COMPARISON OF DIFFERENT SNOW STABILITY TESTS INCLUDING THE EXTENDED COLUMN TEST

Kurt Winkler* and Jürg Schweizer
WSL Swiss Federal Institute for Snow and Avalanche Research SLF, Davos, Switzerland

ABSTRACT: Several tests have been proposed in the past for evaluating snow stability. However, their performance is presently unclear since few comparative studies have been done. During winter 2007-2008 we have collected a dataset of 146 snow profiles, consisting of snow stratigraphy, a rutschblock test (RB), one to two extended column tests (ECT) and in most of the cases also one to two compression tests (CT). We studied whether the tests were able to predict stability. Study slopes were classified as rather unstable, when either signs of instability such as whumpfs or recent avalanche activity on nearby slopes were observed, or the profile was classified as poor or very poor. The CT had an almost perfect probability of detection, but as the structural stability index (threshold sum), the CT largely overestimated instability (high proportion of false alarms). Of the small scale tests the ECT was best suited to differentiate between stable and unstable situations. By including the ECT score (number of tabs), the number of false alarms was slightly reduced. The performance was similar to the RB which is, however, not independent of the stability classification we used. With two adjoining ECTs it was possible to classify 87% of our test slopes with an accuracy of about 90% in rather stable or rather unstable. Comparing two adjacent stability test results showed that only in about half of the pairs the same weak layer showed up as the most critical one. The ECT proved more difficult to perform than the RB, but was done faster than the RB. This advantage of the ECT contrasts with the lack of an intermediate stability level.

KEYWORDS: snow cover stability, snow stability evaluation, stability test, avalanche forecasting

1. INTRODUCTION

Avalanche forecasting relies on snow stability information. In the absence of signs of instability such as whumpfs, shooting cracks or avalanche activity, snow stability is assessed with the help of stability tests. However, there exists presently no stability test that is easy and fast to do and provides reliable stability information.

The two tests most widely used are the rutschblock test (e.g. Föhn, 1987) and the compression test (e.g. Jamieson, 1999). For both tests it has been shown that the score is related to skier triggered avalanche activity (e.g. Jamieson and Johnston, 1995), but also that the test score can be highly variable. The usefulness of both tests has been improved by noting the type of fracture: the fracture character (van Herwijnen and Jamieson, 2007) for the CT and the release type for the RB (area of the block that releases) (Schweizer and Wiesinger, 2001). Whereas the stability test score depends on the weak layer strength, and hence should be related to fracture initiation, the fracture type relates to fracture propagation propensity and depends, among other things, on the slab properties (Schweizer et al., 2008).

Based on weak layer properties structural instability indices (threshold sum) were developed (e.g. Schweizer and Jamieson, 2007), and popularized as lemons (e.g. McCammon and Schweizer, 2002) or yellows flags (e.g. Jamieson and Schweizer, 2005).

Gauthier and Jamieson (2008) proposed with the propagation saw test a real fracture mechanical beam test, and showed that the test results were related to fracture propagation propensity.

Simenhois and Birkeland (2006, 2007) proposed with the extended column test (ECT) a new test that should also provide information on the two processes of initiating and propagating a fracture. First results were very encouraging; they showed that the ECT was a good indicator of instability.

The aim of this study is to compare the ECT with other well-established tests (RB and CT) and assess its performance for the snow conditions of the Swiss Alps.

Corresponding author address: Kurt Winkler, WSL, Swiss Federal Institute for Snow and Avalanche Research SLF, Flüelastrasse 11, CH-7260 Davos Dorf, Switzerland;
tel: +41 81 4170127; fax: +41 81 4170110;
email: winkler@slf.ch

2. METHODS

2.1 *Observations and tests*

On each study slope we performed a full snow profile in conjuncture with a number of stability tests: a rutschblock test (RB), one to two extended column tests (ECT) and one to two compression tests (CT) (Figure 1). The different tests were arranged as close together as possible. The rearmost wall of all the columns and the RB were cut with a cord. Occasionally, observations were combined with snow micro-penetrometer measurements (SMP) (Pielmeier and Marshall (2008).

