Spatially distributed snow depth, bulk density, and snow water equivalent from ground-based and airborne sensor integration at Grand Mesa, Colorado, USA

Tate G. Meehan[1,2], Ahmad Hojatimalekshah[2], Hans-Peter Marshall[2], Elias. J. Deeb[1], Shad O'Neel[1,2], Daniel McGrath[3], Ryan W. Webb[4], Randall Bonnell[3], Mark S. Raleigh[5], Christopher Hiemstra[6], Kelly Elder[7]

[1] Cold Regions Research and Engineering Laboratory; U.S. Army Corps of Engineers; Hanover, NH, USA
[2] Department of Geosciences; Boise State University; Boise, ID, USA
[3] Department of Geosciences; Colorado State University; Fort Collins, CO, USA
[4] Department of Civil and Architectural Engineering & Construction Management; University of Wyoming; Laramie, WY, USA
[5] College of Earth, Ocean, and Atmospheric Sciences; Oregon State University; Corvallis, OR, USA
[6] Geospatial Management Office; USDA Forest Service; Salt Lake City, UT, USA
[7] Rocky Mountain Research Station; USDA Forest Service; Fort Collins, CO, USA

Correspondence to: Tate G. Meehan (tate.g.meehan@erdc.dren.mil)

Abstract. Estimating snow mass in the mountains remains a major challenge for remote sensing methods. Airborne LiDAR can retrieve snow depth, and some promising results have recently been shown from spaceborne platforms, yet density estimates are required to convert snow depth to snow water equivalent (SWE). However, the retrieval of snow bulk density remains unsolved, and limited data is available to evaluate model estimates of density in mountainous terrain. Toward the goal of landscape-scale retrievals of snow density, we estimated bulk density and length-scale variability by combining ground-penetrating radar (GPR) two-way travel-time observations and airborne LiDAR snow depths collected during the mid-winter NASA SnowEx 2020 campaign at Grand Mesa, Colorado, USA. Key advancements of our approach include an automated layer picking method that leverages GPR reflection coherence and distributed LiDAR–GPR retrieved bulk density with machine learning. The root-mean-square error between the distributed estimates and in situ observations is 11 cm for depth, 27 kg/m^3 for density, and 46 mm for SWE. The median relative uncertainty in distributed SWE is 13 %. Interactions between wind, terrain, and vegetation display corroborated controls on bulk density that show model and observation agreement. Knowledge of the spatial patterns and predictors of density is critical for accurate assessment of SWE and essential snow research applications. Spatially continuous snow density and SWE estimated over approximately 16 km^2 may serve as necessary calibration and validation for stepping prospective remote sensing techniques toward broad-scale SWE retrieval.

1 Introduction

Throughout the past half-century, snowpacks in the western U.S. declined ~ 20 % because of ongoing warming (Pierce et al., 2008; Mote et al., 2018). By the end of the 21st century, projections suggest snow water equivalent (SWE) in this region will decline by an additional ~ 50 % (Siirila-Woodburn et al., 2021). Decreased snow water supply and increased demand motivates new innovations for SWE measurement and modelling (e.g., Lettenmaier et al., 2015). Ground observations of SWE, such as those from snow telemetry (SNOTEL) sites or manual measurements performed during snow surveys provide useful information in the context of a historical record. However, as a strategy for adapting to changing snow-climatological

conditions, building the relationship between these observations and snow distribution patterns across watersheds requires
innovative spatiotemporal datasets and snow hydrology models to advance.

In this pursuit, NASA's snow experiment (SnowEx; 2017-23) campaign tested a suite of remote sensing instruments with potential to measure global SWE if deployed on a future satellite platform (Marshall et al., 2019). The work presented here was part of SnowEx and was designed to expand the spatial scale over which snow depth and density can be observed and reliably extrapolated. Our work provides a validation dataset for SnowEx SWE retrieval methods and yields new insights to spatial patterns and driving factors of snow density at Grand Mesa, Colorado.

Spaceborne snow depth estimates have been achieved from passive microwave sensors (Tedesco et al., 2010), Sentinel-1 radar returns (Lievens et al., 2019, 2022), high-resolution satellite stereo imagery (Marti et al., 2016; McGrath et al., 2019), and light detection and ranging (LiDAR) with ICESat (Treichler and Kääb, 2017) and ICESat-2 (Hu et al., 2021; Deschamps-Berger et al., 2023; Besso et al., 2024). LiDAR and photogrammetry techniques can measure snow depth, by differencing repeated acquisitions during periods with and without snow cover (e.g., Deems et al., 2013). Because of the advantages of greater spatial resolution and flexible scheduling to target acquisitions during periods of interest, airborne LiDAR constitutes a prominent method for estimating snow depth and is being flown operationally for integration with hydrologic modelling at the catchment scale (Painter et al., 2016; Hedrick et al., 2018). Regardless of the choice in snow depth retrieval, an estimate of snow density is required to convert snow depths to SWE, and bulk density often provides the greatest source of uncertainty in SWE estimates, especially in deeper snow (Raleigh & Small, 2017).

Excavating and weighing snow samples of a known volume remains the state-of-the-art approach for measuring snow density, even though the labour-intensive work limits the number of possible observations. Because snow depth varies more in space than density (e.g., Elder et al., 1991; Sturm et al., 2010; López-Moreno et al., 2013) and depth measurements may be collected more rapidly, density is observed far less frequently (e.g., Rovansek et al., 1993; Elder et al., 1998). As a result, snow sampling strategies tend to be too coarse to examine the $10^0 – 10^3$ m scale spatial variability of snow density (e.g., Fassnacht et al., 2010), and the spatial nature of snow density remains largely unknown.

Often, empirical models are used to spatially distribute density in SWE estimates. Linear regression models developed using snow depth alone are often unsuccessful, because the snow load only has a linear effect on bulk density while snow type characteristics (e.g., faceted crystals versus rounded-grain snow) can have an exponential effect (Sturm & Holmgren, 1998). Successful regression models parameterized by snow depth have been split up into elevation and month of year classes (Jonas et al., 2009), accumulation and melt seasons (Hill et al., 2019), or day of year and snow cover classification (Sturm et al., 2010) and account for the effects of snow depth and snow age on density (McCreight & Small, 2014). Snow density often depends on environmental (i.e., slope, aspect, elevation, and vegetation) and climatological (i.e., precipitation, solar radiation, temperature, and wind) factors (Meløysund et al., 2007), which makes these constituents candidates as features for predicting distribution patterns (e.g., Winstral et al., 2002). Machine learning (ML) approaches utilising environmental or climatological features (e.g., Elder et al., 1998; Wetlaufer et al., 2016; Broxton et al., 2019) are often distributed over vast areas with little validation, or consideration to the underlying physical processes, required to gain an acceptable level of model confidence.

Snow density can also be distributed with process-based snow models, which may account for changes in bulk snow density due to new snowfall, metamorphism, and compaction. The representations of snow densification range in complexity, with some models utilising time-dependent compaction curves and other models representing snow compaction dynamically as a function of snow viscosity and overburden pressure (Essery et al., 2013). Dynamic models offer more consistent and accurate characterizations of snowpacks, however even for a single physics-based model, performance in snow density simulations varies across snow climates and watersheds (e.g., Marks et al., 1992; Lv & Pomeroy, 2020). The choice of snow density model (empirical or physical) produces differences in spatial distributions and basin mean estimates of snow density (Raleigh and Small, 2017).

Despite numerous techniques for modelling snow density, few studies characterise spatial variations in snow density and the underlying processes driving variability, largely due to limited density datasets. The labour-intensive nature of in situ observations severely limits spatial analyses, requiring the development of broad scale snow density retrieval. Relationships between snow density, dielectric permittivity, and radar signals (e.g., Tiuri et al., 1984), provide radar-retrieved snow density. Yet, many radar remote sensing retrievals require constraints on the snow depth, density, stratigraphy, and microstructure to be presently reliable (Tsang et al., 2022). Our research utilizes ground-penetrating radar (GPR), LiDAR, and ML to define an approach to map snow density at resolutions appropriate for air- and space-borne remote sensing calibration and validation. Ground-penetrating radar records the amplitude and travel-time of a series of echoes from short-pulse electromagnetic waves as an image in range-time and position coordinates. Provided a constraint on the snow depth, GPR analysis can estimate the snow density, or by exploiting a ray path function of travel-time versus antenna separation (offset) the snow depth and density can be estimated independently (e.g., Griessinger et al., 2018; Meehan et al., 2021). By combining snow depths from drone-based aerial photogrammetry or LiDAR with GPR travel-times, snow density has been estimated along 100 m scale transects, then analysed as a time series to understand the densification process (McGrath et al., 2022; Valence et al., 2022; Bonnell et al., 2023) and explore extrapolation across the study-plot scale (Yildiz et al., 2021).

Our work leverages airborne LiDAR snow depths in process with GPR two-way travel-times (TWT) to facilitate density estimates. These data then become input to multiple ML regression approaches to develop and compare spatially continuous estimates of bulk snow density and SWE across the entire LiDAR domain. Sensitivity testing among regression models informed model repeatability and forcing processes for spatial density patterns at Grand Mesa, Colorado, USA. As part of this workflow, we developed a reliable, automated, radar coherence approach for automatically interpreting TWTs needed to retrieve snow density. This work highlights interactions between snow, terrain, vegetation, and wind in the densification process, as well as the importance of careful ML model parameterizations and validation approaches. Our work addresses the need for high accuracy, distributed density measurements as assimilation data for parameterizations of snow densification, thereby reducing runoff model uncertainty. Additional knowledge of the spatial patterns and predictors of density may improve calibration, validation, and parameterization of radar remote sensing SWE retrievals.

2 Methods

2.1 Study area

Grand Mesa, Colorado, is a high-elevation subalpine plateau with an average elevation of ~3,200 m and an area of ~1,300 km^2. Grand Mesa has a cold and dry continental snow climate, low relief, and varying vegetation cover from shrub steppe and subalpine meadow to dense conifer forest. These factors, along with the proximity to a regional airport make Grand Mesa a near-ideal study area for evaluating airborne snow remote-sensing techniques and developing many challenging snow remote-sensing advancements (e.g., Boyd et al., 2022; Singh et al., 2023).

The Grand Mesa NASA SnowEx Intensive Observation Period (IOP) spanned 27 January – 12 February 2020. During that time, more than 150 snow pits were excavated and nearly 38,000 in situ snow depth measurements were collected. Snow pits were distributed within forested and open areas along the swaths of the airborne remote-sensing campaign flight lines (Fig. 1). As part of the SnowEx campaigns at Grand Mesa, five meteorological stations were installed between 2016 and 2017 and operated through the 2021 water year (Houser et al., 2022). Of these sites, 3 m elevation wind speed and direction data measured at Mesa Middle (MM) and Mesa West (MW) were examined as validation for snow transport potential and to quantify differences in exposed and sheltered terrain (Appendix C.1). The MW station was in exposed western terrain of the mesa, 350 m west of the study domain boundary (Fig. 1). The MM station is sheltered within a dense stand of conifer trees 18.7 km east of the study domain boundary.

2.2 GPR data acquisition

Two GPR instruments were operated during the first week of the Grand Mesa IOP. To acquire data within forested areas of central Grand Mesa, a conventional L-band GPR was pulled by ski during 30 January – 1 February and 5 February (Webb, 2021). This unit was equipped with a Global Positioning Satellite (GPS) receiver with 2.5 m horizontal accuracy. In open areas, we deployed a multi-polarization L-band GPR fastened within a sled that was pulled by a snowmobile at approximately 3 m/s in the open areas of the central and south regions of western Grand Mesa on 28 and 29 January, and 4 February 2020 (Meehan, 2021). The snowmobile was driven along the edges of forested stands but could not travel through densely treed areas. The multichannel L-band GPR was configured with one transmitting antenna and two receiving antennas that were oriented parallel (H) and orthogonal (V) to the transmitter (H). The transmit and receive antennas were separated by 25 cm. Using this GPR configuration we simultaneously acquired the radar imagery in co- and cross-polarizations (HH & HV). A Global Navigation Satellite Systems (GNSS) receiver with approximately 1 m horizontal position uncertainty, was located on the snowmobile 5 m away from the GPR array. We applied a geometric correction to relocate the coordinate positions to the antenna midpoint of each channel. The GPR data were acquired within a few metres of, but not directly beside the snow pit walls, which necessitated a radius for pairing retrieved or modelled data with validation observations. The GPR systems were operated continuously, collecting approximately 30 traces per second, given the duration of the time window for each trace (30 ns), the sample interval (0.1 ns), and the number of stacks acquired (2). Due to differences in the travel speed, the spatial

interval of the GPR traces collected via snowmobile is approximately 10 ± 1 cm, while the interval for traces collected by ski is 5 ± 1 cm. We used piecewise cubic Hermite interpolating polynomials (Kahaner et al., 1989) to fix a geolocation to every acquired trace, as the GPS acquisition rate was 1 Hz. Throughout this week, we acquired 144 km of quasi-gridded and spiralled snowmobile-driven radar transects, and 16 km of skied spiral transects in the forest. Spiral transects were coincident with depth measurements. We used a 4.5 km by 3.5 km portion of the snow-on LiDAR acquisition to bound the GPR transects (Fig. 1) and omitted any transects acquired beyond the LiDAR boundary.

2.3 GPR data processing

Multi-polarization radargrams were processed using the following automated routine. We applied a frequency-wavenumber (F-K) filter as a 2D band-pass filter (Kim et al., 2007). Time-zero correction was performed automatically using the Modified Energy Ratio first break picker (Wong et al., 2009). We removed coherent noise by subtracting the median trace from the radargrams (Kim et al., 2007). The trace amplitudes were corrected for spherical divergence by applying t-squared scaling as a signal gain function (Yilmaz, 2001). In a step of random noise removal, we then applied edge preserving smoothing (Kuwahara et al., 1976). This routine emphasised the continuity and amplitude of the ground reflection, which benefitted the method for automatically picking the travel-times. The GPR data within forests were processed with a bandpass filter, time-zero correction, and background subtraction prior manual interpretation using a semiautomatic algorithm and are available through the National Snow and Ice Data Center (Webb, 2021). The slower paced data acquisition by ski improves the quality of the radargram, which benefits the tracking of the ground surface in the more variable forest environment.

2.3.1 Multi-polarization coherence for automatic two-way travel-time determination

The rough ground depolarized the L-Band radar signal and thus we used the coherence between the co- and cross-polarized channels as a filter that illuminates the ground reflections and removes the planar reflections of the snow stratigraphy. We paired the co- and cross-polarization radargrams into shot gathers, which are the bins of traces that share the same transmitter location. The automatic travel-time pick is determined by maximising the coherence between the co- and cross-polarization shot gathers. For each pair of traces, we applied the unnormalized cross-correlation sum to measure the coherence,

$$C(t) = \frac{1}{2} \sum_{j=t-N/2}^{t+N/2} \left\{ \left[\sum_{i=1}^{M} S_{i,j} \right]^2 - \sum_{i=1}^{M} S_{i,j}^2 \right\}, \qquad (1)$$

which is half of the summed difference between the energy of the stacked traces and the energy of the input traces (Neidell & Taner, 1971). The calculation in Equation 1 is performed in a sliding window over $N = 11$ samples that is evaluated at every sample (j) of the GPR signal ($S_{i,j}$) for channels i ($M = 2$). The HH-HV coherence (C_{HH-HV}) at each shot location is then normalised by the maximum coherence,

$$C_{HH-HV} = \frac{C(t)}{\max C(t)}. \qquad (2)$$

Small (one-wavelength) offsets introduce waves that have approximately normal incidence to the reflection horizons, such that nonlinear effects of travel-time moveout are negligible and snow depth can be directly retrieved from the measurement of

TWT. Because the offsets are equal, the travel-times to the ground for each channel are equal within a small error (due to variability of the ground surface inside the radar footprint), and therefore the two channels sum coherently.

165 We automatically chose the travel-time with the maximum coherence of each trace and subtracted 1 ns (1/2 wavelet) to estimate the first break of the reflection (Booth et al., 2010). We then applied a median filter to remove outliers and reviewed the automatic picks for any systematic errors.

