Author responses in red.

Dear authors,
Thank you for your very thorough response to the reviewers' comments. I have a few very minor points that I would like to see addressed before I can recommend your manuscript for publication.
Thank you for your comments.

- The use of R2 is a bit imprecise. Typically, the coefficient of determination is used as an indicator for the correlation between two datasets. However, in some places (e.g lines 590-594, 719-723) the two datasets are not explicitly stated. Presumably, it is observations vs. model results but it would be a help to the reader if this was clarified (e.g. similar to what is stated in lines 643-646).
We have clarified the datasets that are the basis for these R^2 calculations.

- The figures showing maps of the two glacier (Figs. 3, 8,10 and 11) would be easier to read if a "G" and "W" were added (if possible).
We have added text labels "Wolverine and Gulkana" to Figure 3 to improve readability of this and subsequent figures.

- In response to reviewer #1 's comment to line 642, please state that details on interpolation scheme and geodetic calibrations can be found in van Beusekom et al., 2010 and O'Neel et al., 2014.
We have added these references in the requested location.

- I agree with M. Pelto's request to line 387 re. mentioning how much the observed observations exceeded model results. In view of the large amount of data, it would be sufficient to give an example so the reader has an idea of the magnitude of the values. Something along the lines of "For example, in 2015 observed SWE exceeded modelled SWE by more/less than X amount for Y% of data points".
We have added the suggested calculation for a year as an example of the residual variability.

I hope that you are willing to incorporate these changes so we can move forward with your interesting study.
Best,
Nanna

Interannual snow accumulation variability on glaciers derived from repeat, spatially extensive ground-penetrating radar surveys

Daniel McGrath[1], Louis Sass[2], Shad O'Neel[2] Chris McNeil[2], Salvatore G. Candela[3], Emily H. Baker[2], and Hans-Peter Marshall[4]

[1]Department of Geosciences, Colorado State University, Fort Collins, CO

[2]U.S. Geological Survey Alaska Science Center, Anchorage, AK

[3]School of Earth Sciences and Byrd Polar Research Center, Ohio State University, Columbus, OH

[4]Department of Geosciences, Boise State University, Boise, ID

Abstract

There is significant uncertainty regarding the spatiotemporal distribution of seasonal snow on glaciers, despite being a fundamental component of glacier mass balance. To address this knowledge gap, we collected repeat, spatially extensive high-frequency ground-penetrating radar (GPR) observations on two glaciers in Alaska during the spring of five consecutive years. GPR measurements showed steep snow water equivalent (SWE) elevation gradients at both sites; continental Gulkana Glacier's SWE gradient averaged 115 mm 100 m^{-1} and maritime Wolverine Glacier's gradient averaged 440 mm 100 m^{-1} (over >1000 m). We extrapolated GPR point observations across the glacier surface using terrain parameters derived from digital elevation models as predictor variables in two statistical models (stepwise multivariable linear regression and regression trees). Elevation and proxies for wind redistribution had the greatest explanatory power, and exhibited relatively time-constant coefficients over the study period. Both statistical models yielded comparable estimates of glacier-wide average SWE (1 % average difference at Gulkana, 4 % average difference at Wolverine), although the spatial distributions produced by the models diverged in unsampled regions of the glacier, particularly at Wolverine. In total, six different methods for estimating the glacier-wide winter balance average agreed within ± 11 %. We assessed interannual variability in the spatial pattern of snow accumulation predicted by the statistical models using two quantitative metrics. Both glaciers exhibited a high degree of temporal stability, with ~85 % of the glacier area experiencing less than 25 % normalized absolute variability over this five-year interval. We found SWE at a sparse network (3 stakes per glacier) of long-term glaciological stake sites to be highly correlated with the GPR-derived glacier-wide average. We estimate that interannual variability in the spatial pattern of winter SWE accumulation is only a small component (4–10 % of glacier-wide average) of the total mass balance uncertainty and thus, our findings support the concept

that sparse stake networks effectively measure interannual variability in winter balance on glaciers, rather than some temporally varying spatial pattern of snow accumulation.

1. Introduction

Our ability to quantify glacier mass balance is dependent on accurately resolving the spatial and temporal distributions of snow accumulation and snow/ice ablation. Significant advances in our knowledge of ablation processes have improved observational and modelling capacities (Hock, 2005; Huss and Hock, 2015; Fitzpatrick et al., 2017), yet comparable advances in our understanding of the distribution of snow accumulation have not kept pace (Hock et al., 2017). Reasons for this discrepancy are two-fold: (i) snow accumulation exhibits higher variability than ablation, both in magnitude and length scale, largely due to wind redistribution in the complex high-relief terrain where mountain glaciers are typically found (Kuhn et al., 1995) and (ii) accumulation observations are typically less representative (i.e., one stake in a few hundred meter elevation band) or less effective than comparable ablation observations (i.e., precipitation gage measuring snowfall vs. radiometer measuring short-wave radiation). This discrepancy presents a significant limitation to process-based understanding of mass balance drivers. Furthermore, a warming climate has already modified – and will continue to modify – the magnitude and spatial distribution of snow on glaciers through a reduction in the fraction of precipitation falling as snow and an increase in rain-on-snow events (McAfee et al., 2013; Klos et al., 2014; McGrath et al., 2017; Beamer et al., 2017; Littell et al., 2018).

Deleted: Knowles et al., 2006;

Significant research has been conducted on the spatial and, to a lesser degree, the temporal variability of seasonal snow in mountainous and high-latitude landscapes (e.g., Balk and Elder, 2000; Molotch et al., 2005; Erickson et al., 2005; Deems et al., 2008; Sturm and Wagner, 2010; Schirmer et al., 2011; Winstral and Marks, 2014; Anderson et al., 2014; Painter et al., 2016). Although major advances have occurred in applying physically-based snow distribution models (i.e., iSnobal (Marks et al., 1999), SnowModel (Liston and Elder, 2006), Alpine 3D (Lehning et al., 2006)), the paucity of required meteorological forcing data proximal to glaciers limits widespread application. Many other studies have successfully developed statistical approaches that rely on the relationship between the distribution of snow water equivalent (SWE) and physically-based terrain parameters (also referred to as physiographic or topographic properties or variables) to model the distribution of SWE across entire basins (e.g., Molotch et al., 2005; Anderson et al., 2014; Sold et al., 2013; McGrath et al., 2015).

A major uncertainty identified by these studies is the degree to which these statistically derived relationships remain stationary in time. Many studies (Erickson et al., 2005; Deems et al., 2008; Sturm and Wagner, 2010; Schirmer et al, 2011; Winstral and Marks, 2014; Helfricht et al., 2014) have found 'time-stability' in the distribution of SWE, including locations where wind redistribution is a major control on this distribution. For instance, a climatological snow distribution pattern, produced from the mean of nine standardized surveys, accurately predicted the observed snow depth in a subsequent survey in a tundra basin in Alaska (~4–10 cm root mean square error (RMSE); Sturm and Wagner, 2010). Repeat LiDAR surveys over two years at three hillslope-scale study plots in the Swiss Alps found a high degree of correlation (r=0.97) in snow depth spatial patterns (Schirmer et al., 2011). They found that the final snow depth distributions at the end of the two winter seasons were more similar than the distributions of any two individual storms during that two-year period (Schirmer et al., 2011). Lastly, an 11-year study of extensive snow probing (~1200 point observations) at a 0.36 km^2 field site in southwestern Idaho found consistent spatial patterns (r=0.84; Winstral and Marks, 2014). Collectively, these studies suggest that in landscapes characterized by complex topography and extensive wind redistribution of snow, spatial patterns are largely time-stable or stationary, as long as the primary drivers are stationary.

Even fewer studies have explicitly examined the question of interannual variability in the context of snow distribution on glaciers. Spatially-extensive snow probe datasets are collected by numerous glacier monitoring programs (e.g., Bauder et al., 2017; Kjøllmoen et al., 2017; Escher-Vetter et al., 2009) in order to calculate a winter mass balance estimate. Although extensive, such manual approaches are still limited by the number of points that can be collected and uncertainties in correctly identifying the summer surface in the accumulation zone, where seasonal snow is underlain by firn. One study of two successive end-of-winter surveys of snow depth using probes on a glacier in Svalbard found strong interannual variability in the spatial distribution of snow, and the relationship between snow distribution and topographic features (Hodgkins et al., 2006). Elevation was found to only explain 38–60 % of the variability in snow depth, and in one year, snow depth was not dependent on elevation in the accumulation zone (Hodgkins et al., 2006). Instead, aspect, reflecting relative exposure or shelter from prevailing winds, was found to be a significant predictor of accumulation patterns. In contrast, repeat airborne LiDAR surveys of a ~36 km^2 basin (~50% glacier cover) in Austria over five winters found that the glacierized area exhibited less interannual variability (as measured by the interannual standard deviation) than the non-glacierized sectors of the basin (Helfricht et al., 2014). Similarly, a three-year study of snow distribution on Findelgletscher in the Swiss Alps using ground-penetrating radar (GPR) found low interannual variability, as 86 % of the glacier area experienced less than 25 % normalized relative variability (Sold et al., 2016). These latter studies suggest that seasonal snow distribution on glaciers likely exhibits 'time-stability' in its distribution, but few datasets exist to robustly test this hypothesis.

The 'time-stability' of snow distribution on glaciers has particularly important implications for long-term glacier mass balance programs, as seasonal and annual mass balance solutions are derived from the integration of a limited number of point observations (e.g., 3 to 50 stakes), and the assumption that stake and snow pit observations accurately represent interannual variability in mass balance rather than interannual variability in the spatial patterns of mass balance. Previous work has shown 'time-stability' in the spatial pattern of annual mass balance (e.g., Vincent et al., 2017) and while this is important for understanding the uncertainties in glacier-wide mass balance estimates, the relative contributions of accumulation and ablation to this stability are poorly constrained, thereby hindering a process-based understanding of these spatial patterns. Furthermore, accurately quantifying the magnitude and spatial distribution of winter snow accumulation on glaciers is a prerequisite for understanding the water budget of glacierized basins, with direct implications for any potential use of this water, whether that be ecological, agricultural, or human consumption (Kaser et al., 2010).

To better understand the 'time-stability' of the spatial pattern of snow accumulation on glaciers, we present five consecutive years of extensive GPR observations for two glaciers in Alaska. First, we use these GPR-derived SWE measurements to train two different types of statistical models, which were subsequently used to spatially extrapolate SWE across each glacier's area. Second, we assess the temporal stability in the resulting spatial distribution in SWE. Finally, we compare GPR-derived winter mass balance estimates to traditional glaciological derived mass balance estimates and quantify the uncertainty that interannual variability in spatial patterns in snow accumulation introduces to these estimates.

