

## **Letter of Response to Reviewers**

Dear Florent Dominé and Reviewers,

First, the authors would like to thank the reviewers for their constructive input that will greatly improve this manuscript. We have carefully mapped out a plan of action to address the comments. In this response letter, we consolidated the reviews into Primary and Editorial Concerns, where primary topics were requested by more than one reviewer. A summary of the comments is listed first in italics, followed by our plan to address the concern.

### **Primary Concerns**

*The primary concerns raised by the referees are that the manuscript is unfocused, has a wandering discussion, and is perhaps too rich in methods and analysis throughout.*

In an annotated draft of the preprint, we have color coded our suggested revisions to guide the editor quickly to the proposed changes. Paragraphs requiring simplification/condensing are coded **blue**, results in the methods section are **purple**, and methods in the results/discussion sections are **pink**. These will be moved into the appropriate sections as described in the comments attached to the highlighted text. Figures that are currently referenced in the methods section and convey results will be moved to the results.

The introduction will be significantly shortened to emphasize our work. Editorial emphasis will be placed on the final paragraphs to concisely provide the purpose of the work, downplaying the details of the approach and results. We will clearly state that the purpose of the manuscript is to present a method for efficient snow density estimates over broad spatial scales by integrating GPR and lidar. These observations are used to estimate density and SWE across the entire study domain using regression modeling and supported with evidence-based reasoning. This final point is one which lacks from the body of literature in recent and important works, such as Broxton et al. (2019).

The following Methods subsections will be relocated from the main body of the paper to the appendix or an additional Supplemental Material (at the discretion of the editor according to the format of the journal): 2.4.3.1 *Error analysis of LiDAR – GPR estimated density*; the majority of 2.4.4 *Wind and terrain exposure*. These subsections explain methods which are used to develop our understanding of the spatial data, spatial correlation, and wind processes affecting the observations and models, and are not necessary for many readers. However, information describing our approach of using wind processes as explanatory evidence will remain in the methods section to better frame a key intent of this work.

We will also shorten and focus the discussion, using sub-headings to guide and organize the manuscript. We propose the following discussion sections: deriving density and related uncertainties, distributing observations and related uncertainties, length scales of variability, and physically based explanations of spatial patterns. Presentation of uncertainties early in the discussion should help readers to gain confidence in the work before progressing to more interpretive sections. We can move considerations in regression model development, model transferability, and labor reduction to the Supplemental Material. The purpose-oriented topics – novelty in GPR processing, errors and uncertainties, comparable modeling studies, and explainable spatial patterns – will remain.

Spaceborne: *The link between snow density measurements at the field site and remote sensing is unclear*

We will emphasize the need for distributed density estimates to resolve SWE from space in 1-2 concise sentences. This will be linked to the premise of the SnowEx mission. In short, spatial density estimates are required to convert any snow depth remote sensing product (e.g. lidar or SfM) to SWE, and density is also a major control on snow retrievals from narrow-band SAR.

Point-Cloud Differencing: *Why develop Lidar-based snow depths using point cloud differencing, rather than using existing, raster differenced data products.*

We produced an in-house snow depth product, because to our knowledge no comparable publicly available data product exists. We have since received data privately from Airborne Snow Observatories and agree that to deliver confidence in our approach, it would be a valuable opportunity to compare to other aerial data products, if the ASO data will be public in the near future. We will clarify our approach and the M3C2 jargon.

Co-registration/Geolocation Uncertainty: *Clarify the magnitude of geolocation errors and increase the descriptiveness of the approach.*

Section 3.7 will be subdivided to include: 1. uncertainty analysis of the rasterized SWE product; 2. uncertainty of the derived snow density observations through the combination of radar and lidar information. Much of the reviewer's concerns in subsection 2.4.3.1 can be addressed through reorganization.

Snow Depth Validation: *Provide information and clarification to support data agreement/disagreement.*

A scatter plot which compares the in-situ and lidar derived snow depths in open and forested areas will be added to the appendix/supplement to convey the magnitude of errors which approach the range of snow depths within forested and open areas individually, causing a lack of correlation.

Snow Density Validation: *Data presentation seems contradictory.*

Weak correlation, yet low RMSE of snow density observations – similar to the snow depth validation – will be clarified with additional scatter plots in the appendix/supplement. Points of discussion regarding the uncertainty and known sources of error will be placed in the appropriate sections (i.e., removing results from the methods).

Correlation Length Scale for TWT and SWE: *Clarify the relationship between SWE and TWT, and how this is represented by length-scale analysis.*

The point that TWT and SWE are both dependent on snow depth and density will be clarified and organized into a subsection of the discussion.

## **Editorial Concerns**

Most editorial comments will be accepted and noted in the revision through tracked changes except for the few detailed below.

### **RC1**

Line 449-450: It is hard to reconcile how the errors in SWE found using this methodology at this field site relate to a global SWE uncertainty target.

*Developing uncertainties at this site, using methodologies which potentially represent a best-case-scenario for broad-scale SWE retrieval, suggest that achieving such a high standard may not be feasible for the global target.*

### **RC2**

L24 maybe keep depth and SWE with the same unit, cm?

*The units presented are standard.*

Fig 3, 4, 5, 7: Divergent color map (red, yellow, blue) is not really appropriated for variables which are always positive. It is traditionally used for variables centered on zero (yellow).

*The divergent colormap is useful in this presentation to show the clear differences in snow properties between open and forested areas.*

### **RC3**

Line 330: The Regression models were used to “fill-the-gaps” in the density field provided by the GPR. I’m curious as to what these maps look like.

*It is unclear what is being asked by this comment. The regression models are presented in Figure 7, while the Lidar—GPR derived densities are presented in Figure 5.*

The authors miss a fruitful opportunity to compare their SWE map with other spatially distributed estimates of SWE.

*ASO snow depth or SWE products are not within the public domain, such a comparison cannot yet be made openly. Data are available upon request, which the authors have received, and will agree to these comparisons, provided that the data are made public in a timely manner.*

## **References**

Broxton, P. D., Leeuwen, W. J. D., & Biederman, J. A. (2019). Improving Snow Water Equivalent Maps With Machine Learning of Snow Survey and Lidar Measurements. *Water Resources Research*, 55(5), 3739–3757. <https://doi.org/10.1029/2018WR024146>

With Kind Regards,

Tate G. Meehan (Corresponding Author)



## 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<sup>1,2</sup>, Ahmad Hojatimalekshah<sup>2</sup>, Hans-Peter Marshall<sup>2</sup>, Elias. J. Deeb<sup>1</sup>, Shad O’Neel<sup>1,2</sup>, Daniel McGrath<sup>3</sup>, Ryan W. Webb<sup>4</sup>, Randall Bonnelli<sup>3</sup>, Mark S. Raleigh<sup>5</sup>, Christopher Hiemstra<sup>6</sup>, Kelly Elder<sup>7</sup>

<sup>1</sup> Cold Regions Research and Engineering Laboratory; U.S. Army Corps of Engineers; Hanover, NH, USA

<sup>2</sup> Department of Geosciences; Boise State University; Boise, ID, USA

<sup>3</sup> Department of Geosciences; Colorado State University; Fort Collins, CO, USA

<sup>4</sup> Department of Civil and Architectural Engineering & Construction Management; University of Wyoming; Laramie, WY, USA

<sup>5</sup> College of Earth, Ocean, and Atmospheric Sciences; Oregon State University; Corvallis, OR, USA

<sup>6</sup> Geospatial Management Office; USDA Forest Service; Salt Lake City, UT, USA

<sup>7</sup> Rocky Mountain Research Station; USDA Forest Service; Fort Collins, CO, USA

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

**Abstract.** ~~Spaceborne remote sensing of snow currently enables landscape-scale snow covered area, but estimating snow mass in the mountains remains a major challenge from space.~~ 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. Knowledge of the spatial patterns and predictors of density is critical for accurate assessment of SWE and essential snow physics, such as energy balance and mechanics related to hazards and over-snow mobility. 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 co- and cross-polarization coherence and distributed LiDAR–GPR inferred bulk density with machine learning. The root-mean-square error between the distributed estimates is 12 cm for depth, 27 kg/m<sup>3</sup> for density, and 42 mm for SWE, and the median relative uncertainty in distributed SWE is 7 %. Wind, terrain, and vegetation interactions display corroborated controls on bulk density that show model and observation agreement. The spatially continuous snow density and SWE estimated over approximately 16 km<sup>2</sup> represents ~~the next step~~ towards broad-scale SWE retrieval.

### 1 Introduction

Anthropogenic climate change has contributed to the declining western U.S. snowpack over the past 50 + years (e.g., Pierce et al., 2008), and by the end of the 21st century, snow water equivalent (SWE) in the western U.S. is projected to decline by ~50 ± 10 % (Siirila-Woodburn et al., 2021). 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 with changing conditions, the relationship between these sparse observations and snow distribution patterns across



watersheds is changing. Snow varies over such short length scales that spatial interpolation from ground observations is not possible. In response to declining snowpacks in the western U.S. (Mote et al., 2018) and the increased demand on water resources (e.g., Achhemi et al., 2018), attention has been given to measuring, modelling, and developing new techniques for SWE estimates (e.g., Lettenmaier et al., 2015). The objective of NASA's snow experiment (SnowEx) campaign was to test a suite of remote sensing instruments, which measure SWE, and can be deployed on a satellite platform for global monitoring (Marshall et al., 2019). Accurate space-borne snow depth estimates have been achieved from passive microwave sensors (Tedesco et al., 2010), Sentinel-1 radar returns (Lievens et al., 2019, 2022), WorldView stereo digital surface models (McGrath et al., 2019), and light detection and ranging (LiDAR; e.g., Deschamps-Berger et al., 2023; Hu et al., 2021) aboard ICESat-2 (Abdalati et al., 2010). LiDAR techniques have the ability to measure snow depth, by differencing repeated acquisitions during periods with and without snow cover (e.g., Deems et al., 2013), whereas active microwave retrievals are complicated by signal penetration, depolarization, and backscattering. Because of the advantages of greater spatial resolution and flexible scheduling to target acquisitions during periods of interest, airborne LiDAR has become a prominent methodology for estimating snow depth and is being flown operationally for integration with hydrologic modelling at the catchment scale (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).

Snow density is typically measured in a time consuming and spatially limited manner by excavating and weighing snow samples of a known volume from a snow pit or snow core. Because snow depth varies in space more significantly than density (e.g., Elder et al., 1991; Sturm et al., 2010) and depth measurements may be collected more rapidly, density is observed far less frequently (e.g., Elder et al., 1998; Rovanešek et al., 1993). 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).

Often, empirical models provide the means 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 grain-bond characteristics can have an exponential effect (Sturm & Holmgren, 1998). The accuracy varies among linear snow density models that are parameterized by features such as net radiation, elevation, slope, curvature, and snow depth, as the success of such approaches is dependent on the time of year and snow climate (e.g., Elder et al., 1998; López-Moreno et al., 2013). 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) to account for the effects of snow depth and snow age on density. Over timescales of days to weeks, densification processes of freshly accumulated snow result in negative correlation between snow depth and density, while over timescales of months or longer, depth and density tend to be positively correlated (McCreight & Small, 2014). By differentiating between the short and long timescales of densification, McCreight and Small (2014) developed a linear density model capable of accurate daily density estimates for converting depth to SWE at distributed measurement stations. Hill et al. (2019) showed



improvements in accuracy compared to previous models when using sets of linear models based on snow depth and climate predictors from snow telemetry data from the western U.S., Canada, and Alaska.

70 Snow density can also be estimated 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 more simple time-dependent compaction curves and other models representing snow compaction dynamically as a function of snow viscosity and overburden pressure. Essery et al. (2013) found that the dynamic models offer more consistent and accurate characterizations of snowpacks. However, there has been a range of performance in snow density simulations, even for a single physics-based model. For example, Snobal (Marks et al., 1992) yielded low errors (mean absolute

75 difference of  $24 \text{ kg/m}^3$ ) in a study of the California Sierra Nevada (Painter et al., 2016) but higher errors (root-mean-square error up to  $142 \text{ kg/m}^3$ ) in a study of the Canadian Rockies (Lv & Pomeroy, 2020). Egli et al. (2009) found similar capabilities in estimating snow density with physics-based models and empirical models at a point location in Switzerland. In contrast, Raleigh and Small (2017) found that the choice of snow density model (empirical or physical) produced differences in spatial distributions and basin mean estimates of snow density in California, and one of their main conclusions was that the major

80 source of uncertainty in SWE estimates from airborne lidar was due to the accuracy of the density estimate used. Despite numerous techniques for modelling snow density, there are few studies comparing how they characterise spatial variations in snow density and the underlying processes related to density variations, largely due to limited density datasets. Spatial validation has been limited because few techniques exist for continuous spatial measurement of snow density, and the limited spatial support of in situ observations proves too intensive for analyses of spatial variability in snow density.

85 Relationships between dielectric permittivity (a main constituent of electromagnetic wave propagation) and snow density (e.g., Matzler, 1996), provide radar remote sensing an advantage for snow density retrieval. Passive microwave emission measurements combined with radiative transfer modelling is an established theoretical basis for retrieving snow density and ground permittivity (Schwank et al., 2015). Though this technique has been proven experimentally at the plot scale (Lemmetyinen et al., 2016), initialising the ground temperature and roughness remains challenging for satellite application.

