Optimal snow-survey design for the estimation of winter balance on alpine glaciers

Alexandra PULWICKI,[a] Gwenn E. FLOWERS,[b]

[a] *Department of Earth Sciences, Simon Fraser University, 8888 University Drive, Burnaby, BC, V5A 1S6, Canada* *<apulwick@sfu.ca>*

[b] *Department of Earth Sciences, Simon Fraser University, 8888 University Drive, Burnaby, BC, V5A 1S6, Canada* *<gflowers@sfu.ca>*

ABSTRACT. Efficient collection of snow depth and density data is critical to a successful snow measurement campaign and to accurately estimate glacier winter balance. Extensive, high resolution and accurate snow accumulation measurements on glaciers are almost impossible to achieve so surveys need to optimize the extent and spacing of snow measurements to obtain reliable estimates of winter balance. To address this need, we estimate winter balance and root mean squared error (RMSE) using synthetic and real data from subsets of extensive surveys on three glaciers in the St. Elias Mountains, Yukon. We generate six different survey designs, which encompass possible snow sampling patterns and various numbers of measurement locations. We then use linear regression with topographic parameters to interpolate measurements. Analysis of both synthetic and real data indicates that an 'midline & transect' and'hourglass' sampling patterns result in the most efficient and accurate estimate of winter balance, while the midline pattern results in poor estimates of winter balance for all glaciers. RMSE decreases with increased sample size, with no further reduction after about 30 measurement locations. This study highlights the ability for future winter balance and snow survey studies to optimize snow data collection within a glacierized basin.

Keywords: glacier; alpine; snow survey design; optimize; St. Elias Mountains; snow probing

INTRODUCTION

Estimates of basin-wide seasonal snow accumulation are critical for monitoring glacier mass balance and for predicting the availability and timing of surface runoff, especially in mountainous regions. The net accumulation of snow on a glacier over a winter season is known as the winter surface mass balance, or "winter balance" (WB) and is typically reported in meters of water equivalent (m w.e.) (Cogley and others, 2011). Winter balance accounts for half of the seasonally resolved mass balance, initializes ablation conditions and affects energy and mass exchange between the land and atmosphere (e.g. Hock, 2005; Réveillet and others, 2016).

Optimal survey designs for snow depth are central to accurately estimating snow distribution and mass balance from *in situ* measurements. Measuring snow depth and travelling between measurement locations is both time consuming and can disturb the snow, so care must be taken to choose a survey design that avoids bias, allows for the greatest variability to be measured and minimizes distance travelled (e.g. Shea and Jamieson, 2010; Kinar and Pomeroy, 2015). Moreover, the period of seasonal maximum snow depth can be brief, further motivating the need for time-efficient survey designs.

There are a number of different spatial sampling patterns that have been employed to obtain point measurements of snow depth, including pure random (e.g. Elder and others, 1991), linear random (e.g. Shea and Jamieson, 2010), nested (e.g. Schweizer and others, 2008), gridded random (e.g. Bellaire and Schweizer, 2008; Elder and others, 2009; Bellaire and Schweizer, 2011) and gridded (e.g. Molotch and Bales, 2005; Kronholm and Birkeland, 2007; López-Moreno and others, 2011). Sampling schemes that incorporate randomness are favourable because they limit sampling bias by varying sample spacing and direction. However, they are less efficient than designs that incorporate grids. Grid-based designs minimize travel distance but measurements are biased by regularly spaced intervals and linear orientations which can result in an under-representation of snow-depth variability (Kronholm and Birkeland, 2007).

Snow surveys on glaciers are conducted to estimate winter balance, and multi-year sampling programs are often established to monitor changes in winter balance with time. An optimized survey design requires (1) a sampling scheme that captures spatial variability and minimizes travel distance and (2) knowledge of the minimum number of measurement locations needed to estimate WB to the desired precision. Optimization of winter-balance survey design is rarely investigated because the locations of snow-depth measurements are often dictated by field resources and logistics. Few studies have investigated the number of measurement locations needed to effectively sample

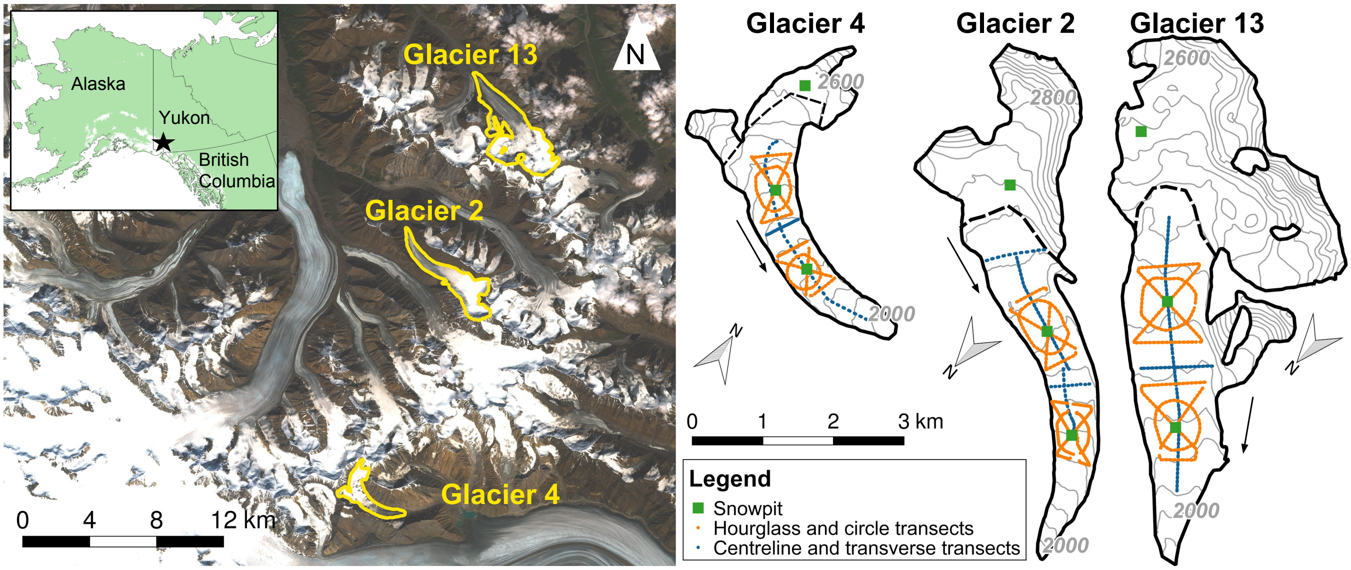

Fig. 1: Study area location and sampling patterns for Glaciers 4, 2 and 13. (Left) Study region in the Donjek Range of the St. Elias Mountains of Yukon, Canada (inset). Imagery from Landsat8 (29 August 2013, data available from the U.S. Geological Survey). (Right) Details of the spatial sampling patterns, with midline and transverse profiles (blue dots), hourglass with inscribed circle designs (orange dots) and locations of snow-density measurements (green squares). Arrows indicate ice-flow directions. Approximate location of ELA on each glacier is shown as a black dashed line. Contour interval is 50 m (grey).