2. Study Area

During the spring seasons of 2013 – 2017, we conducted GPR surveys on Wolverine and Gulkana glaciers, located on the Kenai Peninsula and eastern Alaskan Range in Alaska (Fig. 1). These glaciers have been studied as part of the U.S. Geological Survey's Benchmark Glacier project since 1966 (O'Neel et al., 2014). Both glaciers are ~16 km^2 in area and span ~1200 m in elevation (426 – 1635 m asl for Wolverine, 1163 – 2430 m asl for Gulkana). Wolverine Glacier exists in a maritime climate, characterized by warm air temperatures (mean annual temperature = –0.2 °C at 990 meters; median equilibrium line altitude for 2008 – 2017 is 1235 m asl) and high precipitation (median glacier-wide winter balance = 2.0 m water equivalent (m w.e.)), while Gulkana is located in a continental climate, characterized by colder air temperatures (mean annual temperature = –2.8 °C at 1480 meters; median equilibrium line altitude for 2008 – 2017 is 1870 m asl) and less precipitation (median glacier-wide winter balance = 1.2 m w.e.) (Fig. 2). The cumulative mass balance time series for both glaciers is negative (~ –24 m w.e. between 1966–2016), with Gulkana showing a more monotonic decrease over the entire study interval, while Wolverine exhibited near equilibrium balance between 1966 and 1987, and sharply negative to present (O'Neel et al., 2014; O'Neel et al., 2018).

3. Methods

The primary SWE observations are derived from a GPR measurement of two-way travel time (*twt*) through the annual snow accumulation layer. We describe five main steps to convert *twt* along the survey profiles to annual distributed SWE products for each glacier. These include (i) acquisition of GPR and ground-truth data, (ii) calculation of snow density and associated radar velocity, which are used to convert measured *twt* to annual layer depth and subsequently SWE, and (iii) application of terrain parameter statistical models to extrapolate SWE across the glacier area. We then describe approaches to (iv) evaluate the temporal consistency in spatial SWE patterns and (v) compare GPR-derived SWE and direct (glaciological) winter mass balances.

3.1. Radar data collection and processing

Common-offset GPR surveys were conducted with a 500 MHz Sensors and Software pulseEkko Pro system in late spring close to maximum end-of-winter SWE and prior to the onset of extensive surface melt. GPR parameters were set to a waveform-sampling rate of 0.1 ns, a 200-ns time window, and “Free Run” trace increments, where samples are collected as fast as the processor allows, instead of at uniform temporal or spatial increments.

Deleted: P

Deleted:

In general, GPR surveys were conducted by mounting a plastic sled behind a snowmobile and driving at a near-constant velocity of 15 km h^{-1} (Fig. 3, S1, S2), resulting in a trace spacing of ~20 cm. Coincident GPS data were collected using a Novatel Smart-V1 GPS receiver (Omnistar corrected, L1 receiver with root-mean-square accuracy of 0.9 m (Perez-Ruiz et al., 2011)). We collected a consistent survey track from year-to-year that minimized safety hazards (crevasses, avalanche runouts) but optimized the sampling of terrain parameter space on the glacier (e.g., range and distribution of elevation, slope, aspect, curvature, etc.). However, in 2016 at Wolverine Glacier, weather conditions and logistics did not allow for ground surveys to be completed. Instead, a number of radar lines were collected via a helicopter survey. To best approximate the ground surveys completed in other years, we selected a subset of helicopter GPR observations within 150 m of the ground-based surveys. Previous comparisons between ground and helicopter platforms found excellent agreement in SWE point observations (coefficient of determination (R^2)=0.96, root mean square error=0.14 m; McGrath et al., 2015).

Radargrams were processed using the ReflexW-2D software package (Sandmeier Scientific Software). All radargrams were corrected to time zero, taken as the first negative peak in the direct wave (Yelf and Yelf, 2006), and a dewow filter (mean subtraction) was applied over 2 ns. When reflectors from the base of the seasonal snow cover were insufficiently resolved, gain and band-pass filters were subsequently applied. Layer picking was guided by ground-truth efforts and done semi-automatically using a phase-following layer picker. For further details, please see McGrath et al. (2015).

3.2. Ground truth observations

We collected extensive ground-truth data to validate GPR surveys, including probing and snowpit/cores. In the ablation zone of each glacier, we probed the snowpack thickness every ~500 m along-track. In addition, we measured seasonal snow depth and density at an average of five locations (corresponding to the glaciological observations; see Section 3.5) on each glacier in each year. Typically these locations include one or two in the ablation zone, one near the long-term ELA, and two or more in the accumulation zone. We measured snow density using a gravimetric approach in snowpits (at 10 cm intervals) and with 7.25 cm diameter cores (if total depth >2 m; at 10–40 cm intervals depending on natural breaks) to the previous summer surface. We calculated a density profile and column-average density, ρ_{site}, at each site.

As snow densities did not exhibit a consistent spatial nor elevation dependency on the glaciers (e.g., Fausto et al., 2018), we calculated a single average density, ρ, of all ρ_{site} on each glacier and each year, which was subsequently used to calculate SWE:

$$SWE = \left(\frac{twt}{2}\right) \cdot v_s \cdot \rho, \qquad (1)$$

where twt is the two-way travel time as measured by the GPR and v_s is the radar velocity. v_s was calculated for each glacier in each year as the average of two independent approaches: (i) an empirical relationship based on the glacier-wide average ρ (Kovacs et al., 1995) and (ii) a least-squares regression between snow depth derived by probing and all radar twt observations within a 3-m radius of the probe site. An exception was made at Wolverine in 2016 as no coincident probe depth observations were made during the helicopter-based surveys. Instead, we estimated the second radar velocity by averaging radar velocities calculated from observed twt and snow depths at three snowpit/core locations.

3.3. Spatial Extrapolation

Extrapolating SWE from point measurements to the basin scale has been a topic of focused research for decades (e.g., Woo and Marsh, 1978; Elder et al., 1995; Molotch et al., 2005). Most commonly, the dependent variable SWE is related to a series of explanatory terrain parameters, which are proxies for the physical processes that actually control SWE distribution across the landscape. These include orographic gradient in precipitation (elevation), wind redistribution of existing snow (slope, curvature, drift potential), and aspect with respect to solar radiation and prevailing winds (eastness, northness). We derived terrain parameters from 10-m resolution digital elevation models (DEMs) sourced from the ArcticDEM project (Noh and Howat, 2015) for Gulkana and produced from airborne Structure from Motion photogrammetry at Wolverine (Nolan et al., 2015). Both DEMs were based on imagery from August 2015. Specifically, these parameters include elevation, surface slope, surface curvature, northness (Molotch et al., 2005), eastness, and snow drift potential (*Sb*) (Winstral et al., 2002; Winstral et al., 2013; Fig. S3, S4). The *Sb* parameter is commonly used to identify locations where airflow separation occurs based on both near and far-field topography and are thus likely locations to accumulate snow drifts (Winstral et al., 2002). For specific details on this calculation, please refer to Winstral et al. (2002). In the application of *Sb* here, we determined the principle direction by calculating the modal daily wind direction during the winter (October – May) when wind speeds exceeded 5 m s^{-1} (~minimum wind velocity for snow transport; Li and Pomeroy, 1997). The length scales for curvature were found using an optimization scheme that identified the highest model R^2.

Prior to spatial extrapolation, we aggregated GPR observations to the resolution of the DEM by calculating the median value of all observations within each 10 m pixel of the DEM. We then utilized two approaches to extrapolate GPR point observations across the glacier surface: (i) least-squares elevation gradient applied to glacier hypsometry and (ii) statistical models. For (i), we derived SWE elevation gradients in two ways; first, solely on observations that followed the glacier centerline and second, from the entire spatially-extensive dataset. For (ii), we utilized two different models: stepwise multivariable linear regressions and regression trees (Breiman et al., 1984). All of these approaches produced a spatially-distributed SWE field over the entire glacier area. Individual points in this field are equivalent to point winter balances (b_w; m w.e.). From the distributed b_w field, we calculated a mean area-averaged winter balance (B_w; m w.e.).

Deleted: both

Additionally, we implemented a cross-validation approach to the statistical models (multivariable regression and regression tree), whereby 75 % of the aggregated observations were used for training and 25 % were used for testing. However, rather than randomly selecting pixels from across the entire dataset, we randomly selected a single pixel containing aggregated GPR observations and then extended this selection out along continuous survey lines until we reached 25 % of the total observational dataset, thus removing entire sections (and respective terrain parameters) from the analysis (Fig. S5). This approach provided a more realistic test for the statistical models, as the random selection of individual cells did not significantly alter terrain-parameter distributions. For each glacier and each year, we produced 100 training/test dataset combinations, but rather than take the single model with the highest R^2 or lowest RMSE (between modelled SWE and the GPR-derived test dataset), we produced a distributed SWE product by taking the median value for each pixel from all 100 model runs and a glacier-wide median value that is the median of all 100 individual *Bw* estimates. We chose the median-value approach over a highest R^2/lowest RMSE approach that is often utilized because, despite being randomly selected, some training datasets were inherently advantaged by a more complete sampling of terrain parameter distributions. These iterations resulted in the highest R^2/lowest RMSE when applied to the training dataset, but weren't necessarily

Deleted: extrapolations

Deleted: from the resulting

Deleted:

Deleted: test dataset

Deleted: distribution of

Deleted:

Deleted: s

indicative of a better model, particularly in the context of being able to predict SWE at locations on the glacier where the terrain parameter space had not been well sampled.

Deleted: .

3.3.2. Stepwise Multivariable Linear Regression

We used a stepwise multivariable linear regression model of the form,

$$SWE_{(i,j)} = c_1 x_{1(i,j)} + c_2 x_{2(i,j)} + \ldots + c_n x_{n(i,j)} + \varepsilon_{(i,j)}, \quad (2)$$

where $SWE_{(i,j)}$ is the predicted (standardized) value at location i,j and c_1, c_2, c_n are the beta coefficients of the model, x_1, x_2, x_n are terrain parameters which are independent variables that have been standardized and ε is the residual. We applied the regression model stepwise and included an independent variable if it minimized the Akaike information criterion (AIC; Akaike, 1974). We present the beta coefficients from each regression (each year, each glacier) to explore the temporal stability of these terms.

3.3.3. Regression Trees

Regression trees (Breiman et al., 1984) provide an alternative statistical approach for extrapolating point observations by recursively partitioning SWE into progressively more homogenous subsets based on independent terrain parameter predictors (Molotch et al., 2005; Meromy et al., 2013; Bair et al., 2018). The primary advantage of the regression tree approach is that each terrain parameter is used multiple times to partition the observations, thereby allowing for non-linear interactions between these terms. In contrast, the MVR only allows for a single "global" linear relationship for each parameter across the entire parameter-space. We implemented a random forest approach (Breiman, 2001) of repeated regression trees (100 learning cycles) in Matlab, using weak learners and bootstrap aggregating (bagging; Breiman, 1996). Each weak learner omits 37% of observations, such that these "out-of-bag" observations are used to calculate predictor importance. The use of this ensemble/bagging approach reduces overfitting and thus precludes having to subjectively prune the tree and provides more accurate and unbiased error estimates (Breiman, 2001). Prior to implementing the regression tree, we removed the SWE elevation gradient from the observations using a least-squares regression. As described in the results, elevation is the dominant independent variable and as our observations (particularly at Wolverine) did not cover the entire elevation range, the regression tree approach was not well suited to predicting SWE at elevations outside of the observational range.