90 Ground-penetrating radar (GPR) 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. With information about the propagation velocity of the snow, GPR analysis can estimate the snow depth, or by exploiting a ray path function of travel-time versus antenna separation (offset) the velocity can be estimated, and thereby the snow depth and density. Interpreting GPR transect imagery is a laborious process that requires expert knowledge of digital signal filtering and manual picking of travel-times. Though intricate, multi-offset

95 GPR analysis has continuously measured snow depth and density along transects of hundreds of metres in alpine mountains (Griessinger et al., 2018) to tens of kilometres in Greenland (Meehan et al., 2021). By combining drone-based aerial photogrammetry with GPR travel-times, snow density has been estimated along 100 m scale transects, which have been analysed for densification time-series (McGrath et al., 2022; Valence et al., 2022) and spatial interpolation across the study-plot scale (Yildiz et al., 2021). Radar derived snow observations require calibration and validation, which remains limited due



100 to the challenges of in situ sampling. Further hindering comparisons, in situ snow observations exhibit shorter length-scale  
variability than can be inferred with radar techniques, which tend to reveal lower spatial-frequency patterns.  
Here, we evaluated whether combining GPR two-way travel-time (TWT) observations with airborne LiDAR snow depths can  
accurately estimate density, and then be used to derive spatially continuous estimates of bulk snow density (and thus SWE),  
and whether these estimates reveal repeatable and physically realistic spatial patterns at Grand Mesa, Colorado, USA. We  
105 developed a novel approach for automatically interpreting TWTs using co- and cross-polarization coherence of GPR cross-  
section images and implemented a fast decision tree search to pair co-located GPR and LiDAR observations. Using variogram  
analysis, we determined the spatial length scales of variability in LiDAR snow depth, GPR TWT, the resulting bulk density,  
and SWE, to quantify differences in the snow distribution between open and forested areas. Snow density inferred from the  
integration of GPR and LiDAR information was then distributed at 1 m resolution over a ~16 km<sup>2</sup> study domain using machine  
110 learning regression techniques trained on LiDAR features. Spatial covariance derived from the snow density variogram was  
developed into a baseline, statistically distributed snow density model representative of the mean and spread of density  
observed in snow pits, using random normal field theory. We assessed the random field model and snow density models  
distributed via linear (MLR), random forest (RF), and artificial neural network (ANN) regression for accuracy, structural  
similarity, and ability to represent wind shelter and exposure. We evaluated the degree of improvement in SWE at the sub-  
115 catchment scale by comparing distributed SWE estimates from regression techniques and random-field synthesis to those  
estimated from mean density measured in snow pits with the airborne LiDAR snow depths. We set 10 % average uncertainty  
as a benchmark for our approach (National Academies of Sciences, Engineering, and Medicine, 2018), and identified areas  
where uncertainties exceed this standard.

Our work addresses the need for high accuracy, distributed density measurements to improve parameterizations of snow  
120 densification processes and reduce model uncertainty. The measurement and spatial distribution of snow density will increase  
the snow-water resources community awareness of the spatial variability in snow density and its importance. Knowledge of  
the spatial patterns and predictors of density is applicable for radar remote sensing retrievals of snow, essential snow physics  
such as mechanics related to hazards and over snow mobility, and thermal conductivity as it pertains to the energy balance and  
modelled snowpack evolution.

## 125 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<sup>2</sup>.  
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  
130 near-ideal study area for evaluating airborne snow remote-sensing techniques and developing many challenging snow remote-  
sensing advancements, such as our density retrieval.



135 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. The snow pits were distributed in forested and open areas along the swaths of the three airborne remote-sensing campaign flight lines (Fig. 1).

## 2.2 GPR data acquisition

140 Two GPR instruments were operated during the first week of the Grand Mesa IOP. In the forested areas, a conventional L-band GPR was pulled by ski in forested areas of central Grand Mesa and into open areas on the forest perimeters during 30 January – 1 February and 5 February (Webb, 2021), which was equipped with a Global Positioning Satellite (GPS) receiver with 2.5 m horizontal accuracy. Whereas, in open areas, we deployed multi-polarization L-band GPR fastened within a sled that was pulled by 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 the many forested stands in the survey domain but did not travel through densely treed areas. The multichannel L-band GPR 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 sub-meter accuracy was located on the snowmobile 5 m away from the GPR array, so we applied a geometric correction to relocate the coordinate positions to the antenna midpoint of each channel.

150 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  $10 \pm 1$  cm, while the interval for traces collected by ski is  $5 \pm 1$  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. For the multi-polarization GPR, we estimated the accuracy of the georeferencing at 70 cm, which is on the order of the GPR footprint. This estimate follows from adding the horizontal dilution of precision (50 cm) and uncertainty in the sled location (50 cm) in quadrature. 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

160 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





165 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 emphasises the continuity and amplitude of the ground reflection, which benefits the method for automatically picking the travel-times. The GPR data within forests were manually interpreted and picked 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.

### 170 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,

$$175 \quad C(t) = \frac{1}{2} \sum_{j=1}^N \left\{ \left[ \sum_{i=1}^M S_{i,t} \right]^2 - \sum_{i=1}^M S_{i,t}^2 \right\}, \quad (1)$$

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 ( $t$ ) of the GPR signal ( $S_{i,t}$ ) for channels  $i$  ( $M = 2$ ). The HH-HV coherence ( $C_{HH-HV}$ ) at each shot location is then normalised by the maximum coherence, ( $C$ ):

$$180 \quad C_{HH-HV} = \frac{c}{(C)}. \quad (2)$$

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 inferred 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 footprint), and therefore the two channels sum coherently.

185 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, less than 1 % of the automatic picks requiring manual correction. To illustrate this, automated picks are overlaid on the radargrams of a 900-metre-long transect in Fig. 2. 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).



## 2.4 Snow observations

### 2.4.1 In situ measurements

Snow pit observations and manual depth probe measurements were collected throughout the 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<sup>3</sup> wedge sampler, with duplicate samples. If the difference between the two measurements at a given depth exceeded 10 %, the density was sampled a third time, and bulk density was then calculated by averaging all measurements for each snow pit. The average density sampled from each of the columns shows high repeatability with a root-mean-squared-difference of 2.5 %. By regressing average density against the time elapsed throughout the campaign, we observed linear trends in snow densification: 0.07 kg/m<sup>3</sup>/hr in forested areas and nearly double in open areas (0.13 kg/m<sup>3</sup>/hr). 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 these rates. 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 measurements ( $h_{s,probe}$ ) were collected using geolocated probes ( $\pm 3$  m spatial accuracy) along spiral transects ( $\sim 60$  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 using the Multiscale Model to Model Cloud Compare (M3C2) method (Lague et al., 2013). The M3C2 method operates directly on point cloud data, computes the local distance between two point clouds at a scale that is appropriate for the surface roughness, and estimates a confidence interval for each distance measurement. The Airborne Snow Observatory (ASO) performed the snow-free acquisition on 26 September 2016 (Painter et al., 2016; Painter & Bormann, 2020), and NV5 Geospatial acquired a time-series of snow-covered surface elevations during the IOP, both with a point density of approximately 20 points/m<sup>2</sup>. We selected the 1 February 2020 flight to minimise temporal differences with the GPR and resulting errors due to snow redistribution and densification. We removed vegetation following methods in Štroner et al. (2021) and 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 M3C2 method to estimate snow depth on a 1 m grid. The relative accuracy of the snow depth measurement was estimated at 7 cm, based on the maximum standard deviation of the M3C2 method, which agrees well with previous lidar error assessments (Hojatimalekshah et al., 2021). After computing the snow depth, the 3 m ASO bare-earth and vegetation data products were resampled 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.



### 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 co-register 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, and the estimated velocities were inverted for density. By calculating the maximum cross-correlation lag on continuous segments of transect data, we determined that the overall accuracy of the spatial registration between the LiDAR and GPR varies on the order of a few metres. We found that errors in the co-registration of these data are the leading source of error in the estimated densities.

The average electromagnetic wave speed of the snowpack was estimated using

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

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

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

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 ( $v_s$ ; Equation 3), and the density of ice ( $\rho_i = 917$  kg/m<sup>3</sup>) to determine the dry snow density ( $\rho_{s, \text{LiDAR-GPR}}$ ; Equation 4).

#### 2.4.3.1 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 snow density. This involved establishing a level curve through the average snow values of the study area (mean density 276 kg/m<sup>3</sup>, mean TWT 8 ns, and mean snow depth 96 cm) and applying perturbations to evaluate the density error (Fig. 3). Perturbations of up to  $\pm 1$  ns were added to the TWT and  $\pm 15$  cm were added to the depth. After the TWT and depth perturbations were applied, the densities were evaluated, and the mean (276 kg/m<sup>3</sup>) was subtracted from this result to measure the density perturbations. The error bars of Fig. 3 represent the LiDAR root-mean-square difference (RMSD) 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<sup>3</sup> can be expected from this sensor integration method. Using variogram analysis (Section 2.5), we determined that beyond 10 m these errors are spatially uncorrelated suggesting that the errors are random and can be treated with filtering.

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. To extract a meaningful density signal, we reduced the random error to  $\pm 30$  kg/m<sup>3</sup> by filtering outliers. We chose the interquartile range of the LiDAR – GPR inferred densities as the threshold for determining outliers, because the 25<sup>th</sup> and 75<sup>th</sup> 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 inferred density observations.

#### 2.4.4 Wind and terrain exposure

##### 2.4.4.1 Meteorological station wind data and snow transport

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. The MW station was in exposed western terrain of the mesa, 350 m west of the study domain boundary. The MM station is sheltered within a dense stand of conifer trees 18.7 km east of the study domain boundary.

Wind data were examined from 1 October 2019 through the end of the SnowEx IOP on 12 February 2020. The MW station wind rose is presented in the Appendix C (Fig. C.1a). Hourly air temperature data parameterized 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 as 6.4 m/s at 200 degrees (Fig. C.1b).

##### 2.4.4.2 Maximum upwind slope and wind factor parameters

Serving as an additional comparison, we computed two parameters (maximum upwind slope and wind factor parameters; Winstral et al., 2002) from the 1 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 ( $S_x$ ) was calculated for each azimuth from 0 to 355 degrees in 5-degree increments, averaged over  $\pm 15$ -degree overlapping bins (Winstral et al., 2002). A search distance of 25 m was applied in the calculation of local  $S_x$ , while a 250 m distance was used to calculate the outlier length scale  $S_x$  between 25 and 250 m from the pixel. The local and outlying  $S_x$  parameters were differenced to calculate the slope break parameter,  $S_b$  (Winstral et al., 2002).

The wind factor parameter determines the degree of wind exposure or shelter for a given pixel, based on  $S_x$ ,  $S_b$ , vegetation proximity, and the average scalar multiple by which wintertime winds at the MW station exceed 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  $S_x$  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<sup>th</sup> percentile were arbitrarily reduced by 10 % to enhance the effective wind shelter provided by vegetation.

290 Rather than incorporating the  $S_x$  and wind factor parameters as model predictors of snow density, we utilised this information  
as explanation of the patterns and processes captured by the machine-learned regression ensemble. To verify the GPR-LiDAR  
estimated training data and the model results, we calculated the correlation between the model input and output and the  $S_x$  and  
wind factor rasters for all wind directions. Because pixels with high wind exposure have negative  $S_x$  values, and we  
hypothesise that wind exposed terrain will have greater snow density (e.g., due to formation of wind slabs), the wind direction  
295 expressing the largest negative correlation was recognized as the best explanation of density patterns determined by  $S_x$ .  
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. Agreement  
between the prevailing wind direction responsible for snow transport measured at the MW station and the wind direction  
expressing the strongest correlation with model inputs and outputs, we suggest, would bolster the method of solving for snow  
300 density by integrating GPR and LiDAR information and support the spatial patterns predicted by regression model outputs,  
which were not trained with wind information.

## 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  
305 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  
310 Carlo simulation with 10 % random subsampling to assess the mean and standard deviation of the variogram parameters (Efron  
& Tibshirani, 1986).

To serve as a baseline for model assessment, a 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<sup>2</sup> domain. Using Cholesky  
315 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.



## 2.6 Regression modelling of spatial snow density

320 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  $\sim 16$  km<sup>2</sup> 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 ( $Z_g$ ),  
325 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  $\geq 0.5$  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  
330 were applied to the surrounding terrain. For detail on the parameter estimation and predictor importance see Appendix A.

## 3 Results

### 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.  
335 4). 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  
mean LiDAR snow depth is  $96 \pm 15$  cm, while in the forest ( $H_{veg} \geq 0.5$  m), the mean LiDAR snow depth is  $79 \pm 23$  cm.  
The in situ snow depth observations ( $h_{s,Pit}$  and  $h_{s,Probe}$ ) compare well with the LiDAR snow depths throughout the entire  
domain ( $R^2 = 0.61$ , RMSE = 11 cm, ME = 0 cm), however within the open and forested domains individually, LiDAR and  
340 GPR estimated snow depths are uncorrelated with in situ snow depths (Table 1).

### 3.2 LiDAR – GPR estimated density

The LiDAR–GPR inferred 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. 5). 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  
345 compared to snow pit observations, the relative RMSE among the 37 snow pits that are within 12.5 m of the GPR transects is  
13 % (Table 2). The GPR data were acquired within a few metres of, but not directly beside the faces of snow pits, which  
necessitated a radius for pairing observations. The choice of 12.5 m matches that of the filtering during the sensor integration



step (Section 2.4.3). Measurements accumulated over 12.5 m distance introduce inherent variability on the order of 10 % (Section 3.3), which, along with differences between representative observation scales may explain the weak correlation between estimated density and the in situ measurements. The standard deviation of  $\rho_{s,Pit}$  encompasses the mean of  $\rho_{s,LiDAR-GPR}$  and vice versa, and the data are both normally distributed as evidenced by Z-test (Appendix B). 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 (see Table C.1). Within the width of bins used to calculate the wind parameters, the optimal wind direction uniquely agrees with the prevailing winds able to transport snow and suggests that the method retrieves density patterns which are related to the degree of exposure and shelter due to topography and vegetation.

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

The generalised relative semi-variogram allows us to examine the expected percent variability between observations at a given distance separation and the differences in the length scales of variability among depth, density, SWE, and TWT within the forested and open areas of Grand Mesa (Fig. 6). Table 3 overviews the generalised relative semi-variogram parameter estimates (nugget, sill, and correlation length). The variability in depth, SWE, and TWT are lower, while variability in density is greater within the open areas than within forest stands and the length scale of variability is longer. 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. However, we find that the expected variability among co-located  $\rho_{s,LiDAR-GPR}$  is approximately 2 %, which is consistent with replicated in situ density observations ( $\rho_{s,Pit}$ ) having approximately 2.5 % variability on average. Compared to the estimated correlation length of density ( $\sim 100$  m), the median distance between snow pits is  $\sim 150$  m, which indicates that average snow pit observations are independent of each other and unable to resolve spatial patterns. Supporting the basis that in dry snow, depth and density formulate TWT and SWE, the relative length-scales of variability for SWE closely resemble that of TWT and indicate that TWT is a better informer of SWE than either depth or density, independently.

