REVISITING AVALANCHE DESTRUCTIVE SCALE AND ASSOCIATED MASS ESTIMATES

Erich H. Peitzsch[1,2]*, John M. Sykes[3,1], Zachary M. Guy[4], Ron Simenhois[5], Bruce Jamieson[6]

[1] U.S. Geological Survey, Northern Rocky Mountain Science Center, West Glacier, MT, USA
[2] Montana State University, Department of Earth Sciences, Bozeman, Montana, USA
[3] Simon Fraser University, Avalanche Research Program, Burnaby, BC, Canada
[4] Crested Butte Avalanche Center, Crested Butte, CO, USA
[5] Colorado Avalanche Information Center, Denver, CO, USA
[6] Dept of Civil Engineering, University of Calgary, AB, Canada

ABSTRACT: Avalanche destructive (D) size classifications are essential for forecasting, communication, and risk management, yet size estimation remains one of the more subjective elements of avalanche assessment. Practitioners typically rely on visual impressions of destructive potential, path length, and estimated mass and volume. Recent work indicates that size is often underestimated even by experienced observers, especially when based on photographs or distant observations. Although mass is central to defining destructive size, the empirical foundations of mass–size relationships have rarely been revisited, and expert variability in size assignment could be decreased if more consistent operational tools were available. We evaluated whether a simple, practitioner-focused calculation tool using slab dimensions, snow density, and snow water equivalent can provide reliable mass estimates that correspond meaningfully with existing destructive size classes. Using this tool, we estimated avalanche mass and compared the results to observer-assigned D sizes. We also simulated mass outputs using a dense flow avalanche simulation module and compared observer estimates and simple-tool outputs to recent volume-based methods for D-size classification. Across 143 well-documented dry snow avalanches from multiple avalanche climates, the size estimation tool tends to classify avalanches larger than field-based observers. We found a mean bias of 0.47 D-size units and a typical (mean absolute) disagreement of 0.56. Our results show that 95% of events landed within one whole D-class of each other and 71% within a half-class. Agreement between the simple tool mass estimates and dense flow avalanche simulations was generally strong, though discrepancies emerged in cases with substantial entrainment or complex terrain. We compared the size tool to the simulation module with entrainment (n=4) and without entrainment (n=8). With entrainment, the two methods estimate the same D-size class in 4 of 4 comparable events and 7 of 8 events without entrainment. Also, the simulation module estimates larger sizes than the observer in 7 of 8 cases without entrainment and all 4 cases with entrainment. These findings support that development of practical, mass-based tools could improve D-size consistency and emphasize the value of revisiting historical mass–size relationships. The accompanying application offers practitioners an accessible and operationally useful resource for improving D-size estimation from field observations.

KEYWORDS: Destructive Scale, Avalanche Size, Avalanche Mass.

1. INTRODUCTION

Snow avalanche size is a key component of avalanche hazard and important for forecasting, land use planning, infrastructure protection, and avalanche mitigation (Statham et al., 2017). Avalanche size classification underpins operational risk management, professional communication, and the long-term datasets that support avalanche research. In North America, the destructive size scale (D-scale) has served as the standard since McClung and Schaerer (1981) introduced it, and it is included in both the U.S. observation guidelines (Snow Weather and Observation Guidelines for the United States (SWAG); Greene et al., 2022) and the Canadian guidelines (Observation Guidelines and Recording Standards for Weather, Snowpack, and Avalanches (OGRS); Canadian Avalanche Association, 2024).

The scale originated as an extension of an earlier destructive-effects scale (U.S.D.A., 1961), and McClung and Schaerer (1981) formalized it using 744 avalanches recorded over two winters at Rogers Pass, British Columbia. The scale defines five classes by destructive potential, asking observers to estimate the harm an avalanche could cause to an object positioned at terminal velocity near the middle of the path. Each class carries typical values for mass, path length, and impact pressure that increase by roughly an order of magnitude per class. McClung and Schaerer (1981) championed an "all observables" approach, arguing that no single quantity captures

Erich Peitzsch
U.S. Geological Survey
215 Mather Dr., West Glacier, MT, USA 59936
epeitzsch@usgs.gov

size adequately, and they limited the scale to five classes because subjectivity precluded finer resolution.

Both current standards codify these typical values. SWAG (Greene et al., 2022) lists typical mass (10 t to 10^5 t), path length, and impact pressure, and it retains the R-scale (relative to path) to preserve continuity with older records in the United States. OGRS (Canadian Avalanche Association, 2024) adds typical deposit volumes for each class. Avalanche and snow safety operations often use the most relevant parameters (i.e. destructive potential, mass, volume, impact pressure, and R-scale) to suit the needs of their operation and maintain consistency (Jamieson et al., 2014). Despite the assumption of reasonable agreement, recent research documents substantial variability in size estimates. Moner et al. (2013) found that experts agreed with the most common size only 62% of the time. Jamieson et al. (2014) surveyed 22 practitioners rating 18 avalanches and reported 79% agreement on full sizes. However, they cautioned that less experienced observers and larger avalanches would drive more variability. Hafner et al. (2023) surveyed 170 professionals and found that any two estimates agreed only 53% of the time with an individual-estimate reliability of 0.72. This result alone has implications for research on avalanche size because a model validated against single-observer labels can reach only about 73% accuracy by chance alone.

Carson et al. (2023) identified that interpretation differences and operational context could contribute to inconsistent ratings. They also noted historical drift in practitioner application of the scale as risk-based decision-making and measurement technology advanced over time. Observers frequently underestimate avalanches viewed from a helicopter or from a distance (Hafner et al. 2023). Hafner et al. (2023) also found that North American practitioners rate avalanches systematically smaller than European practitioners, partially due to differing scale definitions. The North American scale specifies mass whereas the European scale specifies volume. Carson et al. (2023) further pointed to stress, operational pressure, and observer bias, and argued that measurable indicators such as mass and volume offer more reliable ground for resolving disagreements than the more subjective destructive potential classification.

