankful for lab member Zachary Morrow and office members Natali Kragh, Rebekah Kennedy, and Brandon Botha for additional feedback and discussion.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

ABSTRACT

Snow droughts in the western United States threaten water availability to municipalities and ecosystems, highlighting the need for accurate snowpack monitoring in regions that depend on seasonal snowmelt. SNOTEL and snow course networks provide valuable long-term records, but their sparse coverage in complex alpine terrain limits comprehensive assessments. Remote sensing helps close this gap: the Sentinel-2 mission offers high-resolution imagery for snowpack mapping, and NOAA's High-Resolution Rapid Refresh (HRRR) model supplies meteorological data where ground stations are absent. This thesis develops a Gradient Boosting Machine (GBM) ensemble that integrates Sentinel-2 snow indices, SNOTEL observations, HRRR atmospheric variables, and DEM-derived terrain metrics at a consistent 30-meter resolution. Building on recent methods (Malygin et al., 2024), the model estimates snow water equivalent (SWE) and Snow Depth with low error across the Absaroka–Beartooth Wilderness of Montana. The results of this research enhance our understanding of how machine learning can capture snowpack variability in complex alpine terrain and improve monitoring efforts in regions without expansive monitoring networks.

CHAPTER ONE

INTRODUCTION

The western United States relies on snowmelt for nearly 60% of its water supply, sustaining municipalities, agriculture, and ecosystems (USDA, n.d.). Globally, snowmelt is vital to billions of people and has an economic value estimated in trillions of dollars (Dunmire et al., 2024; Sturm et al., 2017). In recent decades, snow accumulation has declined, with earlier melting and reduced snowpack at lower elevations (Mote et al., 2006, 2018). These shifts are driven by rising temperatures and changing atmospheric circulation, which alter both the timing and magnitude of snowpack processes in the western United States (Hamlet et al., 2005; Mote et al., 2006).

The Snow Telemetry (SNOTEL) network, established in the 1960s by the U.S. Department of Agriculture and operated by the Natural Resources Conservation Service (NRCS), provides real-time measurements of snowpack and meteorological conditions. These records are critical for monitoring snow water equivalent (SWE), snow depth, and related variables that inform water resource management, avalanche forecasting, and climate research (Serreze et al., 1999). Today, approximately 890 SNOTEL sites are monitored across the western United States and Alaska (Perkins et al., 2009; Schaefer & Paetzold, 2000; Strobel et al., 2009).

Although technological advances and expanded coverage have improved the accuracy of SNOTEL measurements, significant limitations remain. In complex alpine terrain, point-based stations are unable to capture the full spatial and temporal variability of snowpack, which restricts their usefulness for large-scale assessments (Serreze et al., 1999; Bormann et al., 2018;

Dozier et al., 2016). As a result, a critical gap is persistent in quantifying snow depth and SWE across topographically diverse landscapes.

High-resolution airborne methods such as LiDAR provide detailed snow depth measurements with exceptional vertical accuracy and help bridge data gaps, but their high cost and limited spatial coverage constrain large-scale use (Largeron et al., 2020). In contrast, satellite-based systems, including MODIS, AVHRR, Landsat, and Sentinel-2, offer scalable and accessible solutions for snow monitoring over broad regions but often lack the spatial and temporal resolution needed to resolve fine-scale variability in mountainous terrain. MODIS (which is stationed on NASA's Terra and Aqua satellites) provides close to daily global coverage (1–2 day revisit) at 250–500 m resolution, which is ideal for suitable for regional assessments but too coarse for steep topographic gradients and small alpine catchments (Hall et al., 2002). The AVHRR sensor, first launched on NOAA's TIROS-N satellite in 1978, offers similar daily coverage but coarser spatial resolution at 1 km, limiting its application in heterogeneous terrain (Robinson & Kukla, 1985). Landsat sensors achieve 30 m spatial detail but have a 16-day revisit time, creating temporal gaps and greater chance of cloud interference (Roy et al., 2014). The Sentinel-2 mission offers an effective alternative, combining 10–20 m spatial resolution with a 5-day revisit, enabling improved detection of snowpack dynamics (Gascoin et al., 2019; Hofmeister et al., 2022). These sensors rely on the spectral contrast between visible and shortwave infrared (SWIR) reflectance to be able to distinguish snow from other bright surfaces on the terrain through the Normalized Difference Snow Index (NDSI) (Dozier, 1989; Hall et al., 2002). NDSI is computed as: NDSI=((Green-SWIR))/((Green+SWIR)).

Where snow exhibits high reflectance in the green spectrum and strong absorption in SWIR wavelengths. The band configurations for NDSI vary slightly across sensors: MODIS (B4, 0.55 μm; B6, 1.64 μm; Hall et al., 2002), AVHRR (Ch1, 0.58–0.68 μm; Ch3A, 1.58–1.64 μm; Robinson & Kukla, 1985), Landsat (0.53–1.65 μm; Roy et al., 2014), and Sentinel-2 (B3, 0.56 μm; B11, 1.61 μm; Gascoin et al., 2019; Hofmeister et al., 2022). Despite recent advancements in the past decade of spatial and spectral resolution of satellite sensors, cloud cover and temporal mismatch of sensors remain key challenges in the ability to capture short-term snow accumulation and ablation events across complex mountain landscapes, which are essential to understanding these processes.

No snow-specific satellite mission currently exists, and existing methods fail to capture the detailed patterns necessary for comprehensive snowpack monitoring (Dozier et al., 2016). To address these challenges, researchers have increasingly utilized remote sensing techniques to monitor snow accumulation in regions between SNOTEL sites. The MODIS sensor has provided significant value to global snow monitoring due to its legacy dataset, radiometric quality, and near-daily temporal coverage (Hall & Riggs, 2007; Riggs et al., 2017). However, with the Terra and Aqua platforms entering decommissioning phases (NASA 2023), and the NASA–ISRO NISAR mission delayed until 2025 (NASA JPL 2024), researchers increasingly view Sentinel-2 as the most capable satellite for high-resolution snowpack mapping (Gascoin et al., 2019; Hofmeister et al., 2022).

Sentinel-2 offers high spatial resolution (20 meters) and a five-day revisit period, which provides an ideal framework that will advance snow monitoring efforts (Gascoin et al., 2019; Hofmeister et al., 2022).

Many prior studies estimating snow water equivalent (SWE) have relied on statistical interpolation or regression approaches that do not explicitly incorporate high-resolution atmospheric data. These methods, including regional regression models (Mizukami et al., 2011) and spatial interpolation techniques such as kriging and inverse distance weighting (e.g., Fassnacht et al., 2003; Venäläinen et al., 2023), were primarily developed to extrapolate ground-based observations rather than represent dynamic meteorological processes. Common statistical frameworks such as kriging, inverse distance weighting, and physiographic regression have been applied to interpolate SNOTEL observations across varying elevations in mountainous terrain (Fassnacht et al., 2003; Zheng et al., 2018). While these methods have contributed to understanding snow distribution patterns, their limited integration of temporally and spatially resolved atmospheric forcing restricts their ability to capture snowpack variability driven by elevation-dependent climate processes.

Additionally, empirical relationships between MODIS-derived fractional snow cover and in-situ SWE have been developed to reconstruct snowpack characteristics and variability across complex landscapes (Schneider et al., 2016; Guan et al., 2013). While these statistical analyses have been integral to improving understanding of snowpack accumulation and distribution, they often neglect the influence of atmospheric variables that have the ability to significantly drive snow processes. Consequently, these methods tend to produce static spatial estimates that lack the temporal and physical consistency achievable through integration with high-resolution atmospheric reanalysis or numerical weather prediction data (Venäläinen et al., 2023; PRISM Climate Group, 2025).

This research addresses these limitations by incorporating NOAA's high-resolution HRRR atmospheric reanalysis data into a machine learning-based framework to improve snowpack estimates. Specifically, this study integrates remote sensing data, SNOTEL observations, and meteorological inputs using a Gradient Boosting Machine (GBM) ensemble approach, a technique shown to improve predictive accuracy for complex environmental variables (Arumugam et al., 2024).

GBMs, including XGBoost, LightGBM, and CatBoost, are ensemble algorithms that iteratively refine predictions by constructing decision trees, with each new tree correcting errors from the previous iteration (Chen & Guestrin, 2016; Dorogush et al., 2018; Ke et al., 2017). Unlike traditional regression models or Random Forest, GBMs prioritize the most important predictors using feature importance scoring, making them particularly effective for modeling complex, nonlinear relationships in environmental data such as snowpack characteristics and behavioral predictions (Breiman, 2001; Friedman, 2001; Arumugam et al., 2024).

In this research, a GBM model is developed to estimate snow water equivalent (SWE) by integrating multiple datasets, including Sentinel-2 imagery, SNOTEL station observations from nine locations across the Absaroka–Beartooth Wilderness, and atmospheric variables from NOAA's High-Resolution Rapid Refresh (HRRR) model, building off of the workflow conducted by Malygin et al., (2024).

Topographic variables such as elevation, slope, and aspect strongly influence snow accumulation and melt dynamics, making them essential for modeling SWE (Dozier, 1989; Fassnacht et al., 2018). These attributes can be derived from DEMs and paired with satellite indices such as the NDSI to improve model accuracy (Hall et al., 2002; Molotch & Bales, 2005).

Traditional workflows in ArcGIS Pro allow these parameters to be extracted and joined, but such approaches are often time-intensive and limit scalability. To address this challenge, we developed a Python-based workflow utilizing open-source code that automates the extraction of slope, aspect, and elevation at a consistent 30-meter resolution. This shift not only ensured congruency with Sentinel-2 imagery but also improved efficiency, making it possible to integrate terrain metrics directly into machine learning models (Gorelick et al., 2017; Zhu & Woodcock, 2014).

