Water Resources Research

RESEARCH ARTICLE
10.1029/2018WR024146

Key Points:
- It is important to consider the spatial variability of snow density to create spatially distributed maps of SWE from lidar data
- Relative to that of snow depth, snow density variability is higher earlier in the snow season
- The representativeness of snow depth, snow density, and SWE observations from SNOTEL stations varies at different points in the snow season

Supporting Information:
- Supporting Information S1

Correspondence to:
P. D. Broxton,
broxtopd@email.arizona.edu

Citation:
Broxton, P. D., van 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*. https://doi.org/10.1029/2018WR024146

Received 24 SEP 2018
Accepted 6 APR 2019
Accepted article online 18 APR 2019

©2019. American Geophysical Union. All Rights Reserved.

Improving Snow Water Equivalent Maps With Machine Learning of Snow Survey and Lidar Measurements

Patrick D. Broxton[1], Willem J. D. van Leeuwen[1,2], and Joel A. Biederman[3]

[1]School of Natural Resources and the Environment, University of Arizona, Tucson, AZ, USA, [2]School of Geography and Development, University of Arizona, Tucson, AZ, USA, [3]Southwest Watershed Research Center, USDA ARS, Tucson, AZ, USA

Abstract In the semiarid interior western USA, where a majority of surface water supply comes from mountain forests, high-resolution aerial lidar-based surveys are commonly used to study snow. These surveys provide rich information about snow depth, but they are usually not accompanied with spatially explicit measurements of snow density, which leads to uncertainty in the estimation of snow water equivalent (SWE). In this study, we use a novel approach to distribute ~300 field measurements of snow density with artificial neural networks. We combine the resulting density maps with aerial lidar snow depth measurements, bias corrected with a very large and precisely geolocated array of field-measured snow depths (~4,000 observations), to create and validate maps of snow depth, snow density, and SWE over two sites along Arizona's Mogollon Rim in February and March 2017. These maps show differences between midwinter and late-winter snow conditions. In particular, compared to that of snow depth, the spatial variability of snow density is smaller for the later snow survey than the earlier snow survey. These gridded data also show that the representativeness of Snow Telemetry and other point measurements is different for the midwinter and late-winter snow surveys. Overall, the lidar artificial neural network SWE estimates can be as much as 30% different than if Snow Telemetry density were used with lidar snow depths to estimate SWE.

Plain Language Summary In the western USA, a majority of surface water originates from mountain snowmelt. Knowing the quantity of water in the snowpack, called snow water equivalent (SWE), is critical for water supply forecasts and management of rivers and streams for water delivery and hydropower. In this study, we develop a new method to estimate SWE by combining aerial remote sensing maps of snow depth with snow density maps generated through machine learning of hundreds of field measurements of snow density. This study finds that on a given date, snow density can vary widely, highlighting the importance of considering its spatial variability when estimating SWE. These gridded data show that the representativeness of Snow Telemetry and other point measurements is different for the midwinter versus late winter snow surveys. In addition, we show that using spatially variable maps of snow density can impact watershed-scale SWE estimates by up to 30% as compared to using snow density measurements from commonly used snow monitoring stations. The method described in this study will be useful for generating SWE estimates for water supply monitoring, evaluating snow models, and understanding how changing mountain forests might impact SWE.

1. Introduction

Melt water from snowpacks in mountain forests such as those in the western USA is vital to water supply, as it is a dominant source of water for many mountain streams (Kampf & Lefsky, 2016) and can account for more than 75% of water resources used for agriculture, human consumption, and hydroelectric power generation for downstream areas (Bales et al., 2006; Mankin et al., 2015). However, our understanding of mountain snowpacks is challenged by a sparsity of high quality, spatially explicit snow observations. On one hand, point measurements of snowpack, which are relatively easy to make, can be a poor indicator of snowpack over larger areas in mountainous environments because of the extreme heterogeneity of snow (Guan et al., 2013; Molotch & Bales, 2005). On the other hand, snowcover data sets produced by moderate-spatial-resolution multispectral and especially coarser-resolution microwave satellite remote sensing can be unreliable in mountainous regions due to many issues including cloud cover, heterogeneous forest cover, terrain variability, and incomplete snowcover (Dietz et al., 2012; Dozier et al., 2016; Foster et al., 2005;

Lettenmaier et al., 2015; Nolin, 2010). Snow monitoring for smaller regions (e.g., managed watersheds) can be more reliably done using statistical snow modeling along with multitemporal fine spatial resolution data (Czyzowska-Wisniewski et al., 2015).

Recently, observations using light detection and ranging (lidar) from fixed-wing aircraft are increasingly being used to measure snow depth in mountainous environments (e.g., Baños et al., 2011; Cline et al., 2009; Dadic et al., 2010; DeBeer & Pomeroy, 2010; Deems et al., 2006; Fassnacht & Deems, 2006; Grünewald et al., 2013; Kim et al., 2018; Painter et al., 2016; Tinkham et al., 2014; Trujillo et al., 2007). Lidar is ideal for measuring snow in forested environments because individual lidar pulses can penetrate a canopy and create multiple measurement returns (Baltsavias, 1999). In addition, lidar data has sufficiently high spatial resolution to observe the effects of fine-scale snow-forest interactions such as snow variability associated with forest gaps and edges (Golding & Swanson, 1986; Molotch et al., 2009; Musselman et al., 2008; Veatch et al., 2009). Lidar snow maps have been applied for water resource understanding (e.g., Henn et al., 2018) as well as to train and validate snow models (e.g., Bair et al., 2016; Broxton et al., 2015; Hedrick et al., 2015), which can be used for hydrological monitoring, predicting how regional warming may influence water supply (Foster et al., 2016; Mankin et al., 2015), and quantifying feedbacks between water availability and forest disturbance.

Despite its strengths, there are challenges affecting lidar snow surveys. For example, lidar data are sometimes evaluated in a statistical rather than spatially explicit fashion. In addition, they are sometimes paired with ground surveys conducted on different days from the lidar. However, probably the largest challenge facing most lidar snow surveys is a lack of spatially explicit snow density data from which to make maps of SWE (Raleigh & Small, 2017; Wetlaufer et al., 2016), which is the snow variable of greatest interest for most hydrological applications (Dozier, 2011). Some studies do not consider the spatial variability of snow density (e.g., Molotch & Bales, 2005), which is known to differ across a landscape due to spatial variability in several processes (e.g., compaction, melt refreeze, and wind-caused densification) (López-Moreno et al., 2013; Wetlaufer et al., 2016). Because of this, the spatial distribution of snow density would also be expected to change through time. Furthermore, not accounting for this spatial variability might lead to biases of snow density estimates if too few or poorly distributed snow density observations are used.

One method to account for these spatial and temporal variations in snow density is to use a physically based snow density model to convert the lidar snow depth observations to SWE. Most prominently, NASA's Airborne Snow Observatory (Painter et al., 2016) uses an energy balance snow model (Marks et al., 1999) with spatially variable snow density estimates to convert lidar depths to SWE (Hedrick et al., 2018). However, there are uncertainties with many snow density models themselves (Avanzi et al., 2014; Dawson et al., 2017; Raleigh & Small, 2017) which can lead to error in SWE estimates. For example, ASO uses observed values of snow density to remove both elevation-specific and overall biases from the modeled snow density fields (Painter et al., 2016). These biases indicate that there is still substantial uncertainty when using these models to estimate snow density across lidar domains.

Another common approach is to generate maps of estimated snow variables using statistical models trained by field measurements of these same variables paired with geophysical or climatic predictor variables. These statistical models, which include a variety of approaches including regression models (such as multiple linear regression—MLR), binary regression trees, and lookup tables, have been applied using observations that span both larger (e.g., continental) scales (Bormann et al., 2013; Sturm et al., 1995, 2010) and smaller (e.g., watershed) scales (Jonas et al., 2009; Wetlaufer et al., 2016). Other machine learning approaches such as Random Forests and Artificial Neural Networks have also become popular for estimating snow quantities (particularly SWE and snow cover) using a variety of input data including data from satellite sensors (e.g., Bair et al., 2018; Dobreva & Klein, 2011; Tedesco et al., 2004), land surface models (e.g., Snauffer et al., 2018), and ground observations (e.g., Tabari et al., 2010; Buckingham et al., 2015; Gharaei-Manesh et al., 2016). These approaches have been shown to be highly adaptable to capture nonlinear relationships involved in snow measurement (Czyzowska-Wisniewski et al., 2015), allowing them to outperform linear approaches such as MLR. However, Artificial Neural Networks have not yet been applied to estimate snow density.

In this study, we use Artificial Neural Networks (ANNs) to estimate snow density across lidar domains: We create spatial maps of snow depth, snow density, and SWE by combining data from airborne lidar along and

extensive ground snow surveys (~4,000 manual snow depth measurements and ~300 snow density measurements) through ANN modelling, thus relying entirely on observational data. The ground snow survey data were precisely geolocated and explicitly associated with 1-m pixels of snow depth and other geophysical attributes derived from the lidar data such as aspect, forest cover, and solar forcing index. This enabled both explicit validation and improvement of the lidar snow depth data as well as generation of gridded snow density estimates.

These data allow us to assess how the spatial variability of snow density and SWE compare to that of snow depth at two different snapshots in time representing midwinter and late-winter (near peak SWE for the higher terrain) conditions. Additional temporal context is provided by a network of daily snow depth observations along transects spanning forest and adjacent clearings using time lapse photography (a.k.a. snowtography). These field observations also provide spatial context for point snow observations in terms of snow depth, snow density, and SWE. In particular, we evaluate the spatial representativeness of all three of these variables measured at Snow Telemetry (SNOTEL) stations (which are used by local water managers for snow monitoring and streamflow forecasting) at two different times in the snow season.

2. Materials and Methods

2.1. Study Areas

Our study areas include two ~100-km^2 snow-on lidar boxes (Figure 1) in the highlands of central Arizona (AZ), including a mid-elevation lidar box (located ~20-km north of Payson, AZ; Figure 2) and a high-elevation lidar box (located ~50-km east of Show Low, AZ; Figure 3). The mid-elevation lidar box has an environment typical of much of the seasonally snow-covered highlands of northern and central AZ. Its elevations range from 2,100- to 2,400-m above sea level (a.s.l.; Figure 2a), and it is relatively densely forested with Ponderosa Pine (*Pinus ponderosa*) forests (the average canopy closure based on lidar data is 56.2%; Figure 2d). Precipitation amounts from December to March, which encompasses the typical snow-covered period, range from 360 mm at lower elevations to 470 mm at the higher elevations (estimated from PRISM climate data). During this period, most precipitation falls as snow, yet temperatures are mild for alpine seasonally snow-covered environments, averaging 3.1 °C at a SNOTEL station (Baker Butte SNOTEL, 2,225 m a.s.l.). Within the mid-elevation lidar box, there are five intensively sampled field sites including snow survey transects and other monitoring data used in this study (Figure 1).

The high-elevation lidar box has an environment typical of the highest elevation areas in central and northern AZ. It contains a slightly larger range of elevations than the mid-elevation box, with low elevations in the south and high elevations in the north (ranging from ~2,400 to 2,950 m a.s.l., Figure 3a). The forest is more mixed with *Pinus ponderosa* forests dominating the mid-elevation southern end of the box and mixed conifer forests with stands of Aspen (*Populus tremuloides*) at the higher elevations. There is also a significant amount of burned area (from a 2011 wildfire) as well as preexisting large montane meadows in the central and northern regions, resulting in less forest canopy cover (32.4%; Figure 3d) than in the mid-elevation box. Overall, the climate is colder and drier than the mid-elevation box. December–March average temperature and precipitation is 1.3 °C and 232 mm at a SNOTEL located on the southern end of the lidar box (Wildcat SNOTEL, 2,393 m a.s.l.), and − 1.0 °C and 278 mm at a SNOTEL located on the northern end of the lidar box (Maverick Fork SNOTEL, 2,804 m a.s.l). Although the high-elevation box receives less winter precipitation than the mid-elevation box, it accumulates more snow because of its colder temperatures. Within the high-elevation lidar box, there are two intensively sampled field sites (Figure 1).