In addition to observer variability, typical mass and volume values are also a source of uncertainty. Jamieson et al. (2014) investigated a Size 5 case study on Mt. Tumbledown in British Columbia, Canada, and showed that the typical masses for Size 4 and Size 5 avalanches in the original scale are roughly an order of magnitude lower than values from other sources. They suggested that explosive control at Rogers Pass, the study site dataset used in McClung and Schaerer (1981), reduced deposit masses during calibration, thus artificially biasing mass for the largest avalanches. Jamieson et al. (2014) noted that accurately measuring avalanche mass is extremely challenging with most mass estimates relying on measured volume combined with an assumed or estimated snow density. They also found that width-length-depth calculations overestimate deposit volume by about 86% on average. Additionally, they identified the phrase "largest snow avalanches known" as a deterrent to classifying extreme events as Size 5, a point Carson et al. (2023) echoed in recommending its removal. Grist et al. (2023) illustrated these principles operationally, calculating a deposit mass of 2.4×10^8 kg for the Jack MacDonald avalanche in British Columbia, Canada in 2022. They used mass, volume, run length, and destroyed forest area to classify it as Size 5. Entrainment further complicates mass estimations. For large avalanches, mass can increase by a factor of 4 to 10 after it releases (McClung and Schaerer, 2006; Sovilla et al., 2006). Entrainment contributes to mass, destructive potential, and runout but is difficult to quantify.

These limitations and uncertainty highlight the need for greater objectivity when classifying avalanche size. Carson et al. (2023) proposed volume descriptors and annotated imagery, and Hafner et al. (2023) recommended standardized training, consensus estimates, and eventually automation. Mass and volume offer common objective approaches to size classification that avalanche practitioners can refer to when disagreement or uncertainty arises. Here, we introduce a simple tool (https://avystormtracker.com/Avalanche_Destructive_Size) that converts routine field observations into mass estimates to aid in assigning avalanche sizes. In this study, our objectives are:

1. Develop a simple tool that can be used to calculate mass as a means of helping with size estimates and to improve consistency within and across avalanche operations.

2. Assess mass of individual avalanches using a mass-based tool to estimate D-size and compare against field observer D-size estimates.

3. Revisit original mass associations categorized using the D-scale and determine how using the tool to calculate the D-size compares to the original estimate.

To achieve these objectives, we attempt to answer these questions: 1) Can we use simple estimates of snow water equivalent or density, slab

dimensions (volume), and entrainment to calculate mass for improving accuracy of D-size estimates? and 2) How do the resultant size tool estimates align with observer D-scale classifications and physics-based simulated mass estimates?

2. DATA

We collected a robust dataset of well-documented dry slab avalanches by sourcing previous avalanche events from regional avalanche forecast centers in the western United States and Austria. For quality control, we included only avalanches that were observed and documented by avalanche professionals, and avalanches where dimensions and snowpack characteristics could be reasonably estimated using photos or on-site measurements. A portion of the data comes from avalanche fatality and accident datasets, which are typically the most thoroughly documented avalanches.

The dataset included 143 unique avalanche events comparing original field-rated D-sizes to mass and volume estimates using our size tool. Of the 143 events, 130 were assessed using a mass-based calculation from start zone characteristics, and 13 used a volume-based calculation from runout characteristics. Thirty-three events had both representative snow profiles and snow water equivalent estimates from their start zones, allowing comparison between the two mass-based approaches. The dataset consists of unique detailed avalanche events from several snow climates throughout Colorado (29%), Alaska (24%), Montana (11%), Austria (10%), Idaho (8%), Wyoming (6%), Washington (5%), and Utah (3%), The dataset included a broad range of field-rated D-sizes (Table 1).

Table 1. D-sizes as reported by field observers.

D-size (Observer reported)	n	%
D1	6	4.2
D1.5	10	7.0
D2	41	28.7
D2.5	20	14.0
D3	39	27.3
D3.5	8	5.6
D4	16	11.2
D4.5	1	0.7
D5	2	1.4

3. METHODS

3.1 D-scale mass estimation tool

We used the size tool to estimate the mass of a slab release, and optionally, additional mass from entrainment, using the dimensions and density of the avalanche release. We, or the observers of the avalanches in the dataset, estimated the dimensions of the start zone from on-site field measurements or by mapping the release area using crown photos and satellite or aerial imagery. We defined the width of the slab by the crown and flanks and estimated the downslope length of the slab using terrain and slope angle tools, or the stauchwall if it was still visible. Some users included entrainment for avalanche events with appropriate data where they estimated the area of the track and runout using mapping software, photos, and field measurements.

In the tool, the density of the slab was calculated using two approaches: a snow water equivalent (SWE) method or a snow profile method. In situations where the slab releases on a known weak layer with a known burial date, the cumulative SWE since weak layer burial serves as an estimate for the density of the slab. This cumulative SWE is tallied from nearby snow telemetry (SNOTEL) or manual snow study sites, and can be adjusted to account for anticipated differences caused by wind loading or scouring and orographics based on expert judgment and local terrain familiarity. For example, when planning for extreme avalanches, Schaerer (2002) proposed that height of snow water equivalent (HSW) can be reduced by 30% on windward or scoured slopes, and that HSW can be increased by 50% for a lee start zone (Jamieson, 2018). When a crown profile or representative snow profile was conducted, we estimated slab density using empirical relations to hand hardness and grain type (Kim and Jamieson, 2014).

