SNOW SLOPE STABILITY EVALUATION USING CONCEPTS OF FRACTURE MECHANICS

Jürg Schweizer[1,] *, Ian McCammon[2] and J. Bruce Jamieson[3]

[1] WSL Swiss Federal Institute for Snow and Avalanche Research SLF, Davos, Switzerland
[2] SnowPit Technologies, Salt Lake City UT, U.S.A.
[3] Department of Civil Engineering, Department of Geophysics and Geology, University of Calgary, Calgary AB, Canada

ABSTRACT: Dry snow slab avalanche release is generally believed to proceed in three stages: 1) initiation of a local failure, 2) widespread propagation of that fracture beneath the slab, and 3) detachment of the slab from its margins. To date, most field stability tests primarily assess the strength of the weak layer and thus relate to the first stage of avalanche release – fracture initiation. But field methods that comprehensively evaluate the second stage – fracture propagation – have remained elusive. In this paper, we explore evidence that field estimates of stability can be improved by integrating three elements: test score, fracture character or release type, and a simple index of structural stability (the threshold sum or "lemon count" across the fracture interface). Using field data collected from skier triggered avalanches and skier tested slopes that did not release, we show that when these three elements fall into a critical range the accuracy of predicting the probability of a skier triggered avalanche is higher than when any one element is used alone. Further, we show through a qualitative analysis that these three elements fulfill, at least partially, the criteria for fracture propagation prior to avalanching. As with any field stability method that relies on localized snowpack data, the approach presented here is not intended to be used in isolation, but in conjunction with other measurements and observations that relate to the probability and consequences of avalanche release.

KEYWORDS: snow stability evaluation, avalanche formation, avalanche release, skier triggering, stability test, fracture

1. INTRODUCTION

Dry snow slab avalanche release is ultimately a fracture process involving catastrophic failure of the sloping snow cover. Seen on the slope scale, a snow slab loses its shear support and slides downslope. These kinds of shear (or mode II and III) fractures are not typical of homogeneous materials. Under mixed mode loading conditions cracks in brittle, isotropic, homogeneous materials grow in most cases by kinking in a direction such that the advancing tip is in mode I. Shear fractures under mixed mode loading conditions are more common in heterogeneous, composite or layered materials where the interface presents a low-toughness fracture path through joined solids. The competition between crack advance within the interface and kinking out of the interface depends on the relative toughness of the interface to that of the adjoining material (Hutchinson and Suo, 1992).

The snow cover is a layered structure; one layer originating from a snowfall is bonded to the layer below originating from the previous snowfall. The snowpack obviously offers opportunities for preferential interfacial crack growth, i.e. the crack is restricted to move within a plane. Stiffness changes – as indicated by hardness changes – across layer boundaries act as stress concentrators so that cracks will preferentially grow on, or near, the interface between two layers of dissimilar hardness.

If there is an avalanche prone structure (distinct differences in layering), failure initiation and fracture propagation are required for a snow slab to release. Failure initiation results from introducing a local failure, for example, when locally the strength of the weak layer or interface is overcome by the additional stress imparted by a skier or snow boarder while moving on the snow surface. Failure initiation is therefore related to strength. Fracture propagation occurs as the initial local failure spreads out below the slab. Fracture

* *Corresponding author address:* Jürg Schweizer, WSL Swiss Federal Institute for Snow and Avalanche Research SLF, Flüelastrasse 11, CH-7260 Davos Dorf, Switzerland;
tel: +41 81 4170164; fax: +41 81 4170110;
email: schweizer@slf.ch

mechanics tries to answer the question of how tolerant is the snow to that local flaw. The material property describing this flaw bearing capacity is called fracture toughness. In a material with a high fracture toughness, small cracks will generally not lead to catastrophic failure. We use fracture toughness in the general sense of resistance to propagation so that it also applies to the weak layer collapse model (Heierli, 2005; Heierli and Zaiser, 2006).