2.2. Data

2.2.1. Lidar Data

Two lidar flights (by Quantum Spatial Inc. or QSI) were flown on 1 February and 7 March 2017 for each box. The first flight captured peak snowpack conditions for the mid-elevation lidar box and midwinter snowpack conditions for the high-elevation lidar box and the second flight captured midablation snowpack conditions for the mid-elevation box and peak-snowpack conditions for portions of the high-elevation box.

The lidar surveys used a Leica ALS80 system mounted on a fixed wing aircraft flying at about 1,800-m above ground level. The 1 February lidar survey had an average first-return point density of 15.02 points/m^2 and an average ground classified point density of 4.40 points/m^2 (Quantum Spatial, 2017a). For the 7 March lidar

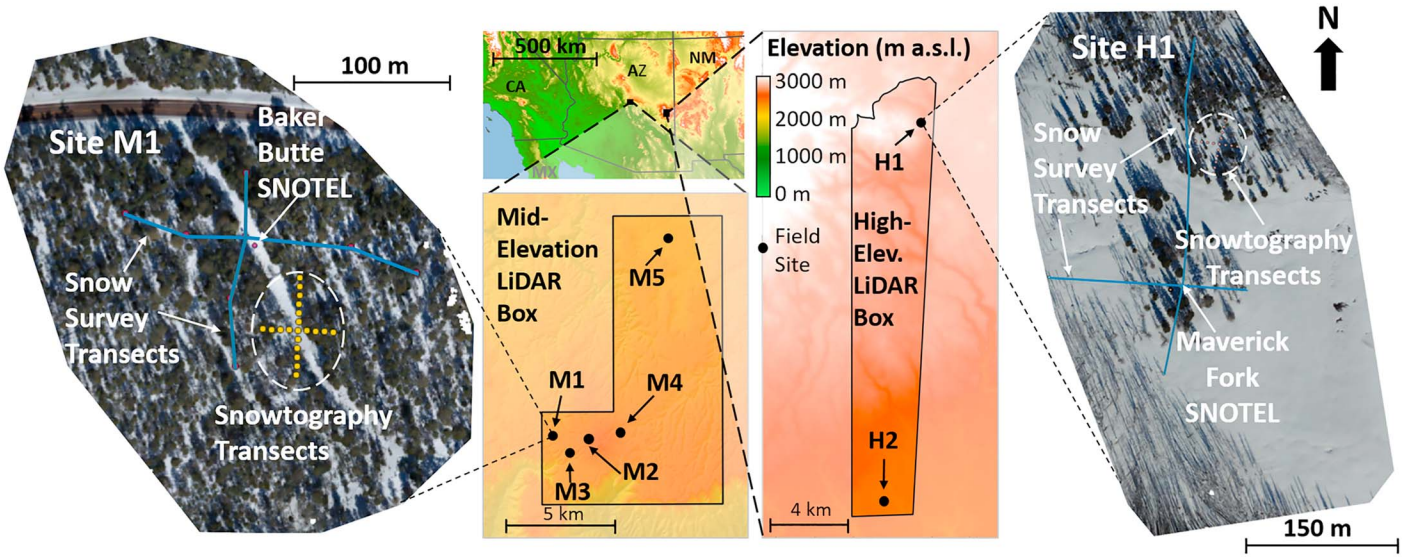

Figure 1. Site map showing the locations of the field data used in this study. The center panels show the locations of the mid-elevation and high-elevation lidar boxes, as well as the locations of the sites with ground snow survey data. The outer panels show site maps for the two most intensively studied research sites including the locations of snow survey transects, SNOTELs, and snowtography transects. SNOTEL = Snow Telemetry.

survey, the average first-return density was 12.18 points/m^2 and the average ground classified density was 3.68 points/m^2 (Quantum Spatial, 2017b). These snow-on survey data were compared with the snow-off lidar collection (also flown by QSI) in 2013 for an area encompassing the mid-elevation lidar box and 2014 for an area encompassing the high-elevation lidar box (Quantum Spatial, 2013, 2014). Point count return densities for the 2013 and 2014 data, respectively, were 9.38 and 15.41 points/m^2 (ground point counts averaged 3.97 and 5.95 points/m^2). For all collections, the absolute vertical accuracy was estimated to be on the order of 3–6 cm.

QSI did all of the processing tasks required to produce the lidar data deliverables (raw point cloud data as well as bare earth, highest hit, and intensity raster maps at 1-m resolution). Snow depth was computed as the difference between the bare-earth digital terrain model (DTM) generated using ground-classified lidar points in the 2013 and 2014 snow-off lidar data and those in the 2017 snow-on lidar collections.

2.2.2. Ground Snow Survey Data

We conducted snow survey field campaigns simultaneously with the lidar flights on both dates (within 24 hr of the lidar flights) involving more than 24 employees and volunteers from five public and private partner entities (see Acknowledgements). In this study, we use 3,860 measurements of snow depth, 300 measurements of snow density, and 17 snow pits to ground reference the lidar data and to support the development of SWE estimates across the lidar boxes. Surveys were designed to capture the widest feasible ranges of the landscape variables expected to regulate snow depth and density including elevation, slope, aspect, forest height, cover, and geometry (i.e., size and orientation of canopy and intercanopy gaps). For the 1 February campaign, we collected 2,965 snow depth probe and 245 snow tube (giving snow depth, density, and SWE) samples at all seven sites in Figure 1, and for the 7 March campaign, we collected 895 snow depth probe and 55 snow tube samples at M1, M2, and H1. Measurements were taken along premarked transects at each site (see examples in the inset maps in Figure 1).

An important advance of the present work is the precise geolocation of ground measurements, which allows spatially explicit (~1 m) comparison with lidar measurements of snow depth, SWE, terrain, and forest structure. Prior to the snow season, all endpoints of transect segments were marked and georeferenced using two methods. First, high-quality lidar-derived maps of canopy height were loaded onto mobile devices equipped with Global Positioning System (GPS), and large trees were identified both in the field and in the lidar data. This ensured that the positioning of the transects and the lidar data matched. Second, multiple differential GPS readings were taken within 1 m of each transect segment endpoint to ensure that the correct tree was identified. As an additional check, distances were measured (using a taut tape measure) between the

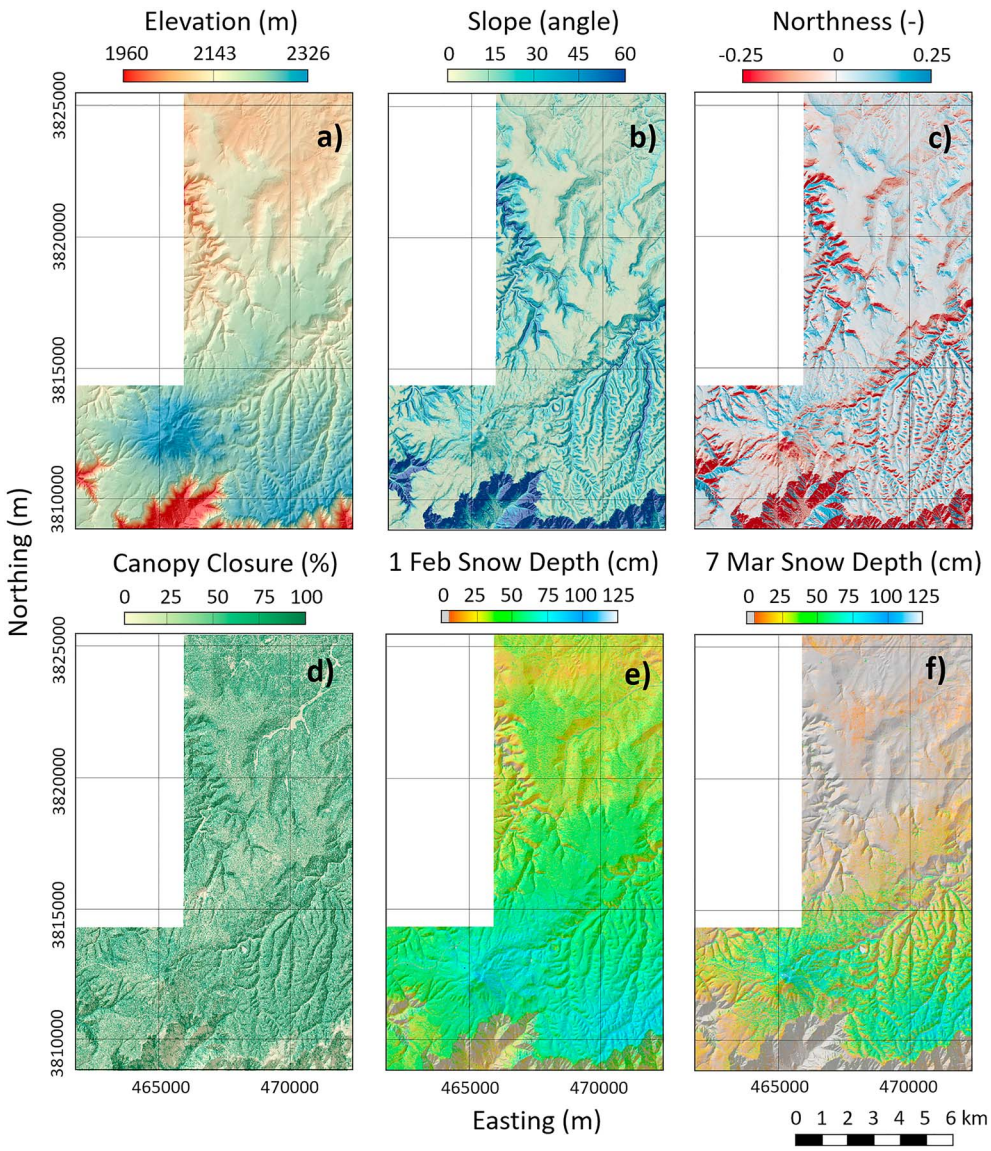

Figure 2. Maps showing (a) elevation, (b) slope, (c) northness [sin (slope) × sin (aspect)], (d) canopy closure, and bias-corrected lidar-derived snow depth maps for (e) 1 February 2017 and (f) 7 March 2017 for the mid-elevation lidar survey box. For northness, positive values indicate north facing slopes.

endpoints of transect segments and all distances were compared with map distances between the transect endpoints.

Snow measurement campaigns generally followed the methods described by Biederman et al. (2014). Surveyors collected data at regular intervals along tape measures deployed between the transect endpoints. To ensure spatial representativeness, snow depth measurements were taken with snow probes every five meters along each transect in a star pattern (i.e., one measurement at exactly the fifth meter and measurements taken 1 m away forward, backward, to the right, and to the left). At each site, SWE measurements were taken (using federal core samplers) every 10 or 20 m along each transect, depending on the transect lengths. We also excavated 17 snow pits to derive layer-integrated bulk snow density estimates. Snow pit locations were selected to represent the dominant regimes in variables regulating snow density (i.e., elevation, canopy cover, shading, and topography). Density estimates from pits were compared against federal sampler core estimates representing similar regimes (i.e., expected to have similar density).