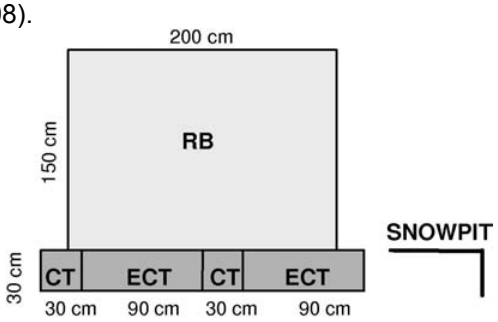

Figure 1: Set-up for slope observations. Either one or two ECT and CT were done adjacent in front of the RB.

The snow profile observations included grain type and size, hand hardness index, snow temperature, and in 68 % of the profiles also ram hardness, all corresponding to standard methods (e.g. CAA, 2002; Greene, 2004).

Based on observed snow stratigraphy, we calculated the threshold sum as indicator of structural instability using threshold values as described by Schweizer and Jamieson (2007). A threshold sum of ≤ 4 indicated rather stable, ≥ 5 rather unstable conditions.

For the rutschblock test (RB) the score and the release type were recorded. For the RB score, values ≤ 3 indicated rather unstable, higher values rather stable conditions. Only the release type "whole block" indicated unstable conditions (e.g. Schweizer et al., 2008). RB score and RB release type were combined with the threshold sum (Schweizer et al., 2008). If at least two of the three variables were in the critical range (RB score ≤ 3; RB release type "whole block"; threshold sum: 5 or 6) rather unstable conditions have to be expected.

For the compression test (CT), the number of taps (score) and the fracture character were recorded according to van Herwijnen and Jamieson (2007). A CT score ≤ 13 indicated rather unstable, > 13 rather stable conditions. "Sudden Collapse" (SC) and "Sudden Planar" (SP) were assumed to be related to unstable slopes, "Resistant Planar" (RP), "Progressive Compression" (PC), "Non-planar Break" (B) and no fracture were assumed to indicate rather stable conditions. As for the RB, we combined the CT test results with the threshold sum. Rather unstable conditions were assumed if all three variables were in the critical range (CT score ≤ 13; fracture character SC or SP; threshold sum 5 or 6).

The extended column test (ECT) was performed according to Simenhois and Birkeland (2006). The slope was assumed to be rather unstable when a fracture crossed the entire column in one layer and during the same or the next loading tap when the fracture initiated.

2.2 *Classification of study slopes*

Study slopes were classified as rather unstable, when at least one of the following three criteria was satisfied: 1) signs of instability such as whumpfs or cracks on the study slope. 2) recent (less then one day old) naturally or human triggered avalanches on nearby slopes. 3) the profile was classified as poor or very poor according to Schweizer and Wiesinger (2001). For analyzing the reproducibility of the stability tests and the most critical weak layer, we subdivided the class of rather stable slopes into either "fair" if the RB score was ≤ 3 or the RB release type was "whole block", and "good" otherwise.

2.3 *Data analysis*

For each observation on a slope the stability estimate was compared to the results of the various stability tests. For this purpose, only slabs in the range from 0.13 to 0.89 m were considered. Thinner or thicker slabs were assumed to be not critical, because they are less frequently triggered. For each individual test the result of the most critical weak layer was used. Thus, for a given test location different layers may have been considered as critical based on the different test results. The most critical failure layer was determined as described below. If finally, still multiple fractures remained in a single stability test, all of them were assumed to be critical.

- For the scores, the first fracture was assumed to be decisive. If multiple fractures occurred at the same tap, we choose in this order: 1. the fracture that initiated first. 2. the fracture with the more critical release type or fracture character.

- For the RB release type as well as for the CT fracture character, we selected the failure layer depth with the most critical fracture type. If a critical fracture type was observed at various depth, we selected the most critical one as follows: 1. the layer with the lower score. 2. the layer that was observed to fracture first.
- For the threshold sum, the layer with the highest score in the profile was considered as the critical one. In case of ties, all layers were considered to be critical.
- For the ECT, the layer where the fracture propagated furthest was assumed to be the critical failure layer. If more than one fracture crossed the entire column, the layer for which the number of taps between the beginning and reaching the rear end of the column was lowest, was selected. In case of ties, the critical layer was selected as follows: 1. the layer where the fracture started at the lowest score. 2. the layer that fractured first.

To determine whether the critical layers found in the various tests agreed, all the available stability tests were used. For each pair of tests, we checked whether the critical layer found in the first test, coincided with the critical layer found in the second test.