3.4. Interannual variability in spatial patterns

We quantified the stability of spatial patterns in SWE across the five-year interval using two approaches: (i) normalized range and (ii) the coefficient of determination. In the first approach, we first divided each pixel in the distributed SWE fields by the glacier-wide average, B_w, for each year and each glacier, and then calculated the range in these normalized values over the entire five-year interval. For example, if a cell had normalized values of 84 %, 92 %, 106 %, 112 % and 120 %, the normalized range would be 36 %. A limitation of this approach is that it is highly sensitive to outliers, such that a single year can substantially increase this range. This is similar to an approach presented by Sold et al. (2016), but unlike their calculation (their Fig. 9), the normalized values reported here have not been further normalized by the normalized mean of that pixel over the study interval. Thus, the values reported here are an absolute normalized range, whereas Sold et al. (2016) report a relative normalized range. In the coefficient of determination (R^2) approach, we computed the least-squares regression correlation between the SWE in each pixel and the glacier-wide average, B_w, derived from the MVR model over the five-year period. For this approach, cells with a higher R^2 scale linearly with the glacier-wide average, while those with low R^2 do not.

Deleted: has

3.5. Glaciological mass balance

Beginning in 1966, glacier-wide seasonal (winter, B_w; summer, B_s) and annual balances (B_a) were derived from glaciological measurements made at three fixed locations on each glacier. The integration of these point measurements was accomplished using a site-index method – equivalent to an area-weighted average (March and Trabant, 1996; van Beusekom et al., 2010). Beginning in 2009, a more extensive stake network of seven to nine stakes was established on each glacier, thereby facilitating the use of a balance profile method for spatial extrapolation (Cogley et al., 2011). Systematic bias in the glaciological mass balance time-series is removed via a geodetic adjustment derived from DEM differencing over decadal timescales (e.g., O'Neel et al., 2014). For this study, glaciological measurements were made within a day of the

GPR surveys, and integrated over the glacier hypsometry using both the historically applied site-index method (based on the long-term three stake network) and the more commonly applied balance profile method (based on the more extensive stake network). We utilized a single glacier hypsometry, derived from the 2015 DEMs, for each glacier over the entire five-year interval. Importantly, in order to facilitate a more direct comparison to the GPR-derived B_w estimates, we used glaciological B_w estimates that have not been geodetically calibrated.

4. Results

4.1. General accumulation conditions

Since 1966, Wolverine Glacier's median B_w (determined from the stake network) exceeds Gulkana's by more than a factor of two (2.3 vs. 1.1 m w.e.), and exhibits greater variability, with an interquartile range more than twice as large (0.95 m w.e. vs. 0.4 m w.e.). Over the five-year study period, both glaciers experienced accumulation conditions that spanned their historical ranges, with one year in the upper quartile (including the 5th greatest B_w at Wolverine in 2016), one year within 25% of the median, and multiple years in the lower quartile (2017 at Gulkana and 2014 at Wolverine had particularly low B_w values) (Fig. 2). In all years, B_w at Wolverine was greater, although in 2013 and 2014, the difference was only 0.1 m w.e.

Average accumulation season (taken as October 1 – May 31) wind speeds over the study period were stronger (~7 m s^{-1} vs. ~3 m s^{-1}) and from a more consistent direction at Wolverine than Gulkana (northeast at Wolverine, southwest to northeast at Gulkana) (Fig. S6). On average, Wolverine experienced ~50 days with wind gusts >15 m s^{-1} each winter, while for Gulkana, this only occurred on ~7 days. Over the five-year study period, interannual variability in wind direction was very low at Wolverine (2016 saw slightly greater variability, with an increase in easterly winds). In contrast, at Gulkana, winds were primarily from the northeast to east in 2013–2015, from the southwest to south in 2016–2017, and experienced much greater variability during any single winter.

4.2. *In situ* and GPR point observations

Glacier-averaged snow densities across all years were 440 kg m^{-3} (range 414–456 kg m^{-3}) at Wolverine and 362 kg m^{-3} (range 328–380 kg m^{-3}) at Gulkana (Table S1). Average radar velocities were 0.218 m ns^{-1} (range 0.207–0.229 m ns^{-1}) at Wolverine and 0.223 m ns^{-1} (0.211–0.231 m ns^{-1}) at Gulkana. Over this five-year interval, the GPR point observations revealed a general pattern of increasing SWE with elevation, along with fine-scale variability due to wind redistribution (e.g., upper elevations of Wolverine) and localized avalanche input (e.g., lower west branch of Gulkana) (Fig. S1, S2). The accumulation season (hereafter, winter) SWE elevation gradient was steeper (~440 vs. ~115 mm 100 m^{-1}) and more variable in its magnitude at Wolverine than Gulkana. Gradients ranged between 348 – 624 mm 100 m^{-1} at Wolverine, and 74 – 154 mm 100 m^{-1} at Gulkana (Fig. 4). Over all five years at both glaciers, elevation explained between 50 % and 83 % of the observed variability in SWE (Fig. 4).

4.3. Model performance

To evaluate model performance in unsampled locations of the glacier, both extrapolation approaches were run 100 times for each glacier and each year, each time with a unique, randomly selected training (75 % of aggregated observations) and test (remaining 25 % of aggregated observations) dataset. The median and standard deviation of the coefficients of determination (R^2) between modeled SWE and the test datasets for the 100 models runs are shown in Fig. 5. Model performance ranged from 0.25 to 0.75, but on average, across both glaciers and all years, was 0.56 for the MVR approach and 0.46 for the regression tree. Model performance was higher and more consistent at Wolverine, whereas 2015 and 2017 at Gulkana had test dataset R^2 of ~0.4 and 0.3, likely reflecting the lower winter SWE elevation gradients and coefficients of determination with elevation during these years (Fig. 4). The wide range in R^2 across the 100 model runs reflects the variability in training and test datasets that were randomly selected. When the test dataset terrain parameter space was captured by the training dataset, a high coefficient of determination resulted, but when the test dataset terrain parameter space was exclusive (e.g., contained only a small elevation range), the model performance was typically low. This further highlights the importance of elevation as a predictor for these glaciers.

Deleted: from these

Deleted: ,

Deleted: ,

At Gulkana, the model residuals (Fig. S1) exhibited spatiotemporal consistency, with positive residuals (i.e., observed SWE exceeded modeled SWE by ~0.2 m w.e.) at mid-elevations of the west branch, and at the very terminus of the glacier. The largest negative residuals typically occurred at the highest elevations. In both cases, these locations deviated from the overall SWE elevation gradient. At Wolverine, observations at the highest elevations typically exceeded the modeled SWE (i.e., positive residuals), particularly at the highest elevations of the northeast corner where wind drifting is particularly prevalent (Fig. S2). For example, in 2015, nearly 80% of the residuals in this section were positive and had a median value of 0.4 m. Elsewhere at Wolverine, the residuals often alternated between positive and negative values over length scales of 10s to 100s of meters (Fig. S2), which we interpret as zones of scour/drift not captured by the MVR model.

Deleted: in the northeast quadrant of the glacier

Deleted: that were better captured by the regression tree models.

The beta coefficients of terrain parameters from the MVR were fairly consistent from year-to-year at both glaciers (Fig. 6). At Wolverine, elevation was the largest beta coefficient, followed by *Sb* and curvature. At Gulkana, elevation was also the largest beta coefficient, followed by curvature. Gulkana experiences much greater variability in wind direction during the winter months (Fig. S6), possibly explaining why *Sb* was either not included or had a very low beta coefficient in the median regression model. As our surveys were completed prior to the onset of ablation, terrain parameters related to solar radiation gain (notably the terms that include aspect: northness and eastness) had small and variable beta coefficients.

4.4. Spatial Variability

A common approach for quantifying snow accumulation variability across a range of means is the coefficient of variation (CoV), which is calculated as the ratio of the standard deviation to the mean (Liston et al., 2004; Winstral and Marks, 2014). The mean and standard deviation of CoVs at Wolverine were 0.42 ± 0.03 and at Gulkana, 0.29 ± 0.05, indicating relatively lower spatial variability in SWE at Gulkana (Fig. 7). CoVs were fairly consistent across all five years, although 2017 saw the largest CoVs at both glaciers. Interestingly, 2017 had the lowest absolute spatial variability (i.e., lowest standard deviation), but also the lowest glacier-wide averages during the study period, resulting in greater CoVs.

Qualitatively, both Wolverine and Gulkana glaciers exhibited consistent spatiotemporal patterns in accumulation across the glacier surface, with elevation exerting a first-order control (Fig. 8, S7, S8). Overlaid on the strong elevational gradient are consistent locations of wind scour and deposition, reflecting the interaction of wind redistribution and complex – albeit relatively stable year to year – surface topography (consisting of both land and ice topography). For instance, numerous large drifts (~2 m amplitude, ~200 m wavelength) occupy the northeast and northwest corners of Wolverine Glacier, where prevailing northeasterly winds consistently redistributed snow into sheltered locations in each year of the study period (Fig. 8). The different statistical extrapolation approaches produced nearly identical B_w estimates (4 % difference on average at Wolverine and 1 % difference on average at Gulkana) (Fig. 9). The MVR B_w estimate was larger in 4 out of 5 years at Wolverine (Fig. 9), while neither approach exhibited a consistent bias at Gulkana.

Although the glacier-wide averages between these approaches showed close agreement, we explored the differences in spatial patterns by calculating a mean SWE difference map for each glacier by differencing the five-year mean SWE produced by the regression tree model from the same produced by the MVR model (Fig. 10). As such, locations where the MVR exceeded the regression tree are positive (yellow). At Gulkana, where the two approaches showed slightly better glacier-wide B_w agreement, the magnitude in individual pixel differences were substantially less than at Wolverine (e.g., color bar scales range ± 0.2 m at Gulkana vs. ± 0.5 m at Wolverine). At Wolverine Glacier, there were three distinct elevation bands where the MVR approach predicted greater SWE, namely the main icefall in the ablation zone, a region of complex topography centered around a normalized elevation of 0.65, and lastly, at higher elevations, where both approaches predicted a series of drift and scour zones, although in sum, the MVR model predicted greater SWE.

We used two different approaches to quantify the 'time-stability' of spatial patterns across these glaciers. By the first metric, normalized range, we found that both glaciers exhibited very similar patterns (Fig. 11), with either ~65 or 85 % (regression tree and MVR, respectively) of the glacier area experiencing less than 25 % absolute normalized variability (Fig. 12). The R^2 approach provides an alternative way of assessing the time stability of SWE, essentially determining whether SWE at each location scales with the glacier-wide value. By this metric, 80 % of the glacier area at Wolverine and 96 % of the glacier area at Gulkana (based on MVR model) had a coefficient of determination greater than 0.8 (Fig. 12), suggesting that most locations on the glacier have a consistent relationship with the mean glacier-wide mass balance. By both metrics, the MVR output suggests greater 'time-stability' (e.g., lower normalized range or higher R^2) compared to the regression tree.

