

# APPLICATION OF FIELD STABILITY EVALUATION METHODS TO THE SNOW CONDITIONS OF THE EASTERN PYRENEES.

Ivan Moner<sup>1\*</sup>, Jordi Gavalda<sup>1</sup>, Montserrat Bacardit<sup>2</sup>, Carles Garcia<sup>3</sup>, Glòria Martí<sup>3</sup>

<sup>1</sup> Consell Generau d'Aran (CGA), Val D'Aran, Spain; <sup>2</sup> Centre de Recerca d'Alta Muntanya, Val d'Aran, Spain; <sup>3</sup> Institut Geològic de Catalunya (IGC), Barcelona, Spain

**ABSTRACT:** Field methods for stability evaluation were recently developed and validated on the basis of skier-triggered slopes data from Switzerland and North America. In this paper, some of these methods (structural stability indexes, shear quality, and extended column test) were tested and adapted when necessary to the snow conditions of the Eastern Pyrenees. This was done by analyzing a dataset of 86 snow profiles and stability tests (from the last 8 winter seasons) observed next to skier-triggered slabs (including remotely triggered slab avalanches and whumpfs) and on skier-tested slopes that did not release a slab avalanche. ECT and shear quality were validated, whereas the profile field analysis methods had to be revised and new thresholds which provide a better adjustment for this part of the Pyrenees were established.

**KEYWORDS:** slab avalanche, snow stability, stability evaluation, skier triggering, Pyrenees

## 1. INTRODUCTION

In the Pyrenees, as in most countries of Europe and North America, skier-triggered dry slabs are responsible for most of the deaths by avalanche (83% of all avalanche accidents in Catalonia, IGC Avalanche Statistics).

In the past few years, a better understanding of the skier-triggering mechanisms of a dry slab avalanche has allowed an important progress in the tools available for the evaluation of the snow stability. McCammon and Schweizer (2002), Schweizer and Jamieson (2003, 2007) and Schweizer et al (2004) have suggested objective methods for the interpretation of manual snow profiles –lemons and yellow flags- on the basis of extensive datasets taken in the Swiss Alps and North America. They are the so-called structural stability indexes and they result from the statistical comparison between profiles made on slopes where skier-triggered slab avalanches had been released with profiles made on skier-tested slopes which did not release. Johnson and Birkeland (2002) and van Herwijnen and Jamieson (2002) set forth two new variables to be taken into account when performing stability tests: shear quality and fracture character. Both of them are concerned with the tendency of the

surrounding snowpack to concentrate shear stresses at the weak layer or interface and to propagate a shear fracture along that layer or interface. During the last ISSW, Simenhois and Birkeland (2006) presented the extended column test. Based upon a large dataset –though limited in its geographical and temporal scope-, the results of this new test show a promising correlation with stability.

All these methods have been developed and validated on the basis of snow profiles and stability tests taken in mountain ranges which are colder and more northern than the Pyrenees. This paper deals with the testing and adaptation of the said snow stability evaluation methods to the particular snow climate of the Eastern Pyrenees (e.g. Oller et al. 2006).

## 2. DATA

We used data from 43 snow pits made near the fracture line of or on slopes adjacent to skier-triggered avalanches, including whumpfs and remotely triggered slabs. These were called “unstable” profiles and tests, and they were compared to a “stable” dataset, collected in skier-tested slopes that did not release. In approximately half of the pits, stability tests were carried out in addition to snow profiles. For unstable tests we have: 27 CT/RB scores, 18 ECT scores and 23 shear quality / fracture character scores.

The data were collected in the Eastern Pyrenees, mainly in the Aran Valley, Catalonia, Spain (see Fig. 1) from the winter 2000-2001 to 2007-2008 (more than 50% of the profiles were taken between 2004-2008). All snow pits were made within the first 24 hours after the release,

---

*Corresponding author address:* Ivan Moner.  
Conselh Generau d'Aran, Vielha, Spain;  
25530; Tel./fax: +34 973 641 801; email:  
[i.moner@aran.org](mailto:i.moner@aran.org)

some of them a few minutes after it. 75% of the data were taken by the first two authors.

