Available online at www.sciencedirect.com

Cold Regions Science and Technology 39 (2004) 205–218

cold regions science and technology

www.elsevier.com/locate/coldregions

Variations in snow strength and stability on uniform slopes

C. Landry[a,*], K. Birkeland[a,b], K. Hansen[a], J. Borkowski[c], R. Brown[d], R. Aspinall[a]

[a] *Department of Earth Sciences, Montana State University, P.O. Box 366, Silverton, CO 81433, USA*
[b] *U.S. Forest Service National Avalanche Center, USA*
[c] *Department of Mathematical Sciences, Montana State University, USA*
[d] *Department of Civil Engineering, Montana State University, USA*

Received 30 September 2002; accepted 30 November 2003

Abstract

This research investigated whether single snowpits could reliably represent snowpack strength and stability conditions throughout apparently ‘uniform’ slopes. Seven slopes were selected by experienced avalanche forecasters, three each in the Bridger and Madison Ranges of Southwest Montana (USA), and one in the Columbia Mountains near Rogers Pass, British Columbia (Canada). Teams performed 10 quantified loaded column tests in each of five snowpits within a 900 m^2 sampling site on each ‘uniform’ slope, measuring shear strength in a single weak layer. Collection of slab shear stress data enabled the calculation of a stability index, S_{QLCT}. Altogether, eleven trials were performed during 2000/2001 and 2001/2002, testing several weak layer types exhibiting a wide range of strengths. Weak layer strength and slab stress conditions varied widely across the sampling sites, with coefficients of variation in weak layer shear strength ranging from 10% to 50%, coefficients of variation in shear stress from 2% to 48%, and stability indices ranging from 1.8 to 5.7. Of the 54 snowpits completed, 10 pits were empirically rejected as unrepresentative of the stability index at their sampling sites. Of the remaining 44 statistically analyzed pits, 33 pits were statistically representative of their site-wide stability index, and the other 11 pits were found statistically unrepresentative of their site. All five snowpits at a site were statistically representative of their site-wide stability index in three of the eleven trials. The frequent inability of single pits to reliably represent stability on those eleven 900 m^2 sampling sites, located on apparently ‘uniform’ slopes, highlights the importance of improving our understanding of the processes affecting the variability of snowpack stability on any given day and the uncertainties associated with ‘point’ stability data.

Keywords: Avalanche forecasting; Stability; Extrapolation; Spatial variation

1. Introduction

Avalanche forecasting has been described as “… the prediction of current and future snow instability in space and time relative to a given triggering (deformation energy) level …” (McClung, 2002). It follows, then, that among the many objectives of a forecaster is to “… minimize the uncertainty about instability introduced by the temporal and spatial variability of the snow cover (including terrain influences) …” (McClung, 2002).

Avalanche forecasters seek a variety of data in order to minimize uncertainty regarding instability. Evidence of instability obtained from the observation

---

* Corresponding author. Tel.: +1-970-387-5080; fax: +1-970-387-5082.

E-mail address: clandry@snowstudies.org (C. Landry).

0165-232X/$ - see front matter © 2004 Published by Elsevier B.V.
doi:10.1016/j.coldregions.2003.12.003

of actual avalanches is considered relevant, 'low entropy' (unambiguous) information collected at the appropriate scale. Such data are most effective in reducing uncertainty and, as such, are given the highest weighting among the multiple and redundant types of data collected (LaChapelle, 1980). In the absence of actual avalanche observations, or to corroborate the evidence they present, field measurements of snowpack stability obtained from in-situ stability tests are also considered relevant Class I data (McClung and Schaerer, 1993). In-situ stability tests measure a critical load required to cause snowpack fracture in a small snowpack sample.

However, it is often unsafe to conduct in-situ stability tests within avalanche starting zones, particularly when conditions approach the threshold of avalanching (Föhn, 1987; CAA/NRCC, 1995). Further, it is infeasible to obtain stability test data from every starting zone of interest, given the magnitude of terrain that most forecasters evaluate (Armstrong, 1991). For those reasons, avalanche forecasters routinely perform stability tests at carefully selected and safe sites presumed to be 'representative' and predictive of snowpack conditions in nearby avalanche terrain, and then extrapolate their results (McClung and Schaerer, 1993; Fredston and Fessler, 1994). For instance, Chalmers and Jamieson (2001) correlate a stability index collected at a level study plot to skier triggered avalanching within 100 km. Skilled avalanche forecasters are clearly capable of reliably extrapolating sparse data over broad regions, probably because of their ability to effectively filter large amounts of diverse data (McClung, 2002).

The underlying premise of this stability-sampling practice is that if and/or when a 'representative' slope is a good proxy for snowpack and stability conditions in nearby avalanche terrain, then reliable extrapolation of stability test results is facilitated. According to a fundamental tenet of geography, the closer the location and characteristics of the stability test site are in space to the location and characteristics of the extrapolated point, the more successful extrapolation should be. It follows, then, that extrapolating stability test results from one location on an apparently 'uniform' slope (selected to minimize known sources of spatial variability on snowpack properties across the slope) to a nearby point on the same 'uniform' slope should be the most reliable form of stability extrapolation possible.

While actual avalanches present comparatively unambiguous stability information, stability test results may contain substantial informational ambiguity regarding strength/stress relationships caused by the unknown scale of spatial variations in snowpack characteristics within the stability-sampling site itself. Several studies have documented spatial variation in snowpack stability within actual avalanche terrain, wherein terrain and snowpack characteristics were known to vary (Bradley, 1970; Conway and Abrahamson, 1984; Föhn, 1989; Jamieson and Johnston, 1993a; Birkeland, 2001; Kronholm et al., 2002; Stewart and Jamieson, 2002).

Less attention has been given, however, to variation within study slopes specifically selected to maximize the chances of sampling a snowpack that is homogeneous throughout the sampling site. Our study investigated spatial variations in snowpack weak layer strength and stability indices across sampling sites specifically selected to minimize the effects of spatial variations in terrain, aspect, substrate, vegetation and wind on snowpack processes. We tested short-range extrapolations of weak layer strength and stability measurements across sampling sites in order to assess whether a set of 10 (quantified loaded column) stability tests from a single snowpit could reliably predict the stability index of an apparently uniform slope.

In addition to varying across space at a given moment in time, stability also changes through time at a given location in space. This study also measured temporal variations in stability in a series of three 900-m^2 trials conducted at three side-by-side locations on the same slope, over a period of 18 days. Those results, and their interpretation, are the subject of a companion article (Birkeland and Landry, 2002). The current article will focus on the spatial variations in snowpack strength and stability indices observed during this study.