Despite the critical role of terrain and satellite-derived indices in snowpack modeling (Dozier, 1989; Fassnacht et al., 2018), few studies have focused on the Absaroka–Beartooth Wilderness, a region with complex topography and limited ground observations (Malygin et al., 2024; Mote et al., 2018). This thesis develops a Python-based workflow that integrates Sentinel-2 NDSI data (Hall et al., 2002), SNOTEL observations (Serreze et al., 1999), HRRR atmospheric variables (Benjamin et al., 2016), and DEM-derived terrain attributes (Molotch & Bales, 2005; Winstral et al., 2002) into an ensemble machine learning framework. The goal is to improve snow water equivalent (SWE) and snow depth estimation and extend predictive capacity into areas lacking ground-truthing measurements..

CHAPTER TWO

STUDY AREA

TThe Absaroka–Beartooth Wilderness in south-central Montana is a high-elevation landscape with the highest elevation of 3,658 m and shaped by glaciers and heavy snowfall, averaging more than 1,000 cm annually (University of Montana, n.d.). Its steep terrain creates strong gradients in elevation, slope, and aspect, making it an ideal natural laboratory for studying how topography influences snowpack accumulation and melt. Long-term monitoring is supported by several SNOTEL stations that measure SWE, snow depth, temperature, and precipitation (Serreze et al., 1999) in the region surrounding the wilderness boundaries. These ground observations are complemented by Sentinel-2A and 2B satellites, which provide multispectral imagery every five days to support high-resolution snowpack mapping (Copernicus/ESA, n.d.; Drusch et al., 2012).

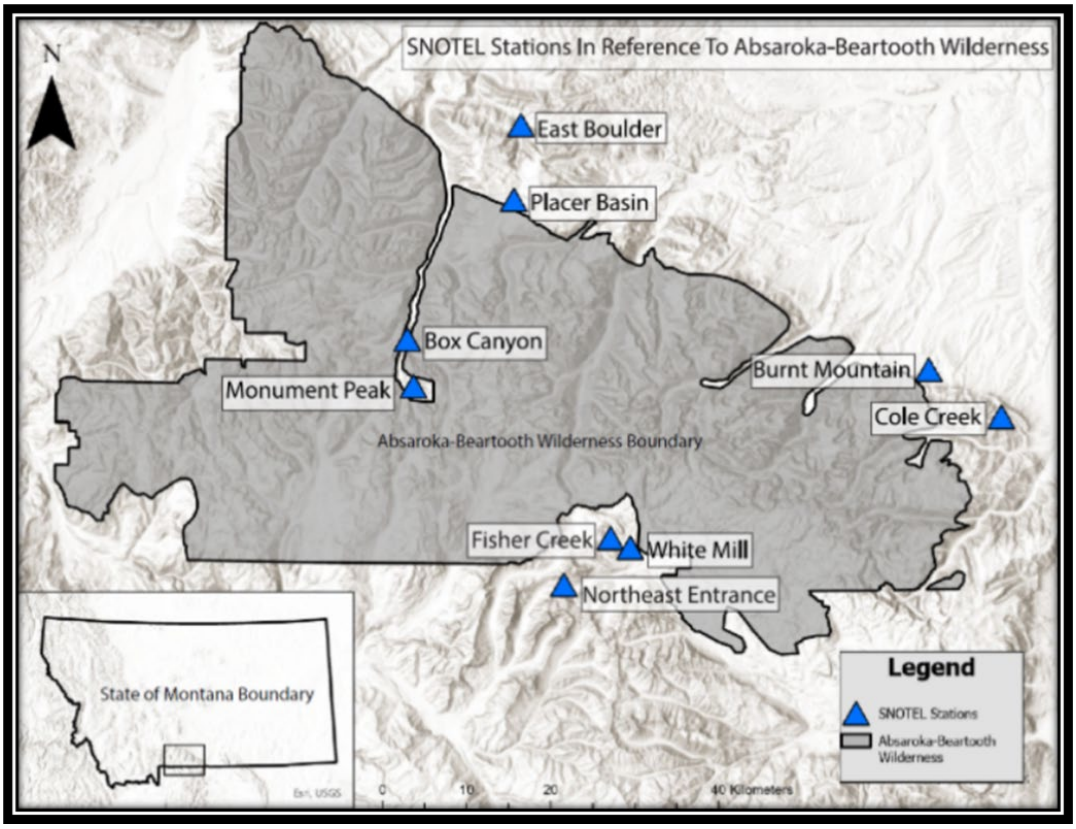

Figure 1. Absaroka–Beartooth Wilderness boundary in relation to SNOTEL locations; Monument Peak, Box Canyon, East Boulder, Placer Basin, Fisher Creek, White Mill, Northeast Entrance, Burnt Mountain and Cole Creek.

Compared to regions such as the Sierra Nevadas, Cascades, and Colorado Rockies, the Absaroka–Beartooth Wilderness has received far less research attention despite its hydrological significance (Malygin et al., 2024; Mote et al., 2018). In the Sierra Nevada, snowpack research has been advanced by decades of monitoring at the Central Sierra Snow Laboratory and by major investments such as NASA's Airborne Snow Observatory (ASO). Airborne LiDAR mapping in basins like the Tuolumne and San Joaquin, coupled with the iSnobal energy balance model, has enabled precise SWE estimation and greatly improved understanding of snow distribution and runoff contributions (Margulis et al., 2016; Painter et al., 2016).

The Cascades and Pacific Northwest also benefit from long-standing snow and hydrology records, with multi-decadal SWE datasets extending back to the mid-20th century. These records have been critical for identifying climate-driven trends, including widespread snowpack decline across the western United States (Luce, 2009; Mote et al., 2018; Pederson et al., 2011). In addition, the Colorado Rockies have been the focus of intensive campaigns, combining ASO, airborne LiDAR, and iSnobal modeling in basins such as the Senator Beck Basin and Grand Mesa (Hedrick et al., 2018). NASA's ongoing SnowEx program has further expanded monitoring in the region through multi-platform field measurements and modeling (Kim et al., 2017).

Montana has fewer long-term snowpack records and less spatial monitoring sites. Some recent work in Yellowstone National Park and by the Montana Snow Survey has expanded its data collection missions, which include adding instrumentation at SNOTEL sites and improving the resolution and reliability of SWE datasets (NRCS, n.d.; Pederson et al., 2011). Many large areas of complex mountainous terrain in the northern Rocky Mountains remain under-sampled due to limited accessibility and the logistical constraints of field-based snow monitoring (Geddes et al., 2005; Moran et al., 2022).

By conducting research in the Absaroka-Beartooth Wilderness, this study extends snowpack research in a comparatively underexplored region. Monitoring advances in other mountain ranges, such as fractional snow cover assimilation with Landsat/Sentinel (Baba et al., 2018; Yang et al., 2022; Gascoin et al., 2020), ASO-guided updates within iSnobal (Hedrick et al., 2018; Painter et al., 2016), and SNOTEL-calibrated SWE reconstructions (Schneider & Molotch, 2016; Bair et al., 2016), provide methodological frameworks that can be adapted to data-sparse alpine settings. Ultimately, the results from this study will help improve SWE and

snowpack estimation in complex alpine terrain and support water resource management and climate adaptation in snow-dependent regions with limited monitoring infrastructure.

CHAPTER THREE

DATA

Ground-based snow observations were retrieved from the USDA NRCS Automated Water Data Base (AWDB) REST API, which provides continuous daily records of snow water equivalent (SWE, WTEQ), snow depth (SNWD), precipitation (PREC), and air temperature (TAVG). The Python-based extraction was adapted from Hayden Libby's SNOTEL data scripts and configured to query by station triplet ID (e.g., "836:MT:SNTL" for Twin Lakes) within a defined date range (2000–2024) (Hayden Libby et al., in prep., 2024). This approach ensured complete temporal coverage across the historical SNOTEL record while maintaining full reproducibility through scripted HTTP (GET) requests.

To maintain consistent spatial alignment with remote sensing and atmospheric datasets, all SNOTEL station coordinates (latitude, longitude, and elevation) were projected to UTM Zone 12N (EPSG:32612). Daily SNOTEL data were used both as training targets for model development and as independent validation datasets. SNOTEL daily values represent conditions recorded at midnight local time. To ensure temporal alignment, HRRR, Sentinel-2, and terrain-derived predictors were sampled or aggregated to the nearest preceding analysis hour so that model inputs correspond to SNOTEL's midnight timestamp. This correction prevents temporal bias that would otherwise arise from mixing morning/afternoon HRRR fields with overnight SNOTEL observations. Gaps shorter than three days were linearly interpolated, while longer gaps were excluded to preserve temporal integrity. Erroneous zero values during known accumulation periods were corrected using adjacent station trends following the quality control procedures outlined in Aragon et al. (2021).

We selected nine SNOTEL stations within and surrounding the Absaroka–Beartooth Wilderness (Montana, USA) based on spatial proximity, data completeness, and elevation representativeness. The SNOTEL network, maintained by the USDA Natural Resources Conservation Service (NRCS), provides high-elevation meteorological and snowpack observations essential for model calibration in mountainous terrain (NRCS, n.d.). To ensure consistency and maximize temporal overlap, daily records were retrieved for the period 2003–2024, which encompasses the full operational era of modern SNOTEL instrumentation.

Station Name	Longitude	Latitude	Elevation
Box Canyon	45.27	-110.25	2033 m
Burnt Mountain	45.24	-109.25	1792 m
Cole Creek*	45.19	-109.35	2393 m
East Boulder	45.5	-110.08	1931 m
Fisher Creek	45.06	-109.94	2774 m
Monument Peak	45.22	-110.24	2697 m
Northeast Entrance	45.01	-110.01	2240 m
Placer Basin	45.42	-110.09	2691 m
White Mill	45.05	-109.91	2652 m

Table 1. SNOTEL locations, names and elevations; *Cole Creek was excluded.