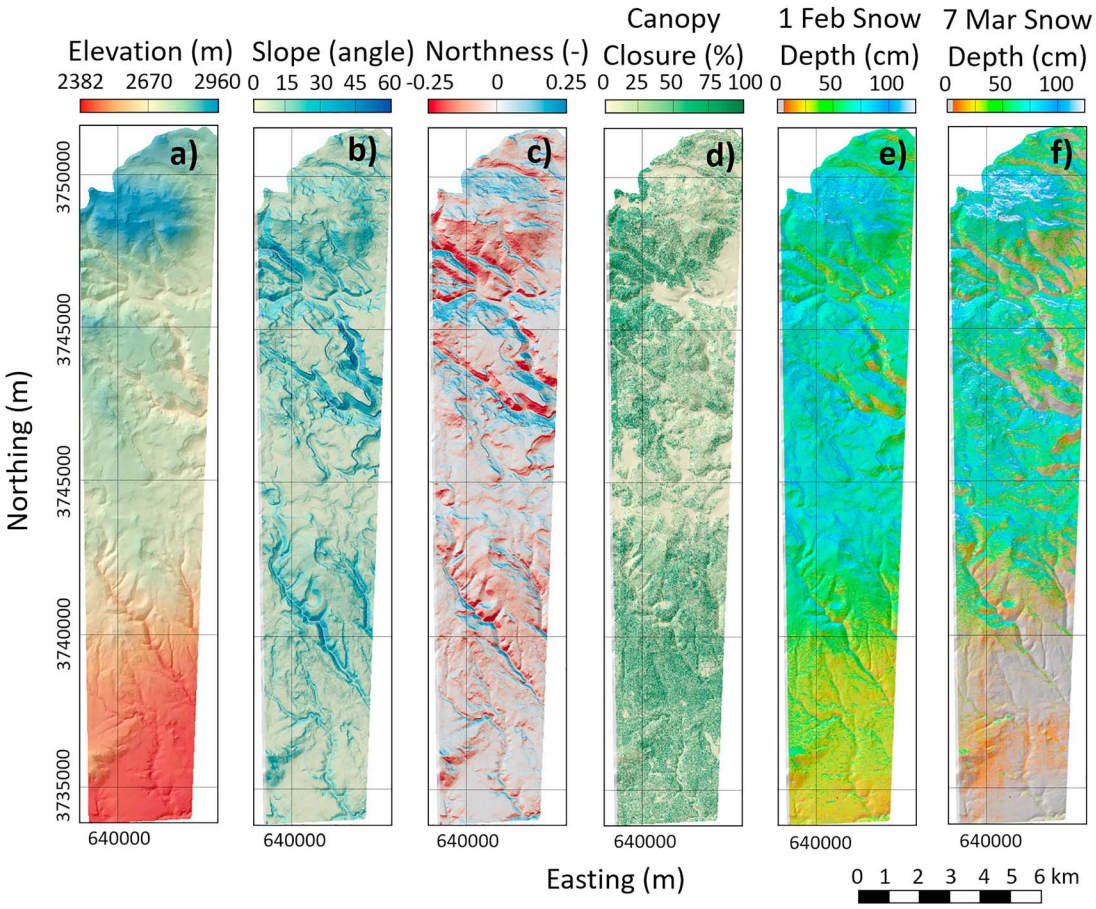

Figure 3. Maps showing (a) elevation, (b) slope, (c) northness [sin (slope) × sin (aspect)], (d) canopy closure, and bias-corrected lidar-derived snow depth maps for (e) 1 February 2017 and (f) 7 March 2017 for the high-elevation lidar survey box. For northness, positive values indicate north facing slopes.

2.2.3. Snowtography Measurements

We also collected "snowtography" measurements or daily pictures of snow measurement poles by trail cameras mounted in trees to construct daily snow depth time series throughout the winter for ~20 locations each at two of our field sampling sites (M1 and H1 in Figure 1). They provide a temporal complement to the spatial SWE and snow depth data collected during the lidar snow surveys, and they are used as data points to train and evaluate gridded snow density data generated in this study. These measurements are located within 50 m of SNOTEL stations at each site. The snowtography measurements (which were precisely georeferenced using differential GPS measurements and drone imagery) were set up along transects in cardinal directions across clearings and extend beyond the boundaries of the clearings into the forest. These locations represent gradients of both snowfall inputs (e.g., between canopy-covered and open areas) and snowpack ablation (sublimation and/or melt) in areas with varying amounts of exposure to sun and wind. Daily photographs were analyzed by hand to construct snow depth time series data at each measurement pole (in each image, two to four measurement poles could be seen). In addition, snow density measurements were made using a federal sampler at every other measurement pole during the snow surveys.

2.3. Bias Correcting Lidar Snow Depth Data

Lidar snow depth data are corrected using field snow depth measurements. First, the precise coordinates of each snow depth measurement along each snow survey transect are identified by computing map locations geometrically, relative to transect endpoints. The average of the five snow depth measurements taken within 1 m of each transect location are compared to the corresponding 1-m pixel of the snow-on minus snow-off lidar DTM (as delivered by QSI) difference maps. These comparisons are used to construct correction factors to remove the biases between the lidar and field-measured snow depth data (generally, lidar data

underrepresents the field-measured snow depths, Figure S1 in the supporting information) by fitting power law curves to the relationship between the lidar depths and the difference between lidar and field depths. Different correction factors are developed individually for each lidar coverage (Figure S2). These correction models are evaluated using a tenfold cross validation (samples for each lidar box on each date are broken into ten subsets, and each subset is evaluated independently with models trained with data from the other nine). Next, the lidar snow depth data are multiplied by these correction factors. Note that even though this multiplicative factor is higher at lower values of snow depth, its impact on the absolute value of snow depth is not necessarily larger at lower snow depths. Maps showing the corrected snow depths for each lidar box are shown in Figures 2 and 3 (panels e–f).

2.4. Artificial Neural Network Modeling of Snow Density Measurements

Next, field snow density measurements are used to construct maps of snow density based on ANN modelling of the field measurements. First, snow density data are reviewed for suspect values. We discard snow density values less than 0.2 g/cm^3 for the 1 February lidar survey and 0.25 g/cm^3 for the 7 March lidar survey to eliminate outliers. This results in 4% of snow density measurements being discarded on 1 February and 7% being discarded on 7 March. For comparison, none of snow pits that we dug during our snow surveys recorded snow densities below 0.26 g/cm^3 on 1 February or 0.29 g/cm^3 on 7 March. Figure S3 shows histograms of snow density measurements used in our snow density modeling (which are all based on federal sampling) for the middle- and high-elevation boxes on the two dates, with vertical lines representing snow density measurements from snow pits for comparison. Although there are too few snow pit observations for meaningful statistical comparisons with federal sampler snow density observations, snow pit densities are also compared with snow cores representing similar environments at each sample location (e.g., M1, M2, and M3) in section 3.3.1 below.

As with the snow depth data, the snow density measurements are precisely georeferenced for comparison with lidar attribute data. At each measurement location, geophysical attributes (elevation, slope, northness—or sin (slope) × sin (aspect) —where positive values indicate north facing slopes, canopy closure, and canopy height) are extracted from 1-m raster coverages for each lidar box (many of these are plotted in Figures 2 and 3), and these attributes are associated with the snow density measurements. Elevation, slope, and northness maps are derived from the bare-earth DTMs delivered by QSI using the Geospatial Data Abstraction Library software, the canopy height maps are computed as the difference between the bare-earth DTMs and the highest hit DTMs delivered by QSI, and canopy closure is generated in U.S. Forest Service FUSION/LDV software (McGaughey, 2012). In addition to these maps, we also use maps of snow depth, skyview factor, and below-canopy solar forcing index, or the ratio of incident solar radiation reaching the ground surface over a given period to that hitting a flat surface with no obstructions. Skyview factor and below-canopy solar forcing index are generated from 1-m lidar maps of canopy and terrain using the SnowPALM energy balance snow model (Broxton et al., 2015) integrating from the date of first snowfall to the date of each lidar flight at each site. Figures 4 and S4 show how representative the sample locations are of these predictor variables across the lidar boxes. For the 1 February (larger) surveys (Figure 4), the samples span much of the dynamic range of many of the predictor variables. The notable exception is elevation, which was limited by our site locations at which snow was measured in the field. For the 7 March (smaller) surveys (Figure S4), despite the lower number of samples taken, the dynamic range of many of these predictor variables is still sampled fairly well (with the exception of elevation).

Next, the ANN models are generated and trained with the snow density data (and associated predictor variables) using the Artificial Neural Network Toolbox® in Matlab®. These trained ANNs are reapplied using the predictor maps to generate maps of snow density. Like the snow depth data, on each date and for each area, the snow density models are evaluated using a tenfold cross validation. Note that all snow density data from the snow survey and snowtography transects are combined to create a single data set for each lidar box and for each area. We use a relatively simple network structure (one hidden layer with 10 neurons) because increasing the complexity of the ANNs does not improve model performance. Levenberg–Marquardt backpropagation (Marquardt, 1963) is used for model fitting. For each set, we run the optimization 50 times and average the resulting maps, as this technique further reduces model overfitting and improves the performance and consistency of the ANN results. Finally, the resulting 1-m resolution snow density maps are multiplied by the corresponding corrected lidar snow depths to produce 1-m SWE maps.

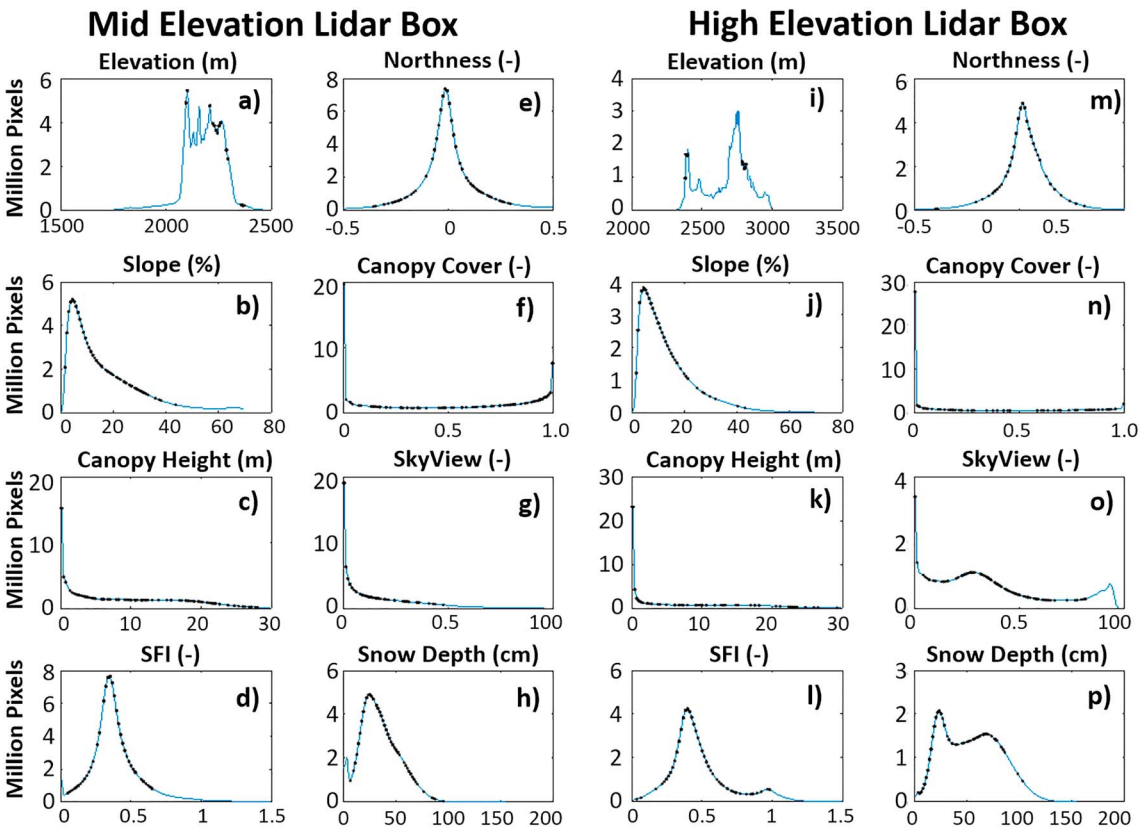

Figure 4. (a–h) Plots showing the distributions of landscape predictors within the mid-elevation lidar box (lines) and those predictors at individual sampled locations (dots) on 1 February. (i–p) The same but for the high-elevation lidar box on 1 February.

These gridded maps (of snow depth, snow density, and SWE) are used to evaluate the spatial representativeness of SNOTEL stations that are used for water supply forecasting by the Salt River Project (SRP), a major water stakeholder in AZ that provides hydropower and water supply to millions of customers in central AZ from snowmelt-fed streams that originate in our study areas. Understanding how these point measurements relate to the surrounding landscape in terms of snow depth, density, and SWE is important when using them to monitor snowpack.

