

# Snow Distribution Patterns from Satellite Laser Altimetry

Zhihao Liu<sup>a</sup>, Désirée Treichler<sup>a,\*</sup>, Simon Filhol<sup>a</sup>

<sup>a</sup>University of Oslo, Department of Geosciences, Sem Sælands vei 1, Oslo, 371

---

## Abstract

Estimating the variability of seasonal snow cover, in particular snow depth in remote areas, poses significant challenges due to limited spatial and temporal data availability. This study uses snow depth measurements from ICESat-2 satellite laser altimetry, which are sparse in both space and time, and incorporates them with climate reanalysis data in a downscaling-calibration scheme to produce monthly gridded snow depth maps at the hillslope scale ( $<= 100$  m), thereby capturing the intricate patterns of snow distribution in mountain environment. Snow surface elevation measurements obtained from ICESat-2 along profiles are compared to a digital elevation model (DEM) to determine snow depth at each point. To efficiently manage this process, a high-performance co-registration algorithm and a bias correction are employed. The bias correction model uses ICESat-2 snow-free measurements to correct bias in the used DEMs. Subsequently, regression analysis is conducted to establish a relationship between the retrieved snow depth and the corresponding ERA5 Land snow depth. This relationship, referred to as subgrid variability, is then applied to downscale the monthly ERA5 Land snow depth data. We observe that the generic output should be calibrated by a small number of localized control points to restore full range snow depth map. The validation of downscaled snow depth data using dataset from meteorological stations as well as airborne laser scanning (ALS) in the Hardangervidda region of southern Norway show that the snow depth prediction achieved  $R^2$  values ranging from 0.74 to 0.88 for aggregated results. While the validation specifically focus on the Hardangervidda area which lies above the treeline, the snow depth distribution patterns derived from ICESat-2 and ERA5 Land exhibit potential applications in other regions. This approach optimizes snow surveys over large areas, reducing costs and risk exposure. Though requiring area-specific calibration, it builds upon and expands existing surveys, offering valuable data for permafrost, hydrological, avalanche, and climate ecological modeling tasks.

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

---

## 1. Introduction

In a warming world, understanding the spatiotemporal variations of seasonal snow is increasingly vital for climate impact assessments, meltwater supply (Immerzeel et al., 2020; Livneh and Badger, 2020), permafrost modeling (Gisnås et al., 2016) and ecological responses (Callaghan et al., 2011). Seasonal snow accumulates and melts away once a year in a number of climatic zones from, for instance, the forested regions of the taiga (largest terrestrial ecosystem), the open tundra of the Arctic and many high mountain ranges throughout the world (Sturm and Liston, 2021). Being white, snow is easily observable from space. However, its thin nature (typically less than 2 m) makes it very challenging to measure its thickness or mass from space at large scale and with high repeatability. The complex processes driving snow metamorphism and precipitation estimation are hurdles too to rely on model for assessing seasonal snow dynamic. Observing and modeling

---

\*Corresponding author

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

seasonal snow at large scale therefore remains a challenge (Tsang et al., 2022; Mudryk et al., 2020). This is exacerbated in remote and complex terrain with limited data availability (Bormann et al., 2018).

A major persistent gap in contemporary is observing reliable snow mass over mountain regions. The point-based station measurements are generally located in gentle terrain at lower or mid-elevation, which makes it difficult to represent the region with varying topography (Fassnacht et al., 2018). Space-borne radar technologies have several issues with footprint size, penetration, and volume scattering problems. For instance, the passive microwave radar, e.g. with a footprint size of up to 25 km, is insufficient to capture the heterogeneous snow depth in mountainous terrain (Tsang et al., 2022; Mudryk et al., 2020). Despite recent developments in C-band radar technology have enabled snow depth retrieval at sub-kilometer resolution (Lievens et al., 2019, 2022), the even finer-scale snow depth remains poorly addressed (Grünewald et al., 2010; Mott et al., 2018). Typically, high-resolution snow depth data can be obtained using laser scanning or photogrammetry. Either method does not directly provide snow depth but operates by collecting surface height data pre- and post-snowfall, enabling snow depth mapping through comparative analysis of height differences. Although both methods, employing airborne sensors on airplanes or UAVs (unmanned aircraft vehicles), offer very high-resolution data (Bühler et al., 2016), they also have several downsides: the high cost and localized focus of airborne surveys render it most effective for watershed-scale snow depth mapping (Deems et al., 2013), and the use of UAV requires access to the site and is limited to even smaller catchments. Recently, there has been a significant effort to retrieve snow depth by combining snow-free digital elevation models (DEMs) with space-borne lidar measurements of the snow surface from ICESat (Treichler and Kääb, 2017) and its successor ICESat-2 (Neuenschwander and Pitts, 2019). This cutting-edge satellite offers high-resolution, accurate lidar measurements of the Earth's surface, including snow-covered terrains. 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, which achieved MAD 0.14 m to 0.20 m and root mean square error (RMSE) of 0.18 m to 0.33 m for the Tuolumne Basin and Methow Valley, USA. These studies have uncovered ICESat-2 as an emerging and cost-efficient data source for snow depth but also have brought attention to challenges associated with the data and elevation differencing workflow, primarily stemming from discrepancies and spatially/temporally varying inconsistencies between reference DEMs and ICESat-2. Therefore, implementing and improving this workflow requires careful co-registration and bias correction on DEMs & ICESat-2. Additionally, the sparse nature of ICESat-2 measurements presents another significant challenge to creating comprehensive snow-depth maps: how can we extrapolate both spatially and temporally to areas outside of the ICESat-2 measurement profiles?

Another approach to understanding snow dynamics is through snow modeling. Researchers primarily use two modeling strategies to study these dynamics, namely process-based and statistical approaches. Process-based models (Lehning et al., 2006; Liston and Elder, 2006; Kim et al., 2021) incorporate physical processes, which are driven by meteorological forcing data and yield gridded snow depth products. However, these models are computationally costly and thus often struggle to cover large areas or offer fine resolution, and are 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; Parr et al., 2020). The consistent recurrence of this pattern supports the use of computationally efficient statistical approaches. Many studies seek to establish parameterizations for subgrid variability, such as snow depletion curves, snow depth elevational gradients, snow probability distribution (Mendoza et al., 2020b), subgrid snow depth coefficient of variation (Liston, 2004; He et al., 2019; Gisnås et al., 2016) or topographic correlations (Helbig and van Herwijnen, 2017). To produce a snow depth map from predictors at fine scales typically involves the use of regression models. Multiple-linear regression (Grünewald et al.,

2013; Dvornikov et al., 2015), binary regression trees (Revuelto et al., 2014), random forests (Revuelto et al., 2020) or convolutional neural network (Daudt et al., 2023) have been used to predict snow distribution patterns with varied 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 hardly be transferred to other catchments or season (Grünewald et al., 2013; Revuelto et al., 2020). Another category of statistical models capable of generalizing subgrid variability is more commonly known as downscaling models. These models are adept at refining data from coarse, broad-scale grids to finer, localized subgrid levels (Maraun, 2019), and they can extend this analysis both retrospectively and into the future given the stationary assumption of input snow depth data. Currently, there is a limited number of statistical downscaling models applied for snow depth (Helbig et al., 2024; Tryhorn and DeGaetano, 2013). The primary obstacles in this area include efficiently obtaining comprehensive snow depth measurements and accurately characterizing variability through relevant model features.

In light of these challenges and advancements, this work hypothesizes that it is possible to leverage ICESat-2 data, combined with ERA5-land data and high-resolution DEMs, to effectively generate snow depth distribution patterns, particularly at the hillslope scale, through a downscaling-calibration scheme. The objectives and workflow of this study are, as follows:

1. Retrieving snow depth from ICESat-2 laser altimetry data across mainland Norway.
2. Using this data to train a machine learning-based downscaling model that accommodates spatial and temporal variations of snow depth in mountain environments.
3. Applying calibration based on control points of downscaled snow distribution and validating the rest of the area at different scales through comparisons with in-situ observations, gridded snow model products, and meteorological stations.
4. Exploring the challenges encountered in our workflow.

The study area is chosen based on the availability of validation data, but the downscaling-calibration scheme is designed to be globally applicable where an accurate DEM is available and a proper calibration dataset is available. To retrieve snow depth measurements, we employed a DEM co-registration algorithm (GDC, Gradient Descent Co-registration) and corrected the bias of DEM using ICESat-2 snow-free measurements. Subsequently, we statistically downscaled ERA5 Land from its ~9 km native resolution into a finer scale using the XGBoost 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 use ICESat-2 data to statistically downscale ERA5 Land data.

## 2. Study Area and Data Setting

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 topography, 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 mountainous environments.

Hardangervidda, the largest mountain plateau in northern Europe, spans approximately  $6,500 \text{ km}^2$ . Its comparably flat terrain is nevertheless covered by hills and troughs exposed to high winds and heavy snowfall (Figure 1). The plateau predominantly lies above 1,000 meters above sea level (m a.s.l.), featuring a low alpine ecosystem with grass heaths, dwarf shrubs, and higher areas with bare rock or lichen marsh tundra. The eastern region is characterized by its open terrain with numerous lakes and streams. The western and southern areas, reach up to 1,700 m a.s.l. and act as significant orographic barriers to the prevailing westerly wind flow. As moist air masses encounter this mountain range, they are lifted and cooled, leading to increased precipitation on the windward slopes and a subsequent decrease of precipitation on the leeward side. Snow accumulation typically begins in mid to late September at higher elevations, peaking around late April. Mean annual precipitation ranges from 750 mm to as much as 3,000 mm over relatively short distances, with approximately 50–60% of this annual precipitation falling as snow (Ketzler et al., 2021).

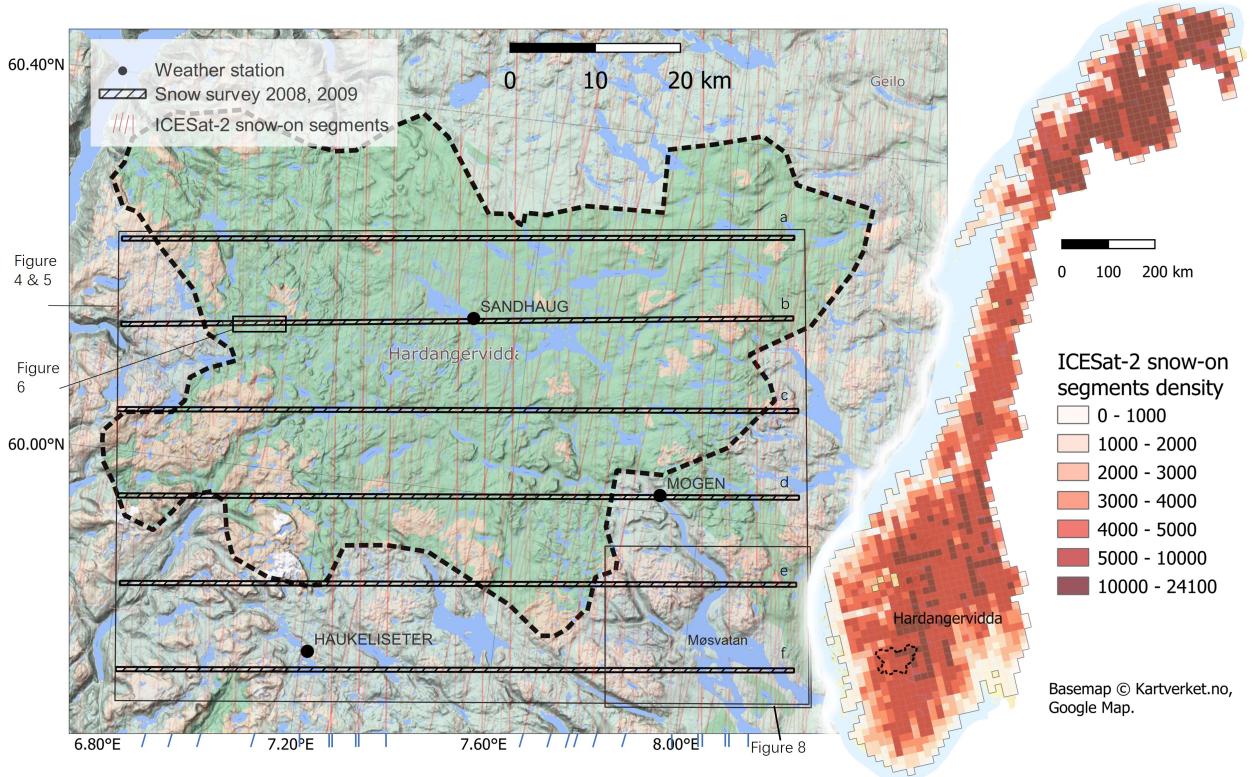


Figure 1: Map of the study area. Left panel: Snow-on data (ICESat-2 ATL08 segments) are displayed in red (all data from October 2018 to October 2022). The ALS survey has 6 flight lines (a-f). The blue ticks at the bottom highlight the ground tracks acquired during April. The right panel shows the total number of available snow-on data segments for mainland Norway.

### 2.1. ICESat-2 ATL08 elevation data

Launched in September 2018, ICESat-2 is equipped with the Advanced Topographic Laser Altimeter System (ATLAS), which provides photon-counting lidar measurements at a global scale (Neuenschwander and Pitts, 2019). ATLAS comprises three parallel beam pairs separated by 3.3 km on the ground. Each beam pair includes a strong and weak beam separated by 90 m. The ATLAS system emits a pulse every 0.7 m along the track, covering a circular footprint with a diameter of ~15 m. At mid-latitudes, ICESat-2 ground tracks are generally not repeated but shifted for each overpass to maximise global spatial coverage. The ICESat-2 ATL08 product (level L3A, version 5) (Neuenschwander et al., 2021) 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 height referring to best-fit terrain elevation at the midpoint location of the segment ( $h_{te\_best\_fit\_20m\_2}$ ). The subsegment heights are 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 tracks from April of one of the four years (as ground tracks are not repeated).

## 2.2. Snow-off elevation data

As reference ground for snow depth retrieval, we employ the Norwegian DTM1 elevation model (DTM1) ([Kartverket, 2022](#)), 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 from the same data provider, DTM10 ([Kartverket, 2022](#)). To demonstrate the workflow's applicability in areas without lidar-based elevation products, we incorporate global DEMs such as Copernicus GLO-30 ([European Space Agency, 2021](#)), hereafter referred to as COP30, and FABDEM ([Hawker et al., 2022](#)) (Forest And Buildings Removed Copernicus DEM, hereafter referred to as FAB) as reference ground. COP30 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 COP30, eliminates buildings and trees using the random forest algorithm, enhancing accuracy. FAB serves as a reference for comparison with COP30.