4.5. Winter mass balance

In order to examine systematic variations between the approaches we outlined in Section 3 for calculating the glacier-wide winter balance, B_w, we first calculated a yearly mean from the six approaches (including four based on the GPR observations: MVR, regression tree, elevation gradient derived from centerline only observations, elevation gradient derived from all point observations, and two based on the *in situ* stake network: site-index and profile). In general, Gulkana exhibited greater agreement (4 % average difference) among the approaches, with most approaches agreeing within 5 % of the six-approach mean (Fig. 13; Table S2). Wolverine showed slightly less agreement (7 % average difference), as the two terrain parameters statistical extrapolations (MVR and regression tree) produced B_w estimates ~9 % above the mean, while the two stake derived estimates were ~7 % less than the mean. On average across all five years at Wolverine, the MVR approach was the most positive, while the glaciological site-index approach was always the most negative (Fig. 13). At both glaciers, the estimates using elevation as the only predictor yielded B_w estimates on average within 3 % of the six-method mean, with the centerline only based estimate being slightly negatively biased, and the complete observations being slightly positively biased.

To examine the systematic difference between the glaciological site-index method and GPR-based MVR approach, we compared stake-derived b_w values from the three long-term stakes to all GPR-based MVR b_w values within that index zone (Fig. 14). Both the stakes and the GPR-derived b_w values have been normalized by the glacier-wide value to make these results comparable across years and glaciers. It is apparent that Wolverine experienced much greater spatial variability in accumulation, with larger interquartile ranges and a large number of positive outliers in all index zones. Importantly, the stake weight in the site-index solution is dependent on the hypsometry of the glacier, and for both glaciers, the upper stake accounts for ~65 % of the weighted average. In years that the misfit between GPR B_w and site-index B_w was largest (2015 and 2016 at Gulkana, 2013 and 2017 at Wolverine), the stake-derived b_w at the upper stake was in the lower quartile of all GPR-derived b_w values, explaining the significant difference in B_w estimates in these years. Potential reasons for this discrepancy are discussed in Section 5.3.

In situ stake and pit observations traditionally serve as the primary tool for deriving glaciological mass balances. However, in order for these observations to provide a systematic and meaningful long-term record, they need to record interannual variability in mass balance rather than interannual variability in spatial patterns of mass balance. To assess the performance of the long-term stake sites, we examined the interannual variability metrics for the stake locations. By both metrics (normalized absolute range and R^2), the middle and upper elevation stakes at both glaciers appear to be in locations that achieve this temporal stability, having exhibited ~10 % range and R^2>0.95 over the five-year interval. The lower elevation stake was less temporally stable and exhibited opposing behavior at each glacier. At Gulkana, this stake had a high R^2 (0.93) and moderate normalized variability (26 %), which in part, reflects the lower total accumulation at this site and the ability for a single uncharacteristic storm to alter this total amount significantly. In contrast, Wolverine's lowest site exhibited both low R^2 (<0.01) and normalized range (2 %), a somewhat unlikely combination. The statistical models commonly predicted zero or near-zero cumulative winter accumulation at this site

Deleted: spatial

Deleted: in

Deleted: extrapolation approaches frequently (i.e., mid-winter rain and/or ablation is common at this site), so although the normalized range was quite low, predicted SWE values were uncorrelated with B_w over the study interval.

Discussion

5.1. Interannual variability in spatial patterns

Each glacier exhibited consistent normalized SWE spatial patterns across the five-year study, reflecting the strong control of elevation and regular patterns in wind redistribution in this complex topography (Fig. 11, S7, S8). This is particularly notable given the highly variable magnitudes of accumulation over the five-year study and the contrasting climate regions of these two glaciers (wet, warm maritime and cold, dry continental), with unique storm paths, timing of annual accumulation, wind direction and wind direction variability, and snow density. At both glaciers, the lowest interannual variability was found away from locations with complex topography and elevated surface roughness, such as crevassed zones, glacier margins, and areas near peaks and ridges.

In the most directly comparable study using repeat GPR surveys at Switzerland's Findelgletscher, 86 % of the glacier area experienced less than 25 % range in relative normalized accumulation over a three-year interval (Sold et al., 2016). As noted in Section 3.4., we reported an absolute normalized range, whereas Sold et al. (2016) reported a relative normalized range. Following their calculation, we found that 81 and 82 % of Wolverine and Gulkana's area experienced a relative normalized range less than 25 %. Collectively, our results add to the growing body of evidence (e.g., Deems et al., 2008; Sturm and Wagner, 2010; Schirmer et al., 2011; Winstral and Marks, 2014) suggesting 'time-stability' in the spatial distribution of snow in locations that span a range of climate zones, topographic complexity, and relief. While the initial effort required to constrain the spatial distribution over a given area can be significant, the benefits of understanding the spatial distribution are substantial and long-lasting, and have a wide range of applications.

5.1.1 Elevation

Elevation explained between 50 and 83 % of the observed SWE variability at Gulkana and Wolverine, making it the most significant terrain parameter at both glaciers every year (Fig. 4, 6). Steep winter SWE gradients characterized both glaciers throughout the study period (115 – 440 mm 100 m^{-1}). Such gradients are comparable to previous results for glaciers in the region (Pelto, 2008; Pelto et al., 2013; McGrath et al., 2015), but exceed reported orographic precipitation gradients in other mountainous regions by a factor of 2–3 (e.g., Anderson et al., 2014; Grünewald and Lehning, 2011). These steep gradients are likely the result of physical processes beyond just orographic precipitation, including storm systems that deliver snow at upper elevations and rain at lower elevations (common at both Wolverine and Gulkana) and mid-winter ablation at lower elevations (at Wolverine). These processes have also been shown to steepen observed SWE gradients relative to orographic precipitation gradients in a mid-latitude seasonal snow watershed (Anderson et al., 2014). Unfortunately, given that we solely sampled snow distribution at the end of the accumulation season, the relative magnitude of each of these secondary processes is not constrained.

Wolverine and Gulkana glaciers exhibited opposing SWE gradients at their highest elevations, with Wolverine showing a sharp non-linear increase in SWE, while Gulkana showed a gradual decrease. This non-linear increase was also noted at two maritime glaciers (Scott and Valdez) in 2013 (McGrath et al., 2015), and perhaps reflects an abundance of split precipitation phase storms in these warm coastal regions. The cause of the observed reverse gradient at Gulkana may be the result of wind scouring at the highest and most exposed sections of the glacier, or in part, a result of where we were able to safely sample the glacier. For instance, in 2013, when we were able to access the highest basin on the glacier, the SWE elevation gradient remained positive (Fig. 4). Reductions in accumulated SWE at the highest elevations have also been observed at Lemon Creek Glacier in southeast Alaska and Findel Glacier in Switzerland (Machguth et al., 2006), presumably related to wind scouring at these exposed elevations.

5.1.2. Wind redistribution

Both statistical extrapolation approaches found terrain parameters *Sb* and curvature, proxies for wind redistribution, to have the largest beta coefficients after elevation (Fig. 6, S9). The spatial pattern of SWE estimated by each model clearly reflects the dominant influence of wind redistribution and elevation (Fig. 8), as areas of drift and scour are apparent, especially at higher elevations. However, these terms do not fully capture the redistribution process, as the model residuals (Fig. S1, S2) show sequential positive and negative residuals associated with drift/scour zones. There are a number of reasons why this might occur, including variable wind directions transporting snow (this is likely a more significant issue at Gulkana, which experiences greater wind direction variability (Fig. S6)), complex wind fields that are not well represented by a singular wind direction (Dadic et al., 2010), changing surface topography (the glacier surface is dynamic over a range of temporal scales, changing through both surface mass balance processes and ice dynamics), and widely varying wind velocities. This is particularly relevant at Wolverine, where wind speeds regularly gust over 30 m s^{-1} during winter storms, speeds that result in variable length scales of redistribution that would not be captured by a fixed length scale of redistribution. All of these factors influence the redistribution of snow and limit the predictive ability of relatively simple proxies. Significant effort has gone into developing physically-based snow-distribution models (e.g., Alpine3D and SnowModel), however, high-resolution meteorological forcing data requirements generally limit the application of these models in glacierized basins. Where such observations do exist, previous studies have illuminated how the final distribution of snow is strongly correlated to the complex wind field, including vertical (surface normal) winds (Dadic et al., 2010).

5.1.3. Differences with non-glaciated terrain

Although our GPR surveys did not regularly include non-glaciated regions of these basins, a few key differences are worth noting. First, the length scales of variability on and off the glacier were distinctly different, with shorter scales and greater absolute variability (snow-free to >5 m in less than 10 m distance) off-glacier (Fig. S10). This point has been clearly shown using airborne LiDAR in a glaciated catchment in the Austrian Alps (Helfricht et al., 2014). The reduced variability on the glacier is largely due to surface mass balance and ice flow processes that act to smooth the surface, leading to a more spatially consistent surface topography, and therefore a more spatially consistent SWE pattern. For this reason, establishing a SWE elevation gradient on a glacier is likely much less prone to terrain-induced outliers compared to off-glacier sites, although the relationship of this gradient to off-glacier gradients is generally unknown.

5.2. Spatial differences between statistical models

The two statistical extrapolation approaches yielded comparable large-scale spatial distributions and glacier-wide averages, although there were some notable spatial differences (Fig. 10). The systematic positive bias of the MVR approach over the regression tree at Wolverine was due to three sectors of the glacier with both complex terrain (i.e., icefalls) and large data gaps (typically locations that are not safe to access on ground surveys). The difference in predicted SWE in these locations is likely due to how the two statistical extrapolation approaches handle unsampled terrain parameter space. The MVR extrapolates based on global linear trends, while the regression tree assigns SWE from terrain that most closely resembles the under-sampled location. Anecdotally, it appears that the MVR may overestimate SWE in some of these locations, which is most evident in Wolverine's lower icefall, where bare ice is frequently exposed at the end of the accumulation season (Fig. S11) in locations where the MVR predicted substantial SWE. Likewise, the regression tree models could be underestimating SWE in these regions, but in the absence of direct observations the errors are inherently unknown. The regression tree model captures more short length scale variability while the MVR model clarifies the larger trends. Consequently, smaller drifts and scours are captured well by the regression tree model in areas where the terrain parameter space is well surveyed, but the results become progressively less plausible as the terrain becomes distinctly different from the sampled terrain parameter space. In contrast, the MVR model appears to give more plausible results at larger spatial scales. This suggests that there is some theoretical threshold where the regression tree is more appropriate if the terrain parameter space is sampled sufficiently, but that for many glacier surveys the MVR model would be more appropriate.