Optionally, entrainment density can be calculated using either hand hardness or SWE. Observers estimated the depth and hardness of entrained surface snow or the amount of SWE available for entrainment in the upper snowpack in situations where the avalanche gains significant mass as it moves through the track and into the runout.

We incorporated user uncertainty estimates into the tool for each input parameter section. This included uncertainty for slab dimensions and density for the start zone method, and volume in the runout method. We then used the root sum square (RSS) to calculate an overall uncertainty that minimizes inflation and captures the most influential sources of error, assuming all inputs are independent. We used the uncertainty estimates to calculate a confidence band around the mass

estimates, helping users visualize the range of potential D-sizes for each event.

With these mass estimates, the tool assigns D-sizes using the typical mass values reported in SWAG and OGRS (D1 = 10 t, D2 = 100 t, D3 = 1,000 t, D4 = 10,000 t, D5 = 100,000 t). These typical values serve as the geometric midpoints of logarithmic bins (see https://avystormtracker.com/Avalanche_Destructive_Size for classification bins). Each full D-size class and each half-step class therefore spans a factor of √10 (≈ 3.16) on either side of its typical value. For example, masses from roughly 56–178 t are classified as D2, 178–562 t as D2.5, 562–1,778 t as D3, and so on. We also included a calculated volume estimate for the deposit in the tool. Because debris density is highly variable, the tool calculates D-size directly from measured deposit volume using the typical volume classification as per Jamieson et al. (2014). We used the same binning techniques for volume as we did for mass, and we also allowed users to input custom debris density for an optional mass reference.

3.2 AvaFrame com1DFA

For a subset of eight individually investigated avalanche case studies, we compared size tool and observer-assessed D-size estimates against dense-flow avalanche simulations using AvaFrame. AvaFrame is an open-source software package for avalanche dynamics simulation and hazard mapping (Oesterle et al., 2026). Simulations were run using AvaFrame's com1DFA module, a depth average dense-flow avalanche model based on a Lagrangian, smoothed particle numerical scheme (Tonnel et al., 2023). We simulated each event both with and without a snow entrainment module active along the avalanche track. Given the small sample size (n = 8), we reported per-event differences and simple agreement counts (e.g., number of events for which the modeled and observer-assessed D-size classes matched exactly) as descriptive case-level results. Our intention was to provide an independent, physics-based benchmark against which to contextualize the size tool's mass-based estimates, distinct from the subjective destructive-force judgments used in the observer comparison.

3.3 Statistical Analysis and Comparisons

We compared size tool-calculated D-size estimates against observer-assessed D-size (following the destructive-force scale of the American Avalanche Association's SWAG guidelines (Greene et al., 2022) for 143 independent avalanche events. We quantified agreement using mean bias (bias), mean absolute error (MAE), and the percentage of events agreeing within one-half and one whole D-size class. Because D-size classes are ordinal and approximately log-scaled with respect to avalanche mass, we additionally converted MAE into an approximate multiplicative mass ratio via 10^{MAE} to express typical disagreement in an interpretable way. We tested directional bias using the Wilcoxon signed-rank test (Wilcoxon, 1945), and size-dependent disagreement using a Bland-Altman analysis, comparing two estimates against their difference and regressing the difference on estimated magnitude (Bland and Altman, 1986).

4. RESULTS

4.1 Analysis Set

In our dataset, 130 (91%) events were user-submitted events that applied a mass-based start-zone estimate and 13 (9%) volume-based runout estimates. We conducted a separate dual-run analysis of SWE-based method versus snow profile (layer density+grain type) method, resulting in 33 clean event pairs with both density approaches. Both density estimate approaches were similar when compared to observer estimates with a mean tool spread of 0.09 D-size units when compared to observer estimates (snow profile MAE=0.56; SWE-based density MAE=0.55). For event pairs, we chose the snow profile density method as the approach when comparing to observer estimates for this analysis because layer density and grain type reflect directly observed snowpack properties measured in the profile and SWE-based density can depend on a remote or estimated SWE source. Additionally, we focused most of the analysis and interpretation on mass estimates because we have a much larger sample size when compared to volume estimates.

4.2 Overall Agreement Between Tool-Calculated and Observer-Assessed D-size

We focused on three metrics: bias direction, MAE, and percent-within-one-class. These metrics provide insight into the magnitude and direction of disagreement between the size tool and the field observer assessment. Across the 143 events, the size tool estimated slightly larger values of D-size than the observer's field estimate (Figure 1). The tool produced a mean bias of 0.47 D-size units and an MAE (disagreement) of 0.56 classes. Overall, 95% of size tool estimates fell within one whole D-class of the observer estimate, and 71% fell within one-half D-size class of the observer's estimate. The disagreement was not a constant offset. Mass-based start-zone events showed a mean bias of 0.48, but the much smaller volume/runout subset showed a mean

bias of 0.35 (n=13), which we interpreted cautiously. The disagreement gap also widened at larger avalanche sizes, where the size tool provided progressively larger estimates relative to the observer estimates as size increased.

