

Assessing Subgrid Snow Depth Variability with Satellite Laser Altimetry Potential and challenges

Zhihao Liu^a, Désirée Treichler^{a,*}, Simon Filhol^a

^a University of Oslo, Department of Geosciences, Sem Sælands vei 1, Oslo, 371

Abstract

Observing and estimating subgrid variability of seasonal snow cover, in particular snow depth in remote areas, poses significant challenges. This study uses ICESat-2 satellite laser altimetry to capture this variability at the hillslope scale and incorporates it into a downscaling model. The model excels in refining coarse-resolution snow depth data to hillslope scale. Validation tests in the Hardangervidda region of southern Norway show that the downscaled snow depth achieved R^2 values ranging from 0.74 to 0.88 (post-calibration) for aggregated prediction. However, the study also highlighted challenges in accurately predicting snow depth at the microscale (< 50 m), attributable to factors such as the use of non-snow-free reference ground, oversimplified wind fields, and sampling imbalances, necessitating the post-calibration. The research contributed a workflow for snow depth retrieval from the ICESat-2 ATL08 product at national wide, featuring an efficient co-registration algorithm and a bias correction model that utilizes ICESat-2 snow-free measurements correcting DEMs. Moreover, our scalable and explainable downscaling model, which incorporates temporal variability, advances the understanding of snow depth dynamics at the hillslope scale. While the validation was specific to Hardangervidda, the methodology shows potential for broad applications. Further validation in vegetated areas is crucial. Future development of this workflow should focus on downscaling wind fields, and developing more adaptable calibration methods for varied scenarios.

Keywords: Snow depth, Subgrid variability, Laser altimetry, Statistical downscaling, ICESat-2, XGBoost

1. Introduction

Seasonal snow is a critical component of the Earth's system, influencing the surface energy balance, water availability, vegetation and trace gas fluxes. In a warming world, understanding the spatiotemporal variations of seasonal snow is increasingly vital for climate impact assessments, meltwater supply, permafrost modeling, ecological responses, and rain-on-snow flood events (Immerzeel et al., 2020; Kraaijenbrink et al., 2021; Livneh and Badger, 2020; Myers-Smith et al., 2020; Callaghan et al., 2011; Musselman et al., 2018). While remote sensing has effectively addressed snow cover, accurately observing and modeling snow depth (or snow mass) remains a formidable challenge (Tsang et al., 2022;

Mudryk et al., 2020), especially in remote and complex terrain with limited data availability.

Mountainous areas exhibit pronounced spatial heterogeneity in snow depth, a variability across multiple scales that are not adequately captured by traditional station measurements or radar remote sensing (Fassnacht et al., 2018; Tsang et al., 2022). Although recent developments in C-band radar technology have enabled snow depth retrieval at sub-kilometer resolution (Lievens et al., 2019, 2022), the fine-scale snow depth, such as hillslope levels (under 100 meters, Grünwald et al. (2010); Mott et al. (2018)) displays much greater spatial variability that remains poorly addressed. Recent years have witnessed the availability of laser scanning (or Lidar), which does not directly provide snow depth. Lidar operates by collecting data pre- and post-snowfall, enabling snow depth mapping through comparative analysis of height differences. While Lidar offers high-resolution data, its high cost and localized focus render it most effective for watershed-scale snow

*Corresponding author
Email addresses: zhihao1@geo.uio.com (Zhihao Liu),
desiree.treichler@geo.uio.no (Désirée Treichler),
simon.filhol@geo.uio.no (Simon Filhol)

depth mapping (Deems et al., 2013). With more available high-resolution digital elevation models (DEMs), much effort has been put into retrieving snow depth from space-borne lidar measurements, such as ICESat (Treichler and Kääb, 2017) and its successor ICESat-2. Deschamps-Berger et al. (2023) derived snow depth from ICESat-2 ATL06 products and reported an accuracy of 0.2 m (bias) and a precision (normalized median absolute deviation; NMAD) of 0.5 m for low slopes and 1.2 m for steeper areas when compared to airborne lidar measurements over the upper Tuolumne basin, California, USA. Enderlin et al. (2022) found that snow depth estimates based on ICESat-2 data had a median absolute deviation (MAD) ranging from 0.2 m for slopes $< 5^\circ$ to over 1 m for slopes $> 20^\circ$. Besso et al. (2024) questioned the varying accuracy of the ICESat-2 ATL08 product and developed a self-defined processed elevation product. These studies have uncovered ICESat-2 as an emerging and cost-efficient data source for snow depth but also have brought attention to the challenges associated with elevation differencing workflow, primarily stemming from discrepancies between reference DEMs and ICESat-2. Therefore, implementing this workflow requires careful co-registration and non-stationary bias correction on DEMs & ICESat-2. Additionally, deriving comprehensive snow-depth maps from sparse ICESat-2 measurements presents another significant challenge.

Another approach to understanding snow dynamics is through snow modeling. Researchers primarily use two modeling methods to study these dynamics. Process-based models (Lehning et al., 2006; Liston and Elder, 2006; Kim et al., 2021) are driven by meteorology forcing data and yield gridded snow depth products. However, these models often struggle to cover large areas or offer fine resolution, hindered by complex near-surface atmospheric processes and limited data on precipitation and wind fields (Freudiger et al., 2017), leading to a new question: how to quantify the subgrid variability of snow depth (Clark et al., 2011). The distribution patterns of snow exhibit a notable resemblance year after year due to their dependence on topography, vegetation, and consistent synoptic weather patterns (Sturm and Wagner, 2010; Revuelto et al., 2014; Freudiger et al., 2017; Pflug et al., 2021; Parr et al., 2020). This recurring similarity supports the use of computationally efficient statistical approaches. Many study cases seek to define parameterizations for sub-

grid variabilities, such as depletion curves, elevational gradients, probability distribution (Mendoza et al., 2020), coefficient of variation (Liston, 2004; He et al., 2019; Gisnås et al., 2016), scaling proprieties (Clark et al., 2011; Melvold and Skaugen, 2013; Mott et al., 2011) or topographic correlations (Helbig and van Herwijnen, 2017). To produce a snow depth map at fine scales still requires regression models. Multiple-linear regression (Grünwald et al., 2013; Dvornikov et al., 2015), binary regression trees (Revuelto et al., 2014), random forests (Revuelto et al., 2020) have been used to extract snow distribution patterns with varies performance (R^2 of 0.25-0.91). However, these statistical models typically require a substantial amount of training data from terrestrial or airborne sensors. Therefore, most spatial regression models can not be transferred to other catchments (Grünwald et al., 2013; Revuelto et al., 2020) with sparse temporal variability. In contrast, statistical models capable of generalizing subgrid variability are more accurately defined as downscaling models. These models are adept at refining snow depth data from coarse, broad-scale grids to finer, localized subgrid levels, and they can extend this analysis both retrospectively and into the future. There are currently few statistical downscaling models for snow depth (Pflug et al., 2021; Fiddes et al., 2022). The primary challenges include the efficient acquisition of comprehensive snow depth measurements and the accurate characterization of variability through relevant model features.

In light of these challenges and advancements, this work aims to further characterize the subgrid variability of snow depth from both observation and statistical modeling perspectives. The objectives of this study are as follows:

1. Retrieve snow depth from ICESat-2 laser altimetry data across mainland Norway.
2. Use this data to train a machine learning-based downscaling model that accommodates spatial and temporal variations of snow depth in mountain environments.
3. Validate the downscaled snow depth at different scales through comparisons with in-situ observations, gridded snow model products, and meteorological stations.
4. Explore the challenges encountered in our workflow and propose directions for future research.

To retrieve snow depth, we employed an efficient DEM co-registration algorithm (GDC, Gradient

Descent Co-registration) and corrected the non-stationary bias of DEM with ICESat-2 snow-free measurements. Subsequently, we statistically down-scaled ERA5 Land from its 9 km native resolution into a hillslope scale using a machine learning algorithm. The downscaling model is capable of capturing preferential snow deposition patterns at the hillslope scale with monthly temporal resolution. To our knowledge, this study marks the first attempt to retrieve snow depth from ICESat-2 nationwide and use it in statistically downscaling ERA5 Land.

2. Study Area and Data Setting

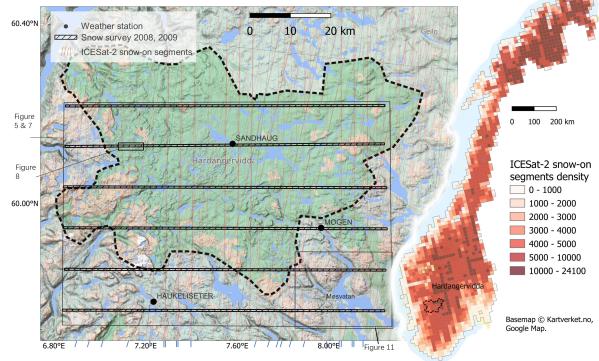


Figure 1: Map of the study area. Snow-on segments are displayed in red (from October 2018 to October 2022), while the blue ticks at the bottom annotate the ground tracks in April.

Norway, located on the western side of the Scandinavian Peninsula in northwestern Europe, spans latitudes from $57^{\circ} 58' N$ to $71^{\circ} 11' N$ and longitudes from $4^{\circ} 40' E$ to $30^{\circ} 58' E$. The country features a diverse climate, ranging from lowland valleys to high mountains. In winter, westerly winds bring moisture, resulting in substantial snowfall from the coast to inland areas. This snow acts as a crucial reservoir for hydropower, emphasizing the importance of estimating snow mass in mountain environments. Coastal areas experience strong winds due to temperature differences between the land and sea, while mountains serve as wind barriers, creating calmer conditions inland.

Hardangervidda, the largest mountain plateau in northern Europe, spans approximately $6,500 \text{ km}^2$. Its flat terrain is exposed to high winds and heavy snowfall, making it an ideal study site for snow depth variability (Figure 1). The plateau predominantly lies above 1,000 meters above sea level (m

a.s.l.), featuring low alpine regions with grass heaths, dwarf shrubs, and higher areas with bare rock or lichen/marsh tundra. The eastern region is characterized by its flat, open terrain with numerous lakes and streams. Conversely, the western and southern areas, reaching up to 1,700 m a.s.l. serves as a significant orographic feature to the prevailing westerly wind flow. As moist air masses encounter this mountain range, they are lifted, leading to increased precipitation on the windward slopes and a subsequent decrease on the leeward side. Snow accumulation typically begins in mid-September at higher elevations, peaking around late April. Mean annual precipitation ranges from 750 mm to less than 3,000 mm over relatively short distances, with approximately 50-60% of this precipitation falling as snow.

2.1. ICESat-2 ATL08 elevation product

Launched in September 2018, ICESat-2 is equipped with the Advanced Topographic Laser Altimeter System (ATLAS), which provides photon-counting Lidar measurements globally ([Neuenschwander and Pitts, 2019](#)). ATLAS comprises three beam pairs separated by 3.3 km on the ground. Each beam pair includes a strong and weak beam within a 90 m distance. The ATLAS system emits a pulse every 0.7 m along the track, covering a circular footprint with a diameter of ~ 15 m. The ICESat-2 ATL08 product (level L3A, version 5) offers elevation data in fixed segment sizes of 100 meters along the ground track ([Neuenschwander et al., 2022](#)). For each segment, there are 5 geolocations (subsegments) in 20 m intervals. Instead of using the mean elevation of the segment (h_{te_mean} , e.g. [Enderlin et al. \(2022\)](#)), we utilized the subsegment heights at the midpoint location of the segment, referring to as best-fit terrain elevation ($h_{te_best_fit_20m_2}$). The height is determined through polynomial fitting to terrain photons with slope correction and weighting applied ([Neuenschwander et al., 2022](#)). Mainland Norway has a total of 3,968 ATL08 granules available for analysis from October 14, 2018, to October 12, 2022. After removing invalid data, and excluding permanent ice and inland water, our dataset consists of 13,197,376 segments, including 4,778,904 snow-free segments on land and 8,418,472 segments with snow cover over land. The coverage of snow-on segments is displayed in Figure 1 with blue ticks representing the ground track in April.

2.2. Snow-off elevation product

For reference ground in snow depth retrieval, we employ the Norwegian DTM1 elevation model (DTM1), a 1 m Lidar-based product acquired by Kartverket between 2016 and 2022. As a sensitivity test of DEM resolution, we also utilize the 10-meter resolution variant, DTM10. To demonstrate the workflow's applicability in areas without Lidar-based elevation products, we incorporate global DEMs such as Copernicus GLO-30 (GLO30) and FABDEM (Forest And Buildings Removed Copernicus DEM, FAB) as reference ground. GLO30 ([European Space Agency, 2021](#)) is a 30-meter-resolution Digital Surface Model (DSM) acquired between December 2010 and January 2015 through Synthetic Aperture Radar interferometry (TanDEM-X mission). FAB, a variant of GLO30, eliminates buildings and trees using the random forest algorithm, enhancing accuracy ([Hawker et al., 2022](#)). FABDEM serves as a suitable reference for comparison with GLO30.

2.3. Large-scale reanalysis data

ERA5 Land (version 5) is an ECMWF (European Centre for Medium-Range Weather Forecasts) reanalysis product covering the period from 1950 to the present. It offers a spatial resolution of approximately 9 km ([Hersbach et al., 2020](#); [Muñoz-Sabater et al., 2021](#)). ERA5 Land describes water and energy cycles over global land areas, providing hourly data on 50 variables. This reanalysis data supplies the necessary forcing data for the downscaling model to generate sub-grid products while also accounting for input errors in the model ([Günther et al., 2019](#)). ERA5 Land's snow depth data represents an instantaneous average of the snow thickness on the ground, excluding snow on the canopy ([Muñoz Sabater, 2021a](#)). To align it with ICESat-2-derived snow depth, ERA5 Land has been resampled at daily resolution and linearly interpolated, excluding permanent snow cover. Additionally, the ERA5 Land monthly ([Muñoz Sabater, 2021b](#)) dataset contributes wind speed and wind direction fields at a height of 10 m above the land surface.

2.4. Validation data

Our validation methods include Airborne Laser Scanning (ALS) surveys, seNorge snow model data, Sentinel-2 satellite imagery, and meteorological station data.. These diverse sources offer both spatial and temporal contexts for evaluating model performance:

- ALS survey: The survey provides high-resolution snow depth data over Hardangervidda ([Melvold and Skaugen, 2013](#)). The ALS encompassed six flight lines apart in 10 km intervals, each extending 80 km in a west-east direction with a crossline scanning width of 500 m (Figure 1). The data were collected between 3-21 April 2008, 21-24 April 2009, and 21 September 2008. During the autumn collection period, the ground was nearly bare condition except for perennial snow patches. The snow depth is retrieved by elevation differencing workflow, excluding inland water areas. This work resampled the ALS to a spatial resolution of 10 m.
- Open accessed snow product: The seNorge ([www.senorge.no](#)) employs a snow model that predicts snow depth based on interpolated precipitation and temperature data (seNorge2018 v23.09) from various meteorological stations ([Salaranta, 2012](#)). It offers daily snow depth maps at a 1 km × 1 km grid resolution and is available through MET Norway's public archive service¹. This work aggregated daily snow depth from seNorge to monthly mean.
- Sentinel-2 satellite imagery: Sentinel-2 imagery (L2A), with a spatial resolution of 20 m, was utilized to confirm the presence or absence of snow in the Lake Møsvatn area, which is located in the southeast of Hardangervidda.
- Meteorological stations: This study compared the snow depth time series with three available meteorological stations in the region. The weather station Sandhaug is located 50 m north of one of the lidar flight lines at an elevation of 1,250 m above sea level (a.s.l.). Another station, Mogen (954 m a.s.l.), is situated directly along a flight line. Additionally, Haukeliseter (990 m a.s.l.) is positioned 1.9 km north of the flight lines. Monthly mean snow depth data were collected from Frost.met.no². Due to harsh observing conditions in our validation area, all station observations carry a median confidence level indicated by a quality flag of 2.