3. Results

3.1. Temporal Context for the Lidar Data

In the mid-elevation lidar box, the snow-on lidar maps show snow conditions roughly 1 week following large snowstorms, which occurred on 20–24 January 2017 and 28 February 2017 (Figures 5a and 5b). The two storms both brought more than 100 mm of SWE to the mid-elevation box. In the lower elevations, this snowfall occurred on top of very shallow (for the first storm) or nonpresent (for the second storm) preexisting snowpack. In these areas, the snowpack depicted in the lidar maps was composed mostly (for the February flight) or entirely (for the March flight) of snow no more than ~1 week old. In other areas of the mid-elevation box, there was snowpack on the ground prior to the two flights, so the lidar maps depict a mix of newer and older snow.

Throughout much of the high-elevation lidar box, there was continuous snowpack persisting from December to March (Figures 5c and 5d). Unlike most areas within the mid-elevation box, the period preceding the January storm had significant snow accumulation in the high-elevation box. Furthermore, the storms on 20–24 January and especially 28 February were smaller, especially relative to the snowpack on the ground. Therefore, for much of the high-elevation box, most of the snowpack consisted of older snow on both survey dates.

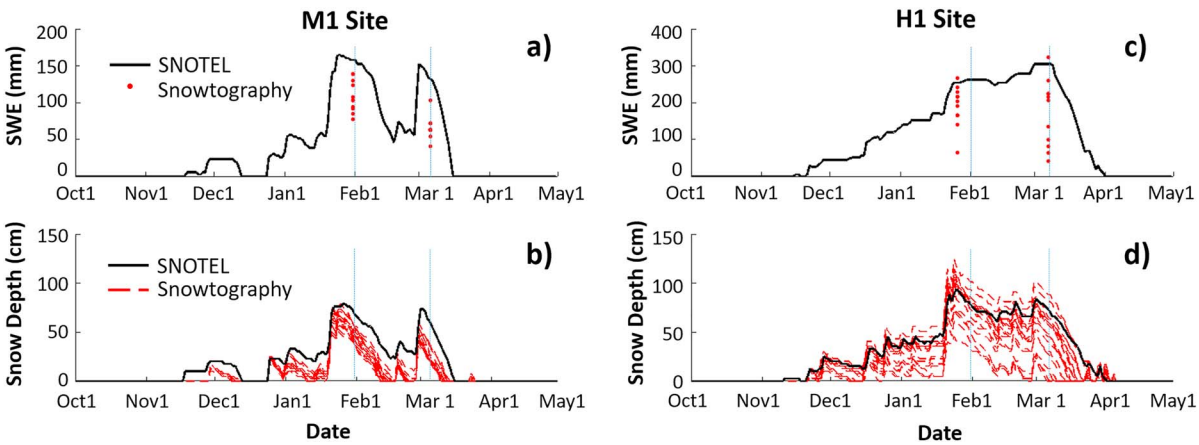

Figure 5. (a) SWE time series at the Snow Telemetry (SNOTEL; black solid line) and manual SWE measurements along snowtography transects (red dots) at the M1 site. (b) Snow depth time series at the SNOTEL (black line) and snowtography measurement locations (red lines). (c and d) The same information at the H1 site. Snowtography measurements capture a broad range of environmental variables expected to control snow depth and density, while SNOTEL sensors are located in areas of high snow accumulation. SWE = snow water equivalent.

Temperatures were much warmer and sun angles were higher prior to the second (7 March) survey, so significantly more snow ablation occurred between the 28 February snowstorm and the 7 March survey than between the 20–24 January snowstorm and the 1 February survey. Therefore, there was substantially less/no snow in the lower reaches of the lidar boxes on 7 March (Figures 2f and 3f). At the higher elevations, snow depths were fairly similar between the two lidar surveys, though there was a much larger difference between snow depths on differently oriented slopes. There are large differences between the two lidar surveys on the south- versus north facing slopes in the northern half of the high-elevation box (Figures 2f and 3f), as well as on different sides of forest stands (Figures 6 and S5). For the mid-elevation lidar box, the coefficient of variability (CV) of the 7 March snow depth map is 2.8 times as large as that of the 1 February map (0.49 vs. 1.35). For the high-elevation lidar box, the CV of the 7 March snow depth map is 2.1 times that of the 1 February map (0.42 vs. 0.87).

3.2. Comparison Between Lidar and Snow Survey Snow Depth Data

For both surveys, the uncorrected lidar snow depths (those generated by differencing the snow-on and snow-off lidar DTMs), are, in general, a little shallower than field-measured snow depths at all of the sites that we sampled, indicated by the negative biases in Table 1, column 6. These biases represent up to 33% of the average snow depth in some areas. However, after bias correction, the absolute value of these biases is reduced for both lidar boxes on both dates (Table 1, column 9), ranging from +1.6 to −2.2 cm (compared to −3.6 to −12.1 cm before bias correction. Root-mean-square errors (RMSEs) between the lidar and field snow depth data also decrease, with RMSEs between 7.8 and 11.5 cm for the bias-corrected data (compared to 9.1 to 18.7 cm before bias correction). $\boldsymbol{R}^2$ values between field-measured and lidar snow depths are very similar before and after bias correction.

Figure 6 shows the agreement between the lidar and field-measured snow depth data on 1 February at a site where the snow depth variability was particularly dramatic (site H1 in the high-elevation lidar box). This site had a persistent snowpack lasting from December to March and so developed more snow variability than a majority of the other sites. On 1 February, the site had substantial variability of snow depths (CV = 0.33) and was close to 100% snow covered. The deepest snow was found in locations that received less solar radiation, such as north facing slopes (in the lower left corner of Figure 6) or areas shaded by canopy. Shallow snow was found on south facing slopes (in the upper right portion of Figure 6), along the southern margin of canopy stands, and in areas that became marshy with standing water (the area in the middle of the clearing just northeast of the transect endpoint labelled B2). By 7 March, the variability increased (CV = 0.56), with deep snow persisting in the shaded areas and becoming shallower, or completely disappearing in the sunniest areas (Figure S5). The lidar data captured the snow depth patterns that were observed in the field well (transects shown in Figures 6 and S5) with site-wide R^2 values between the two ranging from 0.8 to 0.9 (not shown).

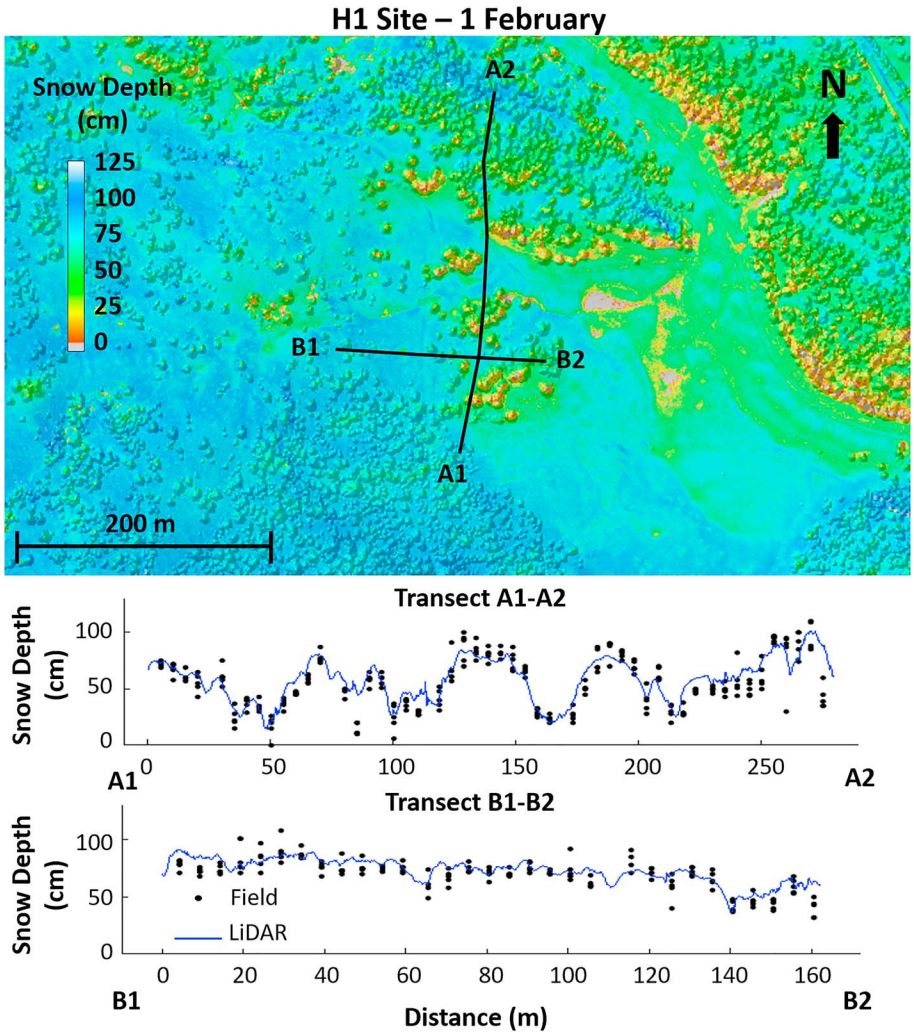

Figure 6. Snow depth maps derived from the 1 February lidar snow surveys at the H1 field site and comparison between lidar and field-measured snow depths along the transects shown in the maps. In the transect plots, all five field depth measurements are shown at each measurement location along each transect (black dots), and the blue line represents bias-corrected lidar snow depths.

Table 1
Snow Depth Measurements Along the Snow Survey Transects in Each Lidar Box for Each Date

						Uncorrected data			Bias-corrected data		
	Lidar box	Samples	Mean (cm)	Std (cm)	CV (–)	R^2 (–)	Bias (cm)	RMSE (cm)	R^2 (–)	Bias (cm)	RMSE (cm)
1 Feb	Mid	1,690	45.8	23.3	0.51	0.85	−12.1	15.2	0.84	−0.9	9.4
	High	1,275	39.7	24.1	0.61	0.91	−3.6	9.1	0.90	+1.6	7.8
7 Mar	Mid	325	44.0	32.4	0.74	0.89	−14.8	18.7	0.88	−2.2	11.5
	High	570	49.3	33.5	0.68	0.90	−8.4	13.8	0.90	+1.4	10.8

Note. Shown for each site are the total number of depth samples along the transects (column 1), the mean, standard deviation (Std), and CV of the samples (columns 2–4) and the squared correlation coefficient, mean bias (where negative biases indicate that lidar snow depths are less than observed snow depths), and RMSE between measured and lidar snow depths for both the uncorrected (columns 5–7) and bias corrected data (columns 8–10). CV = coefficient of variability; RMSE = root-mean-square error.

Table 2
Snow Density and SWE Measurements Within Each Lidar Box

			Density					SWE				
	Lidar box	# Sample	Mean (g/cm^3)	Std (g/cm^3)	CV (−)	R^2 (−)	RMSE (g/cm^3)	Mean (cm)	Std (cm)	CV (−)	R^2 (−)	RMSE (cm)
1 Feb	Mid	124	0.27	0.039	0.14	0.38	0.031	14.0	6.7	0.48	0.79	3.1
	High	121	0.28	0.039	0.14	0.29	0.033	11.1	6.9	0.62	0.90	2.2
7 Mar	Mid	33	0.28	0.035	0.12	0.34	0.028	10.9	9.4	0.86	0.88	3.2
	High	22	0.32	0.031	0.10	0.39	0.024	19.4	7.8	0.40	0.79	4.1

Note. Shown are the number of snow core samples along the transects (column 1), the mean, standard deviation (Std), and CV of snow density measurements (columns 2-4), and the squared correlation coefficient and RMSE between measured densities and ANN-generated snow density data using a tenfold cross validation (columns 5-6). The same information is shown for SWE measurements (columns 7–9) and their agreement to the lidar-ANN SWE data (columns 10–11). CV = coefficient of variability; RMSE = root-mean-square error.