### 3.4 Learned Regression and Random Field Modelled Density

Using supervised learning techniques three models were generated from LiDAR information (Fig. 7 a, b, and c). Using prior information from the in situ snow pit observations a randomly distributed density was synthesised (Fig. 7 d). The mean of the regression-based ensemble was taken to generate  $\rho_{s,\overline{EnS}}$  (not shown). 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). As discussed in Appendix A.1.2, the model hyperparameters were developed such that the variance of the predictions in pixels where training data exists matches that of the predictions. This coincided with an  $R^2$  of approximately 0.8. Evaluation against snow pit density does not depict an obvious best model choice. However, we have selected the average of the regression models, as the random field contains little meaningful spatial



380 information, and the average of the ensemble spreads the strengths of each regression model (i.e., accuracy of training data and representativeness of wind-induced densification). The spatial similarity of these models is presented in Appendix A.3.

### 3.5 Model representation of wind, terrain, and vegetation effects

385 Snow density inferred by the LiDAR – GPR retrieval method and modelled using regression approaches displays spatial patterns that we hypothesise are representative of the interactions between small-scale topography, vegetation, and wind. To test this hypothesis, we explored the correlation between the maximum upwind slope and wind factor parameters developed by Winstral et al. (2002). For each wind direction, the three learned regression models and the ensemble mean were correlated to the wind exposure parameters (Fig. 8). Based on the maximum correlation observed for the wind factor, and the minimum observed for the upwind slope, the patterns of snow density agree with the prevailing south-southwest wind direction (Appendix C). No correlation was evident between these wind exposure parameters and the randomly distributed snow densities (Table C.1). 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 lack of a large-scale topographic trend in density, such as one driven by elevation or aspect, evinces the role of forest vegetation on density.

### 3.6 Spatially distributed snow water equivalent

395 SWE was distributed within the  $\sim 16 \text{ km}^2$  domain by multiplying the mean snow density of the regression ensemble estimated with the LiDAR snow depths (Fig. 9). Decreased SWE and increased variability within forest stands corroborates previous work on wind-terrain-vegetation characterization of Grand Mesa (Webb et al., 2020). Wind redistribution is also evidenced by snow drifts which tend to have less SWE on the windward side and increased SWE on the leeward side. Metre-scale stippling patterns are the effect of low-stature vegetation ( $H_{veg} < 0.5 \text{ m}$ ) and boulders, which tend to reduce snow depth and to a lesser degree reduce the modelled average density. The averages and standard deviations of SWE measured in snow pits and distributed throughout the study area domain by combining either the average snow pit density ( $276 \text{ kg/m}^3$ ) or the ensemble modelled density with LiDAR snow depths is presented in Table 4. It is unclear which method best represents the true water equivalent for the study domain because the true SWE distribution is not observed.

400 We compared the SWE distributed using LiDAR snow depths multiplied by the average density in each respective domain measured from snow pits, the regression modelled mean ensemble densities, and the random field densities to observations at all snow pits within the study domain (Table 4). SWE distributed from a single average density has lower bias (by  $9 \text{ kg/m}^3$ ) and root-mean-squared-error (by  $4 \text{ kg/m}^3$ ) than  $b_{s, LiDAR - \overline{EnS}}$ . The ensemble-modelled densities explain more variation in the distributed SWE estimates than those distributed randomly with an appropriate correlation length and prior mean and spread but maintain a larger bias. Using an average measured value performs slightly better, which suggests that depth is primary to SWE in this environment, yet covariance exists between snow depth and regression modelled density. Assessed at the average value of SWE for all 96 snow pits, the SWE distributed using  $H_{s, LiDAR}$  and  $\rho_{s, \overline{EnS}}$  has a negative bias of approximately 7 %, 410





which is within the uncertainty of SWE distributed from LiDAR snow depths and the mean of snow pit density measurements (Appendix B.2).

### 3.7 Contributions of SWE uncertainty

415 The errors between  $h_{s, \text{LiDAR}}$  and  $h_{s, \text{Probe}}$  are positively correlated when compared to  $h_{s, \text{LiDAR}}$  ( $R^2 = 0.16$ ), as are the errors between  $\rho_{s, \overline{EnS}}$  and  $\rho_{s, \text{Pit}}$  are when compared to  $\rho_{s, \overline{EnS}}$  ( $R^2 = 0.35$ ), but the errors among snow depth and density are uncorrelated with negligible covariance. Using simple linear regression, we modelled the errors as a function of the LiDAR measured snow depth or the regression ensemble mean density. Following from the propagation of errors for relative errors in snow depth and density, we estimated the SWE uncertainty to first order (Raleigh & Small, 2017). The distributed relative SWE uncertainty is presented in Fig. 10 and is negatively correlated with the distributed SWE ( $R^2 = 0.44$ ). The median SWE uncertainty is 7.6 %, 420 which breaks down to 10.6 % median uncertainty in the forest and 6.9 % median SWE uncertainty in the open areas. Snow depth and density had approximately equal contributions to SWE uncertainty at 4 %. Within the open areas LiDAR snow depth contributes 3 % uncertainty, while in the forested areas snow depth contributes 8 % median uncertainty. Snow density uncertainty is 4 % in both open and forested areas.

## 4. Discussion

425 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. These derived data and the validation data acquired during the NASA SnowEx 2020 IOP at Grand Mesa, Colorado, USA, provide a core set of observables that can better inform essential snow research. 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 depolarization on L-Band GPR signals, we developed a new, 430 automated GPR processing workflow that accurately 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.

435 Sensitivity analysis showed how measurement errors propagate into the LiDAR – GPR measured snow density (Section 2.4.3). We found that measurement errors 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<sup>3</sup> or greater, though these estimates were reduced to 30 kg/m<sup>3</sup> by median filtering and interpolating through outliers (Section 2.4.3). While the signal of each instrument is 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 440 transformations) that are on the scale of the 1 m resolution data products. In some locations the co-registration may 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. 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 where LiDAR—  
445 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<sup>3</sup> error in the modelled density on average (up to 2 % in forest stands), that outlier filtering is robust to sensor integration errors, and this error is small relative to the overall SWE uncertainty.

Our work characterised the measurement uncertainties and the resulting SWE uncertainty in pursuit of the goal for 10 %  
450 uncertainty in global SWE estimation (National Academies of Sciences, Engineering, and Medicine, 2018). Based on the evaluation of the remotely sensed or modelled snow properties with in situ measurements, we used simple linear regression to model uncertainties spanning both forested and open areas. The uncertainty in LiDAR snow depth varies 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 were greater beneath the tree canopy (Table  
455 1). The choice of uncertainties propagated through the SWE uncertainty analysis (Section 3.7) dictates which factor, depth or density, will have the greater contribution to the overall SWE uncertainty. Uncertainty in midwinter SWE tends to reduce at peak SWE, where snow depth and density are greater. Our findings (Fig. 10) are within the remarkably difficult to achieve 10 % goal and point to the success and accuracy of the joint LiDAR – GPR methodology for SWE retrieval at plot to forest stand scales.

We tested the model sensitivity to training and learned how much data is required for accurate density estimation. Using  
460 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. The large GPR grid  
465 in open areas captured the high degree of spatial heterogeneity and improved LiDAR spatial predictor importance, while GPR acquired in forests added necessary data for estimating sub-canopy snow density.

The Grand Mesa IOP is 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  
470 campaign can be treated as a random normal variable with mean and standard deviation of  $276 \pm 21$  kg/m<sup>3</sup>, despite differences of roughly 25 kg/m<sup>3</sup> 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  
475 within the study domain are sufficient to reduce the median uncertainty in distributed SWE to within  $10 \pm 2$  % (Appendix



B.2). 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 3.6) but resolves no information about spatial patterns of snow density, and is therefore less useful for understanding density patterns across the landscape. In situ snow campaigns targeting average SWE require far fewer pits than needed to  
480 resolve the spatial patterns. Without prior information regarding the correlation length of snow density, the random field synthesised density has unfounded spatial context, as the required covariance information cannot be retrieved by manual observations.

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 radar remote sensing. As estimated from the variogram analysis, the median  
485 distance between snow pit observations in our study is beyond the length scale of variability for snow density, and therefore the observations are spatially uncorrelated. 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 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  
490 160 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. Though it is worth noting 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 density measurements inferred from GPR profiles permitted us to quantify the spatial length scales of density variability, albeit insensitive to subpixel or sub-aggregated pixel length scales on which greater variability may exist. Using variogram analysis, we determined that measurements of density up to ~100 m are correlated. These findings significantly differ from a previous variogram analysis that found correlation lengths for snow density of less than 10 m (Yildiz et al., 2021). However, these findings are study site dependent, and it may be that we have identified an additional longer, lower spatial frequency  
500 scaling of snow density. 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). 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 and nugget variability in the forested and open areas. This finding supports TWT as an informer of spatial SWE variability. We found that snow depth and TWT reached comparable maximum variability  
505 in open areas, while depth variability is greater in the forested areas. Snow density exhibited 4 % greater variability in the open areas than in the forests, indicating that wind exposure increases the variability, and conversely, shelter provided by terrain and vegetation tends to reduce spatial density variability.

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



510 as maximum upwind slope and wind factor. Instead, these wind parameters were utilised as a corroboratory metric for  
explaining spatial patterns predicted in regression modelling. For each regression model, we identified the most important  
LiDAR features used to distribute density and found dependencies on 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  
515 weakened. We used a “kitchen-sink” approach to the regression modelling presented but found comparable accuracy in models  
using fewer parameters. To capture the range of processes (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 artificial neural network combined with airborne LiDAR-derived snow depth to estimate SWE. Broxton et al. (2019)  
520 highlighted the importance of representing the broader landscape with distributed densities for estimating SWE, finding ~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 ~1000  
525 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. The low relief and shallow slopes  
of Grand Mesa remove much of the elevation and aspect dependency on snow density but evince the interactions between  
wind, local terrain, and vegetation. Our approach in validating wind effects on density is an explanatory simplification of the  
530 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).

The regression models developed from the LiDAR and GPR acquisitions during the SnowEx 2020 Grand Mesa IOP will likely  
535 have weak predictive capability at other field sites. 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). It may be necessary to recalibrate the model using GPR or another instrument to measure radar travel-time such as  
airborne FMCW radar (e.g., Yan et al., 2017). Provided radar-measured snow surface elevations, it is possible to infer density  
540 along FMCW flight-lines when combined with a snow-free digital surface model. However, high resolution elevation and  
snow depth data significantly improves modelled spatial heterogeneity in snow 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



545 permittivity (e.g., Valence et al., 2022), and in situ measurement campaigns combined with learned-regression models (e.g.,  
Wetlaufer et al., 2016; Broxton et al., 2019) should be utilised where appropriate and examined for the physical basis. Empirical  
models of this type are often distributed over vast areas with little validation or consideration to the underlying physical  
processes.

## 5. Conclusion

550 We developed an innovative approach to estimate SWE across a  $\sim 16$  km<sup>2</sup> 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  $\sim 160$  km of GPR  
555 transects. From these along-track estimates, we calculated the length-scales of variability for depth, density, SWE and  
found that snow depth had 25 % reduced variability, SWE had 12 % reduced variability, and density had 4 % increased  
variability in open areas compared to that exhibited within forested stands. In dry snow, we found TWT informs SWE, better  
than either depth or density independently. Snow density distributed by regression techniques revealed anomalies associated  
with localised terrain features and forest stands that shelter the snowpack from wind densification. Spatial patterns show the  
best agreement in the direction of prevailing winds strong enough to transport snow. Roughly 60 % of density variability in  
our single mid-winter survey can be accounted for using a wind factor analysis. Additionally, we estimated that snow density  
560 and depth contribute equally to SWE uncertainty for a typical mid-winter snowpack on Grand Mesa, Colorado, USA. On  
average, distributed relative SWE uncertainty was less than 10 % and tends to be greatest in the shallower and lower density  
snow beneath tree canopies. The total SWE is not observable, making it unclear which method is more accurate, yet analysis  
suggests that one snow pit per km<sup>2</sup> may reduce the SWE uncertainty to within  $10 \pm 2$  %. However, using such a sampling  
strategy would not resolve spatial patterns and variability in snow properties. This pilot study provides a useful method  
565 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 space-borne radar remote sensing  
retrievals of SWE.

## Appendix A.

### A.1 Regression Parameter Optimization

#### 570 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,  $\beta$  is the vector of the regression coefficients which we seek to estimate, and  $\epsilon$  represents the model residual. From the method of least squares, the multiple linear regression coefficients are estimated as

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

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 RMSD between the two models is less than 1 kg/m<sup>3</sup>.

### 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 toolbox, 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, calculating the averaged standard deviations of density data for each ensemble predicted along the GPR transects ( $\overline{\sigma_{train}}$ ) and the densities predicted 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

$$\varphi = \frac{1}{R^2} \frac{\overline{\sigma_{pred}}}{\overline{\sigma_{train}}} \quad . \quad (\text{A.3})$$

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, yet 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<sup>3</sup>. 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 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

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” 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

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 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 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 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.

### B.1 Statistics of in situ, LiDAR – GPR Inferred, and Modelled Snow Density

To show that the inferred 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.1). The means of the distributions are overlapping within the standard deviations of the datasets. The LiDAR – GPR inferred densities suggest a broader distribution of densities than observed in pits or modelled. Sampling biases may explain 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  $-10 \text{ kg/m}^3$  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.2 Sample Uncertainty of Density and SWE

The Grand Mesa IOP is 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, an order of magnitude fewer snow pits are available, and uncertainty arises from spatial sampling of the underlying density distribution. We found that snow pit density measurements have an average uncertainty of 2.5 %. 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.2). 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 3.7 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  $10 \pm 2$  %.

## 665 **Appendix C.**

### **C.1 Ancillary Data: Wind Speed and Direction**

Figure C.1 presents the wind speed and direction data from the Mesa West meteorological station (Houser et al., 2022). Table C.1 lists the accompanying data presented in Fig. 8.

#### **Code availability**

670 The processing software developed for the multi-polarization GPR analysis is available at [https://github.com/tatemeehan/SnowEx2020\\_BSU\\_pE\\_GPR](https://github.com/tatemeehan/SnowEx2020_BSU_pE_GPR).