## 2.3. Large-scale reanalysis data

ERA5 Land hourly data (version 5) ([Muñoz Sabater, 2021a](#)) is an ECMWF (European Centre for Medium-Range Weather Forecasts) reanalysis product covering the period from 1950 to the present. It describes water and energy cycles over global land areas with over 50 variables at a spatial resolution of approximately 9 km ([Muñoz-Sabater et al., 2021](#)). 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; Pflug et al., 2021](#)). ERA5 Land's snow depth data (*sde*) represents an instantaneous average of the snow thickness on the ground, excluding snow on vegetation canopy ([Muñoz Sabater, 2021a](#)). Additionally, the ERA5 Land monthly ([Muñoz Sabater, 2021b](#)) dataset contributes instantaneous wind fields (*u10*, *v10*) at a height of 10 m above the land surface.

## 2.4. Validation data

Our validation methods include 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:

- The ALS survey: The survey by [Melvold and Skaugen \(2013\)](#) provides 10 m gridded snow depths data for two winters over Hardangervidda. The survey encompassed six flight lines apart in 10 km intervals, each extending 80 km in a the east-west 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 (snow-free reference). During the autumn collection period, the ground was in nearly bare condition except for few perennial snow patches ([Melvold and Skaugen, 2013](#)).
- The seNorge data ([www.senorge.no](http://www.senorge.no)) employs a snow model that predicts snow depth based on interpolated precipitation and temperature station observations (seNorge2018 v23.09) ([Saloranta, 2012, 2016](#)). It offers daily snow depth maps at a 1 km × 1 km grid resolution and is available via the public archive service Thredds<sup>1</sup> of the Norwegian Meteorological Institute (MET Norway). We aggregated the daily snow depth from seNorge into monthly average values.
- Meteorological stations: This study compared the snow depth timeseries with three available meteorological stations in the region (Figure 1). The weather station Sandhaug is located 50 m north of one of the ALS flight lines at an elevation of 1,250 m above sea level (a.s.l.). The other station Mogen (954 m a.s.l.) is situated directly along one of the flight line. Additionally, Haukeliseter (990 m a.s.l.) is positioned between two flight lines. Monthly mean snow depth data were retrieved from MET Norway's Frost API<sup>2</sup>. Due to harsh observing conditions in our validation area, all station observations carry a median confidence level indicated by a quality flag of 2.
- Sentinel-2 satellite imagery (L2A) ([Sentinel-2 \(processed by ESA\),, 2021](#)) was used for visual data quality checks and validation of the presence/absence of snow in the Lake Møsvatn area (Figure 1).

<sup>1</sup>MET Norway's Thredds API: <https://thredds.met.no>. Last access: Sep 11, 2023.

<sup>2</sup>Frost API, MET Norway's archive of historical weather and climate data: <https://frost.met.no>. Last access: Sep 11, 2023.

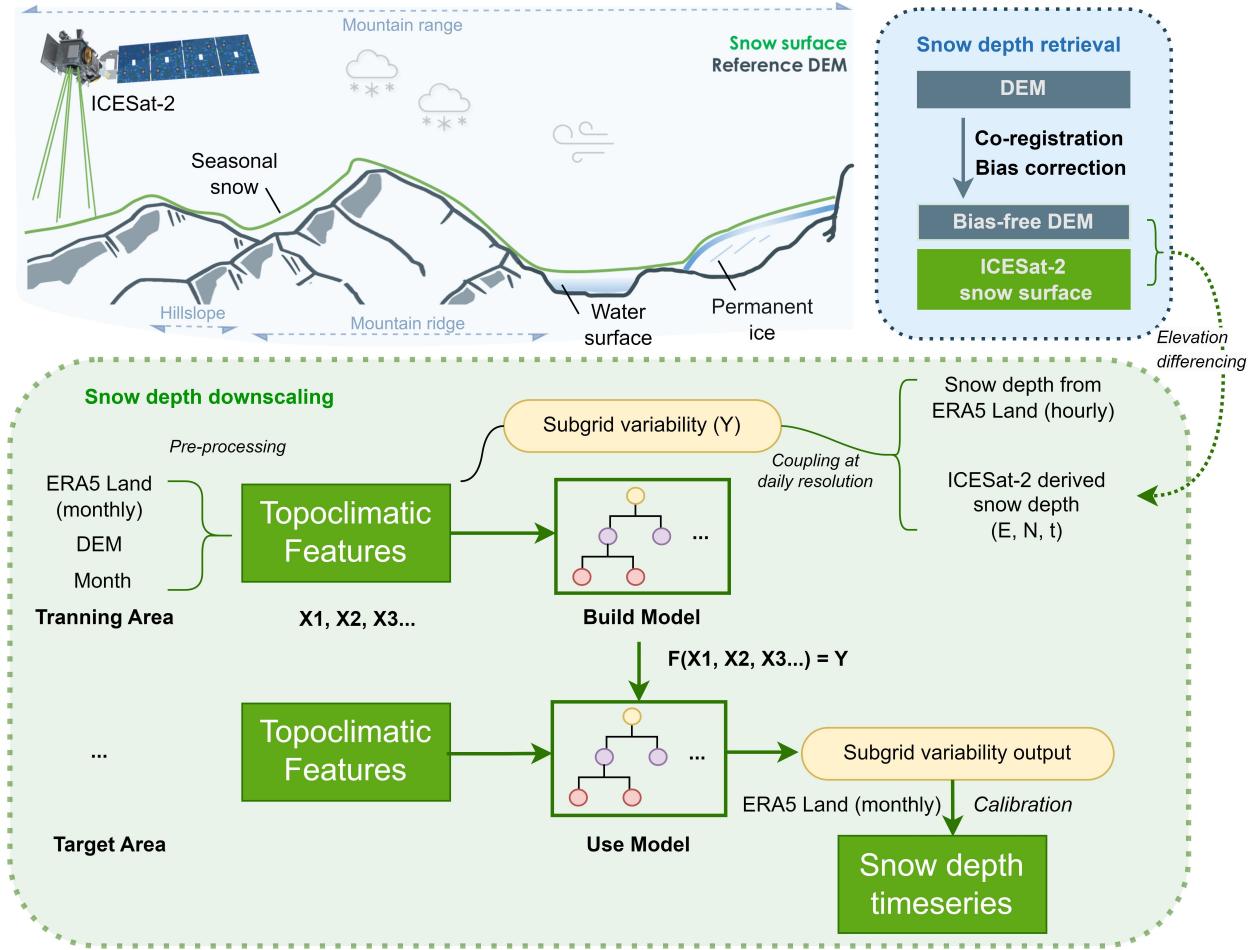


Figure 2: Flow chart of the snow depth retrieval and downscaling-calibration scheme. First, snow depth is extracted by elevation differencing. Consequently, a tree-structure-based regressor for downscaling is trained and implemented to predict local variability of snow depth in any location and at any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

### 2.5. Calibration control points

For calibration purposes, we identified 7103 points where ICESat-2 tracks overlapped with ALS strips collected in April 2008 (Figure 1). These points, distributed across the entire study area, were selected as control points to represent the study area while the rest of the area remains unseen to the downscaling model. The downscaling model's output is then calibrated against these selected control points to ensure accuracy.

## 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 mask. We used both the snow mask from NOAA daily snow cover product (Neuenschwander et al., 2022), and the brightness flag indicating the existence of snow cover at 100 m resolution. Snow-free segments were used for DEM co-registration and subsequent bias correction (Figure 2, see 3.1.1, 3.1.2). This method involved estimating the discrepancy ( $\hat{\Delta}_h$ ) between the reference DEM

( $DEM_{ref}$ ) and ICESat-2 Snow-free measurements:

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

Moving terrains such as water surface and permanent ice were excluded from the analysis using a land cover reference mask from the ATL08 product (based on 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 spatially and temporally incomplete information on snow dynamics. To overcome this, we resample snow depth from ERA5 Land at daily resolution and interpreted linearly on each desired point ( $E, N, t$ ), differentiating with the retrieved snow depth, to create a localized dependent variable, denoted as subgrid variability ( $Y$ ) (Figure 2). This variable represents the deviation from the gridded mean snow depth at each point and is used in the subsequent downscaling model (see Section 3.2). Our model predicts this subgrid variability and applies it to ERA5 Land monthly snow depth data to generate snow depth time series. Due to the inherent nature of the employed model, the raw outputs are biased towards average values while extreme values are underrepresented. Thus, a calibration step (see Section 3.2.3) was added to better represent the full range of the snow depth distribution. Data processing was done in Python with custom scripts that are available in a public GitHub repository (see Section 6) with the help of Xarray (Hoyer and Hamman, 2017) and Pandas (pandas development team, 2024). The retrieved snow depth for mainland Norway, spanning from October 2018 to October 2022 is available on Zenodo (DOI: 10.5281/zenodo.10048875).

### 3.1.1. Co-registration

Co-registration aligns a DEM with a reference dataset, with ICESat-2 snow-free data serving as a highly precise and spatially consistent reference. We used a computationally efficient gradient descent-based co-registration algorithm (for details see Supplement A). The process, facilitated by the open-source xDEM tool (Xdem contributors, 2021), was applied to each DEM tile individually, ensuring accurate alignment across all datasets.

### 3.1.2. Bias correction

Bias correction estimates discrepancies ( $\hat{\Delta}_h$ ) between DEMs and ICESat-2 data. Given that DEMs are often patched together from multiple datasets and various sensors, captured in different seasons and at different resolutions, they must be cautiously used as a reference ground surface (Hugonnet et al., 2022). Magruder et al. (2021) used ICESat-2 elevations to correct DEM, taking canopy and slope into account. On the other side, Tian and Shan (2021); Enderlin et al. (2022) found that the ICESat-2 ATL08 has more terrain height underestimation (negative skewness) compared to reference DEMs over steep terrain and thus proposed slope-dependent bias correction. This study does not assert which dataset represents ground truth accurately but focuses on quantifying the discrepancies so that we can exclude it from snow depth elevation differencing. Essentially, estimating the discrepancies ( $\hat{\Delta}_h$ ) between the two datasets is a regression problem, which is dependent on terrain parameters, vegetation conditions, and quality metrics from ICESat-2. To tackle this, we employ the XGBoost<sup>3</sup> (Chen and Guestrin, 2016), a gradient-boosted decision tree (GBDT) algorithm. The regression model trained using ICESat-2 snow-free measurements, is used to predict discrepancies ( $\hat{\Delta}_h$ ) for all other DEM grid cells where no ICESat-2 snow-free data exists, further explained in Supplement B. Bias correction successfully removed a significant negative skewness in the elevation difference histograms (Figure B.14) that was present before bias correction. In addition, considerable dependence on first and second-order DEM derivatives (slope and curvature) was detected and removed. After bias correction, NMAD values between ICESat-2 and all DEM datasets are reduced to below 0.62 m.

### 3.2. Snow depth downscaling-calibration

The XGBoost algorithm is also subsequently employed for the downscaling task to derive fine-scale, localized snow depth predictions from coarser-scale input data. XGBoost has demonstrated its effectiveness

---

<sup>3</sup>The XGBoost (version 2.0.0) library can be accessed at <https://xgboost.readthedocs.io/>

in downscaling tasks, such as total water storage anomaly from satellite gravimeter (Ali et al., 2023), precipitation (Zhu et al., 2023) or wind speed (Hu et al., 2023). We use 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; see section 3.2.1), curvature, planform curvature (*planC*), profile curvature (*profC*), cumulative wind-aspect factor ( $W_{uf}$ , see section 3.2.2) 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) at 10 m resolution, consistent with the model's output resolution. The wind fields are from ERA5 Land (monthly), the other features don't change over time. We assume the stationarity of ERA5 land, allowing us to apply the downscaling model trained on snow observations from 2018 to 2022 to other periods.

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) Square error (reg:squareerror) as a loss function to estimate the conditional mean of the target variable, this was the main method used to generate the spatially distributed maps. (2) Quantile regression (reg:quantileloss) to give probabilistic predictions, such as conditional median (Q50), Q25 and Q75, which are used for point-based downscaling at weather station locations, to gain insights into the uncertainties of downscaling over time and validate the representativeness of weather stations compared to the model output. The adoption of quantile regression is inspired by its successful application in probabilistic forecasting (Meinshausen, 2006; Zhang et al., 2018), where it is valued for its non-reliance on the distributional assumptions of the target variable and its robustness to outliers. This makes it an ideal choice for our downscaling task, where capturing the full range of possible snow depth values under varying climatic conditions.

### 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 neighbouring 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 neighbouring 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, we used two additional indices: tpi\_9 (calculated in 9 x 9-pixel windows, equivalent to 90 m x 90 m) and tpi\_27 (270 m x 270 m).

### 3.2.2. Cumulative wind-aspect factor

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  $\text{dir}_{wind}$  is the direction of the wind origin with northerly wind (blowing from north to south) referred to  $0^\circ$ . This study further divides  $W_f$  into leeward  $W_{f_{positive}}$  and windward factors  $W_{f_{negative}}$ , multiplied by wind speed to the power of three (Figure 3 c) to capture the cumulative effect of wind redistribution for each water year period.

$$W_{uf_{positive}} = \sum W_{f_{positive}} u_{wind}^3$$

$$W_{uf_{negative}} = \sum W_{f_{negative}} u_{wind}^3$$

where  $u_{wind}$  is the monthly average wind speed from ERA5 Land, linearly interpolated to 10 meters resolution. The accumulation begins in September from zero until the next August 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.