2. Methods

2.1. Stability sampling design and site selection

We adopted a systematic sampling design for this study distributing five snowpits in a regular pattern across a 30 × 30 m, or 900 m^2 test site (Fig. 1). Sys-

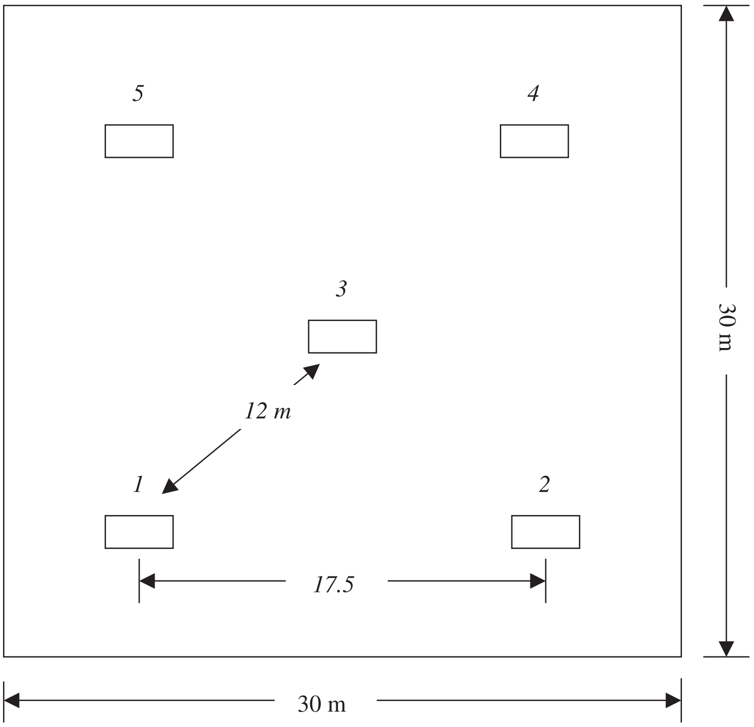

Fig. 1. Showing the 900 m^2 stability-sampling site pit layout. The five sampling snowpits are numbered and represented as rectangles (see Landry, 2002 for exact pit location coordinates and pit QLCT layout).

tematic sampling assured coverage throughout the site.

Seven 900 m^2 sampling sites were selected based on several desired attributes: no prior disturbance by skiers, snowmobiles, etc.; planar slope profile; no or minimal vegetation, besides grass; adequate distance from nearby trees to prevent shading, tree-drop 'bombs' and other vegetation effects; smooth substrate, without large scree or protruding bedrock; slope angle from 25° to 30°, for safety; protection from wind. Satisfying all of these conditions was difficult, and some sites suffered from more wind exposure than other, well-sheltered sites.

Nonetheless, in the opinion of the observers conducting the trials, all of whom were experienced avalanche forecasters, the selected sites seemed to be nominally 'uniform' slopes capable of exhibiting consistent snowpack characteristics throughout the 900 m^2 stability-sampling site. Three of the stability sampling sites were in the Bridger Range, northeast of Bozeman, Montana, three in the Madison Range, southwest of Bozeman, and the final site was at Rogers Pass, British Columbia, near Fidelity Station in Glacier National Park.

2.2. Measuring and calculating stability

Shear strength data for a single weak layer was collected at each site using the quantified loaded column stability test (QLCT) method (Landry et al., 2001) (Fig. 2). Ten QLCT were performed in each of the five snowpits at a site, using two rows of five, 50 cm-wide test cells, with the front of the second row of five cells 1 m uphill of the front of the first row. QLCT results were size-adjusted (Landry et al., 2001; Jamieson and Johnston, 2001; Föhn, 1987) to calculate shear strength τ_∞. Slab properties above the weak layer were also measured once at each pit in order to calculate the shear stress τ_{Slab} acting upon the weak layer at that pit location:

$$\tau_{\mathrm{Slab}} = \rho g h \sin\psi \tag{1}$$

Fig. 2. A surface mode QLCT being performed during the 27 Jan. 2001 stability-sampling trials at the Bradley Meadow site in the Bridger Range, MT. The observer is applying a rapid vertical load to a 0.08 m^2 isolated column of snow at the centroid of a plywood load plate using a mechanical force gauge (Landry et al., 2001).

where h represents the thickness of the slab (m), measured perpendicular to the slope, r is the density of the slab (kg/m^3), g is gravity, and ψ is slope angle. A QLCT strength/stress stability index, similar in principle but not strictly comparable to other stability ratios and indices (e.g., Roch, 1966; Föhn, 1987; Jamieson and Johnston, 1993b) due to differences between the QLCT and shear frame methods, was calculated:

$$S_{\mathrm{QLCT}} = \tau_{\infty}/\tau_{\mathrm{Slab}} \tag{2}$$

Mean and/or median stability ratio S_{QLCT} values were calculated for each pit and for each sampling site (by pooling all valid S_{QLCT} results).

2.3. Data analysis

Coefficients of variation for strength, stress, and the stability index (of the general form $\mathrm{CV}=s/\bar{x}$, where s represents the sample standard deviation and $\bar{x}$ the sample mean) were calculated within individual pits and across entire sites. Jamieson and Johnston (2001) found that 20 of 28 sets of shear frame measurements exhibited normal distributions of shear strength, and recommended the coefficient of variation as the best measure of variability in shear strength since the standard deviation of shear strength is known to increase as mean shear strength increases (Jamieson and Johnston, 2001). Some of our trials also exhibited normal distributions in the stability index, when the results from all five pits at the site were pooled (Birkeland et al., in press). However, the small number of strength and stability index measurements (nominally 10) at any individual pit did not allow us to conclusively show that our pit-wise results were normally distributed.

In order to compare data from pits to the pooled data collected at a site, we used the non-parametric Mann–Whitney test to evaluate the hypothesis of "no difference" between pit-wise and site-wise shear strength and stability index results. We pooled results from a single snowpit with the remaining four snowpits at a site to obtain site stability $\bar{S}_{\mathrm{QLCT(Site)}}$ or site shear strength $\bar{\tau}_{\infty(\mathrm{Site})}$. If a particularly strong/weak or stable/unstable pit were not pooled with the remaining four pits, site-wide variability in strength or stability would have been understated and made to appear more consistent than was actually measured. Thus, our analyses conservatively evaluated whether the strength and/or stability in any single snowpit within a site reliably represented site-wide strength and/or stability and, therefore, whether that study site represented a single strength (or stability) population.

3. Results

3.1. Stability-sampling trials

Altogether, eleven 900 m^2 sampling trials were performed over the course of the 2000/2001 and 2001/2002 winter seasons yielding data from 54 pits (Table 1). During the Round Hill trial the entire slope collapsed with a loud 'whumpf' during the preparation of the final pit (pit #5) and no data were obtained. Hence, we completed 54 total pits rather than a full set of 55.