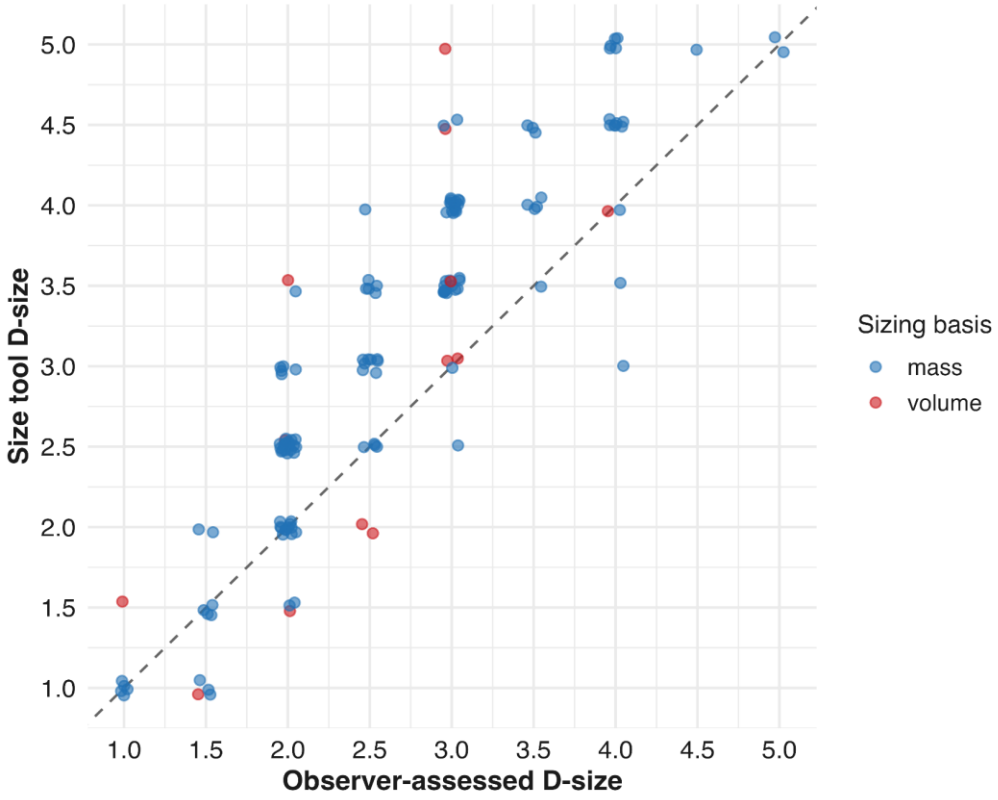


Figure 1: Tool-calculated D-size plotted against field-assessed D-size for n=143 avalanche events (points jittered slightly to reduce overplotting at repeated integer/half-integer values). Point color indicates sizing basis (blue=mass-based, red=volume/runout-based). The dashed diagonal line indicates exact agreement between the tool and the field observer.

The direction of disagreement between the size tool and observer estimates was also statistically significant (p<0.001, Table 2) using the start zone approach. The majority of cases consisted of the size tool estimating larger D sizes than the observer in 98 of 143 cases (69%).

Table 2. Disagreement of size tool and observer estimates of D-size. Direction of bias shows whether the tool estimated the avalanche larger than (>), smaller than (<), or the same (Exact match) as the field observer for each event. *n*= the raw counts of events in each direction, and Percent is the share of events in each direction.

Sizing Basis	***Direction of Bias***	***n***	***Percent***
Mass	Tool > Observer	92	70.8
Mass	Exact match	30	23.1
Mass	Tool < Observer	8	6.2
Volume	Tool > Observer	6	46.2
Volume	Exact match	3	23.1
Volume	Tool < Observer	4	30.8

4.3 Does Disagreement Depend on Size?

We used the Bland-Altman test to examine size tool and observer estimate disagreement with avalanche magnitude (Figure 2). Disagreement increased significantly with avalanche size (slope = 0.27, p<0.001). The difference between size tool and observer estimates averaged approximately 0.27 D-size units among the smallest third of events (D-size range ≈ 1–2) but increased to approximately 0.76 units among the largest third (D-size range ≈ 3–5).

However, our sample size at the largest classes (D4 and 5) is limited thereby influencing the weight of the upper limit of the range so this should be interpreted with caution.