Snowpack observations for stability evaluation should ideally focus on the above mentioned essential elements (or ingredients): layering, fracture initiation (strength) and fracture propagation (toughness) (McCammon and Sharaf, 2005). Not surprisingly, this is and has long been the standard procedure. A snow stability test following a snow profile exactly does the job. The stability test results include the location of potential failure layers, test scores and fracture character (or release type or shear quality) (e.g., van Herwijnen and Jamieson, 2005). To evaluate whether the snowpack has the critical layering, structural instability indices based on threshold sums such as the lemons or yellow flags have recently been introduced (Jamieson and Schweizer, 2005; McCammon and Schweizer, 2002).

In this paper, we will relate the three variables (field observations) to skier triggering probability. The variables thought to be predictors of snow slope stability are threshold sum (corresponding to the release element layering), the RB score (corresponding to failure initiation) and the RB release type (corresponding to fracture propagation). Results will be discussed in terms of snow slope stability evaluation with a particular view on the fracture process. We will analyze a dataset of more than 500 snow profiles with adjacent stability tests from the Swiss Alps and the Columbia Mountains of western Canada.

2. DATA

Our data come from snow profiles completed with a stability test, usually a rutschblock test. Profiles were done on skier tested slopes (no avalanche released) or on slopes where a recent skier triggered slab avalanche occurred. For the sake of simplicity, we called these two categories "stable" (= non-skier triggered) and "unstable" (= skier triggered), although we are aware that a single snowpack observation may not represent slope stability, and that under so-called "stable" conditions similar slopes (in particular unsupported ones) may have had the potential to avalanche as a result of skier triggering. The stable dataset did include quite a number of profiles from slopes with poor stability, sometimes even whumpfs and shooting cracks were recorded on the same day as the so-called stable profile was taken. However, as the slope was skier tested and did not release an avalanche, these cases were classified as stable.

Overall, the dataset included 514 cases (228 from Canada, 286 from Switzerland) and was almost balanced in terms of skier triggered (259 cases) vs. non-skier triggered cases (255 cases). The data were collected during the last 17 winters (1988-1989 to 2005-2006). As the release type was recorded routinely only in the more recent winters (but occasionally since the early 1990s), there were only 184 cases that included RB score, RB release type and complete structural information (threshold sum).

3. METHODS

Standard methods were applied for snowpack observations (e.g., CAA, 2002; Greene, 2004). In most cases, rutschblock tests were performed along with a snow profile. Not only the rutschblock score but also the release type was recorded: "whole block", "part of the block", "edge only" (Schweizer, 2002). To take into account the structural instability (lemons, yellow flags), we calculated the threshold sum using six unweighted variables: difference in grain size, failure layer grain size, difference in hardness, failure layer hardness, failure layer grain type and slab thickness (or failure layer depth). We used the same threshold values (or critical ranges) as described by Schweizer et al. (2005).

To evaluate the performance of predictors, various categorical statistics scores were used (Wilks, 1995). The scores were described in Schweizer et al. (2005) with the exception of the threat score (also called critical success index) that measures the fraction of observed and/or forecast events that were correctly predicted. In the notation typically used in contingency tables (Schweizer et al., 2005) it is defined as: TS = hits / (misses + false alarms + hits).

When comparing variables from the stable/unstable datasets, the non-parametric Mann-Whitney *U*-Test was used. When comparing categorical variables such as release type, the data were cross-tabulated and a Yates' corrected Pearson χ^2 statistic was calculated. A level of significance $p = 0.05$ was chosen to decide