#### **Data availability**

675 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 estimated density observations; LiDAR estimated snow depth, terrain, and vegetation model predictors, the ensemble of modelled snow density, and the SWE and uncertainty estimated therefrom will be archived within the NSIDC at DOI: 10.5067/LANQ53RTJ2DR pending the review of this manuscript.

#### **Author contribution**

680 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.

#### 685 **Competing interests**

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 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 staff of the SnowEx 2020 campaign, who helped make this research possible.

## References

- Abdalati, W., Zwally, H. J., Bindschadler, R., Csatho, B., Farrell, S. L., Fricker, H. A., Harding, D., Kwok, R., Lefsky, M., Markus, T., Marshak, A., Neumann, T., Palm, S., Schutz, B., Smith, B., Spinhirne, J., & Webb, C. (2010). The ICESat-2 laser altimetry mission. *Proceedings of the IEEE*, 98(5), 735–751. <https://doi.org/10.1109/JPROC.2009.2034765>
- Achhami, A., Kalra, A., & Ahmad, S. (2018). Dynamic Simulation of Lake Mead Water Levels in Response to Climate Change and Varying Demands. *World Environmental and Water Resources Congress 2018*, 1(i), 260–268. <https://doi.org/10.1061/9780784481400.023>
- Andrews, D. F. (1974). A robust method for multiple linear regression. *Technometrics*, 16(4), 523–531. <https://doi.org/10.1080/00401706.1974.10489233>
- Bentley, J. L. (1975). Multidimensional Binary Search Trees Used for Associative Searching. *Communications of the ACM*, 18(9), 509–517. <https://doi.org/10.1145/361002.361007>
- Bonner, H. M., Raleigh, M. S., & Small, E. E. (2022). Isolating forest process effects on modelled snowpack density and snow water equivalent. *Hydrological Processes*, 36(1). <https://doi.org/10.1002/hyp.14475>
- 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>
- Breiman, L. (2001). Random Forests. In *Machine Learning* (Vol. 45, pp. 5–32). <https://doi.org/10.1023/A:1010933404324>
- Broxton, P. D., Leeuwen, W. J. D., & Biederman, J. A. (2019). Improving Snow Water Equivalent Maps With Machine Learning of Snow Survey and Lidar Measurements. *Water Resources Research*, 55(5), 3739–3757. <https://doi.org/10.1029/2018WR024146>
- Cartographic Boundary Files*. (2020). U.S. Census Bureau. <https://www.census.gov/geographies/mapping-files/time-series/geo/cartographic-boundary.html>
- Cressie, N. (1985). Fitting variogram models by weighted least squares. *Journal of the International Association for*



*Mathematical Geology*, 17(5), 563–586. <https://doi.org/10.1007/BF01032109>

- 720 Deschamps-Berger, C., Gascoïn, S., Shean, D., Besso, H., Guiot, A., & López-Moreno, J. I. (2023). Evaluation of snow depth retrievals from ICESat-2 using airborne laser-scanning data. *The Cryosphere*, 17(7), 2779–2792. <https://doi.org/10.5194/tc-17-2779-2023>
- Deems, J. S., Fassnacht, S. R., & Elder, K. J. (2006). Fractal Distribution of Snow Depth from Lidar Data. *Journal of Hydrometeorology*, 7(2), 285–297. <https://doi.org/10.1175/JHM487.1>
- 725 Deems, J. S., Painter, T. H., & Finnegan, D. C. (2013). Lidar measurement of snow depth: A review. *Journal of Glaciology*, 59(215), 467–479. <https://doi.org/10.3189/2013JoG12J154>
- Efron, B., & Tibshirani, R. (1986). Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science*, 1(1), 77–77. <https://doi.org/10.1214/ss/1177013817>
- Egli, L., Jonas, T., & Meister, R. (2009). Comparison of different automatic methods for estimating snow water equivalent. *Cold Regions Science and Technology*, 57(2–3), 107–115. <https://doi.org/10.1016/j.coldregions.2009.02.008>
- 730 Elder, K., Dozier, J., & Michaelsen, J. (1991). Snow accumulation and distribution in an Alpine Watershed. *Water Resources Research*, 27(7), 1541–1552. <https://doi.org/10.1029/91WR00506>
- Elder, K., Rosenthal, W., & Davis, R. E. (1998). Estimating the spatial distribution of snow water equivalence in a montane watershed. *Hydrological Processes*, 12(10–11), 1793–1808. [https://doi.org/10.1002/\(SICI\)1099-1085\(199808/09\)12:10/11<1793::AID-HYP695>3.0.CO;2-K](https://doi.org/10.1002/(SICI)1099-1085(199808/09)12:10/11<1793::AID-HYP695>3.0.CO;2-K)
- 735 Essery, R., Morin, S., Lejeune, Y., & B Ménard, C. (2013). A comparison of 1701 snow models using observations from an alpine site. *Advances in Water Resources*, 55, 131–148. <https://doi.org/10.1016/j.advwatres.2012.07.013>
- Fassnacht, S. R., Heun, C. M., López-Moreno, J. I., & Latron, J. (2010). Snow Density Variability in the Rio Esera Valley, Pyrenees Mountains, 2. Study Site. *Cuadernos de Investigación Geográfica*, 36(1), 59–72.
- 740 Goh, A. T. C. (1995). Back-propagation neural networks for modeling complex systems. *Artificial Intelligence in Engineering*, 9(3), 143–151. [https://doi.org/10.1016/0954-1810\(94\)00011-S](https://doi.org/10.1016/0954-1810(94)00011-S)
- Griessinger, N., Mohr, F., & Jonas, T. (2018). Measuring snow ablation rates in alpine terrain with a mobile multioffset ground-penetrating radar system. *Hydrological Processes*, 32(21), 3272–3282. <https://doi.org/10.1002/hyp.13259>
- Hapfelmeier, A., Hothorn, T., Ulm, K., & Strobl, C. (2014). A new variable importance measure for random forests with missing data. *Statistics and Computing*, 24(1), 21–34. <https://doi.org/10.1007/s11222-012-9349-1>
- 745 Hedrick, A. R., Marks, D., Havens, S., Robertson, M., Johnson, M., Sandusky, M., Marshall, H., Kormos, P. R., Bormann, K. J., & Painter, T. H. (2018). Direct Insertion of NASA Airborne Snow Observatory-Derived Snow Depth Time Series Into the iSnobal Energy Balance Snow Model. *Water Resources Research*, 54(10), 8045–8063. <https://doi.org/10.1029/2018WR023190>
- 750 Hiemstra, C. A., Vuyovich, C. M., & Marshall, H.-P. (2021). *SnowEx20 Grand Mesa Reference GIS Data Sets, Version 1*. <https://doi.org/https://doi.org/10.5067/YDZXY4Q79VIJ>
- Hiemstra, C., Marshall, H. P., Vuyovich, C., Elder, K., Mason, M., & Durand, M. (2020). *SnowEx20 Community Snow Depth*



*Probe Measurements, Version 1*. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. <https://doi.org/https://doi.org/10.5067/9IA978JIACAR>

755 Hill, D., Burakowski, E., Crumley, R., Keon, J., Hu, J. M., Arendt, A., Wikstrom Jones, K., & Wolken, G. (2019). Converting Snow Depth to Snow Water Equivalent Using Climatological Variables. *Converting Snow Depth to Snow Water Equivalent Using Climatological Variables*, 1–34. <https://doi.org/10.5194/tc-2018-286>

760 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., & Enterkine, J. (2021). Tree canopy and snow depth relationships at fine scales with terrestrial laser scanning. *The Cryosphere*, 15(5), 2187–2209. <https://doi.org/10.5194/tc-15-2187-2021>

Homer, C., Dewitz, J., Jin, S., Xian, G., Costello, C., Danielson, P., Gass, L., Funk, M., Wickham, J., Stehman, S., Auch, R., & Riitters, K. (2020). Conterminous United States land cover change patterns 2001–2016 from the 2016 National Land Cover Database. *ISPRS Journal of Photogrammetry and Remote Sensing*, 162(November 2019), 184–199. <https://doi.org/10.1016/j.isprsjprs.2020.02.019>

765 Houser, P., Rudisill, W., Johnston, J., Elder, K., Marshall, H. P., Vuyovich, C., Kim, E. J., & Mason, M. (2022). *SnowEx Meteorological Station Measurements from Grand Mesa, CO, Version 1*. NASA National Snow and Ice Data Center Distributed Active Archive Center. <https://doi.org/https://doi.org/10.5067/497NQVJOCBEX>

Hu, X., Hao, X., Wang, J., Huang, G., Li, H., & Yang, Q. (2021). Can the Depth of Seasonal Snow be Estimated From ICESat-2 Products: A Case Investigation in Altay, Northwest China. *IEEE Geoscience and Remote Sensing Letters*, 1–5. <https://doi.org/10.1109/LGRS.2021.3078805>

770 Isaaks, E. H., & Srivastava, R. M. (1989). *Applied Geostatistics*. Oxford University Press. <https://books.google.com/books?id=vC2dcXFLI3YC>

Jain, A. K., Jianchang Mao, & Mohiuddin, K. M. (1996). Artificial neural networks: a tutorial. *Computer*, 29(3), 31–44. <https://doi.org/10.1109/2.485891>

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

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

780 Lague, D., Brodu, N., & Leroux, J. (2013). Accurate 3D comparison of complex topography with terrestrial laser scanner: Application to the Rangitikei canyon (N-Z). *ISPRS Journal of Photogrammetry and Remote Sensing*, 82, 10–26. <https://doi.org/10.1016/j.isprsjprs.2013.04.009>

Lemmetyinen, J., Schwank, M., Rautiainen, K., Kontu, A., Parkkinen, T., Mätzler, C., Wiesmann, A., Wegmüller, U., Derksen, C., Toose, P., Roy, A., & Pulliainen, J. (2016). Snow density and ground permittivity retrieved from L-band radiometry: Application to experimental data. *Remote Sensing of Environment*, 180, 377–391. <https://doi.org/10.1016/j.rse.2016.02.002>

785 Lettenmaier, D. P., Alsdorf, D., Dozier, J., Huffman, G. J., Pan, M., & Wood, E. F. (2015). Inroads of remote sensing into hydrologic science during the WRR era. *Water Resources Research*, 51(9), 7309–7342.



<https://doi.org/10.1002/2015WR017616>

- Li, L., & Pomeroy, J. W. (1997). Estimates of Threshold Wind Speeds for Snow Transport Using Meteorological Data. *Journal of Applied Meteorology*, 36(3), 205–213. [https://doi.org/10.1175/1520-0450\(1997\)036<0205:EOTWSF>2.0.CO;2](https://doi.org/10.1175/1520-0450(1997)036<0205:EOTWSF>2.0.CO;2)
- 790 Lievens, H., Brangers, I., Marshall, H.-P., Jonas, T., Olefs, M., & De Lannoy, G. (2022). Sentinel-1 snow depth retrieval at sub-kilometer resolution over the European Alps. *The Cryosphere*, 16(1), 159–177. <https://doi.org/10.5194/tc-16-159-2022>
- Lievens, H., Demuzere, M., Marshall, H.-P., Reichle, R. H., Brucker, L., Brangers, I., de Rosnay, P., Dumont, M., Giroto, M., Immerzeel, W. W., Jonas, T., Kim, E. J., Koch, I., Marty, C., Saloranta, T., Schöber, J., & De Lannoy, G. J. M. (2019).  
795 Snow depth variability in the Northern Hemisphere mountains observed from space. *Nature Communications*, 10(1), 4629. <https://doi.org/10.1038/s41467-019-12566-y>
- López-Moreno, J. I., Fassnacht, S. R., Heath, J. T., Musselman, K. N., Revuelto, J., Latron, J., Morán-Tejeda, E., & Jonas, T. (2013). Small scale spatial variability of snow density and depth over complex alpine terrain: Implications for estimating snow water equivalent. *Advances in Water Resources*, 55, 40–52. <https://doi.org/10.1016/j.advwatres.2012.08.010>
- 800 Lukas, V., & Baez, V. (2021). *3D Elevation Program—Federal best practices: U.S. Geological Survey Fact Sheet 2020–3062. February*. <https://doi.org/https://doi.org/10.3133/fs20203062>
- Lv, Z., & Pomeroy, J. W. (2020). Assimilating snow observations to snow interception process simulations. *Hydrological Processes*, 34(10), 2229–2246. <https://doi.org/10.1002/hyp.13720>
- Marks, D., Dozier, J., & Davis, R. E. (1992). Climate and energy exchange at the snow surface in the Alpine Region of the  
805 Sierra Nevada: 1. Meteorological measurements and monitoring. *Water Resources Research*, 28(11), 3029–3042. <https://doi.org/10.1029/92WR01482>
- Marshall, H.P., Vuyovich, C., Hiemstra, C., Brucker, L., Elder, K., Deems, J., Newlin, J., Bales, R., Nolin, A., & Trujillo, E. (2019). *NASA SnowEx 2020 Experiment Plan. August 2019*, 1–100. <https://snow.nasa.gov/campaigns/snowex/experimental-plan-2020>
- 810 Marshall, Hans Peter, Koh, G., Sturm, M., Johnson, J. B., Demuth, M., Landry, C., Deems, J. S., & Gleason, J. A. (2006). Spatial variability of the snowpack: Experiences with measurements at a wide range of length scales with several different high precision instruments. *Proceedings ISSW 2006. International Snow Science Workshop, Telluride CO, U.S.A., 1-6 October 2006*, 359–364. <http://arc.lib.montana.edu/snow-science/item/947>
- Matzler, C. (1996). Microwave permittivity of dry snow. *IEEE Transactions on Geoscience and Remote Sensing*, 34(2), 573–  
815 581. <https://doi.org/10.1109/36.485133>
- McCreight, J. L., & Small, E. E. (2014). Modeling bulk density and snow water equivalent using daily snow depth observations. *Cryosphere*, 8(2), 521–536. <https://doi.org/10.5194/tc-8-521-2014>
- McGrath, D., Bonnell, R., Zeller, L., Olsen-Mikitowicz, A., Bump, E., Webb, R., & Marshall, H.-P. (2022). A Time Series of Snow Density and Snow Water Equivalent Observations Derived from the Integration of GPR and UAV SfM  
820 Observations. *Frontiers in Remote Sensing*, 3(May), 1–15. <https://doi.org/10.3389/frsen.2022.886747>