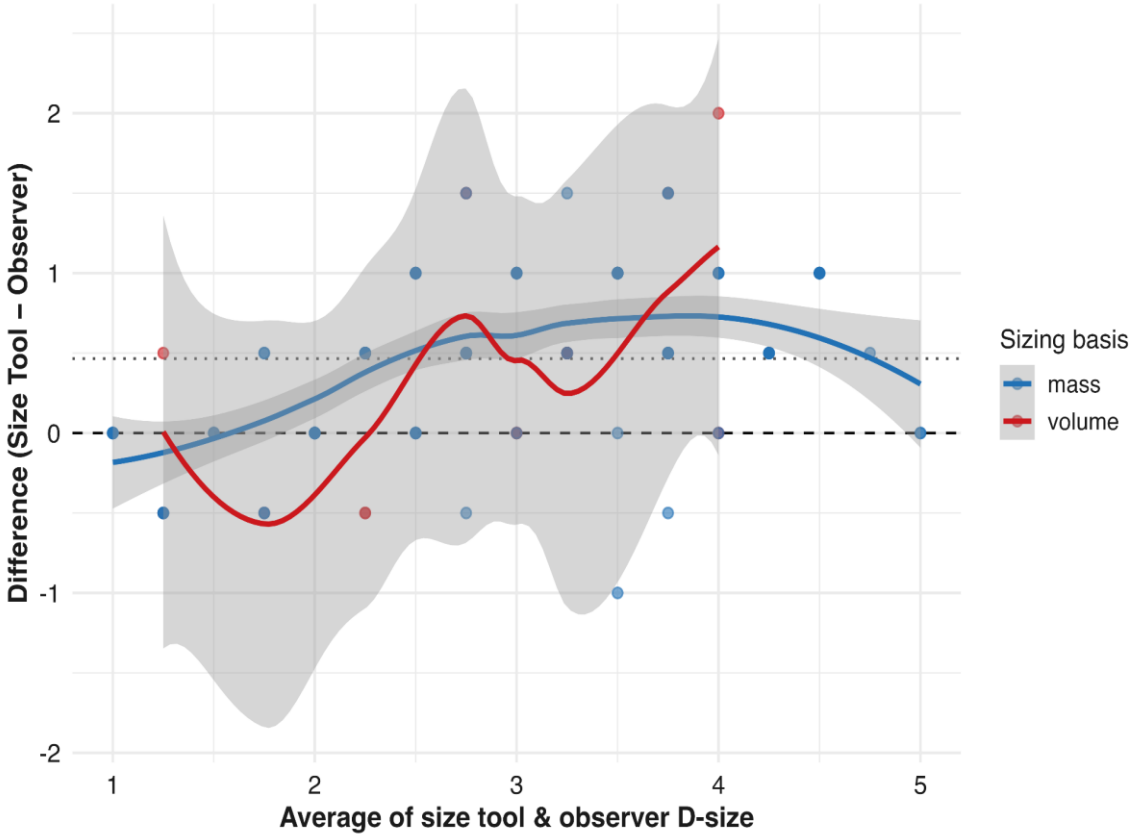


Figure 2: Bland-Altman comparison of tool-calculated and field-assessed D-size across n=143 avalanche events. The x-axis shows the average of the two estimates; the y-axis shows their difference (size tool minus observer). Point color indicates sizing basis (blue = mass-based start-zone events, red = volume-based runout events). The dashed horizontal line marks zero difference, and the dotted line marks the observed mean bias (0.47 D-size units) across all events. Shaded bands are loess-smoothed 95% confidence intervals, shown separately by sizing basis.

When examining the distribution of differences in size estimates between the size tool and observers, we found most of the differences to be one-half of a D-size unit for mass estimates (Figure 3). Additionally, in the few cases of volume estimates, the tool typically estimates one-half size smaller than observers' estimates.

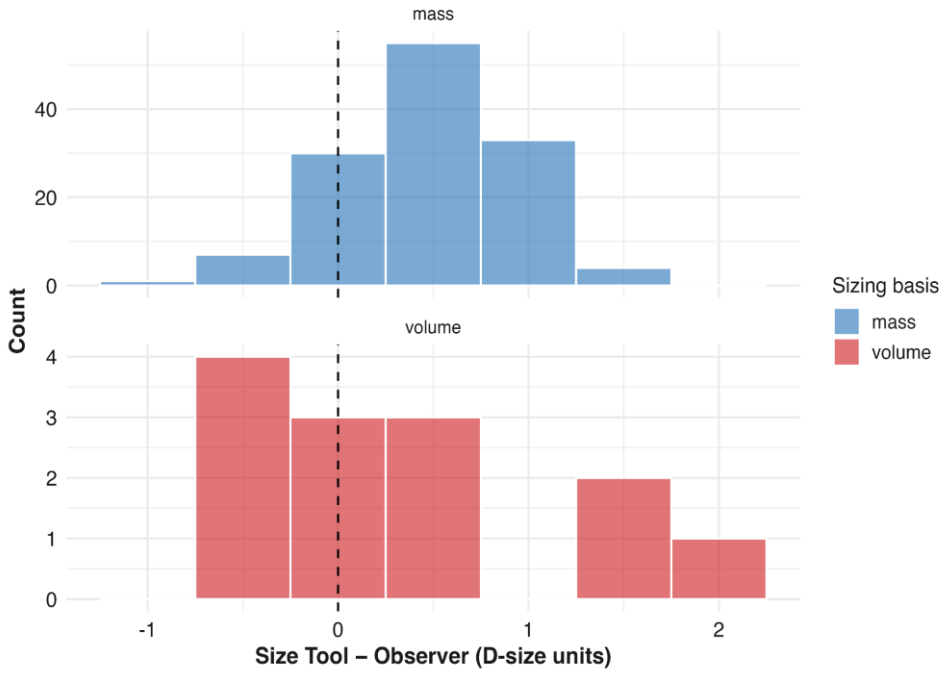


Figure 3: Distribution of tool-minus-field D-size differences for n = 143 independent avalanche events, classified by sizing basis. The top plot shows mass-based start-zone events (n=130) and the bottom plot shows volume-based runout events (n=13). The dashed vertical line marks zero difference.

4.4 Comparison with Physics-Based Simulations

We compared the size tool and observer estimates against physics-based AvaFrame com1DFA simulations for 8 individually investigated case studies with entrainment added in 5 of the events (Figure 4a). The sample size is too small for any formal statistical test, so we present these results as illustrative. In each of the 8 events, the size tool (both with and without entrainment) classified D-sizes as either a half or full size larger than observer estimates. We compared the size tool to com1DFA with entrainment (n=4 for both tool and simulation for-comparison) and without entrainment (n=8). With entrainment, the two methods estimated the same D-size class in 4 of 4 comparable events and 7 of 8 events without entrainment. When comparing com1DFA estimates to observer estimates, com1DFA estimated larger sizes than the observer in 7 of 8 cases without entrainment and all 4 cases with entrainment.