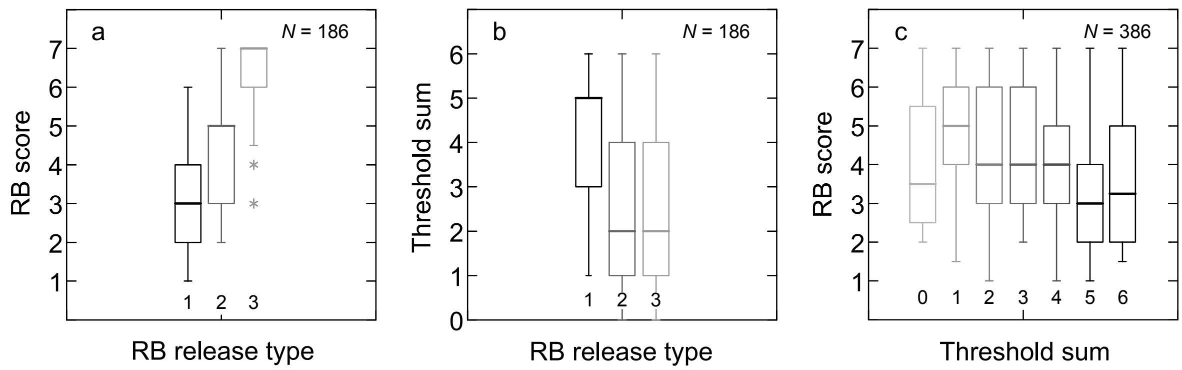

Figure 1: Distribution of (a) RB score and (b) threshold sum with RB release type (1: whole block, 2: part of the block, 3: edge only), and (c) distribution of RB score with threshold sum.

whether the observed differences were statistically significant. Split values between two categories were determined with the classification tree method (Breiman et al., 1998). To check for correlations between variables Spearman rank-order correlation coefficients were calculated.

4. RESULTS

The three predictor variables RB score, RB release type and threshold sum were all highly correlated ($p < 0.001$). For whole block releases, lower RB scores and higher threshold sums were found than for rutschblocks where only an edge was triggered (Figure 1). The RB score decreased with increasing threshold sum.

The RB score as well as the threshold sum slightly increased with increasing failure layer depth ($p < 0.001$ and $p = 0.009$, respectively). Slab thickness was lowest for the RB release type "part of the block" (median: 37 cm), and very similar for "whole block" (48 cm) and "edge only" (52 cm).

Figure 2 and Table 1 show how well the three predictors RB score, RB release type and threshold sum discriminate between stable and unstable cases. All three classifiers are highly significant variables (non parametric *U*-test, level of significance $p < 0.001$). Their classification accuracy varied between 71% for the RB release type to 66% for the threshold sum (Table 2). For the RB score, values < 4 indicated rather unstable, values ≥ 4 rather stable conditions. For the threshold sum, the critical range (rather unstable) included the values of 4, 5 and 6. However, when the threshold sum was 4, 48% of the cases in the database were rather stable, 52% were rather unstable. Only the release type "whole block" indicated rather unstable conditions, whereas the other two types were clearly more frequently found

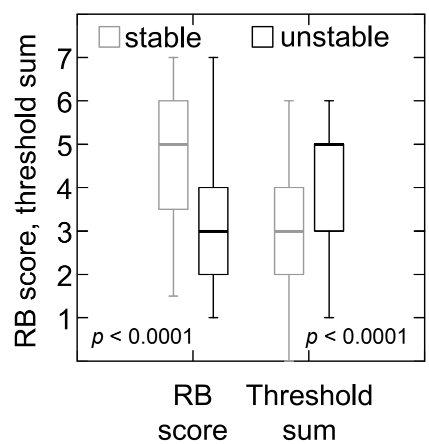

Figure 2: Distributions of RB score ($N = 459$) and threshold sum ($N = 428$) for the stable and unstable samples.

Table 1: Frequency of release type with snowpack stability ($N = 188$)

Release type	Snowpack		Total
	stable	unstable	
Whole block	42	51	93
Part of the block	50	9	59
Edge only	33	3	36
Total	125	63	188

Table 2: Univariate classification results

Variable or classifier	N (stable/ unstable)	Critical range or threshold	Accuracy (%)	Probability of detection (%)	False alarm ratio (%)	True skill score (%)	Threat score (%)	False-stable predictions (%)
RB score	459 (255/204)	< 4	68	61	35	35	46	30
RB release type	189 (125/63)	whole block	71	81	45	47	49	13
Threshold sum	416 (204/212)	≥ 4	66	74	36	33	53	31

with rather stable conditions. The RB release type was the best classifier in terms of the true skill score (47%). It had the highest probability of detection (81%) and the lowest portion of false-stable predictions (13%).

A multivariate analysis using the classification tree method included all three variables with the threshold sum as first node (< 5 stable), and the RB score and RB release type as second and third node to improve the classification of the rather unstable cases (Fig. 3). Figure 3 shows that a RB score < 4 was an indicator of rather unstable conditions only, if the block released as a whole. Low RB scores (≤ 3) with partial release were considered as less critical.