3.3. Distributed Maps of Snow Density and SWE From Measurements and Artificial Neural Network Modeling

3.3.1. Snow Density Data

During both snow surveys, measured snow densities varied widely. For both lidar boxes, the observed snow densities vary from 0.21 to 0.36 g/cm^3 for the 1 February snow survey and from 0.24 to 0.37 g/cm^3 on the 7 March snow survey (Figure S2). For the February snow survey data, the CV of snow density measurements was 0.14 for both lidar domains (Table 2). For the 7 March snow survey, snow densities have a slightly smaller range (CV = 0.10–0.12; Table 2). As a check on the accuracy of the federal sampler snow core values, we compared core densities with layer-integrated density values from snow pits, considered a highly accurate method. Snow density showed no clear pattern of bias between pits and cores collected in environments expected to have similar density based on elevation, topography, snow depth, canopy density, and solar forcing (mean bias -0.009 ± 0.022 g/cm^3; Table S1). At a given survey site (e.g., Site M1 in the mid-elevation lidar box), multiple cores from a given environment (e.g., level terrain with dense canopy) had lower variability than the aggregate of cores from across the site (Figure S6). Core densities matched well with layer-integrated pits at most sites, increasing confidence in core accuracy, although a low bias (−0.03–0.05 g/cm^3) was observed at Site M5, which had the least snow depth of all sites (20–30 cm).

Overall, the ANN modelling was able to represent snow densities fairly well in terms of RMSE, while the spatial representation of ANN-generated snow density is poorer than that of the lidar snow depths. The $\boldsymbol{R}^2$ agreement between snow density measurements and the ANN-generated snow density data (using tenfold cross validation) range from 0.29 to 0.39 for the middle- and high-elevation lidar boxes on 1 February and 7 March and RMSE's range from 0.024 to 0.033 g/cm^3 (Table 2). These RMSEs are 8%–12% of the average snow density values for each lidar box (Table 2).

To get a sense of model performance when fewer samples are used to train the ANN snow density models (indicating the effect of sample size on model performance), the models were also trained with fewer data. The same tenfold cross validation was used in all cases, except that only part of the data in the training subsets were used to train the models for each evaluation subset. In all cases, we find that as the number of training samples increases, the model performance increases, suggesting that larger training samples are better than smaller ones (Figure 7), though the performance (at least for the larger 1 February snow surveys) begins to plateau above 75–100 samples. There is also interesting variation between the sites. For example, the spatial performance ($\boldsymbol{R}^2$ statistic) for the snow density model is always higher for the mid-elevation box than for the high-elevation box on 1 February. In addition, the increase in $\boldsymbol{R}^2$ occurs faster for both areas on 7 March.

Figure 8 shows 1-m maps of ANN-generated snow density and SWE. These snow density maps show values ranging from 0.2 to 0.35 g/cm^3 for the mid-elevation lidar box and mostly above 0.25 g/cm^3 for the high-elevation lidar box. In the mid-elevation box, lower-elevation areas have lower snow density (Figure 8a; as the low elevations had snowpacks that were predominantly composed of snow no older than 1 week), while in the high-elevation lidar box, areas under trees tend to have lower values of snow density than open areas (Figures 8c and 8f). These trends can be seen when assessing the importance of individual factors on the snow density models: for 1 February, elevation and terrain slope appear dominant over other variables in

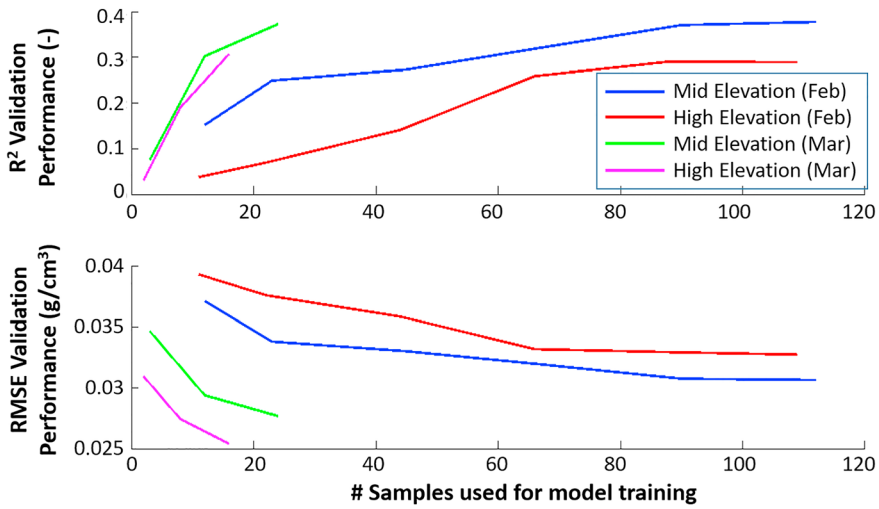

Figure 7. R^2 and RMSE of snow density samples using a tenfold cross validation showing model performance showing different numbers of training samples to generate each model for each lidar box on each date. RMSE = root-mean-square error.

the mid-elevation box, while in the high-elevation box, solar forcing index at the snow surface and skyview (which would distinguish between sunny areas in gaps and shaded areas near to and underneath trees) appear to be important (Table 3). On 7 March, the snow densities are higher overall (generally above 0.25–0.3 g/cm^3; Figure S7 in the supporting information). In the mid-elevation box on 7 March, elevation seems to be much less important, while parameters related to the presence of vegetation become much more important. For the high-elevation box on 7 March, terrain slope seems to be the most important factor. Note that there is colinearity between some of the predictors in Table 3 (e.g., between snow depth and elevation) or vegetation cover and vegetation height, and so this analysis does not necessarily assign unique importance to all of the predictors.

In general, there is more variability of snow densities on 1 February than 7 March. For the mid-elevation box, the CV of the 7 March snow density map is 95% that of the 1 February snow density map (0.066 vs. 0.069), while for the high-elevation box, the CV of the 7 March snow density map is 59% that of the 1 February snow density map (0.049 vs. 0.072).

3.3.2. SWE Data

There is generally good agreement between the SWE maps (generated by multiplying the bias-corrected lidar snow depth maps with the ANN snow density maps) and observed SWE values. R^2 values between the Lidar-ANN SWE data and field SWE measurements (ranging from 0.79 to 0.90; Table 2) are much higher for than between the ANN snow density data and field density measurements. RMSEs between the Lidar-ANN SWE data and field SWE measurements range from 2.2 to 4.1 cm (which are 28%–37% of the RMSEs for snow depths). Note that these RMSEs are also up to 17% lower than they would be if the snow density was assumed to be constant (as the average value of the snow density measurements). For example, if a constant snow density that represents the average of all snow density measurements at the site (0.27 g/cm^3) were used to predict SWE at the sampled locations for the mid-elevation box on 1 February, the RMSE for predicted versus measured SWE values would be 3.6 cm (as opposed to 3.1 cm listed in Table 2). For the remaining areas, these RMSEs are 6% to 13% higher than the values listed in Table 2.

Although variability of the SWE maps is heavily influenced by variability in the lidar snow depth maps, there are differences between the SWE and snow depth maps. For example, because there are higher values of snow density at higher elevations (where there is deeper snow) on 1 February for the mid-elevation lidar box, SWE differences are larger between the lower and higher elevations of both boxes than would be indicated from the lidar snow depth data alone (and assuming a constant snow density). However, in general, because the variability of the snow density maps is lower for 7 March than 1 February (section 3.3.1) while the variability of the snow depth maps is higher 7 March (section 3.1), the relative importance of the snow depth variability is much larger for 7 March. For example, for the mid-elevation box, the ratio between the

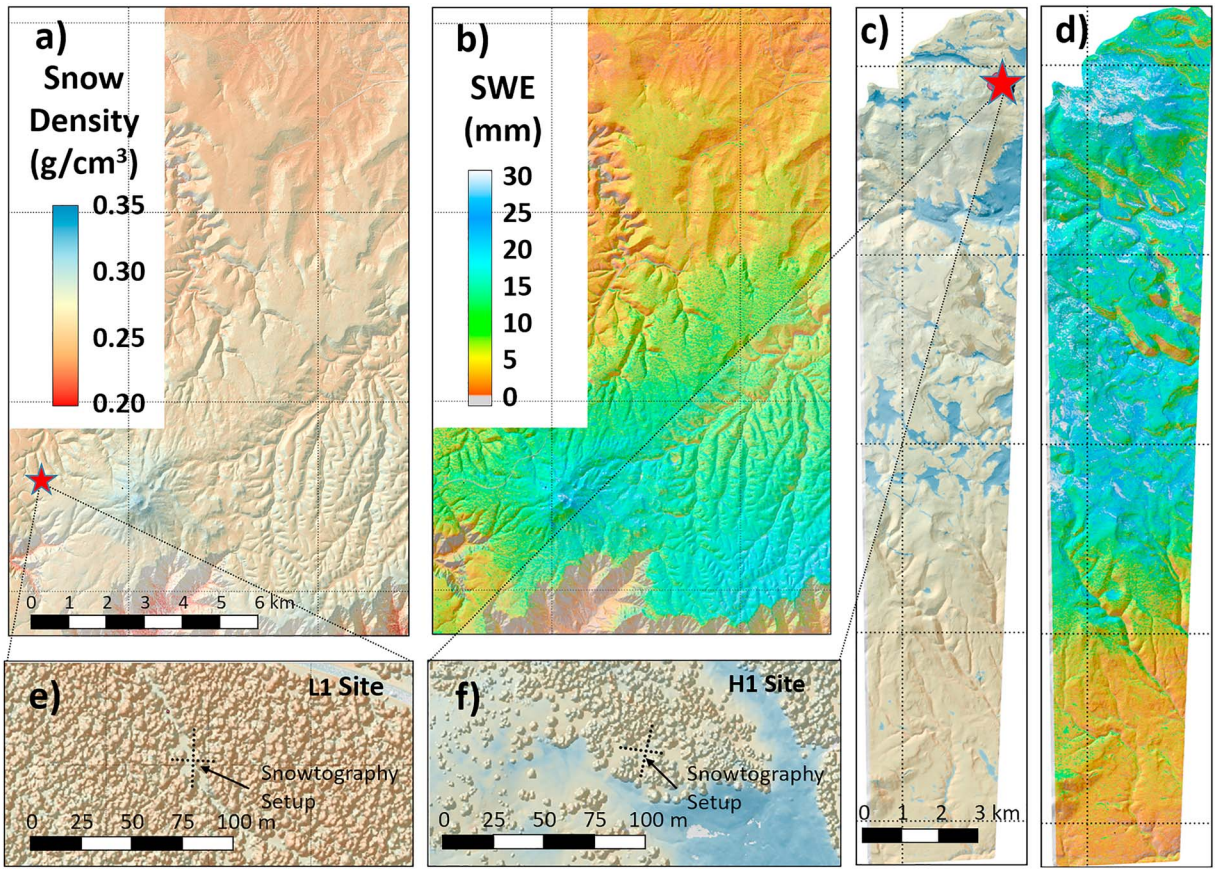

Figure 8. (a) ANN-generated map of snow density for the mid-elevation lidar box on 1 February. (b) SWE map computed by multiplying the ANN-generated snow density map (from a) by the bias-corrected lidar snow depth map (Figure 2e). (c and d) Same except for the high-elevation box. (e and f) Zoomed in versions of the snow density maps around the snowtography setup at sites M1 and H1. Areas with less than 1 cm of SWE are masked out in the snow density maps (appearing as gray) and also appear as grey in the SWE maps (note the color bar).

snow density and snow depth CV decreases from 0.15 to 0.05 from 1 February to 7 March. A similar decrease (from 0.17 to 0.06) occurs for the high-elevation box. Overall, the CV of the SWE maps is similar to that of the snow depth maps (0.50 and 1.34 for the mid-elevation box on 1 February and 7 March, respectively, and 0.45 and 0.88 for the mid-elevation box on 1 February and 7 March, respectively).