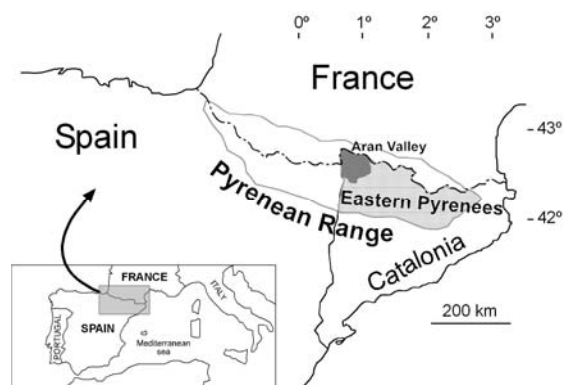


Figure 1: Study area.

In the case of the unstable sample, the failure layer was known, and in the case of the stable sample, we used the failure layer provided by rutschblock (RB), extended column test (ECT), compression test (CT) or shovel test (ST), in this order. We should point out that a weak layer –sometimes as thin as a few millimeters- was found in all the profiles except two (less than 5% of the cases).

### 3. METHODS

The following variables were analyzed: failure interface (FI) depth, failure layer (FL) thickness, FL grain type, FL grain size, FL hardness, adjacent layer (AL) grain type, AL grain size, AL hardness, grain size difference, hardness difference, RB/CT score, ECT score, shear quality / fracture character.

The variables were obtained by standard snowpack observations as described in Colbeck et al (1990). For grain size we used the average value in mm. In those cases where the grain size of a crust was unknown we tried two methods: using 1 mm as proposed by Jamieson and Schweizer (2005), or assigning a value of 0 mm. Usually, when the grain size is not measured it is because the grains are so bonded together that no separate crystals are distinguishable. Since grain size could be an indirect measure of bonding (the more grains per unit of volume, the more points of contact and, therefore, the more bonds), we thought that assigning a value of 0 mm to these layers could improve the classification power of this variable.

We used the hand hardness index from 1 to 5 for fist (F), four-finger (4F), one-finger (1F), pencil (P) and knife (K).

Intermediate values were allowed (e.g. 2-3 → 2.5) (e.g.. F+ → 1.3; P- → 3.7).

For grain type analysis, primary grain type was considered. Grains were classified as persistent (facets, depth hoar, surface hoar) and non-persistent (the rest) as proposed by Jamieson and Johnson (1995). We also tried to include decomposing forms –a grain type frequently associated to weak layers in the Pyrenees (9% of unstable cases)- into the first category.

Differences in average grain size and hardness were calculated between the failure layer and the adjacent layer, i.e. across the failure interface. When the failure interface was known, we chose as failure layer the one with the lowest hand hardness. When there was no difference in hardness, we chose the one with the highest average grain size. When none of these values differed, the lowest layer was chosen. Whenever the failure layer was known but not the failure interface (as frequently occurs with very thin FL), we first took into account the hardness differences, and then the average grain size differences in order to define the failure interface.

Rutschblock tests were performed as described in Schweizer (2002); compression tests as in Jamieson and Johnson (1996); extended column tests as in Simenhois and Birkeland (2006, 2007); shear quality as in Johnson and Birkeland (2002); fracture character as in van Herwijnen and Jamieson (2004). CT scores were converted into comparable RB scores as in Schweizer and Jamieson (2003). Similarly, fracture character results were converted into their shear quality equivalent (van Herwijnen and Jamieson 2003)

Data analysis was performed using Statistica 6.0 software package. For each variable, the differences between stable and unstable datasets were examined first. In order to decide whether two distributions were different or not, non-parametric tests were used, since they are independent from population distributions. We considered  $p < 0.05$  to indicate a significant difference between parameter distributions. Mann-Whitney U-test was used to compare samples of ordinal data, while the nominal variables distributions (grain type, shear quality, test scores) were compared by cross-tabulating the data and calculating the Pearson  $\chi^2$  statistic.

The significant stratigraphic variables were then used as predictor variables in a classification tree analysis, in order to establish the threshold value that best divided the stable and unstable populations.

Stability test scores and associated parameters were not included in the multivariate analysis. As shown in Schweizer and Jamieson (2003), including rutschblock test scores in the

threshold sum functions does not improve significantly the results of the classification. Therefore, treating them separately proves to be more useful. Following the said authors, in order to obtain a simple and robust field profile analysis method, a binary threshold function was established by using the significant variables. The outcomes (0 or 1 for each classifier) were then combined by using an unweighted sum. This sum results in an index of stability ranging from 0 to 6, from less to more unstable. In order to divide the unstable and stable population, a split value was then determined by applying the classification tree method to the threshold sum.