To describe the performance of the different tests, the following measures for categorical forecasts were used (e.g. Wilks, 1995). With the definitions used in contingency tables (Tab. 1), the measures are calculated as follows:
Probability to detect a stable slope:

$$\text{specificity (= 1-POFD)} = \frac{a}{a+c} \quad (1)$$

Probability to detect an unstable slope:

$$\text{sensitivity: POD} = \frac{d}{b+d} \quad (2)$$

$$\text{Overall accuracy or hit rate: PC} = \frac{a+d}{N} \quad (3)$$

Unweighted average accuracy

$$= 0.5\left(\frac{a}{a+c} + \frac{d}{b+d}\right) \quad (4)$$

Table 1: Contingency table. Total of samples: N=a+b+c+d

		Observed stability	
		Stable	Unstable
Result of the stability test	Stable	a	b
	Unstable	c	d

The overall accuracy measures the success of a model, but is not a good measure if the samples sizes (stable/unstable) are different. With only about 25% unstable observations (Tab. 2) our dataset is not balanced and the unweighted average accuracy is preferred. The probability of a false alarm is POFD and of false-stable predictions it is 1–POD. Although stability tests are only one factor considered in avalanche forecasting, false-stable predictions can have more serious consequences than false alarms.

If at a single snow pit location two ECT or CT were made, they were considered as not being independent of each other. As a mean value cannot be derived, we randomly selected one of the two tests for the statistical analysis. We then repeated this procedure nine times and calculated the mean.

To check for differences in the performance of the various tests we used the two-proportion Z-test (SYSTAT, 2007).

3. DATA

Data from 146 profiles were collected during winter 2007-2008 by researchers, forecasters and observers. All profiles were from the Alps, mainly from the Grisons region in Switzerland. The elevations at the profile site range from 1936 m to 3184 m a.s.l. with a median elevation of 2450 m a.s.l. Profiles were performed prevailingly on shady slopes (NW, N and NE) (Fig. 2) where more frequently poor snow stability can be found and a large part of the avalanche accidents occur.

The profile type was classified according to Schweizer and Wiesinger (2001) mostly based on the ram hardness (in 68% of the cases), otherwise based on the hand hardness. The dataset contained all different profile types; profile type 7 was found most frequently (Fig. 3). Snowpacks with consolidated basal layers were dominantly found (62%) during the winter 2007-2008, but snowpacks with weak basal layers were still well represented (38%).

Table 2: Dataset and proportion of unstable slopes

stability test	Number of samples	proportion unstable
RB	146	0.25
ECT	225	0.25
	(67 profiles with 1 test)	*(0.24)*
	(79 profiles with 2 tests)	*(0.25)*
CT	240	0.27
	(32 profiles with 1 test)	*(0.16)*
	(104 profiles with 2 tests)	*(0.29)*
Lemons	146	0.25

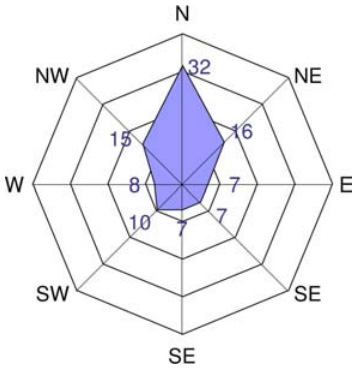

Figure 2: Frequency of the aspects [%] (*N* = 146)

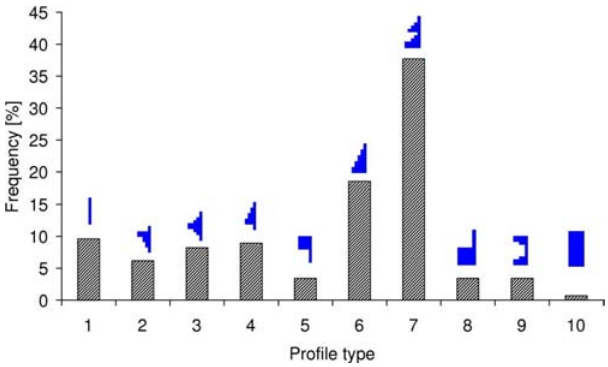

Figure 3: Frequency of the profile types (profiles reaching the ground only, N = 137).