In ten of the 11 trials, the weak layer tested was, in fact, the weakest weak layer within the snowpack. In the eleventh, at Baldy Mountain on 18 Feb 2001, a total of five 'weakest' weak layers were eventually identified during the course of the trial. To provide a margin of safety for the sampling teams, we attempted to avoid sampling sites known to be approaching a state of instability susceptible

Table 1
Stability-sampling trials summary. Weak layer types are: 'DH' = depth hoar; 'BF' = basal facets; 'NF' = new forms; 'SH' = surface hoar; 'NSF' = near-surface facets

Site (weak layer type)	Trial date	Site median stability index	CV site stability	Site mean strength (Pa)	CV site shear strength	CV site shear stress	Total depth (m)	Weak layer age (days)
Bacon Rind (DH)	4 Jan. 2001	1.85	31%	530	32%	4%	0.48	≅ 60
Bradley Mdw. (NF)	27 Jan. 2001	5.14	22%	590	23%	32%	1.24	7
Round Hill (SH)	4 Feb. 2001	2.05	44%	830	50%	10%	1.82	7
Baldy Mtn. (DH/BF)	18 Feb. 2001	nr	nr	1130[a]	26%[a]	28%[a]	1.11	≅ 75
Saddle Peak (DH/BF)	18 Feb. 2001	1.90	26%	1490	25%	16%	1.08	≅ 75
Bradley Mdw. (DH/BF)	18 Feb. 2001	nr	nr	1660[b]	21%[b]	48%[b]	2.16	≅ 75
Bradley Mdw. (NSF)	17 Mar. 2001	3.00	27%	430	27%	2%	1.15	≅ 5
Middle Basin (DH/BF)	7 Dec. 2001	2.07	24%	700	23%	12%	0.89	≅ 14
Lionhead Mtn. (SH)	9 Jan. 2002	2.50	10%	380	10%	4%	1.20	14
Lionhead Mtn. (SH)	15 Jan. 2002	3.08	11%	520	10%	4%	1.10	20
Lionhead Mtn. (SH)	26 Jan. 2002	2.38	16%	1080	18%	4%	1.51	31

'CV' indicates coefficient of variation. Total snowpack depth (HS) is given.

[a] Five different weak layers were revealed during the trial. Strength, stress and stability data above are for 19 valid QLCT results obtained from the targeted depth-hoar weak layer only.

[b] Results are for 20 valid QLCT results in the targeted depth-hoar weak layer; 30 tests exceeded the range of the QLCT equipment.

to the substantial disturbance these sampling sessions produced. Weak layer types tested included depth hoar and/or basal facets (five trials), near-surface facets (one trial), surface hoar (four trials), and precipitation particles (one trial) (Table 1). Weak layers ranged in age from 5 to 75 days, and snowpack depth ranged from 48 to 216 cm (Table 1). In most trials, shear strength was more variable, sometimes by an order of magnitude, than shear stress. However, in three trials—Bradley Meadow on 27 Jan. 2001, Baldy Mountain on 18 Feb. 2001, and Bradley Meadow on 18 Feb. 2001—shear stress was somewhat more variable than shear strength (Table 1). During two trials—Baldy Peak and Bradley Meadow on 18 Feb. 2001—inconsistent QLCT results from the targeted weak layer resulted in all pits being found empirically unrepresentative of their sites. Since those sites had been carefully selected to minimize known sources of spatial variability, and to optimize the chance of finding uniform snowpack characteristics throughout an apparently 'uniform' slope, we chose to include those empirically rejected pits in our analyses of their representativeness of site-wide strength and stability; we also report our results without including those two sites. Excluding those pits from our analysis would have understated the variability we observed during this research.

3.2. Variability in shear strength

Among the snowpits in our study for which a valid coefficient of variation in shear strength could be calculated $CV\tau_{\infty}$, ranged from 6% to 37%, with a mean of 17%. While the QLCT and shear frame methods do apply different types of stress to a measured weak layer, these values are similar to results reported by Jamieson and Johnston (2001). They had a mean coefficient of variation in shear strength of 18%, and a range from 4% to 54%, in 114 sets of shear frame tests conducted on slopes of at least 35°. The QLCT also is capable of detecting low variability in shear strength; at Lionhead Mountain we obtained $5.7\% \leq CV\tau_{\infty} \leq 6.2\%$ in six separate snowpits over two sampling trials. Further, during two side-by-side trials of the QLCT and shear frame conducted in buried surface hoar at Rogers Pass, British Columbia, in March 2000, coefficients of variation for the QLCT and shear frame were, respectively, 15% versus 13%, and 11% versus 12% (Landry et al., 2001).

3.3. Pit-to-site differences in shear strength

Of the 54 total pits performed, 10 pits (19%) were deemed unrepresentative of site-wide strength based on conclusive empirical evidence. Such evidence included QLCT results (or the lack thereof) which

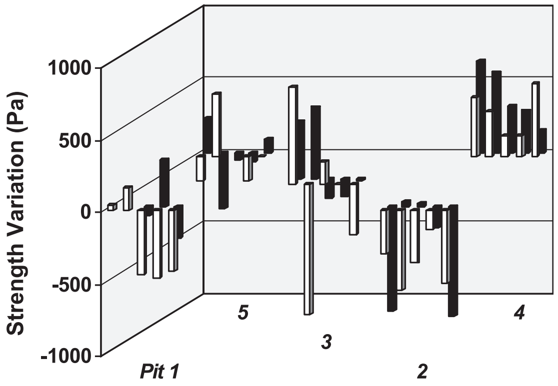

Fig. 3. Variations from median strength (1500 Pa), Saddle Peak trials of 18 Feb. 2001 (CV$\bar{\tau}_{\infty(\text{Site})}$ = 25%). White bars represent QLCT results from the first (front) row of five tests in a pit and black bars represent results from the second (back) row (N = 40).

could not be interpreted (Bradley Meadows, 18 Feb. 2001), or the presence of multiple 'weakest' weak layers (Baldy Mountain, 18 Feb. 2001). A statistically significant difference (a = 0.05) was found between pit-mean and site-wide strength at 14 (26%) of the remaining 44 pits. Thus, 30 (56%) of the 54 pits were statistically representative (percentages do not equal 100% due to rounding errors). Alternatively, if the rejected pits were not considered the proportion of representative pits increased to 68% (30 out of 44).

Only two trials (Bradley Meadow on 17 Mar. 2001, and Middle Basin on 7 Dec. 2001) yielded full sets of five pits that were statistically 'representative' of site-wide strength. The Bradley Meadows 17 Mar. 2001 trials tested a layer of near-surface facets formed by diurnal recrystallization (Birkeland, 1998) lying underneath a thin frozen-rain crust. The Middle Basin trial tested laminated layers of depth hoar and basal facets. One other trial produced four statistically 'representative' pits, the Lionhead Mountain trial of 9 Jan. 2002. At the other end of the spectrum, at Round Hill three of the four completed pits were found statistically unrepresentative of the site-wide shear strength $\bar{\tau}_{\infty(\text{Site})}$.

Charts of individual QLCT measurements of strength revealed several interesting patterns of spatial variation from site-wide strength (Figs. 3 and 4). These charts present an oblique view of all five pits at a site, with a key indicating relative pit locations (refer also to Fig. 1). Strength for each valid QLCT is shown as an individual 'bar' rising above or descending below the site-median strength 'surface'.

Three general types of pit-wise variation from site-wide shear strength were observed, and the Saddle Peak trials exemplified all three (Fig. 3). A weak layer of 1–2 mm faceted grains and mixed forms overlying a stronger layer of faceted forms was tested, with 'Q2' (average, mostly smooth) shear fractures (Johnson and Birkeland, 2002) occurring at the interface between the two faceted layers approximately 20 cm above-ground. Total snowpack depth was 108 cm. Pit 3 at Saddle Peak, the cluster at the center of the site, contained the strongest and the weakest individual QLCT results from the entire site. This typified many pits in which the extremes of above- and below-average site-wide shear strength were both present in the same pit. Interestingly, such strong and weak results were often obtained from adjoining tests, as seen in the front row of Pit 3. Pits 2 and 4 (lower right and upper right clusters) were typical of pits showing a consistent departure, either above or below, site-wide strength. And, finally, Pits 1 and 5 (lower and upper left clusters) were characteristic of pits exhibiting modest scatter about the site-median strength. Saddle Peak Pits 1, 3 and 5 were all statistically representative of site-wide shear strength.