The R-script utilized R packages including httr, glue, and purrr to dynamically construct API calls for the selected variables (SWE, snow depth, precipitation, average temperature, relative humidity, wind speed, and incoming shortwave radiation). Retrieved JSON responses were parsed into tidy tabular formats using dplyr and lubridate, with only 00:00 (midnight) observations retained to ensure calibration consistency. A master sequence of daily timestamps

was generated to synchronize all variables temporally, and missing records were flagged for quality control.

All individual element tables were merged into a single wide-format dataset and exported as .csv files for downstream integration with Sentinel-2 and HRRR datasets in the Python-based ensemble modeling workflow. This automated pipeline produced reproducible, high-quality in-situ datasets suitable for long-term snowpack modeling and validation.

Accurate snowpack modeling requires high-resolution atmospheric variable inputs that capture temperature, humidity, wind, and shortwave radiation fluxes across complex terrain. Previous studies have utilized various atmospheric and at times downscale snowpack variables. Coarser products such as the North American Regional Reanalysis (NARR), ERA5, and MERRA-2, provide valuable long-term climate data at spatial resolutions of 10–30 km, which limit their ability to be used in complex alpine regions (Lundquist et al., 2015; Wrzesien et al., 2019). More advanced regional models such as WRF (Weather Research and Forecasting Model) and NLDAS (North American Land Data Assimilation System) have improved local meteorological variable projections but still require empirical correction or downscaling to accurately resolve near-surface variability in steep mountainous environments (Vionnet et al., 2022; Quéno et al., 2024).

The High-Resolution Rapid Refresh (HRRR) model, developed by NOAA's Earth System Research Laboratory, represents a major advance in both spatial and temporal modeling. Operating at a 3-km grid spacing with hourly outputs, HRRR provides consistent, assimilated atmospheric fields including temperature, humidity, wind, pressure, and shortwave radiation (Blaylock, 2022). The frequent reinitialization and assimilation of radar and satellite data

improve the model's representation of mesoscale dynamics, enabling finer resolution of processes like orographic precipitation, wind channeling, and diurnal snowmelt variability (Dowell et al., 2022).

Data were obtained using the Herbie Python package (Blaylock, 2022), which provides automated access to NOAA's HRRR GRIB2 archives through streamlined URL construction, caching, and integration with libraries, xarray and cfgrib for parsing files into labeled datasets. The open-source framework enabled efficient retrieval of multi-year HRRR data.

The following atmospheric variables were extracted from HRRR to represent key physical drivers of snowpack accumulation and ablation. Near-surface air temperature (TMP2m, °C) and daily average temperature (TAVG, °C) are primary thermal controls on snow accumulation and melt, determining precipitation phase and governing energy exchange at the surface (Barnett et al., 2005; Vionnet et al., 2022). Relative humidity (RH2m, %) modulates sublimation rates and surface energy fluxes through latent heat exchange (Liston & Sturm, 1998; Pomeroy et al., 1998). Snow water equivalent (WEASD, mm) and snow depth (SNOD, mm) represent modeled indicators of snow storage and physical depth, both influenced by micro-meteorological and terrain conditions (Sturm et al., 2010; Mott et al., 2018). Ten-meter wind speed (WSPD10, m s^{-1}) affects snow redistribution, compaction, and sublimation, shaping spatial heterogeneity in snow cover (Winstral & Marks, 2002; Lehning et al., 2008). Downward shortwave radiation (DSWRF_inst, W m^{-2}) provides the dominant energy input for snowmelt through absorption of solar energy, with magnitude and duration governed by aspect and slope orientation (Marks et al., 1999; Dozier & Painter, 2004).

All data were subset to a bounding box encompassing the Absaroka–Beartooth Wilderness (longitude −112° to −108.5°, latitude 44.5° to 46.0°) with a buffer to ensure full coverage of the adjacent SNOTEL stations. Raw GRIB2 outputs were processed in xarray (Hoyer & Hamman, 2017), flattened into a data table format containing latitude, longitude, and UTM coordinates, and joined with terrain and remote-sensing variables. Missing values were interpolated using nearest-neighbor methods implemented in scikit-learn (Pedregosa et al., 2011), ensuring spatial consistency within each daily composite.

All HRRR data were organized by water year (WY), spanning WY2020–WY2024 (October 2019 through September 2024). This approach aligns with snow accumulation and melt cycles typical of hydrologic studies (Stewart, 2009) and allowed for modular processing and model retraining on a yearly basis.

Although HRRR provides hourly, high-resolution atmospheric fields, several constraints remain. The model's limited CONUS domain and substantial data volume make multi-year analyses computationally demanding and constrain long-term data accessibility. Additionally, the 3-km grid spacing can introduce subgrid-scale elevation mismatches, reducing the accuracy of near-surface temperature, precipitation, and radiation estimates in complex mountain terrain (Hoyer & Hamman, 2017). Additional uncertainty arises from precipitation phase partitioning near 0 °C (Vionnet et al., 2022) and potential resampling errors introduced when HRRR data are interpolated to 30 m grids for integration with Sentinel-2 and DEM layers. Consequently, regional and elevation-dependent biases may persist, as model representations of surface energy balance and snowpack processes remain simplified and not fully optimized for steep alpine environments (English et al., 2021; Quéno et al., 2024).

The Sentinel-2 mission, developed by the European Space Agency under the Copernicus program, consists of two polar-orbiting sister satellites, Sentinel-2A (2015) and Sentinel-2B (2017). Both carry a multispectral instrument (MSI) with 13 spectral bands (443–2190 nm) at spatial resolutions of 10 m, 20 m, or 60 m (Copernicus/ESA, n.d.; Drusch et al., 2012). The sister satellite configuration provides a ~5-day revisit at the equator (more frequent at higher latitudes), which makes the satellite well-suited for monitoring dynamic cryosphere processes. Compared to coarser sensors such as MODIS (250–500 m), Sentinel-2's higher spatial resolution allows for the improved characterization of snowpack variability across elevation, slope, and aspect gradients and has been shown to detect finer changes in accumulation and ablation patterns in mountain regions (Gascoin et al., 2019; Hofmeister et al., 2022).

A multitude of optical satellite sensors have been used for snow cover monitoring, and each offer advantages and limitations in terms of spatial and temporal resolution and monitoring. The Moderate Resolution Imaging Spectroradiometer (MODIS) provides daily global snow products at 250–500 m resolution and has maintained a continuous archive since 2000, which makes it valuable for long-term climatological studies. However, its coarse resolution leads to mixed-pixel uncertainties in mountainous terrain and cloud contamination issues (Hall et al., 2002). The Landsat satellite series offers a fine spatial resolution (30 m) and an extensive historical record, but its 16-day revisit time limits the ability to capture rapid snowpack dynamics. Cloud interference also remains a challenge (Roy et al., 2014). The Visible Infrared Imaging Radiometer Suite (VIIRS) provides daily global coverage with 375 m resolution, an improvement from MODIS. However, it still lacks sufficient spatial detail for narrow valleys and small alpine basins (Justice et al., 2013).

Sentinel-2 achieves a balance between spatial and temporal resolution, providing 10 - 20 m multispectral imagery with an approximate five-day revisit time. Its utility for snow accumulation detection in alpine environments has been well validated (Gascoin et al., 2019; Hofmeister et al., 2022). However, challenges related to cloud cover and its comparatively shorter temporal archive remain.

Snow cover was mapped from Sentinel-2 imagery utilizing the Normalized Difference Snow Index (NDSI), defined as NDSI = (Green − SWIR) / (Green + SWIR), where Band 3 (green, 560 nm) and Band 11 (SWIR, 1610 nm) were used. NDSI exploits snow's high reflectance in the visible spectrum and strong absorption in the shortwave infrared to effectively differentiate snow from clouds, vegetation, and bare ground (Hall et al., 2002). Data processing was conducted in Google Earth Engine (GEE) using the Sentinel-2 Surface Reflectance collection (COPERNICUS/S2_SR). The area of interest encompassed the Absaroka–Beartooth Wilderness and surrounding SNOTEL stations. Image scenes were filtered to less than 10 percent cloud cover, with clouds masked using the MSK_CLDPRB band. The Five-day median NDSI composites were then generated to minimize cloud contamination and temporal noise.

All Sentinel-2 imagery was resampled to 30 m spatial resolution to maintain consistency with the DEM-derived terrain layers and to facilitate spatial alignment with the coarser 3 km HRRR atmospheric grids. This approach reduced computational demand while preserving the spatial detail necessary for mountain snow modeling. Final outputs included 30 m GeoTIFF rasters and tabular CSV files containing summary statistics for each date and variable, with dates containing insufficient valid pixels automatically flagged and excluded from further analysis. Missing Sentinel-2 NDSI observations caused by cloud or shadow masking were filled using

temporally interpolated NDSI_filled values derived from adjacent clear-sky composites utilizing a Nearest Pixel Analysis approach. Only pixels with paired Sentinel-2, HRRR, and SNOTEL observations were retained for model training, which resulted in reduced sample counts for early water years.

The processed Sentinel-2 NDSI products were integrated with SNOTEL observations of snow water equivalent (SWE) and snow depth (SNWD), along with HRRR-derived atmospheric variables including temperature, precipitation fields, radiation, humidity, and wind speed. Together, these datasets formed the predictor matrix used to train Gradient Boosting Machine (GBM) models (XGBoost, LightGBM, and CatBoost) for estimating SWE across the Absaroka–Beartooth region.