To describe the performance of the proposed methods, we applied the measures for categorical forecast used by Schweizer et al (2004): accuracy (or probability of correct forecast: PCF), probability of detection: POD, false alarm rate: FAR, and true skill score: TSS. With the definitions showed in Table 1, they are calculated as follows:

$$\text{Probability of correct forecast PCF} = \frac{a + d}{n}$$

$$\text{Probability of detection POD} = \frac{d}{b + d}$$

$$\text{False alarm rate FAR} = \frac{c}{c + d}$$

$$\text{True skill score TSS} = \text{POD} - \text{FAR}$$

Table 1: Contingency table  
(Total number of cases  $n = a + b + c + d$ )

		Observed	
		Stable	Unstable
Forecasted	Stable	a: correct stables	b: False stables
	Unstable	c: False alarms	d: hits

#### 4. RESULTS

The stable/unstable samples are first compared in Table 2. We then report on the univariate analysis, choosing the stratigraphic variables with high significance ( $p < 0.005$ ) for the multivariate classification methods.

From the analysis of table, regarding FL grain type, we should point out that grouping decomposing forms with persistent grains increased the significance of the variable in one order of magnitude. Concerning grain size difference, assigning a grain size value of 0 mm to the crusts for which grain size was unknown also improved significance (reduced  $p$  from 0.001 to 0.0004).

Failure layer depth was the selected variable with lowest significance ( $p = 0.004$ ).

Table 2: Stable-unstable comparisons of snow profile variables and stability tests. Very highly significant variables ( $p < 0.005$ ) are given in bold. For variables in which two alternatives were tested, the one with a higher significance is marked in bold. For categorical variables, mode is indicated instead of median. p: persistent grains; np: non persistent grains; df: decomposing forms.

Variable	Stable		Unstable		$P$
	$N$	Median	$N$	Median	
<b>Failure interface depth (cm)</b>	43	26	43	38	0.004
Failure layer thickness (cm)	33	5	39	4	0.58
<b>Failure layer grain type</b>	p/np	p	43	p	0.04
<b>p+df/np</b>	43	p+df	43	p+df	0.003
<b>Failure layer grain size (mm)</b>	43	0.6	43	1.0	<0.0001
<b>Failure layer hardness</b>	43	2 (4F)	43	1 (F)	<0.0001
Adjacent layer grain type	p/np	Np	43	Np	0.60
	p+df/np	Np	43	Np	0.17
Adjacent layer grain size (mm)	Crusts=1	0.4	43	0.5	0.07
	Crusts=0	0.4	43	0.4	0.37
Adjacent layer hardness	43	2.7 (1F-)	43	3 (1F)	0.27
<b>Grain size difference (mm)</b>	Crusts=1	0.2	43	0.5	0.001
	<b>Crusts=0</b>	0.2	43	0.6	0.0004
<b>Hardness difference</b>	43	0.6	43	1.3	0.0008
<b>Rutschblock score</b>	36	5.5	27	2	<0.0001
<b>Extended column test score (P/N)</b>	29	N	18	P	<0.0001
<b>Shear quality</b>	33	Q2, Q3	23	Q1	<0.0001

Surprisingly, both the median and the mean of this variable were lower in the stable sample than in the unstable one (Fig. 2).

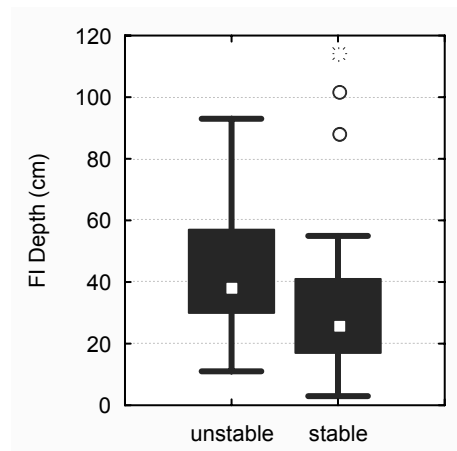


Figure 2: Failure layer depth for stable and unstable samples. Boxes span the interquartile range from 1st to 3th quartile with a white square showing the median. Whiskers show the range of observed values that fall within 1.5 times the interquartile range above and below the interquartile range.

#### 4.1 Establishing thresholds

The thresholds and critical ranges shown in Table 3 were obtained by applying

the classification tree methods to the variables, one by one. From within the stratigraphic variables, hardness difference is the one with the highest true skill score, followed by FL hardness and FL grain size.