The overall classification accuracy was 83% and the true skill score was 51%, but due the unbalanced dataset (124 stable cases vs. only 60 unstable cases) the classification accuracy for the unstable cases was fairly poor. The probability of detection was 53%, hence 47% of the unstable

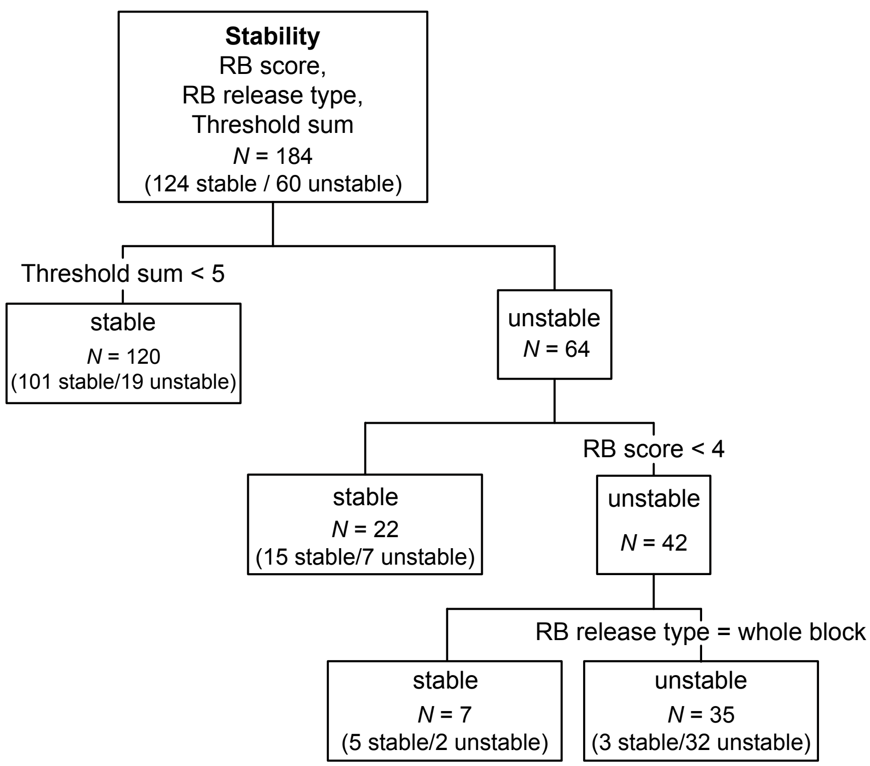

Figure 3: Classification tree (N = 184). Snow profiles were classified into rather stable or rather unstable cases based on RB score, RB release type and threshold sum. Classification accuracy was 83%.

Table 3: Frequency of stable and unstable cases depending on the value of the three variables RB score, RB release type and threshold sum. The values of 0 or 1 indicate whether the variables' values were in the critical range. Bold indicates a clear majority of either stable or unstable cases.

Variables			Total number in critical range	Observation	
RB score (< 4)	RB release type (whole block)	Threshold sum (≥ 5)		stable	unstable
0	0	0	0	59 (**97%**)	2 (3%)
0	0	1	1	6 (60%)	4 (40%)
0	1	0	1	15 (**79%**)	4 (21%)
1	0	0	1	12 (**80%**)	3 (20%)
1	0	1	2	5 (**71%**)	2 (29%)
0	1	1	2	9 (**75%**)	3 (25%)
1	1	0	2	15 (60%)	10 (40%)
1	1	1	3	3 (9%)	32 (**91%**)

cases were not recognized, and 19% of the cases predicted to be rather stable were in fact unstable cases.

As the dataset was unbalanced, we ran a series of ten classification tree analyses where we randomly chose about half of the stable cases for each analysis. This resulted in fairly different classification trees. On one hand, the number of splits varied (3 trees with one split, 6 trees with two splits and 1 tree with three splits. On the other hand, the variable at the first split varied and was in seven cases the RB score, in two cases the threshold sum, and in one case the RB release type. Overall, the threshold sum appeared 8 times, the RB score 6 times and the release type 3 times as a split variable. The split values consistently were < 4 for the RB score, ≥ 5 for the threshold sum, and "whole block" for the release type.