- McGrath, D., Webb, R., Shean, D., Bonnell, R., Marshall, H. P., Painter, T. H., Molotch, N. P., Elder, K., Hiemstra, C., & Brucker, L. (2019). Spatially Extensive Ground-Penetrating Radar Snow Depth Observations During NASA's 2017 SnowEx Campaign: Comparison With In Situ, Airborne, and Satellite Observations. *Water Resources Research*, 55(11), 10026–10036. <https://doi.org/10.1029/2019WR024907>
- 825 Meehan, T. G. (2021). *SnowEx20 Grand Mesa IOP BSU 1 GHz Multi-polarization GPR, Version 1*. NASA National Snow and Ice Data Center Distributed Active Archive Center. <https://doi.org/https://doi.org/10.5067/Q2LFK0QSVGS2>
- Meehan, T. G., Marshall, H. P., Bradford, J. H., Hawley, R. L., Overly, T. B., Lewis, G., Graeter, K., Osterberg, E., & McCarthy, F. (2021). Reconstruction of historical surface mass balance, 1984–2017 from GreenTrACS multi-offset ground-penetrating radar. *Journal of Glaciology*, 67(262), 219–228. <https://doi.org/10.1017/jog.2020.91>
- 830 Mote, P. W., Li, S., Lettenmaier, D. P., Xiao, M., & Engel, R. (2018). Dramatic declines in snowpack in the western US. *Npj Climate and Atmospheric Science*, 1(1). <https://doi.org/10.1038/s41612-018-0012-1>
- National Academies of Sciences Engineering and Medicine. (2018). Thriving on Our Changing Planet: A Decadal Strategy for Earth Observation from Space. In *Thriving on Our Changing Planet: A Decadal Strategy for Earth Observation from Space*. National Academies Press. <https://doi.org/10.17226/24938>
- 835 Neidell, N. S., & Taner, M. T. (1971). Semblance and Other Coherency Measures for Multichannel Data. *Geophysics*, 36(3), 482–497. <https://doi.org/10.1190/1.1440186>
- NOAA. (2021). *VDatum 4.3 Vertical Datum Transformation*. <https://vdatum.noaa.gov/>
- 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., & Winstral, A. (2016). The Airborne Snow Observatory: Fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo. *Remote Sensing of Environment*, 184(June), 139–152. <https://doi.org/10.1016/j.rse.2016.06.018>
- 840 Painter, T. H., & Bormann, K. J. (2020). *ASO L4 Lidar Point Cloud Digital Terrain Model 3m UTM Grid, Version 1*. NASA National Snow and Ice Data Center Distributed Active Archive Center. <https://doi.org/https://doi.org/10.5067/2EHMWG4IT76O>
- 845 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., & Nozawa, T. (2008). Attribution of declining Western U.S. Snowpack to human effects. *Journal of Climate*, 21(23), 6425–6444. <https://doi.org/10.1175/2008JCLI2405.1>
- Raleigh, M. S., & Small, E. E. (2017). Snowpack density modeling is the primary source of uncertainty when mapping basin-wide SWE with lidar. *Geophysical Research Letters*, 44(8), 3700–3709. <https://doi.org/10.1002/2016GL071999>
- 850 Rovaneck, R. J., Kane, D. L., & Hinzman, L. D. (1993). Improving estimates of snowpack water equivalent using double sampling. In *Proceedings of the 61st Western snow conference* (pp. 157–163).
- Schwank, M., Matzler, C., Wiesmann, A., Wegmuller, U., Pulliainen, J., Lemmetyinen, J., Rautiainen, K., Derksen, C., Toose, P., & Drusch, M. (2015). Snow Density and Ground Permittivity Retrieved from L-Band Radiometry: A Synthetic

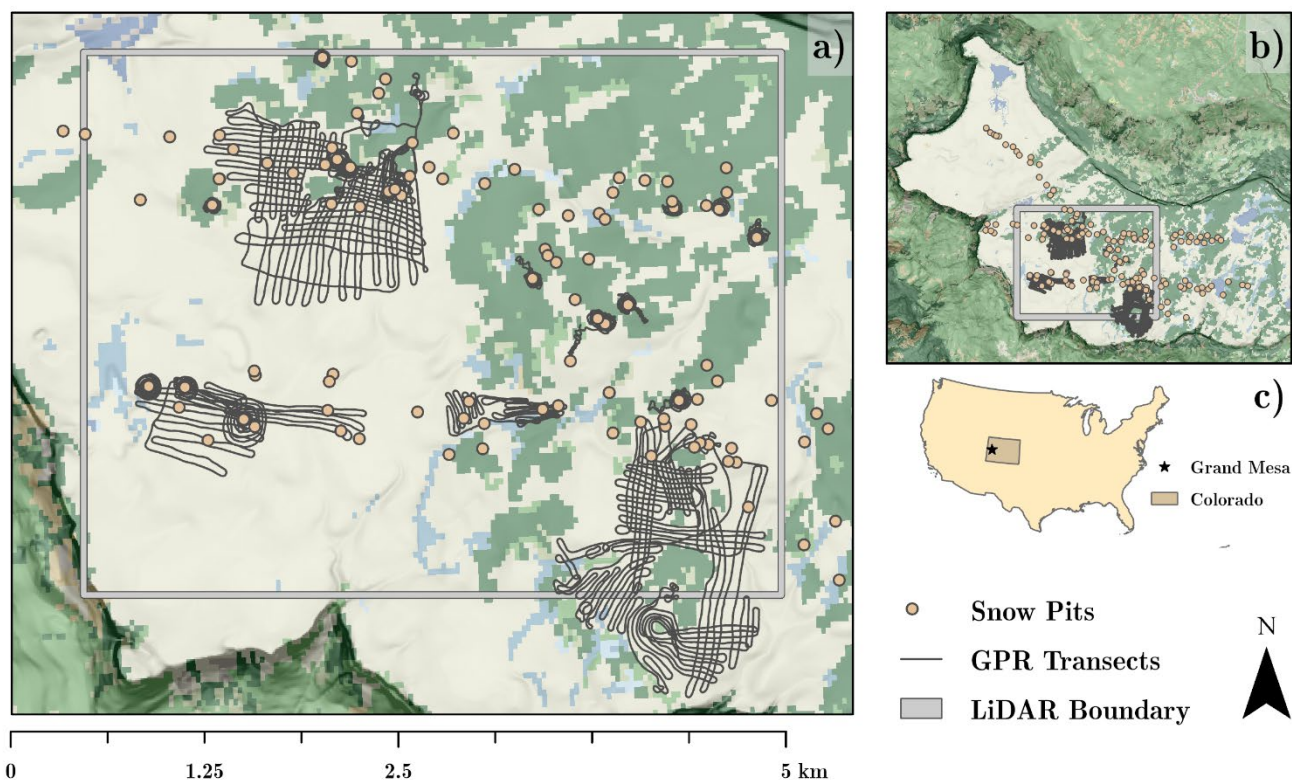


- 855 Analysis. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(8), 3833–3845.  
https://doi.org/10.1109/JSTARS.2015.2422998
- Siirila-Woodburn, E. R., Rhoades, A. M., Hatchett, B. J., Huning, L. S., Szinai, J., Tague, C., Nico, P. S., Feldman, D. R.,  
Jones, A. D., Collins, W. D., & Kaatz, L. (2021). A low-to-no snow future and its impacts on water resources in the  
western United States. *Nature Reviews Earth and Environment*, 2(11), 800–819. https://doi.org/10.1038/s43017-021-  
860 00219-y
- Štroner, M., Urban, R., Lidmila, M., Kolář, V., & Křemen, T. (2021). 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  
Sensing*, 13(15), 3050. https://doi.org/10.3390/rs13153050
- Sturm, M., & Holmgren, J. (1998). Differences in compaction behavior of three climate classes of snow. *Annals of Glaciology*,  
865 26, 125–130. https://doi.org/10.3189/1998AoG26-1-125-130
- Sturm, M., Taras, B., Liston, G. E., Derksen, C., Jonas, T., & Lea, J. (2010). Estimating snow water equivalent using snow  
depth data and climate classes. *Journal of Hydrometeorology*, 11(6), 1380–1394.  
https://doi.org/10.1175/2010JHM1202.1
- Tedesco, M., Reichle, R., Low, A., Markus, T., & Foster, J. L. (2010). Dynamic Approaches for Snow Depth Retrieval From  
870 Spaceborne Microwave Brightness Temperature. *IEEE Transactions on Geoscience and Remote Sensing*, 48(4), 1955–  
1967. https://doi.org/10.1109/TGRS.2009.2036910
- Trujillo, E., Ramírez, J. A., & Elder, K. J. (2009). Scaling properties and spatial organization of snow depth fields in sub-  
alpine forest and alpine tundra. *Hydrological Processes*, 23(11), 1575–1590. https://doi.org/10.1002/hyp.7270
- Valence, E., Baraer, M., Rosa, E., Barbecot, F., & Monty, C. (2022). Drone-based ground-penetrating radar (GPR) application  
875 to snow hydrology. *The Cryosphere*, 16(9), 3843–3860. https://doi.org/10.5194/tc-16-3843-2022
- Vecherin, S., Meyer, A., Quinn, B., Letcher, T., & Parker, M. (2022, August 16). Simulation of Snow Texture for Autonomous  
Vehicle Numerical Modeling. *Ground Vehicle Systems Engineering and Technology Symposium (GVSETS)*.  
http://gvsets.ndia-mich.org/publication.php?documentID=928
- Vuyovich, C. M., Marshall, H. P., Elder, K., Hiemstra, C., Brucker, L., & McCormick, M. (2021). *SnowEx20 Grand Mesa  
880 Intensive Observation Period Snow Pit Measurements, Version 1*. Boulder, Colorado USA. NASA National Snow and  
Ice Data Center Distributed Active Archive Center. https://doi.org/https://doi.org/10.5067/DUD2VZEVBJ7S
- Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image Quality Assessment: From Error Visibility to  
Structural Similarity. *IEEE Transactions on Image Processing*, 13(4), 600–612.  
https://doi.org/10.1109/TIP.2003.819861
- 885 Webb, R. W., Raleigh, M. S., McGrath, D., Molotch, N. P., Elder, K., Hiemstra, C., Brucker, L., & Marshall, H. P. (2020).  
Within-Stand Boundary Effects on Snow Water Equivalent Distribution in Forested Areas. *Water Resources Research*,  
56(10). https://doi.org/10.1029/2019WR024905
- Webb, Ryan W. (2021). *SnowEx20 Grand Mesa IOP UNM 800 and 1600 MHz MALA GPR, Version 1*. Boulder, Colorado