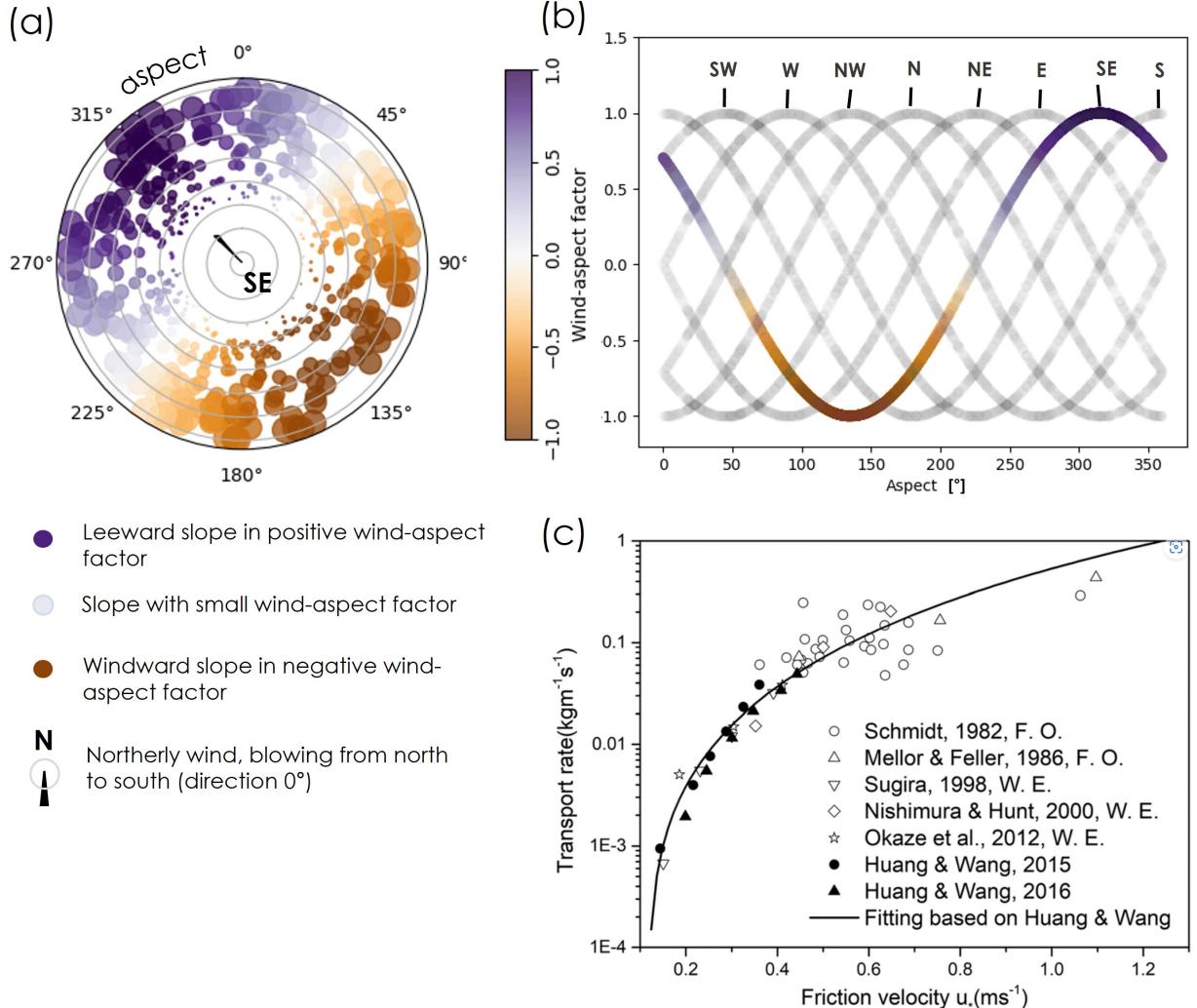


Figure 3: Quantifying the relationship between wind, aspect and snow redistribution. The prevailing wind, e.g. from SE, results in negative value on the windward side and positive values on the lee side(a). Eight cardinal directions are plotted but the function works on any wind direction (b). Research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

### 3.2.3. Calibration

In our study, the XGBoost is tasked with predicting snow depth under a variety of conditions. We observed that XGBoost tends to produce a more conservative conditional mean, particularly when the feature set does not fully capture variables critical to extreme snow depth. This conservative tendency is a natural outcome of the model's objective to minimize overall prediction error, often resulting in a distribution that skews

towards average conditions and under-represents the extremes (known as scaling bias, further discussed in 5.2). To address this issue, while preserving the relative snow depth predictions by the downscaling model, the calibration removes the scaling bias for each specified quantile by scaling factors calculated from control points, ensuring a localized full range of snow depths. Such calibration, based on quantile mapping(Cannon et al., 2015; Li et al., 2010), is done as follow:

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

$$\Delta = \frac{x_{adj}}{x}$$

Here,  $x_{adj}$  denotes the calibrated prediction. For a given percentage, as determined by the cumulative distribution function (CDF) of the downscaling model output  $F_M(x)$  and that of the observational control points  $F_O(x_{adj})$ , the discrepancy between two distribution's quantiles is encapsulated by a fixed scaling factor ( $\Delta$ ). This scaling factor is subsequently applied to the model output, aligning it with the statistical characteristics of the observational control points, and thereby mitigating scaling biases.

### 3.3. Model interpretation 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 feature can either play for or against the prediction (positive or negative SHAP value), and the magnitude of its SHAP value shows how significant its role is. The sum of all these contributions, plus the base prediction, gives us the final prediction.

For equitable comparisons, we aggregated high-resolution validation datasets to the same resolutions. Model validation was conducted across two different scales. At the microscale/ site scale, we captured snow distribution over typical micro-terrain features such as rills and gullies using the original model output resolution of 10 m. At the mesoscale, we aggregated data into a 100 m x 500 m grid, reflecting the 500 m width of the ALS survey swath, with 100 m intervals in the transection direction to enable effective depiction of snow variation across different aspects of hills and ridges. Successful prediction results in a (near-) perfect data match in probability distribution and ranking correlation. To quantify downscaling performance, we employ four key statistical metrics. RMSE and  $R^2$  scores evaluate the overall accuracy and fitness of the model. Spearman's rank correlation coefficient ( $\rho$ ) is used as a fidelity metric, with a high  $\rho$  indicating good similarity in spatial distribution. The Kolmogorov-Smirnov D statistic (KSD) quantifies the degree of probability distribution matching, with KSD = 0 indicating a perfect match.

Additionally, we used variograms to quantify the model's ability to capture the spatial heterogeneity of snow depth. The semi-variance ( $\gamma$ ) 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 ( $l$ ) (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 variogram indicates the rate at which correlation decreased with distance. By fitting variograms to the sum of the spherical model and Gaussian model from short to long ranges (as did by (Rolstad et al., 2009; Hugonnet et al., 2022) ), we identified spatial correlation of snow distribution at different scales.

## 4. Results

### 4.1. Mesoscale snow depth Variability

Figure 4 shows the snow depth maps for April 2008. Model input data, i.e., linearly interpolated snow depth from ERA5 Land (a), only represents large-scale variability. After downscaling, fine-scale snow

depth variability aligns with the topography both at microscale (10 m, b) and aggregated 1 km (d). In comparison, seNorge data (c) at the same spatial resolution appears smoother with less spatial variability and overestimates the depth in the western mountains (less so in April 2009, see [Supplement E](#)).

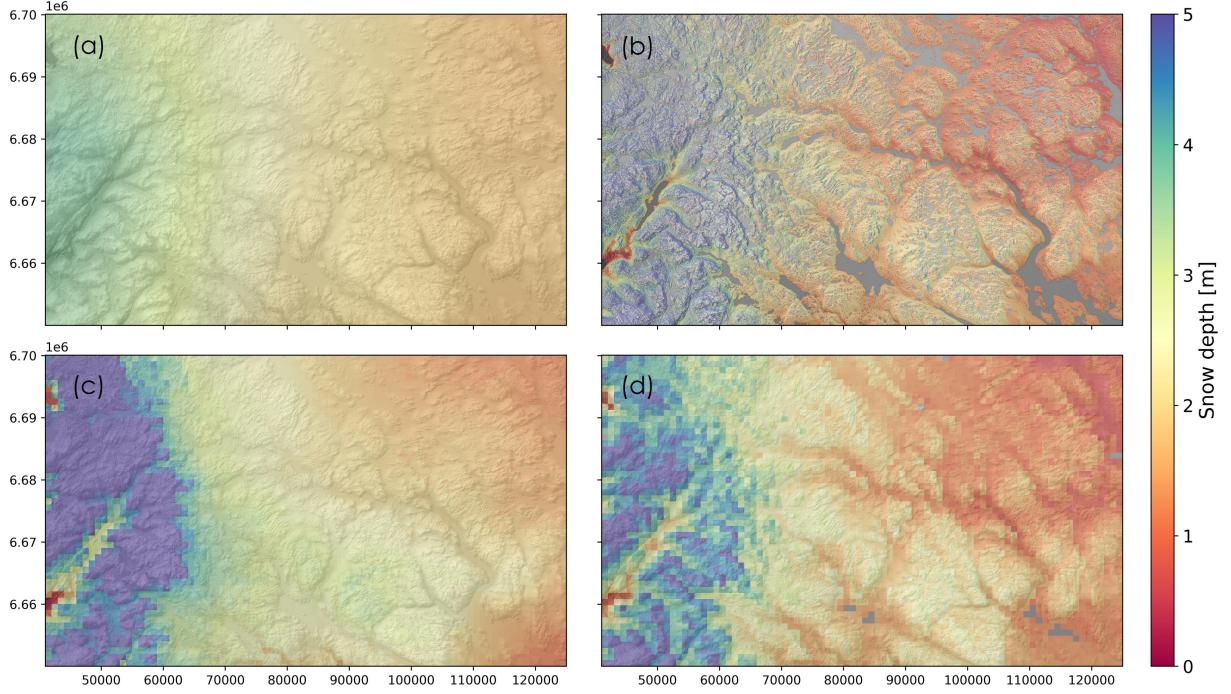


Figure 4: Spatial distribution of snow depth in the Hardangervidda area, April 2008. (a) Downscaling model input, linearly interpolated snow depth from ERA5 Land. (b) Downscaled snow depth output at 10 m resolution. (c) seNorge snow depth at 1 km resolution. (d) Downscaled snow depth aggregated to 1 km resolution.

Figure 5 shows snow depth transects at the mesoscale for April 2008. The ALS snow survey, downscaled output (based on DTM1, after calibration) and elevation are aggregated to the mean value of 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 depths exceeding 6 m to around 2 m.

The ALS snow survey reveals impressive details over slopes that are not depicted from elevation profile. The downscaled snow depth captures such variability among various slopes in close proximity. The seNorge shows about 40 % overestimation on the western mountain ridges but not in the East. [Melvold and Skaugen \(2013\)](#) attributed the overestimation to the reliance of the seNorge model on weather stations located in low-elevation populated areas. Moreover, the variability on either side of the ridges is not captured by seNorge data (1km resolution), addressing the importance of the resolution for snow distribution over hillslopes. Among the six flight lines, the calibrated downscaling model scored  $R^2$  between 0.71 and 0.88, and RMSE 0.76 m to 0.44 m, outperforming seNorge as expected.

Upon closer inspection, ALS ground truth data exhibits many spikes on the east-facing slopes instead of the crest, likely resulting from wind redistribution and gravity processes. Our downscaling model acknowledges the importance of these factors (Figure 9 c,g) but does not reproduce corresponding snow distribution in the correct places. Meanwhile, ERA5 land snow depth data has a systematic bias in the north-south direction, from overestimation to underestimation relative to ALS, which is also noticed by the downscaling model (Figure 9 e). For instance, the perfect match of the first flight line in the eastern open area between the ALS survey and downscaled output indicates that the overestimation by ERA5 Land is the stationary basis fixed

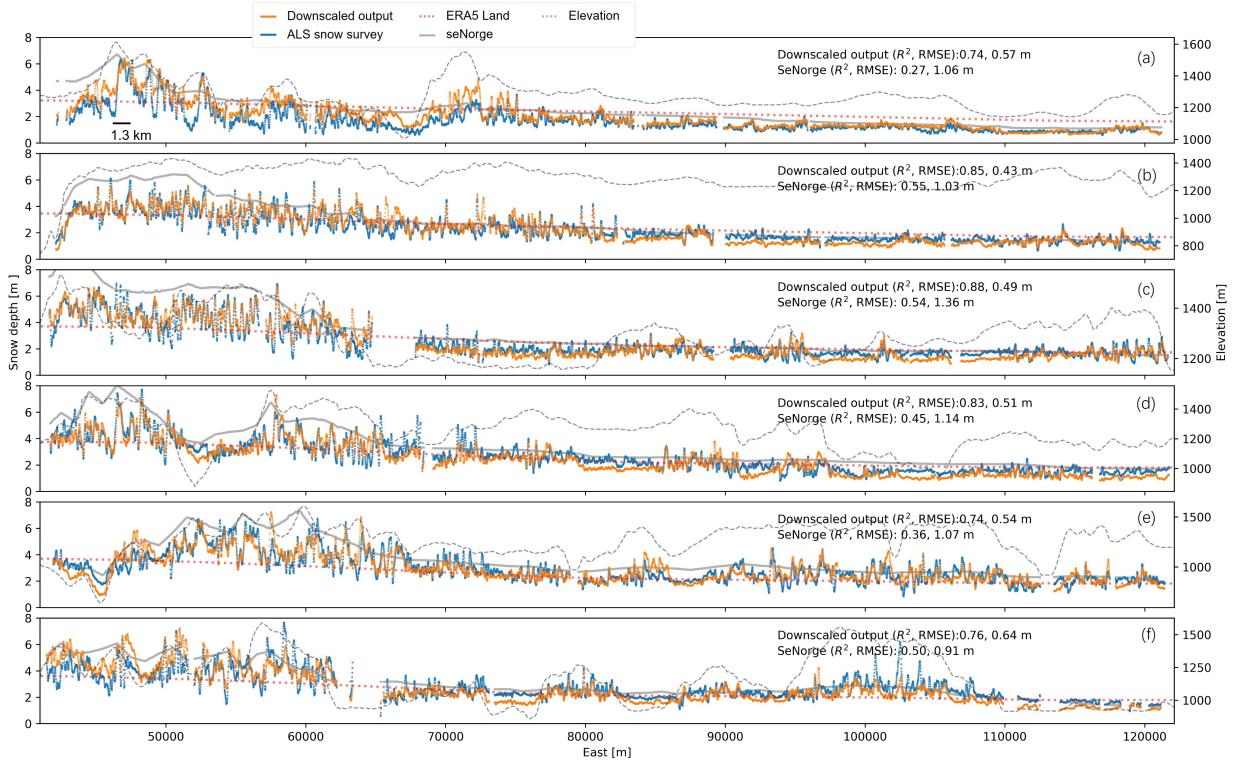


Figure 5: Snow depth profiles across Hardangervidda in April 2008 at mesoscale along six 500 m-wide ALS flight lines in W–E direction, from north to south (panels a-f). ALS/downscaled data points are aggregated to 100 m x 500 m cells (because of the ALS survey's 500 m transect width).

by the model. However, 10 km southward on the second flight line, the ERA5's 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.