the winter-balance distribution (c.f. Fountain and Vecchia, 1999; Walmsley, 2015). The sampling patterns used for most winter-balance programs do not include randomness, and measurements are typically made along the glacier centreline (e.g. **?**) to capture expected orographic effects (e.g. Grünewald and others, 2014). However, centreline surveys are known to underestimate winter balance, so transverse profiles are often added to improve the reliability of the sampling scheme (e.g. Walmsley, 2015). An hourglass with an inscribed circle (personal communication from C. Parr, 2016) is an alternative sampling scheme that captures changes in winter balance with elevation while avoiding the centreline bias, and is easy to travel. To our knowledge, no study has yet compared the performance of these different sampling patterns in estimating the spatially distributed winter balance.

The goal of this work is to determine the optimal survey design for estimating distributed and glacier-wide (integrated) winter balance. We consider both the spatial sampling scheme and the number of measurement locations in defining the optimal survey design. For three alpine glaciers in the St. Elias Mountains of Yukon, Canada, we explore the effects of survey design on: (1) the accuracy of estimated glacier-wide (integrated) winter balance in tests with synthetic data, and (2) the accuracy of estimated winter balance at distributed locations across the glacier by comparisons with real data.

STUDY SITE

We investigate winter-balance survey design for three unnamed glaciers in the Donjek Range of southwest Yukon, Canada. Situated on the northern flanks of the St. Elias Mountains, which rise sharply from the Pacific Ocean, the Donjek Range experiences a continental climate. Monitoring of snow distribution and glacier mass balance in the St. Elias Mountains began in the 1950s and 1960s with a series of research programs, including Project "Snow Cornice" and the Icefield Ranges Research Project (Wood, 1948; Danby and others, 2003). More recent studies have focused on studies of individual alpine glaciers (e.g. Clarke, 2014; Flowers and others, 2014) as well as regional glacier mass balance and dynamics (e.g. Arendt and others, 2008; Burgess and others, 2013; Waechter and others, 2015).

Glacier 4, Glacier 2 and Glacier 13 (labelling adopted from Crompton and Flowers (2016)) are small alpine glaciers (3.8–12.6 km^2) with simple geometries. The elevation of these glaciers ranges from 1900 to 3100 m a.s.l. and ELAs are located at ∼2500 m. The glaciers are generally oriented southeast-northwest in steep-walled valleys. We suspect that all three glaciers are polythermal, based on a targeted study of Glacier 2 (Wilson and others, 2013) and related theoretical modelling (Wilson and Flowers, 2013). A detailed analysis of winter-balance estimation on these three glaciers is presented by Pulwicki and others (2017).

METHODS

Below we outline the process of determining point-scale and grid-scale values of winter balance from snow depth and density measurements. We then describe the production of gridded synthetic distributions of winter balance for each of the three study glaciers, from which synthetic sample datasets can be extracted. Finally, we present our strategy