A chart of the Bacon Rind trials graphically depicts the variability of strength observed there (Fig. 4). Consistent 'Q1' shear fractures (unusually

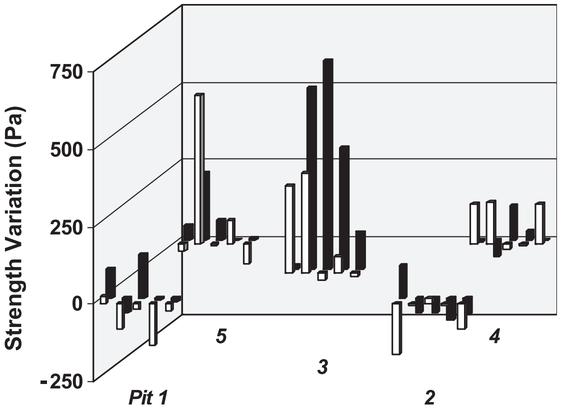

Fig. 4. Variations from median strength (470 Pa), Bacon Rind trials of 4 Jan. 2001 (CV$\bar{\tau}_{\infty(\text{Site})}$ = 32%). White bars represent QLCT results from the first (front) row of five tests in a pit and black bars represent results from the second (back) row (N = 50).

clean, planar, smooth) occurred at the top of a layer of basal depth hoar approximately 17 cm aboveground; total snowpack depth was 48 cm. Pit 3, the center cluster, contains mostly above-average strength results.

3.4. Pit-to-site differences in stability

When the variability of shear stress at a sampling site was low (i.e., $\mathrm{CV} < 10\%$), the spatial patterns of pit-to-site differences in stability closely resembled

Table 2
Mann–Whitney tests of pit-to-site stability ($\alpha = 0.05$), shown by site and pit number

Site	Trial date	Pits statistically representative of site stability	Pits statistically unrepresentative of site stability	Pits empirically unrepresentative of site stability
Bacon Rind	4 Jan. 2001	Pit 1 ($p=0.271$)	Pit 3 ($p=0.023$)	
		Pit 2 ($p=0.153$)		
		Pit 4 ($p=0.606$)		
		Pit 5 ($p=0.445$)		
Bradley Meadow	27 Jan. 2001	Pit 3 ($p=0.219$)	Pit 1 ($p=0.014$)	
		Pit 4 ($p=0.885$)	Pit 2 ($p=0.001$)	
			Pit 5 ($p=0.041$)	
Round Hill	4 Feb. 2001	Pit 3 ($p=0.125$)	Pit 1 ($p=0.002$)	
			Pit 2 ($p=0.024$)	
			Pit 4 ($p=0.013$)	
Baldy Mountain[a]	18 Feb. 2001			Pits 1–5
Saddle Peak	18 Feb. 2001	Pit 1 ($p=0.173$)	Pit 4 ($p=0.001$)	
		Pit 2 ($p=0.161$)	Pit 5 ($p=0.023$)	
		Pit 3 ($p=0.055$)		
Bradley Meadow[b]	18 Feb. 2001			Pits 1–5
Bradley Meadow	17 Mar. 2001	Pit 1 ($p=0.223$)		
		Pit 2 ($p=0.087$)		
		Pit 3 ($p=0.942$)		
		Pit 4 ($p=0.336$)		
		Pit 5 ($p=0.724$)		
Middle Basin[c]	7 Dec. 2001	Pit 1 ($p=0.565$)	Pit 2 ($p=0.039$)	
		Pit 3 ($p=0.147$)		
		Pit 4 ($p=0.953$)		
		Pit 5 ($p=0.921$)		
Lionhead Mountain	9 Jan. 2002	Pit 1 ($p=0.427$)		
		Pit 2 ($p=0.112$)		
		Pit 3 ($p=0.858$)		
		Pit 4 ($p=0.565$)		
		Pit 5 ($p=0.122$)		
Lionhead Mountain	15 Jan. 2002	Pit 1 ($p=0.294$)	Pit 3 ($p=0.048$)	
		Pit 2 ($p=0.104$)		
		Pit 4 ($p=0.241$)		
		Pit 5 ($p=0.178$)		
Lionhead Mountain	26 Jan. 2002	Pit 1 ($p=0.351$)		
		Pit 2 ($p=0.068$)		
		Pit 3 ($p=0.968$)		
		Pit 4 ($p=0.219$)		
		Pit 5 ($p=0.766$)		
Totals		33	11	10
		61%	20%	19%

[a] Five different weak layers were revealed during the trial.

[b] Only 20 valid QLCT results in the targeted weak layer; 30 tests exceeded the range of the QLCT equipment.

[c] ‘Representativeness’ results reflect estimation of 9 QLCT tests results where strength exceeded the range of the equipment.

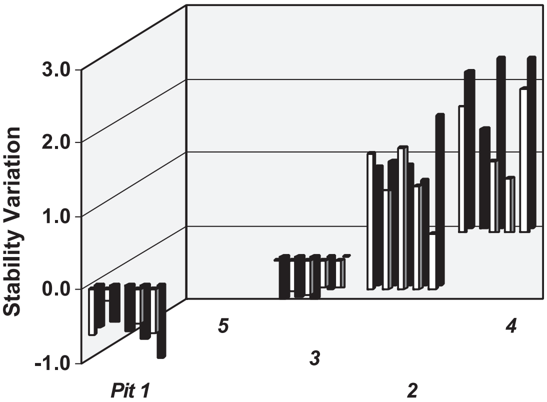

Fig. 5. Variations from median stability index (2.1), Round Hill trials of 4 Feb. 2001 ($CV\bar{S}_{QLCT(Site)} = 44\%$). White bars represent S_{QLCT} results from the first (front) row of five tests in a pit and black bars represent results from the second (back) row ($N = 37$).

pit-to-site differences in shear strength. But, where shear stress showed larger variations across a given site, patterns of pit-to-site differences in stability changed, as compared to pit-to-site differences in strength.

Of the 54 total pits sampled, 10 pits (19%) exhibited conclusive empirical evidence (described in 3.3 above) of 'unrepresentativeness' of site-wide stability. We found no statistically significant pit-to-site differences ($\alpha = 0.05$) in stability in 33 (61%) of the 54 total pits, given that the 10 empirically rejected pits are retained in the sample set of 54 pits (Table 2). Alternatively, if the rejected pits were not considered the proportion of representative pits increased to 75% (33 out of 44). Three of the 11 trials produced full sets of five pits exhibiting no statistically significant difference in pit-to-site stability (Bradley Meadow on 17 Mar. 2001, Middle Basin, and Lionhead Mountain on 26 Jan. 2002).