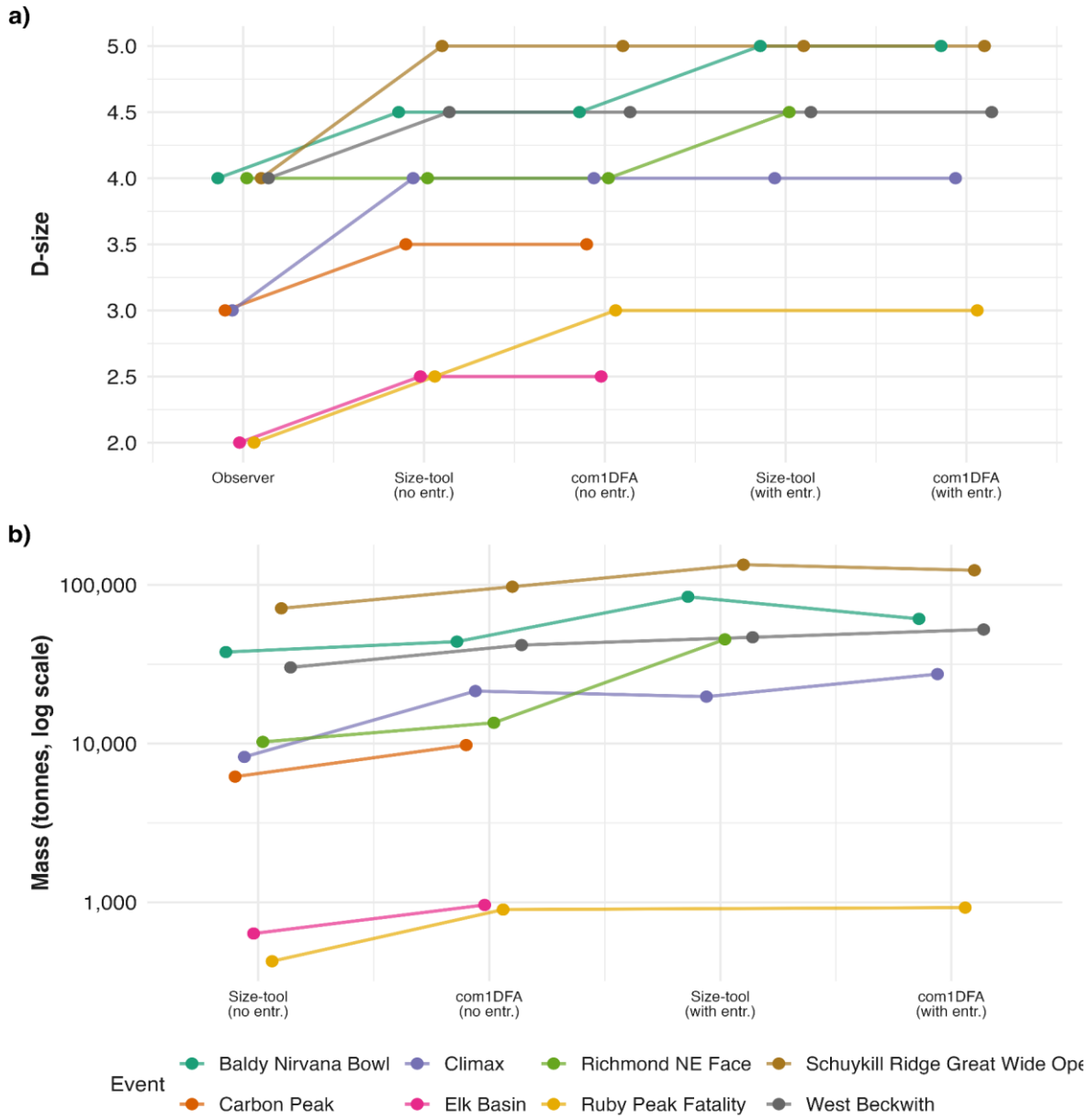


Figure 4: a) D-size estimate by method for n=8 individually investigated avalanche case studies: observer field assessment, Size-tool, and AvaFrame com1DFA simulation with and without snowpack entrainment. b) Estimated mass (tonnes, log scale) by method for n=8 avalanche case studies: Size-tool, and AvaFrame com1DFA simulation with and without snowpack entrainment. The observer's field assessment is a D-size class rather than a mass estimate and is not shown. Each line is one avalanche, tracked across methods within each panel; line color identifies the event and is consistent between panels. Points and lines are offset by a small, fixed horizontal amount per event (the same offset for that event in both panels and at every category) so that events sharing an identical D-size or mass value at a given category remain individually visible rather than overlapping into a single dot. Note that all 8 cases were run without entrainment, and 4 cases where entrainment was included in the analysis for each method.

5. DISCUSSION

5.1 Agreement and the Direction of Disagreement

The size tool reproduced the observer's D-size estimate within one full class in 95% of events and within a half class in 71%, indicating that a mass calculation built from routine field observations can reliably estimate the size class most of the time. This level of agreement is notable in light of the lack of the "ground truth" it is being measured against. Hafner et al. (2023) found that any two practitioners agree on a full D-size

class only 53% of the time, with an individual-estimate reliability of 0.72, which caps the accuracy any model can reach against single-observer labels at roughly 73% by chance alone. A tool that lands within a half class of a single observer in 71% of cases is therefore performing near the ceiling implied by inter-observer agreement itself, and generally above it at the full-class level.

The size tool estimated a larger size than the observer in 71% of cases, matched exactly in 23%, and was smaller in only 6% using mass estimates. This distribution was significantly skewed toward the size tool estimating a larger D-size than observers' estimates ($p<0.001$). This directional bias is the most challenging to interpret. We provide two potential explanations consistent with previous literature. First, observers systematically under-classify avalanches, especially when viewed from a distance or from a helicopter and especially at the largest sizes (Hafner et al., 2023; Carson et al., 2023). This level of disagreement matches how coarse and judgment-based the observer's destructive-force estimation is by design. Additionally, a potential scale mismatch exists when using destructive potential, which can include observer bias (Carson et al., (2023) compared to more objective mass estimates within the same destructive scale (Greene et al., 2022; OGRS, 2024). Second, the mass thresholds in the original D-scale may be set too low, so that a correctly measured mass corresponds to a higher class than an observer would assign from destructive potential descriptions alone. Jamieson et al. (2014) drew similar conclusions, suggesting "that the typical values of deposit mass [from McClung and Schaerer (1981)] for Size 4 and 5 avalanches could be increased." Both explanations move in the same direction and our data cannot fully separate them. However, neither explanation requires that the size tool itself be bias-adjusted in either direction.