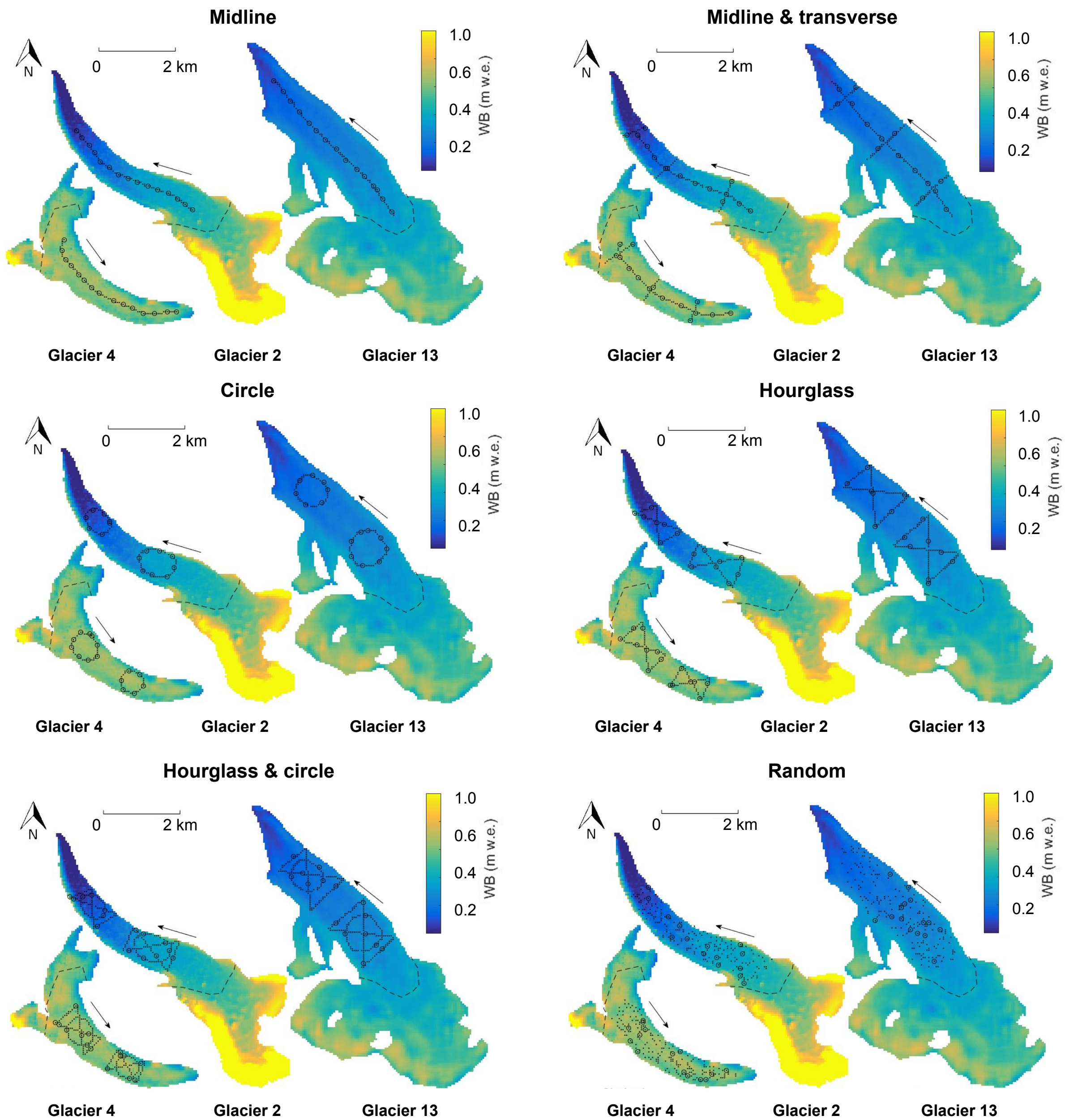

Fig. 2: Synthetic sampling patterns on the three study glaciers. All synthetic sampling locations (small black dots) and a subset of synthetic sampling locations ($n = 15$, open circles) are shown for each sampling pattern. The random pattern shows a total of 200 synthetic sampling locations. Sampling patterns are overlain on the synthetic distribution of winter balance (WB) in colour.

for evaluating survey design with real and synthetic data, using specific performance metrics.

Field measurements

Point-scale values of winter balance are obtained from direct measurements of snow depth and density (Figure 1). Snow depth was measured using 3.2 m graduated aluminium avalanche probes. Measurement locations followed linear and curvilinear transects in various spatial patterns termed "midline-" and "transverse profiles" and "hourglass with inscribed circle". Midline profiles, alone or combined with transverse profiles, comprise the most common survey designs used in winter-balance studies (e.g. Kaser and others, 2002; Machguth and others, 2006). The midline profile aims to capture changes in winter balance with elevation, while transverse profiles provide some characterization of lateral variations in winter balance. The hourglass with inscribed circle allows for sampling in multiple directions and easy travel (personal communication from C. Parr, 2016).

Sampling patterns were similar between study glaciers, with a sample spacing of 10–60 m dictated by protocols for safe glacier travel. Each observer made 3–4 individual