¹Norwegian Meteorological Institute (MET)'s Thredds API (<https://thredds.met.no>) Last access: Sep 11, 2023.

²Frost API provides MET Norway's archive of historical weather and climate data. Last access: Sep 11, 2023.

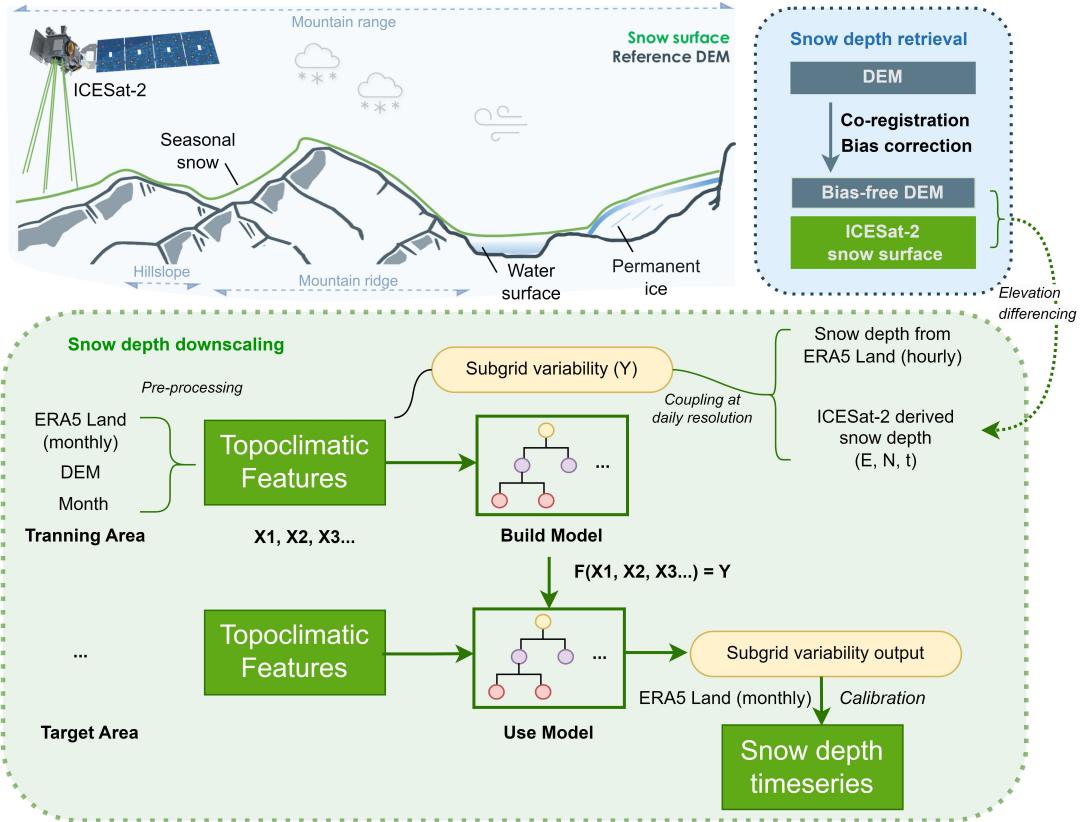


Figure 2: Flow chart of the snow depth retrieval and downscaling. The snow depth is extracted by elevation differencing workflow. Consequently, a tree-structure-based regressor for downscaling is trained and implemented to predict local variability of snow depth in any location and at any time. Satellite graphics source: NASA.gov

3. Methodology

3.1. ICESat-2 Snow Depth Retrieval

Snow depth (SD_{snow}) was derived from ICESat-2 high-resolution elevation measurements through an elevation differencing workflow. ICESat-2 ATL08 data were categorized into snow-on IS_{snow} and snow-free segments based on a snow mask³. Snow-free segments were used for DEM co-registration and subsequent bias correction (Figure 2). This method involved estimating the discrepancy (Δ_h) between the reference DEM (DEM_{ref}) and ICESat-2 Snow-free measurements:

$$SD_{snow} = IS_{snow} - (DEM_{ref} + \hat{\Delta}_h)$$

³ICESat-2 ATL08 offers the reference snow mask from NOAA daily snow cover product (Neuenschwander et al., 2022) but also a brightness flag indicating the existence of snow cover at 100 m resolution.

Moving terrains such as water surface and permanent ice were excluded from the analysis using a land cover reference mask from the ATL08 product (the mask is originally from the Copernicus Global Land Cover dataset at a 100-meter spatial resolution, Neuenschwander et al. (2022)).

The snow depth retrieved at each point (E, N, t) provides spatial and temporal incomplete information, insufficient to fully describe snow dynamics. To overcome this, we integrated snow depth data from ERA5 Land to create a localized dependent variable, denoted as subgrid variability (Y) (Figure 2). This variable represents the deviation from the mean snow depth of the grid at each time and location and is used in subsequent downscaling workflows. Our model predicts this subgrid variability and applies it to ERA5 Land monthly snow depth data to generate snow depth time series at the hillslope scale. However, we found that currently imperfect model outputs required calibration.

3.1.1. Co-registration

The co-registration aligns the DEM with a reference, with ICESat-2 data serving as a highly precise reference. Laser altimetry offered ground elevation control points (Gruber et al., 2012; Yamazaki et al., 2017), making it a valuable resource for co-registration, quality assessment of the DEM (Chen et al., 2022; Li et al., 2022), and bias correction (Magruder et al., 2021). To efficiently address georeferencing errors, we introduced a gradient descent-based co-registration (GDC) algorithm. This GDC method turns the co-registration into an optimization problem, standing out for its ability to handle large datasets with ease, such as DTM1 over national-wide in this study (see Section Appendix A.1). The method is accessible from an open-source repository xDEM⁴.

3.1.2. Bias correction

Bias correction estimates non-stationary discrepancies (Δ_h) between DEMs and ICESat-2 data. DEMs, produced by various sensors and representing different seasons and resolutions, should be cautiously used as a reference ground surface (Hugonnet et al., 2022). The negative skewness between ICESat-2 and DEMs has been noted in many studies (Tian and Shan, 2021; Moudry et al., 2022) and requires correction. Enderlin et al. (2022) has done bias correction based on slope. Magruder et al. (2021) has done a two-dimensional bias correction considering the canopy and slope. Building on these, this study extends the correction to multi-dimensions, paying particular attention to curvature-related bias. A supervised regression model, trained on ICESat-2 ATL08 snow-free measurements, incorporated terrain features, vegetation parameters, and quality metrics from ICESat-2 (see Section Appendix A.2). The same model structure is subsequently employed for the downscaling task.

3.2. Snow Depth Downscaling

The downscaling model aims to derive fine-scale, localized snow depth predictions from coarser-scale input data. For this purpose, we employ the XGBoost⁵, a scalable, robust gradient-boosted decision tree (GBDT) algorithm (Chen and Guestrin, 2016).

⁴The xDEM library support GDC since version 0.0.11
<https://xdem.readthedocs.io/>

⁵The XGBoost (version 2.0.0) library can be accessed at
<https://xgboost.readthedocs.io/>

XGBoost has demonstrated its effectiveness in downscaling tasks, such as satellite gravity (Ali et al., 2023), precipitation (Zhu et al., 2023) and wind speed (Hu et al., 2023). In this study, the XGBoost model's parallel tree-boosting structure captured nonlinear relationships between snow depth subgrid variability and topo-climatic features. Two types of loss functions have been considered in the regression task: (1) using square error (reg:squareerror) as a loss function to estimate the conditional mean of the target variable. (2) using quantile regression (reg:quantileloss) to give probabilistic predictions, such as conditional median (Q50). The idea of using quantile regression with a forest algorithm stems from quantile regression forest (Meinshausen, 2006) and has been popular in probabilistic forecasting (Zhang et al., 2018). The advantage of quantile regression is that it makes no assumptions about the distribution of the target variable. Additionally, it has better resistance to outlying observations.

The regression task considers a comprehensive set of topo-climatic features, including snow depth from ERA5 Land (*sde_era*), east, north, elevation(*h_te_best_fit*), slope, aspect, topographic position index (TPI, Weiss (2001)), curvature, planform curvature (*planC*), profile curvature (*profC*), wind-aspect index (W_{uf}) and month. These features offer valuable insight into the physical conditions governing snow dynamics. For example, the slope characterizes the steepness of the terrain surface. The planform curvature and profile curvature further classify the surface by convergence and divergence, which serve different positions for snow accumulation. We compute these terrain attributes using xDEM (Xdem contributors, 2021) based on DTM10 with Zevenbergen & Thorne algorithm (Zevenbergen and Thorne, 1987). These calculations yield 10 m resolution features, consistent with the model's output resolution. The wind fields are from ERA5 Land (monthly) so the temporal resolution is a month. Details regarding the pre-processing of TPI and W_{uf} are elaborated upon in the following section. In addition, we assume the stationarity of ERA5 land performance, allowing us to apply the downscaling model trained on snow observations from 2018 to 2022 to other periods.

3.2.1. Topographic position index (TPI)

The TPI is a metric used to access slope position and classify different landforms. It quantifies the difference between the elevation of a central pixel

and the average elevation of its neighboring pixels (3 x 3 pixels). A TPI value of zero or near zero indicates a flat or nearly continuous slope. Positive TPI values suggest that the central pixel is significantly higher than the surrounding areas, forming a ridge or hill. Conversely, negative TPI values indicate that the central pixel is notably lower than its neighboring areas, signifying a valley. TPI has proven effective in predicting snow distribution in alpine environments (Revuelto et al., 2014). To represent landforms at different scales, this study used two additional indices: tpi_9 (calculated in 9 x 9-pixel windows, equivalent to 90 m x 90 m) and tpi_27 (using a window size of 27 x 27 pixels, equivalent to 270 m x 270 m).

3.2.2. Wind-aspect index

The Wind-aspect factor (W_f) (Bennett et al., 2022; Dvornikov et al., 2015) serves as a proxy for snow accumulation and erosion on topographic obstacles. It assigns positive values on the leeward side and negative values on the windward side of these features. The relationship between wind and aspect follows a cosine function that ranges from -1 to 1 for any prevailing direction (see Figure 3):

$$W_f = -\cos(\text{aspect} - \text{dir}_{wind})$$

Where, dir_{wind} is defined by the direction of the wind origin with northerly wind blowing from the north to the south and referred to as 0° . This study further divides W_f into leeward factors and windward factors, multiplied by wind speed to the power of three (Figure 3 c) to capture the accumulative index of wind redistribution for each water year period.

$$W_{uf+} = \sum W_{f+} u_{wind}^3$$

$$W_{uf-} = \sum W_{f-} u_{wind}^3$$

Where u_{wind} is the monthly average wind speed at 10 meters from ERA5 Land. The accumulation begins in September from zero until the next September when the value reaches its maximum. The value does not accumulate when the monthly average snow depth falls below 0.1 m during the annual cycle.

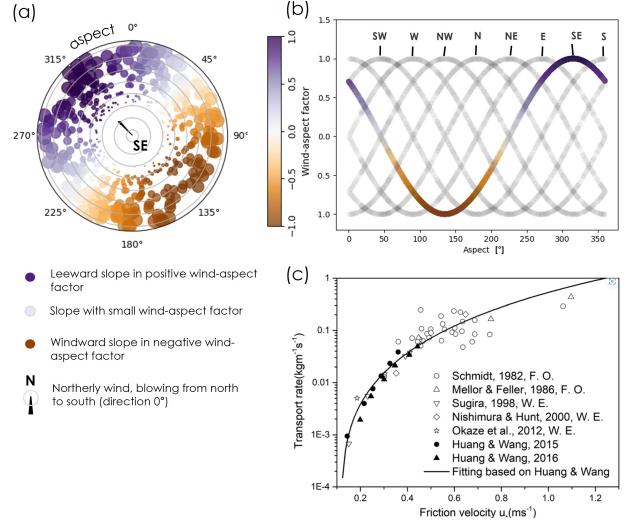


Figure 3: Quantifying the relationship between wind, aspect and snow redistribution. The prevailing wind, e.g. from NE, SE result in negative value on the windward side and positive values on the lee side (a). Eight cardinal directions are plotted but the function works on any wind direction (b). The aeolian research suggests that as wind speed increases, the transport rate increases exponentially with a power of 3 (adapted from Freudiger et al. (2017); Li et al. (2018))(c).

3.2.3. Calibration

In regression tasks, XGBoost predicts the conditional mean of the snow depth for a given set of conditions. This approach often leads to conservative predictions that are close to average values when certain critical variables that influence extreme snow depths are not fully included in the existing feature set. As the model is trained to minimize overall error, it tends to stay within its “comfort zone” around average conditions. As a result, extreme values are less represented, leading to a ‘flat’ distribution. The model still provides reasonable predictions in terms of ranking but suffers from scaling bias of the distribution. In different scenarios, there may be additional causes of scaling errors, which will be further discussed in Section 5.1

To address this issue, we propose quantile mapping to adjust the scaling bias. The scaling factors for adjustment are calculated based on ALS snow depth from 2009 (Section Appendix A.3 provided the details and model output without applying adjustment). This adjustment does not change the rank correlation but adjusts the scaling bias. We have not yet tested if our calibration is valid in other catchments.

3.3. Model Interpolation and Assessment

To interpret the contribution of topo-climatic features in our tree-based models, we employed SHAP values (Lundberg et al., 2020), an interpretability metric derived from game theory. The SHAP value attributes the contributions of each feature to individual predictions. A positive (negative) SHAP value indicates that the presence of a feature pushes the model's output higher (lower). The magnitude of the SHAP value signifies the difference between the actual prediction and the average prediction. The bigger the SHAP value, the stronger the feature's effect. SHAP values adhere to an additive property, where the sum of all SHAP values for a given prediction, when added to the base prediction, equals the actual model output. The SHAP value splits the contribution to correlated features.

The success of the prediction will result in a perfect match in data distribution and data ranking correlation. To evaluate the performance of the downscaling model, we employ four key statistical metrics. RMSE and R^2 score evaluate the overall accuracy and fitness of the model. Spearman's rank correlation coefficient (ρ) has been used as fidelity metrics, with a high ρ indicating good similarity in structure. The Kolmogorov-Smirnov D statistic (KSD) quantifies the degree of probability matching, with KSD = 0 indicating a perfect match.

Additionally, to quantify the model's ability to capture the spatial heterogeneity of snow distribution, we use variograms to assess the downscaling model's ability to capture the spatial characteristics of snow depth. The semi-variance (γ) is a measure of spatial variability, calculated for pairs of observations as half the average squared difference between values separated by a specific lag distance (l). Mathematically, for lag l , the semi-variance is defined as (Oliver and Webster, 2014):

$$\gamma(l) = \frac{1}{2N(l)} \sum_{i=1}^{N(l)} (z(x_i) - z(x_i + l))^2$$

where $z(x_i)$ represents the snow depth at location x_i . The spatial correlations of snow distribution are observed at different scales, which requires fitting a sum of models at multiple ranges (introduced in Rolstad et al. (2009) for glaciology applications and well explained in Hugonet et al. (2022) for DEM uncertainty analysis). This study fits variogram models to the sum of the spherical model

and Gaussian model from short to long ranges for more reliable estimates.