4. RESULTS

4.1 Classification

Table 3 shows the performance of the different stability tests. Of all tests, with 90% (score) or 84% (release type), the RB had the highest probability to detect a stable slope (specificity, eq. 1) but also the highest rate of undesired false-stable predictions (1-sensitivity, eq. 2). Combining the score with the release type did not improve the unweighted average accuracy. If score *and* release type had critical values, the specificity increased to 99% and the sensitivity decreased. This very low false alarm rate was not attained by any other of the instability indicators ($p \leq 0.003$) but came at the price of a significantly reduced sensitivity ($p \leq 0.035$, except for the combination of CT and Lemons, for which the difference was not significant). With the RB score *or* the RB release type in the critical range, the probability to detect an unstable slope was large, and consequently the proportion of false-stable prediction was low (6%), but the specificity decreased and was significantly lower than for the RB score alone (p=0.003). With the combination of RB score, RB release type and threshold sum a good, balanced performance was reached with 14% false stables and 20% false alarms.

With 98%, the CT fracture character showed the highest sensitivity (=probability to detect unstable slopes) of all tests. The sensitivity was significantly better than with the RB (p=0.041); the difference to the ECT was marginally not significant (p=0.051). However, the specificity of the CT was low, i.e. there were more than 50% false alarms. This deficiency of the CT cannot easily be compensated.

Table 3: Classification results for the different stability indicators. Asterisk (*) indicate results for random selection if two tests were available; numbers in italic denote min and max values, respectively.

Stability Test/indicator	specificity (correct stables)	sensitivity (correct unstables)	unweighted average accuracy
RB score	0.90	0.78	0.84
RB release type	0.84	0.78	0.81
RB score *and* RB release type	0.99	0.61	0.80
RB score *or* RB release type	0.75	0.94	0.84
RB score, RB release type, threshold sum	0.80	0.86	0.83
ECT*	0.79 *(0.77 / 0.82)*	0.83 *(0.81 / 0.86)*	0.81 *(0.79 / 0.84)*
ECT (score ≤ 21)*	0.82 *(0.80 / 0.85)*	0.83 *(0.81 / 0.86)*	0.83 *(0.80 / 0.85)*
CT score*	0.45 *(0.40 / 0.50)*	0.91 *(0.89 / 0.94)*	0.68 *(0.64 / 0.72)*
CT fracture character*	0.22 *(0.20 / 0.26)*	0.98 *(0.97 / 1.00)*	0.60 *(0.60 / 0.61)*
CT score *and* CT fracture character)*	0.50 *(0.45 / 0.53)*	0.90 *(0.86 / 0.94)*	0.70 *(0.67 / 0.74)*
CT score *or* CT fracture character*	0.17 *(0.15 / 0.20)*	1.00 *(1.00 / 1.00)*	0.59 *(0.58 / 0.60)*
CT score, CT fracture character, threshold sum*	0.63 *(0.57 / 0.66)*	0.74 *(0.71 / 0.77)*	0.68 *(0.66 / 072)*
Threshold sum	0.38	0.86	0.62

With 86%, the threshold sum reached a high sensitivity and proved to be helpful in combination with the RB test. On its own, the threshold sum showed as the CT a poor specificity.

In our dataset, all the correct unstable predictions from the ECT occurred at ≤ 21 taps. All unstable predictions with taps > 21 were false-alarms. Thus we can enhance the specificity of the ECT without reducing the sensitivity by considering only the fractures occurring up to the 21st tap. However, this threshold might be specific to our dataset.

The ECT had neither the best probability of detecting stable nor unstable slopes, but showed balanced values for specificity (82%) and sensitivity (83%). The specificity was significantly better than with the CT, the threshold sum and all the combinations of these two indicators ($p \leq 0.007$). The unweighted average accuracy was similarly high for the ECT and the RB test and its combinations (including the threshold sum).

4.2 Reproducibility: test result

When stability tests are repeated side by side as in our set-up, similar test results should be expected (Table 4). In fact, for the ECT, in 87% of the cases the same stability class was found, i.e. twice stable or twice unstable. The reproducibility increased to 92% if only the test pairs were considered that were done on either a slope that was classified as "unstable" or "good". On the slopes classified as "fair" the reproducibility was with 72% significantly lower (p=0.048).