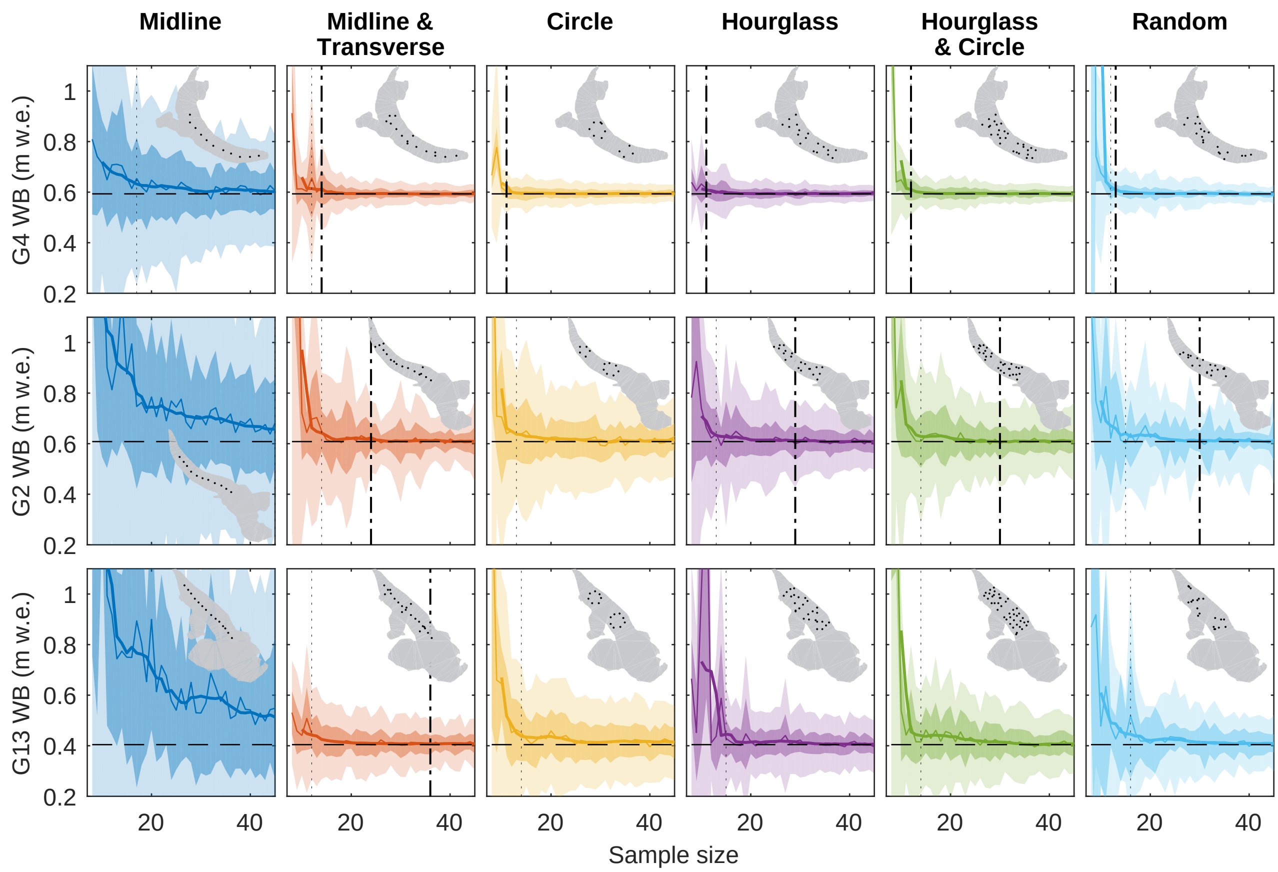

Fig. 3: Glacier-wide winter balance (WB) estimated with various sampling patterns (columns) and sample sizes (x-axes) for Glacier 4 (top row), Glacier 2 (middle row) and Glacier 13 (bottom row). Insets show example sampling patterns. Bold solid lines are smoothed mean values of WB estimated from 100 linear regressions of synthetic samples of WB on topographic parameters for each spatial sampling patterns (columns) and sample sizes from 8 to >45 (x-axes). Dark and light shading indicate the standard deviation in estimated WB for low- and high-noise cases, respectively. Horizontal dashed lines indicate the synthetic target value of glacier-wide WB. Vertical dotted lines show sample size n_c that achieves an accuracy of 5% (difference between smoothed mean and target value). Vertical dot-dashed lines show sample size n_v for which the standard deviation comes within 25% of the target value in the high-noise case. Not all sampling patterns have n_c and n_v within the range of samples sizes shown.

Table 1: Ranking of survey designs from 1 to 6 (top row) for Glaciers 4, 2 and 13 (G4, G2, G13) based on three different metrics (see text) from synthetic tests: n_c (top three rows), n_v (middle three rows) and travel distance (d) required to perform survey (bottom three rows). Sampling patterns are: midline (M), midline and transverse (MT), circle (C), hourglass (H), hourglass and circle (HC) and random (R). Colours follow figures. Values in parentheses are n_c, n_v or travel distance (km). Travel distances given for random sampling scheme are averages.

Metric		**1**		**2**		**3**		**4**		**5**		**6**	
n_c	**G4**	C, H	(8)			MT, HC	(9)			R	(10)	M	(14)
	G2	C, H	(10)			MT, HC	(11)			R	(12)	M	(43)
	G13	MT	(9)	HC	(11)	C, H	(12)			R	(13)	M	(101)
n_v	**G4**	C, H	(8)			HC	(9)	R	(10)	MT	(13)	M	(–)
	G2	MT	(21)	H	(26)	HC, R	(29)			C	(47)	M	(–)
	G13	MT	(34)	H	(44)	R	(49)	HC	(50)	C	(96)	M	(–)
d	**G4**	C	(4.8)	M	(6.3)	H	(6.9)	R	(∼7.9)	MT	(8.3)	HC	(11.1)
	G2	M	(4.3)	C	(5.7)	H	(8.4)	MT	(8.6)	R	(∼9.2)	HC	(12.5)
	G13	C	(7.0)	M	(8.0)	H	(10.6)	MT	(11.0)	R	(∼11.3)	HC	(16.8)

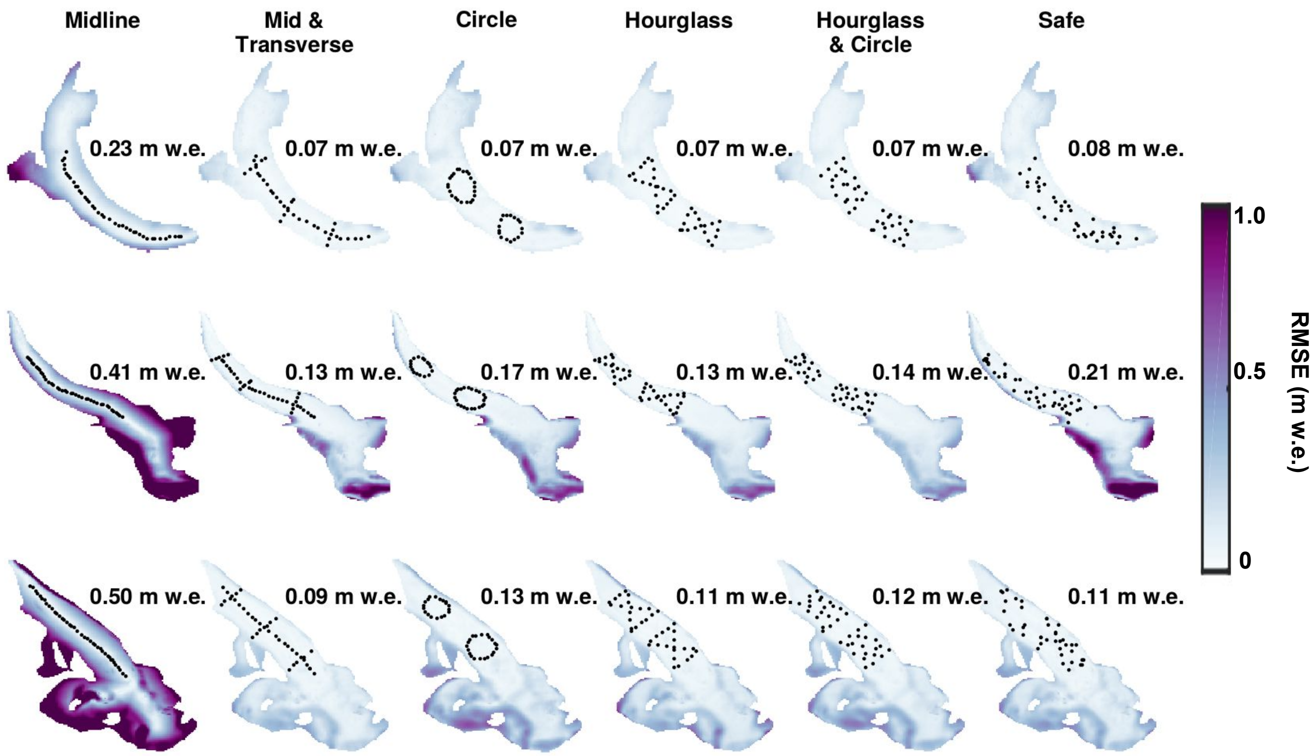

Fig. 4: RMSE between synthetic target distribution of WB and distributed WB estimated with various sample patterns, $n = 40$ and 100 regressions of the high-noise sample set. Sampling patterns are shown in columns for Glacier 4 (top row), Glacier 2 (middle row) and Glacier 13 (bottom row). RMSE for the entire glacier is displayed. Synthetic sampling locations are shown as black dots.