4. Results

The retrieved snow depth for mainland Norway, spanning from October 2018 to October 2022.10 is available on Zenodo. Evaluating the accuracy of high-resolution, nationwide snow depth retrieval is challenging due to the scarcity of distributed observations at this scale. Therefore, our results are primarily presented for downscaled snow depth at the designated validation area. We begin by interpreting the downscaling models, followed by comparing snow maps and profiles from the downscaling output with various datasets, including ALS snow surveys, seNorge simulations, meteorological station data and Sentinel-2 imagery. These comparisons of variability are conducted across three different scales. In mesoscale, we aggregated the data into a 100 m x 500 m grid cell. In the microscale and site scale, we used the native output resolution of 10 m. To ensure equitable comparisons, the validation datasets were aggregated to the same resolutions.

4.1. Interpreting the Downscaling Model

The most informative relationships between the input features and the predicted subgrid variability are listed in Figure 4 in descending order. Elevation ($h_{te_best_fit}$) emerges as the most significant factor, with higher elevations positively influencing subgrid snow depth. When snow depth from ERA5 Land (sde_era) is high, both very thin snow depth and thick localized snow depth exist. Notably, relatively thicker snow is estimated by the downscaling model in concave terrain with an east-facing slope. Conversely, thinner snow is associated with convex and west-facing slopes (Figure 4 b,c). The aspect-dependent preferential accumulation is more apparent in the western mountainous regions, where more diverse subgrid variability is exhibited, contrasting with the relative flatness in the open eastern areas (Figure 4 e). The northern part of the study area appears to be overestimated in snow depth, while the southern part is underestimated (Figure 4 d), a pattern we will further investigate using ALS data in the following section. The positive wind aspect factor ($wf_positive$) contributes positively to preferential snow accumulation. However, the negative wind aspect factor ($wf_negative$) shows a much weaker contribution. The wind aspect

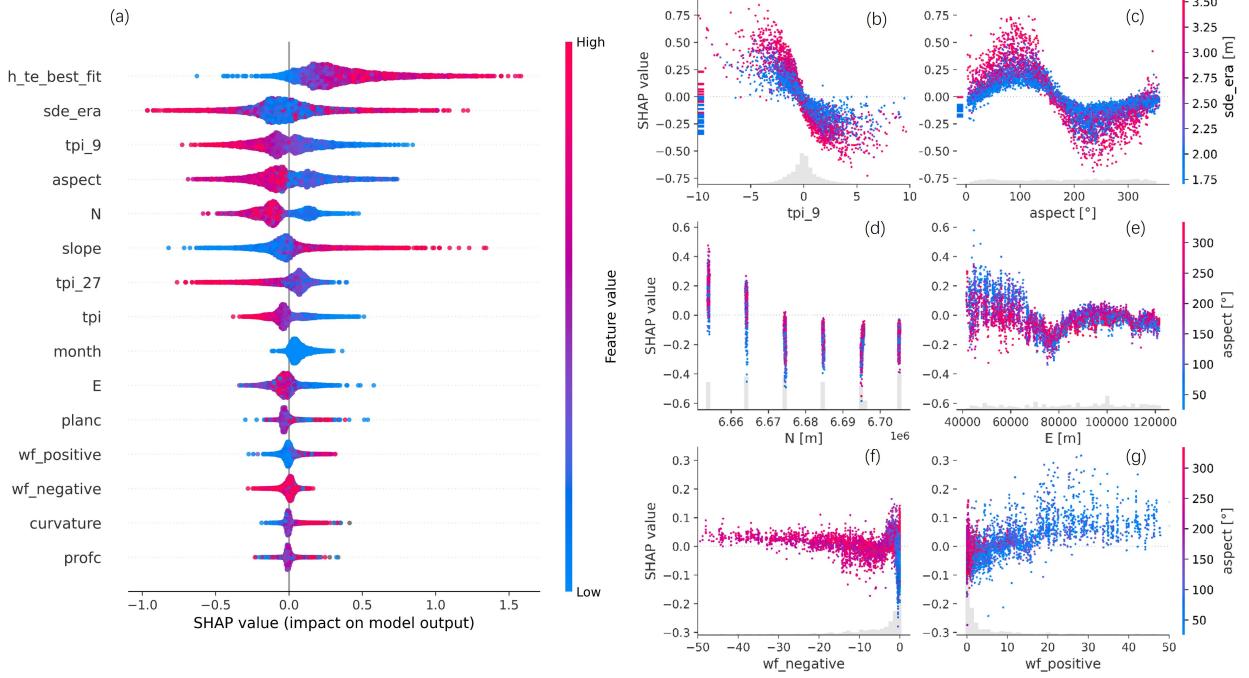


Figure 4: SHAP interpretation for the downscaling model. Panel (a) ranks features by their impact on subgrid variability. Subsequent panels (b-g) depict SHAP dependence plots, illustrating how each feature influences model predictions at the validation site. Each dot represents an individual instance of subgrid variability, providing insights into feature-specific margin effects.

factor at zero, indicative of low wind speeds, shows much variability, indicating unexplainable variance (Figure 4 f,g). However, SHAP cannot distinguish the contributions from correlated features. For instance, slope, while a strong feature, might partly be due to the influence of its correlated counterpart, such as elevation and curvature. The phenomenon of having less snow on very steep slopes (e.g. $>50^\circ$) is not observed.

4.2. Mesoscale Snow Depth Variability

At the mesoscale, variations in snow depth are primarily influenced by local topography and wind redistribution. Figure 5 provides such an example in April 2008. The ALS snow survey and downscaled output are aggregated to the resolution of 100 m x 500 m for each dot. From west to east with the increase of distance to the coastline, all datasets follow a decreasing trend from snow depth above 6 m to around 2 m. The downscaled snow depth effectively captures variability at different hillslopes in close proximity to the ALS snow survey. However, as shown in Figure 5 a,b,c,d,e, seNorge has about 40 % overestimation in the western mountain ridge but

not in the eastern mountain summits. Melvold and Skaugen (2013) attributed the overestimation to the reliance of the seNorge model on weather stations located in low-elevation populated areas. SeNorge has a peak snow depth proximate to the mountain crest, with much less variability on the two sides of the ridges. Among the six flight lines, the downscaling model (based on DTM1, after calibration) scored R^2 between 0.71 and 0.88, and RMSE 0.76 m to 0.44 m, outperforming seNorge, which yields R^2 values between 0.27 and 0.55 and RMSE values between 1.36 m and 0.91 m.

Upon closer inspection, ALS data exhibits many spikes on the east-facing slopes instead of the crest, likely resulting from wind redistribution and gravity processes. While our downscaling model acknowledges these factors (Figure 4 c,g), but does not fully capture them. A more detailed analysis at the microscale will be provided to further explore these differences. Meanwhile, ERA5 land, as a fundamental input to our downscaling model, has a systematic bias in different areas. From north to south, the ERA5 is getting from overestimation to underestimation (Figure 4 e). On the first flight

line in the eastern open area, the overestimation by ERA5 Land is temporally consistent, leading to a good match with the ALS data. However, 10 km southward on the second flight line, the snow depth is much lower than expected (or the experience from the training period), causing the downscaled model's predictions to deviate from the ALS data.

Additionally, we found that global DEMs, such as Copernicus DEM GLO-30 (GLO30), yield results at the mesoscale that are comparable to those obtained using DTM1 and DTM10. The performance of downscaling models based on different DEMs was assessed in Figure 6. The performance metrics, including R^2 , Spearman's ρ , RMSE, and KSD, ranked the DEMs as follows: DTM1, DTM10, FAB, and GLO30, with ALS as a perfect match. DTM1's higher resolution provides clear advantages, and the removal of vegetation in FAB, while not critical in our treeless validation area, still positively impacts the model's performance. The performance of aggregated snow depth is not sensitive to the absolute precision of reference DEMs. This finding supports the possible applicability of our workflow in treeless conditions.

Figure 7, which displays the snow distribution map for April 2008. The linear interpolated snow depth from ERA5 Land is uniformly distributed across the area, representing a large-scale variability from climate. After downscaling, the snow depth exhibits increased variability that aligns with the topography (a, b). As expected, seNorge indicates a decreasing trend in snow depth from west to east but overestimates the depth in the western mountains (in April 2008, not in April 2009, Section Appendix A.4). Notably, it lacks the variability observed on different sides of the mountain ridge compared to the downscaled snow depth, even at the same 1 km resolution. The downscaled output captures the smooth transition in snow depth from valleys to mountain crest (c, d).

4.3. Microscale Snow Depth Variability

Without any aggregation, Figure 8 presents a comparison between the ALS data 2008 and the downscaled snow depth at a 10-meter resolution. The examined region, located in the western part of flight line 2, exhibits distinct microscale landforms, with sheltered depressions hosting thick snow patches (> 8 m) and wind-exposed hilltops featuring thin snow covers (close to 0 m). The transect line (marked in white) of downscaled output across this varied

terrain shows a Spearman ρ of 0.77, an R^2 value of 0.34 and a KSD 0.12. In this area, the downscaling model has an RMSE of 1.33 m and a mean deviation of 0.28 m. The lower metrics observed at the microscale compared to the mesoscale highlight the scale-dependent nature of spatial variability in snow depth.

The variogram analysis, comparing ALS data with downscaling model outputs, reveals critical insights into the spatial correlation of snow depth at the microscale. The variograms are calculated based on all six ALS validation strips. As illustrated in Figure 9, the empirical variogram showed an initial rapid increase in semi-variance with distance, suggesting strong spatial dependence at short ranges. Fitted variograms show that ALS data has 79% variance correlated in 95 m lag, while downscaling models have over 60% variance in 108 m lag, indicating the resolution is a key of variability at the microscale. When the lag distance extends to 1,300 m, the semi-variance for both datasets approaches the sill. Overall, the downscaling model exhibits 13% less variance than ALS data, suggesting a slight underestimation of spatial variability by the model. The 95 m lag is particularly relevant, as it approximates the wavelength of depression features commonly observed in the area, while the 1,300 m distance aligns closely with the typical size of hillslopes in this region, as depicted in Figure 8 (a) and Figure 5 (a).

4.4. Temporal Variability of Snow Depth

In temporal analysis, the output of the time series has not been bias-adjusted because this analysis focuses more on the bias and the possibility of using weather stations for calibration. We used quantile regression in the downscaling model to gain insights into the uncertainties of downscaling and validate the representativeness of weather stations. The quantile regression model captures the central tendency for different groups with similar topoclimate conditions. The predicted value at the median quantile (Q50) serves as a reasonable estimate of the central value for each group. The plot also includes Q25 and Q75 in yellow shadow, with Q75 showing a good match with the weather station data. Figure Figure 10 provides a comparison of time series data at three stations: Sandhaug, Mogen, and Haukeliseter. The R^2 values of Q50 range from 0.23 to 0.84, indicating varying degrees of performance. The primary bias observed is during the peak snow

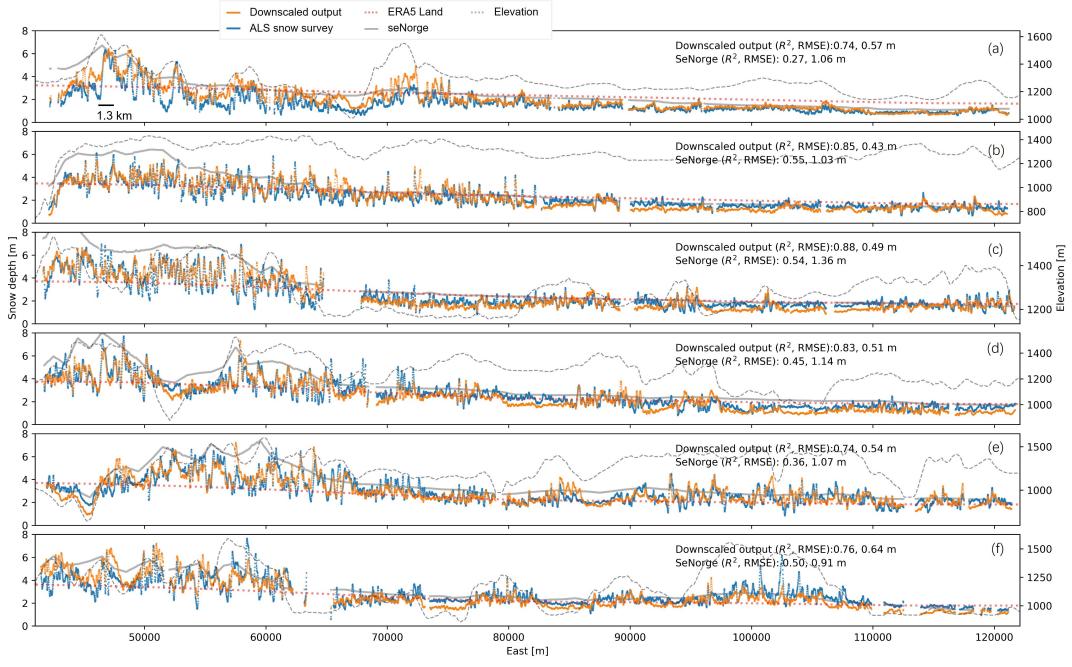


Figure 5: Snow depth profiles across Hardangervidda in April 2008. This figure showcases a mesoscale analysis of snow depth along six ALS flight lines, spanning north to south (panels a-f). Data points, aggregated to represent $100\text{ m} \times 500\text{ m}$ cells (because of the ALS survey's 500 m transect width).

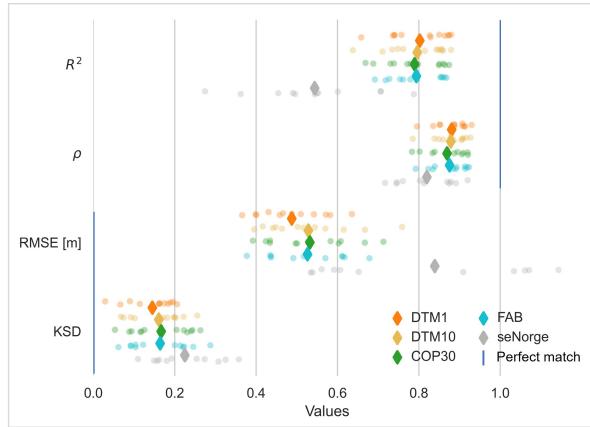


Figure 6: Comparative performance of downscaling models using various DEMs. This statistical summary evaluates model accuracy at the mesoscale, utilizing ALS Snow Survey data from 2008 and 2009. Each dot symbolizes the performance on a specific validation strip, with diamonds indicating average performance across all strips. The analysis benchmarks the effectiveness of different DEMs in capturing mesoscale snow depth variability.

season, with consistent magnitude across each site. For a few years with low ERA5 snow depth, the model gives a closer estimation e.g. 2019 (Figure 1). The possible reason is the training dataset may not contain enough samples in the peak season.

The right panel illustrates the interquartile range (IQR) within a square area measuring 500 m in length. It represents the spread of snow depth predictions by the downscaling model. The red area indicates more typical topoclimate conditions that are easier to predict, while the yellow area shows greater variance under given conditions. In terms of IQR, Sandhaug and Haukeliseter may have better locations than Mogen.

4.5. Validating Snow Occurrence

Figure 11 provides a visual comparison between the model output and Sentinel-2 imagery, showcasing the snow distribution for June 2020. This comparison assesses the model's accuracy in predicting the occurrence of snow during the rapid melt period. The downscaled snow map from our model shows a high degree of agreement with the satellite snow extent, capturing snow patches that align with topographical features (Figure 11 c,d).