CHAPTER FOUR

MODEL DESIGN AND RATIONALE

The GBM ensemble was constructed using an equal-weight average of three machine learning algorithms (XGBoost, LightGBM, and CatBoost) each trained on the complete set of model input variables (Figure 2). Inter-model variance among the three algorithms was used to quantify predictive uncertainty and assess consistency across model outputs. These models were selected because they offer complementary strengths for heterogeneous environmental datasets and nonlinear relationships typical of mountain snowpack modeling. XGBoost provides strong predictive accuracy and reliably captures complex interactions among atmospheric, terrain, and spectral variables while incorporating regularization to limit overfitting. LightGBM contributes computational efficiency through its histogram-based tree construction, allowing fast processing of high-resolution raster inputs and multi-year atmospheric data without loss of performance. CatBoost adds robustness to noisy or irregular patterns in the dataset and handles mixed predictor types effectively, making it well-suited for capturing subtle interactions between snow conditions, weather forcing, and topography. Together, these three algorithms form a balanced ensemble capable of modeling spatially variable snow depth and SWE across complex alpine terrain.

Ensemble predictions were generated annually for Water Years (WY) 2020 through 2024 and extended through WY2030 using future HRRR and PRISM forcing. All datasets, including HRRR, Sentinel-2, and SNOTEL observations, were aggregated to daily values, and annual model training used mean conditions for each water year (October 1 to September 30). Because SNOTEL daily reports represent conditions at midnight, HRRR variables were temporally

aligned by selecting the nearest preceding analysis hour, ensuring that model inputs corresponded to SNOTEL's reporting schedule. Model validation employed both spatial (leave-one-station-out) and temporal (leave-one-year-out) cross-validation to assess generalization across stations and interannual variability. These procedures represent internal validation because the same SNOTEL locations and years contribute to training.

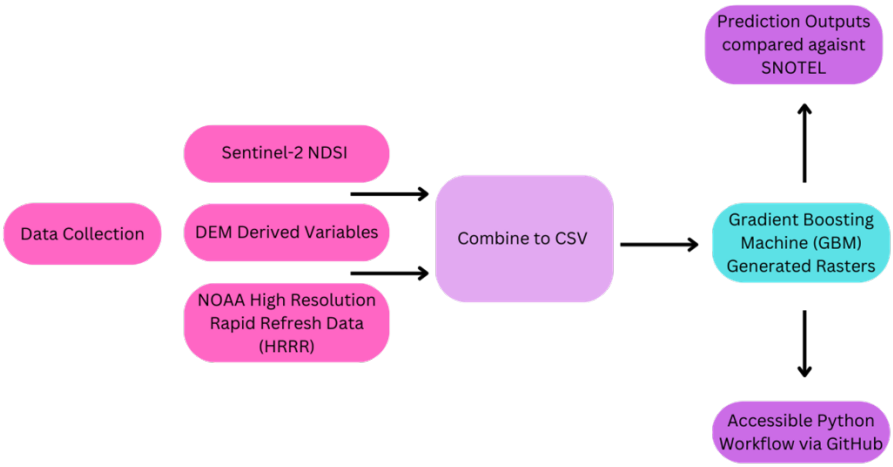

Figure 2. GBM Machine Learning Ensemble Workflow Breakdown.

Comprehensive quality control procedures were implemented prior to model training. All SNOTEL, HRRR, and Sentinel-2 datasets were screened for completeness, plausibility, and unit consistency. SNOTEL SWE (WTEQ) and snow depth (SNWD), originally reported in imperial units, were converted to millimeters and meters to match HRRR and Sentinel-2 inputs, which are expressed in SI units. Raster inputs were checked for spatial mismatches and reprojected to UTM Zone 12N (EPSG:32612). HRRR hourly forecasts, originally in UTC, were converted to

Mountain Standard Time (MST) and merged with daily SNOTEL observations using the merge_asof() function with a ±3-hour tolerance to ensure temporal alignment. Sentinel-2 imagery was screened using the MSK_CLDPRB cloud-probability mask, and cloud-contaminated pixels were replaced when the Normalized Difference Snow Index (NDSI) exceeded 0.4, indicating high-confidence snow cover. Missing data gaps shorter than three days were linearly interpolated; longer gaps were excluded entirely from training.

Following model calibration and validation for WY2020–WY2024, the trained ensemble was extended to generate near-term projections (WY2025–WY2030). During the calibration period, HRRR atmospheric variables and SNOTEL observations jointly informed the model: SNOTEL provided reference SWE and snow depth values for supervised learning, and HRRR variables represented daily meteorological forcing. After training, the model no longer required SNOTEL input because the statistical relationships between terrain, spectral, and atmospheric predictors were embedded in the ensemble parameters. Future projections substituted HRRR atmospheric variables with PRISM temperature and precipitation fields that preserve regional climate gradients, and Sentinel-2 NDSI composites or interpolated equivalents were used to maintain optical snow cover information. Ensemble variance increased during projection years, reflecting the added uncertainty of extrapolating beyond the period of HRRR and SNOTEL availability.

To maintain spatial consistency, all raster predictors were resampled to 30-meter resolution and aligned to the digital elevation model (DEM). Bilinear resampling was used for continuous variables such as temperature and shortwave radiation, while nearest-neighbor resampling was applied to categorical layers. Each annual training dataset included harmonized

predictors derived from SNOTEL, HRRR, and Sentinel-2 sources, including NDSI, near-surface air temperature (TMP2m), relative humidity (RH2m), shortwave radiation (DSWRF_inst), wind speed (WSPD10), and terrain attributes (elevation, slope, aspect).

Separate gradient-boosting models were trained for each water year to reduce temporal autocorrelation and stabilize convergence during years with limited Sentinel-2 availability. Hyperparameter tuning was conducted using randomized grid search with five-fold cross-validation implemented in scikit-learn. Parameter optimization favored moderate tree depths (6–8), learning rates between 0.03 and 0.05, and subsampling and feature-fraction rates between 0.7 and 0.8 to balance bias and variance. These configurations produced stable convergence and limited overfitting across snow seasons. Final predictions for SWE and snow depth were ensembled using the equal-weight mean of the three models, and pixel-wise standard deviation among model outputs represented ensemble uncertainty.

Model evaluation incorporated both statistical and physical validation frameworks. Spatial generalization was assessed using Leave-One-Station-Out Cross-Validation (LOSO-CV), in which each SNOTEL site was withheld from training and independently predicted to evaluate extrapolation across ungauged locations. Temporal generalization was tested using Leave-One-Year-Out Cross-Validation (LOYO-CV), which excluded entire water years to evaluate model stability under interannual climatic variability. RMSE, MAE, and R^2 were computed using SNOTEL observations from WY2020–WY2024, the full period for which temporally aligned SNOTEL, Sentinel-2, and HRRR data were available. Because only eight SNOTEL stations were available, these metrics likely overestimate true predictive accuracy—an expected limitation of

internal validation with a small reference network. SHAP-based feature importance and residual analysis were used to identify key predictors and evaluate terrain-dependent biases.

Predicted SWE and snow depth maps were compared with SNOTEL observations to evaluate spatial accuracy and magnitude correspondence. Multi-year SWE ensemble maps (WY2020–WY2024) were visually examined to verify that modeled snow distribution aligned with known accumulation zones and PRISM climate normals. Atmospheric variables from HRRR, including near-surface temperature (TMP2m, TAVG), relative humidity (RH2m), wind speed (WSPD10), and shortwave radiation (DSWRF_inst), were used as dynamic predictors to represent daily meteorological controls on accumulation and melt. These atmospheric variables, combined with terrain metrics and Sentinel-2 NDSI, enabled the ensemble to represent both spatial and temporal snowpack variability across complex alpine environments. Temporal line plots comparing modeled and observed SWE/SNWD were generated to ensure seasonal agreement, and systematic bias was assessed across elevation, slope, and aspect classes. Ensemble variance maps characterize spatial uncertainty, especially in steep terrain and partially forested basins where illumination and energy balance vary rapidly. PRISM-derived predictors provided an external consistency check by ensuring modeled outputs remained physically consistent with known regional climate gradients.

All model training, raster generation, and analysis were conducted in Python 3.12 within a reproducible conda-managed environment using NumPy (1.26.4), Pandas (2.2.2), Xarray (2024.3), Rasterio (1.3.8), Scikit-learn (1.5.0), LightGBM (4.3.0), XGBoost (2.0.3), and CatBoost (1.2.5). Multi-year HRRR archives were processed in parallel using Dask on Montana State University’s Tempest HPC cluster. Intermediate datasets were stored as .parquet and .tif

files for reproducibility. All workflows were version-controlled within the Snow_ML Jupyter Notebook.

The final workflow balances model sophistication with operational reproducibility and interpretability. By integrating SNOTEL observations, HRRR atmospheric variables, Sentinel-2 spectral indices, and PRISM climatological baselines within an ensemble gradient-boosting framework, the approach produces physically consistent 30-meter SWE and snow depth estimates for the Absaroka–Beartooth Wilderness. Visualization of ensemble maps, uncertainty rasters, and temporal comparisons with SNOTEL records demonstrate that the framework captures the dominant spatial and climatic controls on snowpack evolution while remaining transparent, scalable, and reproducible for future machine-learning applications. .

CHAPTER FIVE

ANALYSIS AND RESULTS

The Gradient Boosting Machine (GBM) ensemble produced spatially explicit estimates of snow depth (SNWD) and snow water equivalent (SWE) across the Absaroka–Beartooth Wilderness, revealing clear spatial and temporal patterns in modeled snowpack conditions. The ensemble combined XGBoost, LightGBM, and CatBoost algorithms trained on integrated HRRR atmospheric variables, Sentinel-2 derived NDSI, and terrain attributes including elevation, slope, and aspect. Model predictions were compared with SNOTEL observations to evaluate performance and bias, and ensemble means were used to generate spatially continuous 30-meter SWE and snow depth maps for each water year. Performance differences among XGBoost, LightGBM, and CatBoost were evaluated using repeated five-fold cross-validation. Although formal significance testing (e.g., Tukey HSD) was not applied, differences in RMSE and R^2 among the models were small and consistent across folds, supporting the use of equal-weight averaging in the ensemble.