Instead of a classification tree analysis, a simple point score (or threshold sum) approach can also be applied for the three predictors RB score, RB release type and threshold sum. The results are shown in Table 3. Only if all three variables are in the critical range (bottom row of Table 3), the unstable cases clearly dominate. At the other end, zero or one variable in the critical range (top four rows in Table 3), the majority of the cases was clearly rather stable (88%). However, when the one variable that was in the critical range was the threshold sum, there was a substantial portion of unstable cases. With two out of the three variables in the critical range, the majority of the cases (66%) remained rather stable.

In the classification tree above (Fig. 3) only the bottom row in Table 3 was selected for the unstable category.

For a classification model with the bottom two rows indicating rather unstable conditions, the overall classification accuracy slightly decreased to 80%, but the true skill score increased to 56%, whereas the proportion of false-stable predictions decreased by about a quarter to 15% compared to the classification tree model (Fig. 3, Table 4).

Alternatively, as the RB release type seems to be the best single classifier (Table 2), the bottom three rows in Table 3 can be chosen to describe rather unstable conditions. This corresponds to the situation when the RB release

Table 4: Multivariate classification scores for various models

Model	Accuracy (%)	Probability of detection (%)	False alarm ratio (%)	True skill score (%)	Threat score (%)	False-stable predictions (%)
Classification tree (Fig. 3)	83	53	9	51	51	19
Bottom two rows (Table 3)	80	70	30	56	54	15
Bottom three rows (Table 3)	77	75	38	53	52	13
0 or 1 vs. 2 or 3	76	78	41	53	51	12

type and at least one of the other two predictors were in their critical range. The accuracy was 77% and the true skill score was about 53% (Table 4).

With an even more conservative approach that included all cases with either two or three variables in their critical range, the number of false-stable prediction slightly decreased (12%), but all other classification scores become worse except the hit rate (or POD) (Table 4).

5. DISCUSSION

We analyzed three snowpack observation variables: the RB score, the RB release type and the threshold sum (metadata) in regard to their classification power to distinguish between cases where a slope had been skier triggered and cases where it had not.

The RB release type proved to be the best single predictor of snow slope stability – it is also the simplest one. It can easily be observed and does not require special experience. This result is in agreement with, for example, findings of Schweizer et al. (2003) and an analysis by van Herwijnen and Jamieson (2006) who showed that the fracture character in a Compression test, which can be considered as the equivalent to the RB release type, related well to the probability of skier triggering. The release of the whole block seems to be a clear indication of instability. It has previously been proposed that the RB release type might be indicative of the fracture propagation propensity (Schweizer, 2002). The argument is based on theoretical and experimental results on the size of the local failure that is critical for rapid fracture propagation. Best estimates indicate that the critical size is on the order of about 0.1 m - 1 m, i.e. on the order of the slab thickness (Bažant et al., 2003; Schweizer et al., 2003). Therefore, fracture propagation has to occur in order for the whole area (3 m^2) of a rutschblock to fracture. In cases where the rutschblock released only below the skis, i.e. the fracture did not propagate uphill, the slab thickness was below average (37 cm) indicating that this release type is more frequently found with new snow instabilities (low RB score but a shallow and not yet well-consolidated slab). The failure often occurs by merely pushing the snow away below the skis rather than initiating a propagating shear fracture.

The character (roughness) of the fracture surface (clean vs. rough/irregular) that is considered as a measure of energy dissipation during fracture propagation, was not related to snow slope stability. It did not discriminate between rather stable/unstable cases ($N = 219$, $p = 0.08$; true skill score: 10%). Although rough or irregular interfaces are more common with rather stable conditions, most fractures were clean for both stable and unstable cases.