2.4 Observed, Derived, and Explanatory Data

2.4.1 In situ measurements

170 Snow pit observations and manual depth probe measurements were collected throughout the 27 January – 12 February 2020 IOP to serve as validation for SWE and snow depth retrieved by airborne remote sensing. Snow pits were measured for the snow depth, density, water equivalent, temperature, wetness, liquid water content, grain size, and stratigraphy (Vuyovich et al., 2021). Snow density ($\rho_{s,pit}$) was measured continuously every 10 cm from the snow surface to the ground using a 1000 cm^3 wedge sampler, with duplicate samples. If the difference between the two measurements at a given depth exceeded 10 %, the

175 density was sampled a third time, and bulk density was then calculated by averaging all measurements for each snow pit. Because the density snapshot we retrieved is valid for the time of the LiDAR flight, we corrected measured density to 12:00 PM on 1 February using densification rates determined by linear regression for both open and forested areas. Liquid water content was estimated by combining the density and in situ measurements of dielectric permittivity in an empirical formula, which showed that the snowpack remained almost completely dry throughout the IOP (Webb et al., 2021). Snow depth

180 measurements ($h_{s,probe}$) were collected using geolocated probes (± 3 m spatial accuracy) along spiral transects (~60 m radius) centred around pits (Hiemstra et al., 2020).

2.4.2 LiDAR snow depth

Snow depth ($H_{s,LiDAR}$) was estimated from repeated airborne LiDAR point cloud surface elevations of snow-free and snow-covered terrain (e.g., Lague et al., 2013). The Airborne Snow Observatory (ASO) performed the snow-free acquisition on 26

185 September 2016 (Painter et al., 2016; Painter & Bormann, 2020), and NV5 Geospatial (formerly Quantum Spatial Inc.) acquired snow-covered surface elevations during the IOP, both with a point density of approximately 20 points/m^2. We selected the 1-2 February 2020 flight to minimise temporal differences with the GPR and resulting errors due to snow redistribution and densification. We transformed the 2016 snow-free vertical datum into NAVD88/Geoid 12B (the same as 2020 snow-on) using NOAA VDatum 4.3 software (NOAA, 2021). Then, we applied the point cloud differencing method to estimate snow

190 depth on a 1 m grid (Appendix B.1). Negative snow depth values were filtered as no data values. After computing the snow depth, the 3 m ASO bare-earth and vegetation data products were resampled using the nearest-neighbour approximation to the 1 m resolution of the snow-covered SnowEx 2020 LiDAR acquisitions and the coordinate system was transformed from UTM

zone 13 N to UTM zone 12 N. As a comparison between our LiDAR snow depths and data processed using raster differencing, we used the 1-2 February 2020 ASO acquired snow depths and upscaled $H_{s,LiDAR}$ to 3 m using the nearest-neighbour method.

2.4.3 LiDAR – GPR estimated density

We combined the LiDAR snow depths with the GPR TWTs to calculate the radar wave velocity, which in dry snow is only a function of density. We applied a k-d tree searcher (Bentley, 1975) to find the LiDAR coordinates within a 1 m radius of the GPR TWTs. We then used the median values of the TWTs within a 1 m radius of these coordinates to interpolate to the LiDAR grid.

The average electromagnetic wave speed in the snowpack was estimated using

$$v_{s,LiDAR-GPR} = 2\frac{H_{s,LiDAR}}{\tau} , \qquad (3)$$

for each of the coincident LiDAR snow depths ($H_{s,LiDAR}$) and GPR two-way travel-times (τ). We then related the electromagnetic wave speed to the dry snow density using the Complex Refractive Index Method (CRIM; Wharton et al., 1980)

$$\rho_{s,LiDAR-GPR} = \rho_i \left(1 - \frac{v_a(v_i - v_{s,LiDAR-GPR})}{v_{s,LiDAR-GPR}(v_i - v_a)}\right) . \qquad (4)$$

The CRIM equation relies on the known wave speeds of the pore-space (v_a = 0.3 m/s) and ice matrix (v_i = 0.169 m/ns), the measured bulk wave speed of the snowpack ($v_{s,LiDAR-GPR}$; Equation 3), and the density of ice (ρ_i = 917 kg/m^3) to determine the dry snow density ($\rho_{s,LiDAR-GPR}$; Equation 4).

2.4.4 Wind and terrain exposure

Wind data were examined from 1 October 2019 through the end of the SnowEx IOP on 12 February 2020. Hourly air temperature data parameterized an empirical relationship to determine the threshold for snow-transportable wind speed (Li and Pomeroy, 1997). For values exceeding this threshold, minus the 95 % confidence interval, we then determined the median wind speed and direction for snow transport (Fig. C.1b). We utilised the maximum upwind slope (Sx) and wind factor parameters (Winstral et al., 2002; Appendix C.1) as explanation of the patterns and processes captured by the ML regression ensemble, rather than incorporating this information as model predictors of snow density. To validate the GPR-LiDAR estimated training data and the modelled results, we calculated the correlation between the model input and output and the Sx and wind factor rasters for all wind directions.

2.5 Spatial scales of variability for snow depth, travel-time, density, and SWE

We examined the differences in snow properties between forested and open areas using generalised relative semi-variograms (Isaaks & Srivastava, 1989). The generalised relative semi-variogram describes the average percent variability, relative to the mean, as a function of separation distance between observations. To estimate the spatial variability of the snow depth, TWT,

density, and the resulting SWE of the 1 m gridded data along the radar transects, the experimental variograms were first calculated in 1 m bins up to a 250 m lag, and then fitted with exponential models via least-squares to estimate the range, sill, and nugget parameters (e.g., Cressie, 1985). We used an exponential variogram model, for which the correlation length is equal to three times the range parameter. We created 250 realisations of the experimental variogram calculation using Monte Carlo simulation with 10 % random subsampling to assess the mean and standard deviation of the variogram parameters (Efron & Tibshirani, 1986).

2.6 Modelling of spatial snow density

2.6.1 Machine learning model ensemble

To distribute the spatial observations of average snow density to areas without GPR observations, we tested three regression techniques: Multiple Linear Regression (MLR; Andrews, 1974; Appendix A.1.1), Random Forest Regression (RF; Breiman (2001); Appendix A.1.2), and Artificial Neural Network Regression (ANN; Jain et al., (1996); Appendix A.1.2). We examined the ~16 km^2 area of the LiDAR domain, which closely bounded the extent of the GPR survey. A set of normalised predictor variables, notated with capital lettering, were developed using the elevations of four LiDAR rasters: bare earth elevation (Z_g), snow-covered elevation (Z_s), snow depth (H_s), and vegetation height (H_{veg}); the aspect, slope, X and Y derivatives of the elevation rasters (excluding H_{veg}); and the distance to the nearest vegetation ≥ 0.5 m (S_{veg}). Aspect rasters were transformed by the cosine to remove wrapping ambiguity around north. We smoothed the elevation, vegetation height, and snow depth rasters using a median filter with a 5 m x 5 m window, and the derivatives of these rasters (slope, aspect, dx, and dy) with a 25 m x 25 m window. Regression models were trained on the LiDAR–GPR estimated snow density using cross-validation and were applied to the surrounding terrain. By retraining the model architectures on random subsets of data, 50 model ensembles were generated and then averaged for both RF and ANN regressions. The model hyperparameters were developed such that the variance of the predictions in pixels where training data exists matches that of the predictions. The appropriate hyperparametrisation coincided with an R^2 of approximately 0.8. A ML snow density ensemble ($\rho_{s,\overline{Ens}}$) was composed by averaging the MLR, RF, and ANN outputs. For detail on the model hyperparameterisation and predictor importance see Appendix A.

2.6.2 Random Gaussian field model

To serve as a baseline for model assessment, a random gaussian field snow density model was synthesised from the statistics of in situ density measurements and the correlation length of snow density estimated via variogram analysis. Provided the empirical variogram function, a covariance matrix was determined between all pairs of points in the ~16 km^2 domain. Using Cholesky decomposition, the large covariance matrix was efficiently inverted to determine a matrix of weights with the desired covariance properties (Vecherin et al., 2022). The synthetic snow density model was then generated by multiplying a normal

random vector with zero mean and standard deviation of the in situ observations by the weighting matrix and adding the mean value of the density observations.

2.7 Distributed snow water equivalent and uncertainty

255   Multiplying $H_{s,LiDAR}$ by $\rho_{s,\overline{Ens}}$ yielded SWE ($b_{s,LiDAR-\overline{Ens}}$) distributed throughout the LiDAR domain. As a benchmark example drawn from in situ sampling, we also distributed SWE using the average snow density (276 kg/m^3), the average density of both open (280 kg/m^3) and forest (257 kg/m^3) areas, and the Gaussian random field model ($b_{s,LiDAR-Rand}$). See Appendix B.3 for additional details. We upscaled $b_{s,LiDAR-\overline{Ens}}$ to 50 m resolution using nearest-neighbour approximation for comparison with the 50 m ASO SWE.

260   Using simple linear regression, we modelled the snow density errors for $\rho_{s,\overline{Ens}}$. No correlation between error and snow depth was found, and the RMSE (11 cm) was used to estimate the random error. Using random errors in snow depth, linear errors in density, and linear error propagation, we estimated the uncertainty in $b_{s,LiDAR-\overline{Ens}}$ to first order (Raleigh & Small, 2017). Appendix B.4 has further information on SWE uncertainties regarding sampling errors estimated from in situ measurements.

3 Results

265   ### 3.1 Snow depth

The LiDAR-derived snow depths show an overall trend increasing from west to east, in addition to smaller scale patterns near vegetation, with deeper snow around the perimeter of treed areas and shallow snow on the ground beneath tree canopies (Fig. 2). This pattern is consistent with previous snow depth distribution studies of Grand Mesa (e.g., McGrath et al., 2019). The mean LiDAR snow depth for the entire domain is 92 cm with a standard deviation of 18 cm. In open areas ($H_{veg} < 0.5$ m), the

270   mean LiDAR snow depth is 96 ± 15 cm, while in the forest ($H_{veg} \geq 0.5$ m), the mean LiDAR snow depth is 79 ± 23 cm. At validation snow depth observations ($h_{s,Pit}$ and $h_{s,Probe}$) the average snow depth over the LiDAR domain is 95 ± 16 cm ($R^2 = 0.61$, RMSE = 11 cm, ME = 0 cm). LiDAR and GPR estimated snow depths, within the open and forested domains individually, are also compared to in situ snow depths (Table 1). In snow depth comparisons between $H_{s,LiDAR}$, $H_{s,ASO}$, and $H_{s,Probe}$, either LiDAR processing method shared similar correlation ($R^2 \approx 0.6$) and root-mean-square error (RMSE $\approx$ 12 cm). However, $H_{s,ASO}$

275   (86 ± 16 cm) underestimates snow depth by 7 cm and $H_{s,LiDAR}$ (93 ± 16 cm) is unbiased on average (Fig. S1; Table S1).

3.2 GPR travel-time

Ground-penetrating radar travel-time data analysed at cross-over locations exhibited a root-mean-square deviation of 1 ns with a bias of 0 ns and no systematic bias between the two GPR instruments was found. Approximately 90 % of the travel-time data applied in this work was automatically determined using the coherence method, where less than 1 % of the automatic

280   picks required manual correction. To illustrate this, automated picks are overlaid on the radargrams of a 900-metre-long

transect in Fig. 3. The resulting TWT data produced from this method and used in this study are available through the National Snow and Ice Data Center (Meehan, 2021).

3.3 LiDAR – GPR estimated density

The LiDAR–GPR retrieved average snow density shows repeatable structure at the many crossover locations and greater variability in the open terrain than areas sheltered by forest canopies (Fig. 4). The integrated LiDAR and GPR data resolve lower spatial frequency patterns than the snow pit observations, which are sparse and have limited spatial support. When compared to snow pit observations, the relative RMSE among the 37 snow pits that are within 12.5 m of the GPR transects is 35 kg/m^3 or 13 % (Table 2). The $\rho_{s,Pit}$ and $\rho_{s,LiDAR-GPR}$ data are both normally distributed as evidenced by Z-test (Appendix B.3) with overlapping standard deviations. The maximum upwind slope and wind factor parameters evaluated on GPR transects show the strongest correlation (R = -0.45 and R = 0.48, respectively) in the directions 225 and 220 degrees (Fig. S2; Table S2).

3.4 Spatial variability of LiDAR snow depth, GPR travel-time, density, and SWE

The generalised relative semi-variogram allows us to examine the differences in the length scales of variability among depth, density, SWE, and TWT within the forested and open areas of Grand Mesa (Fig. 5). Table S3 overviews the generalised relative semi-variogram parameter estimates (nugget, sill, and correlation length). TWT and SWE consistently exhibited similar correlation lengths (~ 100 m) and nugget variability within the forested (~ 35 %) and open (~ 15 %) areas. Snow depth and TWT reached comparable maximum variability in open areas (~ 25 %). Depth variability in the forested areas (~50 %) was greater than that of TWT and SWE (~ 45 %). The median distance between snow pits is ~150 m, which indicates that average snow pit observations are independent of each other and unable to resolve spatial patterns at a finer scale.

3.5 Machine-Learning Regression and Random Gaussian Field Modelled Density

Using supervised ML regression, three models were generated from LiDAR information (Fig. S3 a, b, and c). Using prior information from the in situ snow pit observations a randomly distributed density was synthesised (Fig. S3 d). The mean of the regression-based ensemble was taken to generate $\rho_{s,\overline{Ens}}$ (Fig. 6). Generally, the regression models predict higher snow density in the open and exposed areas than in areas that are protected from the wind by trees. Each of these five models are evaluated against $\rho_{s,LiDAR-GPR}$ and $\rho_{s,Pit}$ (Table 2). The ML regression snow densities are generally uncorrelated ($R^2 \approx 0.05$) having ~ 10 % RMSE with snow pit observations. The spatial similarity of these models is presented in Appendix A.3.

3.6 Model representation of wind, terrain, and vegetation effects

Maximized correlations between terrain exposure parameters for all wind directions and ML regression modelled snow density agree with the median wind direction able to transport snow (6.4 m/s at 200 degrees; Fig. S2). The wind speed and direction data from the Mesa West meteorological station (Houser et al., 2022) is presented for all time-periods October 1 – February 12

(Fig. S4 a) and time-periods where the wind was strong enough to transport snow (Fig. S4 b). Upwind slope and wind factor parameters calculated at 200 degrees wind direction are provided for reference (Fig. S5). No correlation was evident between these wind exposure parameters and the random gaussian field distributed snow densities (Table S2).

3.7 Spatially distributed snow water equivalent

SWE distributed within the ~16 km^2 domain (Fig. 7) underestimated the snow pit average SWE by 20 mm (R^2 = 0.57; RMSE = 46 mm; Table 3; Fig. S6 a-c). Upscaling $b_{s,LiDAR-\overline{Ens}}$ to 50 m resolution resulted in decorrelation (R^2 = 0.03) and increased error (RMSE = 60 mm), and the bias remained nearly the same (-18 mm; Fig. S6 d-f). ASO SWE (b$_{s, ASO}$; 50 m resolution) had similar errors (RMSE = 57 mm), bias (-21 mm), and low corelation (R^2 = 0.1) to measurements (Fig. S6 g-i). Fifty metre $b_{s,LiDAR-\overline{Ens}}$ had mean SWE and standard deviation of 245 ± 33 mm while b$_{s, ASO}$ was 236 ± 45. The average snow density of b$_{s, ASO}$ was 274 kg/m^3 (versus 264 kg/m^3 $\rho_{s,\overline{Ens}}$). Distributing SWE using an average measured value showed lower bias (-9 mm) and root-mean-square error (38 mm) than $b_{s,LiDAR-\overline{Ens}}$. Using a constant density drives SWE spatial patterns and data correlation solely by snow depth. Tested against the random Gaussian distributed density (R^2 = 0.49), the ML regression modelled densities explained more variation in the distributed SWE estimates (R^2 = 0.56), about as much as snow depth alone. For comparison $b_{s,LiDAR-\overline{Ens}}$ is presented alongside H$_{s,LiDAR}$ and $\rho_{s,\overline{Ens}}$ (Fig. S7).

3.8 Contributions of uncertainty

3.8.1 LiDAR snow depth

H$_{s,LiDAR}$ snow depths were computed from point-cloud differencing, rather than raster differencing. The relative accuracy of the snow depth measurement was estimated at 7 cm, based on the maximum standard deviation of the point cloud differencing method, which agrees well with previous LiDAR error assessments using this approach (Hojatimalekshah et al., 2021). The absolute accuracy of the snow depths (11 cm) was determined by comparison to validation snow depth measurements, which shows 0 cm of bias on average over all measurements (Table 1). On ploughed roads, we observed snow depths ranging between 0 and 5 cm (Figure S8).

3.8.2 LiDAR – GPR co-registration

We estimated the horizontal accuracy of the multi-polarization GPR positions, which had clear sky view and GNSS receiver, at ± 1 m. The GPR system operated within the forest stands, had lower GPS fidelity which we conservatively estimate at ± 3 m. Ground validation showed less than 1 cm horizontal position uncertainty within the LiDAR point cloud. We found that errors in the co-registration of the LiDAR and GPR data are the leading source of error in the estimated densities. The overall accuracy of the spatial registration between the LiDAR and GPR varies on the order of a few metres.