3.4. Comparison With Snowtography and SNOTEL Data

As can be seen in Figure 5, the snowtography sites generally have less snow depth and SWE through the winter compared to SNOTEL stations because the snowtography measurements capture a broad range of environmental variables expected to control snow depth and density, while SNOTEL sensors are located in areas of relatively high snow accumulation; that is, they are not necessarily representative of the snow conditions for a majority of the landscape at the elevation in which they are found. For example, Figure 9 shows how the snowtography and SNOTEL snow depth, snow density, and SWE measurements at sites M1 and H1

Table 3
Factor Importance of the Predictor Variables Used in the ANN-Generated Snow Density Maps

	Lidar box	Snow Depth	Elevation	Northness	Slope	Vegetation Cover	Vegetation Height	SkyView	SFI
1 Feb	Mid	0.23	0.53	0.00	0.45	0.00	0.03	0.01	0.00
	High	0.01	0.07	0.04	0.03	0.01	0.00	0.10	0.34
7 Mar	Mid	0.09	0.11	0.23	0.09	0.50	0.10	0.41	0.08
	High	0.01	0.12	0.04	0.53	0.01	0.02	0.01	0.05

Note. Shown are R^2 values between the ANN-generated snow density maps and these predictor variables. SFI = solar forcing index.

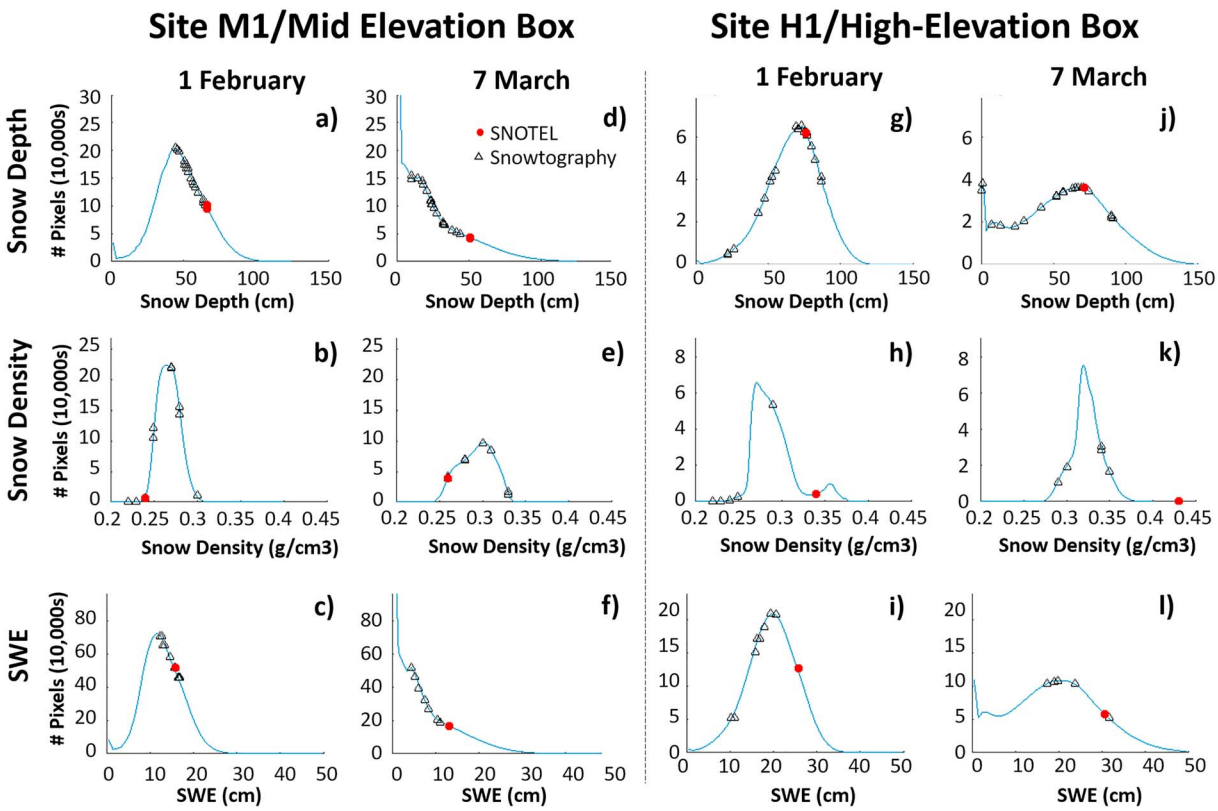

Figure 9. Plots showing the distribution of lidar snow depth (top row), artificial neural network-generated snow density (middle row), and lidar-artificial neural network SWE (bottom row) distributions for the 500-ft (152 m) elevation band used for Salt River Project streamflow forecasting that includes sites M1 and H1. For M1, the band ranges from 2,134 to 2,286 m within the mid-elevation lidar box and for H1, the band ranges from 2,743 to 2,895 m within the high-elevation lidar box. Plotted on the histograms are snow depth, density, and SWE measurements at Snow Telemitry (SNOTEL) sites and along snowtography transects at each location. Note that there are fewer triangles in the middle and bottom rows because snow density was measured (using federal sampling) at every other snowtography location.

compare with the gridded data generated in this study for 500-ft (~152 m) elevation bands in which they are located. These elevation bands are chosen because they correspond to those used by SRP to classify snow amounts for seasonal streamflow forecasts. For M1 (~2,225 m a.s.l.), the band ranges from 2,134 to 2,286 m within the mid-elevation lidar box and for H1 (~2,800 m. a.s.l), the band ranges from 2,743 to 2,895 m within the high-elevation lidar box.

At M1, SNOTEL snow depth measurements on 1 February and 7 March plot toward the higher end of elevation band snow depth, yet snow density measurements (calculated by the common practice of dividing SNOTEL SWE and snow depth measurements) plot toward the low end of the elevation band snow densities. As a result, SNOTEL SWE data, while still high compared to the surrounding area, are not as high as would be predicted if just looking at the SNOTEL snow depth data. Conversely, for the H1 site, the SNOTEL data generally depicts about average snow depth for its elevation band but much higher than average snow density, leading to higher than average SWE. Note that in many cases, in addition to capturing a range of snow depth, density, and SWE that at least captures part of their distributions for the elevation band, the average of the snowtography measurements tends to be closer to the average values for the elevation band in most cases (Table S2).

The representativeness of the point observations also appears to change through time. Figure 10 shows boxplots depicting the interquartile range of SWE data for all 500-ft (152 m) elevation bands within both lidar domains on both dates. Also plotted are the site elevations (vertical lines) and SNOTEL/snowtography SWE measurements (horizontal lines). At both sites, the SNOTEL SWE values are higher than the average SWE for the elevation bands in which they are found, particularly for the 7 March survey. For the later survey date, especially, the SNOTEL SWE measurements are more similar to the average SWE in higher

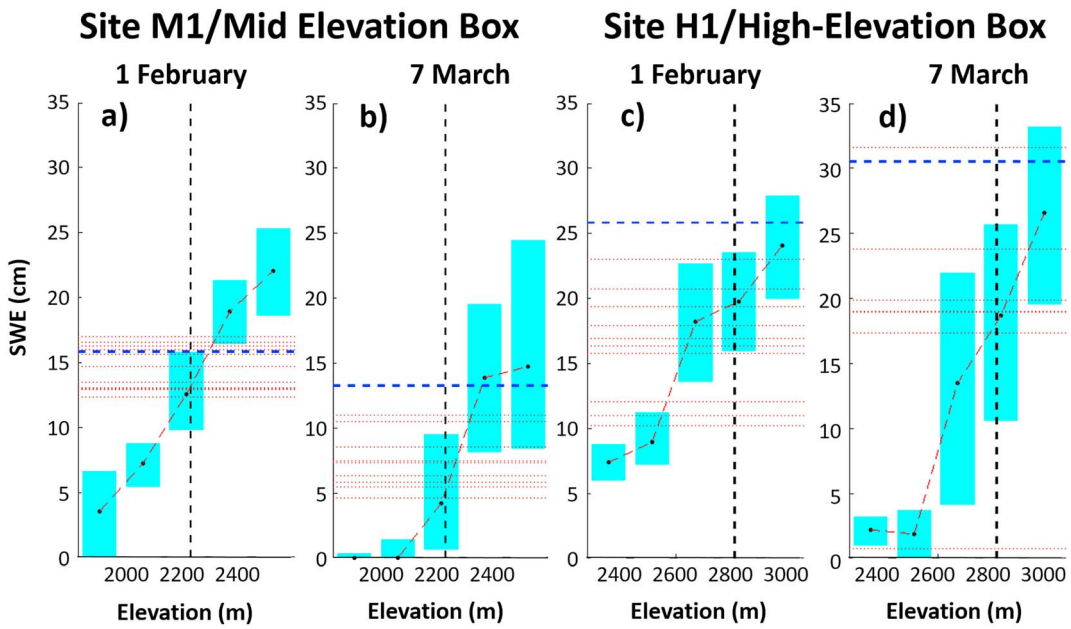

Figure 10. (a and b) Boxplots showing the interquartile range of gridded SWE in the mid-elevation lidar box for 500-ft (152 m) elevation bands for the 1 February and 7 March snow surveys. The connected dots show median SWE for each elevation band. The vertical dashed lines show the elevation of the M1 field site on each panel. The horizontal blue dashed lines show SWE recorded at the Snow Telemetry and the horizontal red dotted lines show SWE measurements along snowtography transects on the two dates. (c and d) The same information for the high-elevation lidar box and H1 site. The elevation bands are chosen to match those used for Salt River Project seasonal streamflow forecasts.

elevation bands. Conversely, the snowtography measurements, which include sunny as well as shaded sites, represent average conditions for their elevation band (Figure 10).

4. Discussion

Several studies have found it important to consider spatial variability of snow density when converting lidar snow depths to SWE (e.g., Jonas et al., 2009; López-Moreno et al., 2013; Raleigh & Small, 2017; Wetlaufer et al., 2016). Here there is a large range of density across the lidar coverages (~0.2–0.35 g/cm^3; spanning ~100 km^2 and ~500 m of elevation), and this variability, while smaller than that of snow depth, is important, especially earlier in the winter. For example, for our midwinter (1 February) survey, the CV of the density maps is 15%–17% of that of the snow depth maps while for our late-winter (7 March) survey, it is only 5%–6% of that of the snow depth maps (driven primarily by increases in snow depth variability). This suggests that measurements of snow depth (e.g., from airborne lidar) are better able represent SWE variability later in the winter. Our results also demonstrate the importance of spatially distributed snowpack measurements because point measurements (e.g., from SNOTEL stations) can be unrepresentative in terms of snow depth, SWE, **and** snow density, and further, their representativeness changes through time (Figure 10).

Spatial patterns in our ANN-generated snow density maps are generally consistent with processes known to regulate density. After snow falls to the ground, compaction from additional snow, metamorphism from aging snow, and melt from liquid and/or refreezing snow increase density, but these processes occur at different rates in different places on the landscape (Bormann et al., 2013; DeWalle & Rango, 2008; Kojima, 1967; Sturm et al., 1995). In general, we expect snowpack density to increase with thickness, age, and exposure to warm temperatures and high radiation. In our ANN maps, we found that older and thicker snowpacks had higher-density snow. For example, snow densities were higher overall for the late-winter snow survey than the midwinter snow survey, particularly in the higher elevation lidar box, where the snowpack was oldest (i.e., persisted all winter) and deepest. Snowpack densities were somewhat lower in

the mid-elevation box versus the high-elevation box during the late-winter snow survey because of the much larger size of the snowstorm that preceded the snow survey relative to the amount of snow on the ground (Figure 5). Similarly, in the mid-elevation box during the midwinter survey, density increased with elevation because snowpack is ephemeral in the lower elevations of the box, and much of what was measured on the survey date had fallen only 1 week before. Meanwhile, the higher elevations of the box had an older, thicker, and denser snowpack. Across areas with snowpack of similar age and thickness, solar radiation and forest cover were important. For example, for the much of the high-elevation box during the midwinter survey, large, sunny clearings had markedly higher snow densities than the shaded areas next to and underneath forest canopy (Figure 8c) perhaps indicating melt refreeze of sun-exposed snowpack contributing to higher densities.