depth measurements within ~1 m at each survey location. Snow-depth measurements were largely restricted to the ablation area, where the clear distinction between snow and ice ensures that only snow from the most recent accumulation season is measured. In total, we collected more than 9000 snow-depth measurements throughout the study area. Snowpit-density profiles were measured using a wedge cutter and spring scale at three locations on each glacier that were distributed in elevation. A mean density was then calculated for each glacier and used to convert snow depth at all measurement locations to values of point-scale winter balance. The mean density and standard deviation was $348 \pm 13\,\mathrm{kg\,m^{-3}}$ on Glacier 4, $333 \pm 26\,\mathrm{kg\,m^{-3}}$ on Glacier 2 and $349 \pm 38\,\mathrm{kg\,m^{-3}}$ on Glacier 13. All point-scale values of winter balance located within a common digital elevation model (DEM) gridcell (40×40 m) were averaged to obtain a single value of winter balance in each sampled gridcell.

Tests with synthetic data

A synthetic glacier-wide distribution of winter balance is obtained for each study glacier by linear regression of the gridcell-averaged values of WB (obtained as described above) on topographic parameters derived from a SPIRIT SPOT-5 DEM (Korona and others, 2009). These parameters include commonly used quantities (e.g. McGrath and others, 2015) such as elevation, slope, aspect, distance from glacier centreline, "northness", curvature and a wind redistribution parameter. The regression is used, along with cross-validation and Bayesian model averaging, to obtain a set of regression coefficients for each glacier (Pulwicki and others, 2017). The resulting distribution of winter balance, hereafter referred to as the "synthetic distribution of winter balance" is determined by multiplying fitted regression coefficients by the corresponding topographic parameters for each gridcell in the DEM. We use the regression as an emulator to generate realistic winter-balance distributions that we can treat as synthetic data.

Synthetic data are used to evaluate survey design, including the sampling pattern and sample number. We investigate various survey designs that are unique combinations of six different sampling patterns: midline, midline and transverse profiles, circle, hourglass, hourglass with inscribed circle and random. The random pattern is restricted to the area encompassed by the original field measurements (Figure 1). For each sampling pattern we investigate the effect of sample size n, where n ranges from a minimum of eight (constrained by the use of seven topographic parameters in the interpolation) to a maximum determined by the number of gridcells sampled by a given pattern (ranging from 57 to 228).

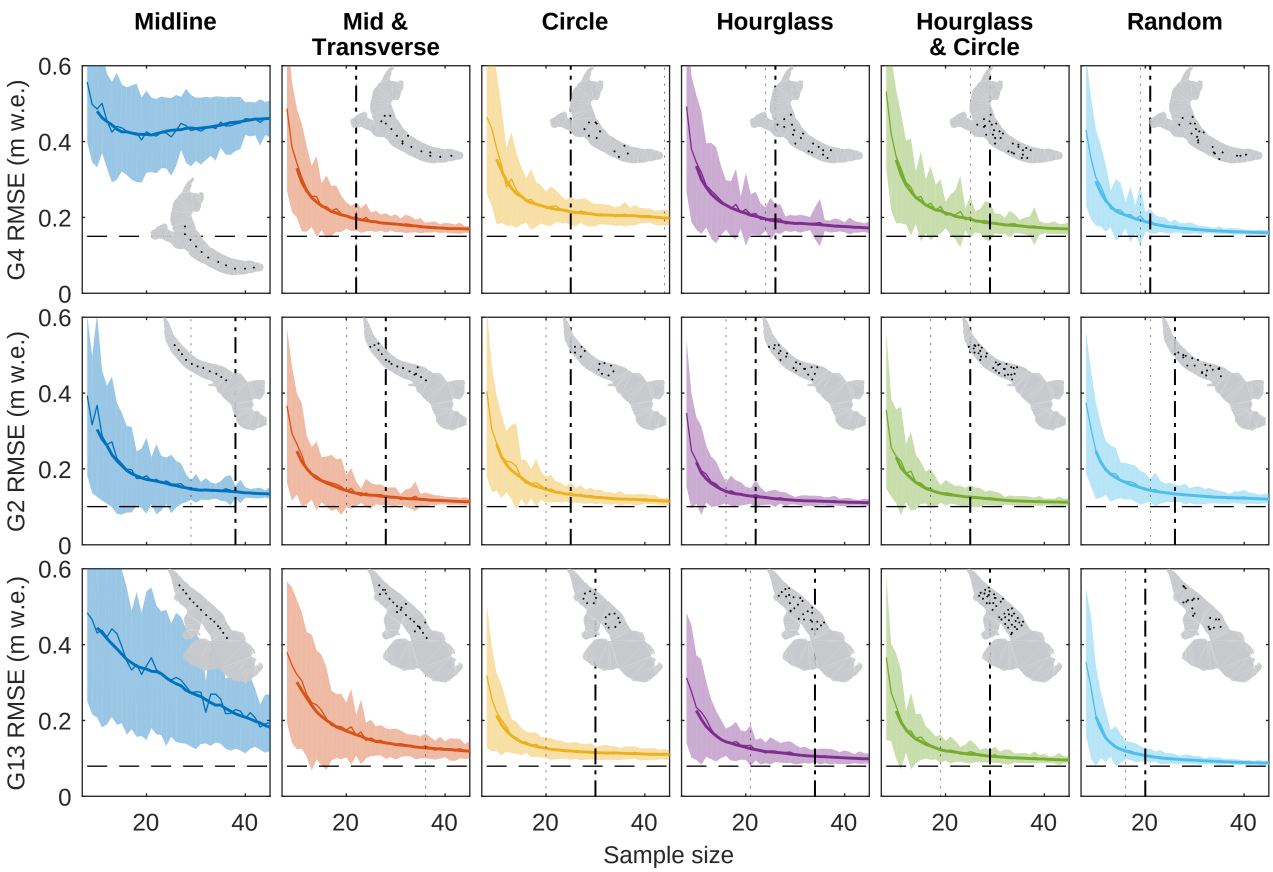

Fig. 5: RMSE between all observed values of winter balance and the co-located estimates of winter balance from randomly selected subsets of the real data, with various sampling patterns (columns) and sample sizes (x-axes) for Glacier 4 (top row), Glacier 2 (middle row) and Glacier 13 (bottom row). Insets show example sampling patterns. Bold solid lines are smoothed mean values of RMSE from 100 linear regressions. Shading indicates the standard deviation of RMSE values arising from different subsets of the randomly chosen data. Horizontal dashed lines indicate RMSE when all data are used to estimate WB. Vertical dotted lines show sample size n_c that achieves RMSE $\leq 5\%$ (difference between smoothed mean and target value). Vertical dot-dashed lines show sample size n_v for which the standard deviation comes within 25% of the target value. Not all sampling patterns have n_c and n_v within the range of samples sizes shown.