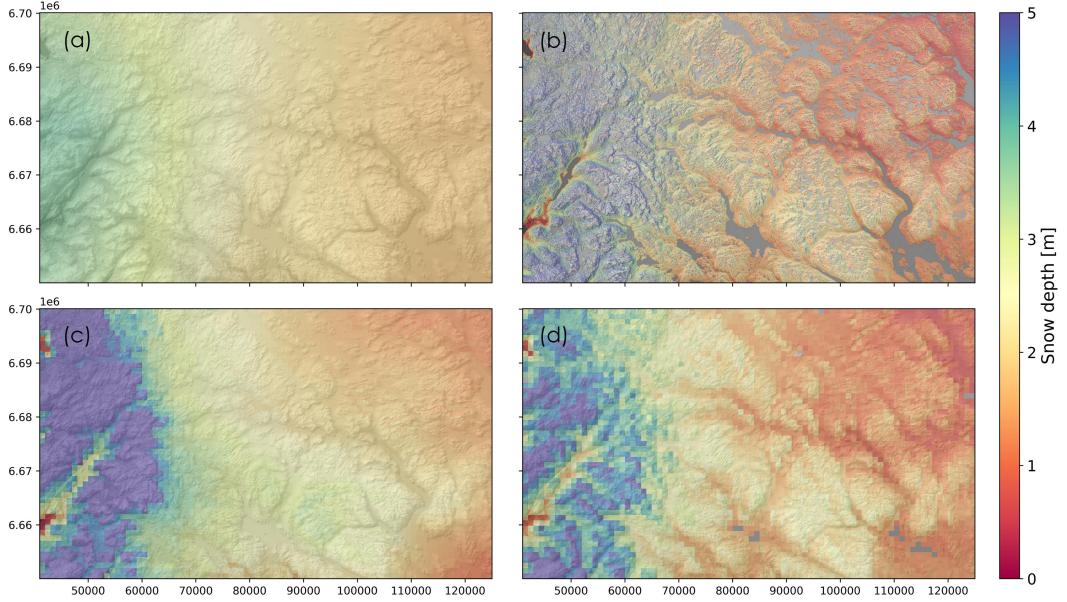


Figure 7: Spatial Distribution of Snow Depth in Hardangervidda, April 2008. (a) Snow depth from ERA5 Land in 10 m resolution (after linear interpolation). (b) Downscaled snow depth output in 10 m resolution. (c) seNorge snow depth at 1 km resolution. (d) Downscaled snow depth aggregated at 1 km resolution.

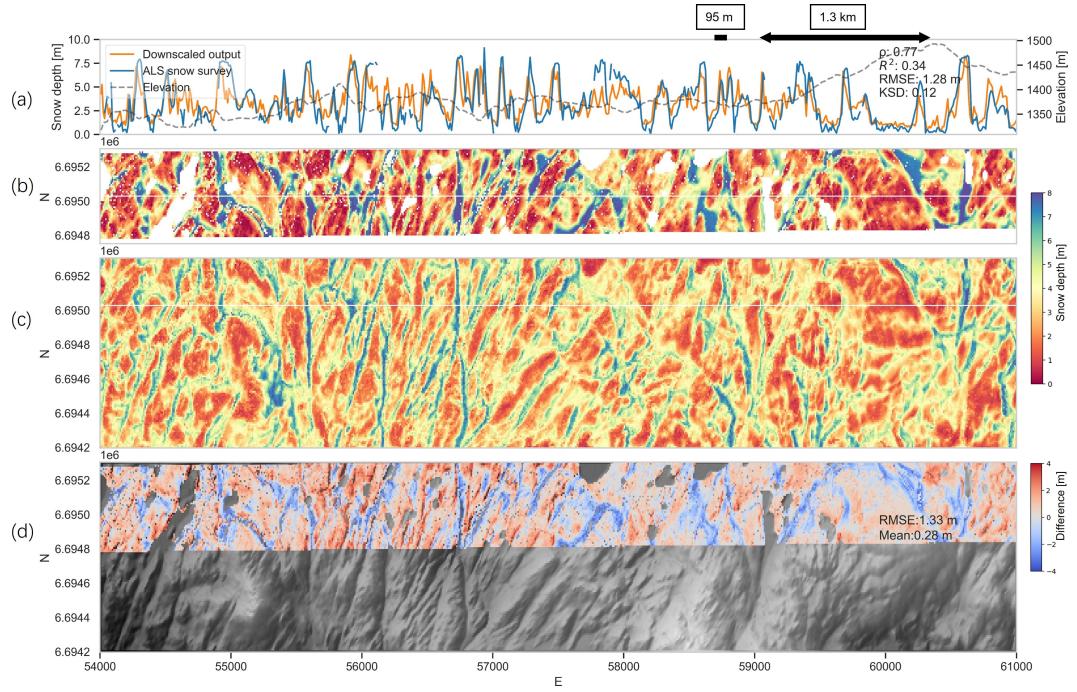


Figure 8: Microscale Snow Depth Comparison in Hardangervidda, April 2008. Panel (b) shows the ALS snow survey validation strip, revealing significant snow depth variations. Panel (c) illustrates the downscaled snow depth from the DTM1 model, capturing most of the observed fidelity but with a low R^2 value of transect line (a). Differences between the ALS data and model output are highlighted in Panel (d). The validation strip is chosen for its representation in the original ALS paper Melvold and Skaugen (2013)'s Figure 3.

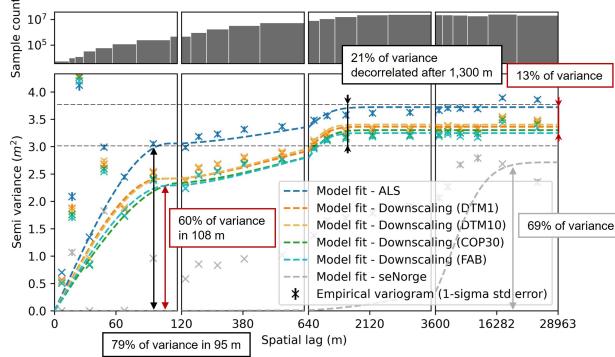


Figure 9: Spatial correlation comparison using variograms. The analysis is based on six ALS validation strips and downscaling model outputs at a 10 m resolution. The empirical variogram shows the degree of spatial correlation in snow depth. The variograms are fitted to the sum of the spherical model and Gaussian model from short to long ranges. Note that the variogram's limited capture of variance in the transverse direction is due to the ALS strips' constrained width.

Panel (e) and (f) reveal that east-facing slopes retain more snow compared to west-facing slopes, a pattern that our model successfully captures. Nonetheless, closer inspection also reveals minor discrepancies, especially in the finer details of snow distribution. These differences are partly attributed to the temporal mismatch, with the Sentinel-2 image taken on June 24, 2020, and the model output representing a monthly mean for June. Notably, the model tends to overestimate snow coverage on specific terrain, such as steep slopes (indicated by the black arrow) and underestimate it in flatter areas (indicated by the white arrow). For a dynamic view of snow occurrence variability, please refer to the supplementary video.

5. Discussion

This study achieved snow depth retrieval from ICESat-2 ATL08 using four different DEMs and subsequently utilized this data in the downscaling of ERA5 Land snow depth. Additionally, the performance of the downscaled snow depth in Hardangervidda, Norway, was analyzed using ALS data, seNorge, Sentinel-2, and weather station observations. We will discuss the challenges, limitations and possible solutions in developing our methodology. Then we comment on how our downscaling model quantifies subgrid variability comparable with validations and other studies.

5.1. Challenges of deriving snow depth from ICESat-2 ATL08

This study confirmed previous efforts on retrieving snow depth from ICESat-2 and further developed a method with an efficient Gradient Descent Co-registration algorithm and a bias correction model based on open-accessed DEMs. The efficient co-registration algorithm offers large training samples. By interpreting our bias correction model based on different DEMs, we noticed that the uncertainties of elevation differencing workflow are highly up to how good we estimate the discrepancy (Δ_h) between reference DEMs and ICESat-2 altimetry data.

5.1.1. Skewness and resolution issues

A critical issue identified with ATL08 and DEMs is negative skewness (Tian and Shan, 2021; Enderlin et al., 2022; Moudrý et al., 2022), which ICESat-2 tends to underestimation surface height under certain conditions. Moudrý et al. (2022) attributed the error to the presence of clouds that increasing atmospheric scattering effects lead to underestimated terrain height. Furthermore, the 100 m segment length of ATL08 is considered insufficient for accurately mapping steep and rugged terrain (Besso et al., 2024). Two issues on ATL08 are closely related since the slope naturally translates horizontal shift (i.e., footprint or geolocation error) into vertical shift (Deems et al., 2013). Therefore, Enderlin et al. (2022) recommended only using ATL08 in unglacierized mountain areas with relatively low slopes and sparse vegetation cover. The consequences of such skewness lead to underestimation of snow depth and a large component of ‘negative snow depth’, making less measurement available. Excluding all negative snow depth resulted in the uneven spread of training samples. In this study, we corrected the negative skewness by bias correction model. Through analysis of the model-captured bias pattern, this study highlights the curvature-related bias because of resolution (Figure A.13). The correction solely based on the slope might be not enough. Further, curvature-related bias reveals that the ATL08 offers elevation product in a resolution better than Copernicus GLO30 when using $h_{te_best_fit}$ of the subset segment (Figure A.13).

5.1.2. Limitations and applicability

The potential limitation is the residual skewness on certain terrains. This is particularly relevant in validation areas with a long snow cover duration. These areas were unlikely to be completely

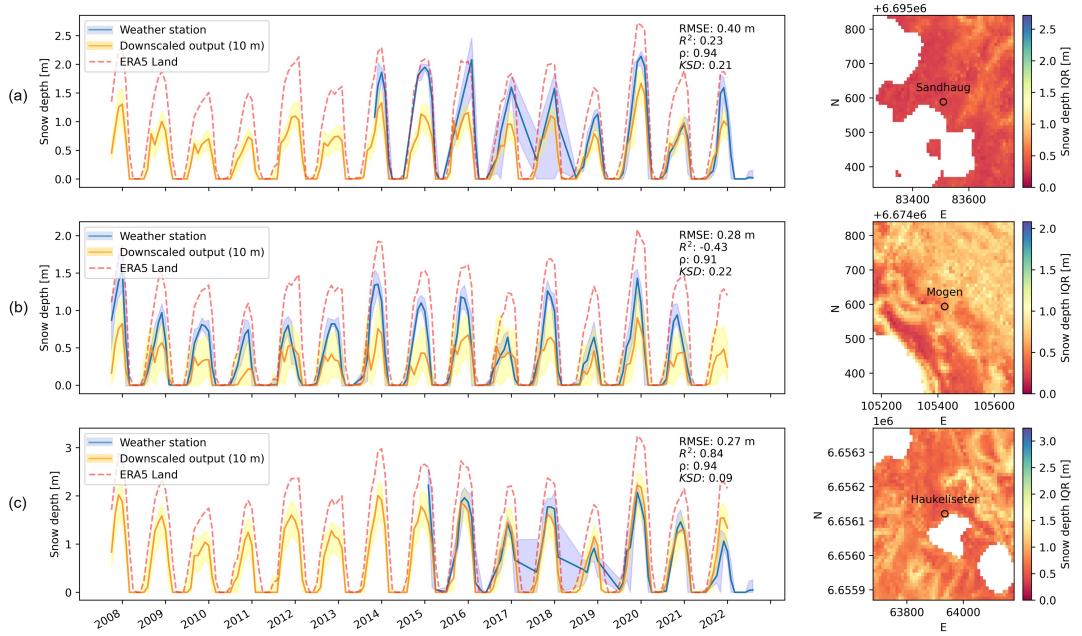


Figure 10: Time-series validation of snow depth across weather station. Presents monthly minimum and maximum values (blue shadow) and the downscaled snow depth at the 25th and 75th quantiles (yellow shadow) for Sandhaug (a), Mogen (b), and Haukeliseter (c). The rightmost plot is the IQR for all seasons to demonstrate the snow depth variability of the surrounding area. Note: Haukeliseter and Sandhaug data availability start from 2015 and 2014, respectively, and some years have incorrect measurements. The inland water is excluded from the map.

free from snow when DEMs were obtained. Furthermore, there are also limited snow-free segments available in such locations. Consequently, the bias correction may not be able to fully remove negative skewness, resulting in an underestimation of snow depth in such an area. On the microscale, the residuals are highly associated with a preferential deposition pattern, indicating that the western mountain area was unlikely to be snow-free when the reference DTM1 was obtained (Figure 8 d). This is evident in the downscaled output without applying calibration (Figure A.15; Figure A.16 e).

Additionally, the use of Copernicus GLO30 for estimating snow depth was considered imprecise (Enderlin et al., 2022). However, we do not directly validate sparsely retrieved snow depth but rather input it into the downscaling model and subsequently validate the downscaled outputs. Encouragingly, our results demonstrate that the Copernicus DEM and FAB perform comparably with DTM1 and DTM10 at mesoscale in treeless areas. This work credits the enhanced performance of GLO30 and FAB to removing skewness through bias correction. The downscaling model is not sensitive to absolute pre-

sion on single-point measurements but skewness on the group of measurements. Other parts of Norway, such as the Finnmark plateau have low skewness on all DEMs (not shown in this paper). However, it is cautious about using GLO30 and FAB in vegetated areas, where the vegetation is hardly been fully removed. The use of lidar-based DTM1 in forested areas also requires cautious validation.

5.2. Challenges of downscaling snow depth

Our approach leaves a significant amount of variability unaccounted for, particularly during periods of high snow accumulation, unless calibration is applied. This section analyzes the reasons for these scaling errors and discusses potential solutions from a downscaling perspective.

5.2.1. Impact of the imbalanced training dataset

Considering the scaling properties of snow depth's spatial distribution (Melvold and Skaugen, 2013; Mott et al., 2018), the resolution of a sensor is critical to address the subgrid variability. Our training dataset, derived from the ICESat-2 ATL08 elevation product, is constrained by its inherent resolution. Given its native footprint size of approximately 15

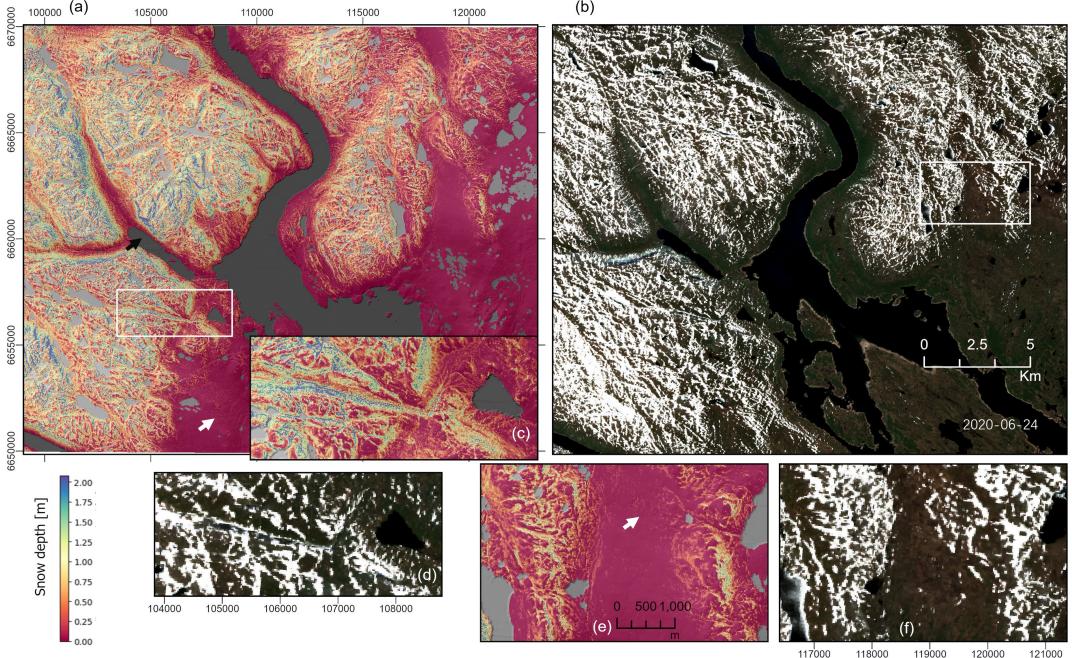


Figure 11: Validation of snow occurrence using Sentinel-2 imagery. Panel (a) illustrates the model’s snow cover distribution. Panel (b) captures the Sentinel-2 image on June 24th, 2020, over the Lake Møsvatn region. Panels (c) and (d) zoom in on selected areas to show detailed discrepancies in snow distribution. Panels (e) and (f) highlight the model’s accuracy in depicting the differential snow retention on east-facing versus west-facing slopes. The overestimations on steep slopes and underestimations in flatter terrains are annotated respectively with black and white arrows. The supplementary video provides a dynamic visualization of this figure, which can be accessed at <https://youtu.be/My1fSNGNx4>.

meters, ATL08 data may average out finer-scale variability. The sampling pattern of ATL08, primarily in the North-South direction with an interval spacing of approximately 3.4 km, further contributes to sparse spatial sampling. This sparsity is compounded by the 92-day revisit period of the satellite, leading to gaps in seasonal data coverage. The temporal and spatial sparsity of the dataset likely contributes to the seasonal underestimation observed in our time series analysis (Figure 10).