Sensitivity analysis showed how measurement errors propagate into the LiDAR – GPR measured snow density (Appendix B.2). At the 1-sigma confidence interval, we found that measurement errors are on the order of 10 cm for LiDAR and 1 ns for

GPR may translate into errors in the density estimate of 150 kg/m^3 or greater. Variogram analyses shows that beyond 10 m these errors are spatially uncorrelated and can be treated with random noise filtering. Via median filtering and interpolating through outliers, error estimates reduced to 30 kg/m^3 (Appendix B.2).

3.8.3 Spatial and temporal support of density measurements

345   Measurements accumulated over 12.5 m distance introduce inherent variability on the order of 10 % (Section 3.4). We found that the expected variability among co-located $\rho_{s,LiDAR-GPR}$ is approximately 2 %, which is consistent with replicated in situ density observations ($\rho_{s,Pit}$). The average density sampled from each of the snow pit columns (within 1 m) shows high repeatability with a root-mean-square deviation of 2.5 %. We observed linear temporal trends in snow densification of 0.07 kg/m^3/hr in forested areas and 0.13 kg/m^3/hr in open areas. These trends were removed and were not considered within

350   uncertainty analyses.

3.8.4 SWE uncertainty

The errors between H$_{s,LiDAR}$ and h$_{s,Probe}$ are uncorrelated when compared to H$_{s,LiDAR}$ (R^2 = 0.04). However, the errors between $\rho_{s,\overline{Ens}}$ and $\rho_{s,Pit}$ are negatively correlated against $\rho_{s,\overline{Ens}}$ (R^2 = 0.35, RMSE = 27 kg/m^3). The errors among snow depth and density are uncorrelated (R^2 = 0.03) with negligible covariance (Fig. S9). The distributed relative SWE uncertainty is presented

355   in Fig. S10 and is negatively correlated with the distributed SWE (R^2 = 0.57). The distributed absolute SWE uncertainty (Fig. S11) is weakly positively correlated with $b_{s,LiDAR-\overline{Ens}}$ (R^2 = 0.22). The median SWE uncertainty is 13 % (30 mm), which breaks down to 15 % (29 mm) median uncertainty in the forest and 13 % (31 mm) median SWE uncertainty in the open areas. Relative uncertainty is greater within forests due to decreased total SWE within stands. Median snow density relative uncertainty equal is 4 % (10 kg/m^3). Median snow depth relative uncertainty is 12 % (11 cm). Median LiDAR snow depth

360   relative uncertainty is 12 % (11 cm) in open areas, while in the median depth uncertainty is 14 % (13 cm). Median snow density uncertainty is 4 % in both open and forested areas, where more extreme density values have greater uncertainty due to linear error modelling. Had snow density been modelled as a random variable, density contributes 10 % error (27 kg/m^3), equating to 16 % uncertainty in SWE on average.

4. Discussion

365   **4.1 Mulitpolarization GPR**

This work advances the utility of GPR for seasonal snow applications, by resolving spatial snow density and SWE through the integration of remotely sensed LiDAR and GPR observations. Grand Mesa is a good site for testing our approach of combining LiDAR and GPR for density and SWE retrieval yet presents many challenges for GPR analysis because of the abrupt discontinuities along reflection horizons due to vegetation and boulders on the ground surface. By exploring effects of

370   depolarization on L-Band GPR signals, we developed a new, automated GPR processing workflow that reliably identifies the ground surface beneath the snow-cover. This advance encourages the collection of large multi-polarization GPR datasets for operational use by removing the subjectivity involved in the GPR post-processing and interpretation and alleviating the labour of manually interpreting radargrams through an objective function.

4.2 SWE uncertainty

375   Our work characterised the measurement uncertainties and the resulting SWE uncertainty in pursuit of the goal for 10 % uncertainty in global SWE estimation (National Academies of Sciences, Engineering, and Medicine, 2018). We found that GPR TWT and LiDAR snow depth contribute approximately equal uncertainty to the retrieved snow density. The uncertainty in LiDAR snow depth tends to vary spatially and is dependent on landscape characteristics such as slope and vegetation (Deems et al., 2013). However, our evaluation of snow depth in forested and open areas did not suggest that LiDAR snow depth errors

380   were greater beneath the tree canopy (Table 1). Relative SWE uncertainty resembles the snow depth distribution, where shallow snow within forest stands yields greater relative uncertainty. Absolute SWE uncertainty resembles the snow density patterns, especially in areas of anomalous snow density where the modelled errors are greatest. Our uncertainty analysis (Fig. S10) suggests that, even using innovatory measurements from airborne and ground-based sensors, 10 % uncertainty is difficult to achieve. Having demonstrated the limitations and success of the joint LiDAR – GPR methodology for SWE retrieval

385   at plot to forest stand scales makes this technique formidable for SWE remote sensing calibration and validation.

4.3 Geolocation errors in sensor fusion

Though the signal of LiDAR and GPR instruments is repeatable and coherent, the leading source of error in our density measurement is spatial misalignments (potentially sourced from geolocation inaccuracies, point cloud to raster processing, and coordinate transformations) that are on the scale of the 1 m resolution data products. In some locations the co-registration may

390   be nearly exact between the two instruments, and the resulting error will be low. We found by cross-correlating the GPR and co-located LiDAR snow depth transects, that misalignments of approximately 1 – 5 m are possible. Utilizing 3 m resolution LiDAR snow depths paired with GPR TWT for snow density retrieval did not reduce the errors. To evaluate how spatial misalignments impact the training data, predictor data, and the regression model output, and to estimate the uncertainties introduced from integrating the cross-platform sensor data, we created multiple sets of training data by effectively perturbing

395   where LiDAR – GPR transects are aligned via cross-correlation lagging, and introduced common practice mistakes in the sensor integration, such as mixing the geographic coordinate system of the data between NAD83 and WGS84. We found that perturbing the sensor integration introduces less than 1 kg/m^3 error in the modelled density on average (up to 5 kg/m^3 in forest stands), that outlier filtering is robust to sensor integration errors, and this error is small relative to the overall SWE uncertainty.

4.4 Spatial variability in snow hydrological properties

Measurements retrieved from GPR profiles permitted us to quantify the spatial length scales of variability in TWT, snow depth, snow density, and SWE. We determined that density measurements up to ~100 m apart are correlated. These findings differ from a previous variogram analysis that found correlation lengths for snow density of less than 10 m at a smaller study site which limited the maximum lag separation to approximately 50 m (Yildiz et al., 2021). Variability is study site dependent (Bonnell et al., 2023), and it may be that we have identified an additional longer, lower spatial frequency scaling of snow density. Corollary to SWE, two-way travel-time in dry snow depends both on snow depth and density. We found that TWT and SWE consistently exhibited similar correlation lengths (~ 100 m) and nugget variability in the forested (~ 35 %) and open (~ 15 %) areas. This finding supports TWT as an informer of spatial SWE variability. We found that snow depth and TWT reached comparable maximum variability in open areas (~ 25 %). While depth variability in the forested areas (~50 %) was greater than that of TWT and SWE (~ 45 %).

In situ snow density observations have limited spatial support and tend to examine shorter length-scales of variability than are expressed in distributed models or retrievable by remote sensing. As a random sampling strategy was targeted for the IOP, the median distance between snow pit observations in our study is beyond the length scale of variability for snow density. The observations are therefore spatially uncorrelated. Snow density exhibited 4 % greater variability in the open areas than in the forests, indicating that wind exposure increased the variability, and conversely, shelter provided by terrain and vegetation tended to reduce spatial density variability. The larger (roughly 10 m) spatial support of the LiDAR – GPR estimated densities cannot directly sense subpixel correlation lengths and potentially missed a zero-to-five-metre scale-break that is more comparable to the spatial support of in situ density observations. Differences between representative observation scales may explain the weak correlation between estimated density and the in situ measurements ($R^2 = 0.01$ in open areas, $R^2 = 0.39$ in forest stands; Table 2 b). Our analysis of the correlation length of LiDAR snow depths generally agrees with scale-breaks identified in previous studies within forested and open areas (Deems et al., 2006; Marshall et al., 2006; Trujillo et al., 2009). We observed decreased correlation and increased error when upscaling rasterized SWE to 50 m resolution. Which suggests, that evaluating 50 m SWE with sparse point measurements, may not be the most representative approach, and greater than 20 % error can be expected due to spatial variability.

4.5 Density modelling using ML regression

The ML regression models developed from the LiDAR and GPR acquisitions during the SnowEx 2020 Grand Mesa IOP will likely have weak predictive capability at other field sites, requiring model recalibration. The SWE predictions, here, represent a single snapshot in time of snow depth and density. The patterns in depth, density, and SWE may be characteristic for mid-winter, dry snow conditions but other times of the year may exhibit different spatial patterns in all three (e.g., due to variable melt or liquid water during ripening). For each regression model, we identified the most important LiDAR features used to distribute density and found dependencies in predictor importance due to model choice and architecture (Appendix A.2).

Vegetation height and proximity to vegetation greater than 0.5 m in height appeared prominent in the three regression models, whereas dependence on elevation, slope, and aspect for snow density at Grand Mesa was weakened. We used a "kitchen-sink" approach to our regression modelling but found comparable accuracy in models using fewer parameters. Evaluations against snow pit density did not depict an obvious best model choice, as so we averaged the regression models.

To capture the range of terrain features (i.e., elevation, slope, aspect, and forest attributes) that influence snow densification, one field campaign in the western U.S. collected density measurements from 300 snow cores at $10 - 20$ m intervals and 17 snow pits (Broxton et al., 2019). From these observations, bulk snow density was distributed at 1 m resolution using an ANN combined with airborne LiDAR-derived snow depth to estimate SWE. Broxton et al. (2019) highlighted the importance of representing the broader landscape with distributed densities for estimating SWE, finding $\sim$30 % differences between the distributed estimates and observations from a nearby SNOTEL station. Elder et al. (1998) used a simpler, three feature (net radiation, slope, and elevation, with an intercept) MLR model that was trained on density observations of five snow pits and averages of five snow core transects to predict basin-wide average density and SWE. More recently, a similar study used a sampling strategy to represent unique classes of basin-wide physiography, acquiring $\sim$1000 snow core observations, and used MLR and binary-classification tree models to distribute density from elevation and incoming radiation (Wetlaufer et al., 2016). The dependence of density on net solar radiation may explain the good performance of these models, whereas terrain parameters, such as slope and aspect, indirectly relate to radiation.

We tested the model sensitivity to training and learned how much data is required for ML density estimation. Using approximately 30,000 LiDAR – GPR derived densities (10 % of the total) from random subsets, we obtained density models that are statistically identical to those generated from the larger data set. Though random sampling is not a practical method for GPR data acquisition and analysis, this exercise showed that the amount of GPR information required to train the model parameters is not as important as collecting data in a variety of landscape and snow-cover characteristics.

4.6 Model representation of snow, terrain, vegetation, and wind interactions

The LiDAR predictors were inspired by theory of wind-terrain-vegetation interactions governing snow distribution (Winstral et al., 2002). Though to keep the model design innate to LiDAR information, we did not include wind data or predictors such as maximum upwind slope and wind factor. Instead, these wind parameters were utilised as a corroboratory metric for explaining spatial patterns retrieved by the LiDAR – GPR density and predicted in ML regression modelling. We support that spatial patterns are representative of snow exposure or shelter to wind provided by topography and vegetation. We found that the prevailing wind direction capable of transporting snow at Mesa West generated upwind slope and wind factor parameters that agree most strongly with the retrieved and ML modelled snow density. A larger correlation is observed for the wind factor than maximum upwind slope, which suggests that the role of wind shelter by vegetation, and not only terrain, has an important quantifiable effect on snow density. The role of forest vegetation on snow density is evinced in this topographically simple environment because a large-scale topographic trend, such as one driven by elevation or aspect, does not saturate the signal in retrieved density. Our approach in validating wind effects on density is an explanatory simplification of the controls on snow

density which may be further impacted by forest stands. Effects such as the blocking of short-wave and emitted long-wave radiation from forest canopies, delivery of canopy intercepted snow to the snowpack, or the loss of snow mass and thereby altered compaction due to sublimation of canopy intercepted snow were unaccounted for (Bonner et al., 2022).

4.7 Value of pit observations for distributed SWE

The Grand Mesa IOP was one of the largest campaigns to examine the spatiotemporal patterns of SWE and provided a rich data source for snow density analysis. In most circumstances few snow pits are dug, and uncertainties arise from spatial sampling of the underlying density distribution. The distribution of density on Grand Mesa during the early February SnowEx 2020 campaign appears as a random normal variable with mean and standard deviation of 276 ± 21 kg/m^3, despite differences of roughly 25 kg/m^3 between the average density of forested and open areas. The large sample size of snow pits allowed us to accurately quantify the mean snow density for distributed SWE estimates. While the uncertainty in any measurement of density was found to be 2.5 % on average, we sought to quantify the degree of uncertainty in the SWE distributed from the sampled population of snow density as a function of sample size. Our analysis suggests that 10 spatially random snow pit observations within the study domain are sufficient to reduce the median uncertainty in distributed SWE to within 10 ± 2 % (Appendix B.4). Although the differences are marginal, we have shown that on average this simpler approach for distributing density more accurately represents the in situ observations than the SWE distributed using our modelled estimates of density (Section 3.7) but resolves no information about spatial patterns of snow density, and is therefore not useful for understanding density patterns across the landscape. In situ snow campaigns targeting average SWE require far fewer pits than needed to resolve the spatial patterns.

Snow pits are an invaluable source of calibration and validation observations but do not adequately scale spatially, incur human errors and biases, and are time intensive to sample. For example, a team of two can fully sample a one-metre-deep SnowEx pit in two hours, which for the approximately 100 snow pits in the study area, amounts to ~ 400 hours of labour (excluding the time to quality control (QC), curate the snow pit logs, and travel to and from the field site). The 160 km of GPR data used in this work required approximately 20 hours to collect and an additional 20 hours to QC TWTs, which amounts to ~ 40 hours, or roughly a 90 % reduction in field labour, excluding the labour for acquisition and processing of airborne LiDAR. We must note the greater financial cost to obtain GPR equipment and outsource airborne LiDAR data collection.

However, densities estimated from GPR TWTs and LiDAR snow depths are objective, repeatable, and offer the spatial continuity and areal coverage to provide insights to the spatial patterns of density. The expense of acquiring airborne remote sensing data is a crux of the technique, and it may not be feasible to fly entire catchments across the breadth of snow climates. Less expensive techniques for estimating SWE distribution, such as drone-based radar retrievals of dielectric permittivity (e.g., Valence et al., 2022), and in situ measurement campaigns combined with ML models (e.g., Wetlaufer et al., 2016; Broxton et al., 2019) should be utilised where appropriate and examined for physical representativeness.

5. Conclusion

We developed an innovative approach to estimate SWE across a ~16 km^2 domain by evaluating GPR travel-times for bulk snow density given a snow depth constraint, then extrapolating across the domain using machine learning. Our automatic and objective technique for interpreting radargrams reduces post-processing labour, which is a primary hindrance to widespread use of GPR in snow science. We leveraged LiDAR estimated snow depth to solve for snow density along ~160 km of GPR transects. From these along-track estimates, we calculated the length-scales of variability for depth, density, SWE and found that snow pit observations following the SnowEx 2020 Grand Mesa IOP sampling strategy are independent and unable to resolve spatial patterns < 150 m scale. Radar travel-time informed dry-snow SWE variability better than either depth or density independently. Snow density distributed by machine learning revealed anomalies associated with localised terrain features and forest stands that shelter the snowpack from wind densification. Density spatial patterns show the best agreement in the direction of prevailing winds strong enough to transport snow, where roughly 60 % of density variability in our single mid-winter survey can be accounted for using a wind factor analysis. On average, distributed relative SWE uncertainty was less than 15 %. While our analysis suggests that measurements from one snow pit per km^2 may reduce the SWE uncertainty to within 10 ± 2 %, using such a sampling strategy would not resolve spatial patterns and variability in snow properties. This pilot study provides a useful method for resolving explanatory spatial patterns in snow depth, density, and water equivalent with comparable uncertainty to in situ methods but with spatial continuity at resolutions practical for calibration or validation of air or spaceborne radar remote sensing retrievals of SWE. Spatial data generated and the validation data acquired during the NASA SnowEx 2020 IOP at Grand Mesa, Colorado, USA, provides a core set of observables that continue to inform essential snow research.

Appendix A.