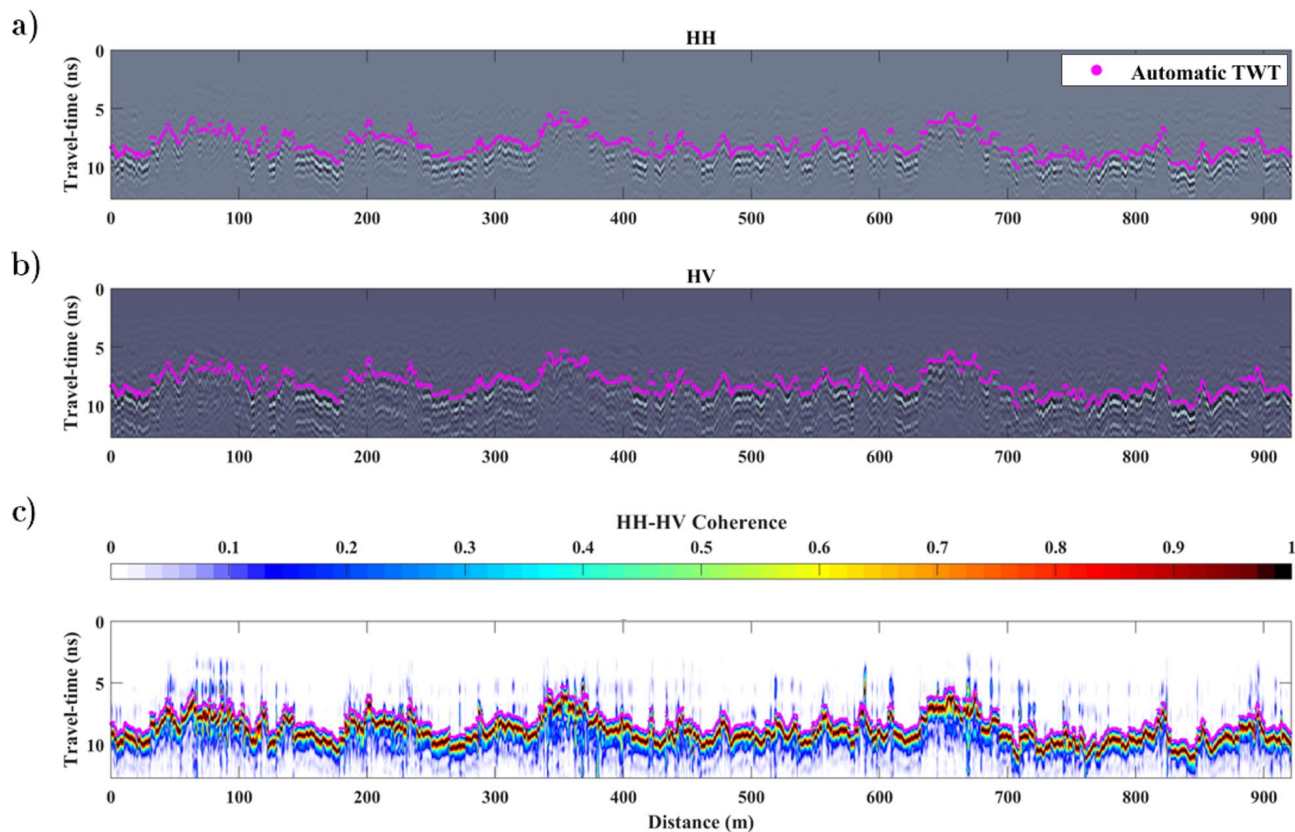


- 890 USA. NASA National Snow and Ice Data Center Distributed Active Archive Center.  
<https://doi.org/https://doi.org/10.5067/WE9GI1GVMQF6>
- Webb, Ryan W., Marziliano, A., McGrath, D., Bonnell, R., Meehan, T. G., Vuyovich, C., & Marshall, H.-P. (2021). In Situ Determination of Dry and Wet Snow Permittivity: Improving Equations for Low Frequency Radar Applications. *Remote Sensing*, 13(22), 4617. <https://doi.org/10.3390/rs13224617>
- 895 Wetlaufer, K., Hendrikx, J., & Marshall, L. (2016). Spatial heterogeneity of snow density and its influence on snow water equivalence estimates in a large mountainous basin. *Hydrology*, 3(1). <https://doi.org/10.3390/hydrology3010003>
- Wharton, R. P., Hazen, G. A., Rau, R. N., & Best, D. L. (1980, April 4). Advancements In Electromagnetic Propagation Logging. *SPE Rocky Mountain Regional Meeting*. <https://doi.org/10.2118/9041-MS>
- Winstral, A., Elder, K., & Davis, R. E. (2002). Spatial Snow Modeling of Wind-Redistributed Snow Using Terrain-Based Parameters. *Journal of Hydrometeorology*, 3(5), 524–538. [https://doi.org/10.1175/1525-7541\(2002\)003<0524:SSMOWR>2.0.CO;2](https://doi.org/10.1175/1525-7541(2002)003<0524:SSMOWR>2.0.CO;2)
- 900 Yan, J. B., Gomez-Garcia Alvestegui, D., McDaniel, J. W., Li, Y., Gogineni, S., Rodriguez-Morales, F., Brozena, J., & Leuschen, C. J. (2017). Ultrawideband FMCW radar for airborne measurements of snow over sea ice and land. *IEEE Transactions on Geoscience and Remote Sensing*, 55(2), 834–843. <https://doi.org/10.1109/TGRS.2016.2616134>
- 905 Yildiz, S., Akyurek, Z., & Binley, A. (2021). Quantifying snow water equivalent using terrestrial ground penetrating radar and unmanned aerial vehicle photogrammetry. *Hydrological Processes*, 35(5), 1–15. <https://doi.org/10.1002/hyp.14190>



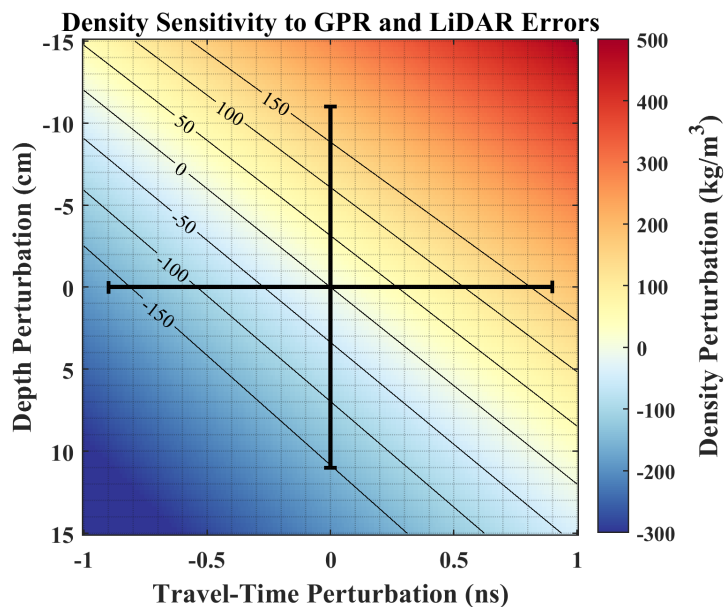


910  
Figure 1: a) Study area map of snow pit locations (yellow circles), GPR transects (black lines), and LiDAR boundary (grey wireframe). Landcover classification identifies the forested areas as green and lakes as blue. b) inset map of Grand Mesa, Colorado depicting the extent of the dataset acquired 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.



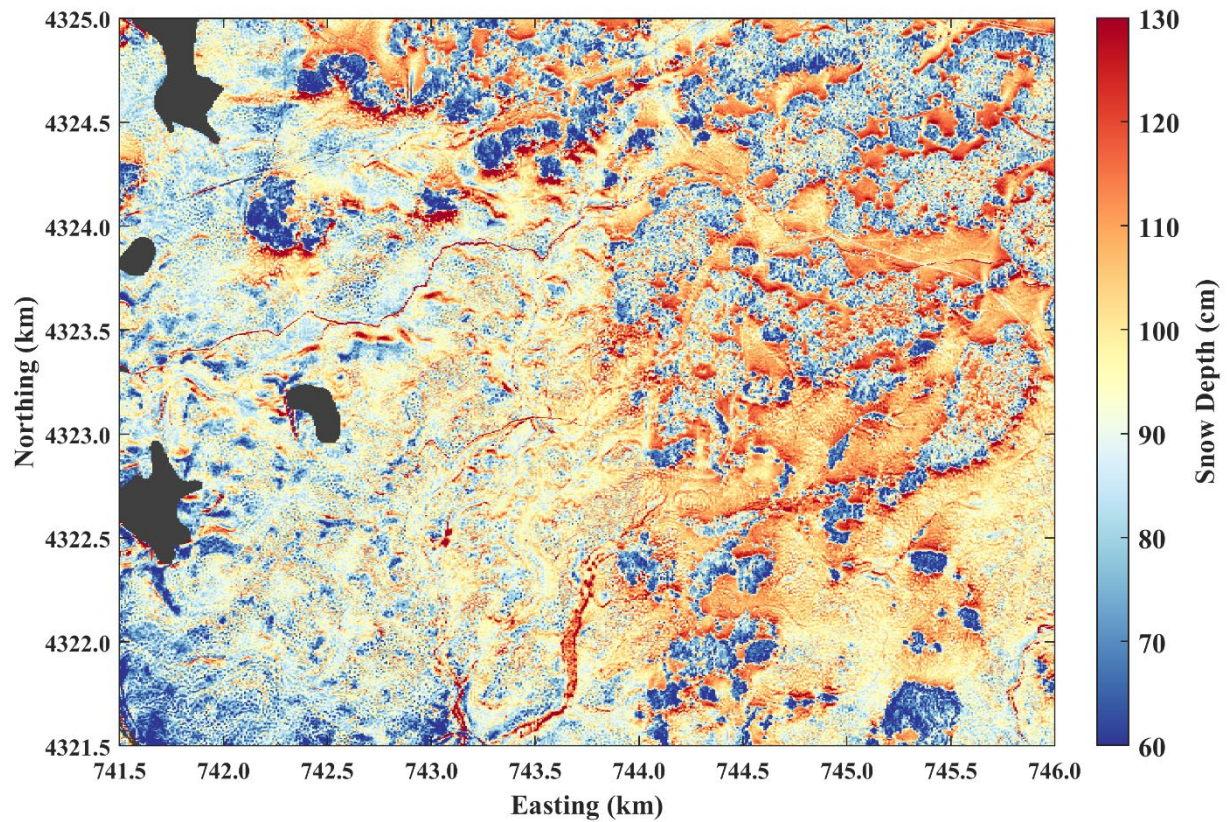
915

Figure 2: 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).



920

**Figure 3:** 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 and the RMSD of the GPR TWT crossovers. At one standard deviation, combined snow density errors of  $\pm 150 \text{ kg/m}^3$  can be expected from sensor integration, which are reduced to within  $\pm 30 \text{ kg/m}^3$  by outlier filtering.



925 **Figure 4: 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).**

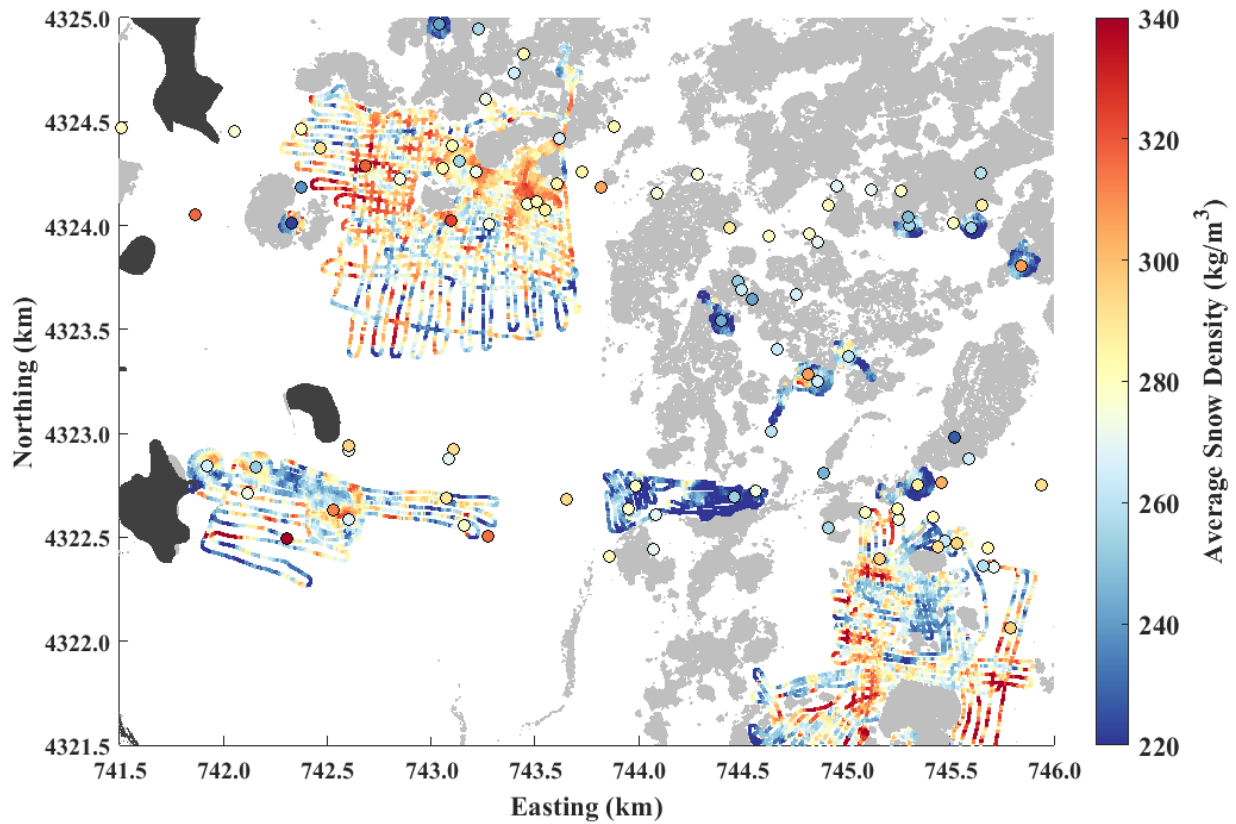
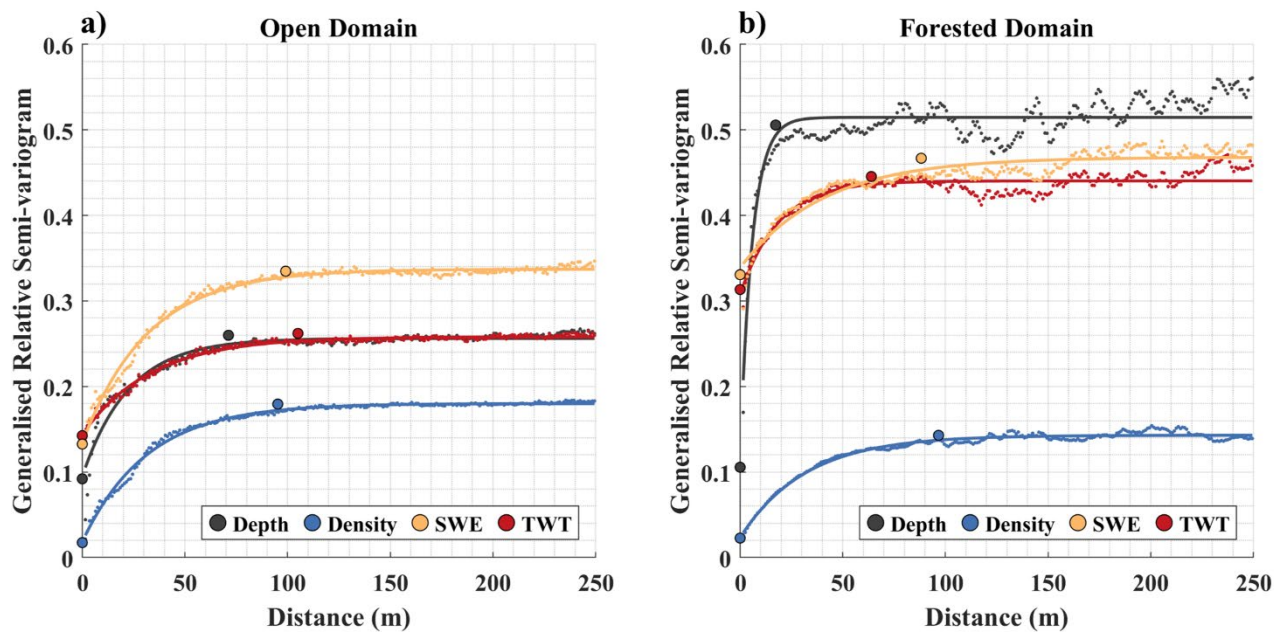
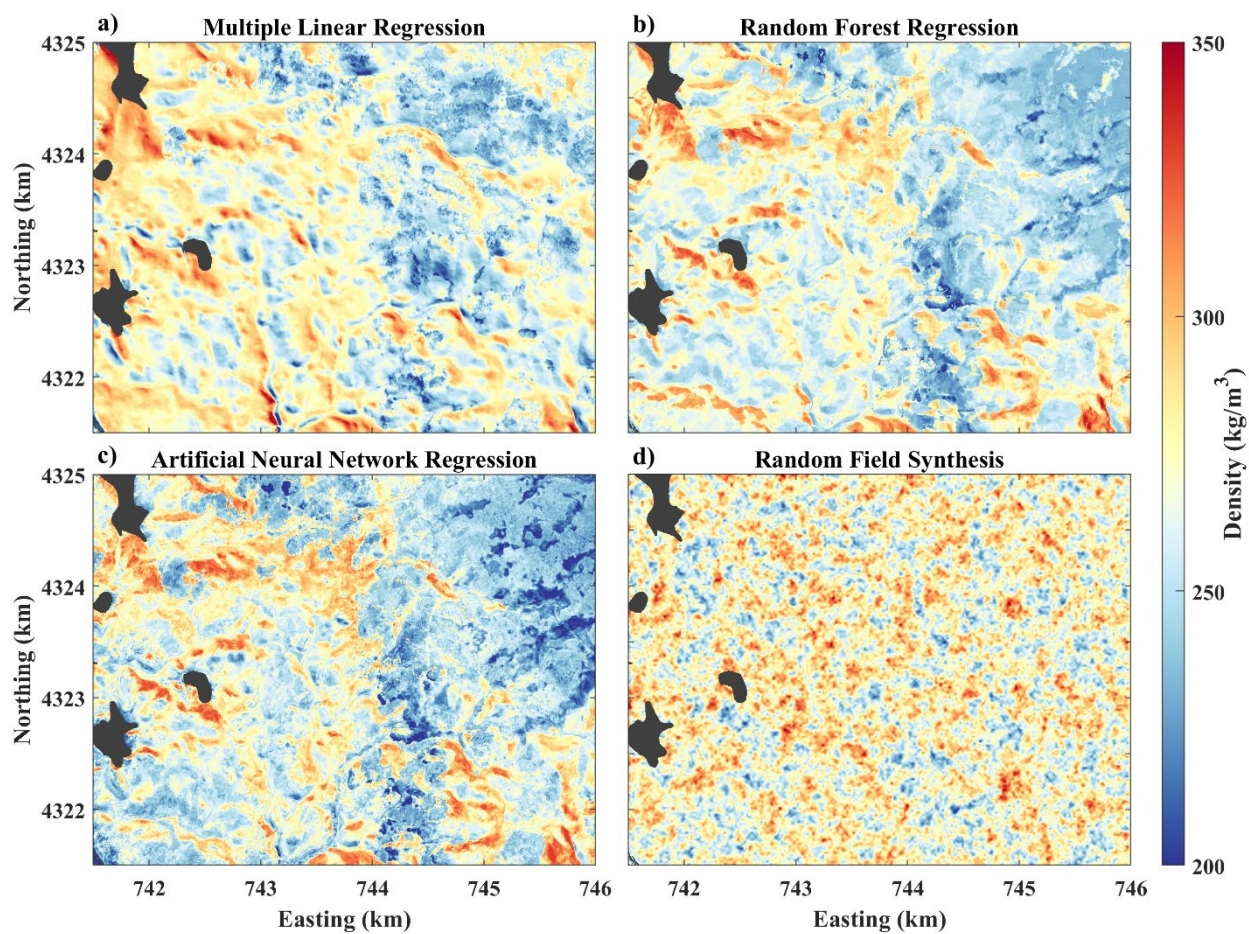