The RB score and the threshold sum were only slightly weaker predictors than the RB release type. The RB score had a relatively low probability of detection (or hit rate), i.e. a substantial number of blocks with low score were observed on slopes where no slab was released. This result might be due to targeted sampling (i.e. seeking instability) (McClung, 2002). In fact, a large number of these so-called stable profiles were rated as poor (Schweizer and Wiesinger, 2001) and occasionally even whumpfs and shooting cracks were observed on the day of observation. Accordingly, we must assume there is some uncertainty in their classification as stable (i.e. non-skier triggered) cases.

The RB score as well as the threshold sum had an intermediate range (RB score: 4, Threshold sum: 4) where neither stable nor unstable cases were clearly dominating. This ambiguity can be accommodated in practice by introducing an intermediate range of extra caution at these values.

It has been previously shown that both the RB score (e.g., Jamieson, 1995) and the threshold sum (Schweizer and Jamieson, 2006) are related to the probability of skier triggering. The rutschblock most closely integrates the fracture

Table 5: Properties of predictor variables in regard to fracture process and spatial variability

Predictor	Relevance to essential elements in the fracture process			Susceptibility to spatial variability
	Layering	Strength	Toughness	
RB score	high	high	low	high
RB release type	high	low	high	low
Threshold sum	high	moderate	moderate	low

process and can be seen as a small slab avalanche; it combines all essential elements (layering, fracture initiation and fracture propagation) and can be considered as class I data (McClung and Schaerer, 1993). In contrast, the threshold sum is secondary to the fracture process, and is more accurately considered class II data. However, the rutschblock test score is more affected by spatial variability than the other two predictors. Also, site selection is crucial and requires considerable expertise (targeted sampling). Table 5 compiles relevant properties of the three predictors.

The univariate classification results (Table 2) suggest that the classification power of the three variables was fairly similar, that they were correlated (and hence included redundant information) (see Figure 1) and that a combination of the three variables might therefore only slightly increase the classification accuracy.

However, combining the predictors improved the overall accuracy by 10-15% and the true skill score (which describes how well a predictor can differentiate between the categories) relatively increased by about 30-40% to values above 50%. Combining any two of the three variables resulted in only marginally lower scores. This means that the method is fairly robust and if one of the predictors is missing, e.g. if not observed, the probability of a correct interpretation is lessened but still possible.

6. CONCLUSIONS

In conclusion, the three predictors RB score, release type and threshold sum at sites selected by experts on or adjacent to avalanche slopes proved to be all highly significant variables in classifying cases as skier triggered or non-skier-triggered and hence are suggested to be indicative of snow slope stability. This result follows from the fact that these variables appear closely related to the three elements that are thought to influence the fracture process: layering, strength and weak layer toughness.

As has long been noted by practitioners, the most robust predictor of skier triggering is the rutschblock release type: a whole block release is a fairly unambiguous indication of instability. It is expected that similar results would be obtained with fracture character or shear quality. Nevertheless, combining the predictors of threshold sum and RB score with RB release type appears to provide a more robust estimate of stability, even in the presence of spatial variability.

Practical application of these results would proceed as follows:

- Rather stable conditions can be expected when none of the predictors is in its critical range (RB score: ≥ 4, RB release type: not whole block, threshold sum < 5). These conditions might roughly correspond to generally good stability.
- Intermediate conditions can be expected when one of the predictors is in its critical range. These conditions might roughly correspond to generally fair stability.
- Unstable conditions can be expected when at least two of the three predictors are in their critical range (RB score: < 4, RB release type: whole block, threshold sum ≥ 5). These conditions might roughly correspond to generally poor stability.

As always snow slope stability evaluation should never rely on a single snowpack observation. The above rating scheme is preliminary, and is intended as an aid towards more objective snow profile interpretation. Future work will be essential in characterizing the uncertainties inherent in such schemes.

ACKNOWLEDGEMENTS

We thank numerous colleagues and staff members for contributing snowpack data. For support, Bruce Jamieson is grateful to the Natural Sciences and Engineering Research Council of Canada, Helicat Canada, Canadian Avalanche Association, Mike Wiegele Helicopter Skiing, Canada West Ski Area Association and Parks Canada.

REFERENCES

Bažant, Z.P., Zi, G. and McClung, D., 2003. Size effect law and fracture mechanics of the triggering of dry snow slab avalanches. J. Geophys. Res., 108(B2): 2119, doi:10.1029/2002JB001884.