Compared to the ECT, the CT fracture character (85%) and the CT score have both a lower reproducibility (78%), but the differences were not significant. Both indicators showed a very high reproducibility (97%) on slopes rated as rather unstable. This follows from the high sensitivity and poor specificity of the CT.

When two ECTs are done close together, the stability classification can be improved by combining their results. If both tests indicated rather unstable conditions the slope was in fact rather unstable indicated by the high sensitivity of 90% in Table 5. Similarly, there was a high probability (90%) that the slope was rather stable if both ECT showed stable test results.

Thus in the 87% of the cases when the two tests had identical results (Table 5), the unweighted average accuracy was 90%. The remaining 13% included 2 cases from rather unstable slopes, and 8 cases from rather stables slopes (5 "fair"; 3 "good"). Theses slopes could not reliably be classified with the ECT. The frequency of these slopes depends on the stability distribution of the dataset analysed. Therefore, a final assessment on the performance of the ECT cannot be done unless a larger and more balanced dataset is available.

Table 4: Pair-wise reproducibility of the ECT, the CT score, the CT fracture character and the RB test. For the RB test no pairs were available; instead it was compared whether the RB score and the RB release type indicated the same level of stability. The asterisk (*) denotes an artefact resulting from the definition of "fair".

stability of the slope / result of the test pairs	ECT			CT score			CT fracture			RB		
	un-stable	stable		un-stable	stable		un-stable	stable		un-stable	stable	
		fair	good		fair	good		fair	good		fair	good
twice unstable	16	4	2	28	12	23	28	16	38	22	1	0*
once unstable, once stable	2	5	3	1	6	16	1	5	10	12	24	3
twice stable	2	9	36	0	3	15	0	0	6	2	2	80
total pairs		79			104			104			146	
same critical layer	16	14	22	26	13	27	25	12	25	35	27	83
different critical layers	5	4	19	6	8	28	7	9	30	1	0	0
total pairs		80			108			108			146	

Table 5: Classification results by using stability test pairs (ECT and CT) depending on the number of unstable results (both unstable, at least one unstable). For the RB, the RB score and release type were compared.

test	specificity	sensitivity	combining the two results: unweighted average accuracy	combining the two results: twice the same result
ECT				
both unstable	0.90	0.80	0.90	0.87
≥ 1 unstable	0.76	0.90		
CT score				
both unstable	0.53	0.97	0.77	0.78
≥ 1 unstable	0.24	1.00		
CT fracture character				
both unstable	0.28	0.97	0.64	0.85
≥ 1 unstable	0.08	1.00		
RB score, RB release type				
both unstable	0.99	0.61	0.97	0.73
≥ 1 unstable	0.75	0.94		

For the CT, combining the pair-wise test results did not improve classification results satisfactorily since the specificity remained low.

For comparison with the RB, we considered RB score and RB release type instead of two different tests. On all the slopes where both test results indicated the same stability class, the unweighted average accuracy was very high (97%). However, in the remaining 27% of the cases when the test results did not agree, the combination did not reduce uncertainty.

4.3 Reproducibility: critical failure layer

When different stability tests are performed close together as in our setup (Figure 1) we would expect that the same layer is the critical failure layer in the various tests.

Of the 80 cases considered, the critical failure layer as found with the first ECT of each pair, coincides in 52 cases (65%) with the critical failure layer found with the second ECT. In the 23 cases where both tests indicated rather unstable conditions, the critical failure layer agreed more often (83%) than in other cases 58% (p=0.036).

For the CT the agreement was slightly lower: 61% for the CT score, 57% for the fracture character. The latter is generally assumed to be less sensitive to spatial variations in snowpack properties. However, these differences were statistically not significant. As in the case of the ECT, the failure layer agreement was higher on unstable slopes (81% for the CT score; 78% for the CT fracture character) than on stable slopes (p≤0.005).