5.2 Size Dependence

The bias varied throughout the D-size scale. Our results demonstrated increasing disagreement with avalanche magnitude (slope = 0.27, $p < .001$). This pattern aligned closely with Jamieson et al. (2014), who showed that the typical masses assigned to Size 4 and Size 5 in the McClung and Schaerer (1981) scale are roughly an order of magnitude lower than values from other sources. They attributed this discrepancy to explosive control at the Rogers Pass calibration site reducing deposit masses during the scale's original development. Our results agree with the notion that if the class-mass associations at the top of the scale are set too low, then a mass-based tool will place large avalanches progressively above the observers' class as size increases. Jamieson et al. (2014) suggested that the phrase "largest snow avalanches known" used to classify Size 5 often serves as a deterrent for classifying an avalanche as a Size 5. This potential deterrent and the tendency to underestimate large avalanches viewed at a distance combine to increase uncertainty in classifying Size 4 and 5 avalanches. This suggests that the widening gap in our dataset is not evidence that the size tool degrades at large sizes, but rather as further evidence for revisiting the mass values that anchor the upper end of the scale. This is also consistent with recent operational case studies (Grist et al., 2024). Furthermore, as avalanche magnitude increases, estimating the potential harm to objects such as trains, trucks, and villages becomes increasingly outside the scope of most practitioners' experience.

5.3 Simulated Comparisons

Our sample size using AvaFrame com1DFA was too small for formal testing, but it offered an independent check on whether the size tool's tendency to estimate larger than observers' estimates is an artifact of the tool itself. The com1DFA simulations' mass output and corresponding D-size also exceeded observer estimates in 7 of 8 cases without entrainment and in all 4 cases with entrainment incorporated. The size tool matched the com1DFA simulated corresponding D-size in 7 of 8 comparable cases (without entrainment) and 4 of 4 events (with entrainment). Two independent methods, one a simple field-input calculation and the other a dynamic, physics-based simulation, both land above the observers' estimates and consistently produced similar mass outputs. This suggests that the disagreement potentially reflects observer under-classification and low scale thresholds rather than a bias specific to the size tool. The agreement between the size tool and com1DFA, both with and without entrainment, also provides qualitative evidence of the size tool's estimates.

5.4 Implications, Limitations, and Potential Tool Improvements

When practitioners disagree on a size or when field conditions limit reliable observations, the size tool can provide an objective and reproducible estimate from measurable inputs. Even in instances when the size tool disagrees with a given observer, the tool's outputs are transparent and repeatable. This supports greater consistency within an operation and comparability across operations. The same properties make it

suitable as a calibration and training aid and as a common reference for resolving disagreement over borderline cases of size. The tool's systematic disagreement from observer estimates, especially at larger avalanche sizes, potentially points toward the known weaknesses of the scale's upper classes or observer inexperience with larger avalanches (Carson et al., 2023; Jamieson et al., 2014) rather than toward tool error.

The most obvious limitation is how the observer D-sizes used as reference are single-observer labels of unknown accuracy. Prior work shows these labels are themselves only moderately reliable (Carson et al., 2023; Hafner et al., 2023; Jamieson et al., 2014). Disagreement between the size tool and an observer cannot necessarily be construed as tool error. The dataset, drawn from a broad range of snow climates, is restricted to well-documented dry slab avalanches and includes a share of fatality and accident cases, which may bias the sample toward particular sizes and conditions. The volume subset (n = 13) and the physics comparison (n = 8) are small and should be read as qualitative comparisons only. Finally, the tool depends on estimated inputs, including SWE or hand-hardness-derived density, slab dimensions, and expert adjustments for wind loading and lee effect. Each of these inputs carries uncertainty that the root-sum-square treatment attempts to reduce but cannot eliminate.

Mass is central to avalanche magnitude, but the destructive force of an avalanche also depends on path length and shape, and the velocity and density of debris. Although our size tool offers an objective approach, using mass alone has limitations in some situations, and requires a thoughtful and holistic assessment (Jamieson et al., 2014; McClung and Schaerer, 1981).

Further work can be completed to improve the size tool and observer size estimates. The size-dependent bias and its consistency with prior findings suggest future research could focus on a formal recalibration of the mass values using entrainment that anchors the D-scale, particularly for Sizes 4 and 5. Second, future validation could be conducted against consensus or multi-observer estimates within an operation rather than from a single observer. Third, the comparison against avalanche simulations could be expanded, and the volume or runout pathway and entrainment estimation refined with larger samples. This could improve the divergence between the start-zone and runout method in our study.

6. CONCLUSION

In this study, we used a robust dataset of 143 avalanches to demonstrate that a simple, practitioner-focused tool can convert routine field observations of slab dimensions, density, and snow water equivalent into mass estimates and corresponding destructive size classes. This tool agrees with subjective-based observer-assigned D-sizes within one full class in 95% of 143 events and within a half class in 71%. The tool estimated a larger size than observers in 71% of cases when using mass estimates, and this gap widened significantly with avalanche magnitude (slope = 0.27, $p < 0.001$). Two independent methods, the size tool and physics-based model simulations, both exceeded observer estimates. This suggests observer under-classification and low mass thresholds, especially at the upper end of the scale, which is consistent with previous work and reinforces the case for revisiting the mass values. The tool supplements observational workflow with an objective, transparent, and reproducible estimate from measurable inputs. As such, it holds practical value for improving consistency within and across avalanche operations, for assisting in borderline size uncertainty, and as a training and calibration aid.