Deleted: more

5.3. Winter mass balance comparisons

On average, all methods for estimating B_w were within ± 11 % of the six-method mean, (Fig. 13). The agreement (as measured by the average percent difference from the mean) between estimates was slightly better at Gulkana than Wolverine, likely reflecting the overall lower spatial variability at Gulkana and the greater percentage of the glacier area where b_w correlates well with the glacier-wide average (Fig. 11 e, f). At both glaciers, B_w solutions based solely on elevation showed excellent agreement to the six-method mean, suggesting that this simple approach is a viable means for measuring B_w on these glaciers. The biggest differences occurred between the GPR-forced MVR model and the glaciological site-index method, which we've shown is attributed to the upper stake (with the greatest weight) underestimating the median SWE for that index zone (Fig. 14). The upper stake location was established in 1966 at an elevation below the median elevation of that index zone, which given the strong elevation control on SWE, is a likely reason for the observed difference. At Gulkana, the relationship between the upper index site and the GPR-forced MVR model is more variable in large part due to observed differences in the accumulation between the main branch (containing the index site) and the west branch of the glacier (containing additional stakes added in 2009). Such basin-scale differences are likely present on many glaciers with complex geometry, and thus illustrate potential uncertainties of using a small network of stakes to monitor the mass balance of these glaciers. In the context of the MVR model, this manifests as a change in sign in the eastness coefficient (which separates the branches in parameter space; Fig. S4). Notably, in the two years where the site-index estimate was most negatively biased at Gulkana (2015 and 2016), the glaciological profile method, relying on the more extensive stake network (which includes stakes in the west branch of the glacier), yielded B_w estimates within a few percent of the GPR-derived MVR estimate.

These GPR-derived B_w results have important implications for the cumulative glaciological (stake-derived) mass balance time-series (currently only based on the site-index method), which is calibrated with geodetic observations (details on the site-index method and geodetic calibrations can be found in Van Beusekom et al., 2010 and O'Neel et al., 2014). It is important to remember that the previous comparisons (e.g., Fig. 13) were based on glaciological B_w values that have not had a geodetic calibration applied. At

Wolverine, the cumulative annual glaciological mass balance solutions are positively biased compared to the geodetic mass balance solutions over decadal timescales, requiring a negative calibration (−0.43 m w.e. a^{-1}; O'Neel et al., 2014) to be applied to the glaciological solutions. The source of this disagreement is some combination of the stake-derived winter and summer balances being too positive relative to the geodetic solution. On average, the GPR-derived B_w results were ~0.4 m w.e. more positive than the site-index B_w results at Wolverine, which would further increase the glaciological-geodetic solution difference and suggest that the stake-derived glaciological solutions are underestimating ablation (B_s) by ~0.8 m w.e. a^{-1}. Preliminary observations at Wolverine using ablation wires show that some sectors of the glacier experience very high ablation rates that are not captured by the stake network (e.g., crevassed zones through enhanced shortwave solar radiation gain (e.g., Pfeffer and Bretherton, 1987; Cathles et al., 2011; Colgan et al., 2016), and/or increased turbulent heat fluxes due to enhanced surface roughness), and/or ice margins (through enhanced longwave radiation from nearby snow-free land cover)). However, these results are not universal, as the assimilation of distributed GPR observations at Findelgletchter significantly improved the comparison between geodetic and modeled mass balance estimates (Sold et al., 2016), suggesting multiple drivers of glaciologic-geodetic mismatch for long-term mass balance programs.

5.3.1. Implications for stake placement

Understanding the spatiotemporal distribution of SWE is useful for informing stake placements and also for quantifying the uncertainty that interannual spatial variations in SWE introduce to historic estimates of glacier-wide mass balance, particularly when long-term mass balance programs rely on limited numbers of point observations (e.g., USGS and National Park Service glacier monitoring programs; O'Neel et al., 2014; Burrows, 2014). Our winter balance results illustrate that stakes placed at the same elevation are not directly comparable, and hence are not necessarily interchangeable in the context of a multi-year mass balance record. Most locations on the glacier exhibit bias from the average mass balance at that elevation and our results suggest interannual consistency in this bias over sub-decadal time scales. As a result, constructing a balance profile using a small number of inconsistently located stakes is likely to introduce large relative errors from one year to the next.

Considering this finding, the placement of stakes to measure snow accumulation is dependent on whether a single glacier-wide winter mass balance value (B_w) or a spatially distributed SWE field is desired as a final product. For the former, a small number of stakes can be distributed over the glacier hypsometry in areas where interannual variability is low. Alternatively, if a distributed field is desired, a large number of stakes can be widely distributed across the glacier, including areas where the interannual variability is higher. In both cases it is important to have consistent locations from year to year, although as the number of stakes increases significantly, this becomes less critical.

We assess the uncertainty that interannual variability in the spatial distribution of SWE introduces to the historic index-method (March and Trabant, 1996) mass balance solutions by first calculating the uncertainty, σ, contributed by each stake as:

$$\sigma_{stake} = \sigma_{model\ residuals} + (1 - R^2) \cdot u \ , \qquad (3)$$

where $\sigma_{model\ residuals}$ is the standard deviation of MVR model residuals over all five years within ± 30 meters of the index site, u is the mean b_w within ± 30 meters of the index site, and R^2 is the coefficient of determination between b_w and B_w over the five-year period (Fig. 11). The first term on the right hand side of Eq. 3 accounts for both the spatial and temporal variability in the observed b_w as compared to the model, and the second term accounts for the variability of the model as compared to B_w. The glacier-wide uncertainty from interannual variability is then:

$$Glacier\ \sigma = \sqrt{\sum_{all\ stakes}(\sigma_{stake} \cdot w_{stake})^2}, \qquad (4)$$

where w_{stake} is the weight function from the site-index method (which depends on stake location and glacier hypsometry). By this assessment, interannual variability in the spatial distribution of SWE at stake locations introduced minor uncertainty, on the order of 0.11 m w.e. at both glaciers (4 % and 10 % of B_w at Wolverine and Gulkana, respectively). This suggests that the original stake network design at the benchmark glaciers does remarkably well at capturing the interannual variability in glacier-wide winter balance. The greatest interannual variability at each glacier is found at the lowest stake sites, but because b_w and the stake weights are both quite low at these sites, they contribute only modestly to the overall uncertainty. Instead, the middle and upper elevation stakes contribute the greatest amount to the glacier-wide uncertainty.

Deleted: slightly

6. Conclusions

We collected spatially extensive GPR observations at two glaciers in Alaska for five consecutive winters to quantify the spatiotemporal distribution of SWE. We found good agreement of glacier-average winter balances, B_w, among the four different approaches used to extrapolate GPR point measurements of SWE across the glacier hypsometry. Extrapolations relying only on elevation (i.e., a simple balance profile) produced B_w estimates similar to the more complicated statistical models, suggesting that this is an appropriate method for quantifying glacier-wide winter balances at these glaciers. The more complicated approaches, which allow SWE to vary across a range of terrain-parameters based on DEMs, show a high degree of temporal stability in the pattern of accumulation at both glaciers, as ~85 % of the area on both glaciers experienced less than 25 % normalized absolute variability over the five-year interval. Elevation and the parameters related to wind redistribution had the most explanatory power, and were temporally consistent at each site. The choice between MVR and regression tree models should depend on both the range in terrain parameter space that exists on the glacier, along with how well that space is surveyed.

Deleted: -

In total, six different methods (four based on GPR measurements and two based on stake measurements) for estimating the glacier-wide average agreed within ± 11 %. The site-index glaciological B_w estimates were negatively biased compared to all other estimates, particularly when the upper-elevation stake significantly underestimated SWE in that index zone. In contrast, the profile glaciological approach, using a more extensive stake network, showed better agreement with the other approaches, highlighting the benefits of using a more extensive stake network.

We found the spatial patterns of snow accumulation to be temporally stable on these glaciers, which is consistent with a growing body of literature documenting similar consistency in a wide variety of environments. The long-term stake locations experienced low interannual variability in normalized SWE, meaning that stake measurements tracked the interannual variability in SWE, rather than interannual variability in spatial patterns. The uncertainty associated with interannual spatial variability is only 4–10 % of the glacier-wide B_w at each glacier. Thus, our findings support the concept that sparse stake networks can be effectively used to measure interannual variability in winter balance on glaciers.

Data Availability. The GPR and associated observational data used in this study can be accessed on the USGS Glaciers and Climate Project website (https://doi.org/10.5066/F7M043G7). The Benchmark Glacier mass balance input and output can be accessed at: https://doi.org/10.5066/F7HD7SRF (O'Neel et al., 2018). The Gulkana DEM is available from the ArcticDEM project website (https://www.pgc.umn.edu/data/arcticdem/) and the Wolverine DEM is available at ftp://bering.gps.alaska.edu/pub/chris/wolverine/. A generalized version of the SWE extrapolation code is available at: https://github.com/danielmcgrathCSU/Snow-Distribution.

Author Contributions. SO, DM, LS, and HPM designed the study. DM performed the analyses and wrote the manuscript. LS contributed to the design and implementation of the analyses, and CM, SC, and EHB contributed specific components of the analyses. All authors provided feedback and edited the manuscript.

Competing Interests. The authors declare that they have no conflict of interest.

Acknowledgments. This work was funded by the U.S. Geological Survey Land Change Science Program, USGS Alaska Climate Adaptation Science Center, and DOI/USGS award G17AC00438 to DM. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. We acknowledge the Polar Geospatial Center (NSF-OPP awards 1043681, 1559691, and 1542736) for the Gulkana DEM. We thank Caitlyn Florentine, Jeremy Littell, Mauri

Pelto, and an anonymous reviewer for their thoughtful feedback that improved the manuscript.

References

Akaike, H.: A new look at the statistical model identification, IEEE Trans. Autom. Control, AC-19(6), 1974.

Anderson, B. T., McNamara, J. P., Marshal, H. P., and Flores, A. N.: Insights into the physical processes controlling correlations between snow distribution and terrain properties, Water Res. Res., 50(6), 4545–4563, doi:10.1002/2013WR013714, 2014.

Bauder, A. (ed): The Swis Glaciers, 2013/14 and 2014/15, Glaciological Report (Glacier) No. 135/136, doi:10.18752/glrep_135-136, 2017.

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

Balk, B. and Elder, K.: Combining binary regression tree and geostatistical methods to estimate snow distribution in a mountain watershed, Water Res. Res., 36(1), 13–26, 2000.

Beamer, J.P., Hill, D.F., McGrath, D., Arendet, A., and Kienholz, C.: Hydrologic impacts of changes in climate and glacier extent in the Gulf of Alaska watershed, Water. Res. Res., 53, doi:10.1002/2016WR020033.

Burrows, R.: Annual report on vital signs monitoring of glaciers in the Central Alaska Network 2011-2013, Natural Resource Technical Report NPS/CAKN/NRTR—2014/905, National Park Service, Fort Collins, Colorado, 2014.

Breiman, L.: Bagging predictors, Mach. Learn., 24, 123–140, https://doi.org/10.1023/A:1018054314350, 1996.