A water year spans from October 1 through September 30 and follows the natural cycle of snow accumulation, peak seasonal storage, and spring melt that characterizes mountain snow regimes in the western United States. Within this context, the modeled snow depth and snow water equivalent (SWE) maps for water years 2020 through 2029 (Fig. 4,5,6,7) depict single-day predictions selected by identifying the closest cloud-free Sentinel-2 acquisition for each year. Each map represents an aggregated snapshot of modeled conditions on one representative day, typically during peak accumulation, rather than a seasonal or monthly mean. These images should therefore be interpreted as spatial snapshots of the snowpack for that specific point in the

water year, allowing consistent comparison across years while preserving terrain-controlled structure.

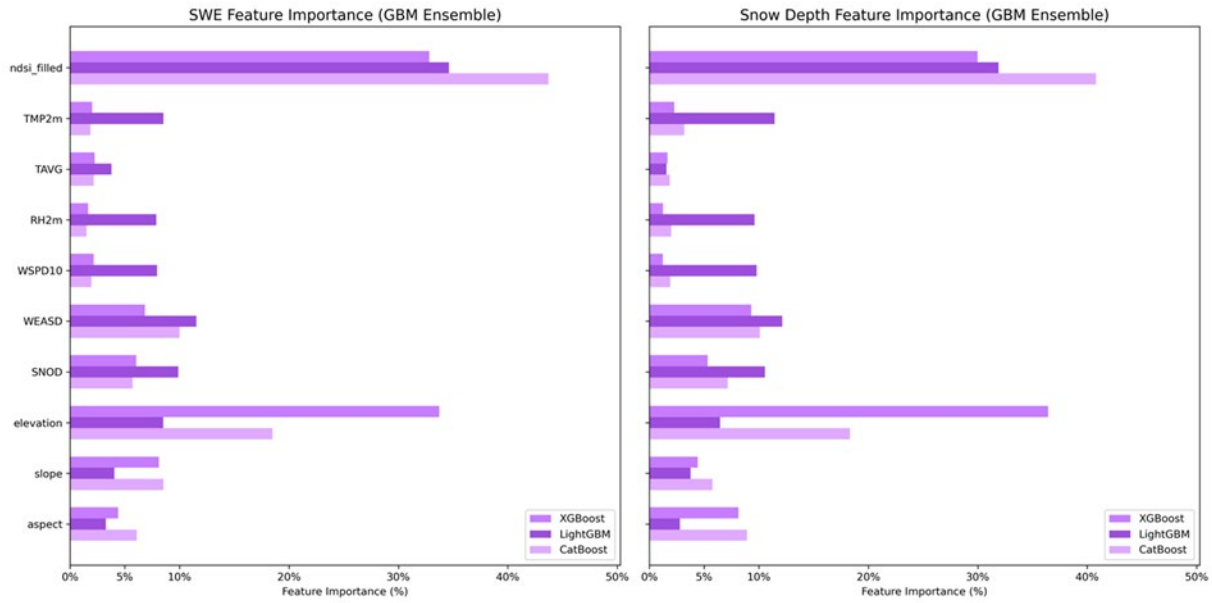

Figure 3. Feature importance across ensemble Gradient Boosting Machine (GBM) models for Snow Water Equivalent (SWE) and Snow Depth prediction.

The ensemble outputs for WY2020 through WY2024 show spatially coherent patterns of snow accumulation across the Absaroka and Beartooth region. High elevation basins and windward slopes exhibit persistent accumulation, while south facing and valley side terrain show reduced snow presence due to enhanced melt and lower retention. Interannual variability in modeled SWE and snow depth aligns with observed differences among years and highlights fluctuations in snowpack extent across the study area.

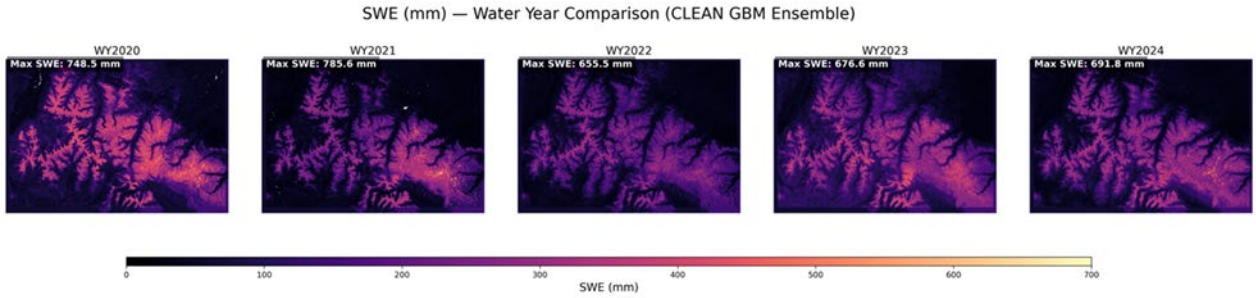

Figure 4. Modeled Snow Water Equivalent (SWE) ensembles across the Absaroka–Beartooth region (WY2020–WY2024, 30 m resolution).

Across the five-year sequence, the ensemble maintains a consistent spatial imprint. WY2020 shows widespread accumulation across upper elevation basins. WY2021 retains this structure with localized variability linked to slope exposure. WY2022 shows a modest reduction in high SWE areas, followed by partial recovery in WY2023. By WY2024, accumulation remains strongest at the highest elevations while mid elevation and leeward zones show reduced continuity. Collectively, these representative daily maps capture year-to-year variability while preserving the dominant terrain-controlled pattern of snow distribution.

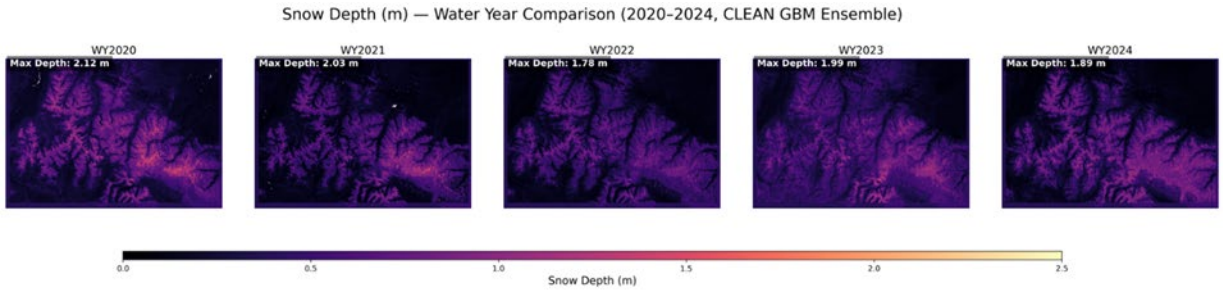

Figure 5. Modeled snow depth for water years 2020 through 2024 across the Absaroka and Beartooth region at 30 meter resolution.

Model predicted snow depth and SWE maps for future water years 2025 through 2029 (Fig. 6,7) follow the same representative-day framework applied to the historical period. Each projected map corresponds to a single day selected to align with the annual Sentinel-2 input for

that run. This consistency allows direct comparison between historical and future years while maintaining the topographic and climate-based controls embedded in the GBM ensemble.

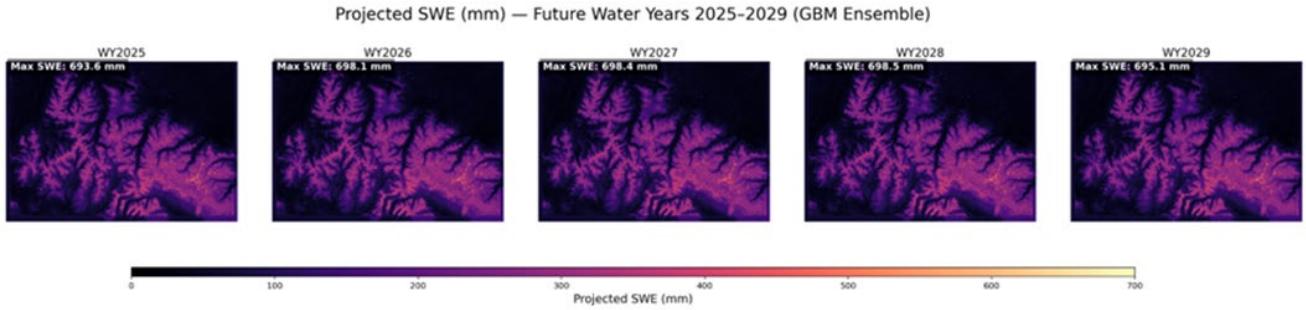

Figure 6. Projected Snow Water Equivalent (SWE) ensembles across the Absaroka–Beartooth region (WY2025–WY2029, 30 m resolution).

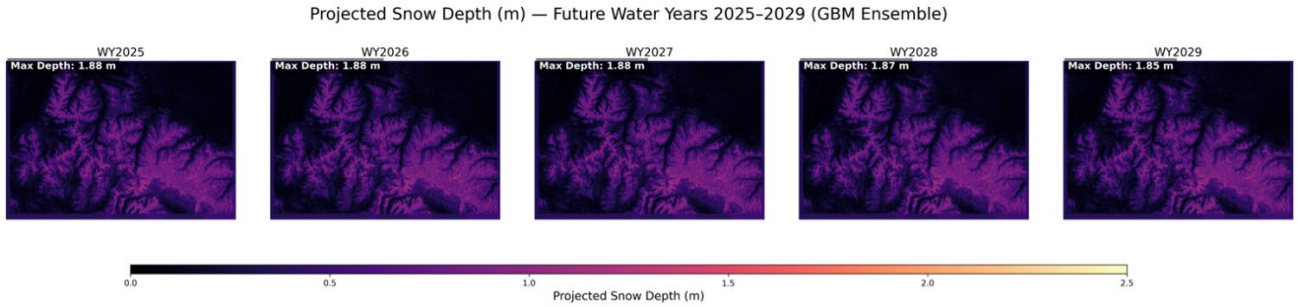

Figure 7. Projected Snow Depth ensembles across the Absaroka–Beartooth region (WY2025–WY2039, 30 m resolution).