A classification tree, taking into account the 6 significant stratigraphic variables with the threshold sum as a single independent variable, suggested a threshold sum of 5 as a split value. As shown in Figure 3, the threshold sum clearly discriminates stable profiles from unstable ones (Mann-Whitney  $U$ -test  $p < 0.0001$ ). The accuracy measures of this model are given in Table 4. When the failure interface depth was not included in the function, the method's true skill improved slightly, but the POD improved considerably. For this model, the classification tree suggests a split value of 4, while the non-parametric Mann-Whitney  $U$ -test shows a very high significance ( $p < 0.0001$ ), one order of magnitude higher than with the preceding method. As in Schweizer and Jamieson (2003), including RB in the threshold sum does not considerably improve the accuracy of the model. Including FI depth beside the threshold sum in the final classification tree does not improve accuracy either. Finally, the slight improvement attained by weighting the variables in the threshold sum does not justify the increase in the complexity of the method.

Table 3: Variable critical ranges and classification accuracy. p: persistent grains; df: decomposing forms;

Variable		Threshold, critical range	Probability of detection POD (%)	False alarm Rate FAR (%)	True skill score POD-FAR
Stratigraphic	Failure layer hardness	$\leq 1.5$	74	23	51
	Failure layer grain size (mm)	$\geq 0.7$	84	39	45
	Hardness difference	$> 1$	84	26	58
	Grain size difference (mm)	$\geq 0.5$	86	53	33
	Failure layer grain type	p + df	88	60	28
	Failure interface depth (cm)	23...97	84	46	38
Test	Rutschblock store	$< 5$	93	33	59
	Extended Column Test	P	89	3	85
	Shear Quality	Q1	78	8	70

Table 4: Classification accuracy of the models. FI: Failure interface

Model	Threshold	Probability of detection POD (%)	False alarm Rate FAR (%)	True skill score POD-FAR	Mann-Whitney $U$ -test $p$
6 variables	$\geq 5$	74.4	9.3	65.1	$< 0.0001$
5 var. without FI depth	$\geq 4$	81.4	14	67.4	$< 0.0001$
6 variables + RB	$\geq 6$	74.0	5.6	68.4	$< 0.0001$

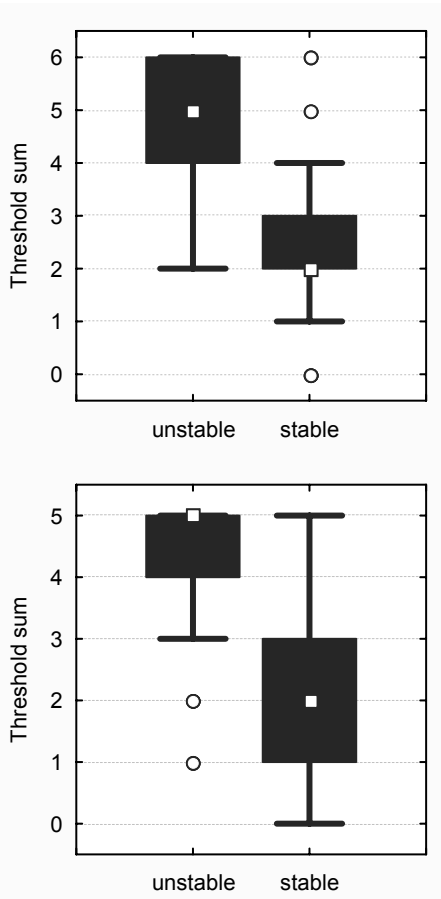


Figure 3: Box-plots showing the distributions of the 6 and 5 variable threshold sums for stable and unstable populations.

## 5. DISCUSSION

The thresholds and critical ranges obtained in this paper are compared to those proposed by Schweizer et al (2004) in Table 5. The values obtained for FL hardness and hardness difference are similar. FI depth cannot be compared, since it was obtained using different methods. The biggest differences were found in FL grain type, FL grain size and grain size difference.

The differences found could be explained, at least partially, by the fact that the Pyrenean sample presents almost no buried surface hoar. While this grain type is an important cause of instability in the Alps and the Columbian Mountains of Western Canada (e.g. Schweizer & Jamieson (2001), a buried surface hoar layer was found in only one case of human-triggered slab (2% of the cases) in the Pyrenees. These crystals, which are often quite big, cause the average grain size and the grain size difference to be higher in the samples taken in the Alps and the Canadian

Rockies than in the ones taken in the Pyrenees.