A.1 Regression parameter optimization

A.1.1 Multiple Linear Regression

The MLR model has the form

$$y = X\beta + \epsilon, \tag{A.1}$$

where y is the observed density along the GPR transects, X is a matrix with columns containing the normalised LiDAR predictors at the coordinates along the GPR transects, β is the vector of the regression coefficients which we seek to estimate, and ϵ represents the model residual. From the method of least squares, the multiple linear regression coefficients are estimated as

$$\beta = (X^T X)^{-1} X^T y . \tag{A.2}$$

Using cross-validation to assess the model accuracy and sensitivity, we estimated the MLR model parameters. We trained the model with 1000 Monte Carlo simulations by randomly sampling 90 % of the density observations and testing on the remaining

10 %. Additionally, we repeated this process and randomly sampled only 10 % of the data and tested on the remaining 90 %. In doing so, we created two sets of parameters that robustly span the parameter space. Using these regression coefficients, Equation A.1 is computed to distribute the predicted densities. The modelled densities are insensitive to the training choice for parameter estimation, as the root-mean-square deviation between the two models is less than 1 kg/m^3.

A.1.2 Random Forest and Artificial Neural Network Regression

Whereas MLR models are relatively inflexible and model overtraining is not a concern, techniques such as Random Forest and Artificial Neural Network regression are highly tuneable and may overfit the data. Hyperparameters determine the model architecture which is often designed subjectively or through an optimization process. The number of trees and the minimum leaf size of a tree were the hyperparameters adjusted for the Random Forest method. Neural networks offer a greater hyperparameter space, allowing for design of the number of and size of hidden layers, the neuron activation function, and model regularisation. The machine learning models were implemented using the MATLAB Regression Learner application, where it was determined that model hyperparameters which minimise the cross-validation mean squared-error overfit the data. Model overfitting was remedied by training ensembles of models with various hyperparameters. We calculated the averaged standard deviations of density data for each ensemble predicted along the GPR transects ($\overline{\sigma_{train}}$) and elsewhere in LiDAR domain ($\overline{\sigma_{pred}}$) and the coefficient of determination (R^2) of the training data and prediction. Optimal hyperparameters which do not overfit the data were then determined by minimising the objective function

$$\varphi = \frac{1}{R^2} \frac{\overline{\sigma_{pred}}}{\overline{\sigma_{train}}} . \tag{A.3}$$

The ratio of the standard deviations asserts that an appropriate model will have similar variance throughout the modelled domain, by penalising overfit data in the training locations and rewarding the model which explains the data accurately. We found that the best model parameterizations that are not overfit scored an $R^2 \cong 0.8$ with $RMSE \cong 15$ kg/m^3. The corresponding hyperparameters for the Random Forest Regression were 10 trees with a minimum leaf size of 200. The ANN architecture had two hidden layers each with 50 neurons and hyperbolic tangent activations and regularisation of $\lambda = 0.015$.

A.2 Predictor importance

A.2.1 Multiple linear regression

We applied the "kitchen-sink" approach because the MLR model that was trained using the LiDAR–GPR densities, which utilised every LiDAR predictor, exhibited the largest correlation ($R^2 = 0.27$) to the observations. However, various model parameterizations which utilised few parameters yielded equivalent accuracies. To assess the importance of the individual predictors, we assembled all combinations of 1 to 17 predictor models, solved the regression for each combination, and cross-validated against a test set of the LiDAR–GPR estimated densities. We considered optimal models as the top 1 % of outputs, and from these we tracked which predictors composed any model. We identified the relative importance of each predictor (Fig. A.1) by summing the number of appearances for a given predictor and dividing by the number of optimal models. Vegetation

parameters and the east-west gradient of the ground surface elevation were featured in all the most accurate modelled predictions. Notably, snow density on Grand Mesa exhibits weak dependence on elevation, aspect, and slope.

A.2.2 Random forest regression

560 The permutation accuracy importance was calculated to determine which of the LiDAR derived predictor variables are most valuable in predicting the response. The permutation importance is assessed by comparing the accuracy of the prediction for a given learner (tree), then randomly permuting the predictor variable of interest and recalculating the prediction accuracy (Hapfelmeier et al., 2014). An important predictor will lose predictive capability after random permutation, while an unimportant predictor will be unaffected by the randomization. The prediction accuracy is calculated using "out of bag"

565 observations that were excluded from the population used to build the decision tree (Breiman, 2001). The relative "out of bag" predictor importance for the ensemble of random forests generated using 10 trees with a minimum leaf size of 200 suggests that the slope of snow depth, snow surface elevation, ground surface elevation, proximity to vegetation, and vegetation height are the five leading predictors of snow density (Fig. A.2).

A.2.3 Artificial neural network regression

570 Approaches which partition the weights between neural connections to determine the relative importance of the predictors within an ANN have a classically simple architecture with a network expressing one hidden layer (Goh, 1995). This technique becomes obfuscated when applying an ANN with multiple hidden layers. To determine the relative importance of predictors within an ensemble of networks with two hidden layers, we simply multiplied the matrices of weights connecting the input to the first hidden layer, the first to the second hidden layer, and second hidden layer to the output. The greater the overall

575 weighting that is assigned to a predictor the greater the importance of the feature. This method suggests that vegetation height, proximity to vegetation, the north derivative of the ground surface elevations, the slope of the ground elevations, and east derivative of the snow surface elevations are the five leading predictors of snow density (Fig. A.3).

A.3 Model similarity intercomparison

Visual inspection reveals apparent structural similarity among the three regression-based models. As a quantified model

580 intercomparison we applied the coefficient of determination measured by Pearson correlation calculated on a pixel-by-pixel basis. The structural similarity index (SSIM; Wang et al., 2004) is a normalized value between 0 and 1 that is defined by image luminance, contrast, and standard deviation. We calculated the SSIM in 100 m radius kernels (comparable to the estimated correlation length of density) as a second means of determining the similarity among the model ensembles. Capturing various structural length scales examines the model similarities throughout the correlation length. Table A.1 overviews the R^2 similarity

585 matrix. Nearly 50 % of the features observed in the MLR model are explained within the RF model, and vice-versa. The SSIM similarity matrix suggests greater structural similarity at larger spatial support (Table A.1). The random field synthesised model

exhibits no repeatable structure and is uncorrelated from the various regression models, as expected since it was randomly generated from the overall snow density statistics only.

Appendix B.

590 **B.1 LiDAR point cloud processing**

Point cloud processing was performed using the Multiscale Model to Model Cloud Compare method (Lague et al., 2013). The point cloud differencing method operates directly on point cloud data by selecting core points from the point clouds and distinguishing between the reference point cloud (representing the ground) and the compared point cloud (depicting snow-covered surfaces). Utilizing a user-defined radius, denoted as D, the algorithm fits a plane within a D/2 radius around each

595 core point and determines the normal vector of the plane. Subsequently, based on the core point and the associated normal vector, the algorithm fits a cylinder with a radius of D/2, oriented along the normal vector, with the cylinder axis passing through the core point. This cylinder intersects both the reference (ground) and compared (snow) point clouds, resulting in two sets of points. The algorithm then projects the points within each set onto the normal vector and calculates the average and standard deviation for each set. These values represent the average position and roughness, respectively. In the presence of

600 outliers, the algorithm substitutes the mean and standard deviation with the median and inter-quartile range, respectively. Ultimately, the local distance between the average positions of the two sets provides the snow depth. However, in cases where surfaces are rough and the surface orientation does not align consistently with the normal direction, the measured distance (snow depth) uncertainty increases. This point cloud differencing method could prove beneficial for flat areas, given the effectiveness of the method on both rough and smooth surfaces.

605 The Cloud Compare Caractérisation de Nuages de Points method was employed to distinguish vegetation from ground and snow returns following methods in Štroner et al., (2021). This process involves a combination of training and classification which defines the dimensionality of point clouds (1D for a line, 2D for a plane, or 3D for a volume) around specific points within a sphere at various scales (radii). The expectation is that branches exhibit a more linear (1D) structure, leaves display 2D surfaces at small scales (e.g., centimetre scales), and trees manifest in 3D at larger scales (e.g., metre scales). Both snow

610 and ground surfaces tend to exhibit more of a 2D nature at both small and large scales. The determination of dimensionality is based on principle component analysis within the algorithm. The combination of point dimensionality at different scales is used to define object categories, and in this instance, remove vegetation.

B.2 Error analysis of LiDAR – GPR estimated density

We conducted a sensitivity analysis to evaluate how the errors in radar travel-time and LiDAR snow depth affect the estimated

615 snow density. This involved establishing a level curve through the average snow values of the study area (mean density 276 kg/m^3, mean TWT 8 ns, and mean snow depth 96 cm) and applying perturbations to evaluate the density error (Fig. B.1). Perturbations of up to ± 1 ns were added to the TWT and ± 15 cm were added to the depth. After the TWT and depth

perturbations were applied, the densities were evaluated, and the mean (276 kg/m^3) was subtracted from this result to measure the density perturbations. The error bars of Fig. B.1 represent the LiDAR root-mean-square deviation evaluated by co-located depth probing (11 cm) and at the GPR TWT crossovers (0.9 ns). At the 1-sigma level, errors of approximately $\pm$ 150 kg/m^3 can be expected from this sensor integration method.

The combined measurement and geolocation errors in LiDAR-derived snow depths and GPR TWTs may translate to errors in the retrieved density that are larger than the range of densities observed in the snow pits. By filtering outliers, random error reduced to $\pm$ 30 kg/m^3 and retrieved a meaningful density signal. We chose the interquartile range of the LiDAR – GPR retrieved densities as the threshold for determining outliers, because the 25th and 75th percentiles envelop the range of snow densities observed in the snow pits. We then applied a 2D median window with a 12.5 m radius (chosen to extend beyond the correlation length of the errors) to smooth the densities and interpolate those at the outlying locations.

Although we only included GPR observations within 100 hours of the 1 February LiDAR flight, seeking a potential bias related to effects of densification and redistribution, we regressed $\rho_{s,LiDAR-GPR}$ against hours elapsed prior to and after 1 February. Separate linear trends were identified in the forested and open regions traversed with the GPR. As conducted for the snow pit density observations, the trends were centred about February 1, and removed from the retrieved density observations.

B.3 Statistics of in situ, LiDAR – GPR retrieved, and modelled snow density

To show that the retrieved and modelled densities are within the range of measurements observed in the snow pits we provided the distribution of these three data sets for the entire study area domain (Fig. B.2). The means of the distributions are overlapping within the standard deviations of the datasets. The LiDAR – GPR retrieved densities suggest a broader distribution of densities than observed in pits or modelled. Sampling biases may explain the small disagreement between mean values. The sample size and spatial representation of each data set varies on many orders of magnitude. The distributions are unequally represented by vegetation class, as 18 % of the snow pits (17 of 96), 23 % of the modelled domain (3,665,343 of 15,753,500), and 7 % of the GPR transects (19,978 of 278,627) were located within the forest stands. However, other than a bias of - 10 kg/m^3 the distribution of modelled estimates closely resembles that of the snow pit measurements. Z-tests confirm that the snow density data are normally distributed about the mean and standard deviations listed with high confidence.

B.4 Sample Uncertainty of Density and SWE

To estimate the uncertainty in average density due to sample size and to propagate this uncertainty in terms of SWE, we conducted Monte Carlo simulations by randomly subsampling density observations. Utilising 1,000 Monte Carlo realisations as a function of sample size, we incrementally increased the number of randomly sampled snow pits to estimate the mean and the standard deviation of the distributed average snow density (Fig. B.3). We set the sampled standard deviation as the spatial uncertainty in density and summed in quadrature the distributed errors in LiDAR snow depth as described in Section 2.7 to propagate the SWE uncertainty (Fig. B.2). Our analysis suggests that 10 snow pit observations are sufficient to reduce the median uncertainty in SWE to within 10 ± 2 %.

Appendix C.

C.1 Maximum upwind slope and wind factor parameters

To infer the physical bases underlying the snow density patterns of $\rho_{s,\overline{Ens}}$, we computed two parameters (maximum upwind slope and wind factor parameters; Winstral et al., 2002) from the 1-2 February LiDAR-derived snow surface elevations. The maximum upwind slope characterises the degree of wind exposure for a given pixel. A wind sheltered pixel has positive slope, indicating that higher terrain exists upwind of the pixel. Conversely, wind exposed terrain has a negative slope value, with lower elevation terrain in the upwind direction. The maximum upwind slope parameter (Sx) was calculated for each azimuth from 0 to 355 degrees in 5-degree increments, averaged over ± 15-degree overlapping bins (Winstral et al., 2002). A search distance of 25 m was applied in the calculation of local Sx, while a 250 m distance was used to calculate the outlier length scale Sx between 25 and 250 m from the pixel. The local and outlying Sx parameters were differenced to calculate the slope break parameter, Sb (Winstral et al., 2002). Because pixels with high wind exposure have negative Sx values, and we hypothesise that wind exposed terrain will have greater snow density (e.g., due to formation of wind slabs), the wind direction expressing the largest negative correlation was recognized as the best explanation of density patterns determined by Sx.

The wind factor parameter determines the degree of wind exposure or shelter for a given pixel, based on Sx, Sb, vegetation proximity, and the average scalar multiple by which wintertime winds at the MW station exceeded those at the MM station (Winstral et al., 2002). The value 2.18 determined from the wind speed data at MW and MM is in close agreement with the value of 2.3 determined by Winstral et al. (2002) and was applied to compute the wind factor by inversely rescaling Sx to the range between 1 and 2.18, where the larger values indicate more wind exposure. Vegetated, wind-sheltered zones were identified as pixels within a 3 m buffer of LiDAR vegetation greater than 0.5 m height. The wind factors of sheltered vegetated areas and regions of Sb which exceeded the 97.5[th] percentile were arbitrarily reduced by 10 % to enhance the effective wind shelter provided by vegetation. Because the wind factor was inversely rescaled, the wind direction expressing the largest positive correlation with the snow density was recognized as the best explanation of density patterns determined by the wind factor parameter.

Code availability

The processing software developed for the multi-polarization GPR analysis is available at https://github.com/tatemeehan/SnowEx2020_BSU_pE_GPR.

Data availability

Snow pit observations (Vuyovich et al., 2021), snow depth observations (Hiemstra et al., 2020), and GPR travel-time observations (Meehan, 2021; Webb, 2021) acquired during the SnowEx 2020 Grand Mesa IOP are publicly available through

the National Snow and Ice Data Center (NSIDC). Data products resulting from this work, comprised of the LiDAR–GPR
680 estimated density observations; LiDAR estimated snow depth, terrain, and vegetation model predictors, the ensemble of
modelled snow density, and the SWE estimated therefrom are archived within the NSIDC at https://nsidc.org/data/snowex
DOI: 10.5067/LANQ53RTJ2DR. Additional ASO datasets evaluated within this manuscript are hosted on the NSIDC landing
page https://nsidc.org/data/aso.

Author contribution

685 HPM, CH, and KE organized and led the SnowEx 2020 Grand Mesa IOP; TM, AH, HPM, DM, RW, RB, CH, and KE collected
field measurements and curated various datasets utilized in this work; HPM, ED, CH, and KE acquired funding for the field
effort and labour in preparing this manuscript; TM, AH, HPM, RB, MR, and KE conceptualized the goals of this research; TM
developed the methodology, conducted the formal analysis and prepared the manuscript; all co-authors provided feedback to
the manuscript through review and editing.

Competing interests

690

The authors declare that they have no conflict of interest.

Acknowledgements

This research was funded under the U.S. Army Engineer Research and Development Center (ERDC) Basic Research Program
through Program Element 601102/Project AB2/Task 1. Permission was granted by the Director, Cold Regions Research and
695 Engineering Laboratory, to publish this information with unlimited distribution. The findings of this document are not to be
construed as an official Department of the Army position unless so designated by other authorized documents. The use of
trade, product, or firm names in this document is for descriptive purposes only and does not imply endorsement by the U.S.
Government. Funding was also provided by the NASA Terrestrial Hydrology Program (THP) in coordination with the SnowEx
program awards NNX17AL61G and 80NSSC18K0877. We wish to thank all the scientific participants and logistical support
700 staff of the SnowEx 2020 campaign, who helped make this research possible.

References

Cartographic   Boundary   Files,   https://www.census.gov/geographies/mapping-files/time-series/geo/cartographic-
boundary.html, 2020.

Andrews, D. F.: A robust method for multiple linear regression, Technometrics, 16, 523–531,
705   https://doi.org/10.1080/00401706.1974.10489233, 1974.

Bentley, J. L.: Multidimensional Binary Search Trees Used for Associative Searching, COMMUN ACM, 18, 509–517, https://doi.org/10.1145/361002.361007, 1975.