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J., 1998. Classification and regression trees. CRC Press, Boca Raton, U.S.A., 368 pp.

CAA, 2002. Observation guidelines and recording standards for weather, snowpack and avalanches. Canadian Avalanche

Association (CAA), Revelstoke BC, Canada, 78 pp.

Greene, E. (Editor), 2004. Snow, weather and avalanches: Observational guidelines for avalanche programs in the United States. American Avalanche Association (AAA), Pagosa Springs CO, U.S.A., 136 pp.

Heierli, J., 2005. Solitary fracture waves in metastable snow stratifications. J. Geoph. Res., 110(F2): F02008, doi: 10.1029/2004JF000178.

Heierli, J. and Zaiser, M., 2006. An analytical model for fracture nucleation in collapsible stratifications. Geophys. Res. Lett., 33: L06501, doi:10.1029/2005GL025311.

Hutchinson, J.W. and Suo, Z., 1992. Mixed-mode cracking in layered materials. In: J.W. Hutchinson and T.Y. Wu (Editors), Advances in Applied Mechanics, Vol. 29. Advances in Applied Mechanics. Academic Press, London, pp. 63-191.

Jamieson, J.B., 1995. Avalanche prediction for persistent snow slabs. Ph.D. Thesis, University of Calgary, Calgary AB, Canada, 258 pp.

Jamieson, J.B. and Schweizer, J., 2005. Using a checklist to assess manual snow profiles. Avalanche News, 72: 57-61.

McCammon, I. and Schweizer, J., 2002. A field method for identifying structural weaknesses in the snowpack. In: J.R. Stevens (Editor), Proceedings ISSW 2002. International Snow Science Workshop, Penticton BC, Canada, 29 September-4 October 2002. International Snow Science Workshop Canada Inc., BC Ministry of Transportation, Snow Avalanche Programs, Victoria BC, Canada, pp. 477-481.

McCammon, I. and Sharaf, D., 2005. Integrating strength, energy and structure into stability decisions: So you dig a pit and then what? The Avalanche Review, 23(3): 18-19.

McClung, D.M., 2002. The elements of applied forecasting - Part I: The human issues. Natural Hazards, 26(2): 111-129.

McClung, D.M. and Schaerer, P., 1993. The Avalanche Handbook. The Mountaineers, Seattle, Washington, U.S.A., 271 pp.

Schweizer, J., 2002. The Rutschblock test - Procedure and application in Switzerland. The Avalanche Review, 20(5): 1,14-15.

Schweizer, J., Fierz, C. and Jamieson, J.B., 2005. Assessing the probability of skier triggering from snow layer properties. In: K. Elder (Editor), Proceedings ISSW 2004. International Snow Science Workshop, Jackson Hole WY, U.S.A., 19-24 September 2004, pp. 192-198.

Schweizer, J. and Jamieson, J.B., 2006. A threshold sum approach to stability evaluation of manual snow profiles. Cold Reg. Sci. Technol.: in press.

Schweizer, J., Jamieson, J.B. and Schneebeli, M., 2003. Snow avalanche formation. Rev. Geophys., 41(4): 1016, doi:10.1029/2002RG000123.

Schweizer, J., Kronholm, K. and Wiesinger, T., 2003. Verification of regional snowpack stability and avalanche danger. Cold Reg. Sci. Technol., 37(3): 277-288.

Schweizer, J. and Wiesinger, T., 2001. Snow profile interpretation for stability evaluation. Cold Reg. Sci. Technol., 33(2-3): 179-188.

van Herwijnen, A. and Jamieson, J.B., 2005. Fracture character in compression tests. In: K. Elder (Editor), Proceedings ISSW 2004, International Snow Science Workshop, Jackson Hole WY, U.S.A., 19-24 September 2004, pp. 182-191.

van Herwijnen, A. and Jamieson, J.B., 2006. Fracture character in compression tests. Cold Reg. Sci. Technol.: in press.

Wilks, D.S., 1995. Statistical methods in the atmospheric sciences: an introduction. International Geophysics, 59. Academic Press, San Diego CA, U.S.A, 467 pp.