Breiman, L.: Random forests, Mach. Learn., 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J.: Classification and Regression Trees, Chapman and Hall, New York, 368 pp., 1984.

Cathles, L. C., Abbot, S. D., Bassis, J. N., and MacAyeal, D.R.: Modeling surface-roughness/solar-ablation feedback: application to small-scale surface channels and crevasses of the Greenland ice sheet, Ann. Glaciol., 52(59), 99–108, 2011.

Cogley, J. G., Hock, R., Rasmussen, L. A., Arendt, A. A., Bauder, A., Braithwaite, R. J., Jansson, P., Kaser, G., Möller, M., Nicholson, L. and Zemp, M.: Glossary of Glacier Mass Balance and Related Terms, IHP-VII Technical Documents in Hydrology No. 86, IACS Contribution No. 2, UNESCO-IHP, Paris, 2011.

Colgan, W., Rajaram, H., Abdalati, W., McCutchan, C., Mottram, R., Moussavi, M. S., and Grigsby, S.: Glacier crevasses: Observations, models, and mass balance implications, Rev. Geophys., 54, doi:10.1002/ 2015RG000504, 2016.

Dadic, R., Mott, R., Lehning, M., and Burlando, P.: Wind influence on snow depth distribution and accumulation over glaciers, J. Geophys. Res., 115, F01012, doi:10.1029/2009JF001261, 2010.

Deems, J. S., Fassnacht, S. R., and Elder, K. J.: Interannual consistency in fractal snow depth patterns at two Colorado mountain sites, J. Hydromet., 9, 977–988, doi:10.1175/2008JHM901.1, 2008.

Elder, K., Michaelsen, J., and Dozier, J.: Small basin modeling of snow water equivalence using binary regression tree methods, IAHS Publ. 228, 129–139, 1995.

Erickson, T. A., Williams, M.W., and Winstral, A.: Persistence of topographic controls on the spatial distribution of snow in rugged mountain terrain, Colorado, United States, Water Res. Res., 41, W04014, doi:10.129/2003WR002973, 2005.

Escher-Vetter, H., Kuhn, M., and Weber, M.: Four decades of winter mass balance of Vernagtferner and Hintereisferner, Austria: methodology and results, Ann. Glaciol., 50(50), 2009.

Fausto, R.S., and 11 others.: A Snow Density Dataset for Improving Surface Boundary Conditions in Greenland Ice Sheet Firn Modeling, Front. Earth Sci., 6(51), doi:10.3389/feart.2018.00051, 2018.

Fitzpatrick, N., Radić, V., and Menounos, B.: Surface energy balance closure and turbulent flux parameterization on a mid-latitude mountain glacier, Purcell Mountains, Canada, Front. Earth Sci., 5(67)), doi:10.3389/feart.2017.00067, 2017.

Grünewald, T., and Lehning, M.: Altitudinal dependency of snow amounts in two alpine catchments: Can catchment-wide snow amounts be estimated via single snow or precipitation stations?, Ann. Glaciol., 52(58), 153–158, 2011.

Helfricht, K., Schöber, J., Schneider, K., Sailer, R., and Kuhn, M.: Interannual persistence of the seasonal snow cover in a glacierized catchment, J. Glaciol., 60(223), doi:10.3189/2014JoG13J197, 2014.

Hock, R.: Glacier melt: a review of processes and their modeling, Prog. Phys. Geog., 29, 362–391, doi:10.1191/0309133305pp453ra, 2005.

Hock, R., Hutchings, J. K., and Lehning, M.: Grand challenges in cryospheric sciences: toward better predictability of glaciers, snow and sea ice, Front. Earth Sci., 5(64), doi:10.3389/feart.2017.00064, 2017.

Hodgkins, R., Cooper, R., Wadham, J., and Tranter, M.: Interannual variability in the spatial distribution of winter accumulation at a high-Arctic glacier (Finsterwalderbreen, Svalbard), and its relationship with topography, Ann. Glaciol., 42, 243–248, 2005.

Huss, M. and Hock, R.: A new model for global glacier change and sea-level rise, Front. Earth Sci., 3, doi:10.3389/feart.2015.00054, 2015.
Kaser, G., Großhauser, M., and Marzeion, B.: Contribution potential of glaciers to water availability in different climate regimes, Proc. Natl. Acad. Sci., 107, 20,223–20,227, doi:10.1073/pnas.1008162107, 2010.

Klos, P. Z., Link, T. E., and Abatzoglou, J. T.: Extent of the rain-snow transition zone in the western U.S. under historic and projected climate, Geophys. Res. Lett., 41, 4560–4568, doi: 10.1002/2014GL060500, 2014.

Deleted: Knowles, N., Dettinger, M. D., and Cayan, D. R.: Trends in snowfall versus rainfall in the Western United States, J. Climate, 19, 4545–4559, 2006.

Kovacs, A., Gow, A. J., and Morey, R. M.: The in-situ dielectric constant of polar firn revisited, Cold Reg. Sci. Tech., 23, 245–256, 1995.

Kuhn, M.: The mass balance of very small glaciers, *Z. Gletscherkd. Glazialgeol.,* 31(1–2), 171–179, 1995.

Lehning, M., Grünewald, T., and Schirmer, M.: Mountain snow distribution governed by altitudinal gradient and terrain roughness, Geophys. Res. Lett., 38, L19504, doi:10.1029/2011GL048927, 2011.

Li, L. and Pomeroy, J. W.: Estimates of threshold wind speeds for snow transport using meteorological data, J. Applied Met., 36, 205-213, 1997.

Liston, G. E., and Elder, K.: A distributed snow-evolution modeling system (SnowModel), J. Hydromet., 7, 1259-1276, 2006.

Littel, J. S., McAfee, S. A., and Hayward, G. D.: Alaska snowpack response to climate change: statewide snowfall equivalent and snowpack water scenarios, Water, 10 (5), doi: 10.3390/w10050668, 2018.

Kjøllmoen, B. (ed), Andreassen L.M., Elvehøy, H., Jackson, M., and Melvold, J: Glaciological investigations in Norway 2016, NVE Rapport 76 2017, 108 pp, 2017.

Marks, D., Domingo, J., Susong, D., Link, T., and Garen, D.: A spatially distributed energy balance snowmelt model for application in mountain basins, *Hydrol. Processes,* 13, 1935–1959, 1999.

Machguth, H., Eisen, O., Paul, F., and Hoelzle, M.: Strong spatial variability of snow accumulation observed with helicopter-borne GPR on two adjacent Alpine glaciers, Geophys. Res. Lett., 33, L13503, doi:10.1029/2006GL026576, 2006.

March, R. S., and Trabant, D. C.: Mass balance, meteorological, ice motion, surface altitude, and runoff data at Gulkana Glacier, Alaska, 1992 balance year, Water-Resources Investigations Report, 95-4277, 1996.

McAfee, S., Walsh, J., and Rupp, T. S.: Statistically downscaled projections of snow/rain partitioning for Alaska, Hydrol. Process., 28(12), 3930–3946, doi:10.1002/hyp.9934, 2013.

McGrath, D., Sass, L., O'Neel, S., Arendt, A., Wolken, G., Gusmeroli, A., Kienholz, C., and McNeil, C.: End-of-winter snow depth variability on glaciers in Alaska, J. Geophys. Res. Earth Surf., 120, 1530–1550, doi:10.1002/2015JF003539, 2015.

McGrath, D., Sass, L., O'Neel, S., Arendt, A. and Kienholz, C.: Hypsometric control on glacier mass balance sensitivity in Alaska and northwest Canada, Earth's Future, *5*, 324–336, doi:10.1002/2016EF000479, 2017.

Meromy, L., Molotch, N. P., Link, T. E., Fassnacht, S. R., and Rice, R.: Subgrid variability of snow water equivalent at operational snow stations in the western USA, Hydro. Proc., 27, 2383-2400, doi:10.1002/hyp.9355, 2013.

Molotch, N. P., Colee, M. T., Bales, R. C. and Dozier, J.: Estimating the spatial distribution of snow water equivalent in an alpine basin using binary regression tree models: the impact of digital elevation data and independent variable selection, Hydrol. Proc., 19, 1459–14-79, doi:10.1002/hyp.5586, 2005.

Nolan, M., Larsen, C., and Sturm, M.: Mapping snow depth from manned aircraft on landscape scales at centimeter resolution using structure-from-motion photogrammetry, The Cryosphere, 9, 1445-1463, doi:10.5194/tc-9-1445-2015, 2015.

Noh, M. J. and Howat, I. M.: Automated stereo-photogrammetric DEM generation at high latitudes: Surface Extraction with TIN-based Search-space Minimization (SETSM) validation and demonstration over glaciated regions, GIScience & Remote Sensing, 52(2), 198-217, doi:10.1080/15481603.2015.1008621, 2015.

O'Neel, S., Hood, E., Arendt, A., and Sass, L.: Assessing streamflow sensitivity to variations in glacier mass balance, Climatic Change, 123(2), 329–341, doi:10.1007/s10584-013-1042-7, 2014.

O'Neel, S., Fagre, D. B., Baker, E. H., Sass, L. C., McNeil, C. J., Peitzsch, E. H., McGrath, D. and Florentine, C. E.: Glacier-Wide Mass Balance and Input Data: USGS Benchmark Glaciers, 1966-2016 (ver. 2.1, May 2018), U.S. Geological Survey data release, https://doi.org/10.5066/F7HD7SRF, 2018.

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

Pelto, M.: Utility of late summer transient snowline migration rate on Taku Glacier, Alaska, The Cryosphere, 5, 1127–1133, doi:10.5194/tc-5-1127-2011, 2011.

Pelto, M., Kavanaugh, J., and McNeil, C.: Juneau Icefield Mass Balance Program 1946–2011, Earth Syst. Sci. Data, 5, 319-330, https://doi.org/10.5194/essd-5-319-2013, 2013.

Pérez-Ruiz, M., Carballido, J., and Agüera, J.: Assessing GNSS correction signals for assisted guidance systems in agricultural vehicles, Precision Agric., 12, 639–652, doi:10.1007/s11119-010-9211-4, 2011.

Pfeffer, W. T., and Bretherton, C.: The effect of crevasses on the solar heating of a glacier surface, IAHS Publ., 170, 191–205, 1987.

Pfeffer, W. T., et al.: The Randolph Glacier Inventory: A globally complete inventory of glaciers, J. Glaciol., 60(221), 537–552, doi:10.3189/2014JoG13J176, 2014.

Schirmer, M., Wirz, V., Clifton, A., and Lehning, M.: Persistence in intra-annual snow depth distribution: 1. Measurements and topographic control, Water Resour. Res., 47, W09516, doi:10.1029/2010WR009426, 2011.

Sold, L., Huss, M., Hoelzle, M., Andereggen, H., Joerg, P., and Zemp, M.: Methodological approaches to infer end-of-winter snow distribution on alpine glaciers, J. Glaciol., 59(218), 1047–1059, doi:10.3189/2013JoG13J015, 2013.