Table 6: Agreement of the critical failure layer: probability to find a critical failure layer identified by the first test, in the second test as well. Values in the upper line give the number of pairs and the probability. Values in brackets in the lower line are for tests performed on either rather unstable slopes or stable slopes that were rated as “fair”, again the number of pairs and the probability are given. Numbers are given in italic if the difference to stable slopes rated as "good" was statistically significant (p < 0.05).

first test \ second test	RB score	RB release type	CT score	CT fracture character	ECT	Threshold sum
RB score	147 (64)	147 / 0.99 (64 / 0.98)	254 / 0.43 (117 / *0.59*)	254 / 0.46 (117 / *0.60*)	226 / 0.51 (102 / *0.64*)	147 / 0.32 (64 / *0.41*)
RB (release type	147 / 0.99 (64 / 0.98)	147 (64)	254 / 0.42 (117 / *0.57*)	254 / 0.45 (117 / *0.58*)	226 / 0.51 (102 / *0.64*)	147 / 0.33 (64 / *0.42*)
CT score	255 / 0.42 (125 / *0.55*)	255 / 0.42 (125 / *0.54*)	255 (125)	255 / 0.83 (125 / *0.88*)	266 / 0.45 (128 / 0.50)	255 / 0.29 (125 / 0.33)
CT fracture character	252 / 0.46 (122 / *0.57*)	252 / 0.45 (122 / *0.56*)	252 / 0.84 (122 / *0.90*)	252 (122)	263 / 0.48 (125 / *0.54*)	252 / 0.29 (122 / 0.32)
ECT	229 / 0.51 (105 / *0.62*)	229 / 0.51 (105 / *0.62*)	275 / 0.43 (127 / *0.50*)	275 / 0.45 (127 / *0.54*)	228 (105)	229 / 0.26 (105 / *0.32*)
Threshold sum	297 / 0.17 (116 / 0.22)	297 / 0.18 (116 / *0.23*)	502 / 0.15 (209 / *0.20*)	502 / 0.15 (209 / *0.19*)	457 / 0.13 (181 / *0.19*)	295 (181)

The nearly perfect agreement of the critical failure layers found with the RB test cannot be compared with the above values, because they do not result from two different tests, but from analysing two different test results from one and the same RB test. They do not represent the reproducibility.

Comparing whether the same critical failure layer was identified by two different types of tests at the same profile location, showed lower agreement (Table 6). In slightly more than half of the snow pits (51%), the RB and the ECT identified the same layer as the critical failure layer. Comparing CT score or CT fracture character with the RB or the ECT revealed a slightly lower agreement (42-48%). With agreement scores of 13% to 33%, the threshold sum only relatively rarely identified the same layer as the critical failure layer that was found with the other tests. As often several critical failure layers are identified with the threshold sum the agreement depends on direction of comparison.

Analysing two different test results from the same stability test showed good agreement of the critical failure layers: 99% for the RB test, 84% for the CT. However, these values are not independent from each other. Except for the comparison of the RB score with the RB release type, the agreement of the critical failure layers was always higher between tests performed on either unstable slopes or stable slopes that were rated “fair”, than on stable slopes rated as “good”; the differences were in most cases significant.

4.4 *Ease of use*

For operational use, a stability test has to be easy to perform, and it has to be easy and unambiguous to observe the results. We have not systematically assessed the ease of use, for example, by a questionnaire, but simply report our subjective assessment. In regard to ease of use, the ECT lies in between the RB release type (which is the most simple observation) and the CT fracture character, which some observers had difficulties with. In regard to the time required to do the test, the sequence is the same but in different order: CT, ECT then RB, which uses most time.

5. CONCLUSIONS

We compared various stability tests and indicators of snow instability, in particular to assess the performance of the extended column test (ECT) that has recently been developed by Simenhois and Birkeland (2006, 2007). The data contained 146 sets of various tests performed side by side on potential avalanche slopes above tree line in the Swiss Alps during the winter of 2007-2008.

Based on our limited dataset it has been shown that the ECT was able to well differentiate stable from unstable slopes. By reducing the number of loading-taps to 21, the number of false alarms was slightly reduced, so that the specificity and the sensitivity were 82% and 83%, respectively. This means that the portion of false alarms and false stable prediction was similar. The performance was clearly better than for the CT, which is bothered with a low specificity. The unweighted average accuracy was about 80% comparable to the performance of the RB test. However, the stability classification used for analysis, was partly based on the RB test.

When two different types of stability tests were performed adjacent to each other, in about half of the cases the tests identified the same critical failure layer. On slopes with rather unstable conditions, where prominent weak layers are more frequently expected, the agreement between the tests was higher. Higher agreement was also obtained between the same tests compared to the agreement between different types of tests. The relative low agreement scores represent a challenge for any method aiming at automatically identifying potential failure layers in a snowpack without relying on stability test results.