Water resources stakeholders such as SRP have long relied upon sparse networks of snowpack point measurements, such as SNOTEL stations. Here we find that SNOTEL SWE measurements were biased high compared to the surrounding area at the same elevation. Furthermore, SNOTEL sites became less representative of surrounding snowpack SWE, depth, and density for the late-winter survey. We found that density can be both overestimated and underestimated by SNOTEL measurements ($\pm$0.05–0.10 g/cm^3), perhaps because SNOTEL stations are predominantly in flat, open, and shaded locations with different snow densities than other physiographic settings. This nonrepresentativeness impacts watershed-scale SWE estimates if SNOTEL density is used with lidar snow depths to calculate SWE. For the high-elevation lidar box, applying the SNOTEL point density would increase area-averaged SWE by 18% and 33% compared with the ANN SWE map for the middle- and late-winter surveys, respectively. Conversely, for the mid-elevation box, SNOTEL point density would reduce area-averaged SWE by 15%–18% compared with the ANN map for middle- and late-winter surveys. These results suggest caution when using SNOTELs to evaluate density estimates produced by physically based snow models, as SNOTELS may only be representative for limited physiographic conditions.

4.1. Uncertainty of the ANN Snow Density Data

In general, we find that the ANNs better represent snow density variability than some other machine learning approaches. For our data set, RMSEs between the field-collected and ANN-generated values range from 0.02 to 0.03 g/cm^3, explaining 29%–39% of the observational variance of snow density. This performance is similar in terms of $\boldsymbol{R}^2$ or better in terms of RMSE than regression tree and MLR models used by Wetlaufer et al. (2016) that model snow density similarly based on physiographic factors. For our data set, ANNs better modeled snow density than the MLR approach (Table S3).

A variety of factors likely contribute uncertainty to ANN snow densities. First, they may be related to how well the range of physiographic variables was represented by sampling. Most physiographic variables were represented fairly well (Figures 4 and S4), although a weaker sampling of elevation could have reduced performance of the ANN snow density models at some sites (e.g., for the high-elevation box). Density uncertainty also decreased with larger sample size used to train the models (Figure 7). Finally, it could also be related to uncertainty in field density measurements, themselves. Some previous studies (e.g., Sturm et al., 2010) have suggested that bulk densities measured with coring tubes may be up to 10% too low because of issues such as snow falling out of the sampler unless a soil plug is collected (and removed before weighing) or the sampler hitting an ice layer, which can push snow out of the sampler's path. However, contrasting results suggest core-based densities may actually overestimate snow density but up to 10%, especially in deeper, denser snow (Peterson & Brown, 1975; Farnes et al., 1982). In this study, we find that there is relatively little bias between layer-integrated densities from snow pits (which are considered to be highly accurate) and core-based values taken from similar environments at each sampling site, and furthermore, random uncertainty appears to be smaller than the differences across environments which are we trying to predict (Table S1 and Figure S6). One notable exception is for shallow snowpacks 20- to 40-cm deep at a warm site in the mid-elevation box (site M5) during the midwinter survey, where core-based estimates average ~5% lower density than layer-integrated pit values (Figure S6). The equipment used for this survey (1-m federal sampler tubes and 1,000-ml snow pit layer cutters) was not well adapted for measurement of such thin snowpacks. Because of this, we recommend using 30-cm plastic snow cutter tubes and 250-ml cutters in such low-snow environments.

4.2. Uncertainty of the Lidar Snow Depth Data

Overall, the accuracies of the lidar snow depth data are higher than those of the snow density maps. RMSEs for the bias-corrected lidar snow depth data range from 9.4 to 11.5 cm for the lidar boxes. For comparison, Harpold et al. (2014), who evaluate lidar across four lidar domains with return densities of 8–10 returns/m^2, find RMSEs ranging from 7 to 31 cm (with an average of 23 cm) and Tinkham et al. (2014) report RMSEs of 14 cm in areas with low vegetation and 20–35 cm in forested areas. Painter et al. (2016) who use a laser scanning system having lower point counts (due to high-altitude flight) report snow depth RMSE's of ~8 cm at a coarser (3 m) spatial resolution. Several sources are known to contribute uncertainty to lidar snow depths. For example, either of the lidar retrievals (snow on and snow off) are prone to uncertainties associated with the placement of the lidar points on the order of 1–10 cm because of uncertainties with the onboard differential GPS, inertial measurement unit (IMU), or boresight calibration (e.g., Baltsavias, 1999). In addition, there is some uncertainty about the geolocation of the field data along the transects. However, the largest issue that we identified was that in some areas, lidar snow depths were systematically shallower than field-measured depths, especially in forested areas with shallower snow. A plausible explanation for this bias is that low vegetation such as grasses and shrubs can be misclassified as ground because it is difficult for ground filters to separate low vegetation and fallen logs from the ground surface when there are not enough lidar returns to do so effectively (Gould et al., 2013; Liu, 2008; Meng et al., 2010). This issue would lead to a bias because it would affect the snow-on and snow-off lidar differently (Deems et al., 2013; Harpold et al., 2014; Tinkham et al., 2014). Overall, we find that this bias is higher when there are fewer ground returns and largely disappears when there are more than 8–10 ground returns/m^2 (Figures S1e and S1f).

5. Conclusions

In this study, we combined lidar snow depth data with observation-based gridded snow density estimates to create high-resolution maps of SWE. One of the major methodological advances of this study is the use of a well distributed and precisely geolocated network of snow density measurements along with ANN machine learning to determine how snow density changes based on physiographic factors. Variability of these snow density maps is larger for a midwinter snow survey than for the late-winter snow survey, especially relative to that of snow depth variability. This suggests that is more important to consider the spatial variability of snow density earlier in the winter and that it is also easier to characterize SWE variability later in the season based from snow depth measurements (e.g., from airborne lidar). Even though the CV of our snow density maps is smaller than that of our snow depth maps, it is still important to consider snow density variability when creating distributed maps of SWE. This is particularly true when looking at pixel-to-pixel SWE variability, or when snow density is systematically higher or lower in areas of deeper (e.g., high elevation) or shallower (e.g., low elevation) snowpack.

Perhaps most importantly, though, this study shows that point snow density measurements can be unrepresentative of the broader landscape, impacting watershed-scale SWE estimation by up to ~30%, and furthermore, this unrepresentativeness can be different at different times. At our study sites, snow densities from SNOTEL data were as much as 0.12 g/cm^3 different from the distributed snow densities based on ANN modeling, and in general, the SNOTEL data was less representative later in the winter. This highlights the importance of using a well-distributed and extensive set of snow density samples to estimate snow density across a domain (e.g., a lidar coverage or a watershed), as too few/unrepresentative point measurements of snow density can lead to biased estimates of SWE.

Acknowledgments
This research is supported by a grant from the Salt River Project (SRP) Agricultural Improvement and Power District in Tempe, Arizona. Gridded data sets used in this study are derived from a snow-free lidar acquisition for the Four Forests Restoration Initiative (4Fri) and Snow-on lidar acquisitions acquired through funding by SRP (listed in the references). The 4Fri lidar data can be requested directly from the National Forest Service (https://www.fs.usda.gov/4fri), subject to a nondisclosure agreement to protect cultural resources. All data used to construct and evaluate the models in this study (including field measurement data, snow-on lidar measurement data, and extracted physiographic attribute data) can be accessed in the supplemental data sets. We would like to thank the many volunteers from the NRCS, USFS, USDA-ARS, University of Arizona, and SRP who helped with our field data collections. Finally, we thank Kyle Hartfield at the Arizona Remote Sensing Center for providing geographic information system (GIS) support, as well as those who provided peer-review feedback that greatly improved this manuscript.

References

Avanzi, F., De Michele, C., Ghezzi, A., Jommi, C., & Pepe, M. (2014). A processing-modeling routine to use SNOTEL hourly data in snowpack dynamic models. *Advances in Water Resources*, *73*, 16–29. https://doi.org/10.1016/j.advwatres.2014.06.011

Bair, E. H., Abreu Calfa, A., Rittger, K., & Dozier, J. (2018). Using machine learning for real-time estimates of snow water equivalent in the watersheds of Afghanistan. *The Cryosphere*, *12*(5), 1579–1594. https://doi.org/10.5194/tc-12-1579-2018

Bair, E. H., Rittger, K., Davis, R. E., Painter, T. H., & Dozier, J. (2016). Validating reconstruction of snow water equivalent in California's Sierra Nevada using measurements from the NASA Airborne Snow Observatory. *Water Resources Research*, *52*, 8437–8460. https://doi.org/10.1002/2016WR018704

Bales, R. C., Molotch, N. P., Painter, T. H., Dettinger, M. D., Rice, R., & Dozier, J. (2006). Mountain hydrology of the western United States. *Water Resources Research*, *42*, W08432. https://doi.org/10.1029/2005WR004387

Baltsavias, E. (1999). A comparison between photogrammetry and laser scanning. *ISPRS Journal of photogrammetry and Remote Sensing*, *54*(2-3), 83–94. https://doi.org/10.1016/S0924-2716(99)00014-3

Baños, I. M., García, A. R., Marturià i Alavedra, J., Oller i Figueras, P., Iglesias, J. P., Martínez i Figueras, P., & López, J. T. (2011). Assessment of airborne LIDAR for snowpack depth modeling. *Boletín de la Sociedad Geológica Mexicana*, *63*(1), 95–107. https://doi.org/10.18268/BSGM2011v63n1a8

Biederman, J. A., Brooks, P. D., Harpold, A. A., Gochis, D. J., Gutmann, E., Reed, D. E., et al. (2014). Multiscale observations of snow accumulation and peak snowpack following widespread, insect-induced lodgepole pine mortality. *Ecohydrology*, *7*(1), 150–162. https://doi.org/10.1002/eco.1342

Bormann, K. J., Westra, S., Evans, J. P., & McCabe, M. F. (2013). Spatial and temporal variability in seasonal snow density. *Journal of Hydrology*, *484*, 63–73. https://doi.org/10.1016/j.jhydrol.2013.01.032

Broxton, P. D., Harpold, A. A., Biederman, J. A., Troch, P. A., Molotch, N. P., & Brooks, P. D. (2015). Quantifying the effects of vegetation structure on snow accumulation and ablation in mixed-conifer forests. *Ecohydrology*, *8*, 1073–1094. https://doi.org/10.1002/eco.1565

Buckingham, D., Skalka, C., & Bongard, J. (2015). Inductive machine learning for improved estimation of catchment-scale snow water equivalent. *Journal of Hydrology*, *524*, 311–325. https://doi.org/10.1016/j.jhydrol.2015.02.042

Cline, D., Yueh, S., Chapman, B., Stankov, B., Gasiewski, A., Masters, D., et al. (2009). NASA cold land processes experiment (CLPX 2002/03): Airborne remote sensing. *Journal of Hydrometeorology*, *10*(1), 338–346. https://doi.org/10.1175/2008JHM883.1

Czyzowska-Wisniewski, E. H., van Leeuwen, W. J., Hirschboeck, K. K., Marsh, S. E., & Wisniewski, W. T. (2015). Fractional snow cover estimation in complex alpine-forested environments using an artificial neural network. *Remote Sensing of Environment*, *156*, 403–417. https://doi.org/10.1016/j.rse.2014.09.026

Dadic, R., Mott, R., Lehning, M., & Burlando, P. (2010). Wind influence on snow depth distribution and accumulation over glaciers. *Journal of Geophysical Research*, *115*, F01012. https://doi.org/10.1029/2009JF001261

Dawson, N., Broxton, P. D., & Zeng, X. (2017). A new snow density parameterization for land data initialization. *Journal of Hydrometeorology*, *18*, 197–207. https://doi.org/10.1175/JHM-D-16-0166.1

DeBeer, C. M., & Pomeroy, J. W. (2010). Simulation of the snowmelt runoff contributing area in a small alpine basin. *Hydrology and Earth System Sciences*, *14*(7), 1205–1219. https://doi.org/10.5194/hess-14-1205-2010

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

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

DeWalle, D. R., & Rango, A. (2008). *Principles of snow hydrology*. Cambridge, UK: Cambridge University Press.