A chart of stability indices for the Round Hill trials presents an example of large variations in the stability index between halves of the site, despite generally consistent results within individual pits, and a trend across the site (Fig. 5). This was the only site where we observed such a pattern. This trial was also unique in that two separate observer teams, both highly trained and experienced, collected QLCT data. A layer of 4–6 mm buried surface hoar located approximately 125 cm above-ground yielded consistent 'Q1' shears. Total snowpack depth was 182 cm. During preparation of Pit 5, the final pit performed, the entire slope and test site collapsed thereby precluding valid results from Pit 5. Hence, no cluster of results is shown in the upper-left corner of the chart.

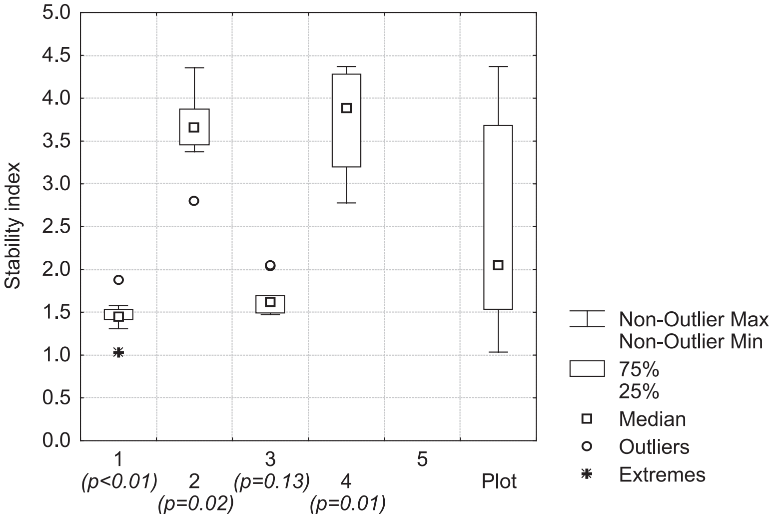

Fig. 6. Box-plot of stability indices, Round Hill trials of 4 Feb. 2001. Pits 1–4 results are shown with Mann–Whitney p-values for individual pits in parentheses below, and site-wide (plot) stability is shown on the far right. Only Pit 3 was statistically representative ($\alpha = 0.05$) of the site-wide stability index. Outliers are greater than 1.5 times the interquartile range away from the ends of the box, and the extreme values are greater than three times the interquartile range away from the ends of the box.

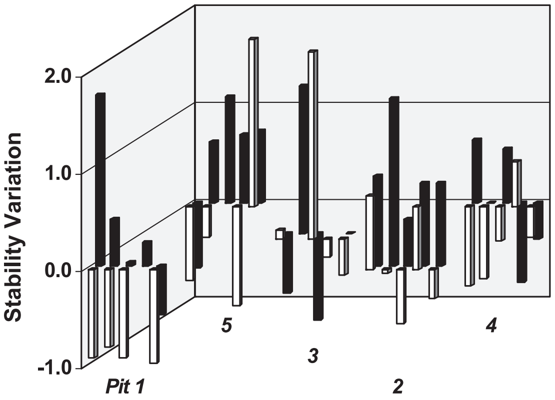

Fig. 7. Variations from median stability index (3.0), Bradley Meadow trials of 17 Mar. 2001 (CV$\bar{S}_{\mathrm{QLCT(Site)}}=27\%$). White bars represent S_{QLCT} results from the first (front) row of five tests in a pit and black bars represent results from the second (back) row ($N=46$).

Both Pits 1 and 3 produced similar stability indices while Pits 2 and 4 were also similar but substantially higher than those of Pits 1 and 3. Our analysis found Pit 3 statistically representative of the site-wide stability index (Table 2). Box-plots of stability results at individual pits at Round Hill reveal why only Pit 3 was statistically representative of site-wide stability (Fig. 6) and the *p*-values also indicate how unrepresentative Pits 1, 2 and 4 were (Table 2 and Fig. 6). At Round Hill, shear stress was generally consistent (CV$\tau_{\mathrm{Slab}}=10\%$), but strength was clearly not consistent throughout the site, with a standard deviation of 420 Pa from a site-wide strength of 830 Pa (CV$\bar{\tau}_{\infty}=50\%$).

In contrast to Round Hill, a chart of stability indices for the Bradley Meadow trials of 17 Mar. 2001 exhibits more scatter about site-wide stability within the pits (Fig. 7). Here, a layer of < 0.5 mm near-surface faceted grains located 95 cm above-ground and immediately below a 1–2 mm ice lens formed during a freezing rain event yielded consistent ‘Q1’ shears. Total snowpack depth was 115 cm. Shear stress was consistent throughout the Bradley Meadow site on 17 Mar. 2001, varying by only 5 Pa (CV$\tau_{\mathrm{Slab}}=2\%$), but shear strength was less consistent, with a standard deviation of 120 Pa from a site-wide strength of 430 Pa (CV$\bar{\tau}_{\infty}=27\%$). Variations in stability indices at Bradley Meadow on 17 Mar. 2001 (Fig. 7) therefore reflected variations in strength.

Nevertheless, our analyses found all five pits at Bradley Meadow on 17 Mar. 2001 statistically representative of site-wide stability, showing that in-pit variability did not preclude a pit from being representative of a site. Box-plots of the Bradley Meadow

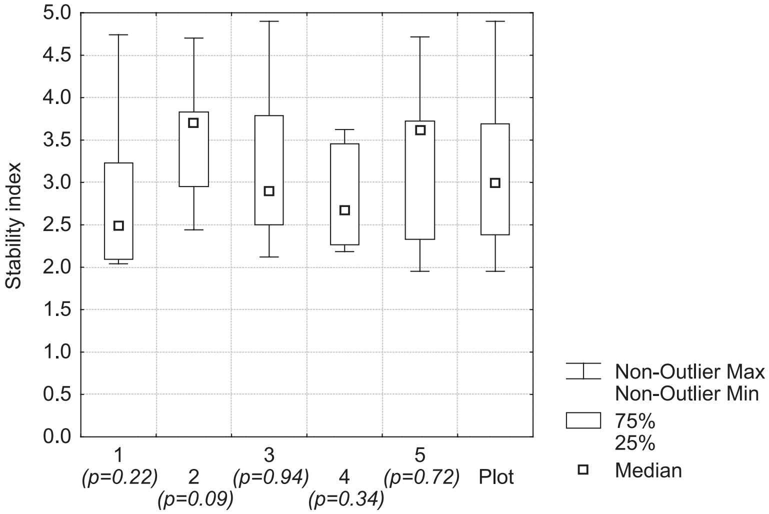

Fig. 8. Box-plot of stability indices, Bradley Meadow trials of 17 Mar. 2001. Pits 1–5 results are shown with Mann–Whitney *p*-values for individual pits in parentheses below, and site-wide (plot) stability is shown on the far right. All five pits were statistically representative ($\alpha=0.05$) of the site-wide stability index. Outliers are greater than 1.5 times the interquartile range away from the ends of the box, and the extreme values are greater than three times the interquartile range away from the ends of the box.

trials of 17 Mar. 2001 also reveal the variability in stability both within and among pits (Fig. 8), yet their *p*-values indicate the pits' representativeness of site-wide stability (Fig. 8 and Table 2).