#### 4.2. Microscale snow depth variability

Figure 6 compares the 2008 ALS snow survey with the downscaled snow depth at a 10-meter resolution. The examined region, located in the western part of flight line 2 (indicated in Figure 1), 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). These terrain features align with the spatial lag that exhibits significant autocorrelation (detailed variogram analysis available in 10 e). The transect line (marked in white) of downscaled output across this varied terrain visually captures most of the observed pattern. While the Spearman's  $\rho$  of 0.77 indicates strong rank correlation, the relatively low  $R^2$  value of 0.34 suggests a reduced statistical agreement compared to the mesoscale analyses. The difference map exhibits an RMSE of 1.33 m, aligning closely with the 1.28 m RMSE observed along the analyzed transect. Given that the calibration is being applied to the entire study area not specifically to the examined region, there is still the KSD 0.12 for the transect line and a mean deviation of 0.28 m for the difference map. These residual errors appear to be correlated with terrain features, raising a concern that the DEM is not entirely snow-free, which is confirmed from comparing the DTM1 with ALS snow-off data Figure C.16 e).

#### 4.3. Temporal variability of snow Depth

The comparison to time series focuses on the downscaling model's ability of generalization to interpolate in time rather than space. Note that for this purpose, the uncalibrated model output is shown in Figure 7, to

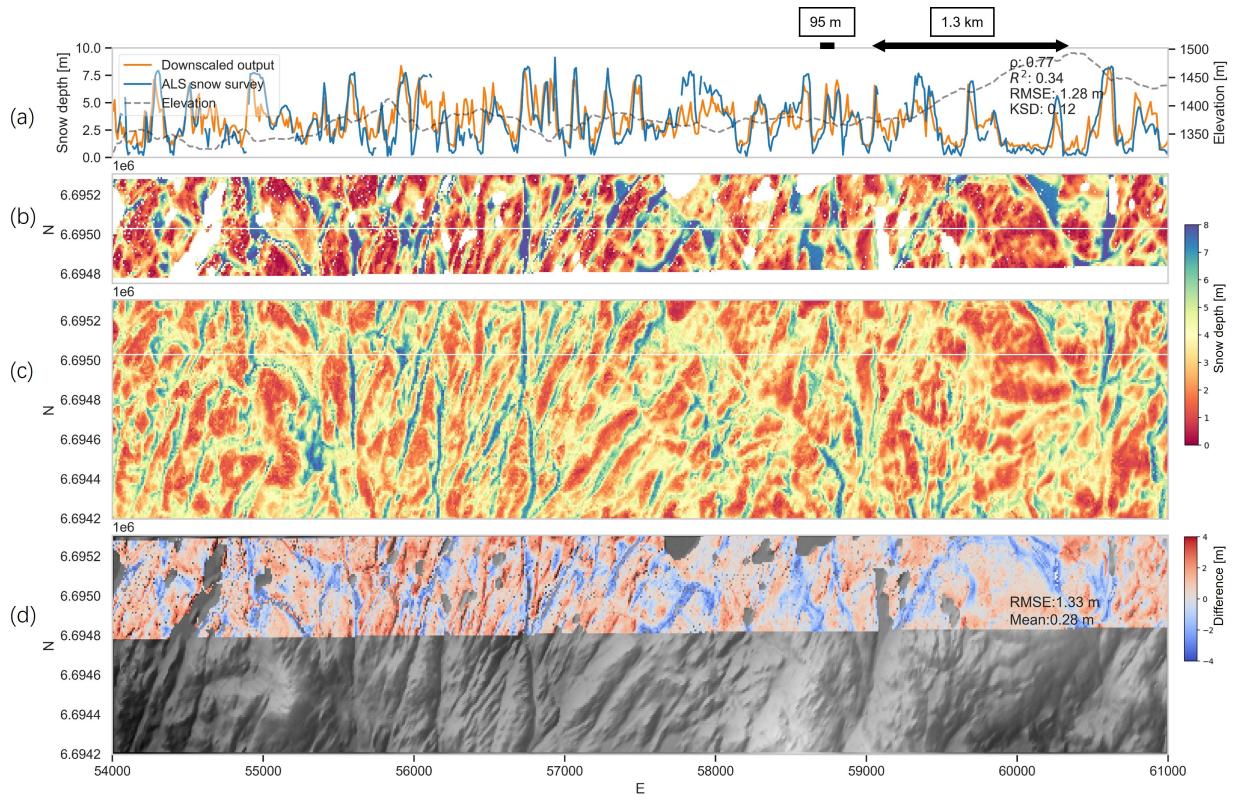


Figure 6: Microscale snow depth comparison in west Hardangervidda, April 2008. (a) Transect line, marked white in panels b/c, (b) ALS snow survey validation strip with significant snow depth variations, (c) downscaled snow depth from the DTM1 model with extended area, (d) differences between the ALS data and model output. The common horizontal scales of landforms are symbolized by black bars.

enable analysis of the local bias and its evolution over time. Time series data are shown for the data cell (10 m pixel size) corresponding to the location of three stations within the Hardangervidda area: Sandhaug (a), Mogen (b), and Haukeliseter (c). The predicted value at the median quantile (Q50) corresponds to the estimated central value and the interquartile range Q25–Q75 (IQR, shown in yellow) to the estimated uncertainty of snow depth predictions. Visually, Q75 shows a good match with the weather station data whereas input ERA5 data over- and Q50 model data underestimate measured snow depths. The  $R^2$  values of Q50 range from -0.43 (Mogen) to 0.86 (Haukeliseter), thus don't describe model performance well. Bias is primarily observed for high snow depth values during the peak snow season whereas the snow-free season is captured accurately. The peak snow bias has a consistent magnitude over time at each site and there are no indications for a better fit for the period with training samples (late 2018–late 2022) compared to earlier years. While the model exhibits bias in predicting the absolute snow depths, the high Spearman's  $\rho$  (0.91 to 0.94) across all three stations suggests that the downscaling successfully maintains the relative ordering of snow depth time series.

The right panels in figure 7 show the IQR of modelled snow depths (temporal average) for the area surrounding the weather stations. Low values indicate more typical topoclimate conditions that are ‘easier’ to predict for the model, while areas with high values show greater model variance. The IQR maps can be seen as a proxy for how similar/representative the weather station location is compared to the surrounding terrain. Our results suggest that Sandhaug may have better locations than Haukeliseter and Mogen (b).

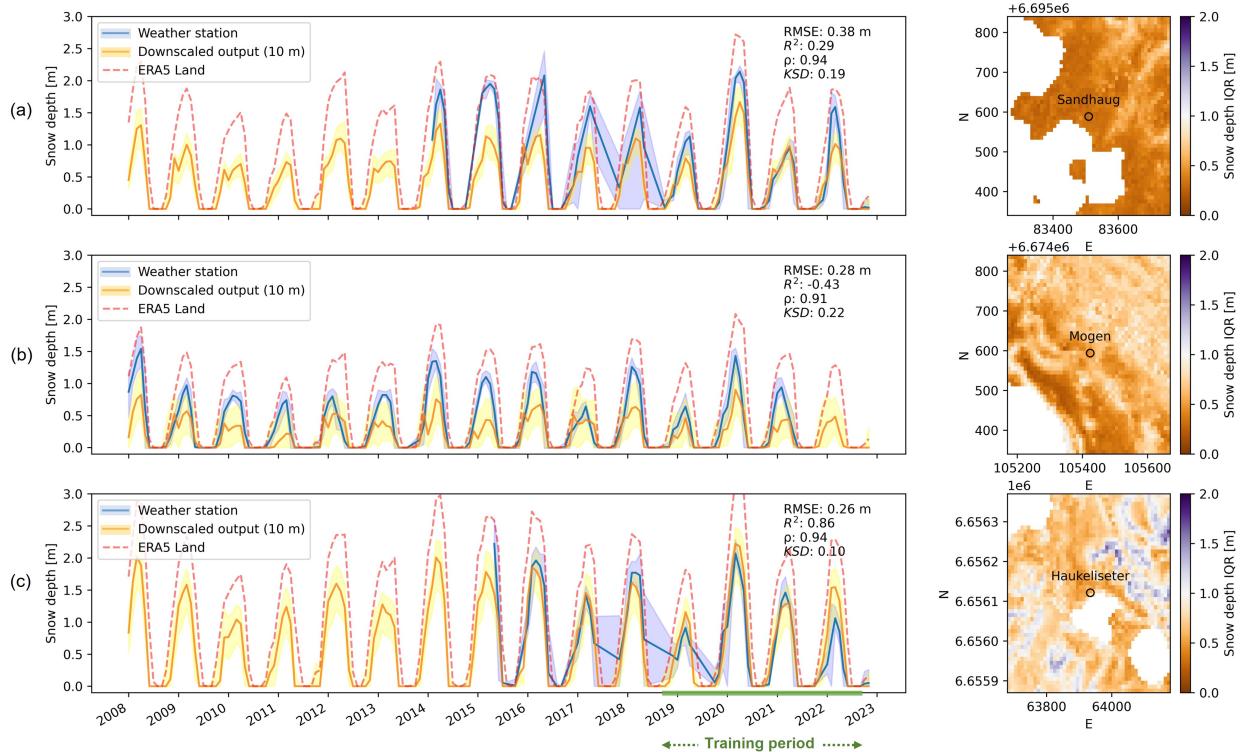


Figure 7: Time-series validation of snow depth for three weather stations Sandhaug (a), Mogen (b), and Haukeliseter (c). Blue: measured monthly mean and minimum/maximum values (blue shadow), orange: median downscaled snow depth (Q50) and interquantile range (IQR) for 25th–75th quantiles (yellow shadow). Note: Haukeliseter and Sandhaug data availability start from 2015 and 2014, respectively, and some years have incorrect measurements. Right panels: average IQR for the time series, a proxy for snow depth variability. Lakes are excluded from the maps.

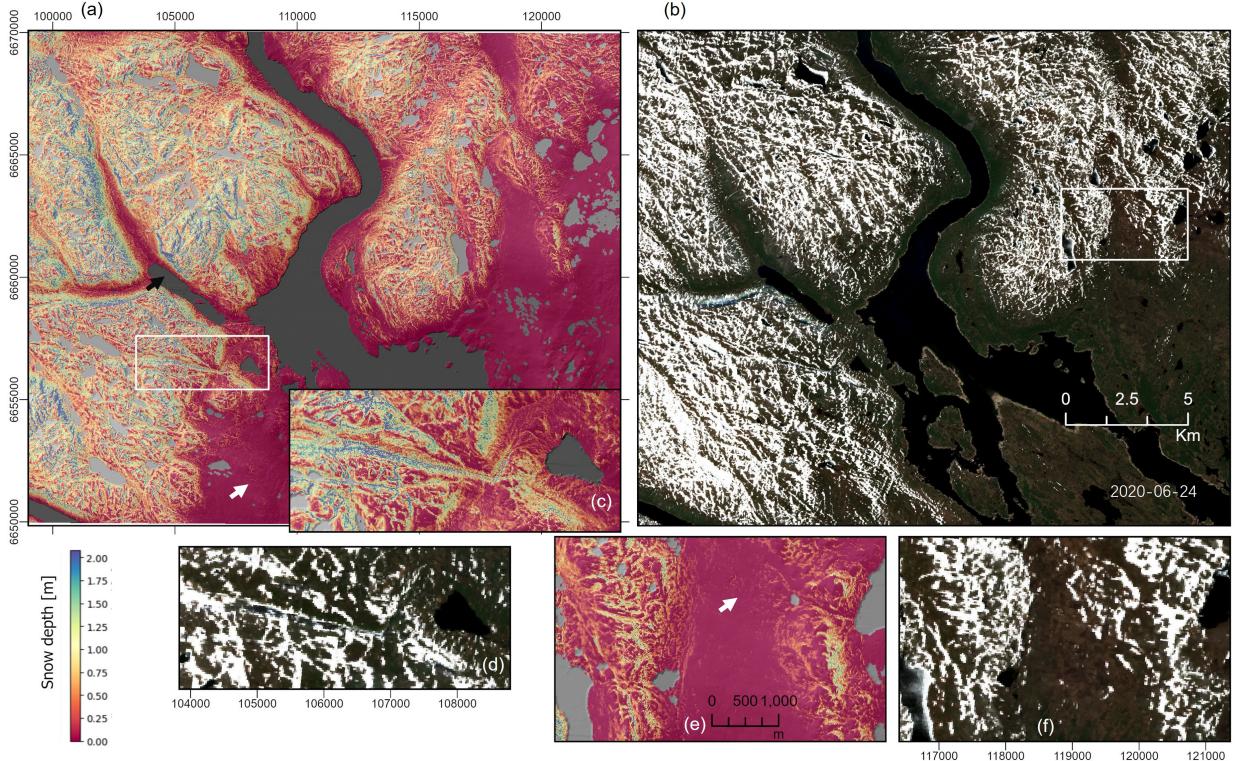


Figure 8: Validation of snow occurrence using Sentinel-2 imagery for the Lake Møsvatn area. (a) modelled snow cover distribution for June 2020, (b) Sentinel-2 image on June 24th, 2020. Panels (c) and (d) show local discrepancies in snow distribution, (e) and (f) show correctly modelled snow retention 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.

#### 4.4. Validating snow occurrence

Figure 8 provides a visual comparison between the model output and Sentinel-2 imagery for June 2020, showing the model's accuracy in predicting the occurrence of snow during the rapid melt period. The downscaled snow map from our model visually shows a high degree of agreement with the satellite snow extent on June 24, 2020, capturing snow patches that align with topographical features (Figure 8 c,d). 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. Some, but not all of these differences can be explained by temporal mismatch, with the Sentinel-2 image taken on a specific day and the model output representing a monthly mean for June. In particular, the model tends to systematically 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.