To generate a synthetic sample size n for a given sampling pattern (e.g. hourglass), we extract values of winter balance from the glacier-wide synthetic distribution that are regularly spaced within the chosen sampling pattern. The geographic locations of the patterns are fixed based on the original field measurements. We then add low or high noise to the synthetic sample set. Low noise is defined by a normal distribution with zero mean and a standard deviation derived from a series of high-density subgrid-scale measurements of winter balance on each glacier (Pulwicki and others, 2017). High noise is defined in the same way, but with the standard deviation three times that of low noise. This value is approximately equal to that derived by Pulwicki and others (2017) to represent multiple sources of uncertainty including subgrid-scale variability (as in the low-noise case), the assignment of snow density in the conversion of snow depth to winter balance and measurement interpolation. For each synthetic sample of winter balance, a randomly chosen value from the high- or low-noise distribution is added to the sample value.

To assess the performance of a given survey design (sampling pattern and sample size), we compare the original synthetic distribution of winter balance to that obtained when the topographic regression is performed using only the noisy synthetic sample set. Since the noise is sampled randomly from a distribution, we repeat the regression to create 100 realizations of "estimated winter balance" from which we can calculate a mean and standard deviation. Comparison of the "synthetic" and "estimated" winter balances allows us to assess both the accuracy (using the mean estimated winter balance) and the precision (using the standard deviation of the estimated winter balances) of a given survey design, with the caveat that the utility of the results is limited by the extent to which the simulated distribution represents a plausible reality.

Tests with real data

Using real data offers an opportunity to evaluate survey design in a different way, and provides a check on the result obtained in the synthetic tests. Here we aim to test how well the estimated winter balance compares with field measurements as a function of sampling pattern and sample number. In this series of tests, we extract subsets of n samples of the observed winter balance (as described in "Field measurements") from individual sampling patterns (e.g. hourglass with inscribed circle) and use each subset of observed values to generate an "estimated winter balance" with a linear regression as described above. Because these values are real data, they already contain noise. Instead of adding noise, as in the synthetic case, we repeat the regression 100 times with randomly selected subsets of n values; this approach means the values will not be regularly spaced. From these 100 realizations we calculate a mean and standard deviation of the estimated winter balance. We then compute the RMSE between all observed values of winter balance and the co-located estimates of winter balance. This comparison allows us to assess both the accuracy and precision [NOT QUITE] of a given survey design in predicting real data, but is limited to the locations of the original field measurements. By its nature, this comparison also combines the performance of the survey design with the performance of the regression.

Performance metrics

To quantify the performance of each survey design, we use two metrics. The first metric we term "convergence", defined as the sample size (n_c) needed to obtain a mean estimated glacier-wide winter balance within 5% of the target value in the synthetic tests, or RMSE $\leq$5% in the comparison with real data. To determine n_c we smooth the mean estimated values of glacier-wide winter balance as a function of n to avoid spurious results arising from the random nature of adding noise to the synthetic samples or selecting the subset of observed values. Convergence is a metric that describes the minimum number of measurement locations required to produce an accurate (within 5%) estimate of winter balance. The second metric we term "variability", defined as the sample size (n_v) needed to obtain a standard deviation of the estimated glacier-wide winter balance within 25% of the target value in the synthetic tests, or RMSE $\leq$25% in the comparison with real data. We again smooth the standard deviation of estimated values of glacier-wide winter balance as a function of n, in order to determine n_v. "Variability" is a metric that describes the sensitivity of the estimated winter balance to noise in the synthetic tests and sampling locations in the comparisons with real data. We also compute the total travel distance required to obtain the measurements required for a given survey design. An efficient survey will have low values of n_c and n_v as well as a short travel distance.

RESULTS

Synthetic data

Based on synthetic data, the most effective sampling patterns are midline & transect and hourglass. Midline & transect and hourglass quickly converge to the synthetic glacier-wide value of WB, have comparatively low standard deviations due to noise and involve the shortest travel distances (Figure 3 and Table 1). Hourglass requires slightly fewer measurements to converge to the synthetic WB and is slightly shorter than midline & transect but midline & transect has slightly lower variability. If using midline & transect or hourglass sampling patterns, a sampling size of $\sim$35 is sufficient to estimate glacier-wide WB on all glaciers with no marked improvement in accuracy or precision at larger sample sizes. With all sampling patterns, WB is significantly overestimated at low sample size.

The remaining sampling patterns are less effective for estimating WB. Circle and hourglass & circle also require small sample sizes for convergence but circle has exceptionally high variability on Glaciers 2 and 13 and hourglass & circle requires a substantially greater travel distance. Notably, hourglass & circle is not a substantial improvement on the hourglass sampling pattern given the considerable increase in travel distance required to sample the along the hourglass & circle. By far the worst sampling pattern is the midline. WB values are highly sensitive to noise and even with a large number of measurement locations the validation values of WB are not achieved.

Although the convergence patterns are similar between glaciers, the standard deviation due to noise and the sample size needed to accurately estimate WB are smallest for Glacier 4 and greatest for Glacier 13 for all sampling patterns. RMSE is also larger on Glaciers 2 and 13 than on Glacier 4, especially in the accumulation area (Figure 4). When midline pattern is used, the errors are exceptionally high in areas that do not fall along the midline for all glaciers.