Regarding stability tests, RB/CT and shear quality/fracture character presented similar results to those obtained in other ranges by other authors. In the case of ECT, results were less accurate than those obtained by Simenhois and Birkeland (2006), but very similar to those published by the same authors in the December 2007 issue of *The Avalanche Review* (Simenhois and Birkeland, 2007). The two false stable cases found in our sample were associated with a very similar stratigraphy: a soft slab over what seemed to be a weak layer made up by graupel or small facets, then a melt-freeze crust not too tough (hardness P, thickness 3 cm), and finally a soft layer of very big deep hoar over the ground. In both cases, it was not clear which was the failure layer and we chose not to cut the crust when carrying out the test, since we considered it to be a hard layer. The result of the test would have probably been positive if we had cut the crust.

Table 5: Variable critical range for the Swiss and Canadian (n=427) (Schweizer et al 2004) and Pyrenean (n=86) datasets. FL: Failure layer; FI: Failure interface. df: decomposing forms.

Variable	Critical range	
	Swiss & Canadian dataset	Pyrenean Dataset
FL hardness	$\leq 1.3$	$\leq 1.5$
FL grain size (mm)	$\geq 1.25$	$\geq 0.7$
Hardness Difference	$\geq 1.7$	$> 1$
Grain size differ. (mm)	$\geq 0.75$	$\geq 0.5$
FL grain type	Persistent	Persist. + df
FI depth (cm)	18...94	23...97

### 5.1 Crusts

The above leads to a discussion on the effects of buried crusts on snow stability. In 16 of the unstable cases (39%), a crust was found next to the failure layer. On one hand, in 5 of these cases the crust was located below the weak layer, fact that might have caused its formation (Jamieson 2006) and the concentration of shear stress on it, contributing to the release (Habermann et al. 2008). On the other hand, in 11 cases (23%) the crust was located on top of the weak layer and it fell in the release. In some of these last cases, the triggered slab was so soft and thin, and the fracture propagation so wide, that it seems that not only did the presence of a weak crust not prevent the release, but it facilitated

it by helping to propagate the fracture in a very weak slab. Nonetheless, bearing in mind that our sample is not very big, the above should be considered strictly a qualitative assertion.

It is as vital as it is difficult to establish the boundary between a crust that may collapse on top of a weak layer and one which can stabilize the snow that lays underneath. Most of the crusts that failed were thin and hard, and they were located on top of layers of very developed deep hoar over the ground. Bridging, defined as the index obtained by multiplying layer thickness (in cm) by hand hardness (Schweizer and Jamieson 2003), gives a measure of the toughness of a layer. In our 11 study cases, the average bridging of the crusts that collapsed was 7, 15 being the highest.

## 6. CONCLUSIONS

The analyzed stability tests and their associated variables show similar results to those obtained in other snow climates. Rutschblock scores < 5 are correlated to unstable slopes, similarly to the results in Schweizer and Jamieson (2003) where the cut value was 4. Extended column test presents a high accuracy, both for false stable and false alarm results, similarly to Simenhois and Birkeland (2007). Q1 scores in shear quality or SC/SP scores in fracture character show a strong correlation with instability.

Structural stability indexes required more adaptation in order to produce good results in the snow climate of the Eastern Pyrenees. Once the new thresholds and critical ranges were established, the method's performance was satisfactory. The model which gives a better accuracy while keeping a level of simplicity that makes it operative is the unweighted threshold sum of 5 variables, without FI depth. Assigning a value of 0 mm to the crusts for which grain size is unknown improves the classification power of this variable. Adding decomposing forms to persistent grains makes the FL grain type reach an acceptable significance. In order to cover the variability of avalanche releases in the Pyrenees, data from more seasons and different areas is needed. These data would also help to improve our knowledge of the effect of crusts on snow stability.

## 7. ACKNOWLEDGEMENTS

Skier-triggered avalanches data was collected in collaboration with the Catalan

Pyrenees SAR services: Pompiers d'Aran and Bombers de la Generalitat de Catalunya. Snow pits were made by the authors and our colleagues of CGA and IGC, as well as by snow filed technicians: Maria Farré, Txema Arteta, Quim Merlos, Enric Nadal and Helena Serred. Neither would it have been possible to obtain our data without the help of the Ski-Patrollers of Baqueira-Beret Ski Resort and of the guards of the refuges of the Aran Valley.