#### 4.5. SHAP analysis and variogram assessment

The most informative relationships between the input features and the predicted subgrid variability are listed in Figure 9 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. In contrast, thinner snow is associated with convex and west-facing slopes (Figure 9 b,c). The aspect-dependent preferential

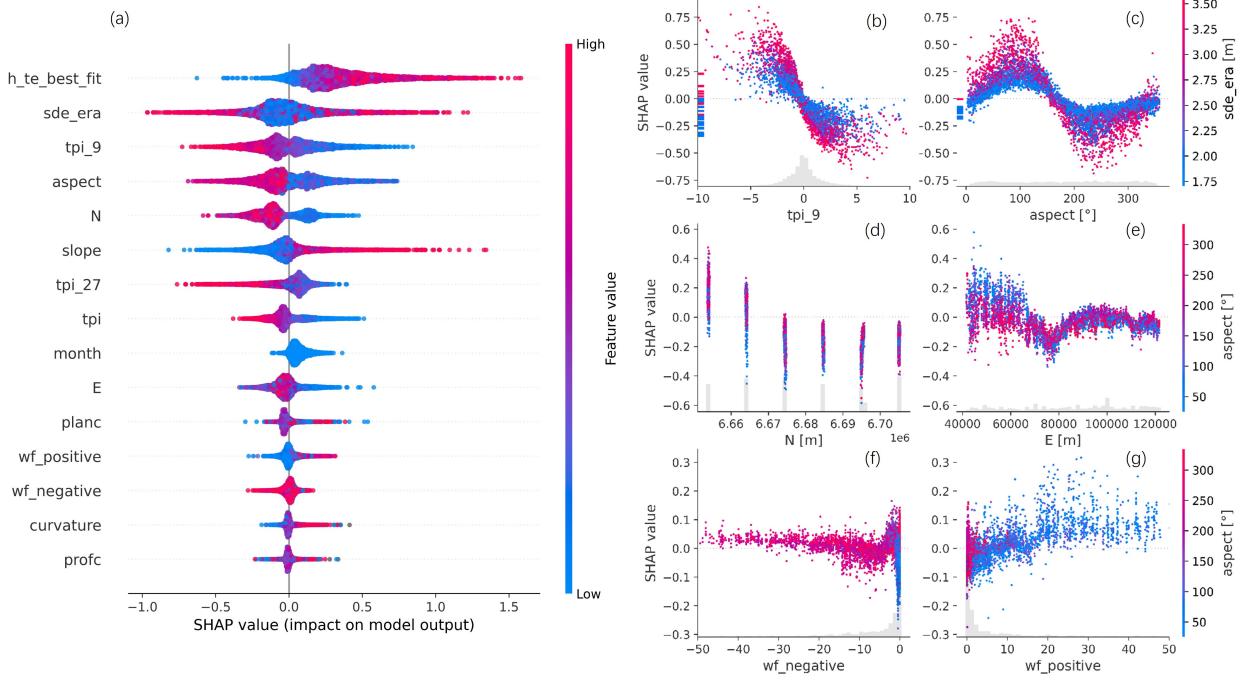


Figure 9: SHAP interpretation for the downscaling model. Panel (a) ranks features by their impact on subgrid variability. Subsequent panels (b-g) depict SHAP dependence plots, illustrating how each feature influences model predictions at the validation site. Each dot represents an individual sample of subgrid variability, providing insights into feature-specific margin effects.

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 9 e). The northern part of the study area appears to be overestimated in snow depth, while the southern part is underestimated (Figure 9 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 factor at zero, indicative of low wind speeds, shows much variability, indicating unexplainable variance (Figure 9 f,g). 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.

The empirical variogram in Figure 10 suggests stronger spatial dependence at short ranges for ALS data (79% variance correlated in 95 m lag) than downscaled snow depths (ca. 60% of variance in 108 m lag), with slightly higher semi-variance values for the national DEMs than the global DEMs. The distance of ca. 100 m corresponds to the wavelength of depression features commonly observed in the area. Overall, the downscaling model exhibits 13% less variance than ALS data, suggesting a slight underestimation of spatial variability by the model. At a distance of ca. 1,300 m, the semi-variance for both datasets approaches the sill. This distance aligns closely with the typical size of hillslopes in this region, as depicted in Figure 6 (a) and Figure 5 (a).

SeNorge is an interpolated product based on precipitation gradients and temperature lapse rate, which are dominated by elevation. Wind processes are not considered in this model. It can explain 69% of the total variance at a spatial lag after 16,000 m. This distance corresponds to the maximum local relief from the mountaintop to the lake surface (Figure 5).

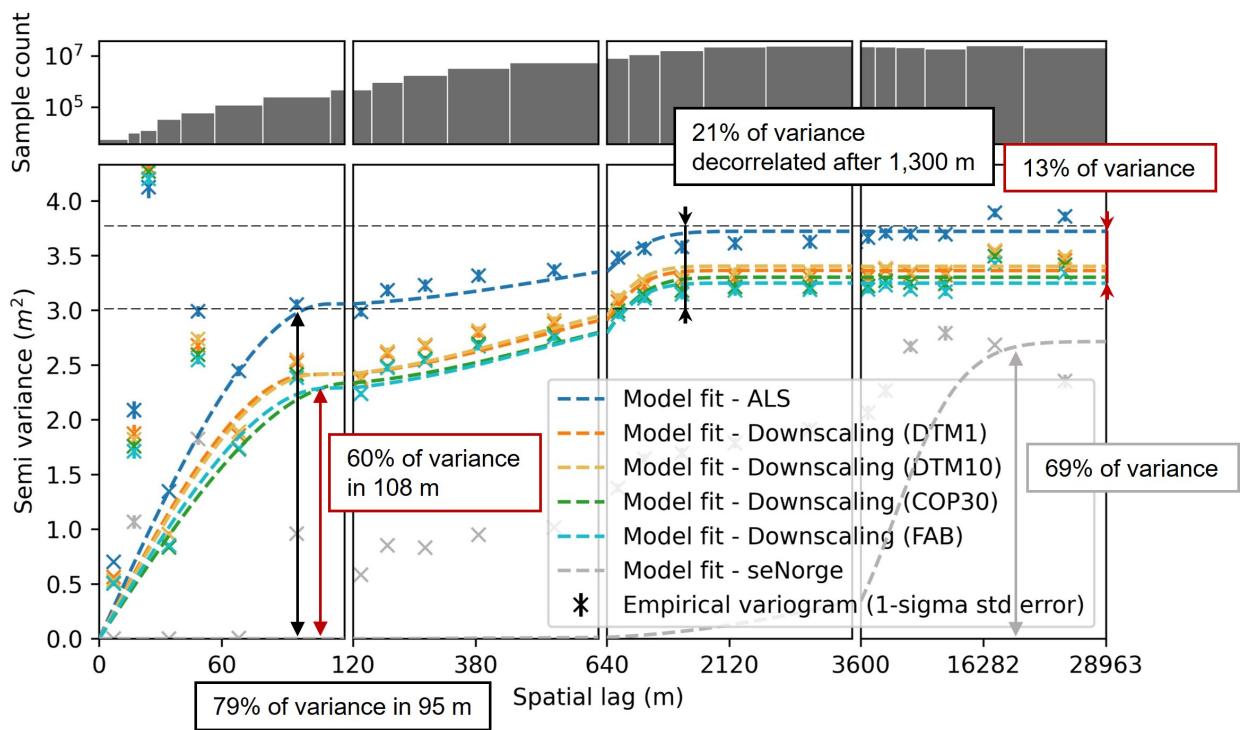


Figure 10: Spatial correlation from variograms based on six ALS validation strips (2008) and downscaling model outputs at 10 m resolution. The variograms are fitted to the sum of the spherical model and Gaussian model from short to long ranges. Note that the variogram's limited capture of variance in the transverse direction is due to the ALS strips' constrained width.

## 5. Discussion

### 5.1. Bias in elevation datasets

This study shows that ICESat2 ATL08 data is not just feasible to measure snow depths, thus confirming previous findings by other authors ([Deschamps-Berger et al., 2023](#); [Enderlin et al., 2022](#); [Tian and Shan, 2021](#); [Besso et al., 2024](#)), but it can be used to produce snow depth time-series at high resolution after proper bias correction and downscaling-calibration process. Previous studies have in common that retrieved snow depths are biased by remaining discrepancies and errors in the DEM/ICESat-2 data that are not removed by co-registration, such as slope-depending bias. The bias correction model introduced in our workflow is able to efficiently remove errors that depend on first or second-order derivatives (slope/curvature). Several studies pointed to ICESat-2 as the primary source of this bias, as ICESat-2 ATL08 data tends to underestimate surface height under certain conditions ('negative skewness'). [Moudrý et al. \(2022\)](#) attributed the error to the presence of clouds and corresponding increasing atmospheric scattering effects that lead to an increased photon travel time and, consequently, underestimated terrain height. Furthermore, the 100 m segment length of ATL08 (e.g. `h_te_mean` used by [Enderlin et al. \(2022\)](#)) is considered insufficient for accurately mapping steep and rugged terrain ([Besso et al., 2024](#)). Therefore, [Enderlin et al. \(2022\)](#) recommended only using ATL08 in mountain areas with relatively low slopes and sparse vegetation cover. We indeed find that bias is higher in the Hardangervidda area than in other, flatter parts of Norway, such as Finnmark (not shown in this paper). And a large number of 'negative snow depth measurements' were present before bias correction, and directly excluding these negative snow depths would result in unbalanced training samples. Or, a correction solely based on the slope, as proposed by ([Enderlin et al., 2022](#); [Tian and Shan, 2021](#)), might be not enough to address the bias. Therefore, in this study (1) we used `h_te_best_fit_20m_2`, which is the subsegment in the middle of the segment, and has a smaller footprint size than other attributes (see [Supplement B](#)). (2) We find that different biasing effects depend on the resolution of the tested DEMs with increasing curvature should be removed to make sure less skewness propagated to snow depth retrieval ([Supplement B](#)).

Also, the snow-off DEMs hold potential limitations for ICESat-2-derived snow depths in certain terrain, especially areas with a long snow cover duration that is unlikely to be completely free from snow when DEMs are obtained. The remaining snow on the supposedly snow-free reference results in an underestimation of snow depth in parts of these areas. Furthermore, there are also fewer snow-free ICESat-2 segments available in such locations. When comparing our downscaled snow depths with the ALS validation data, we find that on the microscale, the residuals are highly associated with a preferential deposition pattern. The systematic underestimation of high snow depths is even more prominent in the downscaled output if the calibration step is not applied (Figures [C.15,C.16,7](#)). This systematic error pattern indicates that the Hardangervidda area was not snow-free when the reference DTM1 was obtained, which is confirmed by comparing the DTM1 with ALS snow-off data Figure [C.16 e](#)) and entirely plausible, given that much of the data for DTM1 in the Hardangervidda was acquired during June or July (DTM1 metadata, available from [hoydedata.no](#)). The presence of remaining snow patches in supposedly snow-free DEMs is not limited to the DTM1 dataset, we observed this for all four tested DEMs and also the Arctic DEM (([Porter et al., 2022](#))) (not shown), as expected for higher elevations or Arctic areas where the snow-free season is short. Here, we indirectly correct for the biasing effect of remaining snow patches on retrieved/downscaled snow depths with the calibration steps, which also address other biasing effects (see section [5.2](#)). Future work should thus include a targeted correction of the DEM data for remaining snow patches. This would likely improve our results and potentially remove the need for the currently included calibration step.

[Enderlin et al. \(2022\)](#) deemed Copernicus GLO30 data too imprecise for estimating snow depth. Encouragingly, our results demonstrate that the COP30 and FAB DEMs have improved after the bias correction. It performs comparably with DTM1 and DTM10 at mesoscale in treeless areas. We attribute this to the bias correction step in our workflow which successfully removes bias for a group of measurements — but less accurately for single data points. We therefore recommend to use global-scale DEMs for aggregated representations of snow depths and prefer curated, high-resolution datasets such as DTM1 for microscale snow depth maps in areas where such data exists.

This study only validates the downscaled snow depths above the tree line and further work is needed to assess the performance of the proposed workflow in forested areas where ALS validation data is available. We expect that more or different features may need to be added to the model workflow for vegetated areas, and results will more strongly depend on the used DEM. High-resolution lidar data allows for removal of vegetation, but global-scale DEMs are usually based on photogrammetry or radar interferometry data that include all or parts of vegetation cover in the elevation data. There are efforts to create global-scale vegetation-free DEMs like the FAB DEM included in this study, and further developments are likely with ever-improving machine-learning algorithms. Currently, for the high elevation accuracy required for snow depth retrieval, we recommend being cautious about using COP30 and FAB in vegetated areas, where the vegetation is hardly been (fully) removed.

### *5.2. Challenges of snow depth spatial regression*

Our downscaling model leaves a significant amount of variability unaccounted for, particularly in periods/areas with high snow depth, unless calibration is applied. These limitations arise from several factors: the model's inherent limitations, sample unbalance, sample density, and features:

Tree-structure models used in this study primarily yield conditional means and struggle with extreme values, especially when key features do not fully explain variations. This often results in underfitting as the models strive to minimize overall prediction errors. For example, while snow avalanche occurrences lead to extreme snow accumulation, and are influenced by terrain parameters, these parameters (avalanche terrain) alone fall short of locating specific avalanche events. Consequently, a model like XGBoost tends to predict avalanche hazard zones, rather than pinpointing individual avalanche occurrences without additional, informative features.

The spatial distribution of snow depth is inherently scale-dependent ([Melvold and Skaugen, 2013](#); [Mott et al., 2018](#)), meaning a sensor's spatial resolution is critical in determining captured variability. Our training samples, derived from the ICESat-2 ATL08 elevation product, corresponding to an area of ca.  $20\text{ m} \times 15\text{ m}$ , which is better than previous studies cases but still averages out finer-scale variability. ICESat-2's sparse spatial sampling pattern with several kilometres gap between sampled snow depth profiles combined with a 92-day revisit period of the satellite and a high chance of cloud cover leads to gaps in seasonal data coverage. Data points with high snow depth measurements are thus more likely to be underrepresented. Furthermore, ICESat-2's elevation data, derived from a ground-finding algorithm that utilizes a probability distribution function (PDF) of reflected photons ([Neuenschwander et al., 2022](#)), can vary in clarity depending on the terrain. Rugged terrains, therefore, tend to be more affected by noise and undersampling, leading to a training dataset that may not reflect the environment's actual variability.

The ERA5 Land has a different snow depth representation from in-situ snow observations, introducing biases into the model. Grid cells adjacent to glaciers, for example, often show exaggerated snow depths, whereas those near large water bodies display reduced values. We suspect that the reason for this may be interpolation artefacts in the production of ERA5 Land as a higher-resolution version of the coarser ERA5 data. Even after excluding glaciers and lake polygons at a resolution of 100 m to mitigate these biases, the influence from adjacent land cover types persists within the ERA5 Land data, affecting the representation of true snow conditions. These biases are spatially dependent and affect the dataset's predictive snow depth accuracy based on location. This highlights the necessity for high-density, local sampling strategies to refine model predictions, especially in heterogeneous terrain. Meanwhile, the model also faces a trade-off between sample size and localization. By expanding the training dataset into a bigger area, the model benefits from more samples but also inherently leans towards a more generalized, averaged representation of snow depth.

We have refined the wind-aspect factor ([Bennett et al., 2022](#); [Dvornikov et al., 2015](#)) into a new cumulative value modeling wind redistribution across topographic barriers by wind fields from ERA5 monthly. While this simple metric explains a considerable amount of the observed preferential snow deposition (Figure 9 g), erosion processes are not equally well captured (Figure 9 f). The spatial-temporal resolution of wind fields from ERA5 Land is likely insufficient for capturing complex wind-topography interactions. 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. Incorporating more informative features on physical processes is a promising idea for future work, because XGBoost does not require normalization of features, but also faces the lack of forcing data.