Furthermore, ICESat-2’s elevation data are derived using a ground-finding algorithm that utilizes a probability distribution function (PDF) of reflected photons. This process tends to disadvantage rugged terrains, which are less likely to be fully sampled under standard quality control criteria, despite their propensity for higher snow accumulation. Given these factors, it is reasonable to infer that the snow depth data obtained from ATL08 may exhibit less variability than actually present in the natural environment, particularly in terms of capturing extreme snow depth values. As a result, the model learns

incorrect prior information from the imbalanced training dataset.

5.2.2. Uncertainties input from ERA5 Land

In this study, the combination of subgrid variability from the downscaling model and the temporal & spatial variability from ERA5 Land is a significant step forward from traditional statistical models. However, as noted by another study (Pflug et al., 2021), the snow depth of ERA5 Land may not exhibit stationarity relative to the baseline (Figure 4 e). The baseline period of 4 years (2018 to 2022) used for training is relatively short, considering that climate studies often necessitate datasets spanning over a decade to accurately represent climate norms (Arguez and Vose, 2011). However, due to the maximum designed lifetime of ICESat-2 being 7 years, obtaining such a long dataset is impractical.

Furthermore, the representativeness of ERA5 Land’s snow depth varies significantly from in-situ snow observations. For example, when grids cover glaciers, they report much higher snow depth val-

ues, while grids covering large water areas tend to report lower values. Although we have excluded estimations above moving terrains in our final results due to a lack of reference DEM, the estimations are located in a grid with substantial moving terrains, whose variability is still interfered with by neighboring land cover. In such grids, the variability of correction is primarily driven by snow representatives, and topo-climatic features might play a lesser role. Consequently, we utilized coordinate features (N, E) to account for this spatially dependent bias. This underscores the importance of dense sampling for the current model.

5.2.3. Performance of Downscaling model

Tree-structure models perform well in capturing nonlinear relationships and providing reasonable predictions for snow depth. XGBoost becomes a more powerful tool when using SHAP to interpret model behavior. However, deterministic regression models primarily yield conditional means and may not adequately represent extreme events, especially when key features do not fully explain variations. This often results in underfitting as the models strive to minimize overall prediction errors. For instance, the gravity process of snow redistribution results in extreme values. Terrain parameters alone fall short in locating these avalanche events but tend to predict above-average snow depth in such terrain.

A notable source of unexplained variance in our model is the limited representation of wind fields, derived from ERA5 monthly data. The spatial-temporal resolution of these fields is insufficient for capturing detailed wind-topography interactions. Despite our efforts to include these effects through a wind-aspect index, currently, the model inadequately captures the scouring effect, as indicated in Figure 4. Additionally, the model does not account for the impact of wind on the energy balance between the snow surface and the atmosphere, involving sensible and latent heat exchanges (Mott et al., 2017), a factor that plays a crucial role in snow dynamics.

To address these issues, we employed a quantile mapping calibration based on ALS data from 2009. This method serves as a straightforward technique to incorporate prior information into the model. Further improvements could include the integration of domain knowledge either in the feature space, using methods like the covariate shift approach KLIEP (Sugiyama et al., 2007), or in the

label space, focusing on removing the residual bias by calibration. Up to date availability, there are downscaling-calibration schematics (Duan and Bastiaanssen, 2013; Zhu et al., 2023) that have been developed for satellite precipitation products. Many studies pointed out that subgrid snow depth follows gamma distribution (Skaugen and Melvold, 2019; Gisnås et al., 2016). However, the sole use of gamma regression does not make progress in handling skewed extreme values in our experiment (not shown).

The lack of high-resolution wind fields makes it challenging for statistical models to accurately reproduce the variability resulting from wind processes. Incorporating more informative features that describe the physical processes of wind-driven processes is a promising solution. Future research could benefit from downscaling wind fields (Fiddes et al., 2022) to enhance the model's ability to capture the variability driven by wind. Meanwhile, as an explainable machine learning model, it can help distinguish the impact of different physical processes on snow depth (Grünwald and Lehning, 2011).

5.3. Subgrid variability in the validation area

The spatial variability is highly scale-dependent because the variability is from the processes at different scales. In our validation area, seNorge does not show variability on different sides of the hillslope (Figure 5) because it is a product interpolated based on precipitation gradients and temperature lapse rate, which are dominated by elevation. If seNorge well represents regional orographic precipitation, it can explain 69% of the total variance at a spatial lag after 16,000 m (seNorge, Figure 9). This distance corresponds to the maximum local relief from the mountaintop to the lake surface (Figure 5).

Snow's spatial variability at smaller scales is primarily driven by wind-driven processes over hillslopes, which can be categorized into three distinct types: (i) enhancement of snowfall by local airflow fields and cloud formation, (ii) preferential deposition of snowfall, and (iii) snow redistribution through saltation and suspension (Mott et al., 2018). As suggested by Mott et al. (2018), a resolution of 50 meters or finer is required to capture the spatial variability associated with these processes. It is a challenge to distinguish the effects of each process separately, as they interact and occur simultaneously. Many studies noticed the scale break smaller than 100 m, with a stronger autocorrelation before

the scale break than beyond (Trujillo et al., 2007; Mott et al., 2011). Such spatial correlation can be attributed to the wavelength of wind-driven snow accumulation features (Mott et al., 2018). In this study, the scale break of ALS data is 95 m, accounting for 79% variance in such a short distance. The downscaling model has scale breaks at 108 m (a pixel wider for the worst scenario), further confirming that the ICESat-2 ATL08-derived snow depth has a resolution better than the segment size of 100 m. This study also finds a scale break at 1,300 m by fitting the empirical variogram at two different ranges, which corresponds to the wavelength of ridges in this area.

A key advantage of the downscaling model is its incorporation of temporal variability (Figure 10). This study's downscaling models include temporal variability from large-scale snow depth, wind-aspect index, and month. In contrast, statistical models that solely depend on topographic features for variability often overlook temporal aspects. Moreover, the ICESat-2 provides more than just snow depth information. They also include valuable data on the occurrence of snow cover (Figure 11). ICESat-2 incorporates a snow flag based on ground brightness and utilizes the NOAA daily snow mask as a reliable label source. This contributes to more accurate results in determining the presence of snow cover. The ability to predict snow-off is rarely addressed by statistical spatial distribution models.

The downscaled snow depth derived from ICESat-2 holds potential for various applications in future research. Many studies noticed the impact of snow depth on the thermal regime of permafrost but fall short in snow depth at hillslope scales (Gisnås et al., 2016). Accurate, cost-efficient subgrid parameterization for snow depth also leads to improvements in discharge modeling (Helbig and van Herwijnen, 2017). We also expect that these methods will aid in risk assessment for avalanches and the study of ecosystems in snow-covered regions.

6. Conclusion

ICESat-2 has captured subgrid variability in snow depth during its overflights of the snow surface. Retrieving and utilizing this information from changes in surface elevation is a challenging task. This study presents a workflow that retrieves and integrates this variability into ERA5 snow depth, utilizing a scalable and explainable downscaling model.

- (1) There are very few snow depth observations available in remote areas, making the subgrid variability rarely addressed. In this context, ICESat-2 stands out as a valuable data source. ATL08 efficiently offers high-resolution measurements by co-registration and bias correction workflow.
- (2) This study trained a downscaling model with sparsely retrieved snow depth, effectively enhancing large-scale snow depth products to finer hillslope scales. This model captures both spatial variability and monthly temporal variability. Post-calibration is currently necessary to compensate for the model's underrepresentation of extreme values.
- (3) This study validates the downscaling model's performance, particularly in predicting snow occurrence and achieving R^2 values ranging from 0.74 to 0.88 (after calibration) for snow depth at mesoscale levels. Notably, the downscaling model is sensitive to skewness but less sensitive to precision at individual measurements. Therefore, similar results can be obtained when using the global DEM Copernicus GLO30 in our validation area.
- (4) Accurate prediction of snow depth at microscale level (< 50 m) continues to pose a significant challenge. These challenges stem from the use of non-snow-free reference DEMs, simplified wind processes, data imbalances, and temporal inconsistencies in ERA5. Future research can downscale wind fields, and develop more adaptable calibration methods for varied scenarios.

Our research contributes to a data-driven solution to enhancing the understanding of spatial and temporal variations in snow depth at the hillslope scale, especially in mountainous regions. While the validation is specific to Hardangervidda in southern Norway, the methodology holds the potential for wide-ranging applicability. Further validation efforts, particularly in vegetated areas, are imperative.

Funding

This work was supported by the Research Council of Norway (Grant number 325519).

CRediT authorship contribution statement

Zhihao Liu: Conceptualization, Methodology, Investigation, Data Curation, Writing - Original Draft, Software, Visualization. **Désirée Treichler:** Conceptualization, Writing – Review & Editing, Resources, Supervision, Funding acquisition, Project Administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The training dataset, which includes retrieved snow depth data from ICESat-2 ATL08 for Mainland Norway spanning the period from October 2018 to December 2022, is openly accessible through Zenodo with DOI: 10.5281/zenodo.10048875. The Gradient Descent Coregistration is open source at <https://github.com/GlacioHack/xdem>. The validation ALS dataset is from the Norwegian Water Resources and Energy Directorate (NVE) at DOI: 10.5281/zenodo.2572731. The rest of the dataset is publicly accessible.

Acknowledgements

We express our gratitude to NASA for generously providing free access to the ICESat-2 data, to ECMWF for the ERA5 Reanalysis dataset. We also extend our appreciation to the Norwegian mapping agency for granting access to the national DEM, and to Kjetil Melvold for the valuable ALS dataset.

References

- Ali, S., Khorrami, B., Jehanzaib, M., Tariq, A., Ajmal, M., Arshad, A., Shafeeqe, M., Dilawar, A., Basit, I., Zhang, L., Sadri, S., Niaz, M.A., Jamil, A., Khan, S.N., 2023. Spatial downscaling of grace data based on xgboost model for improved understanding of hydrological droughts in the indus basin irrigation system (ibis). *Remote Sensing* 15, 873. doi:[10.3390/rs15040873](https://doi.org/10.3390/rs15040873).
- Arguez, A., Vose, R.S., 2011. The definition of the standard wmo climate normal: The key to deriving alternative climate normals. *Bulletin of the American Meteorological Society* 92, 699–704. doi:[10.1175/2010BAMS2955.1](https://doi.org/10.1175/2010BAMS2955.1).
- Bennett, K.E., Miller, G., Busey, R., Chen, M., Lathrop, E.R., Dann, J.B., Nutt, M., Crumley, R., Dillard, S.L., Dafflon, B., Kumar, J., Bolton, W.R., Wilson, C.J., Iversen, C.M., Wullschleger, S.D., 2022. Spatial patterns of snow distribution in the sub-arctic. *The Cryosphere* 16, 3269–3293. doi:[10.5194/tc-16-3269-2022](https://doi.org/10.5194/tc-16-3269-2022).
- Besso, H., Shean, D., Lundquist, J.D., 2024. Mountain snow depth retrievals from customized processing of icesat-2 satellite laser altimetry. *Remote Sens Environ* 300, 113843. doi:[10.1016/j.rse.2023.113843](https://doi.org/10.1016/j.rse.2023.113843).
- Callaghan, T.V., Johansson, M., Brown, R.D., Groisman, P.Y., Labba, N., Radionov, V., Bradley, R.S., Blangy, S., Bulygina, O.N., Christensen, T.R., Colman, J.E., Essery, R.L.H., Forbes, B.C., Forchhammer, M.C., Golubev, V.N., Honrath, R.E., Juday, G.P., Meshcherskaya, A.V., Phoenix, G.K., Pomeroy, J., Rautio, A., Robinson, D.A., Schmidt, N.M., Serreze, M.C., Shevchenko, V.P., Shiklomanov, A.I., Shmakin, A.B., Sköld, P., Sturm, M., Woo, M.k., Wood, E.F., 2011. Multiple effects of changes in arctic snow cover. *AMBIO* 40, 32–45. doi:[10.1007/s13280-011-0213-x](https://doi.org/10.1007/s13280-011-0213-x).
- Chen, T., Guestrin, C., 2016. Xgboost, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, San Francisco California USA. pp. 785–794. doi:[10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785).
- Chen, W., Yao, T., Zhang, G., Li, F., Zheng, G., Zhou, Y., Xu, F., 2022. Towards ice-thickness inversion: An evaluation of global digital elevation models (dems) in the glacierized tibetan plateau. *The Cryosphere* 16, 197–218. doi:[10.5194/tc-16-197-2022](https://doi.org/10.5194/tc-16-197-2022).
- Clark, M.P., Hendrikx, J., Slater, A.G., Kavetski, D., Anderson, B., Cullen, N.J., Kerr, T., Örn Hreinsson, E., Woods, R.A., 2011. Representing spatial variability of snow water equivalent in hydrologic and land-surface models: A review. *Water Resour. Res.* 47. doi:[10.1029/2011WR010745](https://doi.org/10.1029/2011WR010745).
- Deems, J.S., Painter, T.H., Finnegan, D.C., 2013. Lidar measurement of snow depth: A review. *J. Glaciol.* 59, 467–479. doi:[10.3189/2013JoG12J154](https://doi.org/10.3189/2013JoG12J154).
- Deschamps-Berger, C., Gascoin, S., Shean, D., Besso, H., Guiot, A., López-Moreno, J.I., 2023. Evaluation of snow depth retrievals from icesat-2 using airborne laser-scanning data. *The Cryosphere* 17, 2779–2792. doi:[10.5194/tc-17-2779-2023](https://doi.org/10.5194/tc-17-2779-2023).
- Duan, Z., Bastiaanssen, W., 2013. First results from version 7 trmm 3b43 precipitation product in combination with a new downscaling-calibration procedure. *Remote Sensing of Environment* 131, 1–13. doi:[10.1016/j.rse.2012.12.002](https://doi.org/10.1016/j.rse.2012.12.002).
- Dvornikov, Y., Khomutov, A., Mullanurov, D., Ermokhina, K., Gubarkov, A., Leibman, M., 2015. Gis and field data-based modelling of snow water equivalent in shrub tundra. *Fennia* 193, 53–65. doi:[10.11143/46363](https://doi.org/10.11143/46363).
- Enderlin, E.M., Elkin, C.M., Gendreau, M., Marshall, H., O’Neil, S., McNeil, C., Florentine, C., Sass, L., 2022. Uncertainty of icesat-2 atl06- and atl08-derived snow depths for glacierized and vegetated mountain regions. *Remote Sensing of Environment* 283, 113307. doi:[10.1016/j.rse.2022.113307](https://doi.org/10.1016/j.rse.2022.113307).
- European Space Agency, 2021. Copernicus glo-90 digital surface model. doi:[10.5069/G9028PQB](https://doi.org/10.5069/G9028PQB).
- Fassnacht, S.R., Brown, K.S.J., Blumberg, E.J., López Moreno, J.I., Covino, T.P., Kappas, M., Huang, Y., Leone, V., Kashipazha, A.H., 2018. Distribution of snow depth variability. *Front. Earth Sci.* 12, 683–692. doi:[10.1007/s11707-018-0714-z](https://doi.org/10.1007/s11707-018-0714-z).
- Fiddes, J., Aalstad, K., Lehning, M., 2022. Topoclim: Rapid