DISCLAIMER

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. AI-assisted coding tools [Claude Opus 4.8] were employed to generate code used in data processing and analysis. All code was reviewed, tested, and validated by the authors to ensure correctness and reproducibility.

ACKNOWLEDGEMENTS

We greatly appreciate the shared avalanche data from Ben VandenBos, Alex Marienthal, Matthias Welcher, Timothy Glassett, Frank Carus, Gabrielle Antonioli, Christian Jordan, Ryan Huck, Andy Moderow, Jeff Bristow, John Stimberis, Sean Zimmerman Wall, and Mike Cooperstein. We also thank Scott Thumlert and Mike Conlan for their insight and valuable discussions on the Destructive Scale. The U.S. Geological Survey Ecosystems Mission Area supported this work. We also thank Simon Trautman and Ethan Davis for reviewing earlier drafts of this paper.

REFERENCES

Bland, J.M., Altman, D.G., 1986. STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT. The Lancet, Originally published as Volume 1, Issue 8476 327, 307–310. https://doi.org/10.1016/S0140-6736(86)90837-8

Canadian Avalanche Association, 2024. Observation Guidelines and Recording Standards for Weather, Snowpack and Avalanches.

Carson, T., Larson, L., Martland, B., Jamieson, B., 2023. Why observers may not agree on the destructive potential size rating and a draft scale for more consistent ratings. In: Proceedings of the International Snow Science Workshop (Ed. Peitzsch, E., Marhsall, H.P., Savage, S., Trautman, S.) , Bend, OR, USA. October 8-13, 2023.

Greene, E., Birkeland, K.W., Elder, K., McCammon, I., Staples, M., Sharaf, D., Trautman, S., Wagner, W., 2022. Snow, weather, and avalanches: Observation guidelines for avalanche programs in the United States (4th ed), 4th ed. American Avalanche Association, Denver, CO.

Grist, M., 2023. A Large Avalanche in the Ministry of Transportation and Infrastructure's Jack MacDonald Path. The Avalanche Journal 131. https://doi.org/https://avalanchejournal.ca/could-it-be-a-size-5/

Hafner, E.D., Techel, F., Daudt, R.C., Wegner, J.D., Schindler, K., Bühler, Y., 2023. Avalanche size estimation and avalanche outline determination by experts: reliability and implications for practice. Natural Hazards and Earth System Sciences 23, 2895–2914. https://doi.org/10.5194/nhess-23-2895-2023

Jamieson, B., 2018. Planning Methods for Assessing and Mitigating Snow Avalanche Risk. (Contributions by Argue, C., Buhler, R., Campbell, C., Conlan, M., Gauthier, D., Gould, B., Jamieson, B., Johnson, G., Johnston, K., Jones, A., Jonsson, A., Sinickas, A., Statham, G., Stethem, C., Thumlert, S., Wilbur, C.). Canadian Avalanche Association, Revelstoke, British Columbia, Canada.

Jamieson, B., Beglinger, R., Wilson, D., 2014. Case study of a large snow avalanche in the Selkirk Mountains and reflections on the Canadian size classification. Proceedings of Geohazards 6, 15-18.

Kim, D., Jamieson, B., 2014. Estimating the density of dry snow layers from hardness and hardness from density, Proceeding of the 2014 International Snow Science Workshop, Banff, Alberta, Canada. September 29-October 3, 2014. p. 499-506.

McClung, D., Schaerer, P., 2006. The avalanche handbook. Mountaineers Books, Seattle, Wash.

McClung, D.M., Schaerer, P.A., 1981. Snow avalanche size classification. Canadian Avalanche Committee. Avalanche Workshop, 3-5 Nov. 1980. Associate Committee on Geotechnical Research. Technical Memorandum 133, National Research Council of Canada, 12-27.

Moner, I., Orgué, S., Gavaldà, J., Bacardit, M., 2013. How big is big: results of the avalanche size classification survey. Proceedings of the International Snow Science Workshop, Grenoble, France, October 7-11. p. 242-246.

Oesterle, F., Wirbel, A., Fischer, J.-T., Huber, A., Spannring, P., 2026. OpenNHM/AvaFrame: 2.2a1 (Version 2.2a1) [Computer Software]. Zenodo. https://doi.org/10.5281/zenodo.21339428.

Sovilla, B., Burlando, P., Bartelt, P., 2006. Field experiments and numerical modeling of mass entrainment in snow avalanches. Journal of Geophysical Research: Earth Surface 111. https://doi.org/10.1029/2005JF000391

Statham, G., Haegeli, P., Greene, E., Birkeland, K., Israelson, C., Tremper, B., Stethem, C., McMahon, B., White, B., Kelly, J., 2017. A conceptual model of avalanche hazard. Natural Hazards 90, 663–691. https://doi.org/10.1007/s11069-017-3070-5

Tonnel, M., Wirbel, A., Oesterle, F., Fischer, J.-T., 2023. AvaFrame com1DFA (v1.3): a thickness-integrated computational avalanche module – theory, numerics, and testing. Geoscientific Model Development 16, 7013–7035. https://doi.org/10.5194/gmd-16-7013-2023

U.S.D.A., 1961. Snow Avalanches. U.S.D.A. Handbook No. 194, U.S. Govt. Printing Office, Washington, D.C., 84 p.

Wilcoxon, F., 1945. Individual comparisons by ranking methods. Biometrics Bulletin 1, 80–83.