## 8. REFERENCES

- Colbeck, S., Akiyama, E., Armstrong, R., Gubler, H., Lafeuille, J., Lied, K., McClung, D., Morris, E., 1990. International Classification for Seasonal Snow on the Ground. International Commission for Snow and Ice (IAHS), World Data Center for Glaciology, University of Colorado, Boulder, CO, USA.
- Colbeck, S., 2001. Sintering of unequal grains. Proceedings of the International Snow Science Workshop, Big Sky, Montana, USA. Oct. 2000.
- Habermann, M., Schweizer, J., Jamieson, B., 2008. Influence of snowpack layering on human-triggered snow slab avalanche release. Cold Regions Science and Technology, doi:10.1016/j.coldregions.2008.05.003
- van Herwijnen, A. and Jamieson, B., 2002. Interpreting fracture character in stability tests. Proceedings of the International Snow Science Workshop. Pencticton, British Columbia, Canada. Oct. 2002.
- van Herwijnen, A. and Jamieson, B., 2003. An update on fracture character in stability tests. Avalanche News 66: 26-28.
- van Herwijnen, A. and Jamieson, B., 2004. Fracture character in compression tests. Proceedings of the International Snow Science Workshop. Jackson Hole, Wyoming, U.S.A. Sept. 2004.
- van Herwijnen, A. and Jamieson, B., 2007. Snowpack properties associated with fracture initiation and propagation resulting in skier-triggered dry slab avalanches. Cold Regions Science and Technology: 50, 13-22.
- Institut Cartogràfic de Catalunya. 2006. Base de dades d'accidents per allau al Pirineu de Catalunya. [www.igc.cat/allaus](http://www.igc.cat/allaus)
- Jamieson, B and Johnston, C., 1995. Monitoring a shear frame stability index and skier-triggered slab avalanches involving persistent snowpack weaknesses. Proceedings International Snow Science Workshop, Snowbird, Utah, USA. 30 Oct. - 3 Nov. 1994.
- Jamieson, B and Johnson, C., 1996. The compression test for snow stability.

- Proceedings of the International Snow Science Workshop. Banff, Alberta, Canada. Oct. 1996.
- Jamieson, B., 1999. The compression test – after 25 years. *The Avalanche Review*, 18(1): 10-12.
- Jamieson, B. and Schweizer, J., 2005. Using a checklist to assess manual snow profiles. *Avalanche News* 72: 72-61.
- Jamieson, B., 2006. Formation of refrozen snowpack layers and their role in slab avalanche release. *Reviews of Geophysics*, 44, RG2001, doi: 10.1029/2005RG000176
- Johnson, R. and Birkeland, K., 2002. Integrating Shear Quality into stability test results. Proceedings of the International Snow Science Workshop. Pencticton, British Columbia, Canada. Oct. 2002.
- McCammon, I. and Schweizer, J., 2002. A field method for identifying structural weaknesses in the snowpack. Proceedings of the International Snow Science Workshop. Pencticton, British Columbia, Canada. Oct. 2002.
- McCammon, I. and Sharaf, D., 2005. Integrating strength, energy and structure into stability decisions. *The Avalanche Review*, 23 (3) 18–19.
- Moner, I., Gavalda, J., Martí, G. and García, C., 2006. Desencadenament accidental de plaques. Nous conceptes, recerca i aplicació al Pirineu. Proceedings of the 2a Jornada de Neu i Allaus. Barcelona, Spain. Jun. 2006.
- Moner, I. Gavalda, J. Martí, G., García, C., Bacardit, M., 2008. Aplicació al Pirineu dels nous mètodes d'avaluació de l'estabilitat. Proceedings of the 3a Jornada de Neu i Allaus. Betren, Val d'Aran, Spain. Jun. 2008.
- Simenhois, R. and Birkeland, K., 2006. The extended column test: a test for fracture initiation and propagation. Proceedings of the International Snow Science Workshop. Telluride, Colorado, USA. Oct. 2006.
- Simenhois, R. and Birkeland, K., 2007. An update on the Extended Colum Test. New recording standards and additional data analyses. *The Avalanche Review*: 26(2), 17-18.
- Schweizer, J. and Jamieson, B., 2001. Snow cover properties for skier-triggered avalanches. *Cold Regions Science and Technology* 33(2-3), 207-221.
- Schweizer, J., 2002. The rutschblock test. Procedures and applications in Switzerland. *The Avalanche Review*: 20(5), 1, 14-15.
- Schweizer, J. and B. Jamieson. 2007. A threshold sum approach to stability evaluation of manual snow profiles. *Cold Regions Science and Technology*: 47(1-2), 50-59.