4. Discussion

4.1. Potential sources of variability

In some trials, surprising variations in strength, rather than in shear stress, resulted in poor representation of site-wide stability by individual pits. For instance, our first trial at Bacon Rind involved an extremely simple snowpack consisting of a 20-cm thick layer of depth hoar covered by a 30-cm single-layer slab, and initially consistent shear strength and quality. Not until Pit 3, at the center of the site, did we discover what seemed to be 'anomalous' variations in strength, with the site's strongest snow thus far (Figs. 4 and 9). No apparent spatial variations in the substrate, vegetation, aspect, wind effects, or slope shape was observed that might have explained that variability at Bacon Rind at Pit 3.

Overall, we believe our site selection for these eleven stability-sampling trials was successful in minimizing the influence of variations in terrain and substrate on weak layer strength, at least to the extent that we understand the sensitivity of the snowpack over space and time to small differences in those variables (Birkeland and Landry, 2002). Basal weak layers, such as depth hoar, may be the most sensitive to small differences in geothermal heating or in the snowpack's substrate, even when the substrate appears uniform, while 'higher' weak layers forming at the snowpack surface, farther from the ground, may be less sensitive. However, our results at Round Hill, on 4 Feb. 2001, provide evidence that weak layers formed well above the ground surface can also exhibit considerable spatial variation in strength; at this site the targeted surface hoar layer was located approximately 125 cm above the ground.

Variations in the load on a weak layer produced by variations in the overlying slab within a site appeared to explain the variability in strength we observed during other trials. Wind-drift effects were presumed to cause those variations, particularly at the Bradley Meadow and Baldy Mountain sites during the 18 Feb. 2001 trials. Chalmers and Jamieson (2001) found evidence of increases in strength and stability in surface hoar associated with long-term increases in slab load, and Johnson and Jamieson (2000) made a

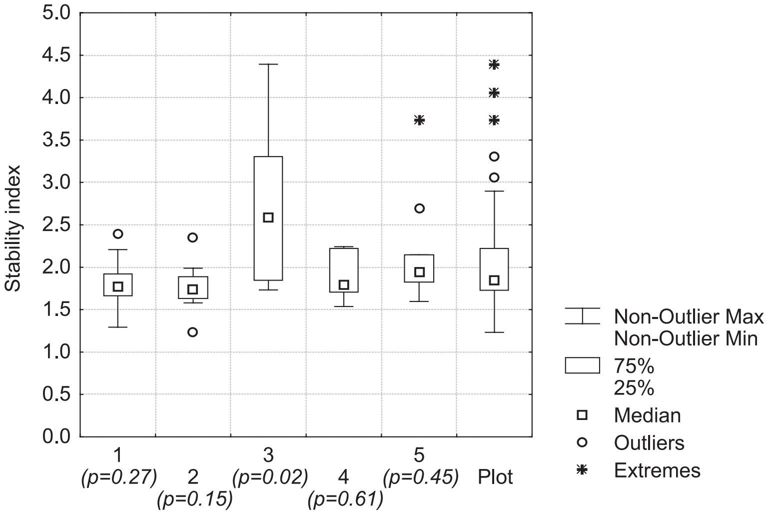

Fig. 9. Box-plot of stability indices, Bacon Rind trials of 4 Jan. 2001. Pits 1–5 results are shown with Mann–Whitney *p*-values for individual pits in parentheses below, and site-wide (plot) stability is shown on the far right. Only Pit 3 was statistically unrepresentative ($\alpha = 0.05$) of the site-wide stability index. Outliers are greater than 1.5 times the interquartile range away from the ends of the box, and the extreme values are greater than three times the interquartile range away from the ends of the box.

similar finding for faceted weak layers. Although those studies measured increases in strength and stability associated with increasing loads over time, the effect of spatial variations in the slab at a given moment in time could help explain spatial variations in weak layer strength and slope stability. However, large spatial variations in weak layer strength also occurred even when shear stress was effectively uniform across a site. For instance, shear stress across the Bacon Rind site varied only slightly, from 240 to 270 Pa, while site-wide shear strength ranged from 300 to 1140 Pa, and from 460 to 1140 Pa within a single pit (Pit 3).

Variation in our observers' QLCT technique, or produced by the QLCT procedure itself, might have offered yet another explanation for the variability in strength we observed. We compared the variability of our QLCT results to a study of variability in shear frame test results and, even though the QLCT method leaves all or a portion of the slab in place above the weak layer, the coefficients of variation in strength using the QLCT method closely resemble those obtained by Jamieson and Johnston (2001) using shear frames. This suggests that the QLCT method may be no more prone to operator-induced variations in test results than the shear frame method. Further, with six pits producing coefficients of variation in shear strength of 6% during the Lionhead trials of 9 Jan. 2002 and 15 Jan. 2002, at least some of which must be attributed to actual variations in snow strength, our results show that the QLCT is capable of detecting low levels of variability in comparatively weak layers. Using the QLCT method, we were also able to detect differences in shear strength at several spatial scales: between side-by-side tests (Pit 3 at Saddle Peak—Figs. 3 and 10), between pits (Pits 2 and 4 at Saddle Peak—Figs. 3 and 10), and within a site (Round Hill—Fig. 5). Therefore, we conclude that when performed by an expert, the QLCT method, like the shear frame, can reliably measure an index of shear strength without introducing overly problematic levels of 'background noise' into the test results (Fig. 10).

Finally, we explored the relationship between the age of a particular weak layer and its variability in strength and stability. Perhaps, the older the weak layer on our apparently uniform slopes had become, the more opportunity they had to experience and reflect spatially differing effects from variations in the overlying slab and from subtle variations in the underlying terrain and snowpack creep. The hypoth-

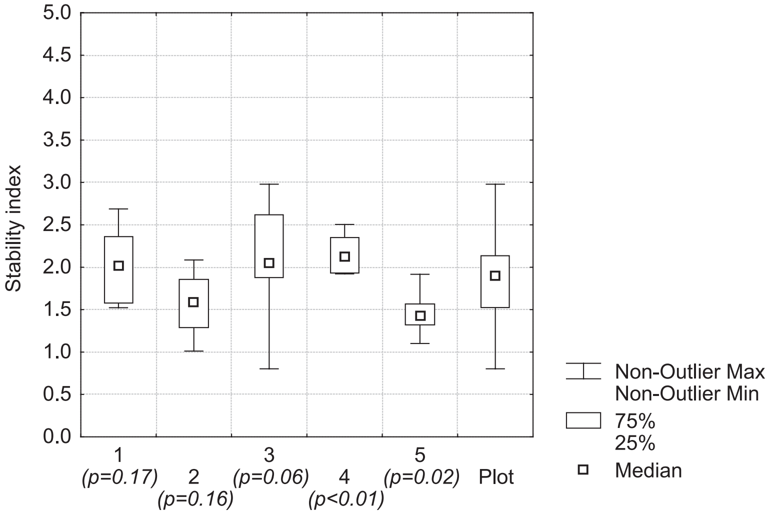

Fig. 10. Box-plot of stability indices, Saddle Peak trials of 18 Feb. 2001. Pits 1–5 results are shown with Mann–Whitney p-values for individual pits in parentheses below, and site-wide (plot) stability is shown on the far right. Only Pits 1, 2 and 3 were statistically representative ($\alpha = 0.05$) of the site-wide stability index. Outliers are greater than 1.5 times the interquartile range away from the ends of the box, and the extreme values are greater than three times the interquartile range away from the ends of the box.