These challenges likely lead to the underestimation of snow depth variability observed in our time series analysis (Figure 7) and snow depth maps (Figures C.15 C.16) if a localized calibration is not applied. To address this, we proposed a calibration process. This step fits the model with selected control points to mitigate residual biases that distort the distribution of predicted snow depths. Although our selected control points cover the entire study area, ensuring a comprehensive correction across all regions analyzed, we acknowledge that such large-scale calibration may not suit all study areas. Localized control points could offer more precise adjustments, particularly in analyses confined to smaller regions.

### *5.3. Characterizing snow depth spatial variability*

Snow depth variability is influenced by distinct processes at various scales. Our model demonstrates high fidelity of snow distribution patterns even at the microscale (Figure 6 c), though with lower performance metrics compared to the mesoscale. Our findings complement the insights from Mott et al. (2018), which recommend a resolution finer than 50 meters to capture the spatial variability inherent in wind-driven processes. Trujillo et al. (2007); Mott et al. (2011); Mendoza et al. (2020a) report a distinct 'scale break' at around 100 m, with a stronger spatial autocorrelation of snow depths before the scale break than beyond. Such spatial correlation can be attributed to the wavelength of wind-driven snow accumulation features (Mott et al., 2018). The ALS data and our model output also identified the presence of scale breaks and dependence on terrain features (95 m accounting for 79% variance for ALS data and 108 m for 60% variance for our model), demonstrating the effectiveness of our scheme on capturing a high level of spatial details and the impact of landscape patterns. Notably, we found a scale break at 1,300 m by fitting the empirical variograms at two different ranges, which corresponds to the wavelength of ridges in this area (Figure 10 e). These scale breaks underscore the critical resolution and minimal survey scope required to accurately capture localized spatial variability. Such insights are pivotal for designing snow surveys or evaluating snow models/products.

### *5.4. Downscaling-calibration scheme and application*

Current limitations of ICESat-2 data, including the month-long data availability lag and significant spatial and temporal gaps, restrict its utility for operational tasks and other applications (Deschamps-Berger et al., 2023). The regression models that rely solely on topographic features to determine snow spatial distribution fall short in capturing temporal variability (Revuelto et al., 2020). Our workflow does not rely on regression for snow depth measurements but focuses on 'subgrid variability', which resolves many of these constraints by incorporating spatial variability from ICESat-2 and temporal variability from ERA5 Land. Additionally, ICESat-2 not only provides detailed spatial distribution of snow depth but also includes a snow flag. This flag is determined by the observations of ground brightness and is labelled by the NOAA daily snow mask. This contributes to more accurate results in determining the presence of snow cover (Figure 8). The integration of these elements makes the workflow hold the potential for a range of applications demanding snow distribution patterns. For example, our workflow can mitigate data gaps in remote areas, facilitating the planning and cost-effective execution of new snow surveys while minimizing outdoor operational risks. The workflow holds the potential to enhance snow models. If the snow model provides not just an average value, but also a probability density distribution that represents the subgrid details, this information can then be translated into a snow depth map using our snow distribution patterns. The snow depth parametrization/distribution at the mesoscale for large areas or snow depth maps at the microscale are crucial for local permafrost studies (Gisnås et al., 2016) or geomorphological processes driven by snowmelt. With careful calibration, accurate, cost-efficient parameterization for snow depth in complex terrain also leads to improvements in discharge modeling. Identifying potential snow accumulation zones is useful for avalanche modeling, helping to mitigate

the risks associated with snow avalanches in mountainous regions. The monthly snow timeseries also benefit ecosystem studies in snow-covered regions e.g. habitat availability, plant phenology.

The performance of aggregated downscaled snow depth is not sensitive to the absolute precision of reference DEMs [Supplement D](#). This finding supports the possible applicability of our workflow in treeless conditions by global DEMs. However, careful consideration is required when applying this workflow to new study areas or applications: (1) The high quality, snow-free, up-to-date DEM is basis while global DEM is not equal in quality across different regions. (2) using enough samples in/around the target area and avoiding the coverage gaps that exist in ICESat-2's measurements. Increasing ICESat-2 data will mitigate these gaps. (3) It is not appropriate to calibrate the entire area using a snow survey conducted at low elevations; instead, it is important to use the most representative control points in calibration. (4) The influence of regional climatic conditions, including the impact of specific weather events, must be accounted for. E.g. some snowfall are heavily influenced by specific weather events.

## 6. Conclusion

This study introduces a workflow for snow depth retrieval from ICESat-2 ATL08 and DEM data to downscale ERA5 Land snow depth data using an XGBoost tree-structure machine learning model. The two datasets have complementary resolutions in space and time that allow the generation of accurate monthly snow depth maps at the microscale (10 m) or mesoscale (100 m). Hereby, ERA5 Land data primarily provides the temporal variability and ICESat-2 the spatial variability in snow depths which are propagated in space using terrain features and other relevant predictors in the model. Advanced bias correction and calibration are part of the workflow in order to address inherently present systematic errors in the data and correct for scaling bias.

- (1) There are very few snow depth observations available in remote areas, and to date no inexpensive ways to map small-scale variability exist. Here, ICESat-2 ATL08 data as presented in our workflow of coregistration-bias-correction stands out as a valuable data source.
- (2) The downscaling-calibration scheme's performance to predict peak snow for  $6 \times 2$  ALS flight strips in the Hardangervidda is very good at mesoscale ( $R^2$  values ranging from 0.74 to 0.88). At microscale, the spatial snow depth pattern is captured very well but absolute values less so. The model is also able to represent the spatial pattern of snow melt-out during late spring as visible from snow cover satellite data.
- (3) We introduce a new cumulative wind-aspect factor in the downscaling model that estimate snow wind re-distribution from ERA5 Land monthly wind fields in a cumulative way. This factor has a high predictive strength for the spatial distribution of snow depth at micro- and mesoscale in the downscaling model.
- (4) The downscaling model is sensitive to systematic bias in the elevation data, like slope- and curvature dependent bias, but less sensitive to precision at individual measurements. Our workflow includes bias correction for systematic errors. Therefore, similar results can be obtained when using the global DEM Copernicus GLO30 compared to the Norwegian national DEM DTM1 at 1 m spatial resolution.
- (5) Post-calibration of downscaling model results is currently necessary to compensate for the model's under-representation of extreme values. This under-representation is likely caused by inherent model behaviour, remaining bias and the nature of the training data. These include persistent snow patches on supposedly snow-free reference DEMs, simplified representation of wind processes, undersampling and underestimation of high snow depths in ICESat-2 data, and temporal nonstationarity in ERA5. Future research and more adaptable calibration methods for varied scenarios may improve results and remove the need of post-calibration.

The result of this work is a scalable and explainable downscaling model that is fully transferrable and applicable in other areas. Our research contributes to a heuristic 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 forested areas, are imperative.

## Funding

This work was supported by the SNOWDEPTH project, 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, Methodology, Writing – Review & Editing, Resources, Supervision, Funding acquisition, Project Administration. **Simon Filhol:** Conceptualization, Methodology, Writing – Review.

## 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 method is open source at <https://github.com/GlacioHack/xdem>. The data processing script can be accessed at [https://github.com/liuh886/subgrid\\_snow](https://github.com/liuh886/subgrid_snow). The validation ALS dataset is from the Norwegian Water Resources and Energy Directorate (NVE) at DOI: 10.5281/zenodo.2572731. The rest of the datasets are publicly accessible (data sources in text).

## 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/Thomas Skaugen from the Norwegian Water Resources and Energy Directorate (NVE) 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).
- 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).
- Bormann, K.J., Brown, R.D., Derksen, C., Painter, T.H., 2018. Estimating snow-cover trends from space. *Nature Clim Change* 8, 924–928. doi:[10.1038/s41558-018-0318-3](https://doi.org/10.1038/s41558-018-0318-3).
- Bühler, Y., Adams, M.S., Bösch, R., Stoffel, A., 2016. Mapping snow depth in alpine terrain with unmanned aerial systems (uass): Potential and limitations. *The Cryosphere* 10, 1075–1088. doi:[10.5194/tc-10-1075-2016](https://doi.org/10.5194/tc-10-1075-2016).

- 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).
- Cannon, A.J., Sobie, S.R., Murdock, T.Q., 2015. Bias correction of gcm precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? *Journal of Climate* 28, 6938–6959. doi:[10.1175/JCLI-D-14-00754.1](https://doi.org/10.1175/JCLI-D-14-00754.1).
- 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).
- 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).
- Daudt, R.C., Wulf, H., Hafner, E.D., Bühler, Y., Schindler, K., Wegner, J.D., 2023. Snow depth estimation at country-scale with high spatial and temporal resolution. *ISPRS Journal of Photogrammetry and Remote Sensing* 197, 105–121. doi:[10.1016/j.isprsjprs.2023.01.017](https://doi.org/10.1016/j.isprsjprs.2023.01.017).
- 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).
- 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).
- 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).
- Gismås, 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).
- 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., Mott, R., Bühler, Y., Le Toumelin, L., Lehning, M., 2024. Snowfall deposition in mountainous terrain: A statistical downscaling scheme from high-resolution model data on simulated topographies. *Front. Earth Sci.* 11. doi:[10.3389/feart.2023.1308269](https://doi.org/10.3389/feart.2023.1308269).
- 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).
- Hoyer, S., Hamman, J., 2017. Xarray: N-d labeled arrays and datasets in python. *JORS* 5, 10. doi:[10.5334/jors.148](https://doi.org/10.5334/jors.148).
- Hu, W., Scholz, Y., Yeligeti, 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).
- Hugonet, 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., Vivioli, 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).
- Kartverket, 2022. Høgdedata og djupnedata. <https://www.kartverket.no/api-og-data/terrengdata>.
- Ketzler, G., Römer, W., Beylich, A.A., 2021. The climate of norway, in: Beylich, A.A. (Ed.), *Landscapes and Landforms of Norway*. Springer International Publishing, Cham. *World Geomorphological Landscapes*, pp. 7–29.
- 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., Pfugl, 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).
- 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., Sheffield, J., Wood, E.F., 2010. Bias correction of monthly precipitation and temperature fields from intergovernmental panel on climate change ar4 models using equidistant quantile matching. *J. Geophys. Res.* 115. doi:[10.1029/2009JD012882](https://doi.org/10.1029/2009JD012882).
- 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., Schöber, 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).
- Maraun, D., 2019. Statistical downscaling for climate science. *Oxford Research Encyclopedia of Climate Science* doi:[10.1093/acrefore/9780190228620.013.712](https://doi.org/10.1093/acrefore/9780190228620.013.712).
- 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., Musselman, K.N., Revuelto, J., Deems, J.S., López-Moreno, J.I., McPhee, J., 2020a. Interannual and seasonal variability of snow depth scaling behavior in a subalpine catchment. *Water Resources Research* 56. doi:[10.1029/2020WR027343](https://doi.org/10.1029/2020WR027343).
- Mendoza, P.A., Shaw, T.E., McPhee, J., Musselman, K.N., Revuelto, J., MacDonell, S., 2020b. 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).
- Moudry, 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., Zsoter, 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-Otin, 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).
- 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., Jelley, 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).
- Neuenschwander, A.L., Pitts, K.L., Jelley, B.P., Robbins, J., Klotz, B., Popescu, S.C., Nelson, R.F., Harding, D., Pederson, D., Sheridan, R., 2021. Atlas/icesat-2 l3a land and vegetation height, version 5. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center doi:[10.5067/ATLAS/ATL08.005](https://doi.org/10.5067/ATLAS/ATL08.005).

- 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.2013.09.006](https://doi.org/10.1016/j.catena.2013.09.006).
- 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).
- Porter, C., Howat, I., Noh, M.J., Husby, E., Khuvis, S., Danish, E., Tomko, K., Gardiner, J., Negrete, A., Yadav, B., Klassen, J., Kelleher, C., Cloutier, M., Bakker, J., Enos, J., Arnold, G., Bauer, G., Morin, P., 2022. Arcticdem - strips, version 4.1. doi:<https://dataVERSE.harvard.edu/citation?persistentId=doi:10.7910/dvn/c98dvs>.
- 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).
- Saloranta, T.M., 2016. Operational snow mapping with simplified data assimilation using the senorge snow model. *Journal of Hydrology* 538, 314–325. doi:[10.1016/j.jhydrol.2016.03.061](https://doi.org/10.1016/j.jhydrol.2016.03.061).
- Sentinel-2 (processed by ESA), C., 2021. Msi level-2a boa reflectance product. collection 1.
- Sturm, M., Liston, G.E., 2021. Revisiting the global seasonal snow classification: An updated dataset for earth system applications. *Journal of Hydrometeorology* 22, 2917–2938. doi:[10.1175/JHM-D-21-0070.1](https://doi.org/10.1175/JHM-D-21-0070.1).
- 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).
- pandas development team, T., 2024. Pandas-dev/pandas: Pandas. Zenodo.
- 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., Ramírez, 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).
- Tryhorn, L., DeGaetano, A., 2013. A methodology for statistically downscaling seasonal snow cover characteristics over the northeastern united states. *Intl Journal of Climatology* 33, 2728–2743. doi:[10.1002/joc.3626](https://doi.org/10.1002/joc.3626).
- Tsang, L., Durand, M., Derkzen, C., Barros, A.P., Kang, D.H., Lievens, H., Marshall, H.P., Zhu, J., Johnson, J., King, J., Lemmetyinen, J., Sandells, M., Rutter, N., Siqueira, P., Nolin, A., Osmanoglu, B., Vuyovich, C., Kim, E., Taylor, D., Merkouriadi, I., Brucker, L., Navari, M., Dumont, M., Kelly, R., Kim, R.S., Liao, T.H., Borah, F., 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).
- 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).

## Supplement A. Gradient descent co-registration

Gradient Descent Coregistration (GDC) works similarly to the well-known NuthKaab DEM coregistration algorithm (Nuth and Käab, 2011) but uses distinct techniques. Where NuthKaab solves the co-registration problem iteratively by minimizing NMAD statistically with the help of terrain parameters, GDC solves the same geo-referencing problem more quickly with comparably good results. GDC regards the mismatch between two datasets 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  $\theta$ , and  $\xi$  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 DEM co-registration, the function to be minimized is the NMAD of all elevation differences between the datasets ( $dh$ ). The parameter is the shift matrix  $(E, N)$  (Figure A.11).

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 that is also corrected. However, this vertical bias correction may not be correct if there is no confirmation that the DEM was produced under snow-free conditions — since a remaining snowpack can bias the actually present vertical shift between the datasets. This study did thus not carry out vertical bias correction during co-registration and left it for the following step of bias correction. Details on the co-registration quality and results for this dataset can be found in Liu (2023) .