Real data

The RMSE rapidly decreases with an increased sample size for all sampling patterns on all three glaciers (Figure5). The sampling patterns trends are similar to those of the synthetic data. Midline & transverse, hourglass, circle, hourglass & circle and random all perform well. However, the higher travelling distance of hourglass & circle and random make these patterns less efficient. Again, the worst sampling pattern is the midline, which results in exceptionally high values of RMSE for all three glaciers. There is no marked decrease in RMSE for sample sizes greater than $\sim$ 30 for all sampling patterns. None of the sampling patterns result in the RMSE reaching that of the full data for all three glaciers.

DISCUSSION

The optimal survey design for our study region is the midline & transverse or hourglass pattern with a sample size of $\sim$ 30. Based on both synthetic and real data, both designs result in low error, is least sensitive to noise and involves relatively small travelling distance. A sample size greater than 30 does not significantly improve the accuracy of WB estimated and does not decrease error. This surprisingly low number of measurement locations indicates that high-resolution sampling is not required to accurately estimate WB. Instead, taking measurements throughout the study basin should be prioritized when choosing a survey design. If the goal is to estimate WB then field resources should be more strongly allocated to distributing measurement locations to capture basin-scale spatial trends in WB since error is greatest in areas far from sampling locations and in areas with extreme values of topographic parameters.

The most effective survey designs capture the dominant WB-elevation trend but also have measurement locations in gridcells that span the range of other topographic parameters. Despite the fact that the synthetic distribution of WB on Glaciers 2 and 13 is largely controlled by elevation, measurements locations that are not along the midline are needed to constrain the regression. When the midline pattern is used, the topographic parameters (excluding elevation) fall within a narrow range so the regression becomes sensitive to noise. Further, the accuracy of estimated WB does not improve at large sample sizes along a midline pattern because the sampling locations do not capture relationships between WB and the remaining topographic parameters. However, if there are a large portion of measurement locations far from the midline then the elevation trend may be less strong resulting in larger errors. Random and hourglass & circle patterns may have larger errors on Glaciers 2 and 13 as a result of the higher proportion of measurement locations far from the midline.

On Glacier 4, the standard deviation of WB due to noise and the number of measurement locations needed to estimate WB is smaller than Glaciers 2 and 13. This trend is likely a result of the small influence of topographic parameters in the distribution of synthetic distribution of WB. The synthetic distribution of WB is derived from a linear regression between field measurements of WB and topographic parameters that explains little of the observed variance (R^2 =0.07). As a result, glacier-wide WB is approximately equal to the mean of WB field measurements. This mostly uniform distribution of snow can therefore be described with few measurement locations, regardless of where the measurements are obtained.

We find that, on average, WB is over-estimated with small sample sizes (< 20), especially on Glaciers 2 and 13 with $n < 10$. We hypothesize that over-estimation results from an exaggeration of the elevation regression coefficient, which is the dominant explanatory parameter for Glaciers 2 and 13. At small sample sizes, the elevation trend over-shadows the relationships with other topographic parameters and is sensitive to noise, which results in large WB values in the accumulation area. Our finds are inconsistent with Walmsley (2015), who found that using only midline probings underestimates WB.

Further work into the optimization of snow survey survey design is needed. The most obvious limitation of our study is the restriction of survey design to accessible locations in the ablation area. Lack of measurements in the accumulation area is a sub-optimal survey design. However, direct measurements of snow depth in the accumulation area are extremely costly because a large amount of time is needed to dig a snow pit, especially in regions with high accumulation. Given the importance of spatially distributed measurements though, there is likely an optimal balance between costly measurements in the accumulation area and less costly measurements in the ablation area. Use of more efficient snow-depth measurement methods, such a ground-penetrating radar (GPR) and lidar, in the accumulation area can also be investigated. Additional extensions to our work include an investigation of the size and placement of hourglass patterns and to optimize measurement locations based on factors such as topographic parameters.

CONCLUSION

From an analysis of synthetic and real WB data, we find that 'midline & transect' and'hourglass' sampling patterns with a sample size of approximately 30 measurement locations is the optimal survey design for estimating glacier-wide WB. Since a relatively low sample size is needed and errors are greatest in the accumulation area, we recommend that field resources be allocated such that measurements locations are distributed throughout the sample basin rather than obtaining high-resolution WB measurements. A midline sampling pattern results in poor estimates of WB and high sensitivity to noise. We find that WB is over-estimated with very small sample sizes ($n < 10$) and along the midline so care should be made to adequately sample the basin to capture changes in elevation as well as transverse variation in WB.

ACKNOWLEDGEMENTS

We thank the Kluane First Nation (KFN), Parks Canada and the Yukon Territorial Government for granting us permission to work in KFN Traditional Territory and Kluane National Park and Reserve. We are grateful for financial support provided by the Natural Sciences and Engineering Research Council of Canada, Simon Fraser University (including the KEY Big Data Initiative) and the Northern Scientific Training Program. We kindly acknowledge Kluane Lake Research Station, Sian Williams, Lance Goodwin and Trans North pilot Dion Parker for facilitating field logistics. We are grateful to Alison Criscitiello and Coline Ariagno for all aspects of field assistance and Sarah Furney for assistance with data entry. Thank you to Etienne Berthier for providing us with the SPIRIT SPOT-5 DEM and for assistance in DEM correction. We are grateful to Derek Bingham and Michael Grosskopf for assistance with statistics. Laura Thomson, Leif Anderson, Dave Bigelow and Erik Young all provided thoughtful and constructive comments on drafts of the manuscript.

REFERENCES

Arendt AA, Luthcke SB, Larsen CF, Abdalati W, Krabill WB and Beedle MJ (2008) Validation of high-resolution

GRACE mascon estimates of glacier mass changes in the St Elias Mountains, Alaska, USA, using aircraft laser altimetry. *Journal of Glaciology*, **54**(188), 778–787 (doi: 10.3189/002214308787780067)

Bellaire S and Schweizer J (2008) Deriving spatial stability variations from penetration resistance measurements. In *Proceedings ISSW*, 188–194

Bellaire S and Schweizer J (2011) Measuring spatial variations of weak layer and slab properties with regard to snow slope stability. *Cold Regions Science and Technology*, **65**(2), 234–241 (doi: 10.1016/j.coldregions.2010.08.013)

Burgess EW, Forster RR and Larsen CF (2013) Flow velocities of Alaskan glaciers. *Nature communications*, **4**, 2146–2154 (doi: 10.1038/ncomms3146)