Besso, H., Shean, D., and Lundquist, J. D.: Mountain snow depth retrievals from customized processing of ICESat-2 satellite laser altimetry, REMOTE SENS ENVIRON, 300, 113 843, https://doi.org/10.1016/j.rse.2023.113843, 2024.

710 Bonnell, R., McGrath, D., Hedrick, A. R., Trujillo, E., Meehan, T. G., Williams, K., Marshall, H. P., Sexstone, G., Fulton, J., Ronayne, M. J., Fassnacht, S. R., Webb, R. W., and Hale, K. E.: Snowpack relative permittivity and density derived from near-coincident lidar and ground-penetrating radar, HYDROL PROCESS, 37, https://doi.org/10.1002/hyp.14996, 2023.

Bonner, H. M., Raleigh, M. S., and Small, E. E.: Isolating forest process effects on modelled snowpack density and snow water 715 equivalent, HYDROL PROCESS, 36, https://doi.org/10.1002/hyp.14475, 2022.

Booth, A. D., Clark, R., & Murray, T. (2010). Semblance response to a ground-penetrating radar wavelet and resulting errors in velocity analysis. *Near Surface Geophysics*, *8*(3), 235–246. https://doi.org/10.3997/1873-0604.2010008 Booth, A. D., Clark, R., and Murray, T.: Semblance response to a ground-penetrating radar wavelet and resulting errors in velocity analysis, NEAR SURF GEOPHYS, https://doi.org/10.3997/1873-0604.2010008, 2010.

720 Boyd, D. R., Alam, A. M., Kurum, M., Gurbuz, A. C., and Osmanoglu, B.: Preliminary Snow Water Equivalent Retrieval of SnowEX20 Swesarr Data, in: Proceedings of the 42nd IEEE International Symposium on Geoscience and Remote Sensing IGARSS vol. 2022-July, pp. 3927–3930, ISBN 9781665427920, https://doi.org/10.1109/IGARSS46834.2022.9883412, 2022.

Breiman, L.: Random Forests, Machine Learning, vol. 45, pp. 5–32, https://doi.org/https://doi.org/10.1023/A:1010933404324, 725 2001.

Broxton, P. D., Leeuwen, W. J. D., and Biederman, J. A.: Improving Snow Water Equivalent Maps With Machine Learning of Snow Survey and Lidar Measurements, WATER RESOUR RES, 55, 3739–3757, https://doi.org/10.1029/2018WR024146, 2019.

Cartographic Boundary Files. (2020). U.S. Census Bureau. https://www.census.gov/geographies/mapping-files/time-730 series/geo/cartographic-boundary.html

Cressie, N.: Fitting variogram models by weighted least squares, J INT ASS MATH GEOL, 17, 563–586, https://doi.org/10.1007/BF01032109, 1985.

Deems, J. S., Fassnacht, S. R., and Elder, K. J.: Fractal Distribution of Snow Depth from Lidar Data, J HYDROMETEOROL, 7,285–297, https://doi.org/10.1175/JHM487.1, 2006.

735 Deems, J. S., Painter, T. H., and Finnegan, D. C.: Lidar measurement of snow depth: A review, J GLACIOL, 59, 467–479, https://doi.org/10.3189/2013JoG12J154, 2013.

Deschamps-Berger, C., Gascoin, S., Shean, D., Besso, H., Guiot, A., and López-Moreno, J. I.: Evaluation of snow depth retrievals from ICESat-2 using airborne laser-scanning data, The Cryosphere, 17, 2779–2792, https://doi.org/10.5194/tc-17-2779-2023, 2023.

740   Efron, B. and Tibshirani, R.: Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy, STAT SCI, 1, 77–77, https://doi.org/10.1214/ss/1177013817, 1986.

Elder, K., Dozier, J., and Michaelsen, J.: Snow accumulation and distribution in an Alpine Watershed, WATER RESOUR RES, 27, 1541–1552, https://doi.org/10.1029/91WR00506, 1991.

Elder, K., Rosenthal, W., and Davis, R. E.: Estimating the spatial distribution of snow water equivalence in a montane
745   watershed, HYDROL PROCESS, 12, 1793–1808, https://doi.org/10.1002/(SICI)1099-1085(199808/09)12:10/11<1793::AID-HYP695>3.0.CO;2-K, 1998.

Essery, R., Morin, S., Lejeune, Y., and Ménard, C. B.: A comparison of 1701 snow models using observations from an alpine site, ADV WATER RESOUR, 55, 131–148, https://doi.org/10.1016/j.advwatres.2012.07.013, 2013.

Fassnacht, S. R., Heun, C. M., López-Moreno, J., and Latron, J.: Snow Density Variability in the Rio Esera Valley , Pyrenees
750   Mountains , 2. Study Site, Cuadernos de iIvestigación Geográfica, 36, 59–72, 2010.

Goh, A.: Back-propagation neural networks for modeling complex systems, ARTIF INTELL ENG, 9, 143–151, https://doi.org/10.1016/0954-1810(94)00011-S, 1995.

Griessinger, N., Mohr, F., and Jonas, T.: Measuring snow ablation rates in alpine terrain with a mobile multioffset ground-penetrating radar system, HYDROL PROCESS, 32, 3272–3282, https://doi.org/10.1002/hyp.13259, 2018.

755   Hapfelmeier, A., Hothorn, T., Ulm, K., and Strobl, C.: A new variable importance measure for random forests with missing data, STAT COMPUT, 24, 21–34, https://doi.org/10.1007/s11222-012-9349-1, 2014.

Hedrick, A. R., Marks, D., Havens, S., Robertson, M., Johnson, M., Sandusky, M., Marshall, H., Kormos, P. R., Bormann, K. J., and Painter, T. H.: Direct Insertion of NASA Airborne Snow Observatory-Derived Snow Depth Time Series Into the iSnobal Energy Balance Snow Model, WATER RESOUR RES, 54, 8045–8063,
760   https://doi.org/10.1029/2018WR023190, 2018.

Hiemstra, C., Marshall, H., Vuyovich, C., Elder, K., Mason, M., and Durand, M.: SnowEx20 Community Snow Depth Probe Measurements, Version 1, NASA National Snow and Ice Data Center Distributed Active Archive Center https://doi.org/10.5067/9IA978JIACAR, 2020

Hiemstra, C. A., Vuyovich, C. M., and Marshall, H.-P.: SnowEx20 Grand Mesa Reference GIS Data Sets, Version 1., NASA
765   National Snow and Ice Data Center Distributed Active Archive Center https://doi.org/10.5067/YDZXY4Q79VIJ, 2021.

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

770   Hojatimalekshah, A., Uhlmann, Z., Glenn, N. F., Hiemstra, C. A., Tennant, C. J., Graham, J. D., Spaete, L., Gelvin, A., Marshall, H.-P., McNamara, J. P., and Enterkine, J.: Tree canopy and snow depth relationships at fine scales with terrestrial laser scanning, The Cryosphere, 15, 2187–2209, https://doi.org/10.5194/tc-15-2187-2021, 2021.

Homer, C., Dewitz, J., Jin, S., Xian, G., Costello, C., Danielson, P., Gass, L., Funk, M., Wickham, J., Stehman, S., Auch, R.,

and Riitters, K.: Conterminous United States land cover change patterns 2001–2016 from the 2016 National Land Cover Database, ISPRS J PHOTOGRAMM, 162, 184–199, https://doi.org/10.1016/j.isprsjprs.2020.02.019, 2020

Houser, P., Rudisill, W., Johnston, J., Elder, K., Marshall, H., Vuyovich, C., Kim, E. J., and Mason, M.: SnowEx Meteorological Station Measurements from Grand Mesa, CO, Version 1, https://doi.org/10.5067/497NQVJ0CBEX, 2022.

Hu, X., Hao, X., Wang, J., Huang, G., Li, H., and Yang, Q.: Can the Depth of Seasonal Snow be Estimated From ICESat-2 Products: A Case Investigation in Altay, Northwest China, IEEE GEOSCI REMOTE S, pp. 1–5, https://doi.org/10.1109/LGRS.2021.3078805, 2021.

Isaaks, E. H. and Srivastava, R. M.: Applied Geostatistics, Oxford University Press, ISBN 9780195050134, 1989.

Jain, A., Mao, J., and Mohiuddin, K.: Artificial neural networks: a tutorial, Computer, 29, 31–44, https://doi.org/10.1109/2.485891, 1996.

Jonas, T., Marty, C., and Magnusson, J.: Estimating the snow water equivalent from snow depth measurements in the Swiss Alps, J HYDROL, 378, 161–167, https://doi.org/10.1016/j.jhydrol.2009.09.021, 2009.

Kahaner, D., Moler, C., & Nash, S. (1989). Numerical methods and software. *Englewood Cliffs: Prentice Hall, 1989*.

Kim, J. H., Cho, S. J., and Yi, M. J.: Removal of ringing noise in GPR data by signal processing, GEOSCI J, 11, 75–81, https://doi.org/10.1007/BF02910382, 2007.

Kuwahara, M., Hachimura, K., Eiho, S., and Kinoshita, M.: Processing of RI-Angiocardiographic Images, pp. 187–202, Springer US, https://doi.org/10.1007/978-1-4684-0769-3_13, 1976.

Lague, D., Brodu, N., and Leroux, J.: Accurate 3D comparison of complex topography with terrestrial laser scanner: Application to the Rangitikei canyon (NZ), ISPRS J PHOTOGRAMM, 82, 10–26, https://doi.org/10.1016/j.isprsjprs.2013.04.009, 2013.

Lettenmaier, D. P., Alsdorf, D., Dozier, J., Huffman, G. J., Pan, M., and Wood, E. F.: Inroads of remote sensing into hydrologic science during the WRR era, WATER RESOUR RES, 51, 7309–7342, https://doi.org/10.1002/2015WR017616, 2015.

Li, L. and Pomeroy, J. W.: Estimates of Threshold Wind Speeds for Snow Transport Using Meteorological Data, J APPL METEOROL, 36, 205–213, https://doi.org/10.1175/1520-0450(1997)036<0205:EOTWSF>2.0.CO;2, 1997.

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., and Lannoy, G. J. M. D.: Snow depth variability in the Northern Hemisphere mountains observed from space, NAT COMMUN, 10, 4629, https://doi.org/10.1038/s41467-019-12566-y, 2019.

Lievens, H., Brangers, I., Marshall, H.-P., Jonas, T., Olefs, M., and Lannoy, G. D.: Sentinel-1 snow depth retrieval at sub-kilometer resolution over the European Alps, The Cryosphere, 16, 159–177, https://doi.org/10.5194/tc-16-159-2022, 2022.

Lukas, V. and Baez, V.: 3D Elevation Program—Federal best practices: U.S. Geological Survey Fact Sheet 2020–3062, https://doi.org/https://doi.org/10.3133/fs20203062, 2021.

Lv, Z. and Pomeroy, J. W.: Assimilating snow observations to snow interception process simulations, HYDROL PROCESS, 34, 2229–2246, https://doi.org/10.1002/hyp.13720, 2020.

810 López-Moreno, J. I., Fassnacht, S. R., Heath, J. T., Musselman, K. N., Revuelto, J., Latron, J., Morán-Tejeda, E., and Jonas, T.: Small scale spatial variability of snow density and depth over complex alpine terrain: Implications for estimating snow water equivalent, ADV WATER RESOUR, 55, 40–52, https://doi.org/10.1016/j.advwatres.2012.08.010, 2013.

Marks, D., Dozier, J., and Davis, R. E.: Climate and energy exchange at the snow surface in the Alpine Region of the Sierra Nevada: 1. Meteorological measurements and monitoring, WATER RESOUR RES, 28, 3029–3042, 815 https://doi.org/10.1029/92WR01482, 1992.

Marshall, H., Vuyovich, C., Hiemstra, C., Brucker, L., Elder, K., Deems, J., Newlin, J., Bales, R., Nolin, A., and Trujillo, E.: NASA SnowEx 2020 Experiment Plan, pp. 1–100, https://snow.nasa.gov/campaigns/snowex/experimental-plan-2020, 2019.

Marshall, H. P., Koh, G., Sturm, M., Johnson, J. B., Demuth, M., Landry, C., Deems, J. S., and Gleason, J. A.: Spatial 820 variability of the snowpack: Experiences with measurements at a wide range of length scales with several different high precision instruments, in: Proceedings ISSW 2006. International Snow Science Workshop, Telluride CO, U.S.A., 1-6 October 2006, pp. 359–364, http://arc.lib.montana.edu/snow-science/item/947, 2006.

Marti, R., Gascoin, S., Berthier, E., de Pinel, M., Houet, T., and Laffly, D.: Mapping snow depth in open alpine terrain from stereo satellite imagery, The Cryosphere, 10, 1361–1380, https://doi.org/10.5194/tc-10-1361-2016, 2016

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

McGrath, D., Webb, R., Shean, D., Bonnell, R., Marshall, H. P., Painter, T. H., Molotch, N. P., Elder, K., Hiemstra, C., and Brucker, L.: Spatially Extensive Ground-Penetrating Radar Snow Depth Observations During NASA's 2017 SnowEx Campaign: Comparison With In Situ, Airborne, and Satellite Observations, WATER RESOUR RES, 55, 10 026–10 036, 830 https://doi.org/10.1029/2019WR024907, 2019.

McGrath, D., Bonnell, R., Zeller, L., Olsen-Mikitowicz, A., Bump, E., Webb, R., and Marshall, H.-P.: A Time Series of Snow Density and Snow Water Equivalent Observations Derived From the Integration of GPR and UAV SfM Observations, Frontiers in Remote Sensing, 3, 1–15, https://doi.org/10.3389/frsen.2022.886747, 2022.

Meehan, T. G.: SnowEx20 Grand Mesa IOP BSU 1 GHz Multi-polarization GPR, Version 1, NASA National Snow and Ice 835 Data Center Distributed Active Archive Center, https://doi.org/https://doi.org/10.5067/Q2LFK0QSVGS2, 2021.

Meehan, T. G., Marshall, H. P., Bradford, J. H., Hawley, R. L., Overly, T. B., Lewis, G., Graeter, K., Osterberg, E., and McCarthy, F.: Reconstruction of historical surface mass balance, 1984–2017 from GreenTrACS multi-offset ground-penetrating radar, J GLACIOL, 67, 219–228, https://doi.org/10.1017/jog.2020.91, 2021.

Mote, P. W., Li, S., Lettenmaier, D. P., Xiao, M., and Engel, R.: Dramatic declines in snowpack in the western US, NPJ CLIM 840 ATMOS SCI, 1, https://doi.org/10.1038/s41612-018-0012-1, 2018.

Meløysund, V., Leira, B., Høiseth, K. V., and Lisø, K. R.: Predicting snow density using meteorological data, METEOROL

APPL, 14, 413–423, https://doi.org/10.1002/met.40, 2007.

Neidell, N. S. and Taner, M. T.: Semblance and Other Coherency Measrues for Multichannel Data, Geophysics, 36, 482–497,
845 https://doi.org/10.1190/1.1440186, 1971.

NOAA: VDatum 4.3 Vertical Datum Transformation, https://vdatum.noaa.gov/, 2021.

National Academies of Sciences Engineering and Medicine: Thriving on Our Changing Planet: A Decadal Strategy for Earth
Observation from Space, National Academies Press, ISBN 978-0-309-46757-5, https://doi.org/10.17226/24938, 2018.

Painter, T. H. and Bormann, K. J.: ASO L4 Lidar Point Cloud Digital Terrain Model 3m UTM Grid, Version 1,
850 NASA National Snow and Ice Data Center Distributed Active Archive Center.
https://doi.org/10.5067/2EHMWG4IT76O, 2020.

Painter, T. H., Berisford, D. F., Boardman, J. W., Bormann, K. J., Deems, J. S., Gehrke, F., Hedrick, A., Joyce, M., Laidlaw,
R., Marks,D., Mattmann, C., McGurk, B., Ramirez, P., Richardson, M., Skiles, S. M. K., Seidel, F. C., and Winstral, A.:
The Airborne Snow Observatory: Fusion of scanning lidar, imaging spectrometer, and physically-based modeling for
855 mapping snow water equivalent and snow albedo, REMOTE SENS ENVIRON, 184, 139–152,
https://doi.org/10.1016/j.rse.2016.06.018, 2016.

Pierce, D. W., Barnett, T. P., Hidalgo, H. G., Das, T., Bonfils, C., Santer, B. D., Bala, G., Dettinger, M. D., Cayan, D. R.,
Mirin, A., Wood, A. W., and Nozawa, T.: Attribution of declining Western U.S. Snowpack to human effects, J
CLIMATE, 21, 6425–6444, https://doi.org/10.1175/2008JCLI2405.1, 2008.