- topography-based downscaling of regional climate model output in complex terrain v1.1. *Geosci. Model Dev.* 15, 1753–1768. doi:[10.5194/gmd-15-1753-2022](https://doi.org/10.5194/gmd-15-1753-2022).
- Freudiger, D., Kohn, I., Seibert, J., Stahl, K., Weiler, M., 2017. Snow redistribution for the hydrological modeling of alpine catchments. *WIREs Water* 4, e1232. doi:[10.1002/wat2.1232](https://doi.org/10.1002/wat2.1232).
- Gisnås, K., Westermann, S., Schuler, T.V., Melvold, K., Etzelmüller, B., 2016. Small-scale variation of snow in a regional permafrost model. *The Cryosphere* 10, 1201–1215. doi:[10.5194/tc-10-1201-2016](https://doi.org/10.5194/tc-10-1201-2016).
- Gruber, A., Wessel, B., Huber, M., Roth, A., 2012. Operational tandem-x dem calibration and first validation results. *ISPRS Journal of Photogrammetry and Remote Sensing* 73, 39–49. doi:[10.1016/j.isprsjprs.2012.06.002](https://doi.org/10.1016/j.isprsjprs.2012.06.002).
- Grünewald, T., Lehning, M., 2011. Altitudinal dependency of snow amounts in two small alpine catchments: Can catchment-wide snow amounts be estimated via single snow or precipitation stations? *Ann. Glaciol.* 52, 153–158. doi:[10.3189/172756411797252248](https://doi.org/10.3189/172756411797252248).
- Grünewald, T., Schirmer, M., Mott, R., Lehning, M., 2010. Spatial and temporal variability of snow depth and ablation rates in a small mountain catchment. *The Cryosphere* 4, 215–225. doi:[10.5194/tc-4-215-2010](https://doi.org/10.5194/tc-4-215-2010).
- Grünewald, T., Stötter, J., Pomeroy, J.W., Dadic, R., Moreno Baños, I., Marturià, J., Spross, M., Hopkinson, C., Burlando, P., Lehning, M., 2013. Statistical modelling of the snow depth distribution in open alpine terrain. *Hydrol. Earth Syst. Sci.* 17, 3005–3021. doi:[10.5194/hess-17-3005-2013](https://doi.org/10.5194/hess-17-3005-2013).
- Günther, D., Marke, T., Essery, R., Strasser, U., 2019. Uncertainties in snowpack simulations—assessing the impact of model structure, parameter choice, and forcing data error on point-scale energy balance snow model performance. *Water Resources Research* 55, 2779–2800. doi:[10.1029/2018WR023403](https://doi.org/10.1029/2018WR023403).
- Hawker, L., Uhe, P., Paulo, L., Sosa, J., Savage, J., Sampson, C., Neal, J., 2022. A 30 m global map of elevation with forests and buildings removed. *Environ. Res. Lett.* 17, 024016. doi:[10.1088/1748-9326/ac4d4f](https://doi.org/10.1088/1748-9326/ac4d4f).
- He, S., Ohara, N., Miller, S.N., 2019. Understanding subgrid variability of snow depth at 1-km scale using lidar measurements. *Hydrological Processes* 33, 1525–1537. doi:[10.1002/hyp.13415](https://doi.org/10.1002/hyp.13415).
- Helbig, N., van Herwijnen, A., 2017. Subgrid parameterization for snow depth over mountainous terrain from flat field snow depth. *Water Resour. Res.* 53, 1444–1456. doi:[10.1002/2016WR019872](https://doi.org/10.1002/2016WR019872).
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellán, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thépaut, J.N., 2020. The era5 global reanalysis. *Q.J.R. Meteorol. Soc.* 146, 1999–2049. doi:[10.1002/qj.3803](https://doi.org/10.1002/qj.3803).
- Hu, W., Scholz, Y., Yelgeti, M., von Bremen, L., Deng, Y., 2023. Downscaling era5 wind speed data: A machine learning approach considering topographic influences. *Environ. Res. Lett.* 18, 094007. doi:[10.1088/1748-9326/aceb0a](https://doi.org/10.1088/1748-9326/aceb0a).
- Hugonnet, R., Brun, F., Berthier, E., Dehecq, A., Mannerfelt, E.S., Eckert, N., Farinotti, D., 2022. Uncertainty analysis of digital elevation models by spatial inference from stable terrain. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing* 15, 6456–6472. doi:[10.1109/JSTARS.2022.3188922](https://doi.org/10.1109/JSTARS.2022.3188922).
- Immerzeel, W.W., Lutz, A.F., Andrade, M., Bahl, A., Biemans, H., Bolch, T., Hyde, S., Brumby, S., Davies, B.J., Elmore, A.C., Emmer, A., Feng, M., Fernández, A., Haritashya, U., Kargel, J.S., Koppes, M., Kraaijenbrink, P.D.A., Kulkarni, A.V., Mayewski, P.A., Nepal, S., Pacheco, P., Painter, T.H., Pellicciotti, F., Rajaram, H., Rupper, S., Sinisalo, A., Shrestha, A.B., Viviroli, D., Wada, Y., Xiao, C., Yao, T., Baillie, J.E.M., 2020. Importance and vulnerability of the world's water towers. *Nature* 577, 364–369. doi:[10.1038/s41586-019-1822-y](https://doi.org/10.1038/s41586-019-1822-y).
- Kim, R.S., Kumar, S., Vuyovich, C., Houser, P., Lundquist, J., Mudryk, L., Durand, M., Barros, A., Kim, E.J., Forman, B.A., Gutmann, E.D., Wrzesien, M.L., Garnaud, C., Sandells, M., Marshall, H.P., Cristea, N., Pflug, J.M., Johnston, J., Cao, Y., Mocko, D., Wang, S., 2021. Snow ensemble uncertainty project (seup): Quantification of snow water equivalent uncertainty across north america via ensemble land surface modeling. *The Cryosphere* 15, 771–791. doi:[10.5194/tc-15-771-2021](https://doi.org/10.5194/tc-15-771-2021).
- Kraaijenbrink, P.D.A., Stigter, E.E., Yao, T., Immerzeel, W.W., 2021. Climate change decisive for asia's snow melt-water supply. *Nat. Clim. Chang.* 11, 591–597. doi:[10.1038/s41558-021-01074-x](https://doi.org/10.1038/s41558-021-01074-x).
- Lehning, M., Völkisch, I., Gustafsson, D., Nguyen, T.A., Stähli, M., Zappa, M., 2006. Alpine3d: A detailed model of mountain surface processes and its application to snow hydrology. *Hydrological Processes* 20, 2111–2128. doi:[10.1002/hyp.6204](https://doi.org/10.1002/hyp.6204).
- Li, G., Wang, Z.S., Huang, N., 2018. A snow distribution model based on snowfall and snow drifting simulations in mountain area. *JGR Atmospheres* 123, 7193–7203. doi:[10.1029/2018JD028434](https://doi.org/10.1029/2018JD028434).
- Li, H., Zhao, J., Yan, B., Yue, L., Wang, L., 2022. Global dems vary from one to another: An evaluation of newly released copernicus, nasa and aw3d30 dem on selected terrains of china using icesat-2 altimetry data. *International Journal of Digital Earth* 15, 1149–1168. doi:[10.1080/17538947.2022.2094002](https://doi.org/10.1080/17538947.2022.2094002).
- Lievens, H., Brangers, I., Marshall, H.P., Jonas, T., Olefs, M., De Lannoy, G., 2022. Sentinel-1 snow depth retrieval at sub-kilometer resolution over the european alps. *The Cryosphere* 16, 159–177. doi:[10.5194/tc-16-159-2022](https://doi.org/10.5194/tc-16-159-2022).
- Lievens, H., Demuzere, M., Marshall, H.P., Reichle, R.H., Brucker, L., Brangers, I., de Rosnay, P., Dumont, M., Girotto, M., Immerzeel, W.W., Jonas, T., Kim, E.J., Koch, I., Marty, C., Saloranta, T., Schöberl, J., De Lannoy, G.J.M., 2019. Snow depth variability in the northern hemisphere mountains observed from space. *Nat Commun* 10, 4629. doi:[10.1038/s41467-019-12566-y](https://doi.org/10.1038/s41467-019-12566-y).
- Liston, G.E., 2004. Representing subgrid snow cover heterogeneities in regional and global models. *J. Climate* 17, 1381–1397. doi:[10.1175/1520-0442\(2004\)017<1381:RSSCHI>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<1381:RSSCHI>2.0.CO;2).
- Liston, G.E., Elder, K., 2006. A distributed snow-evolution modeling system (snowmodel). *Journal of Hydrometeorology* 7, 1259–1276. doi:[10.1175/JHM548.1](https://doi.org/10.1175/JHM548.1).
- Liu, A., 2021. Performance evaluation of gedi and icesat-2 laser altimeter data for terrain and canopy height retrievals. *Remote Sensing of Environment*, 16.
- Liu, Z., 2023. Snow Depth Retrieval and Downscaling Using

- Satellite Laser Altimetry, Machine Learning, and Climate Reanalysis: A Case Study in Mainland Norway. Master's thesis. University of Oslo. Oslo, Norway.
- Livneh, B., Badger, A.M., 2020. Drought less predictable under declining future snowpack. *Nat. Clim. Chang.* 10, 452–458. doi:[10.1038/s41558-020-0754-8](https://doi.org/10.1038/s41558-020-0754-8).
- Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.I., 2020. From local explanations to global understanding with explainable ai for trees. *Nat Mach Intell* 2, 56–67. doi:[10.1038/s42256-019-0138-9](https://doi.org/10.1038/s42256-019-0138-9).
- Magruder, L., Neuenschwander, A., Klotz, B., 2021. Digital terrain model elevation corrections using space-based imagery and icesat-2 laser altimetry. *Remote Sensing of Environment* 264, 112621. doi:[10.1016/j.rse.2021.112621](https://doi.org/10.1016/j.rse.2021.112621).
- Meinshausen, N., 2006. Quantile regression forests. *Journal of Machine Learning Research* 7, 983–999.
- Melvold, K., Skaugen, T., 2013. Multiscale spatial variability of lidar-derived and modeled snow depth on hardangervidda, norway. *Ann. Glaciol.* 54, 273–281. doi:[10.3189/2013AoG62A161](https://doi.org/10.3189/2013AoG62A161).
- Mendoza, P.A., Shaw, T.E., McPhee, J., Musselman, K.N., Revuelto, J., MacDonell, S., 2020. Spatial distribution and scaling properties of lidar-derived snow depth in the extratropical andes. *Water Resour. Res.* 56. doi:[10.1029/2020WR028480](https://doi.org/10.1029/2020WR028480).
- Mott, R., Schirmer, M., Lehning, M., 2011. Scaling properties of wind and snow depth distribution in an alpine catchment. *J. Geophys. Res.* 116. doi:[10.1029/2010JD014886](https://doi.org/10.1029/2010JD014886).
- Mott, R., Schlögl, S., Dirks, L., Lehning, M., 2017. Impact of extreme land surface heterogeneity on micrometeorology over spring snow cover. *Journal of Hydrometeorology* 18, 2705–2722. doi:[10.1175/JHM-D-17-0074.1](https://doi.org/10.1175/JHM-D-17-0074.1).
- Mott, R., Vionnet, V., Grünewald, T., 2018. The seasonal snow cover dynamics: Review on wind-driven coupling processes. *Front. Earth Sci.* 6. doi:[10.3389/feart.2018.00197](https://doi.org/10.3389/feart.2018.00197).
- Moudrý, V., Gdulová, K., Gábor, L., Šárovcová, E., Barták, V., Leroy, F., Špatenková, O., Rocchini, D., Prošek, J., 2022. Effects of environmental conditions on icesat-2 terrain and canopy heights retrievals in central european mountains. *Remote Sensing of Environment* 279, 113112. doi:[10.1016/j.rse.2022.113112](https://doi.org/10.1016/j.rse.2022.113112).
- Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D.G., Piles, M., Rodríguez-Fernández, N.J., Zsotér, E., Buontempo, C., Thépaut, J.N., 2021. Era5-land: A state-of-the-art global reanalysis dataset for land applications. *Earth Syst. Sci. Data* 13, 4349–4383. doi:[10.5194/essd-13-4349-2021](https://doi.org/10.5194/essd-13-4349-2021).
- Mudryk, L., Santolaria-Otín, M., Krinner, G., Ménégoz, M., Derksen, C., Brutel-Vuilmet, C., Brady, M., Essery, R., 2020. Historical northern hemisphere snow cover trends and projected changes in the cmip6 multi-model ensemble. *The Cryosphere* 14, 2495–2514. doi:[10.5194/tc-14-2495-2020](https://doi.org/10.5194/tc-14-2495-2020).
- Muñoz Sabater, J., 2021a. Era5-land hourly data from 1950 to present, copernicus climate change service (c3s) climate data store (cds).
- Muñoz Sabater, J., 2021b. Era5-land monthly data from 1950 to present, copernicus climate change service (c3s) climate data store (cds).
- Musselman, K.N., Lehner, F., Ikeda, K., Clark, M.P., Prein, A.F., Liu, C., Barlage, M., Rasmussen, R., 2018. Projected increases and shifts in rain-on-snow flood risk over western north america. *Nature Clim Change* 8, 808–812. doi:[10.1038/s41558-018-0236-4](https://doi.org/10.1038/s41558-018-0236-4).
- Myers-Smith, I.H., Kerby, J.T., Phoenix, G.K., Bjerke, J.W., Epstein, H.E., Assmann, J.J., John, C., Andreu-Hayles, L., Angers-Blondin, S., Beck, P.S.A., Berner, L.T., Bhatt, U.S., Bjorkman, A.D., Blok, D., Bryn, A., Christiansen, C.T., Cornelissen, J.H.C., Cunliffe, A.M., Elmendorf, S.C., Forbes, B.C., Goetz, S.J., Hollister, R.D., de Jong, R., Loranty, M.M., Macias-Fauria, M., Maseyk, K., Normand, S., Olofsson, J., Parker, T.C., Parmentier, F.J.W., Post, E., Schaeppman-Strub, G., Stordal, F., Sullivan, P.F., Thomas, H.J.D., Tømmervik, H., Trehearne, R., Tweedie, C.E., Walker, D.A., Wilmking, M., Wipf, S., 2020. Complexity revealed in the greening of the arctic. *Nat. Clim. Chang.* 10, 106–117. doi:[10.1038/s41558-019-0688-1](https://doi.org/10.1038/s41558-019-0688-1).
- Neuenschwander, A., Pitts, K., 2019. The atl08 land and vegetation product for the icesat-2 mission. *Remote Sensing of Environment* 221, 247–259. doi:[10.1016/j.rse.2018.11.005](https://doi.org/10.1016/j.rse.2018.11.005).
- Neuenschwander, A., Pitts, K., Jolley, B., Robbins, J., Markel, J., Popescu, S., Nelson, R., Harding, D., Klotz, B., Sheridan, R., 2022. Ice, cloud, and land elevation satellite 2 (icesat-2) algorithm theoretical basis documents (atbd) for land-vegetation along-track products (atl08).
- Nuth, C., Kääb, A., 2011. Co-registration and bias corrections of satellite elevation data sets for quantifying glacier thickness change. *The Cryosphere* 5, 271–290. doi:[10.5194/tc-5-271-2011](https://doi.org/10.5194/tc-5-271-2011).
- Oliver, M., Webster, R., 2014. A tutorial guide to geostatistics: Computing and modelling variograms and kriging. *CATENA* 113, 56–69. doi:[10.1016/j.catena.2013.09.016](https://doi.org/10.1016/j.catena.2013.09.016).
- Parr, C., Sturm, M., Larsen, C., 2020. Snowdrift landscape patterns: An arctic investigation. *Water Resources Research* 56, e2020WR027823. doi:[10.1029/2020WR027823](https://doi.org/10.1029/2020WR027823).
- Pflug, J.M., Hughes, M., Lundquist, J.D., 2021. Downscaling snow deposition using historic snow depth patterns: Diagnosing limitations from snowfall biases, winter snow losses, and interannual snow pattern repeatability. *Water Resources Research* 57, e2021WR029999. doi:[10.1029/2021WR029999](https://doi.org/10.1029/2021WR029999).
- Revuelto, J., Billecocq, P., Tuzet, F., Cluzet, B., Lamare, M., Larue, F., Dumont, M., 2020. Random forests as a tool to understand the snow depth distribution and its evolution in mountain areas. *Hydrological Processes* 34, 5384–5401. doi:[10.1002/hyp.13951](https://doi.org/10.1002/hyp.13951).
- Revuelto, J., López-Moreno, J.I., Azorin-Molina, C., Vicente-Serrano, S.M., 2014. Topographic control of snowpack distribution in a small catchment in the central spanish pyrenees: Intra- and inter-annual persistence. *The Cryosphere* 8, 1989–2006. doi:[10.5194/tc-8-1989-2014](https://doi.org/10.5194/tc-8-1989-2014).
- Rolstad, C., Haug, T., Denby, B., 2009. Spatially integrated geodetic glacier mass balance and its uncertainty based on geostatistical analysis: Application to the western svartisen ice cap, norway. *J. Glaciol.* 55, 666–680. doi:[10.3189/002214309789470950](https://doi.org/10.3189/002214309789470950).
- Saloranta, T.M., 2012. Simulating snow maps for norway: Description and statistical evaluation of the senorge snow model. *The Cryosphere* 6, 1323–1337. doi:[10.5194/tc-6-1323-2012](https://doi.org/10.5194/tc-6-1323-2012).
- Skaugen, T., Melvold, K., 2019. Modeling the snow depth variability with a high-resolution lidar data set and nonlinear terrain dependency. *Water Resour. Res.* 55, 9689–9704. doi:[10.1029/2019WR025030](https://doi.org/10.1029/2019WR025030).