Clarke GK (2014) A short and somewhat personal history of Yukon glacier studies in the Twentieth Century. *Arctic*, **37**(1), 1–21

Cogley J, Hock R, Rasmussen L, Arendt A, Bauder A, Braithwaite R, Jansson P, Kaser G, Möller M, Nicholson L and others (2011) *Glossary of glacier mass balance and related terms*. UNESCO-IHP, Paris

Crompton JW and Flowers GE (2016) Correlations of suspended sediment size with bedrock lithology and glacier dynamics. *Annals of Glaciology*, **57**(72), 1–9 (doi: 10.1017/aog.2016.6)

Danby RK, Hik DS, Slocombe DS and Williams A (2003) Science and the St. Elias: an evolving framework for sustainability in North America's highest mountains. *The Geographical Journal*, **169**(3), 191–204 (doi: 10.1111/1475-4959.00084)

Elder K, Dozier J and Michaelsen J (1991) Snow accumulation and distribution in an alpine watershed. *Water Resources Research*, **27**(7), 1541–1552 (doi: 10.1029/91WR00506)

Elder K, Cline D, Liston GE and Armstrong R (2009) NASA Cold Land Processes Experiment (CLPX 2002/03): Field measurements of snowpack properties and soil moisture. *Journal of Hydrometeorology*, **10**(1), 320–329 (doi: 10.1175/2008JHM877.1)

Flowers GE, Copland L and Schoof CG (2014) Contemporary Glacier Processes and Global Change: Recent Observations from Kaskawulsh Glacier and the Donjek Range, St. Elias Mountains. *Arctic*, **67**, 22–34, copyright - Copyright Arctic Institute of North America 2014; Document feature - Maps; Illustrations; Graphs; ; Last updated - 2015-04-07

Fountain AG and Vecchia A (1999) How many stakes are required to measure the mass balance of a glacier? *Geografiska Annaler: Series A, Physical Geography*, **81**(4), 563–573 (doi: 10.1111/1468-0459.00084)

Grünewald T, Bühler Y and Lehning M (2014) Elevation dependency of mountain snow depth. *The Cryosphere*, **8**(6), 2381–2394 (doi: 10.5194/tc-8-2381-2014)

Hock R (2005) Glacier melt: a review of processes and their modelling. *Progress in Physical Geography*, **29**(3), 362–391 (doi: 10.1191/0309133305pp453ra)

Kaser G, Fountain A, Jansson P and others (2002) *A manual for monitoring the mass balance of mountain glaciers*. IHP-VI- Technical documents in hydrology

Kinar N and Pomeroy J (2015) Measurement of the physical properties of the snowpack. *Reviews of Geophysics*, **53**(2), 481–544 (doi: 10.1002/2015RG000481)

Korona J, Berthier E, Bernard M, Rémy F and Thouvenot E (2009) SPIRIT SPOT 5 stereoscopic survey of Polar Ice: Reference images and topographies during the fourth International Polar Year (2007–2009). *ISPRS Journal of Photogrammetry and Remote Sensing*, **64**(2), 204–212 (doi: 10.1016/j.isprsjprs.2008.10.005)

Kronholm K and Birkeland KW (2007) Reliability of sampling designs for spatial snow surveys. *Computers & geosciences*, **33**(9), 1097–1110 (doi: 10.1016/j.cageo.2006.10.004)

López-Moreno JI, Fassnacht S, Beguería S and Latron J (2011) Variability of snow depth at the plot scale: implications for mean depth estimation and sampling strategies. *The Cryosphere*, **5**(3), 617–629 (doi: 10.5194/tc-5-617-2011)

Machguth H, Eisen O, Paul F and Hoelzle M (2006) Strong spatial variability of snow accumulation observed with helicopter-borne GPR on two adjacent alpine glaciers. *Geophysical Research Letters*, **33**(13), 1–5 (doi: 10.1029/2006GL026576)

McGrath D, Sass L, O'Neel S, Arendt A, Wolken G, Gusmeroli A, Kienholz C and McNeil C (2015) End-of-winter snow depth variability on glaciers in Alaska. *Journal of Geophysical Research: Earth Surface*, **120**(8), 1530–1550 (doi: 10.1002/2015JF003539)

Molotch NP and Bales RC (2005) Scaling snow observations from the point to the grid element: Implications for observation network design. *Water Resources Research*, **41**(11) (doi: 10.1029/2005WR004229)

Pulwicki A, Flowers G and Radić V (2017) Uncertainties in estimating winter balance from direct measurements of snow depth and density on alpine glaciers. *Journal of Glaciology*, (in submission)

Réveillet M, Vincent C, Six D and Rabatel A (2016) Which empirical model is best suited to simulate glacier mass balances? *Journal of Glaciology*, **63**(237), 1–16 (doi: 10.1017/jog.2016.110)

Schweizer J, Heilig A, Bellaire S and Fierz C (2008) Variations in snow surface properties at the snowpack-depth, the slope and the basin scale. *Journal of Glaciology*, **54**(188), 846–856 (doi: 10.3189/002214308787780058)

Shea C and Jamieson B (2010) Star: an efficient snow point-sampling method. *Annals of Glaciology*, **51**(54), 64–72 (doi: 10.3189/172756410791386463)

Waechter A, Copland L and Herdes E (2015) Modern glacier velocities across the Icefield Ranges, St Elias Mountains, and variability at selected glaciers from 1959 to 2012. *Journal of Glaciology*, **61**(228), 624–634 (doi: 10.3189/2015JoG14J147)

Walmsley APU (2015) *Long-term observations of snow spatial distributions at Hellstugubreen and Gråsubreen, Norway*. Master's thesis

Wilson N and Flowers G (2013) Environmental controls on the thermal structure of alpine glaciers. *The Cryosphere*, **7**(1), 167–182 (doi: 10.5194/tc-7-167-2013)

Wilson NJ, Flowers GE and Mingo L (2013) Comparison of thermal structure and evolution between neighboring subarctic glaciers. *Journal of Geophysical Research: Earth Surface*, **118**(3), 1443–1459 (doi: 10.1002/jgrf.20096)

Wood WA (1948) Project "Snow Cornice": the establishment of the Seward Glacial research station. *Arctic*, **1**(2), 107–112