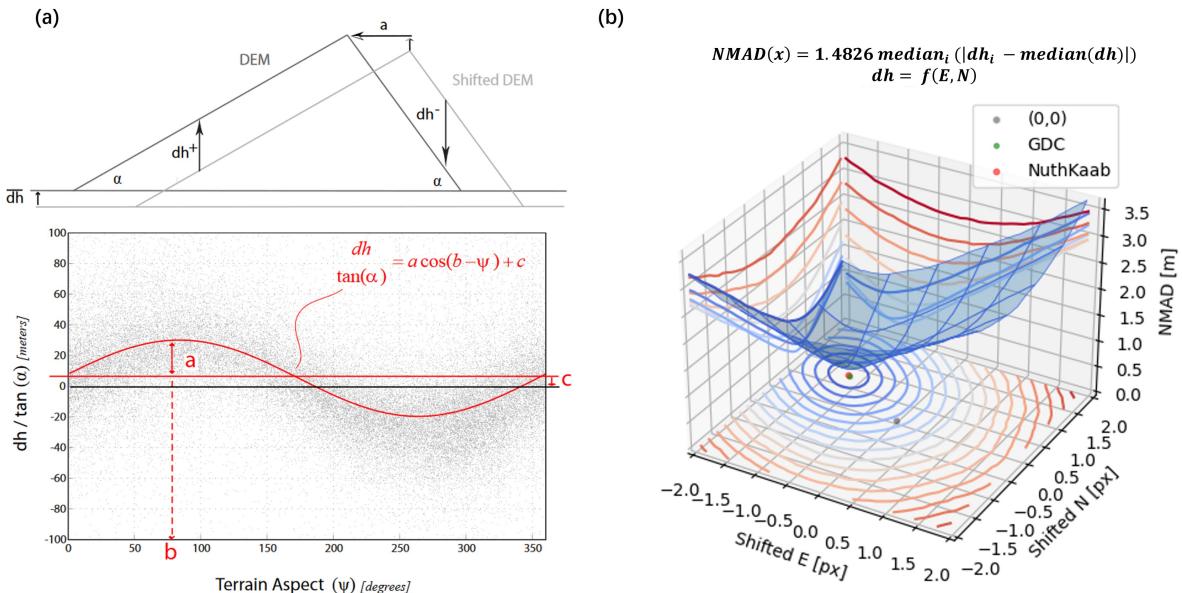


Figure A.11: NuthKaab Co-registration and Gradient Descent Co-registration. The elevation difference ( $dh$ ) induced by mismatch can be expressed in a 2-dimensional curve line (a). This curve suggests a statistical solution where the dispersion is related to aspect ( $\psi$ ), slope ( $\alpha$ ), shifted distance ( $a$ ), shifted direction ( $b$ ) and vertical bias ( $c$ ). By solving the equation, it gives a vector ( $a, b, c$ ) to shift the DEM to the optimal position over several iterations (Nuth and Käab (2011)). On the other hand, GDC presents  $dh$  in a 3-dimensional curved surface. It gets the shift matrix using a gradient descent algorithm, resulting in the local minimal of NMAD (b)

## Supplement B. Bias correction

We employ an XGBoost regression model to estimate the discrepancy for four DEMs using various features such as terrain parameters, vegetation, and quality metrics from ICESat-2 ATL08. Approximately 5 million snow-free segments were utilized as the training dataset. We used squared error as the loss function with hyperparameters suggested by grid search, including *n\_estimators* of 350, *max\_depth* of 10, *learning\_rate* of 0.1, and *subsample* of 0.7. To better inform the model of the characteristics of the discrepancy, we used the following features:

- The *dh\_te\_subsegment* metric, calculated as *h\_te\_best\_fit\_20m\_2* minus *h\_te\_best\_fit*, describes the elevation difference between the midpoint of a segment (which might be an interpolated value when the measurement is missing) and its subsegments.
- Terrain parameters: *slope*, *tpi*, *tpi\_3*, *tpi\_9*, plan curvature (*planc*), *curvature*, profile curvature (*prof*), *aspect*. And coordinates East (E) and North (N).
- Terrain attributes from ICESat-2: *h\_te\_best\_fit*, *h\_te\_std*, *h\_te\_skew*, slope along the track (*terrain\_slope*).
- Other attributes from ICESat-2: *night\_flag* and *cloud\_flag\_atm* describe the observing condition. The *beam* indicates the strong or weak beam. The *region* is about satellite orbit ascending or descending. And, the *subset\_te\_flag* is a quality metric, that has been simplified as an integral number representing the available subsegments.
- The residual vertical bias from co-registration, e.g. *coreg\_bias\_dtm10*.

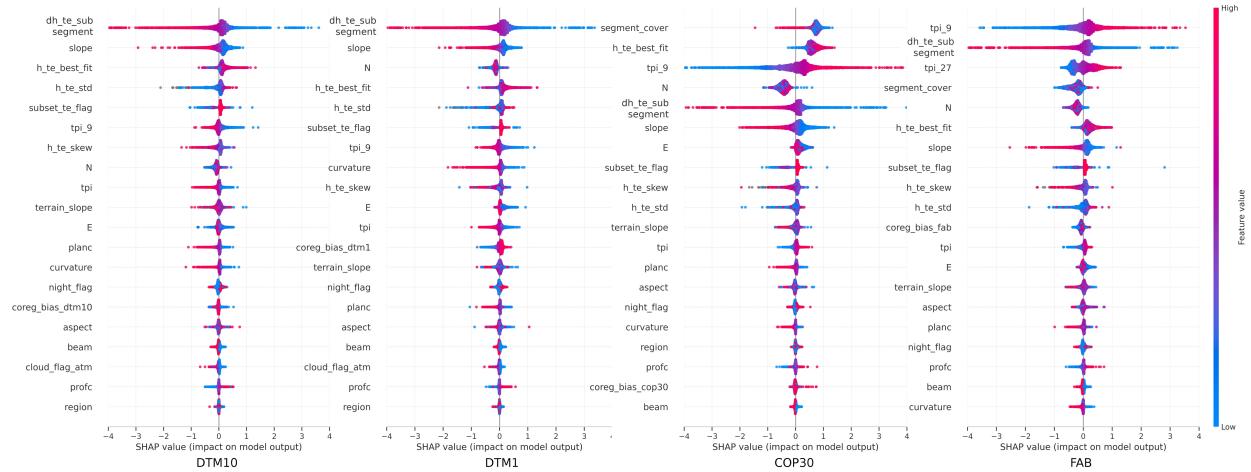


Figure B.12: Feature importance of bias correction model. The features, sorted in descending order of importance, explain the contribution of features to the discrepancy. Each dot represents a sample from the validation area

The feature importance differs for each of the DEMs (Figure B.12), indicating that different biasing effects are at play. The SHAP value is proportional to the elevation difference, which is calculated as ICESat-2 snow-free measurements minus DEM's elevation. Specifically, a negative SHAP value represents the phenomenon of ICESat-2's underestimation relative to DEM. For instance, when the subset geosegment's elevation (*h\_te\_best\_fit\_20m\_2*) exceeds *h\_te\_best\_fit*, it yields a high value (indicated in red) of *dh\_te\_subsegment*, and it still contributes more a negative biasing phenomenon. This indicates that using *h\_te\_best\_fit\_20m\_2* aligns more closely with reference DEM's surface elevation than using *h\_te\_best\_fit*. This justifies this study of using the subsegment instead of *h\_te\_best\_fit* or *h\_te\_mean*. Furthermore, the SHAP importance also shows that not just the slope but also the second-order DEM derivatives (curvature and TPI) rank in the very front. For DTM1 and DTM10, a negative bias resulted by convex terrain (positive *tpi\_9*). Conversely, for COP30 and FAB, a negative bias is observed in concave terrain (negative *tpi\_9*). This is not the result of the interpolation effect of sloppiness but of curvature at different resolutions indicating

that ICESat-2 has a footprint size better than COP30 / FAB and less than DTM1 / DTM10, instead of 100 m (the segment length).

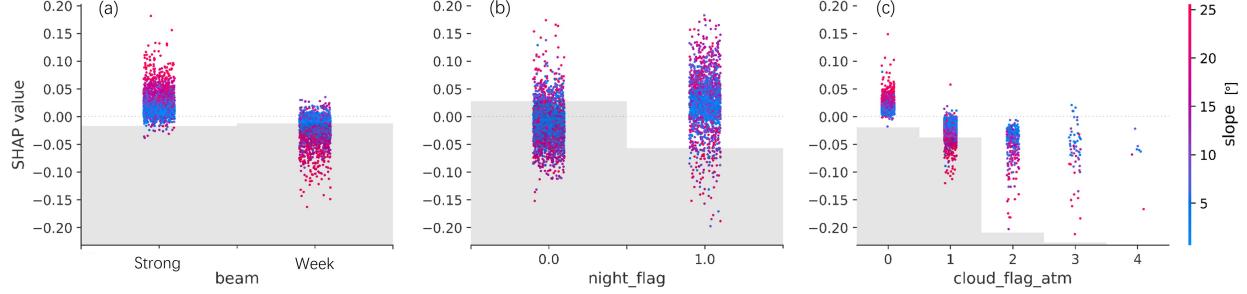


Figure B.13: Interpreting ICESat-2 related bias. Beam intensity (a) night flag, with 0 representing day, and 1 representing night (b). Cloud, 0 means no scattering, 1 means clouds at  $> 3\text{km}$ , 2 means clouds at  $1\text{-}3 \text{ km}$ , and 3 = clouds at  $< 1 \text{ km}$ , 4 = blowing snow (c)

As a comparison, ICESat-2 metrics like beam intensity, cloud existence (*cloud\_flag\_atm*) or illumination conditions (*night\_flag*) exhibit much less biasing effect. The degree of their influence is comparable to that of aspect, as shown in Figure B.12 and B.13.

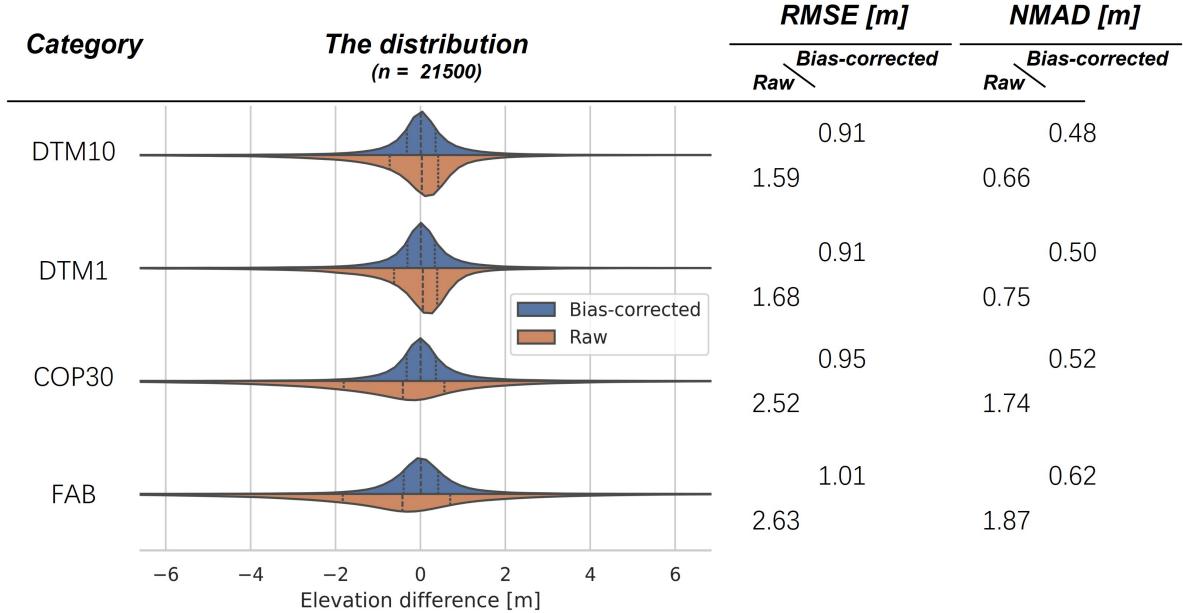


Figure B.14: The distribution before and after applying co-registration and bias correction for a given dataset in the validation area. The Q25, Q50 and Q75 are plotted in vertical lines inside the violin plot.

Figure B.14 shows that, after bias correction, the DTM1 has a reduced NMAD of 0.50 m and COP30 has an NMAD of 0.52 m relative to ICESat-2 snow-free reference in the Hardangervidda area, which is much smaller than the raw DTM1 and COP30 has NMADs of 0.75 m and 1.74 m respectively. Furthermore, all DEMs display a reduction in negative skewness, with COP30 and FAB showing particularly notable improvements. The uncertainty analysis by Liu (2021) on ICESat-2's ATL08 product obtained an RMSE of 0.98 m in Alaska based on airborne lidar elevation ground truth, providing a comparative benchmark metric to the variation observed in this study. We expect the snow depth retrieval that is based on the same bias correction and elevation differencing workflow to have a similar performance.

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, or parts of it, is produced under snow-on conditions (which is common for areas at higher elevations), this correction may not be able to fully remove the snow patch on DEM, as all parameters are calculated from the DEM. And our workflow assumes a consistent bias pattern in ICESat-2 elevation measurements, whether snow is present or not, which is also the assumption of the established method of measuring glacier elevation changes or snow depth from crossing ground tracks. 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 relative canopy height caused by snow cover.

## Supplement C. Pre-calibration

Figure C.15 and C.16 shows the downscaled output at mesoscale and microscale levels before calibration. These figures reveal that the highest snow depth values on the western mountain tops are not well represented, indicating that the DTM1 may still have snow cover. In Figure C.15 e, the difference between DTM1 and ALS DEM acquired in Sep. 2008 is shown (dark blue areas predominantly on eastern/northern slopes), highlighting remaining snow patches on DTM1. This partially explains the underestimation of downscaled snow depth and emphasizes the need for calibration.