Dietz, A. J., Kuenzer, C., Gessner, U., & Dech, S. (2012). Remotes sensing of snow—A review of available methods. *International Journal of Remote Sensing*, *33*(13), 4094–4134. https://doi.org/10.1080/01431161.2011.640964

Dobreva, I., & Klein, A. (2011). Fractional snow cover mapping through artificial neural network analysis of MODIS surface reflectance. *Remote Sensing of Environment*, *115*(12), 3355–3366. https://doi.org/10.1016/j.rse.2011.07.018

Dozier, J. (2011). Mountain hydrology, snow color, and the fourth paradigm. *Eos Transactions American Geophysical Union*, *92*(43), 373–374. https://doi.org/10.1029/2011EO430001

Dozier, J., Bair, E. H., & Davis, R. E. (2016). Estimating the spatial distribution of snow water equivalent in the world's mountains. *WIREs Water*, *3*(3), 461–474. https://doi.org/10.1002/wat2.1140

Farnes, P. E., Peterson, N. R., Goodison, B. E., & Richards, R. P. (1982). Metrication of manual snow sampling equipment. In *Proceedings of 50th Annual Western Snow Conference, Reno, NV, Western Snow Conference* (pp. 120–132).

Fassnacht, S. R., & Deems, J. S. (2006). Measurement sampling and scaling for deep montane snow depth data. *Hydrological Processes: An International Journal*, *20*(4), 829–838. https://doi.org/10.1002/hyp.6119

Foster, J. L., Sun, C., Walker, J. P., Kelly, R., Chang, A., Dong, J., & Powell, H. (2005). Quantifying the uncertainty in passive microwave snow water equivalent observations. *Remote Sensing of Environment*, *94*(2), 187–203. https://doi.org/10.1016/j.rse.2004.09.012

Foster, L. M., Bearup, L. A., Molotch, N. P., Brooks, P. D., & Maxwell, R. M. (2016). Energy budget increases reduce mean streamflow more than snow-rain transitions: Using integrated modeling to isolate climate change impacts on Rocky Mountain hydrology. *Environmental Research Letters*, *11*(4), 044015). https://doi.org/10.1088/1748-9326/11/4/044015

Gharaei-Manesh, S., Fathzadeh, A., & Taghizadeh-Mehrjardi, R. (2016). Comparison of artificial neural network and decision tree models in estimating spatial distribution of snow depth in a semi-arid region of Iran. *Cold Regions Science and Technology*, *122*, 26–35. https://doi.org/10.1016/j.coldregions.2015.11.004

Golding, D. L., & Swanson, R. H. (1986). Snow distribution patterns in clearings and adjacent forests. *Water Resources Research*, *22*(13), 1931–1940. https://doi.org/10.1029/WR022i013p01931

Gould, S. B., Glenn, N. F., Sankey, T. T., & McNamara, J. P. (2013). Influence of a dense, low-height shrub species on the accuracy of a LiDAR-derived DEM. *Photogrammetric Engineering & Remote Sensing*, *79*(5), 421–431. https://doi.org/10.14358/PERS.79.5.421

Grünewald, T., Stötter, J., Pomeroy, J. W., Dadic, R., Baños, I. M., Marturià, J., et al. (2013). Statistical modelling of the snow depth distribution in open alpine terrain. *Hydrology and Earth System Sciences*, *17*(8), 3005–3021. https://doi.org/10.5194/hess-17-3005-2013

Guan, B., Molotch, N. P., Waliser, D. E., Jepsen, S. M., Painter, T. H., & Dozier, J. (2013). Snow water equivalent in the Sierra Nevada: Blending snow sensor observations with snowmelt model simulations. *Water Resources Research*, *49*, 5029–5046. https://doi.org/10.1002/wrcr.20387

Harpold, A. A., Guo, Q., Molotch, N., Brooks, P., Bales, R., Fernandez-Diaz, J. C., et al. (2014). A LiDAR derived snowpack dataset from mixed conifer forests in the Western U.S. *Water Resources Research*, *50*, 2749–2755. https://doi.org/10.1002/2013WR013935.

Hedrick, A., Marshall, H.-P., Winstral, A., Elder, K., Yueh, S., & Cline, D. (2015). Independent evaluation of the SNODAS snow depth product using regional-scale lidar-derived measurements. *The Cryosphere*, *9*, 13–23. https://doi.org/10.5194/tc-9-13-2015

Hedrick, A. R., Marks, D., Havens, S., Robertson, M., Johnson, M., Sandusky, M., et al. (2018). Direct insertion of NASA Airborne Snow Observatory-derived snow depth time series into the iSnobal energy balance snow model. *Water Resources Research*, *54*, 8045–8063. https://doi.org/10.1029/2018WR023190

Henn, B., Painter, T. H., Bormann, K. J., McGurk, B., Flint, A. L., Flint, L. E., et al. (2018). High-elevation evapotranspiration estimates during drought: Using streamflow and NASA airborne snow observatory SWE observations to close the upper tuolumne river basin water balance. *Water Resources Research*, *54*, 746–766. https://doi.org/10.1002/2017WR020473

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

Kampf, S. K., & Lefsky, M. A. (2016). Transition of dominant peak flow source from snowmelt to rainfall along the Colorado Front Range: Historical patterns, trends, and lessons from the 2013 Colorado Front Range floods. *Water Resources Research*, *52*, 407–422. https://doi.org/10.1002/2015WR017784

Kim E, C. Gatebe, D. Hall, J. Newlin, A. Misakonis, K. Elder, et al. (2018), NASA's snowex campaign: Observing seasonal snow in a forested environment, *Geoscience and Remote Sensing Symposium (IGARSS),* 2017 *IEEE International*, doi:https://doi.org/10.1109/IGARSS.2017.8127222.

Kojima, K. (1967). Densification of seasonal snow cover. *Physics of Snow and Ice: proceedings* = 雪氷の物理学: 論文集, *1*(2), 929–952.

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*, 7309–7342. https://doi.org/10.1002/2015WR017616

Liu, X. (2008). Airborne LiDAR for DEM generation: Some critical issues. *Progress in Physical Geography*, *32*(1), 31–49. https://doi.org/10.1177/0309133308089496

López-Moreno, J. I., Fassnacht, S. R., Heath, J. T., Musselman, K. N., Revuelto, J., Latron, J., et al. (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

Mankin, J. S., Viviroli, D., Singh, D., Hoekstra, A. Y., & Diffenbaugh, N. S. (2015). The potential for snow to supply human water demand in the present and future. *Environmental Research Letters*, *10*(11), 114016). https://doi.org/10.1088/1748-9326/10/11/114016

Marks, D., Domingo, J., Susong, D., Link, T., & Garen, D. (1999). A spatially distributed energy balance snowmelt model for application in mountain basins. *Hydrological Processes*, *13*, 1935–1959. https://doi.org/10.1002/(SICI)1099-1085(199909)13:12/13<1935:AID-HYP868>3.0.CO;2-C

Marquardt, D. (1963). An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics*, *11*(2), 431–441. https://doi.org/10.1137/0111030

McGaughey, R. (2012), FUSION/LDV: Software for LIDAR data analysis and visualization, version 3.01. US Department of Agriculture, Forest Service, 151, Pacific Northwest Research Station, University of Washington. Available online at: https://w3.ual.es/GruposInv/ProyectoCostas/FUSION_manual.pdf (last accessed 1 April 2018).

Meng, X., Currit, N., & Zhao, K. (2010). Ground filtering algorithms for airborne LiDAR data: A review of critical issues. *Remote Sensing*, *2*(3), 861–873. https://doi.org/10.3390/rs2030861

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

Molotch, N. P., Brooks, P. D., Burns, S. P., Litvak, M., Monson, R. K., McConnell, J. R., & Musselman, K. (2009). Ecohydrological controls on snowmelt partitioning in mixed-conifer sub-alpine forests. *Ecohydrology*, *2*, 129–142. https://doi.org/10.1002/eco.48

Musselman, K., Molotch, N. P., & Brooks, P. D. (2008). Effects of vegetation on snow accumulation and ablation in a mid-latitude sub-alpine forest. *Hydrologic Processes*, *22*(15), 2767–2776. https://doi.org/10.1002/hyp.7050

Nolin, A. (2010). Recent advances in remote sensing of seasonal snow. *Journal of Glaciology*, *56*(200), 1141–1150. https://doi.org/10.3189/002214311796406077

Painter, T. H., Berisford, D. F., Boardman, J. W., Bormann, K. J., Deems, J. S., Gehrke, F., et al. (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*, 139–152. https://doi.org/10.1016/j.rse.2016.06.018

Peterson, N. R., & Brown, A. J. (1975). Accuracy of snow measurements. *Proceedings of 43rd Western Snow Conference, Coronado, CA, Western Snow Conference* (pp. 1–5).

Quantum Spatial (2013), 4FRI LiDAR: Four forests restoration initiative, Corvallis, OR, December 27, 2013.

Quantum Spatial (2014), 4FRI LiDAR: Four forests restoration initiative technical report, Corvallis, OR, February 17, 2014.

Quantum Spatial (2017a), 4FRI Snow Analysis II, Arizona LIDAR, Technical Data Report, Corvallis, OR, October 5, 2017.

Quantum Spatial (2017b), 4FRI Snow Analysis II, Arizona LIDAR, Technical Data Report, Corvallis, OR, March 24, 2017.

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*, 3700–3709. https://doi.org/10.1002/2016GL071999

Snauffer, A. M., Hsieh, W. W., Cannon, A. J., & Schnorbus, M. A. (2018). Improving gridded snow water equivalent products in British Columbia, Canada: Multi-source data fusion by neural network models. *The Cryosphere*, *12*(3), 891–905. https://doi.org/10.5194/tc-12-891-2018

Sturm, M., Holmgren, J., & Liston, G. E. (1995). A seasonal snow cover classification system for local to global applications. *Journal of Climate*, *8*(5), 1261–1283. https://doi.org/10.1175/1520-0442(1995)008<1261:ASSCCS>2.0.CO;2

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*, 1380–1394. https://doi.org/10.1175/2010JHM1202.1

Tabari, H., Marofi, S., Abyaneh, H. Z., & Sharifi, M. R. (2010). Comparison of artificial neural network and combined models in estimating spatial distribution of snow depth and snow water equivalent in Samsami basin of Iran. *Neural Computing and Applications*, *19*(4), 625–635. https://doi.org/10.1007/s00521-009-0320-9

Tedesco, M., Pulliainen, J., Takala, M., Hallikainen, M., & Pampaloni, P. (2004). Artificial neural network-based techniques for the retrieval of SWE and snow depth from SSM/I data. *Remote Sensing of Environment*, *90*(1), 76–85. https://doi.org/10.1016/j.rse.2003.12.002

Tinkham, W. T., Smith, A. M., Marshall, H. P., Link, T. E., Falkowski, M. J., & Winstral, A. H. (2014). Quantifying spatial distribution of snow depth errors from LiDAR using Random Forest. *Remote Sensing of Environment*, *141*, 105–115. https://doi.org/10.1016/j.rse.2013.10.021

Trujillo, E., Ramírez, J. A., & Elder, K. J. (2007). Topographic, meteorologic, and canopy controls on the scaling characteristics of the spatial distribution of snow depth fields. *Water Resources Research*, *43*, W07409. https://doi.org/10.1029/2006WR005317

Veatch, W., Brooks, P. D., Gustafson, J. R., & Molotch, N. P. (2009). Quantifying the effects of forest canopy cover on net snow accumulation at a continental, mid-latitude site. *Ecohydrology*, *2*(2), 115–128. https://doi.org/10.1002/eco.45

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