Figure 5: Bulk snow density along radar profiles estimated by combining LiDAR snow depths with GPR TWTs. Average density measured in the 96 snow pits within the LiDAR boundary are overlaid as larger markers. Forested areas (grey) and lakes (black) are shown.



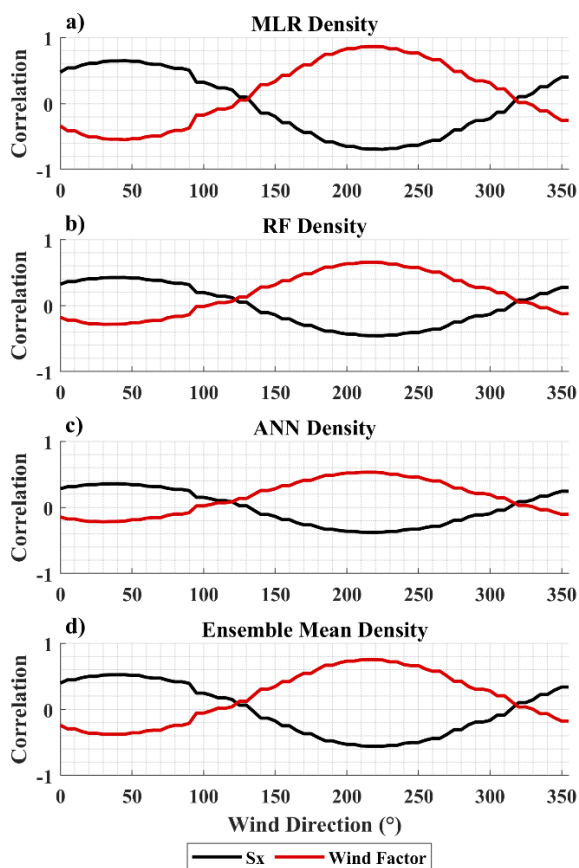
930

Figure 6: Generalised relative semi-variograms in a) open and b) forested areas for LiDAR snow depth, GPR TWT, average density inferred 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.



935

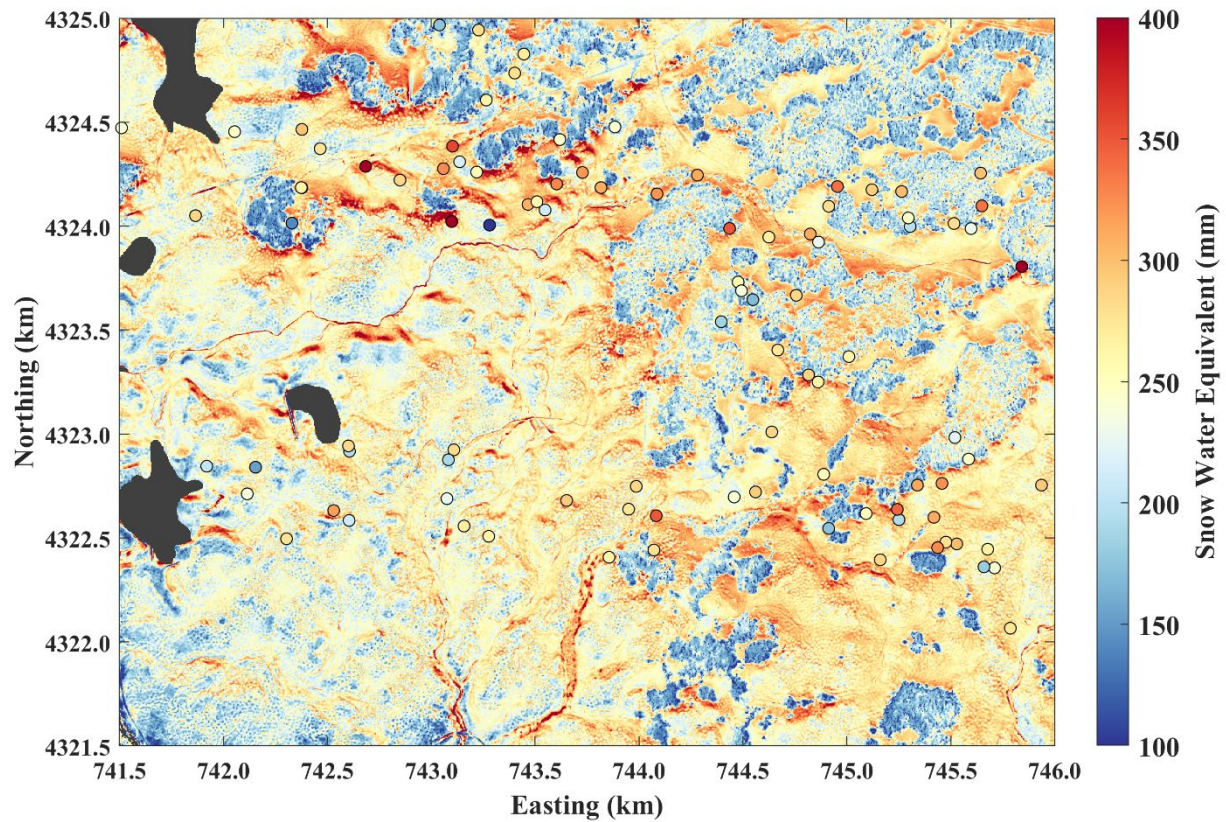
Figure 7: Snow density was distributed spatially using regression techniques a) MLR, b) RF, and c) ANN based on predictive features that were derived solely from LiDAR information (i.e. snow depth, elevation, slope, aspect, and vegetation) and d) a synthesised random field model from the statistics of in situ density measurements ( $275 \pm 20 \text{ kg/m}^3$ ) and the correlation length of snow density estimated via variogram analysis.



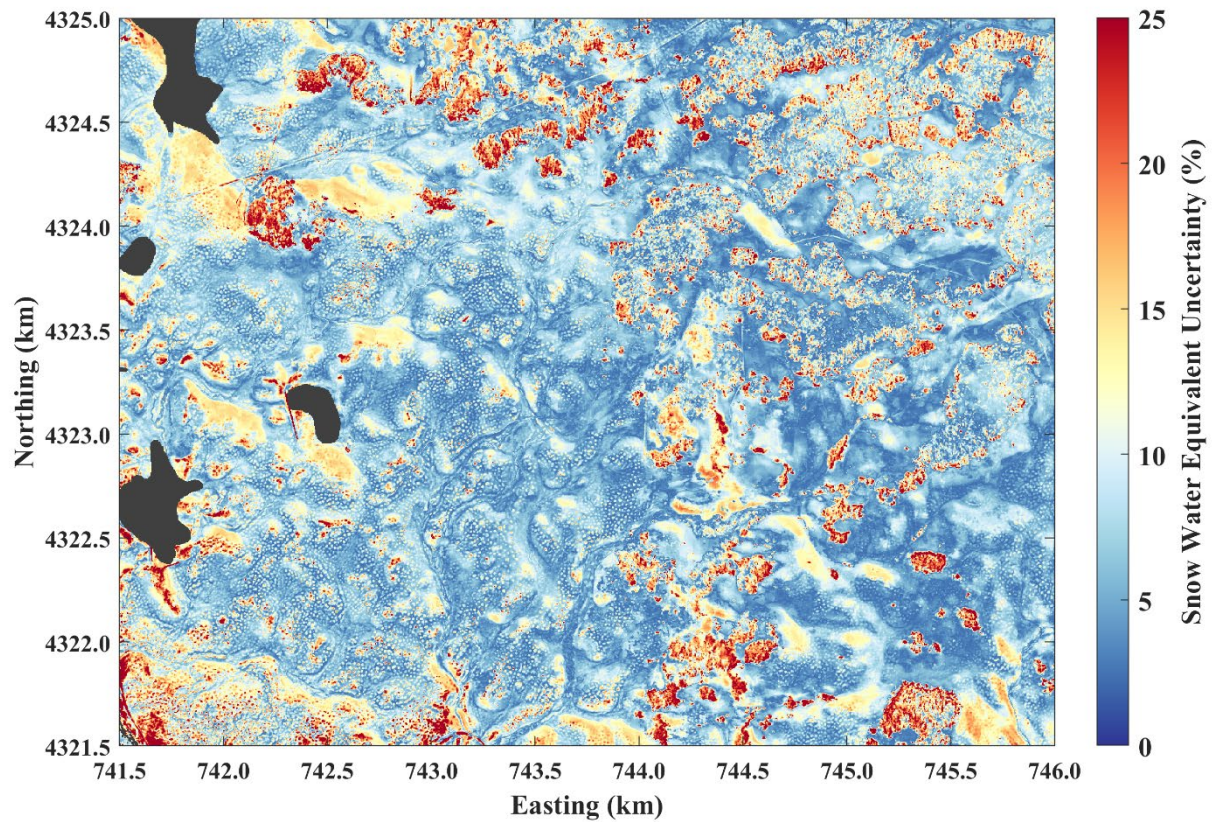
940

**Figure 8: Strong prevailing south-southwest winds (180° – 220°) provide a causal physical mechanism for correlations between the maximum upwind slope (Sx) and wind factor parameters calculated in 5-degree increments.**





945 **Figure 9: Spatially distributed snow water equivalent estimated using the regression ensemble mean density and LiDAR snow depths. 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.**



950 **Figure 10: Estimated uncertainty in snow water equivalent from the propagation of errors in snow depth and density. The spatial distribution of uncertainty tends to be greatest in the shallower and lower density snow underneath tree canopies and least in the deepest snow caught in drifts around the perimeters of forest stands.**

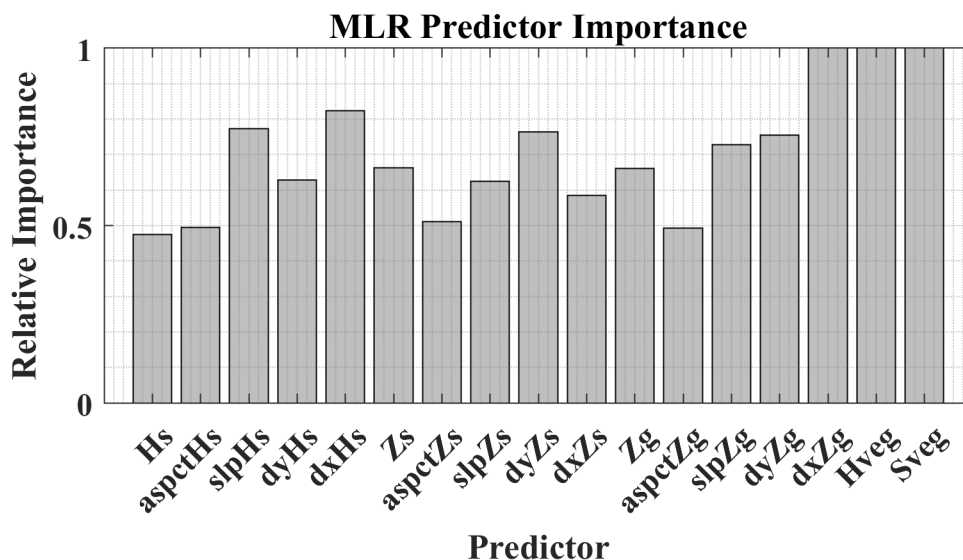


Figure A.1: The relative importance of the LiDAR derived predictors: Hs snow depth, aspcHs aspect of snow depth, slpHs slope of snow depth, dyHs north component of snow depth gradient, dxHs east component of snow depth gradient, Zs the snow surface elevation and derivatives, Zg the ground elevation and 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.