860 Raleigh, M. S. and Small, E. E.: Snowpack density modeling is the primary source of uncertainty when mapping basin-wide
SWE with lidar, GEOPHYS RES LETT, 44, 3700–3709, https://doi.org/10.1002/2016GL071999, 2017.

Rovansek, R. J., Kane, D. L., and Hinzman, L. D.: Improving estimates of snowpack water equivalent using double sampling,
in: Proceedings 61st Western Snow Conference, pp. 157-163, 1993.

Siirila-Woodburn, E. R., Rhoades, A. M., Hatchett, B. J., Huning, L. S., Szinai, J., Tague, C., Nico, P. S., Feldman, D. R.,
865 Jones, A. D., Collins, W. D., and Kaatz, L.: A low-to-no snow future and its impacts on water resources in the western
United States, NAT REV EARTH ENVIRON, 2, 800–819, https://doi.org/10.1038/s43017-021-00219-y, 2021.

Singh, S., Durand, M., Kim, E., Pan, J., Kang, D. H., and Barros, A. P.: A Physical-Statistical Retrieval Framework to Estimate
SWE from X and Ku-Band SAR Observations, vol. 2023-July, pp. 17–20, IEEE., ISBN 9798350320107,
https://doi.org/10.1109/IGARSS52108.2023.10281838, 2023.

870 Sturm, M. and Holmgren, J.: Differences in compaction behavior of three climate classes of snow, ANN GLACIOL, 26, 125
130, https://doi.org/10.3189/1998AoG26-1-125-130, 1998.

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

Tedesco, M., Reichle, R., Low, A., Markus, T., and Foster, J. L.: Dynamic Approaches for Snow Depth Retrieval From
875 Spaceborne Microwave Brightness Temperature, IEEE T GEOSCI REMOTE, 48, 1955–1967,

https://doi.org/10.1109/TGRS.2009.2036910, 2010.

Tiuri, M. E., Sihvola, A. H., Nyfors, E. G., and Hallikaiken, M. T.: The Complex Dielectric Constant of Snow at Microwave Frequencies, IEEE J OCEANIC ENG, 9, 377–382, https://doi.org/10.1109/JOE.1984.1145645, 1984.

Treichler, D. and Kääb, A.: Snow depth from ICESat laser altimetry — A test study in southern Norway, REMOTE SENS ENVIRON, 191, 389–401, https://doi.org/10.1016/j.rse.2017.01.022, 2017.

Trujillo, E., Ramírez, J. A., and Elder, K. J.: Scaling properties and spatial organization of snow depth fields in sub-alpine forest and alpine tundra, HYDROL PROCESS, 23, 1575–1590, https://doi.org/10.1002/hyp.7270, 2009.

Tsang, L., Durand, M., Derksen, 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., and Xu, X.: Review article: Global monitoring of snow water equivalent using high-frequency radar remote sensing, The Cryosphere, 16, 3531–3573, https://doi.org/10.5194/tc-16-3531-2022, 2022.

Valence, E., Baraer, M., Rosa, E., Barbecot, F., and Monty, C.: Drone-based ground-penetrating radar (GPR) application to snow hydrology, Cryosphere, 16, 3843–3860, https://doi.org/10.5194/tc-16-3843-2022, 2022.

Vecherin, S., Meyer, A., Quinn, B., Letcher, T., and Parker, M.: Simulation of Snow Texture for Autonomous Vehicle Numerical Modeling, National Defense Industrial Association, http://gvsets.ndia-mich.org/publication.php?documentID=928, 2022.

Vuyovich, C. M., Marshall, H., Elder, K., Hiemstra, C., Brucker, L., and McCormick, M.: SnowEx20 Grand Mesa Intensive Observation Period Snow Pit Measurements, Version 1., NASA National Snow and Ice Data Center Distributed Active Archive Center. https://doi.org/10.5067/DUD2VZEVBJ7S, 2021.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P.: Image quality assessment: From error visibility to structural similarity, IEEE T IMAGE PROCESS, 13, 600–612, https://doi.org/10.1109/TIP.2003.819861, 2004.

Webb, R. W.: SnowEx20 Grand Mesa IOP UNM 800 and 1600 MHz MALA GPR, Version 1, NASA National Snow and Ice Data Center Distributed Active Archive Center https://doi.org/10.5067/WE9GI1GVMQF6, 2021.

Webb, R. W., Raleigh, M. S., McGrath, D., Molotch, N. P., Elder, K., Hiemstra, C., Brucker, L., and Marshall, H. P.: Within-Stand Boundary Effects on Snow Water Equivalent Distribution in Forested Areas, WATER RESOUR RES, 56, https://doi.org/10.1029/2019WR024905, 2020.

Webb, R. W., Marziliano, A., McGrath, D., Bonnell, R., Meehan, T. G., Vuyovich, C., and Marshall, H.-P.: In Situ Determination of Dry and Wet Snow Permittivity: Improving Equations for Low Frequency Radar Applications, REMOTE SENS, 13, 4617, https://doi.org/10.3390/rs13224617, 2021.

Wetlaufer, K., Hendrikx, J., and Marshall, L.: Spatial heterogeneity of snow density and its influence on snow water equivalence estimates in a large mountainous basin, Hydrology, 3, https://doi.org/10.3390/hydrology3010003, 2016.

Wharton, R. P., Hazen, G. A., Rau, R. N., and Best, D. L.: Advancements In Electromagnetic Propagation Logging, in: Proceedings Society of Petroleum Engineers Rocky Mountain Regional Meeting, Society of Petroleum Engineers,

910 https://doi.org/10.2118/9041-MS, 1980.

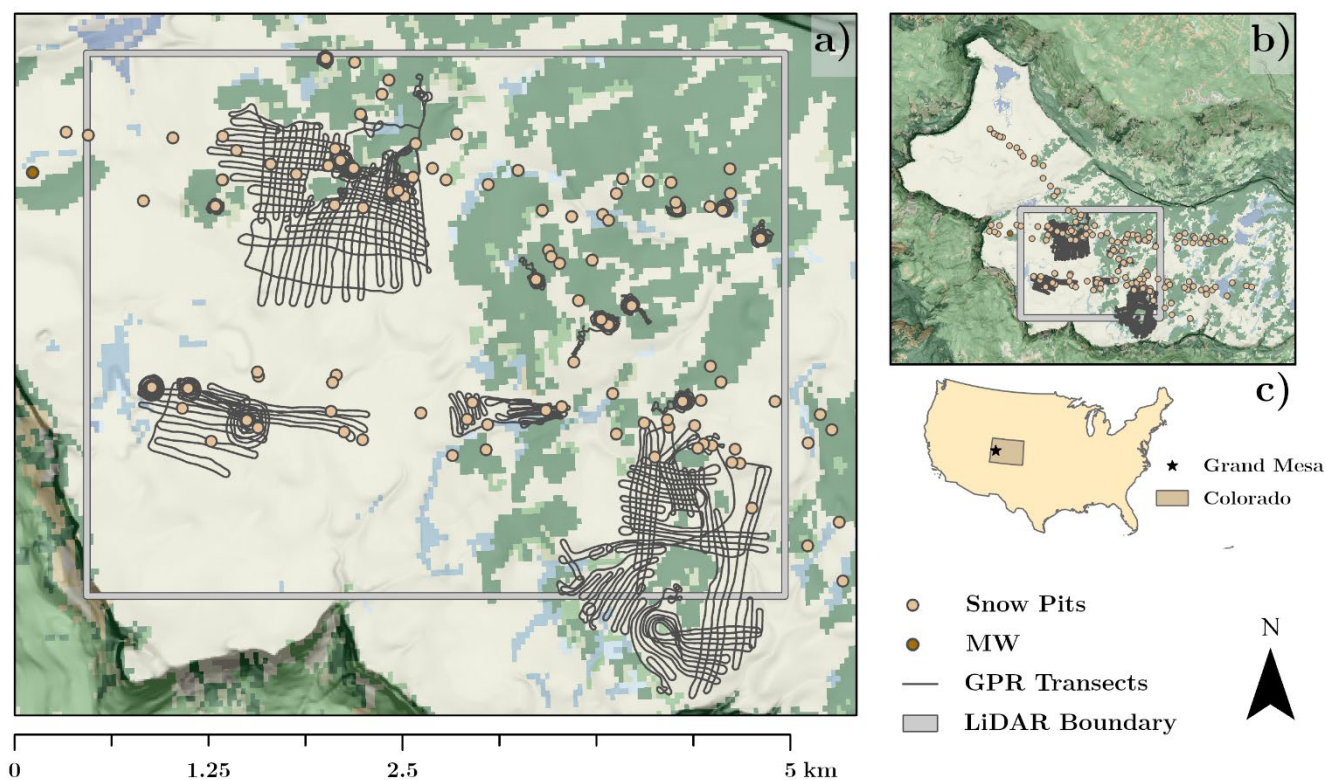
Winstral, A., Elder, K., and Davis, R. E.: Spatial Snow Modeling of Wind-Redistributed Snow Using Terrain-Based Parameters, Journal of Hydrometeorology, 3, 524–538, https://doi.org/10.1175/1525-7541(2002)003<0524:SSMOWR>2.0.CO;2, 2002.

Wong, J., Han, L., Bancroft, J. C., and Stewart, R. R.: Automatic time-picking of first arrivals on noisy microseismic data, in: Proceedings Canadian Society of Exploration Geophysics Meeting, pp. 1–6, 2009.

915

Yildiz, S., Akyurek, Z., and Binley, A.: Quantifying snow water equivalent using terrestrial ground penetrating radar and unmanned aerial vehicle photogrammetry, HYDROL PROCESS, 35, 1–15, https://doi.org/10.1002/hyp.14190, 2021.

Öz Yilmaz: Seismic Data Analysis, Society of Exploration Geophysicists, ISBN 978-1-56080-094-1, https://doi.org/10.1190/1.9781560801580, 2001.

920 Štroner, M., Urban, R., Lidmila, M., Kolář, V., and Křemen, T.: Vegetation Filtering of a Steep Rugged Terrain: The Performance of Standard Algorithms and a Newly Proposed Workflow on an Example of a Railway Ledge, REMOTE SENS, 13, 3050, https://doi.org/10.3390/rs13153050, 2021.

925 **Figure 1: a) Study area map of snow pit locations (yellow circles), Mesa West weather station (brown circle), GPR transects (black lines), and LiDAR boundary (grey wireframe). Landcover classification identifies the forested areas as green and lakes as blue. b) inset map of Grand Mesa, Colorado depicting the extent of the dataset acquired during the NASA SnowEx 2020 Intensive**

Observation Period at Grand Mesa, Colorado (Hiemstra et al., 2021). c) inset map of the contiguous U.S. which identifies the location of Grand Mesa, Colorado. Land cover classification data were accessed from the 2016 National Land Cover Database (Homer et al., 2020). Slope hillshade data were accessed from the USGS 3D Elevation Program (Lukas & Baez, 2021). Cartographic boundary files were accessed from the Census Bureau's MAF/TIGER geographic database (U.S. Census Bureau, 2020). The geographic coordinate projection of these maps is UTM Zone 12 N; EPSG code 32612.

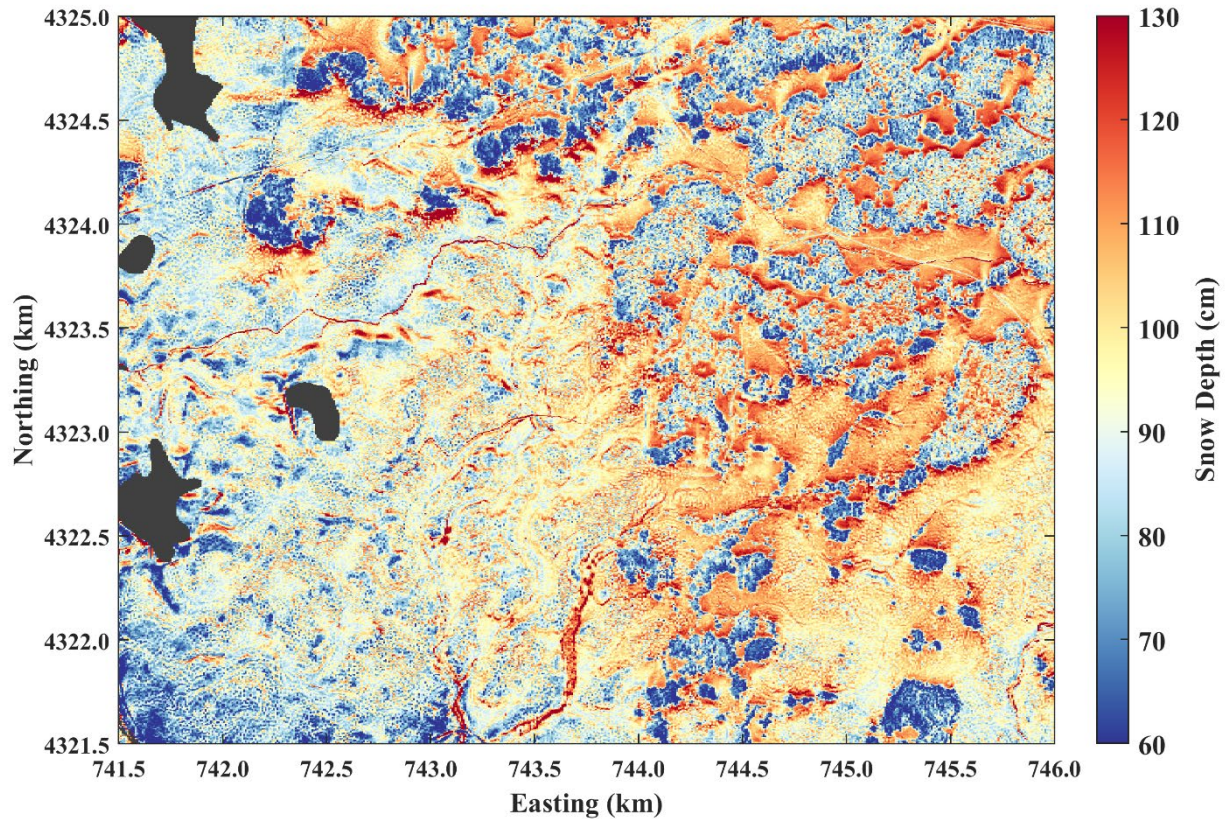



Figure 2: One-metre resolution snow depths from the February 1, 2020 LiDAR flight. The western half of the domain is relatively unforested, shrub steppe (lakes are masked black) while the eastern half has stands of dense forest (see Fig. 1). The colourmap is centred on the mean value.