Future ensemble projections reveal a gradual reduction in modeled snow-depth and SWE across the region through the late 2020s. The ensemble retains strong spatial coherence, with persistent accumulation zones in high elevation basins and earlier melt along lower elevation and south facing terrain. WY2025 and WY2026 display widespread snow cover across upper basins and windward ridges. By WY2027, snowpack reductions become more evident at mid elevations, especially along exposed southern and eastern slopes. WY2028 and WY2029 show continued contraction of high SWE areas, indicating shallower accumulation and earlier ablation.

By WY2029, much of the region exhibits lower snow depth intensity, although isolated high elevation basins retain consistent snow.

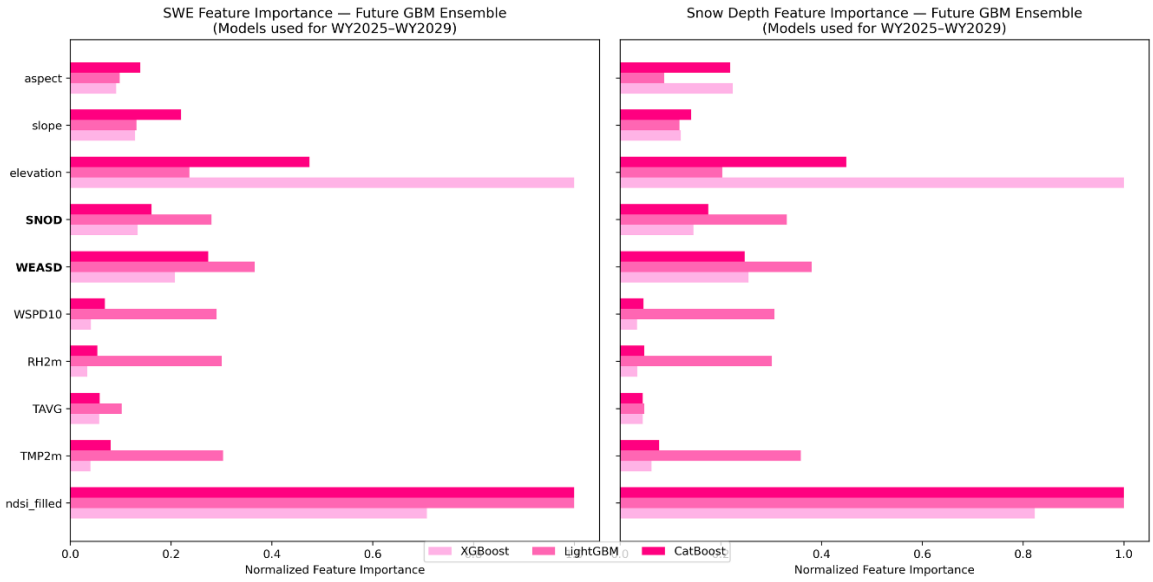

Figure 8. Feature Importance for snow depth and Snow Water Equivalent (SWE) ensembles across the Absaroka–Beartooth region (WY2025–WY2029, 30 m resolution).

The future GBM ensemble produced consistent patterns in predictor importance across XGBoost, LightGBM, and CatBoost (Fig 8). NDSI_filled remained the most influential variable for both SWE and snow depth, indicating that surface snow cover continues to control much of the spatial and temporal variability in snowpack under projected conditions. Elevation was the second-strongest predictor, with LightGBM assigning the greatest weight, reflecting the persistent role of topographic gradients in future snow accumulation and retention.

Snowpack variables (SNOD and WEASD) contributed substantially to the SWE model, showing that antecedent snow conditions remain important in future estimates. Atmospheric variables (TMP2m, RH2m, WSPD10, and TAVG) showed moderate importance but were

consistently secondary to surface and topographic controls, suggesting that large-scale terrain and snow presence remain dominant drivers even as weather patterns shift.

Slope and aspect contributed the least but still provided additional spatial structure related to energy balance differences across terrain. Overall, the ensemble's future feature importance patterns were physically coherent and indicate that snow cover, elevation, and existing snowpack conditions will remain the primary determinants of SWE and snow depth across the Absaroka–Beartooth region under future climate scenarios.

These projected spatial patterns demonstrate the capacity of the ensemble to reproduce terrain governed accumulation and melt processes under evolving atmospheric conditions. They underscore the dominant role of elevation, slope exposure, and terrain shading in shaping snowpack variability across both historical and projected water years.

Across the five modeled water years, SWE distributions exhibited notable interannual variability driven by differences in winter temperature and precipitation. For example, Water Year 2021 displayed below-average SWE across much of the domain, reflecting a warmer and drier winter season, while WY 2023 showed more extensive and deeper snow cover corresponding to colder conditions and higher precipitation totals. This temporal variability demonstrates that the model successfully captured year-to-year changes in snow accumulation linked to meteorological forcing represented in the HRRR input variables.

Validation against SNOTEL observations (Fig. 12) indicated strong agreement between modeled and measured SWE across the eight calibration sites. At most locations, modeled annual mean SWE tracked SNOTEL observations within a similar magnitude and seasonal trajectory, though slight underestimation occurred at lower-elevation sites such as East Boulder and White

Mill. Snow depth followed comparable trends, with higher values recorded at alpine stations such as Fisher Creek and Monument Peak, where mean depths reached 0.5-1.0 m. The consistency between modeled and observed values supports the reliability of the ensemble approach in representing both accumulation and melt dynamics across diverse elevation bands.

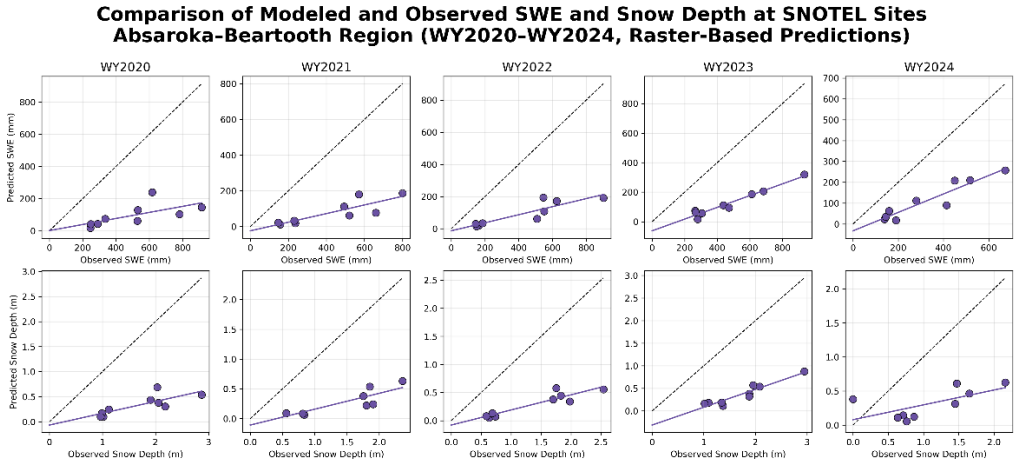

Figure 9. Comparison of modeled and observed Snow Water Equivalent (SWE) and snow depth at SNOTEL sites across the Absaroka–Beartooth region (WY2020–WY2024).

The modeled SWE fields reflected the influence of elevation and slope orientation, with higher SWE values occurring at higher elevations and greater retention on north-facing terrain. These spatial patterns aligned directly with the terrain variables used in the GBM ensemble, and the full distribution is shown in Figure 13. Model performance metrics for the five water years demonstrated a consistent range of predictive error across both SWE and snow depth. SWE RMSE values for individual years fell between roughly 150 and 250 millimeters, and snow depth RMSE values were lower across all sites. R^2 values remained positive each year, and year-by-year metrics are presented in Figure 8.

Predicted and observed SWE values were also compared across elevation for all SNOTEL stations from WY2020 through WY2024 (Figures 14–15). The combined dataset

showed that SWE increased with elevation across all years, and the ensemble reproduced this gradient in each case. Predicted SWE values were generally lower than observed values at most stations, with the largest differences at the highest-elevation sites, including Fisher Creek and Monument Peak above 2600 meters. In years with greater seasonal snow accumulation, such as WY2020 and WY2023, predicted SWE followed the observed elevation trend across stations. In lower-accumulation years, including WY2022, the spread between predicted and observed SWE was larger. Across the full period, higher-elevation stations showed the closest correspondence in elevation-based patterns between modeled and observed SWE.

To evaluate model performance across elevation gradients, predicted and observed SWE values were compared for all SNOTEL sites from WY2020–WY2024 (Figures 8). The model showed that mean predicted versus actual SWE aggregated across all water years, showing that SWE increased sharply with elevation (Fig 14). The ensemble captured this gradient effectively but generally underpredicted SWE magnitude, particularly at mid- and high-elevation stations where snow accumulation was greatest. The underestimation was most apparent at Fisher Creek and Monument Peak, both above 2,600 m, suggesting that the model may underestimate SWE in deep-snow conditions due to limited representation of extreme accumulation within the training data.

In years with above-average accumulation (WY 2020 and WY 2023), the model reproduced spatial trends accurately but remained slightly below observed SNOTEL values across most sites (Fig. 9). Drier years, such as WY 2022, exhibited higher relative error and greater spread between observed and predicted values, consistent with greater uncertainty during low-snow conditions. Across all years, higher-elevation sites (above 2,600 m) maintained the

strongest relative agreement between modeled and observed SWE, despite the consistent underprediction in magnitude.