Sold, L., Huss, M., Machguth, H., Joerg, P. C., Vieli, G. L., Linsbauer, A., Salzmann, N., Zemp, M. and Hoelzle, M.: Mass balance re-analysis of Findelengletscher, Switzerland;

Benefits of extensive snow accumulation measurements, Front. Earth Sci., 4(18), doi:10.3389/feart.2016.00018, 2016.

Sturm, M. and Wagner, A. M.: Using repeated patterns in snow distribution modeling: An Arctic example, Water Res. Res., 46 (12), doi:10.1029.2010WR009434, 2010.

Van Beusekom, A. E., O'Neel, S., March, R. S., Sass, L., and Cox, L. H.: Re-analysis of Alaskan Benchmark Glacier mass balance data using the index method, U.S. Geological Survey Scientific Investigations Report 2010–5247, 16 p., 2010.

Vincent, C., Fischer, A., Mayer, C., Bauder, A., Galos, S.P., Funk, M., Thibert, E., Six, D., Braun, L., and Huss, M.: Common climatic signal from glaciers in the European Alps over the last 50 years, Geophys. Res. Lett., 44, 1376–1383, doi:10.1002/2016GL072094, 2017.

Winstral, A., Elder, K., and Davis, R. E.: Spatial snow modeling of wind-redistributed snow using terrain-based parameters, J. Hydrometeo., 3, 524–538, 2002.

Winstral, A., Marks, D. and Gurney, R.: Simulating wind-affected snow accumulations at catchment to basin scales, Adv. Water Res., 55, 64–79, doi:10.1016/j.advwatres.2012.08.011, 2013.

Winstral, A. and Marks, D.: Long-term snow distribution observations in a mountain catchment: Assessing variability, time stability, and the representativeness of an index site, Water Res. Res., 50, 293–305, doi:1002/2012WR013038, 2014.

Woo, M.-K., and Marsh, P.: Analysis of error in the determination of snow storage for small high Arctic basins, J. Appl. Meteorol., 17, 1537–1541, 1978.

Yelf, R. and Yelf, D.: Where is true time zero?, Electro. Phenom., 7(1), 158–163, 2006.

Figure 1. Map of southern Alaska with study glaciers marked by red outline. All glaciers in the region are shown in white (Pfeffer et al., 2014).

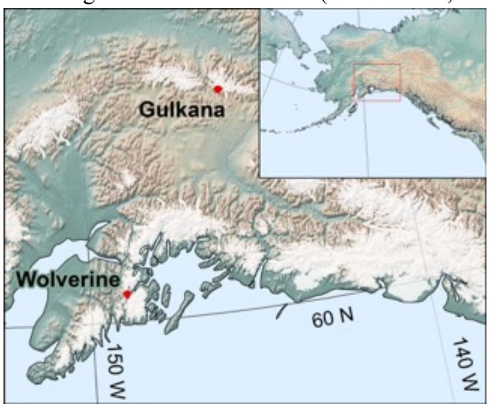

Figure 2. Boxplots of glacier-wide winter balance for Gulkana and Wolverine glaciers between 1966 and 2017. Years corresponding to GPR surveys are shown with colored markers. These values have not been adjusted by the geodetic calibration (see O'Neel et al., 2014).

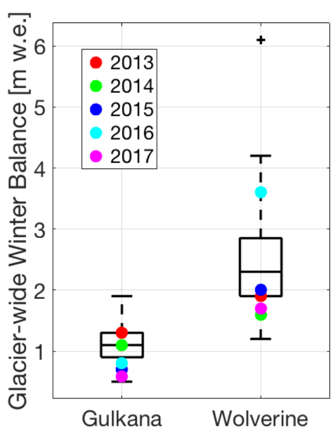

Figure 3. GPR surveys from 2015 at Gulkana (a) and Wolverine (c) glaciers and MVR model residuals (b, d).

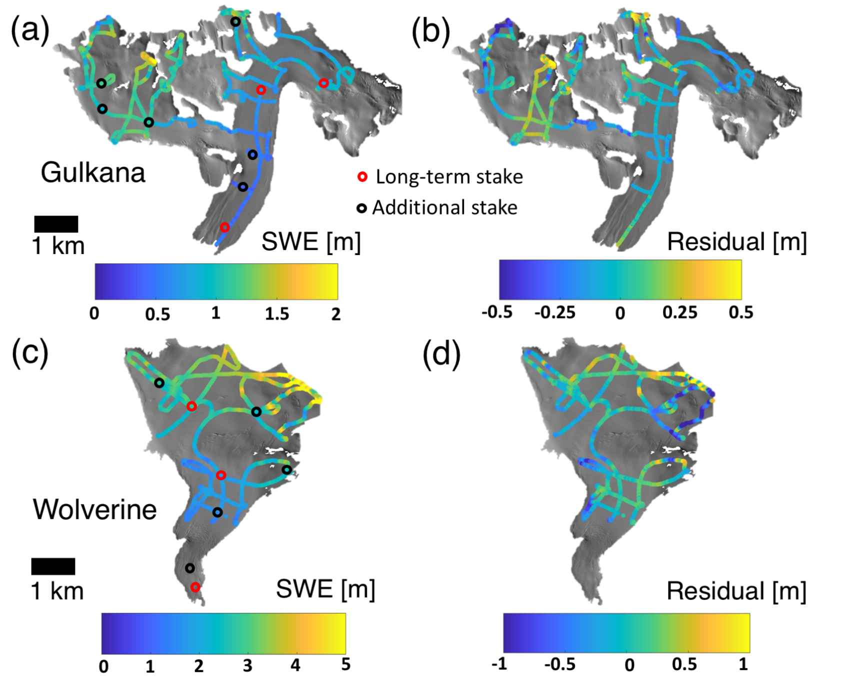

Figure 4. SWE from GPR surveys as a function of elevation, along with least squares regression slope and coefficient of determination for each year of the study period. Wolverine is plotted in blue, Gulkana in red.

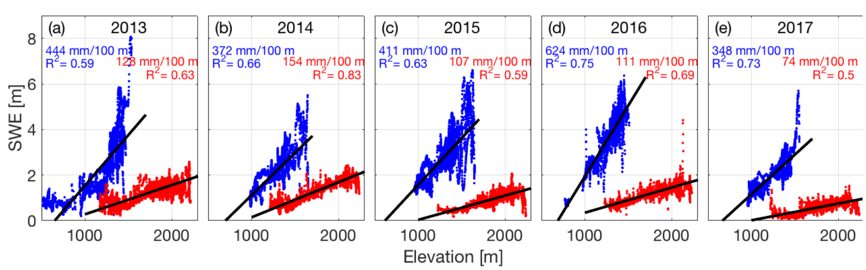

Figure 5. Median and standard deviation (error bars) of coefficient of determination (from 100 model runs) for both extrapolation approaches (circles are MVR, triangles are regression tree) developed on training datasets and applied to test datasets. Symbols and error bars are offset from year for clarity.

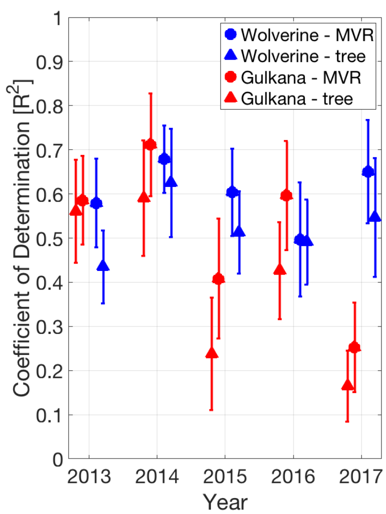

Figure 6. Terrain parameter beta coefficients for (a) Gulkana and (b) Wolverine for multivariable linear regression for each year of the study interval.

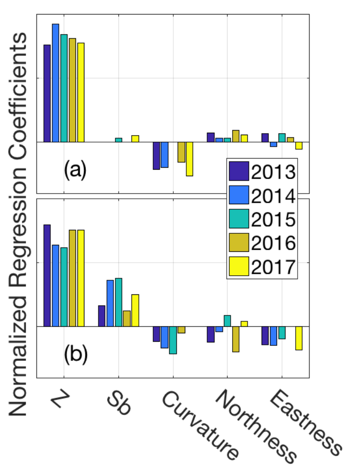

Figure 7. Spatial variability in snow accumulation across the glacier quantified by the coefficient of variation (standard deviation/mean) for each glacier across the five-year interval based on MVR model output.

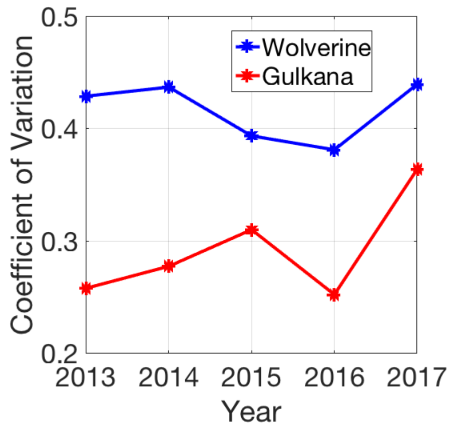

Figure 8. Five-year mean of normalized distributed SWE for Gulkana (a,b) and Wolverine (c,d) for multivariable regression (a,c) and regression tree (b,d).

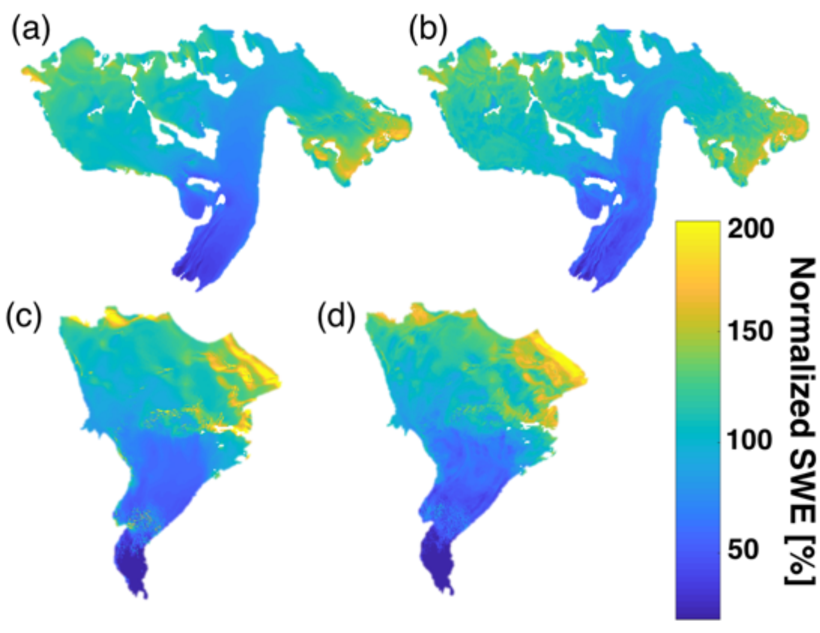

Figure 9. Comparing statistical models for GPR-derived glacier-wide winter balances for both Wolverine (blue) and Gulkana (red) glaciers. For each year and each glacier, two boxplots are shown. The first shows multivariable regression model (MVR) output and the second shows regression tree output (tree). The B_w estimate from the glaciological profile method is shown for each year and glacier as the filled circle.