- Sturm, M., Wagner, A.M., 2010. Using repeated patterns in snow distribution modeling: An arctic example. Water Resour. Res. 46. doi:[10.1029/2010WR009434](https://doi.org/10.1029/2010WR009434).
- Sugiyama, M., Nakajima, S., Kashima, H., Buenau, P., Kawanabe, M., 2007. Direct importance estimation with model selection and its application to covariate shift adaptation, in: Advances in Neural Information Processing Systems, Curran Associates, Inc.. pp. 809–818.
- Tian, X., Shan, J., 2021. Comprehensive evaluation of the icesat-2 atl08 terrain product. IEEE Trans. Geosci. Remote Sensing 59, 8195–8209. doi:[10.1109/TGRS.2021.3051086](https://doi.org/10.1109/TGRS.2021.3051086).
- Treichler, D., Kääb, A., 2017. Snow depth from icesat laser altimetry — a test study in southern norway. Remote Sensing of Environment 191, 389–401. doi:[10.1016/j.rse.2017.01.022](https://doi.org/10.1016/j.rse.2017.01.022).
- Trujillo, E., Ramirez, J.A., Elder, K.J., 2007. Topographic, meteorologic, and canopy controls on the scaling characteristics of the spatial distribution of snow depth fields. Water Resour. Res. 43. doi:[10.1029/2006WR005317](https://doi.org/10.1029/2006WR005317).
- Tsang, L., Durand, M., Derksen, C., Barros, A.P., Kang, D.H., Lievens, H., Marshall, H.P., Zhu, J., Johnson, J., King, J., Lemmettyinen, J., Sandells, M., Rutter, N., Siqueira, P., Nolin, A., Osmanoglu, B., Vuyovich, C., Kim, E., Taylor, D., Merkouriadi, I., Brucker, L., Navari, M., Dumont, M., Kelly, R., Kim, R.S., Liao, T.H., Borah, F., Xu, X., 2022. Review article: Global monitoring of snow water equivalent using high-frequency radar remote sensing. The Cryosphere 16, 3531–3573. doi:[10.5194/tc-16-3531-2022](https://doi.org/10.5194/tc-16-3531-2022).
- Weiss, A., 2001. Topographic position and landforms analysis. Xdem contributors, 2021. Xdem. doi:[10.5281/zenodo.4809698](https://doi.org/10.5281/zenodo.4809698).
- Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J.C., Sampson, C.C., Kanae, S., Bates, P.D., 2017. A high-accuracy map of global terrain elevations. Geophysical Research Letters 44, 5844–5853. doi:[10.1002/2017GL072874](https://doi.org/10.1002/2017GL072874).
- Zevenbergen, L.W., Thorne, C.R., 1987. Quantitative analysis of land surface topography. Earth Surf. Process. Landforms 12, 47–56. doi:[10.1002/esp.3290120107](https://doi.org/10.1002/esp.3290120107).
- Zhang, W., Quan, H., Srinivasan, D., 2018. Parallel and reliable probabilistic load forecasting via quantile regression forest and quantile determination. Energy 160, 810–819. doi:[10.1016/j.energy.2018.07.019](https://doi.org/10.1016/j.energy.2018.07.019).
- Zhu, H., Liu, H., Zhou, Q., Cui, A., 2023. A xgboost-based downscaling-calibration scheme for extreme precipitation events. IEEE Trans. Geosci. Remote Sensing 61, 1–12. doi:[10.1109/TGRS.2023.3294266](https://doi.org/10.1109/TGRS.2023.3294266).

Appendix A. Appendix

Appendix A.1. Gradient descent co-registration

GDC (Gradient Descent Co-registration) and NuthKaab ([Nuth and Kääb, 2011](#)) address the same problem but use distinct techniques. During co-registration with NuthKaab, the DEM to be aligned is iteratively shifted by updating the transformation parameters (a, b, c) in each iteration, requiring interpolation to calculate elevation differences until convergence. NuthKaab solves the co-registration problem by minimizing NMAD statistically and requires calculations of terrain parameters such as slope, aspect, and sometimes curvature at least once. It operates efficiently for low-workload tasks with coarse resolution or small coverage. On the other hand, GDC solves the same geo-referencing problem more quickly and without requiring terrain parameters calculation by regard the mismatch as a bound-restricted minimizing problem with random noise. The stochastic gradient descent algorithm is used to speed up searching:

$$\min_{\mathbf{x} \subseteq \theta} f(\mathbf{x}) = \min_{\mathbf{x} \subseteq \theta} \mathbb{E}[F(\mathbf{x}, \xi)]$$

where x is the parameter that minimizes $f(x)$ within the bounds θ . And ξ represents the noise. The goal of the minimization problem is to find the parameter of \mathbf{x} that minimizes the expected value of $F(\mathbf{x}, \xi)$, which represents the optimal decision under uncertainty. In the context of co-registration. The function to be minimized is the NMAD of elevation difference (dh). The parameter is the shift matrix (E, N) ([Figure A.12](#)).

This study aligned DEMs with ICESat-2 snow-free segments tile by tile. The co-registered DEM tiles exhibit minimal NMAD compared to reference ICESat-2 snow-free segments. In some study cases of co-registration, the remaining median error is usually considered a vertical bias. However, this study cautions against applying vertical bias correction if there is no confirmation that the DEM was produced under snow-free conditions since snowpack on surfaces can lead to a skewed median. This study did not carry out vertical bias correction during co-registration and left it for the following step of bias correction. The removal of aspect-dependant bias can be visualized by an aspect-dh binning plot (see Figure 4.2a in [Liu \(2023\)](#)).

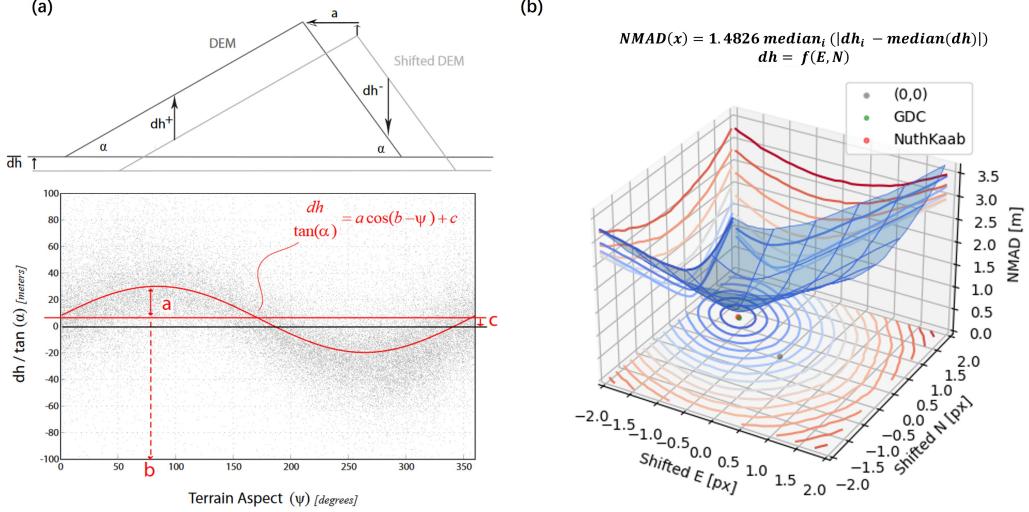


Figure A.12: NuthKaab Co-registration and Gradient Descent Co-registration. The elevation difference (dh) induced by mismatch can be expressed in a 2-dimensional curve line (a). This curve suggests a statistical solution where the dispersion is related to aspect (ψ), slope (α), shifted distance (a), shifted direction (b) and vertical bias (c). By solving the equation, it gives a vector (a, b, c) to shift the DEM back over iterations (Nuth and Kääb (2011)). On the other hand, GDC present dh in a 3-dimensional curved surface, it gets the shift matrix using a gradient descent algorithm, resulting in the local minimal of NMAD (b)

Appendix A.2. Bias correction model based on ICESat-2 snow-free segments

In this work, we employ a regression model to estimate the discrepancy for four DEMs using various features such as terrain parameters, vegetation, and quality metrics from ICESat-2. Approximately 5 million snow-free segments were utilized as the training dataset. Even if DTM1 and DTM10 are high-quality lidar-based DEM, there is a need to correct them because of differences in resolution and surface condition. To better inform the model of the characteristics of the discrepancy, we create the following metrics:

- The ‘Difference’ metric, calculated as $h_{te_best_fit_20m_2}$ minus $h_{te_best_fit}$, describes the elevation difference between the midpoint of a segment and its surrounding terrain.
- The df_dtm1_10 (DTM1 minus DTM10) are describing the elevation difference between DTM1 and DTM10 caused by resolution and interpolation.
- The df_cop30_fab (Copernicus GLO30 minus FABDEM) describes the difference between DEM and DSM caused by the surface condition.

The histograms (a, b) reveal significant negative skewness in DTM1 and GLO30 before bias cor-

rection. The features, sorted in descending order of importance, explain the contribution of features to negative skewness (c). A high difference value (indicated in red) suggests that the subset geosegment’s elevation exceeds $h_{te_best_fit}$ but remains below DTM1. This indicates that using $h_{te_best_fit_20m_2}$ aligns more closely with the DTM1 surface elevation. Convex terrain shows a significant negative bias, with positive values for tpi_9 , tpi , and df_dtm1_dtm10 (f, h). Conversely, when Copernicus GLO30 is used as the reference DEM (d), a negative bias is observed in concave terrain, where tpi_9 and tpi are negatively proportional to the bias (g, i). This shows that ICESat-2 has a resolution better than GLO30 but less than DTM1. After bias correction, the bias-corrected DTM1 has an overall NMAD of 0.43 m and GLO30 has an NMAD of 0.48 m in our validation area. In contrast, the raw DTM1 and GLO30 have NMADs of 0.61 m and 1.75 m respectively when evaluated by ICESat-2 snow-free measurements. This represents a significant improvement for GLO30. Liu (2021) conducted an uncertainty analysis on ICESat-2’s ATL08 and obtained an RMSE of 0.98 m in Alaska. These metrics are comparable to those of this study.

The cloud existence, beam intensity and night_flag are the least strong features for bias. Figure A.14

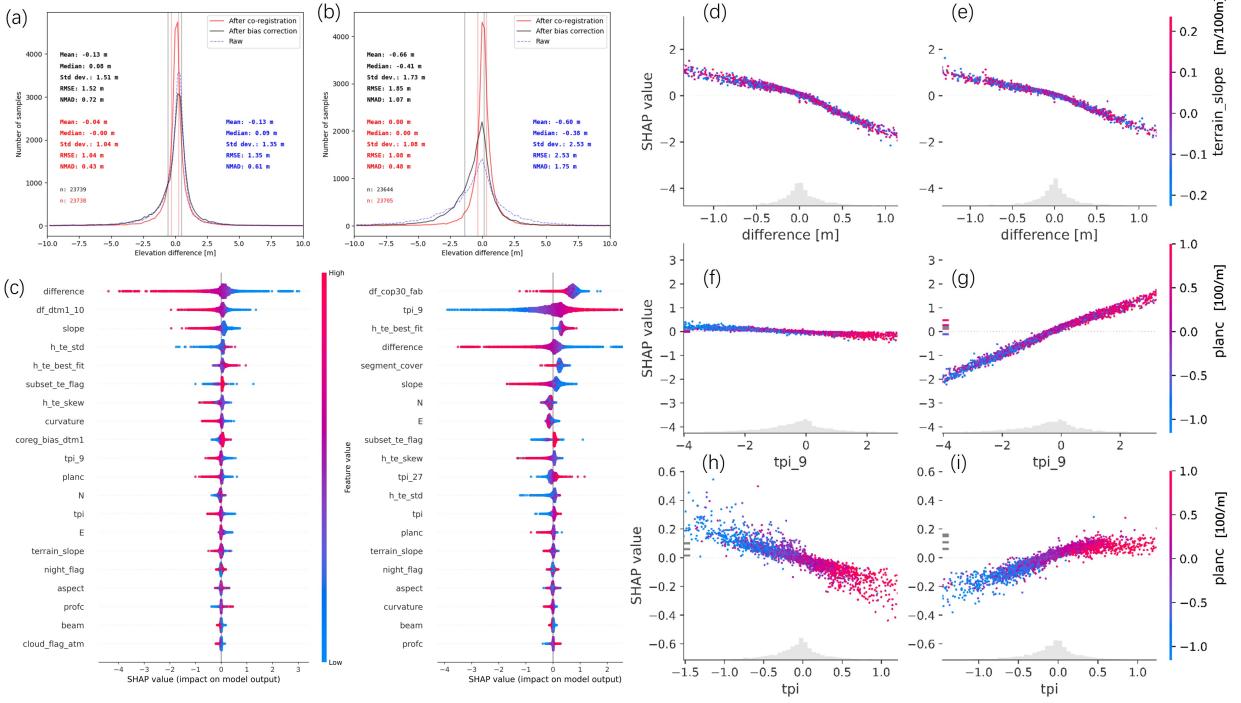


Figure A.13: Interpreting bias correction on DTM1 and GLO30. The discrepancy distribution of DTM1 (a) and GLO30 (b) shows the difference after applying co-registration and bias correction for a given dataset in the validation area. The Q25 and Q75 are plotted in vertical lines.