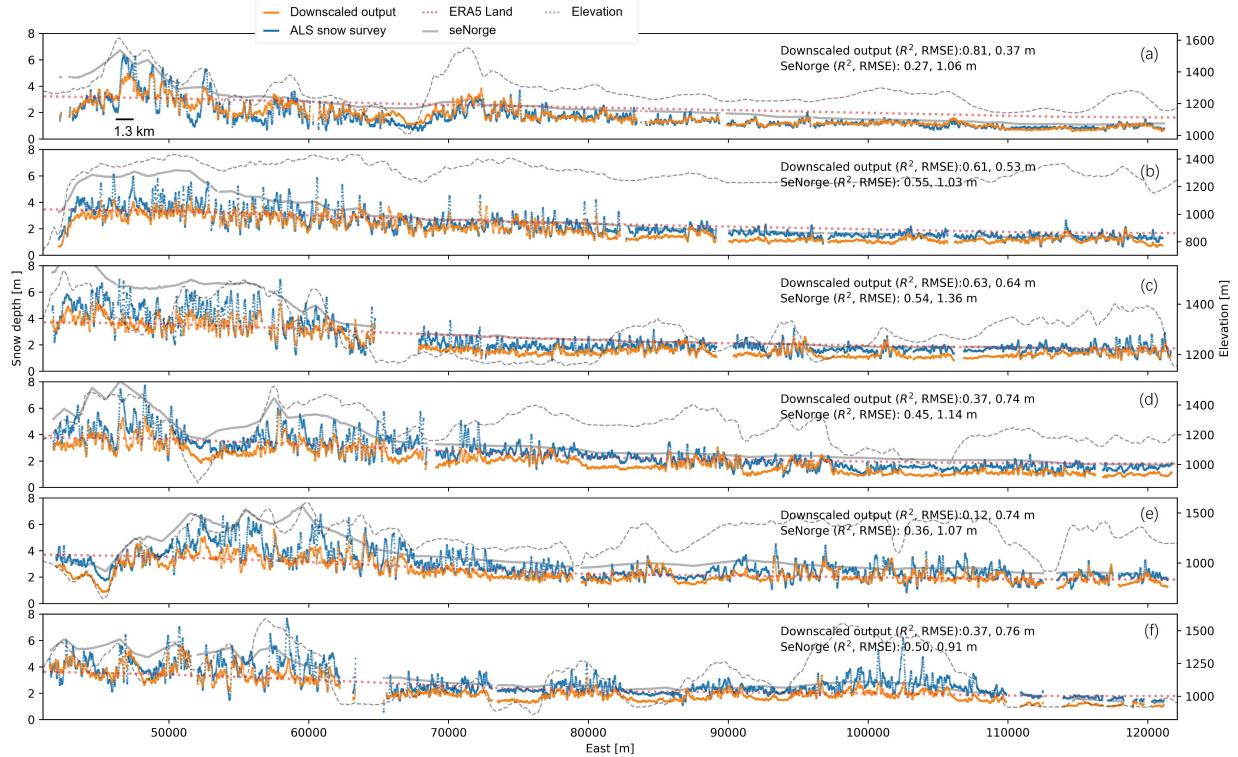


Figure C.15: Snow depth profiles across Hardangervidda in April 2008 (Pre-calibration) at mesoscale. Aggregated downscaled snow depth and ALS flight profiles for April 2008 are shown (each dot has bin size of  $100 \times 500$  m).

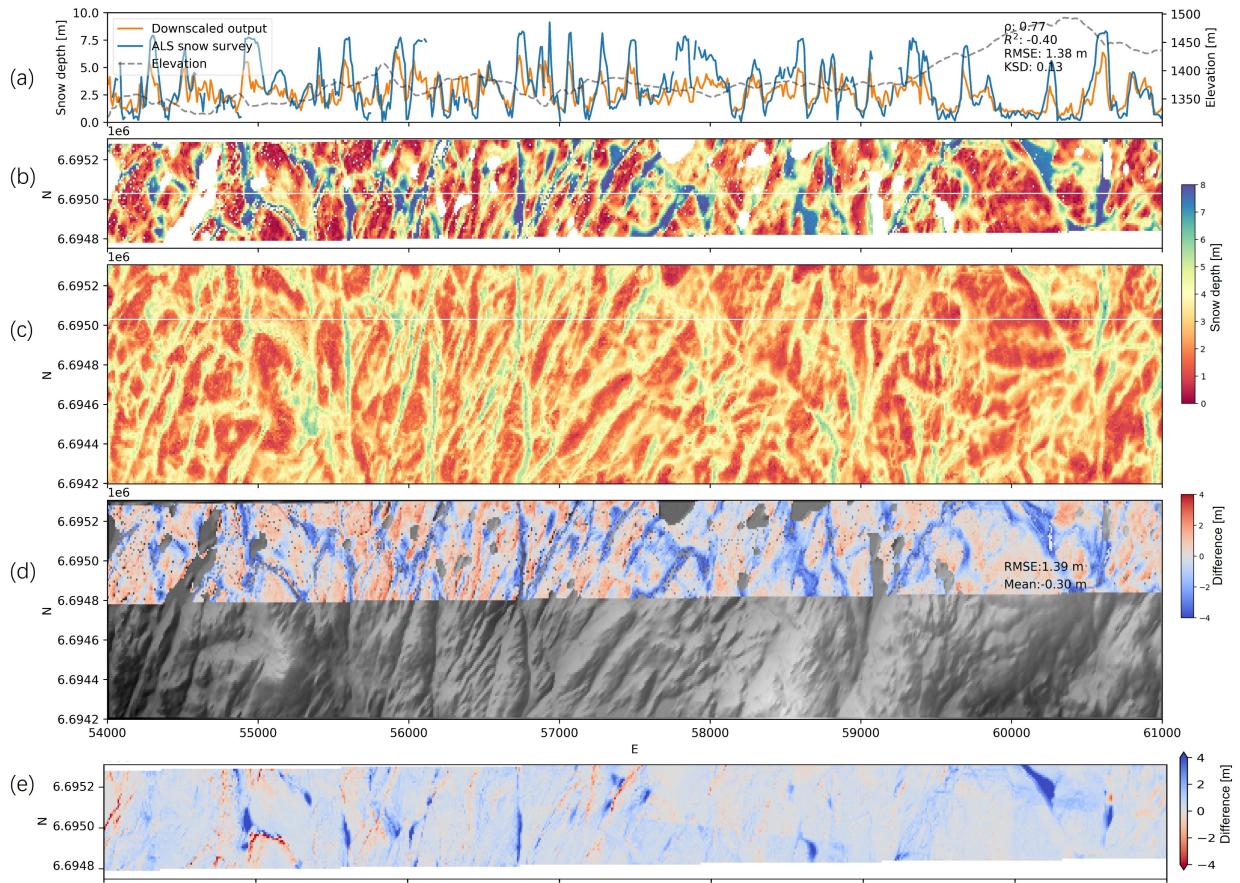


Figure C.16: Microscale Snow Depth Comparison in Hardangervidda, April 2008 (Pre-Calibration) at 10 m spatial resolution. Panel (b) shows the ALS snow survey validation strip, revealing significant snow depth variations. Panel (c) illustrates the downscaled snow depth from the DTM1 model, capturing most of the observed relative variability but with a low  $R^2$  value as shown for the 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.

## Supplement D. The comparison between DEMs

Snow depth maps produced from the other tested DEMs are similar with no visually identifiable systematic biases (not shown). Global DEMs perform similarly to those obtained using the Norwegian high-resolution datasets DTM1 and DTM10 at the mesoscale when compared to the ALS data on Hardangervidda (Figure D.17). DTM1 with its higher resolution slightly outperforms the other DEM datasets with slightly higher  $R^2$  and  $\rho$  values, and lower RMSE and KSD values. On average, the vegetation-corrected FAB DEM yields slightly better results than the COP30 DEM in the mostly treeless Hardangervidda area.

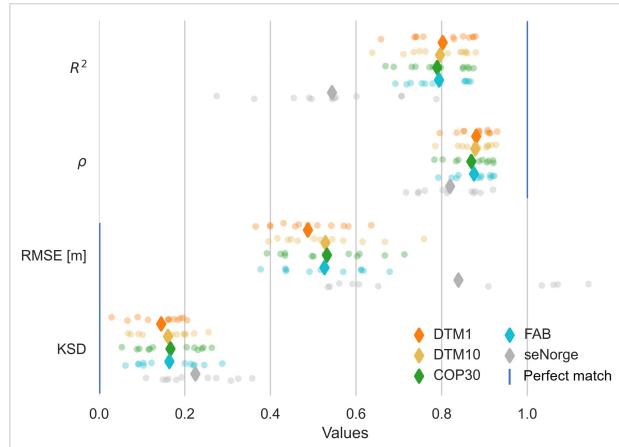


Figure D.17: Statistical performance of the downscaling model using different DEMs at the mesoscale (100 m x 500 m). Dots/stars symbolize the performance compared to ALS snow survey strips from 2008 and 2009, respectively, diamonds indicate average performance across all 12 strips.

## Supplement E. The snow depth 2009

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

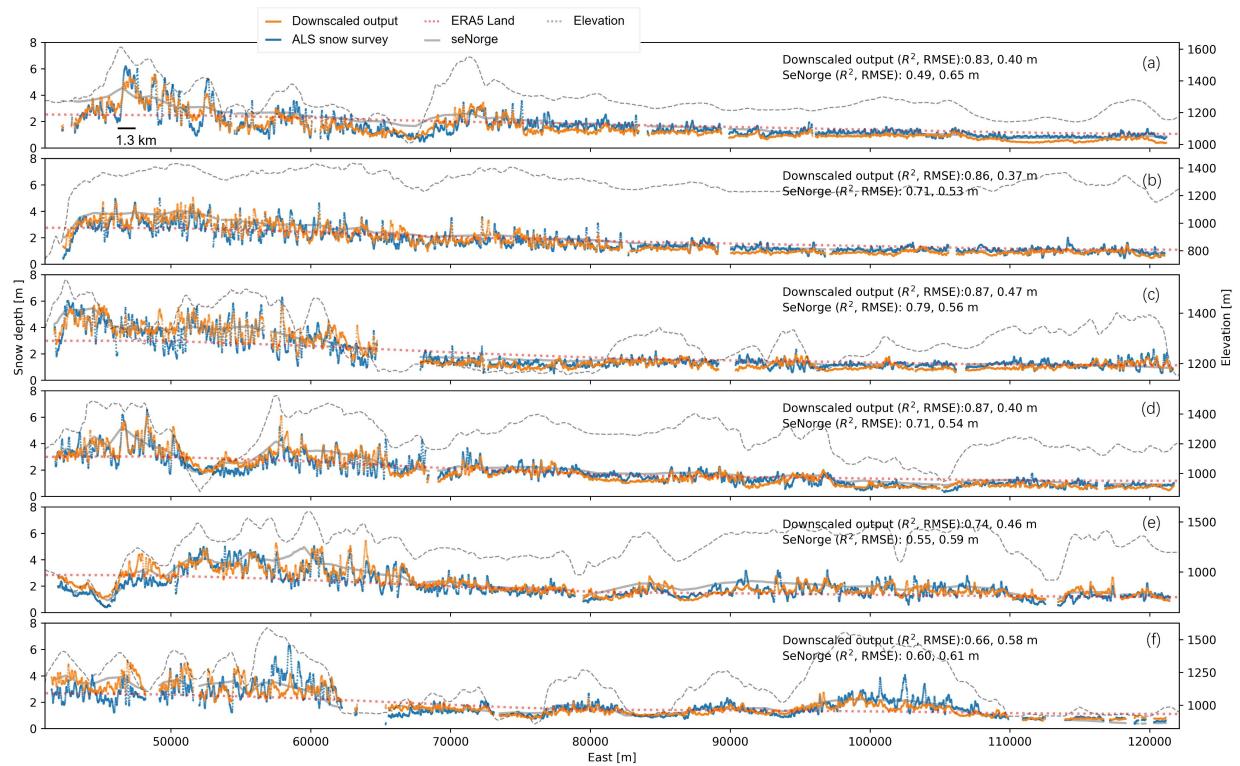


Figure E.18: Snow depth profiles across Hardangervidda in April 2009. This figure showcases a mesoscale analysis of snow depth along six ALS flight lines, spanning north to south (panels a-f). Data points, aggregated to represent 100 m x 500 m cells (because of the ALS survey's 500 m transect width).

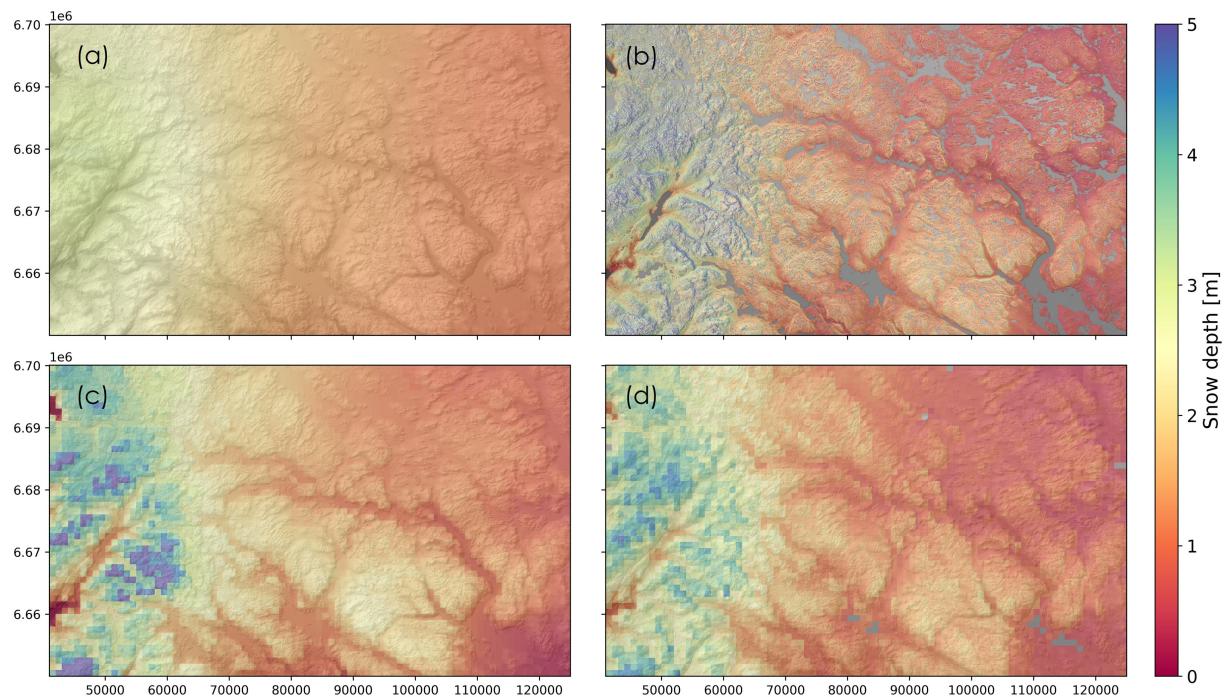


Figure E.19: Spatial Distribution of Snow Depth in Hardangervidda, April 2009. (a) Snow depth from ERA5 Land in 10 m resolution (after linear interpolation). (b) Downscaled snow depth output in 10 m resolution. (c) seNorge snow depth at 1 km resolution. (d) Downscaled snow depth aggregated at 1 km resolution.