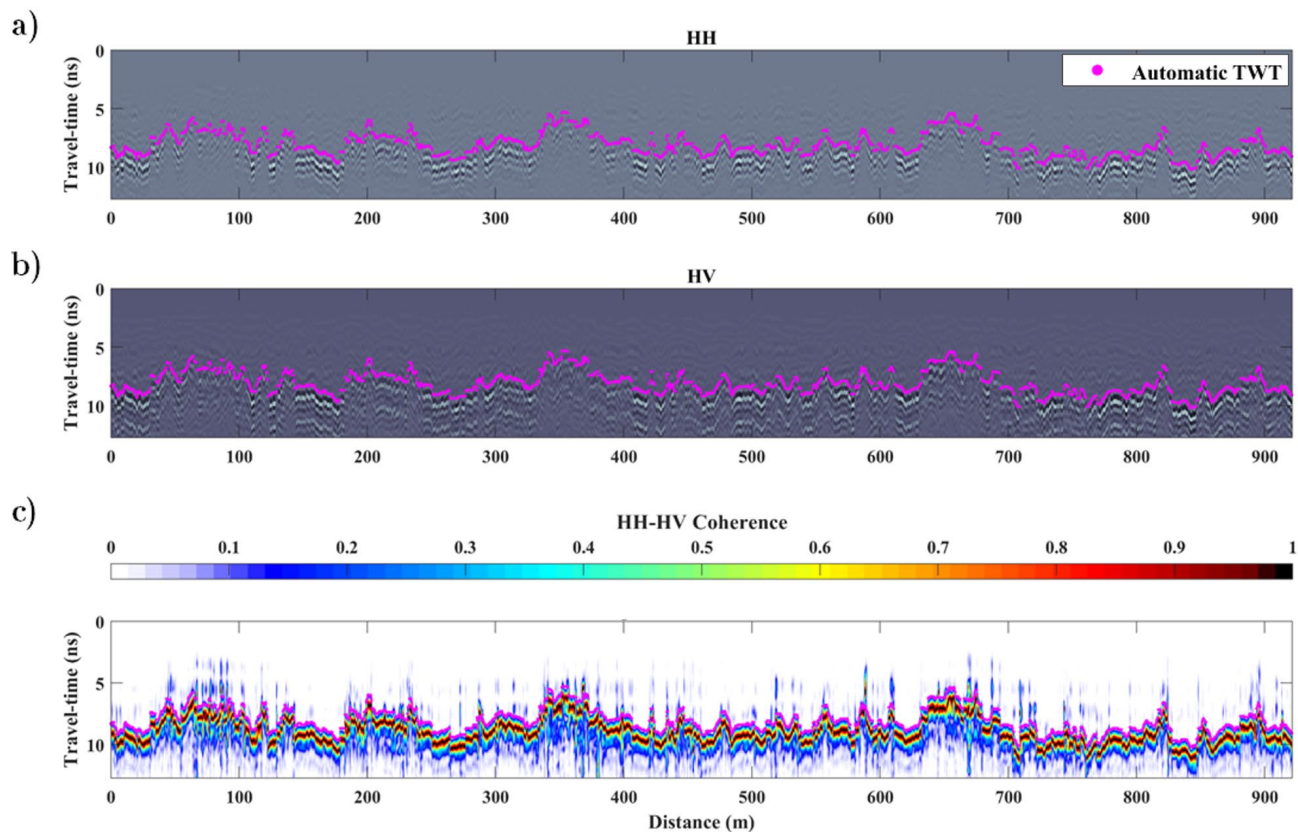



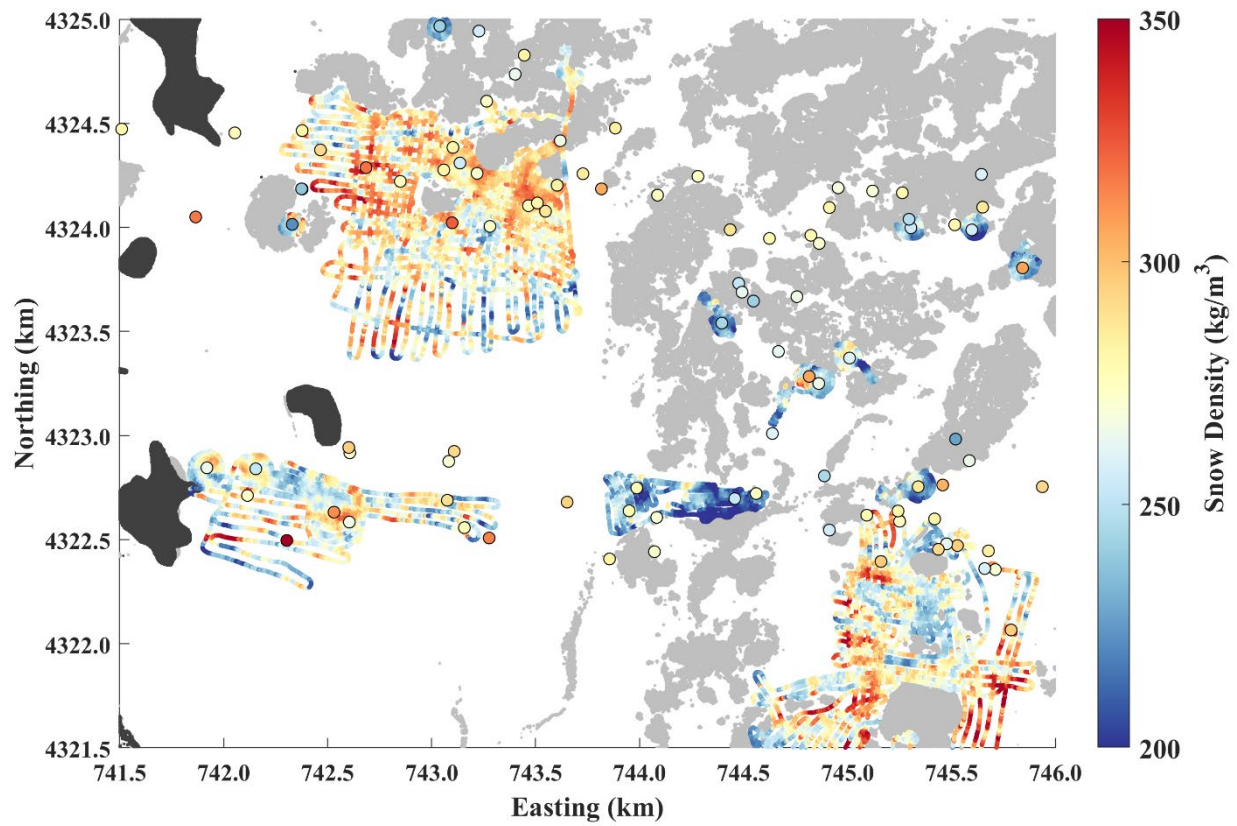
Figure 3: A 900 m GPR transect with autopicks in magenta for a) HH and b) HV profiles of travel time, and c) the coherence of these radargrams (Equations 1, 2).

940

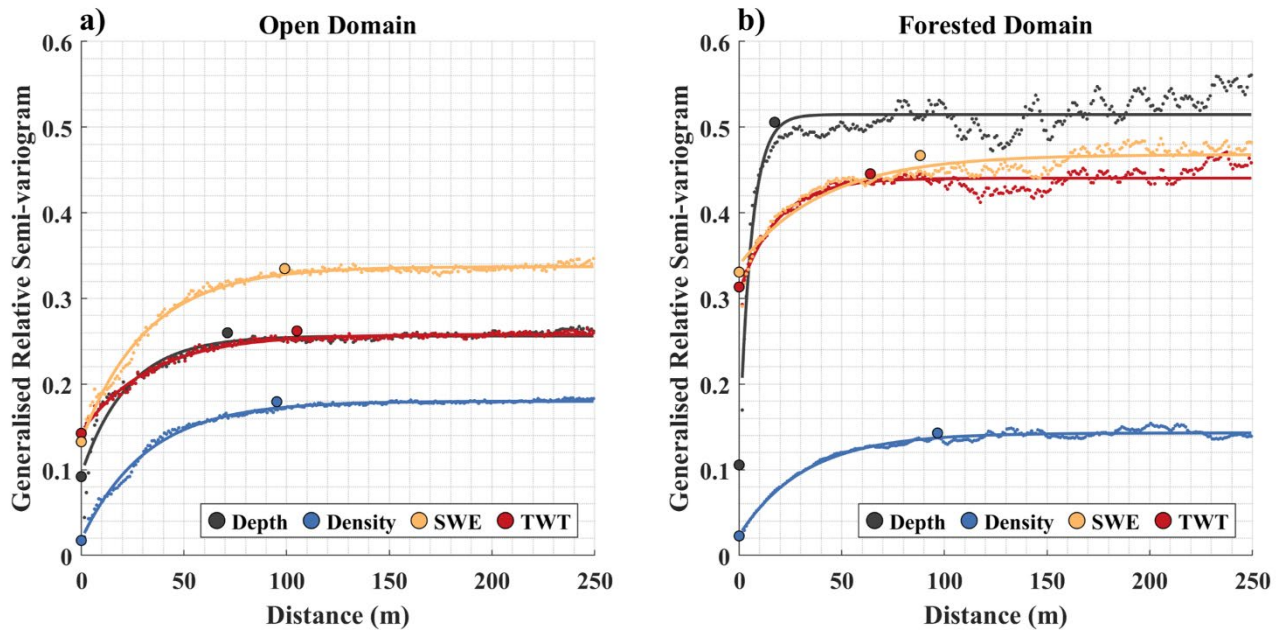
Figure 4: Bulk snow density along radar profiles estimated by combining LiDAR snow depths with GPR TWTs. Average density measured in the 96 snow pits within the LiDAR boundary are overlaid as larger makers. Forested areas (grey) and lakes (black) are shown. The colourmap is centred on the mean value.

945

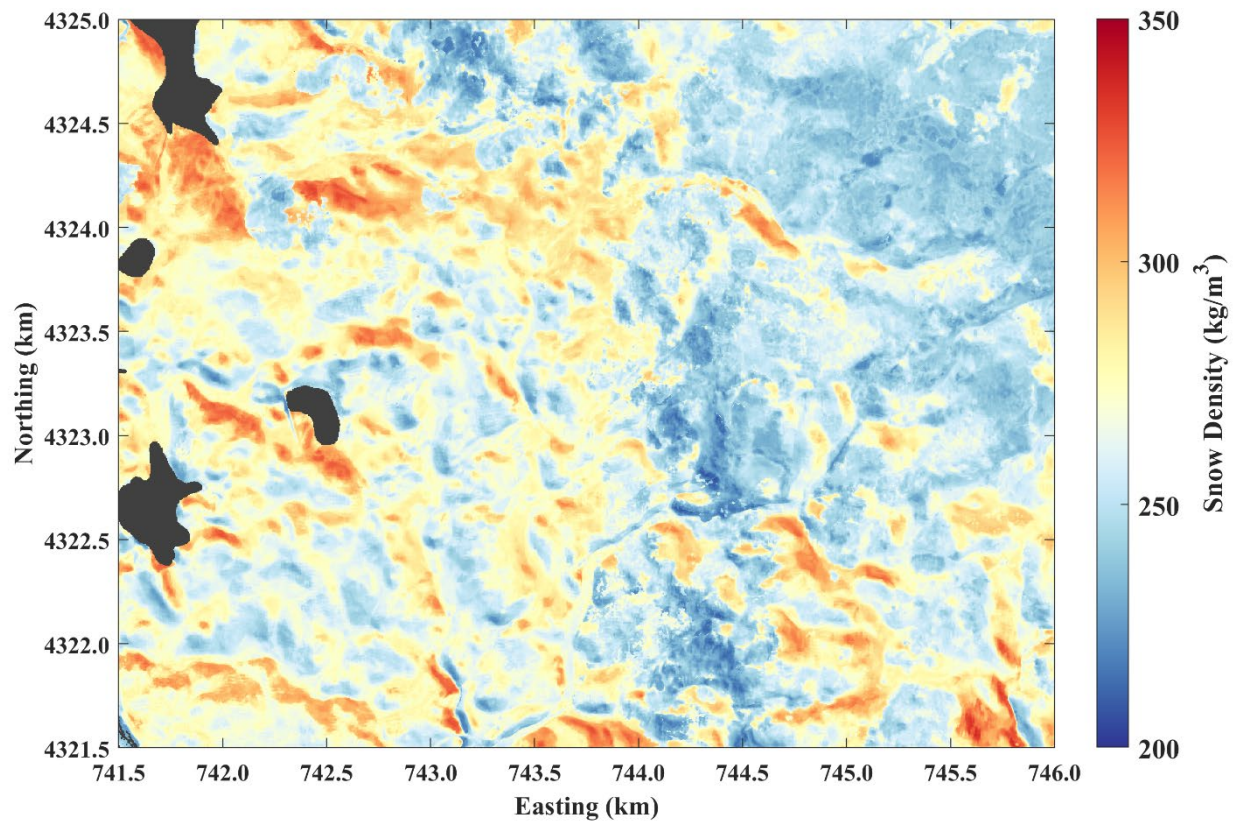
Figure 5: Generalised relative semi-variograms in a) open and b) forested areas for LiDAR snow depth, GPR TWT, average density retrieved along the GPR transects, and resulting SWE. Experimental variograms were fitted with an exponential model to determine the variogram parameters. The larger markers represent the average nugget, sill, and correlation length estimated by Monte Carlo subsampling.

950

Figure 6: Snow density distributed spatially using the ML regression ensemble average ($b_{s,LiDAR-\overline{Ens}}$). The colourmap is centred on the mean value.

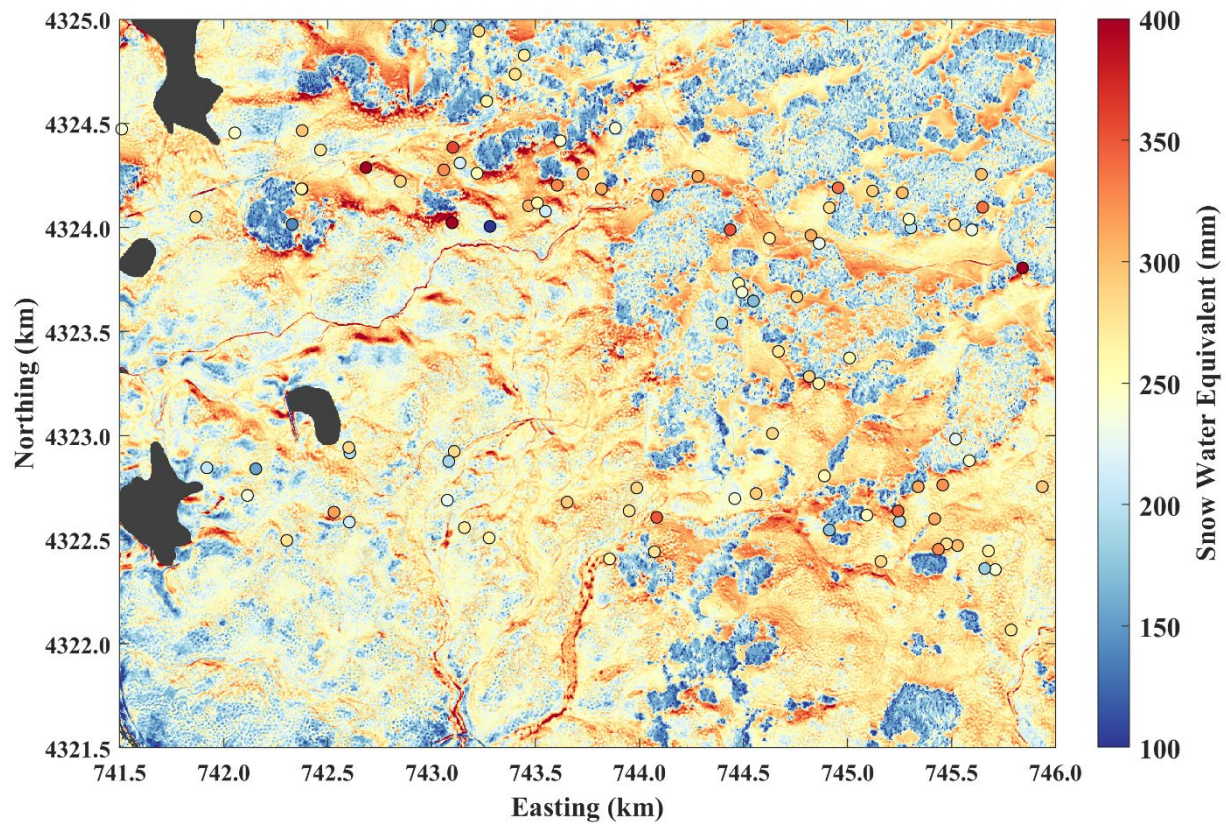



Figure 7: Spatially distributed snow water equivalent estimated using the regression ensemble mean density ($b_{s,LiDAR-\overline{Ens}}$) and LiDAR snow depths ($H_{s,LiDAR}$). Forests and wind scoured areas tend to have lower SWE, and forest perimeters have higher SWE. The metre-scale stippled texture is the result of low-stature vegetation ($H_{veg} < 0.5$) and boulders, which both reduce snow depth and decrease the snow density. Large markers are SWE values measured at snow pits. Lakes are masked in black. The colourmap is centred on the mean value.

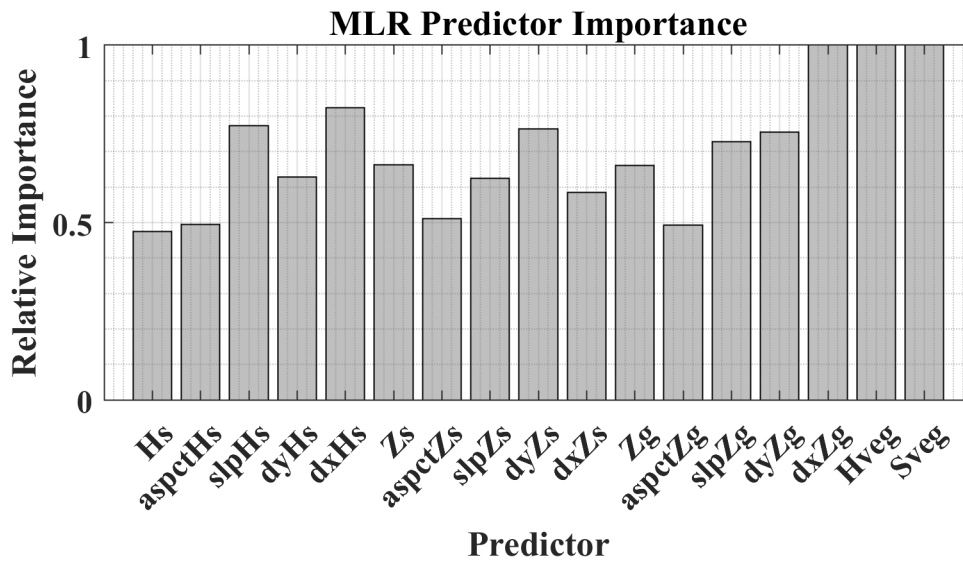



Figure A.1: The relative importance of the LiDAR derived predictors: Hs snow depth, aspctHs aspect of snow depth, slpHs slope of snow depth, dyHs north component of snow depth gradient, dxHs east component of snow depth gradient, Zs the snow surface elevation and derivatives, Zg the ground elevation and derivatives, Hveg vegetation height, and Sveg the distance to vegetation with height greater than 0.5 m. The predictor importance was determined from the top 1 % of models.