As has been shown previously for the case of the RB test (e.g. Jamieson and Johnston, 1993), the stability assessment became more reliable when results from two adjacent ECTs were combined. For our dataset, it was possible to classify 87% of the slopes with accuracy of about 90%.

So far the ECT does only discriminate between rather stable and rather unstable conditions based on whether a fracture propagates fast across the entire column. For operational use, the introduction of an intermediate stability class would be useful, which calls for further work.

In the terms of ease of use, the ECT did not pose a problem though requires from the observer some more skills than the RB test. As the ECT is done faster as the RB test, two ECTs can

easily be done in the same time. Further work has to show whether this obvious advantage balances the above-mentioned lack of detail in the test result.

Finally, snow slope stability evaluation should never rely on the result of a single test whether it is the ECT or any other stability test, but for best results all available information on instability has to be combined.

ACKNOWLEDGEMENTS

We thank Frank Techel, Martin Oberhammer, Jean-Luc Lugon, Peter Diener, Daniele Degiorgi, Pius Henzen, Martin Hepting, Giovanni Kappenberger, Jörg Kindschi, Luca Silvanti, Giorgo Valenti and numerous colleagues from SLF for snowpack observations, and Christine Pielmeier additionally for valuable inputs.

REFERENCES

CAA, 2002. *Observation guidelines and recording standards for weather, snowpack and avalanches*. Canadian Avalanche Association (CAA), Revelstoke BC, Canada, 78 pp

Föhn, P.M.B., 1987: The Rutschblock as a practical tool for slope stability evaluation. *Symposium at Davos 1986 - Avalanche Formation, Movement and Effects, IAHS Publ., 162*, International Association of Hydrological Sciences, Wallingford, Oxfordshire, U.K.: 223-228

Gauthier, D. and B. Jamieson, 2008: Evaluation of a prototype field test for fracture and failure propagation propensity in weak snowpack layers. *Cold Reg. Sci. Technol.*, **51**(2-3): 87-97

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

Jamieson, J.B., 1999: The compression test - after 25 years. *The Avalanche Review*, **18**(1): 10-12

Jamieson, J.B. and C.D. Johnston, 1995: Monitoring a shear frame stability index and skier-triggered slab avalanches involving persistent snowpack weaknesses. *Proceedings ISSW 1994, International Snow Science Workshop, Snowbird, Utah, U.S.A., 30 October-3 November 1994*, 14-21.

Jamieson, J.B. and C.D. Johnston, 1993: Rutschblock precision, technique variations and limitations. *J. Glaciol.*, **39**(133): 666-674

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*: 477-481

Pielmeier C., and Marshall H.P. (2008). Estimating Rutschblock stability from SnowMicroPen Measurements. *Proceedings ISSW 2008. International Snow Science Workshop, Whistler, U.S.A., 21-27 September 2008*, this issue.

Schweizer, J. and Jamieson, J.B., 2007. A threshold sum approach to stability evaluation of manual snow profiles. Cold Reg. Sci. Technol., **47**(1-2): 50-59

Schweizer, J., I. McCammon, and J.B. Jamieson, 2008: Snowpack observations and fracture concepts for skier-triggering of dry-snow slab avalanches. *Cold Reg. Sci. Technol.*, **51**(2-3): 112-121

Schweizer J. and Wiesinger T., 2001. Snow profile interpretation for stability evaluation. Cold Reg. Sci. Techol. **33**: 179-188

Simenhois, R. and Birkeland K.W., 2006. The extended column test: A field test for fracture initiation and propagation. In: Gleason J.A. (Editor), *Proceedings ISSW 2006, International Snow Science Workshop, Telluride CO, U.S.A., 1-6 October 2006*: 79-85.

Simenhois, R. and Birkeland K.W., 2007. An upgrade on the extended column test: New recording standards and additional data analyses. *The Avalanche Review*, **26**(2).

SYSTAT, 2007: *SYSTAT® Statistical Software User Manual*, San Jose CA, U.S.A.

van Herwijnen, A. and Jamieson B., 2007. Fracture character in compression tests. Cold Reg. Sci. Technol. **47**(1-2): 60-68

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