esis that a young surface hoar weak layer should be less likely to exhibit variability in strength than another older layer was belied by our results from Round Hill, where buried surface hoar only 7 days old produced $CV\bar{\tau}_{\infty(\mathrm{Site})}=50\%$, perhaps reflecting a rapid response to variations in initial conditions in the slope or weak layer and resulting in the spatial trend seen in the results (Fig. 5). That result can be juxtaposed to the three Lionhead Mountain trials in older buried surface hoar, with values for $CV\bar{\tau}_{\infty(\mathrm{Site})}$ of 10%, 10%, and 18% at ages 14, 20 and 31 days, which do lend support to the hypothesis of increasing variability with increased weak layer age. Our results in depth hoar/basal facet weak layers do show a more consistent relationship between weak layer age and variability of strength. Since we have a limited amount of data for each weak layer type, and our results show contradictory patterns, we drew no conclusions regarding a relationship relating weak-layer age to variability in strength.

4.2. Stability tests as Class I data

Our findings may provide new insights regarding the principle that Class I stability test results are more easily interpreted than observations such as a snowpack profile (LaChapelle, 1980; McClung and Schaerer, 1993) and may suggest the need to re-visit how stability evaluation is taught to the general public and aspiring avalanche professionals. Experienced avalanche forecasters selectively and conservatively confer ‘reliability’ and ‘representativeness’ to their field observations of stability. Further, Class II (snowpack characteristics) information gleaned in the course of conducting stability tests, such as patterns in the snowpack stratigraphy (McCammon and Schweizer, 2002), or the mere presence/absence of a weak layer, or the quality of shear fracturing (Johnson and Birkeland, 2002), may also receive equal or greater weighting than Class I stability test results (Schweizer and Weisinger, 2001).

Since experience and expert knowledge are required to correctly interpret and appropriately apply stability test results, novice backcountry travelers may be inherently ill-equipped to interpret stability test results. Nonetheless, the concept of the ‘representative location’ for snowpits and stability tests is described for, taught to, and commonly adopted by inexperienced backcountry travelers, as well as aspiring avalanche professionals and mountain guides (McClung and Schaerer, 1993; Fredston and Fessler, 1994; Tremper, 2001). Our results suggest the need for increased emphasis on and awareness regarding the conservative use of ‘representative’ sites and stability test results.

4.3. Representative slopes and extrapolation

While professional avalanche forecasters generally assume spatial variation in stability to be the norm in complex terrain, our study shows that problematic spatial variation in snowpack stability may also exist on apparently nominal, ‘uniform’ slopes. When stability parameter measurements and stability indices cannot be reliably extrapolated within an apparently uniform and presumably ‘representative’ sampling slope, uncertainty may not be reduced by extrapolating data obtained at the ‘representative slope’ to surrounding terrain.

On the other hand, the fact that three of our trials yielded five out of five pits representative of stability ($\alpha=0.05$) provides evidence that single pits on carefully selected, apparently ‘uniform’ slopes, and/or in time-tested study plots, sometimes do provide reliable measures of snowpack characteristics throughout the slope or plot. This result affirms the successful correlation of study plot stability indices to regional skier-triggered avalanching by Chalmers and Jamieson (2001) and other studies of extrapolation (Föhn, 1987). Further, with the complications arising from the dynamics of snow on slopes, more consistent measurements may be possible in level areas than on slopes, as Jamieson and Johnston (1993b) found during their analyses of extrapolation of stability parameters.

Our results lend support to the notion that systematic and/or random sampling of a presumably representative study plot, or other apparently uniform slope, in pursuit of ‘mean’ slope stability information, may not always be as effective in reducing uncertainty as seeking worst-case, ‘instability’ data through ‘targeted sampling’ (McClung, 2002). The ongoing challenge for avalanche forecasters is to learn how to reliably predict those occasions when a single sample will reliably represent an extensive space and/or, alternatively, to learn how to objectively interpret the results of a single pit, inferring or deducing the spatial scale of variation in stability extant on that day, both within the sampled slope and beyond.

5. Conclusions

This research tested the hypothesis that a stability index obtained from a set of 10 stability tests performed at a single snowpit, which was located within a carefully selected and apparently uniform site, would reliably demonstrate a significant probability of predicting the stability index of the entire site. After 11 stability-sampling trials at seven different sites a conservative analysis or our results, retaining ten empirically unrepresentative pits in the sample set, found that 61% of our pits showed no statistically significant difference ($\alpha = 0.05$) between the pit stability index and the site-wide stability index whereas the other 39% of our pits were either statistically or empirically unrepresentative of their respective site's stability index. An alternative interpretation of our results, in which ten empirically unrepresentative pits were withdrawn from the sample set, showed that 75% of our pits were statistically representative ($\alpha = 0.05$) of their site-wide stability indices, while 25% were not.

Three of the eleven trials produced results in which all five pits were representative of their site-wide stability index. We did not detect so-called deficit zones (Conway and Abrahamson, 1984). It was not our purpose to establish a relationship between our stability index S_{QLCT} and actual avalanche activity on our sampling days (Föhn, 1987). Rather, we attempted to optimize the rigorous extrapolation of a strength/stress stability index across reasonably uniform slopes. Since our measurements showed, at best, a one-in-four chance of misrepresenting a slope, we concluded that a single pit on an apparently uniform slope was not shown to be a statistically reliable predictor of sampling-site stability parameters such as weak layer strength during our trials.

Combining stability test results with other observations, such as snow stratigraphy and shear quality, might reduce uncertainty. Irrespective of our results, experienced avalanche forecasters can and do collect stability data, and avalanche forecasts are largely reliable and useful, in spite of the uncertainty of stability test results (McClung, 2002). In part, this is because human avalanche forecasters utilize a great deal of sometimes redundant data to reduce their uncertainty (LaChapelle, 1980).

Additional research is clearly needed to explain why apparently uniform slopes sometimes do and other times do not exhibit uniform stability that sometimes can and other times cannot be reliably sampled with a single set of stability tests. A deeper understanding of the complex processes leading to those conditions, and how they change scales through time, would contribute to reducing uncertainty regarding the spatial and temporal variability of snowpack stability (Birkeland and Landry, 2002).

Acknowledgements

We gratefully acknowledge support of this research by the Montana State University Department of Earth Sciences and the Barry C. Bishop Scholarship for Mountain Research, National Science Foundation and EPSCoR MONTS, American Avalanche Association, Canadian Avalanche Association, Mazamas, the American Alpine Club, and the Geological Society of America Foundation's John Montagne Fund. Bridger Bowl Ski Area, the Parks Canada avalanche control program at Glacier National Park, Life Link International, and Snowmetrics contributed access to research sites and equipment. Many thanks to our Assistant Editor Jürg Schweizer, and to two anonymous reviewers. Their efforts led to substantial improvements and refinements in the article. Frequent discussions with Kalle Kronholm, Christine Pielmeier, Pascal Hägeli, and Jeff Deems also enhanced our understanding of these results. Special thanks also go to Jeff Deems and Ron Johnson for leading data collection teams, and to Chuck Lindsay, Doug Chabot, Stuart Dominick, Lance Riek, Chas Day, Aleph Johnston-Bloom, Zack Matthews, Mark Schaffer, Jim Rasmussen, John Kelly, Johann Schleiss, Dan Miller, Michael Cooperstein, Jeanette Romig, Blase Reardon, and the late Reid Sanders for their good work and cheerful company despite always long days and sometimes miserable conditions.