These elevation-based validation plots confirm that the GBM ensemble effectively reproduced spatial patterns and elevation-dependent gradients in SWE while exhibiting a conservative bias typical of data-driven snowpack models. Minor deviations at low elevations likely reflect transitional rain–snow dynamics and the reduced reliability of HRRR-based temperature variables near the freezing threshold. Despite this bias, the ensemble maintained stable performance across contrasting snow seasons, capturing both the spatial structure and interannual variability of SWE with physically consistent elevation dependencies.

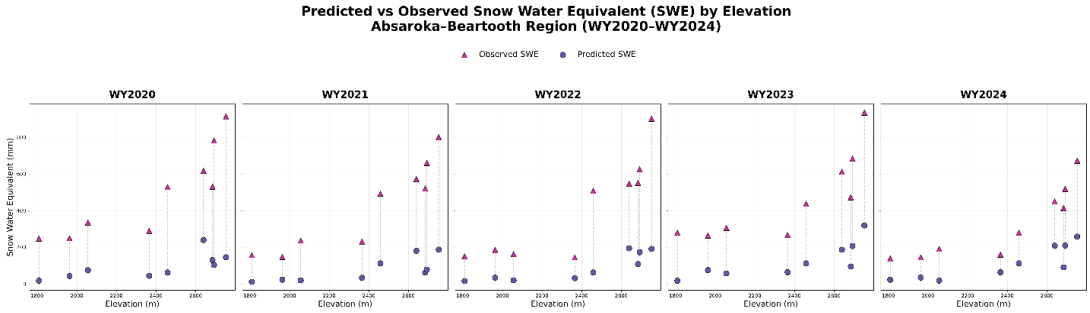

Figure 10. Predicted vs observed Snow Water Equivalent (SWE) by elevation at SNOTEL sites (WY2020–WY2024).

Overall, these quantitative results reinforce the spatial and temporal patterns observed in the preceding sections and confirm that the ensemble model reproduced snowpack variability with sufficient accuracy to support subsequent interpretation and discussion. Building on these validation results, the following discussion interprets the spatial and physical meaning of the modeled snowpack dynamics. Robustness

CHAPTER SIX

DISCUSSION

The ensemble mean of XGBoost, LightGBM, and CatBoost produced the highest overall predictive accuracy, reducing RMSE and increasing R^2 relative to the individual models. Averaging across the three algorithms stabilized predictions in complex terrain and reduced sensitivity to localized anomalies in the SNOTEL record, such as drift-induced snow accumulation or site-specific melt–freeze cycles that do not represent surrounding grid-scale conditions. The resulting ensemble fields produced spatially coherent SWE and snow depth estimates that aligned with observed topographic and climatic gradients across the Absaroka–Beartooth region. Inter-model variance further represented spatial uncertainty, particularly in steep terrain and areas with heterogeneous snow cover.

Modeled accumulation patterns were consistent with established controls on mountain snowpack. The ensemble showed persistent snow retention along high-elevation ridgelines and north-facing slopes, where SWE commonly exceeded 300–400 mm and snow depths reached approximately 0.3–0.4 m. In contrast, low-elevation foothills and south-facing slopes exhibited reduced accumulation, reflecting warmer temperatures, higher solar exposure, and more rapid melt.

The transition from HRRR- and SNOTEL-based calibration to PRISM-based projection maintained continuity in the physical relationships governing snow accumulation and melt. Although future HRRR and SNOTEL data were unavailable for projection, the use of PRISM temperature and precipitation preserved spatial gradients that drive snowpack formation, allowing the ensemble to maintain internally consistent estimates for future water years.

Elevation exerted the strongest control on modeled SWE, with a clear increase in accumulation above approximately 2,400 m. Across all water years, sites above this threshold typically accumulated >200–250 mm of SWE, while values remained below 100 mm at lower elevations. Corresponding snow depth patterns ranged from roughly 0.30 m at low elevations to >0.70 m in alpine terrain. These elevation-dependent gradients agree with prior studies demonstrating the dominant influence of altitude on snowfall, temperature, and accumulation processes in mountain environments (Largeron et al., 2020; Aragon et al., 2021).

Aspect also shaped snow retention by modifying surface energy balance. North- and east-facing slopes retained snow for longer periods, whereas south-facing slopes experienced accelerated melt due to greater exposure to shortwave radiation (Kang et al., 2020). These topographic interactions produced clear spatial patterns in the ensemble maps, especially along high-elevation divides where persistent late-season snow cover was most pronounced.

Feature importance results (Fig. 13) reinforced these controls. Elevation and near-surface temperature variables (TMP2m, TAVG) were consistently the strongest predictors, followed by shortwave radiation (DSWRF_inst) and modeled snow metrics (SNOD, WEASD). These patterns highlight the primary influence of temperature and radiation on snow accumulation and ablation processes (Quéno et al., 2024; Vionnet et al., 2022). Relative humidity and NDSI also contributed meaningfully, indicating that both atmospheric moisture and surface snow indicators improved model sensitivity to snowpack variability.

Interannual differences in modeled SWE corresponded with meteorological variability represented in the HRRR and PRISM inputs. Low-accumulation years, such as WY2021, aligned with warmer winter temperatures and precipitation deficits relative to the five-year mean.

Conversely, WY2023 displayed the highest modeled SWE and snow depth, corresponding to a colder, wetter winter with more frequent frontal storm activity.

Despite strong spatial agreement, the ensemble consistently underpredicted SWE relative to SNOTEL observations across mid- and high-elevation sites (Figures 14–15). This conservative bias is well documented in empirical and hybrid snow models, which often smooth extreme accumulation due to subgrid heterogeneity, coarse meteorological forcing, and uncertainty in precipitation phase partitioning (Broxton et al., 2016; Largeron et al., 2020; Aragon et al., 2021). In this study, the bias likely reflects the downscaling of HRRR inputs and limited representation of deep-snow conditions in the training dataset. As a result, the ensemble favored climatologically consistent values rather than reproducing absolute maxima, a pattern also observed in previous MODIS- and HRRR-based reconstructions (Vionnet et al., 2022; Broxton et al., 2016).

Spatially, SWE distributions during low-accumulation years were more discontinuous, below approximately 2,200 m, whereas high-SWE years exhibited widespread coverage extending into mid-elevation forests. These contrasts demonstrate that the ensemble responded appropriately to interannual temperature and precipitation variability and did not overfit terrain or static predictors. Overall, the model reproduced the dominant spatial patterns, elevation gradients, and temporal variability expected across the Absaroka–Beartooth region.

CHAPTER SEVEN

CONCLUSION

The GBM ensemble effectively estimated Snow Water Equivalent (SWE) and snow depth across the Absaroka Beartooth Wilderness at thirty meter resolution. Elevation, temperature, and radiation consistently emerged as the most influential predictors, producing realistic spatial and temporal patterns of snow accumulation and melt across highly variable mountain terrain. The integration of Sentinel 2 optical data, HRRR atmospheric reanalysis, PRISM climatological baselines, SNOTEL ground observations, and terrain variables derived from the digital elevation model demonstrate that machine learning can help close spatial gaps in mountain snow monitoring while maintaining physical interpretability. The use of XGBoost, LightGBM, and CatBoost within a single ensemble reduced localized bias and minimized variability among the individual models, resulting in smoother and more stable predictions.

Uncertainty in the ensemble was greatest in steep alpine areas and partially forested basins, where Sentinel 2 retrievals were limited by cloud cover, terrain shadows, and mixed canopy and snow scenes. Downscaling HRRR atmospheric variables from three-kilometer resolution to thirty meters introduced interpolation artifacts in locations with rapidly changing temperature and radiation conditions. Limited SNOTEL coverage, especially above three thousand meters, restricted independent validation in the highest elevation zones. The addition of PRISM temperature and precipitation helped maintain broad climatological consistency but could not fully resolve the differences in scale between coarse atmospheric forcing and fine scale snow processes. Even with these challenges, the ensemble produced spatially coherent SWE and

snow depth fields, and variation among the three models provided a useful measure of confidence across the study region.

The results agree with previous research that identifies elevation, air temperature, and incoming shortwave radiation as dominant influences on snow accumulation and melt in mountain environments. This study extends those findings by combining Sentinel 2 NDSI, HRRR atmospheric data, and PRISM climatology within a unified machine learning workflow that represents small scale snowpack variability in complex terrain. The ability to incorporate dynamic atmospheric variables at this resolution represents a meaningful advancement for data driven snow modeling in regions with limited ground observations.

The modeling framework developed here provides a practical and interpretable tool for estimating snow water storage, melt timing, and likely contributions to downstream hydrology. In the northern Rocky Mountains, where mountain snowpack supplies most of the annual streamflow, such tools are increasingly important for water resource planning, ecological forecasting, and public safety. Continued refinement of ensemble modeling methods will support climate readiness as warming temperatures, shifting precipitation, and earlier melt increase uncertainty in seasonal water availability.

Future development of this framework should incorporate additional predictors such as albedo, vegetation cover, and metrics of snow persistence. The use of higher resolution atmospheric datasets would improve representation of local weather patterns. Comparisons with physically based snow models, including SNOWPACK, iSnobal, and SNODAS, would strengthen validation and clarify where empirical and process-based approaches converge. Targeted field measurements remain essential for ensuring that machine learning models remain

physically reliable. Additional sampling at high elevation sites and along gradients in slope and aspect would reduce gaps in the observational record.

A significant opportunity for future work is the implementation of true external validation. Predicting SWE and snow depth for an entire independent snow year, such as Water Year 2025, without retraining the model, would provide a strong evaluation of the model's reliability when applied to an independent snow year. This effort would benefit from a structured sampling plan that includes transects across multiple elevations, slopes, and terrain aspects, as well as repeated measurements following snowfall or rapid melt periods. These data would allow direct comparison with ensemble predictions and would support quantitative estimates of predictive uncertainty.