965

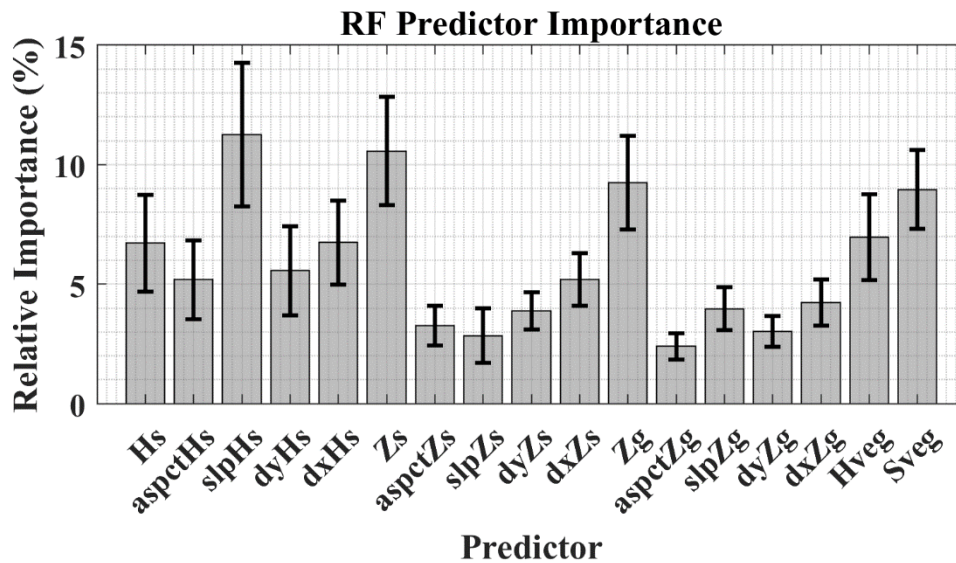



Figure A.2: Relative importance of LiDAR predictors calculated using the "out of bag" technique for random forest regression. Uncertainties were developed using random subsets of training data in a Monte Carlo simulation.

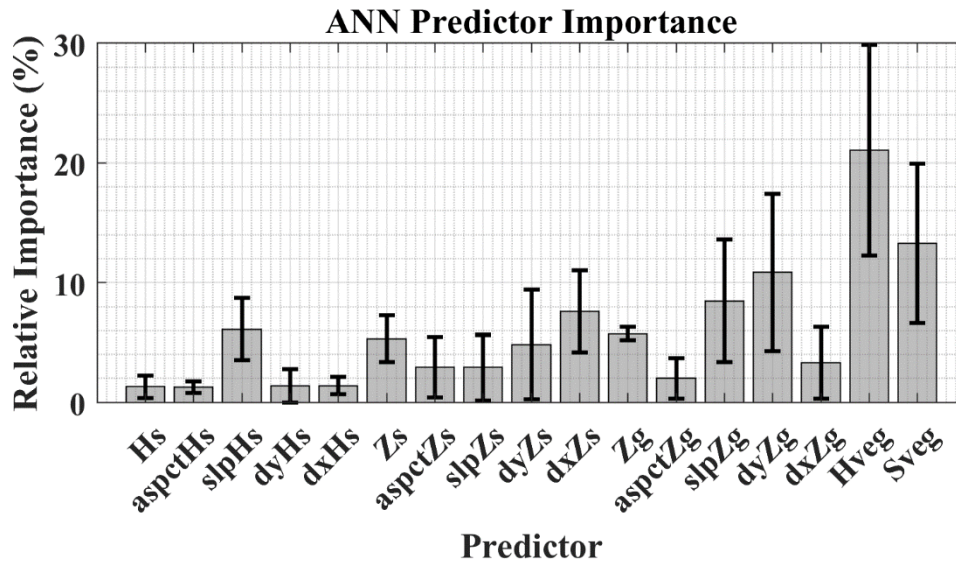



Figure A.3: Relative importance of LiDAR predictors within the ANN comprised of two hidden layers.

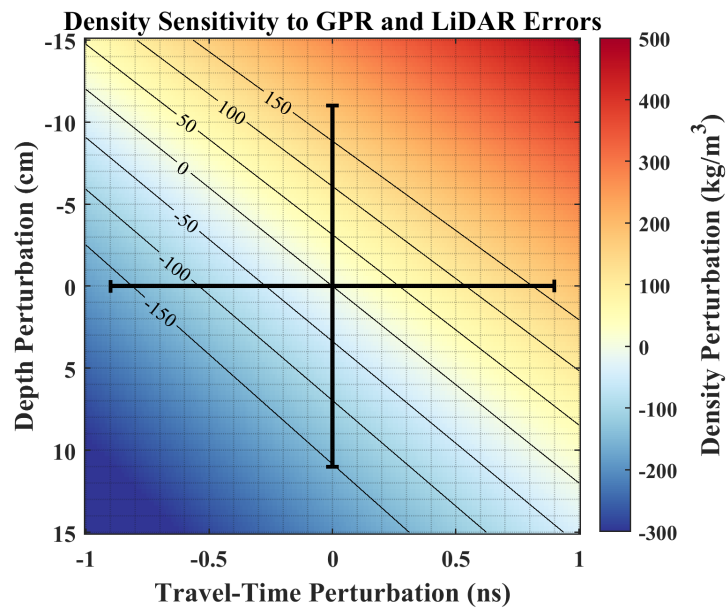



Figure B.1: Perturbations were added to the mean values 8 ns TWT and 96 cm depth then the density was evaluated to estimate potential errors resulting from sensor integration (coloured area and contours). The error bars represent the RMSE of LiDAR evaluated by probing (11 cm) and the RMSD of the GPR TWT crossovers (0.9 ns). At one standard deviation, combined snow density errors of ± 150 kg/m³ can be expected from sensor integration, which are reduced to within ± 30 kg/m³ by outlier filtering.

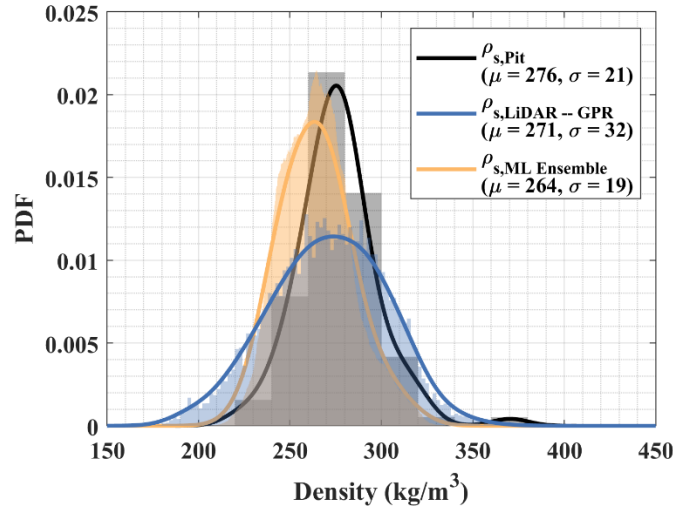



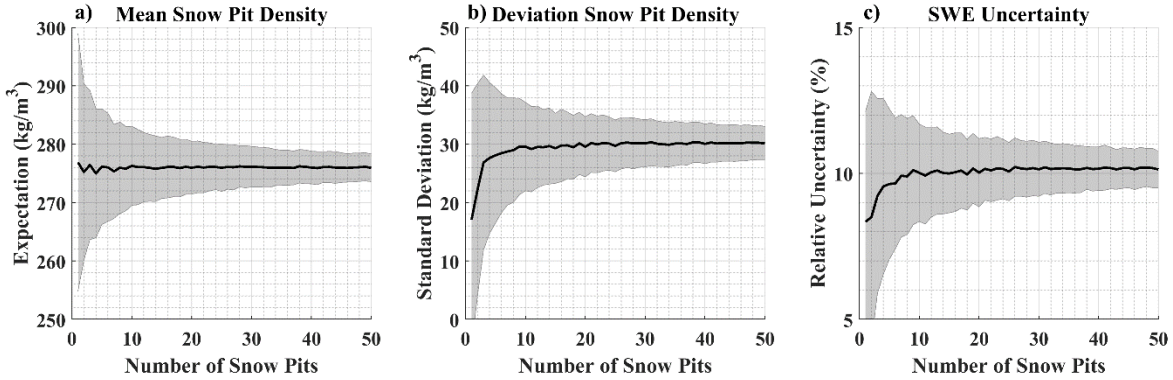
Figure B.2: Histograms of the snow pit measured, LiDAR – GPR retrieved, and regression model ensemble densities.

980   **Figure B.3: Monte Carlo uncertainty analysis for a) mean snow density, b) standard deviation of snow density, and c) propagated SWE uncertainty as functions of sample size.**

Table 1: a) The mean and standard deviation of snow pit and probe measured snow depths. b) Co-located estimates by LiDAR and GPR techniques gathered from within a 1 m radius of the in situ observations are compared to the in situ observations.

a)

Snow Depth (cm) Method	All Domain $\mu \pm \sigma$	Open Domain $\mu \pm \sigma$	Forested Domain $\mu \pm \sigma$
Snow Pit ($h_{s,Pit}$)	97 ± 17	99 ± 17	87 ± 13
Probe ($h_{s,Probe}$)	95 ± 17	98 ± 16	84 ± 18

985

b)

Snow Depth (cm) Method	All Domain $\mu \pm \sigma$ | R^2 | RMSE |Bias	Open Domain $\mu \pm \sigma$| R^2 | RMSE | Bias	Forested Domain $\mu \pm \sigma$ | R^2 | RMSE | Bias
LiDAR ($H_{s,LiDAR}$)	95 ± 16 | 0.61 | 11 | 0	99 ± 14 | 0.57 | 11 | 1	81 ± 16 | 0.60 | 12 | -4

GPR ($h_{s,\overline{Ens}\text{-GPR}}$)	97 ± 19 | 0.26 | 17 | 1	97 ± 19 | 0.26 | 17 | 1	99 ± 17 | 0.22 | 17 | -1

Table 2: a) Mean and standard deviation for snow pit, LiDAR – GPR (Sections 2.4.1 and 2.4.3), and modelled densities. Snow pit mean and standard deviations are estimated from all available data, N = 96 snow pits for the entire domain, N = 79 for the open domain, and N = 17 for the forested domain. b) Comparison between snow pit densities and estimated densities. Statistics (R^2, RMSE, and Bias) are measured from a subset of snow pits within 12.5 m of the GPR transects: N = 42 for all the domain, N = 36 for the open domain, and N = 6 for the forested domain. c) $\rho_{s,LiDAR-GPR}$ estimated densities are evaluated against modelled results.

a)

Snow Density (kg/m^3) Method	All Domain μ ± σ	Open Domain μ ± σ	Forested Domain μ ± σ
Snow Pit ($\rho_{s,Pit}$)	276 ± 21	280 ± 19	257 ± 20
LiDAR – GPR ($\rho_{s,LiDAR-GPR}$)	271 ± 32	273 ± 32	246 ± 19
MLR Model ($\rho_{s,MLR}$)	268 ± 21	273 ± 19	253 ± 18
RF Model ($\rho_{s,RF}$)	269 ± 17	271 ± 18	262 ± 11
ANN Model ($\rho_{s,ANN}$)	260 ± 25	266 ± 25	242 ± 14
ML Ensemble ($\rho_{s,\overline{Ens}}$)	264 ± 19	269 ± 19	248 ± 11
Random Field ($\rho_{s,Rand}$)	275 ± 20	275 ± 20	274 ± 20

b)

Snow Density ($\rho_{s,Pit}$)	All Domain R^2 | RMSE |Bias	Open Domain R^2 | RMSE | Bias	Forested Domain R^2 | RMSE | Bias
LiDAR – GPR ($\rho_{s,LiDAR-GPR}$)	0.02 | 35 | -5	0.01 | 37 | -8	0.39 | 23 | -13
MLR Model ($\rho_{s,MLR}$)	0.05 | 26 | -9	0.01 | 26 | -10	0.0 | 26 | -2
RF Model ($\rho_{s,RF}$)	0.01 | 26 | -8	0.0 | 27 | -10	0.0 | 22 | 4
ANN Model ($\rho_{s,ANN}$)	0.06 | 30 | -11	0.03 | 31 | -12	0.0 | 23 | -8
ML Ensemble ($\rho_{s,\overline{Ens}}$)	0.05 | 27 |-11	0.02 | 28 | -12	0.0 | 22 | -4
Random Field ($\rho_{s,Rand}$)	0.01 | 27 | -2	0.03 | 25 | -6	0.07 | 32 | 20

c)

Snow Density ($\rho_{s,LiDAR-GPR}$)	All Domain R^2 | RMSE |Bias	Open Domain R^2 | RMSE | Bias	Forested Domain R^2 | RMSE | Bias
MLR Model ($\rho_{s,MLR}$)	0.27 | 27 | 0	0.25 | 27 | 0	0.04 | 22 | 5
RF Model ($\rho_{s,RF}$)	0.80 | 15 | 0	0.79 | 15 | 0	0.76 | 10 | 1
ANN Model ($\rho_{s,ANN}$)	0.79 | 15 | 0	0.78 | 15 | 0	0.69 | 11 | 0
ML Ensemble ($\rho_{s,\overline{Ens}}$)	0.72 | 18 | 0	0.71 | 18 | 0	0.59 | 13 | 2
Random Field ($\rho_{s,Rand}$)	0.0 | 32 | 4	0.0 | 37 | 3	0.04 | 39 | 27

Table 3: a) Mean and standard deviation of snow water equivalent measured in snow pits and distributed by LiDAR snow depths using an average snow pit density value (276 kg/m^3), the average of snow pit density in each respective domain (276, 280, 257 kg/m^3), the regression ensemble densities, and the random field densities. In situ mean and standard deviations are estimated from all available data, N = 96 snow pits for the entire domain, N = 79 for the open domain, and N = 17 for the forested domain. b) Evaluation

of SWE between snow pit observations and distributed estimates using LiDAR snow depths multiplied by the average density measured from snow pits in the respective domains, the ML ensemble modelled densities, and the random field synthesized densities.

a)

Snow Water Equivalent (mm) Method	All Domain $\mu \pm \sigma$	Open Domain $\mu \pm \sigma$	Forested Domain $\mu \pm \sigma$
Snow Pit ($b_{s,Pit}$)	269 ± 57	278 ± 55	225 ± 45
Average Pit Density (276 kg/m^3)	255 ± 51	266 ± 41	219 ± 62
Domain Pit Avg. (276, 280 & 257)	255 ± 51	275 ± 37	197 ± 44
ML Ensemble ($b_{s,LiDAR-\overline{Ens}}$)	245 ± 53	259 ± 41	198 ± 57
Random Field ($b_{s,LiDAR-Rand}$)	254 ± 54	265 ± 45	218 ± 64

b)

Snow Water Equivalent (mm) Method	All Domain R^2 | RMSE | Bias	Open Domain R^2 | RMSE | Bias	Forested Domain R^2 | RMSE | Bias
Pit Average (276, 280 & 257)	0.57 | 38 | -9	0.63 | 34 | -3	0.16 | 55 | -29
ML Ensemble ($b_{s,LiDAR-\overline{Ens}}$)	0.56 | 46 | -20	0.61 | 42 | -18	0.19 | 56 | -27
Random Field ($b_{s,LiDAR-Rand}$)	0.49 | 42 | -10	0.52 | 39 | -10	0.10 | 55 | -14

Table A.1: Similarity matrix of R^2 values for a pixel-by-pixel intercomparison and SSIM values for a model intercomparison estimated over 100 m radius (approximate correlation length of snow density).

R^2, SSIM	MLR	RF	ANN	Random Field
MLR	1, 1	-	-	-
RF	0.49, 0.61	1, 1	-	-
ANN	0.31, 0.46	0.39, 0.52	1, 1	-
Random Field	0, 0	0, 0	0, 0	1, 1