955

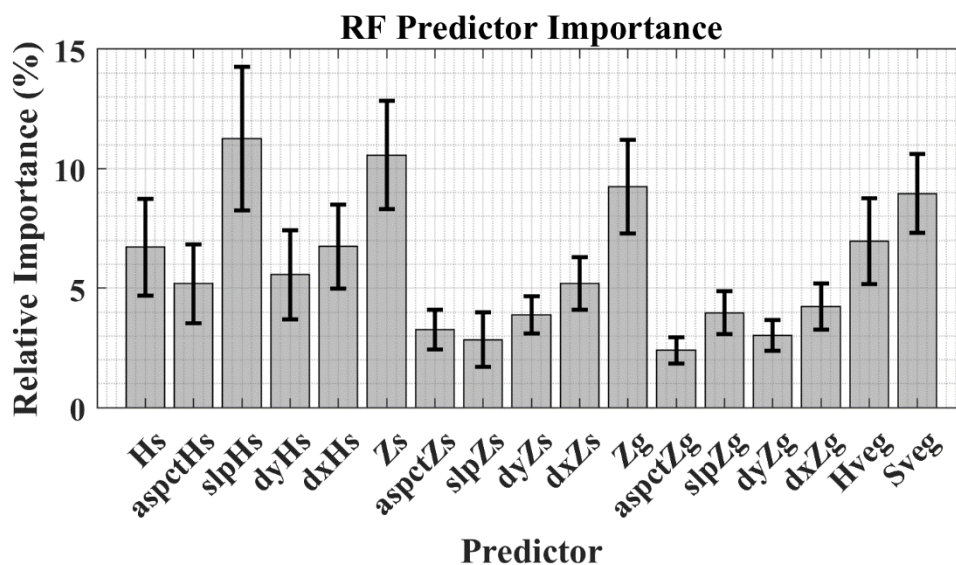
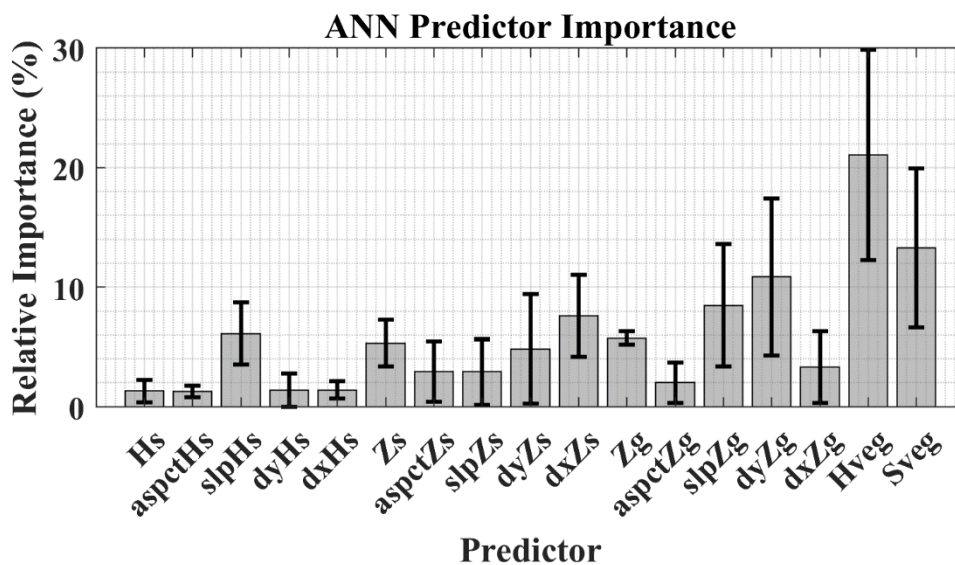


Figure A.2: Relative importance of LiDAR predictors calculated using the “out of bag” technique for random forest regression. Uncertainties were developed using random subsets of training data in a Monte Carlo simulation.



960

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

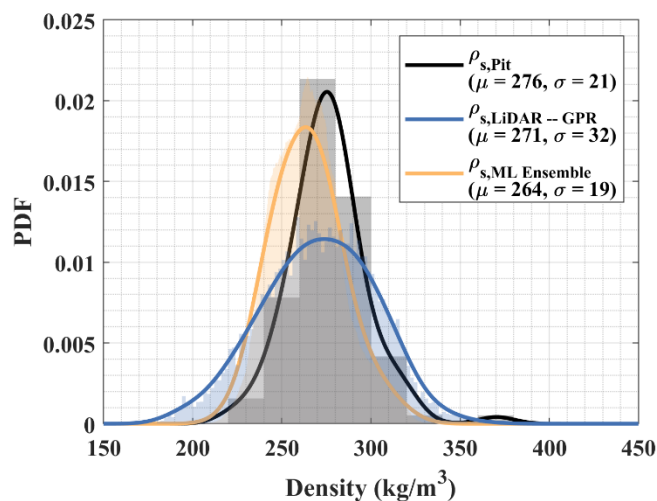
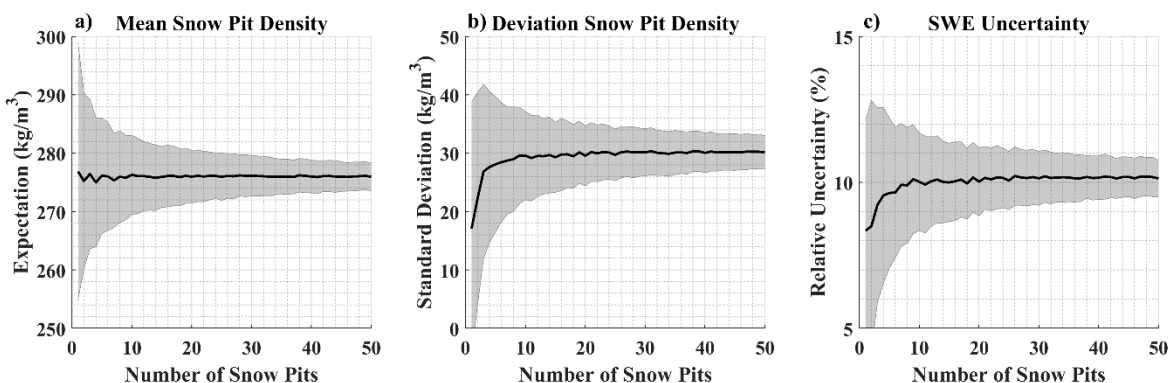
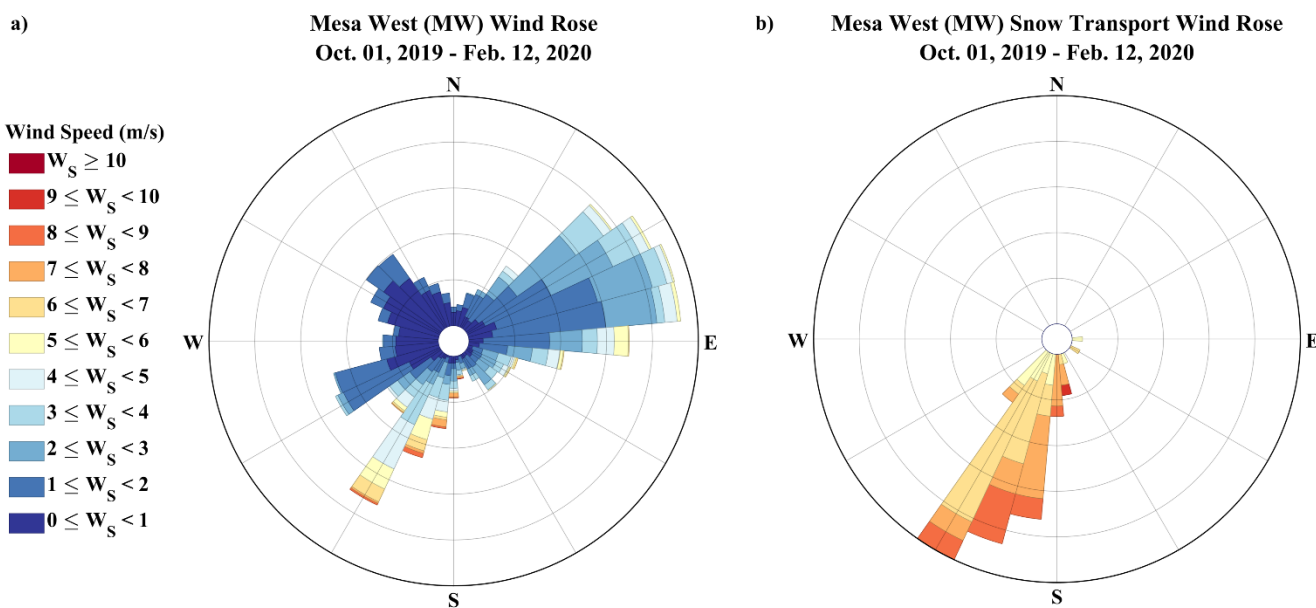


Figure B.1: Histograms of the snow pit measured, LiDAR – GPR inferred, and regression model ensemble densities.



965 **Figure B.2: 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.**



970 **Figure C.1: Wind speed and direction data from the Mesa West meteorological station (Houser et al., 2022). a) wind rose spanning the beginning of the hydrological year through the end of the SnowEx IOP. b) wind rose of winds that are strong enough to transport snow (Li & Pomeroy, 1997) spanning the beginning of the hydrological year through the end of the SnowEx IOP. The median wind direction for snow transport is 200°.**



**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.

975

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

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 ( $H_{s,LiDAR}$ )	96 ± 16   0.61   11   0	94 ± 16   0.00   22   -4	87 ± 14   0.00   23   3
GPR ( $h_{s,E\overline{n}s-GPR}$ )	98 ± 17   0.48   18   1	97 ± 14   0.01   21   -4	89 ± 12   0.01   24   6

980

**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.

985

a)

Snow Density (kg/m <sup>3</sup> ) Method	All Domain $\mu \pm \sigma$	Open Domain $\mu \pm \sigma$	Forested Domain $\mu \pm \sigma$
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 ( $\rho_{s,E\overline{n}s}$ )	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 ( $\rho_{s,E\overline{n}s}$ )	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			Open Domain			Forested Domain		
	R <sup>2</sup>	RMSE	Bias	R <sup>2</sup>	RMSE	Bias	R <sup>2</sup>	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

990

**Table 3:** Variogram parameters are overviewed for the forested and open areas of Grand Mesa. a) The nugget indicates the percentage amount of measurement variability relative to the mean between co-located observations. b) The sill indicates the percentage variability relative to the mean at which measurements are no longer correlated (coefficient of variation). c) The correlation length is the distance in metres above which measurements are no longer correlated.

995

a)

Variogram Parameter: Nugget (% of mean)	Open Domain $\mu \pm \sigma$	Forested Domain $\mu \pm \sigma$
Depth ( $h_{s,LiDAR}$ )	9 ± 0 %	11 ± 2 %
Density ( $\rho_{s,LiDAR-GPR}$ )	2 ± 0 %	2 ± 0 %
SWE ( $b_{s,LiDAR-GPR}$ )	13 ± 0 %	33 ± 1 %
TWT ( $\tau_{s,GPR}$ )	14 ± 0 %	31 ± 1 %

b)

Variogram Parameter: Sill (% of mean)	Open Domain $\mu \pm \sigma$	Forested Domain $\mu \pm \sigma$
Depth ( $h_{s,LiDAR}$ )	26 ± 0 %	51 ± 2 %
Density ( $\rho_{s,LiDAR-GPR}$ )	18 ± 0 %	14 ± 0 %
SWE ( $b_{s,LiDAR-GPR}$ )	33 ± 1 %	47 ± 1 %
TWT ( $\tau_{s,GPR}$ )	26 ± 0 %	45 ± 0 %

1000

c)

Variogram Parameter: Correlation Length (m)	Open Domain $\mu \pm \sigma$	Forested Domain $\mu \pm \sigma$
Depth ( $h_{s,LiDAR}$ )	71 ± 4	17 ± 2
Density ( $\rho_{s,LiDAR-GPR}$ )	95 ± 2	97 ± 4
SWE ( $b_{s,LiDAR-GPR}$ )	99 ± 5	88 ± 17
TWT ( $\tau_{s,GPR}$ )	105 ± 6	64 ± 5

**Table 4:** 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<sup>3</sup>), the average of snow pit density in each respective domain (276, 280, 257 kg/m<sup>3</sup>), 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

1005



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 learned-regression 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 <sup>3</sup> )	255 ± 51	266 ± 41	219 ± 62
Domain Pit Avg. (276, 280 & 257)	255 ± 51	275 ± 37	197 ± 44
ML Ensemble ( $b_{s,LiDAR-EnS}$ )	245 ± 53	259 ± 41	198 ± 57
Random Field ( $b_{s,LiDAR-Rand}$ )	254 ± 54	265 ± 45	218 ± 64

1010

b)

Snow Water Equivalent (mm) Method	All Domain R <sup>2</sup>   RMSE   Bias	Open Domain R <sup>2</sup>   RMSE   Bias	Forested Domain R <sup>2</sup>   RMSE   Bias
Pit Average (276, 280 & 257)	0.57   38   -9	0.63   34   -3	0.16   55   -29
ML Ensemble ( $b_{s,LiDAR-EnS}$ )	0.56   42   -18	0.61   38   -15	0.19   56   -32
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<sup>2</sup> 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 <sup>2</sup> , 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

1015

Table C.1: The maximum correlation and wind direction between snow density and the maximum upwind slope (Sx) and wind factor parameters. Data corresponds to Fig. 8.

Snow Density Method	Sx		Wind Factor	
	R	Direction	R	Direction
LiDAR – GPR	-0.45	225	0.48	220
MLR	-0.69	225	0.86	220
RF	-0.46	225	0.65	215
ANN	-0.38	220	0.53	220
ML Ensemble	-0.56	225	0.75	220
Random Field	-0.03	180	0.02	170