References

Armstrong, R.L., 1991. Snow properties research and data management. Snow Science: Reflections on the past, Perspectives on the Future. Proceedings of the Alta symposium, The Center for Snow Science. Alta, Utah, pp. 1–9.

Birkeland, K.W., 1998. Terminology and predominant processes associated with the formation of weak layers of near-surface faceted crystals in the mountain snowpack. Arct. Alp. Res. 30 (2), 193–199.

Birkeland, K.W., 2001. Spatial patterns of snow stability throughout a small mountain range. J. Glaciol. 47 (157), 176–186.

Birkeland, K.W., Landry, C.C., 2002. Changes in spatial patterns of snow stability through time. Proceedings of the 2002 International Snow Science Workshop, Penticton, British Columbia, Canada. British Columbia Ministry of Transportation, Avalanche Program, Victoria, British Columbia, pp. 482–489.

Birkeland, K., Kronholm, K., Schneebeli, M., Pielmeier, C., in press. Changes in the hardness and shear strength of a buried surface hoar layer. Ann. Glaciol., 38.

Bradley, C., 1970. The location and timing of deep slab avalanches. J. Glaciol. 9 (56), 253–261.

CAA/NRCC, 1995. Weather observation sites and procedures. Appendix A, Observation Guidelines and Recording Standards for Weather, Snowpack and Avalanches. NRCC Technical Memorandum, vol. 132. Canadian Avalanche Association, Revelstoke, British Columbia, pp. 61–62.

Chalmers, T.S., Jamieson, J.B., 2001. Extrapolating the skier stability of buried surface hoar layers from study plot measurements. Cold Reg. Sci. Technol. 33 (2–3), 163–177.

Conway, H., Abrahamson, J., 1984. Snow stability index. J. Glaciol. 30 (106), 321–327.

Föhn, P.M.B., 1987. The stability index and various triggering mechanisms. Avalanche Formation, Movement and Effects. IAHS Publ., vol. 162, pp. 195–214.

Föhn, P.M.B., 1989. Snowcover stability tests and the areal variability of snow strength. Proceedings of the 1988 International Snow Science Workshop, Whistler, British Columbia, Canada. Canadian Avalanche Association, Revelstoke, British Columbia, pp. 262–273.

Fredston, J., Fessler, D., 1994. Snow Sense—a guide to evaluating snow avalanche hazard. Alaska Mountain Safety Center, Anchorage, Alaska, USA, p. 116.

Jamieson, J.B., Johnston, C.D., 1993a. Experience with rutschblocks. Proceedings of the 1992 International Snow Science Workshop, Breckenridge, Colorado, USA. Colorado Avalanche Information Center, Boulder, Colorado, pp. 150–159.

Jamieson, J.B., Johnston, C.D., 1993b. Shear frame stability parameters for large-scale avalanche forecasting. Ann. Glaciol. 18, 268–273.

Jamieson, B., Johnston, C.D., 2001. Evaluation of the shear frame test for weak snowpack layers. Ann. Glaciol. 32, 59–69.

Johnson, G., Jamieson, J.B., 2000. Strength changes of layers of faceted snow crystals in the Columbia and Rocky Mountain snowpack climates in southwestern Canada. Proceedings of the 2000 International Snow Science Workshop, Big Sky, Montana, USA. American Avalanche Association, Pagosa Springs, Colorado, pp. 86–93.

Johnson, G., Jamieson, J.B., 2000. Strength changes of layers of faceted snow crystals in the Columbia and Rocky Mountain snowpack climates in southwestern Canada. Proceedings of the 2000 International Snow Science Workshop, Big Sky, Montana, USA, pp. 86–93.

Johnson, R.F., Birkeland, K.W., 2002. Integrating shear quality into stability test results. Proceedings of the 2002 International Snow Science Workshop, Penticton, British Columbia, Canada. British Columbia Ministry of Transportation, Avalanche Program, Victoria, British Columbia, pp. 508–513.

Johnson, R.F., Birkeland, K.W., 2002. Integrating shear quality into stability test results. Proceedings of the 2002 International Snow Science Workshop, Penticton, British Columbia, Canada, pp. 508–513.

Kronholm, K., Schweizer, J., Schneebeli, M., 2002. Spatial variability of snow stability on small slopes. Proceedings of the 2002 International Snow Science Workshop, Penticton, British Columbia, Canada. British Columbia Ministry of Transportation, Avalanche Program, Victoria, British Columbia, pp. 549–554.

LaChapelle, E.R., 1980. The fundamental processes in conventional avalanche forecasting. J. Glaciol. 26 (94), 75–84.

Landry, C.C., 2002. Spatial variation in snow stability on uniform slopes: implications for extrapolation to surrounding terrain. MS thesis, Department of Earth Sciences, Montana State University, Bozeman, Montana, USA, p. 248.

Landry, C.C., Borkowski, J., Brown, R.L., 2001. Quantified loaded column stability test: mechanics, procedure, sample-size selection, and trials. Cold Reg. Sci. Technol. 33 (2–3), 103–121.

McCammon, I., Schweizer, J., 2002. A field method for identifying structural weaknesses in the snowpack. Proceedings of the 2002 International Snow Science Workshop, Penticton, British Columbia, Canada. British Columbia Ministry of Transportation, Avalanche Program, Victoria, British Columbia, pp. 477–481.

McClung, D.M., 2002. The elements of applied avalanche forecasting: Part I. The human issues. Nat. Hazards 25, 111–119.

McClung, D.M., Schaerer, P., 1993. The Avalanche Handbook. The Mountaineers, Seattle, Washington, USA, p. 272.

Roch, A., 1966. Les variations de la résistance de la neige. Proceedings of the International Symposium on Scientific Aspects of Snow and Ice Avalanches, 1965. IAHS Publ., vol. 69, pp. 86–99.

Schweizer, J., Weisinger, T., 2001. Snow profile interpretation for stability evaluation. Cold Reg. Sci. Technol. 33 (2–3), 179–188.

Stewart, K., Jamieson, B., 2002. Spatial variability of slab stability in avalanche starting zones. Proceedings of the 2002 International Snow Science Workshop, Penticton, British Columbia, Canada. British Columbia Ministry of Transportation, Avalanche Program, Victoria, British Columbia, pp. 544–548.

Stewart, K., Jamieson, B., 2002. Spatial variability of slab stability in avalanche starting zones. Proceedings of the 2002 International Snow Science Workshop, Penticton, British Columbia, Canada, pp. 544–548.

Tremper, B., 2001. Staying Alive in Avalanche Terrain. The Mountaineers, Seattle, Washington, p. 284.