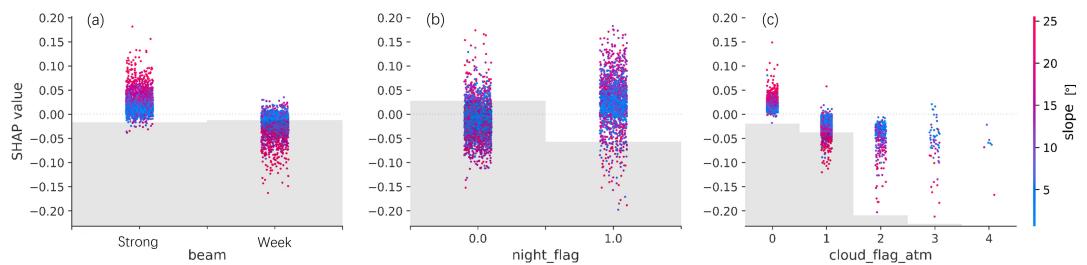


Figure A.14: Interpreting ICESat-2 related bias. Beam intensity (a) night flag, with 0 representing day, and 1 representing night (b). Cloud, 0 means no scattering, 1 means clouds at > 3km, 2 means clouds at 1-3 km, and 3 = clouds at < 1 km, 4 = blowing snow (c)

shows the different tendencies of these features.

When using the vegetation parameter from ICESat-2, the height of vegetation differs significantly between snow-free and snow-covered conditions, with the mean canopy height being substantially higher under snow-free conditions. Moreover, vegetation metrics is given in 100 m interval, which is imprecise for correcting the bias specifically in the middle point of the segment. Consequently, caution is advised when using these vegetation parameters in bias regression for DTM1 and DTM10, particularly in areas with dense vegetation.

It is important to note that the bias correction model is trained to correct the DEMs to their “ICESat-2 ATL08 snow-free condition,” which is not equivalent to the “true surface” but rather still encompasses bias patterns of ICESat-2 ATL08 snow-free segments. If the DEM is produced under heavy snow conditions, this correction may not be able to fully remove the snow patch on DEM, as all parameters are calculated from the DEM.

Our workflow assumes a consistent bias pattern in ICESat-2 elevation measurements, whether snow is present or not, which is also the assumption of using crossover measurement. However, from a snow-free surface to a snow-covered surface, the bias of ICESat-2 may fluctuate due to changes in surface roughness, reflectivity, and canopy height caused by snow cover.

Appendix A.3. Calibration

This study applied quantile mapping to calibrate scaling biases in the model output. This method aligns the cumulative distribution function (CDF) of the model output, $F_M(x)$, with that of the ALS 2009 data $F_O(x)$. For each data point x in the model output, we determined its corresponding quantile q in the model’s distribution as $q = F_M(x)$. We then mapped this quantile to the equivalent value in the observed data distribution, using the inverse CDF of the observed data, $y = F_O^{-1}(q)$. This process adjusts each model data point to match the statistical properties of the observed data, thereby reducing scaling biases:

$$y = F_O^{-1}(F_M(x))$$

Figure A.15 and Figure A.16 show the downscaled output at mesoscale and microscale levels before calibration. These figures reveal that the peak snow on

the western mountain tops is not well represented, indicating that the DTM1 may still have snow cover. In Figure A.15 e, the difference between DTM1 and ALS DEM acquired in Sep. 2008 is shown, highlighting a remaining snow patch on DTM1. This explains the underestimation of downscaled snow depth and emphasizes the need for calibration.

Appendix A.4. The snow depth 2009

The year 2009 experienced less snowfall than 2008. Notably, SeNorge showed improvements in capturing snow depth in the western mountains.

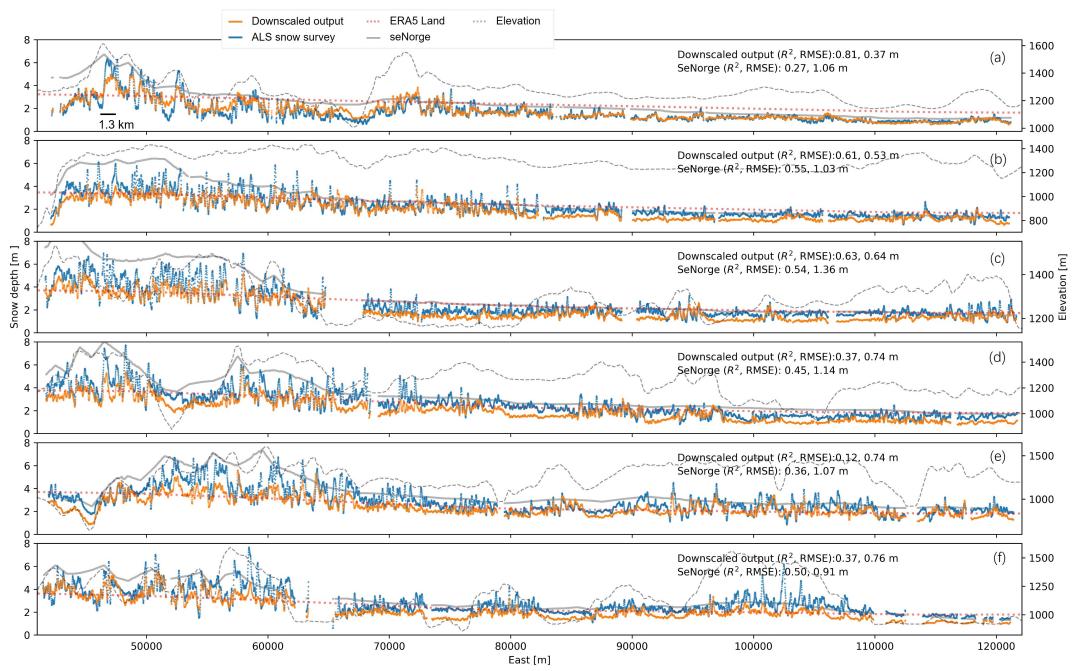


Figure A.15: Snow depth profiles across Hardangervidda in April 2008 (Pre-calibration). This is a mesoscale analysis of aggregated snow depth. Six ALS flight profile for April 2008 has been used for comparison.

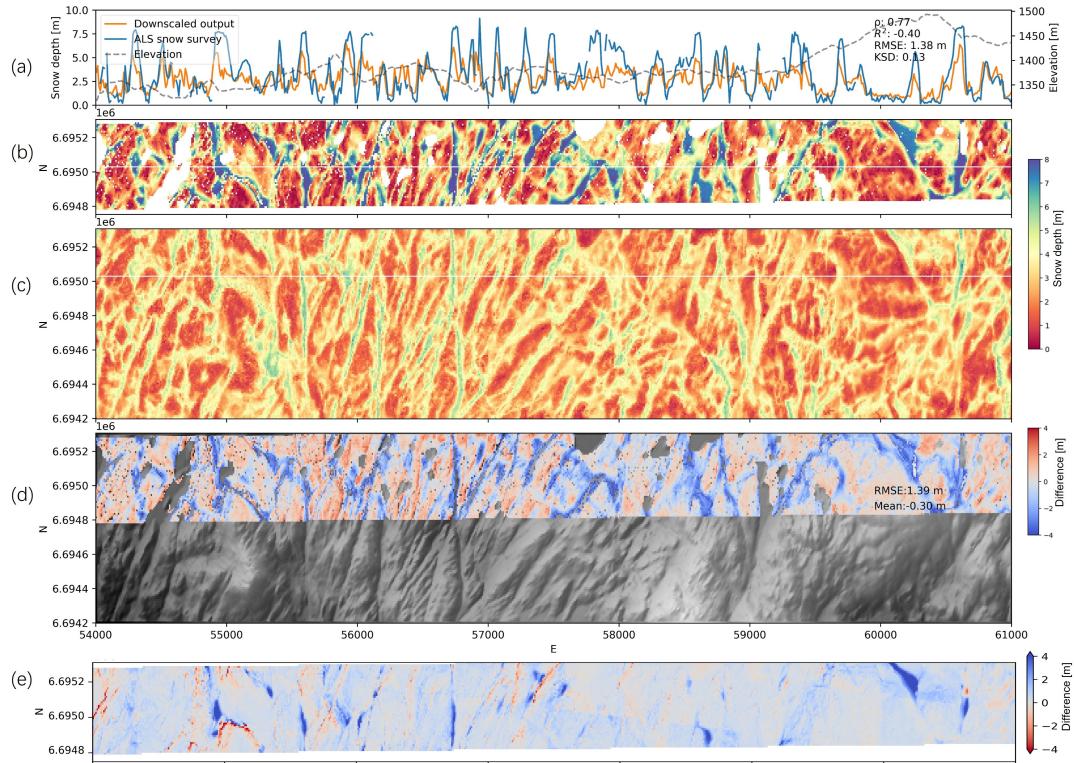


Figure A.16: Microscale Snow Depth Comparison in Hardangervidda, April 2008 (Pre-Calibration). Panel (b) shows the ALS snow survey validation strip, revealing significant snow depth variations. Panel (c) illustrates the downscaled snow depth from the DTM1 model, capturing most of the observed fidelity but with a low R^2 value of transect line (a). Differences between the ALS data and model output are highlighted in Panel (d). Differences between DTM1 and ALS bare ground DEM are presented in Panel (e), showing the patchy snow on DTM1.

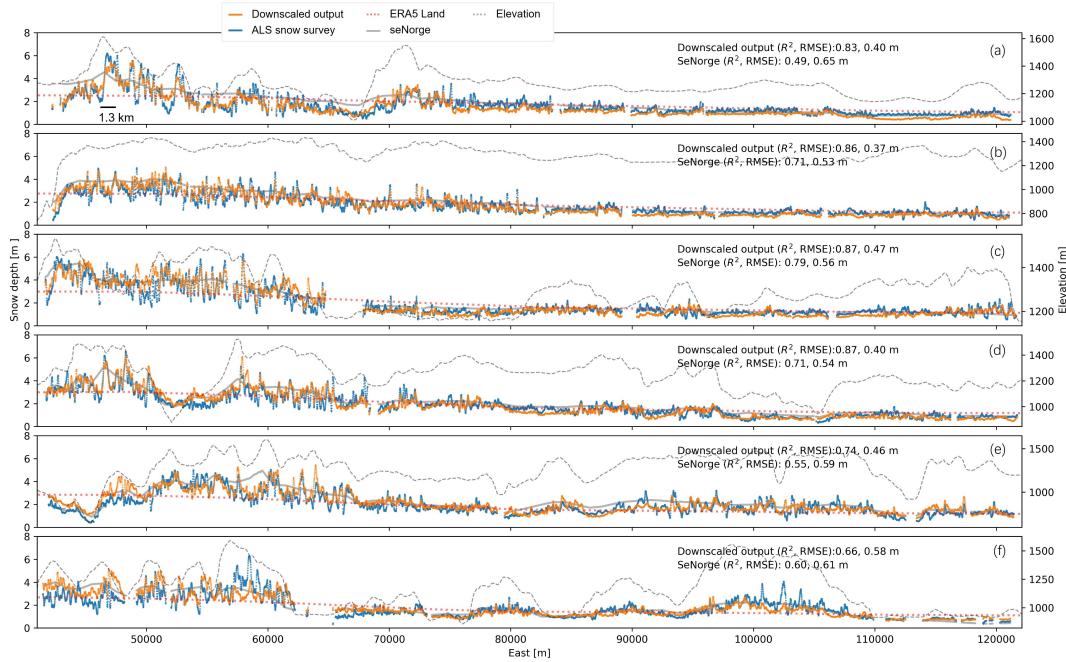


Figure A.17: Snow depth profiles across Hardangervidda in April 2009. This figure showcases a mesoscale analysis of snow depth along six ALS flight lines, spanning north to south (panels a-f). Data points, aggregated to represent 100 m x 500 m cells (because of the ALS survey's 500 m transect width).

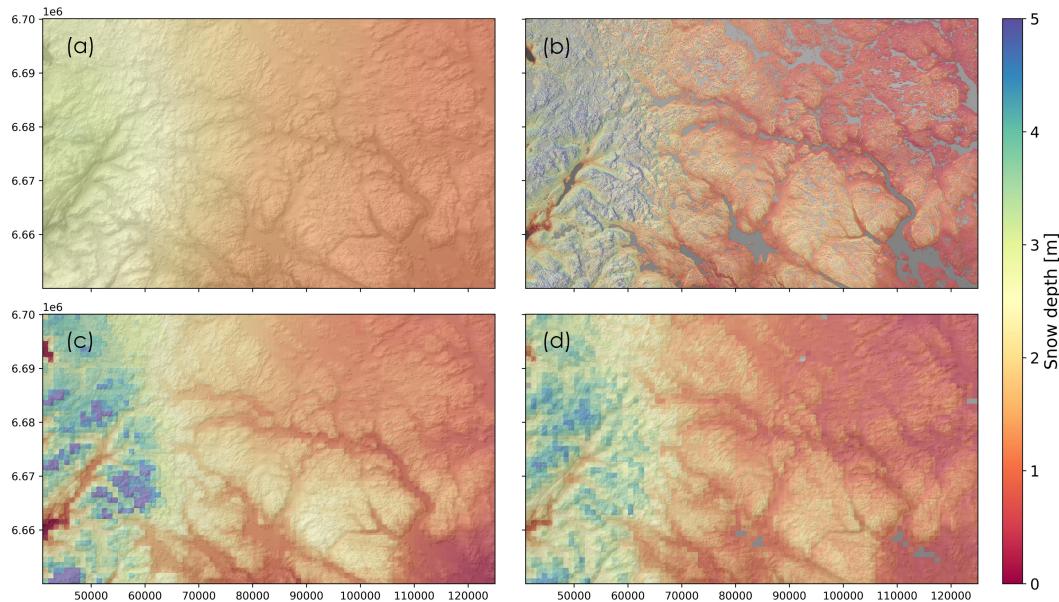


Figure A.18: Spatial Distribution of Snow Depth in Hardangervidda, April 2009. (a) Snow depth from ERA5 Land in 10 m resolution (after linear interpolation). (b) Downscaled snow depth output in 10 m resolution. (c) seNorge snow depth at 1 km resolution. (d) Downscaled snow depth aggregated at 1 km resolution.