The continued development of this method will benefit from engagement with regional water managers and local stakeholders. Their involvement will help ensure that snowpack monitoring tools remain focused on real world needs and maintain relevance for communities that rely on consistent water supply. As environmental observation networks expand and machine learning methods advance, integrated workflows such as this ensemble model hold significant potential for improving our understanding of snow accumulation, persistence, and melt across the mountain West.

REFERENCES CITED

Alexander, C. R., James, E. P., & Dowell, D. C. (2020). Evaluation of HRRR model performance for surface meteorological variables. Weather and Forecasting, 35(3), 919–935. https://doi.org/10.1175/WAF-D-19-0185.1

Aragon, C. M., Molotch, N. P., & Painter, T. H. (2021). Modeling snow water equivalent using MODIS and SNOTEL data. Remote Sensing of Environment, 252, 112136. https://doi.org/10.1016/j.rse.2020.112136

Arumugam, S., et al. (in press). Machine learning approaches for snow hydrology: A review and outlook. [Journal to be determined].

Barnett, T. P., Adam, J. C., & Lettenmaier, D. P. (2005). Potential impacts of a warming climate on water availability in snow-dominated regions. Nature, 438, 303–309. https://doi.org/10.1038/nature04141

Benjamin, S. G., Brown, J. M., Brunet, G., Lynch, P., Saito, K., & Schlatter, T. W. (2016). A North American hourly assimilation and model forecast cycle: The Rapid Refresh. Monthly Weather Review, 144(4), 1669–1694. https://doi.org/10.1175/MWR-D-15-0242.1

Blaylock, B. K. (2022). Herbie: A Python package for downloading and managing numerical weather prediction model output. Journal of Open Source Software, 7(73), 4275. https://doi.org/10.21105/joss.04275

Bormann, K. J., Brown, R. D., Derksen, C., & Painter, T. H. (2018). Estimating snow-cover trends from space. Hydrology and Earth System Sciences, 22(11), 6291–6306. https://doi.org/10.5194/hess-22-6291-2018

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Broxton, P. D., Zeng, X., & Dawson, N. (2016). Why do global reanalyses and land data assimilation products underestimate snow water equivalent? Journal of Hydrometeorology, 17(11), 2743–2761. https://doi.org/10.1175/JHM-D-16-0056.1

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). https://doi.org/10.1145/2939672.2939785

Copernicus/European Space Agency. (n.d.). Sentinel-2 user handbook. https://sentinels.copernicus.eu

Dorogush, A. V., Ershov, V., & Gulin, A. (2018). CatBoost: Gradient boosting with categorical features. In NeurIPS 2018 Workshop on Machine Learning Systems. https://arxiv.org/abs/1810.11363

Dozier, J. (1989). Spectral signature of snow in visible and near-infrared. Cold Regions Science and Technology, 15(1), 49–59. https://doi.org/10.1016/0165-232X(89)90052-9

Dozier, J., & Frew, J. (1990). Rapid calculation of terrain parameters for radiation modeling from digital elevation data. IEEE Transactions on Geoscience and Remote Sensing, 28(5), 963–969. https://doi.org/10.1109/TGRS.1990.572934

Dozier, J., Bair, E. H., & Davis, R. E. (2016). Estimating the spatial distribution of snow water equivalent in the world's mountains. WIREs Water, 3(3), 461–478. https://doi.org/10.1002/wat2.1140

Dowell, D. C., Alexander, C. R., James, E. P., et al. (2022). The HRRR model: History and future of the United States' most advanced real-time weather model. Bulletin of the American Meteorological Society, 103(4), E878–E907. https://doi.org/10.1175/BAMS-D-21-0124.1

Drusch, M., et al. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. Remote Sensing of Environment, 120, 25–36. https://doi.org/10.1016/j.rse.2011.11.026

Dunmire, D., et al. (in press). The global economic value of snowmelt. [Journal to be determined].

English, J. M., Quéno, L., Lafaysse, M., et al. (2021). Evaluation of HRRR winter precipitation forecasts against QPE. NOAA Technical Memorandum. https://repository.library.noaa.gov/

EPSG Geodetic Parameter Dataset. (2023). EPSG: 32612 – WGS 84 / UTM zone 12N. International Association of Oil & Gas Producers. https://epsg.io/32612

Evans, I. S. (1972). General geomorphometry, derivatives of altitude, and descriptive statistics. In R. J. Chorley (Ed.), Spatial analysis in geomorphology (pp. 17–90). Methuen.

Fassnacht, S. R., Deems, J. S., & Painter, T. H. (2018). Topographic controls on snow distribution and melt. Frontiers in Earth Science, 6, 3. https://doi.org/10.3389/feart.2018.00003

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Gascoin, S., Grizonnet, M., Bouchet, M., Salgues, G., & Hagolle, O. (2019). Theia Snow collection: High-resolution operational snow-cover maps from Sentinel-2 and Landsat-8. Earth System Science Data, 11(2), 493–514. https://doi.org/10.5194/essd-11-493-2019

Gillies, S., Van den Brink, L., & Ward, B. (2023). rasterio: Geospatial raster I/O for Python [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.8287423

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment, 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Hall, D. K., Riggs, G. A., & Salomonson, V. V. (2002). MODIS snow-cover products. Remote Sensing of Environment, 83(1–2), 181–194. https://doi.org/10.1016/S0034-4257(02)00095-0

Hedrick, A., et al. (2018). Direct observations of snow depth and SWE using Airborne Snow Observatory. Water Resources Research, 54(8), 6750–6768. https://doi.org/10.1029/2018WR022707

Hofmeister, K., McGowan, M., & Yarleque, C. (2022). Evaluating the utility of Sentinel-2 snow-cover data in complex alpine terrain. The Cryosphere, 16(3), 735–752. https://doi.org/10.5194/tc-16-735-2022

Hoyer, S., & Hamman, J. J. (2017). xarray: N-D labeled arrays and datasets in Python. Journal of Open Research Software, 5(1), 10. https://doi.org/10.5334/jors.148

Justice, C. O., Román, M. O., Csiszar, I., et al. (2013). Suomi NPP VIIRS: A new instrument for studies of the Earth from space. Remote Sensing of Environment, 135, 3–15. https://doi.org/10.1016/j.rse.2013.01.019

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 30, 3146–3154.

Kim, E. J., et al. (2017). NASA SnowEx 2017 overview: Assessing snow mass and snow energy balance. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10(12), 5694–5711. https://doi.org/10.1109/JSTARS.2017.2737452

Largeron, C., Morin, S., & Boone, A. (2020). Snow cover estimation through data assimilation of satellite observations in a snowpack model. The Cryosphere, 14(7), 2145–2167. https://doi.org/10.5194/tc-14-2145-2020

Luce, C. H. (2009). Spatial patterns of snowpack trends in the Pacific Northwest. Geophysical Research Letters, 36(16), L16501. https://doi.org/10.1029/2009GL039090

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774.

Malygin, A., Rets, E., et al. (2024). Operational estimation of snow water equivalent across the western U.S. using gradient boosting machine ensemble. Water Resources Research, 60(5), e2023WR035613. https://doi.org/10.1029/2023WR035613

Margulis, S. A., Cortés, G., Girotto, M., & Durand, M. (2016). A Landsat-era Sierra Nevada snow reanalysis (1985–2015). Water Resources Research, 52(10), 8069–8093. https://doi.org/10.1002/2016WR018704

Molotch, N. P., & Bales, R. C. (2005). Scaling snow observations from the point to the grid element: Implications for observation networks. Water Resources Research, 41(11), W11421. https://doi.org/10.1029/2005WR004229

Mote, P. W., Hamlet, A. F., Clark, M. P., & Lettenmaier, D. P. (2006). Declining mountain snowpack in western North America. Journal of Climate, 19(18), 6209–6220. https://doi.org/10.1175/JCLI3979.1

Natural Resources Conservation Service, U.S. Department of Agriculture. (n.d.). Snow Telemetry (SNOTEL) and snow course data and products. https://www.nrcs.usda.gov

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

PRISM Climate Group. (2004). PRISM gridded climate data. Oregon State University. http://prism.oregonstate.edu

Quéno, L., Lafaysse, M., Vionnet, V., et al. (2024). Snow redistribution in an intermediate-complexity snow hydrology modelling framework. The Cryosphere, 18(7), 3533–3555. https://doi.org/10.5194/tc-18-3533-2024

Rocklin, M. (2015). Dask: Parallel computation with blocked algorithms and task scheduling. In K. Huff & J. Bergstra (Eds.), Proceedings of the 14th Python in Science Conference (SciPy 2015) (pp. 130–136). https://doi.org/10.25080/Majora-7b98e3ed-013

Roy, D. P., Wulder, M. A., Loveland, T. R., et al. (2014). Landsat-8: Science and product vision for terrestrial global change research. Remote Sensing of Environment, 145, 154–172. https://doi.org/10.1016/j.rse.2014.02.001

Stewart, I. T. (2009). Changes in snowpack and snowmelt timing in western North America. Earth-Science Reviews, 89(3–4), 219–248. https://doi.org/10.1016/j.earscirev.2008.12.002

Sturm, M., Goldstein, M. A., & Parr, C. (2017). Snow, water, weather, and climate. Princeton University Press. https://doi.org/10.1515/9781400888892

U.S. Department of Agriculture. (n.d.). Water supply from snowmelt in the western U.S. https://www.usda.gov

Vionnet, V., Quéno, L., & Lafaysse, M. (2022). Post-processing model snow level for improved rain–snow partitioning. Journal of Geophysical Research: Atmospheres, 127(18), e2021JD035391. https://doi.org/10.1029/2021JD035391

Winstral, A., Elder, K., & Davis, R. E. (2002). Spatial snow modeling of wind-redistributed snow using terrain-based parameters. Journal of Hydrometeorology, 3(5), 524–538. https://doi.org/10.1175/1525-7541(2002)003<0524:SSMOWR>2.0.CO;2.