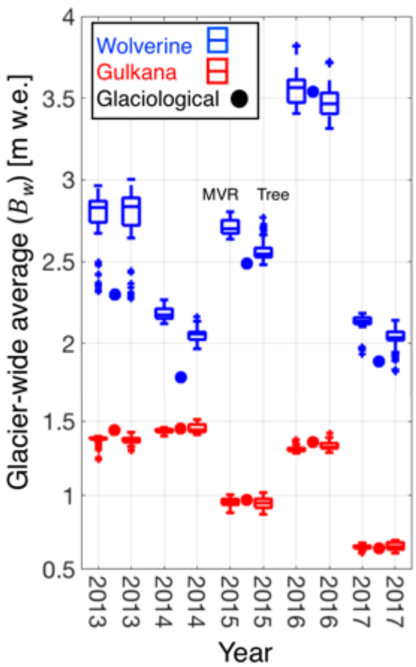

Figure 10. SWE differences between statistical models for Gulkana (a) and Wolverine (b) calculated by differencing the regression tree five-year mean SWE from the multivariable regression (MVR) five-year mean SWE. Yellow colors indicate regions where MVR yields more SWE than decision tree and blue colors indicate the opposite. Note different magnitude colorbar scales. c) Summed SWE difference between methods in bins of 0.05 normalized elevation values.

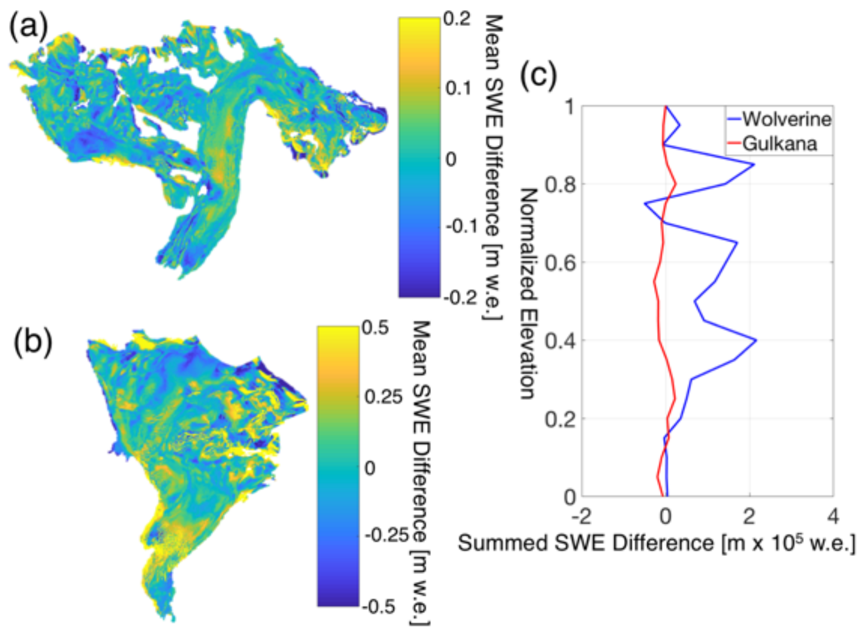

Figure 11. Interannual variability of the SWE accumulation field from 2013–2017, quantified via normalized range (a-d) and R^2 (e-h) approach for median distributed fields from the multivariable regression (left column) and regression tree (right column) statistical models.

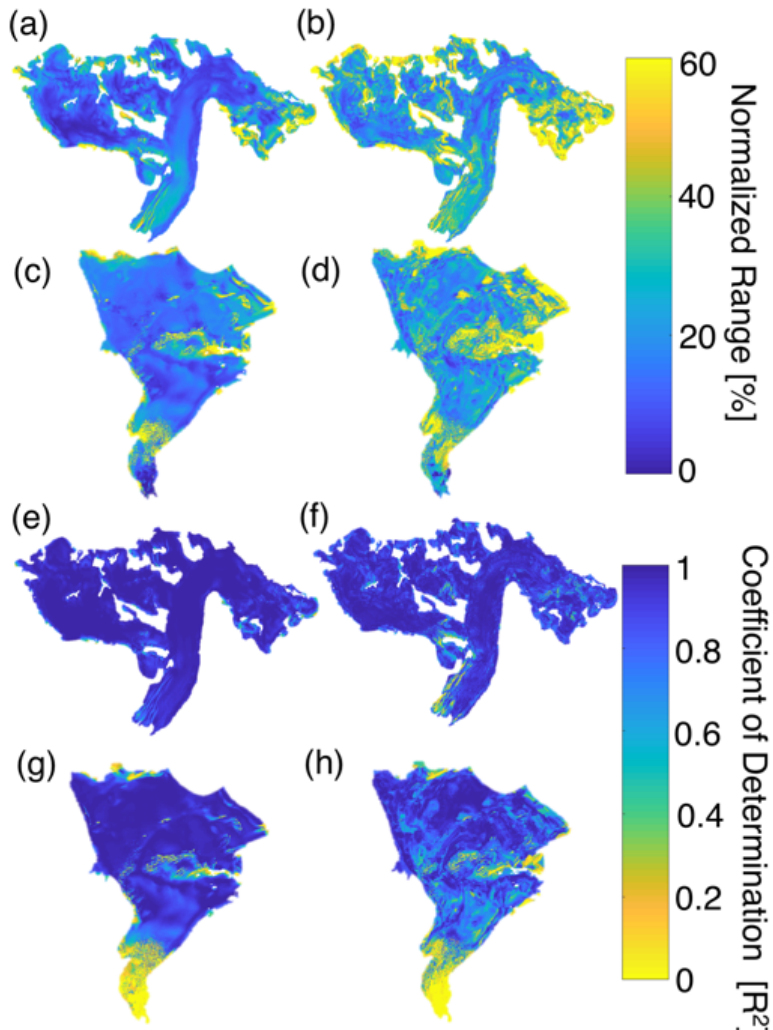

Figure 12. Interannual variability of the SWE accumulation pattern as a function of cumulative glacier area, shown as (a) normalized range and (b) and R^2. Solid lines are for multivariable regression (MVR) and dashed lines are regression tree.

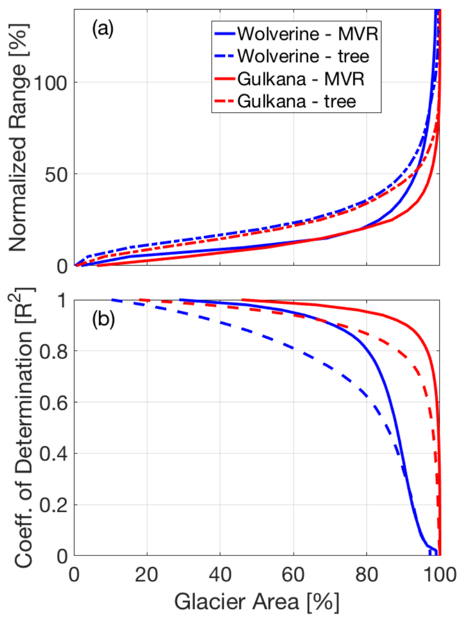

Figure 13. Percent deviation for each estimate from the six-method mean of B_w. Individual years for Gulkana Glacier are shown in panels a-e with the five-year mean shown in f. Individual years for Wolverine Glacier are shown in panels g-k, with the five-year mean shown in l.

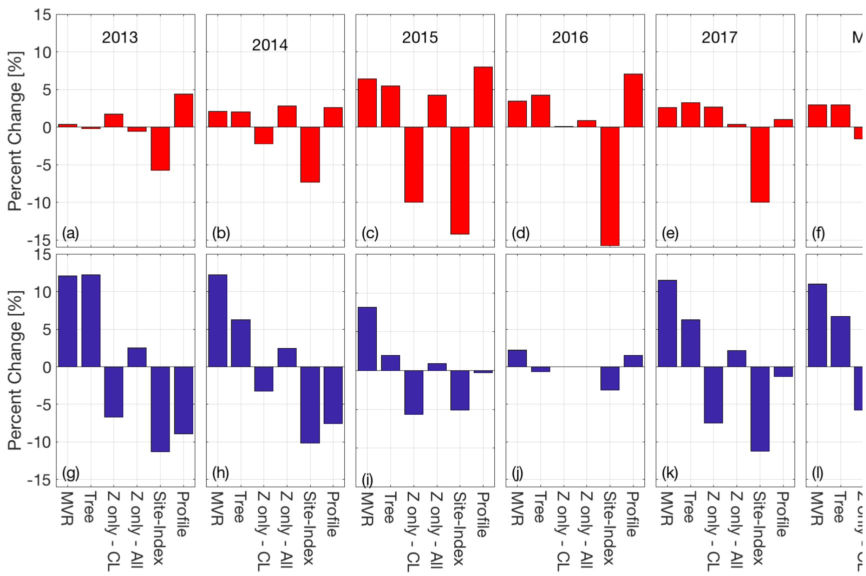

Figure 14. Spatial variability in snow accumulation for individual years (2013-2017) by elevation (lower, middle, upper) compared to stake measurements. Box plot of all distributed SWE values (from multivariable regression) for each index zone of the glacier for Gulkana (a-e) and Wolverine (f-j) for 2013-2017. The filled circles are the respective stake observation for that index zone. SWE is expressed as a percentage of the glacier-wide average, B_w, for that year and glacier.

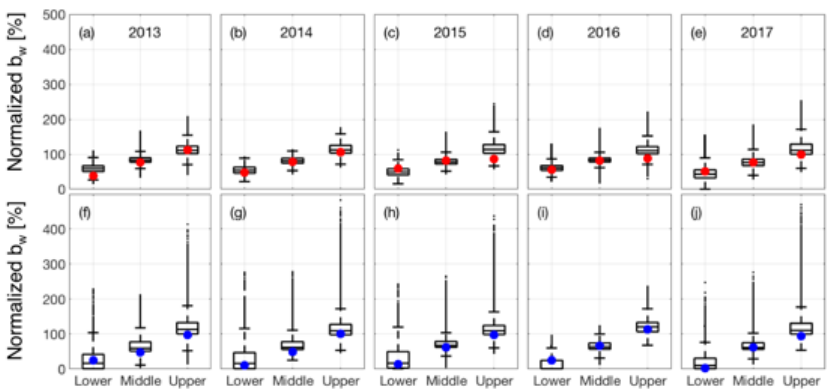

Figure 15. Interannual variability in the spatial pattern of snow accumulation at long-term mass balance stake locations for Wolverine and Gulkana glaciers using a) normalized b_w range and b) coefficient of determination (from Figure 11; MVR model).